

Predictive models for stillbirth and neonatal death: A global scoping review of model performance, validation, transportability and implications for sub-Saharan Africa

Tanida Chaweewong

London School of Hygiene & Tropical Medicine

Audencio Victor

London School of Hygiene & Tropical Medicine

Phillip Wanduru

Makerere University School of Public Health

Miriam Bourke

London School of Hygiene & Tropical Medicine

Peter Waiswa

Makerere University School of Public Health

Hannah Blencowe

London School of Hygiene & Tropical Medicine

Eric O. Ohuma

London School of Hygiene & Tropical Medicine

Joseph Akuze

`joseph.waiswa@lshtm.ac.uk`

London School of Hygiene & Tropical Medicine

Systematic Review

Keywords:

Posted Date: July 7th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10253093/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

Background: Each year 1.9 million stillbirths and 2.3 million neonatal deaths occur worldwide, and sub-Saharan Africa (SSA) accounts for 47% of stillbirths and 46% of neonatal deaths. Predictive models, including classical statistical methods, machine learning (ML) and artificial intelligence (AI), are increasingly used to detect pregnancies and neonates at high risk, but their performance and validation in SSA have not been mapped extensively. **Objective:** To map predictive models for stillbirths and neonatal deaths globally with a focus on SSA, and to characterise the predictors, data sources, modelling approaches, performance metrics and validation. **Methods:** We followed the Arksey and O'Malley framework and PRISMA-ScR guidance. We searched EMBASE, Medline, Scopus, Web of Science and Global Health from 2010 to 2025 plus reference lists. Two reviewers independently screened titles, abstracts and full texts. Data were charted across six domains: predictors, data sources, model type, evaluation metrics, performance and validation. **Results:** Of 1,339 records, 93 studies met the inclusion criteria; 26 (28.0%) were conducted in SSA and 67 (72.0%) elsewhere. Health facility data were the most common source (63/93, 67.7%). Classical methods alone were used in 44 studies (47.3%); ML and AI mostly appeared in hybrid combinations rather than replacing classical models. Internal validation alone was reported in 65 studies (69.9%), internal plus external in 16 (17.2%), external validation only in three (3.2%) and no validation of any kind in nine (9.7%). AUC was reported in 77 studies (82.8%), ranging from 0.610 to 0.996. Calibration-related metrics were reported in 12 studies (12.9%), four in SSA, and decision-curve analysis was reported in only one non-SSA study. **Conclusion:** Only one SSA study reported pragmatic clinical deployment. Most models lack external validation, calibration assessment and clinical-utility evidence for SSA. Future work should follow TRIPOD+AI and PROBAST+AI reporting standards, with pre-registered multi-site external validation in independent African cohorts.

INTRODUCTION

Around 1.9 million stillbirths and 2.3 million neonatal deaths occur each year, and sub-Saharan Africa (SSA) bears approximately 47% and 46% of this burden respectively [1–3]. The global stillbirth rate fell from 21.3 per 1,000 total births in 2000 to 13.9 in 2021 and has since stalled at around 14 per 1,000 in 2023, and the neonatal mortality rate from 31.0 per 1,000 live births in 2000 to 17.0 in 2023, with negligible change since 2020. Progress has lagged behind other child-mortality indicators and most SSA countries are off-track for the Sustainable Development Goals and Every Woman Every Newborn Everywhere (EWENE) 2030 targets [3–5].

The leading causes of stillbirth include obstetric complications, prolonged pregnancy, maternal infections such as malaria, syphilis and HIV, hypertensive disorders, diabetes and fetal growth restriction, alongside behavioural, social and environmental determinants [2, 6, 7]. Neonatal deaths are driven predominantly by prematurity (37%), birth asphyxia and trauma (23%), congenital anomalies (10%), pneumonia (8%) and sepsis (7%) [1] with respiratory distress syndrome, low birth weight, hypothermia and hypoglycaemia contributing further in SSA [8].

Structural barriers compound these clinical causes. Limited access to antenatal and intrapartum care, a high burden of infectious diseases, undernutrition and under-resourced health systems continue to constrain maternal, newborn and child health (MNCH) outcomes [2, 9]. The World Health Organization and national governments have responded by prioritising timely, high-quality care [10, 11]. In these settings, identifying at-risk pregnancies and neonates early is central to clinical decision-making and to directing scarce staffing, referral and monitoring capacity to where it will most reduce mortality; data science is increasingly positioned to support this by flagging elevated risk in time for intervention [12].

Over the past two decades, predictive modelling has been applied increasingly within MNCH to identify pregnancies and neonates at elevated risk [12]. Classical statistical methods, notably logistic regression and generalised estimating equations dominated earlier work, but regression-based approaches underperform when predictors interact non-linearly, when outcomes are rare, or when routine data are incomplete, all of which characterise stillbirth and neonatal death [13].

Machine learning (ML) and artificial intelligence (AI) have since broadened the inputs models can accommodate, from ultrasound imaging to longitudinal electronic and routine records, with modest but consistent gains in discrimination over logistic regression [14–16]. In Zambia, AI estimated gestational age accurately from blind ultrasound sweeps acquired by novice operators, extending basic antenatal screening [17].

Despite this rapid growth, AI applications in MNCH remain at an early stage, particularly in low resource settings, where both the volume of published work and the number of deployed models are a small fraction of those in high-income settings [12, 16, 18, 19]. Models developed in high-income settings often transfer poorly to SSA populations, not because the algorithms fail, but because case-mix, outcome prevalence, care pathways, data completeness and social determinants of risk differ substantially from the contexts in which they were trained [20–22]. Direct comparisons among classical, ML and AI approaches remain uncommon, and the SSA-specific landscape has not been systematically mapped. Without a consolidated view of what has been built, for whom, on what data, and how well it performs, the field risks re-implementing high-income designs that were never calibrated for the populations they are now being asked to serve. We therefore conducted a scoping review to map the global landscape of predictive models for stillbirth and neonatal death, with a specific focus on SSA and on the transportability of existing models to these settings, characterising the predictors, data sources, modelling approaches, performance metrics and validation practices used, and identifying where current evidence falls short of supporting safe translation into African clinical care.

METHODS

Study design

This scoping review followed the methodological framework proposed by Arksey and O'Malley [23] and is reported in line with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) [24]. The protocol was registered on the Open Science Framework (OSF) and published in Wellcome Open Research [25].

Research questions

Five research questions guided this review:

1. Which classical, ML and AI models have been used for predicting stillbirth and neonatal death?
 - a. Which predictive models for stillbirth and neonatal death have been transported or tested in SSA?
2. What predictors have been used in models for predicting stillbirth and neonatal death?
3. What data types and sources have been used to train these models?
4. What performance metrics (including balanced accuracy, sensitivity, specificity, AUC-ROC and F1-score) have been reported?
5. Which type of model (classical, ML, or AI) demonstrates the best predictive performance when comparable metrics are reported?

Eligibility criteria and search strategy

Search strategy: We searched five databases: EMBASE, Medline, Scopus, Web of Science and Global Health, with the final searches conducted on 10-Dec-2025 (Table S1). Searches were extended to reference lists of included studies. No separate grey-literature databases were searched; eligible preprints and grey literature were identified through these sources. Search terms were developed using the Population, Concept, Context (PCC) framework and the full database search strings are reported in Table S1 and our published protocol [25].

Outcome definitions

Stillbirth was defined as a baby born with no signs of life at ≥ 28 weeks of gestation or with a birth weight of $\geq 1,000$ grams[26]. Neonatal death was defined as death within the first 28 days of life among live births. Where included studies used different gestational-age thresholds, the definition applied in each study was recorded during data extraction.

Inclusion and exclusion criteria: Eligibility was pre-specified using the PCC framework (Table S1). The population comprised pregnant women or fetuses with outcomes of stillbirth or neonatal death. The concept was predictive modelling, including model development, validation or application, across three methodological approaches: classical statistical, ML and AI. The context comprised peer-reviewed studies, preprints and grey literature with accessible full text, published between 2010 and 2025. The

2010 start date was chosen a priori to capture the recent era in which machine-learning methods and routine electronic data became widely applied to perinatal outcome prediction. We excluded studies that did not apply a predictive modelling approach (such as descriptive or explanatory analyses), review study designs (scoping reviews, systematic reviews and meta-analyses) and conference abstracts without full-text access. Studies published in languages not accessible to the review team were excluded at retrieval; the team had capacity in English, Portuguese and Thai.

Study selection

Records retrieved from the five databases were imported into Rayyan [27] and duplicates removed. Two reviewers [TC and AV] independently screened titles and abstracts against the pre-specified eligibility criteria, with disagreements resolved by a third senior reviewer [JA]. The same reviewers then independently screened the full texts of candidate records. Data were extracted by one reviewer using a form based on the six predefined domains, with a second reviewer checking the extracted data, and a narrative synthesis was developed. A formal measure of inter-rater agreement (such as Cohen's kappa) was not calculated. This reflects standard scoping-review methodology, in which a formal inter-rater agreement coefficient is not required. Discrepancies were reconciled in consultation with the senior reviewer.

Data charting, collating, and synthesis

Modelling approach classification

Modelling approaches were classified as classical statistical, ML, or AI or combinations of these, following the definitions set out in the published protocol [25]. Classification was based on the analytical approaches used (predictor selection strategy, validation scheme and regularisation or tuning methods) rather than on the algorithm alone. The unit of analysis was the individual study; where a study reported more than one model, all models were recorded but the study remained the unit for classification and synthesis. Where a study applied more than one approach, all approaches were recorded, and each study was assigned to one of seven mutually exclusive family combinations (classical only; ML only; AI only; classical + ML; classical + AI; ML + AI; classical + ML + AI).

Data extraction: Each study's design and data source extracted: designs as retrospective or prospective cohorts, secondary analyses of survey or registry data, or other designs, and data sources as health-facility records (hospital or clinic data on maternity, delivery or newborn care), registry or population sources, multi-country pooled databases, or online data. Extracted data were organised using a six-domain conceptual framework (Fig. 1A) specified in the protocol[25] and covered (1) predictors and risk factors, categorised as maternal, foetal, clinical, sociodemographic or environmental; (2) data sources, including database type, geographical coverage and sample size; (3) modelling approaches and algorithm families; (4) evaluation metrics reported, including discrimination, calibration, classification and rare-outcome metrics; (5) reported model performance, including AUC or C-statistic values and comparisons across model families, regions and outcomes; and (6) validation practices, including validation type (internal, external, both or none), validation design (temporal split, hold-out, geographic, prospective) and transportability to, from or across SSA. Transportability was defined as the development, validation or application of a model across different geographical or population settings, classified by direction relative to SSA, that is into, out of, or across SSA settings.

Quality appraisal and synthesis

Reporting completeness was appraised using the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis + Artificial Intelligence (TRIPOD + AI) statement [28] and risk of bias was assessed using Prediction model Risk Of Bias Assessment Tool + Artificial Intelligence (PROBAST + AI) (Table S2) [29]. Findings were summarised descriptively and narratively across the six domains, with maps, graphs and tables used to illustrate patterns. Reporting followed the PRISMA-ScR checklist [24].

RESULTS

Study selection

The search yielded 1,339 records, of which 93 studies (26 from SSA and 67 from elsewhere) met the eligibility criteria and were retained for narrative synthesis. The PRISMA-ScR flow is shown in Fig. 1B.

Study characteristics, geographic and temporal distribution

Study characteristics are summarised in Table 1, and the full per-study inventory is presented in Tables 2 and S3 [30–122]. Most studies (78/93, 83.9%) were published from 2020 onwards (Fig. 2). Retrospective designs predominated (68/93, 73.1%), followed by prospective studies (13/93, 14.0%), secondary analyses (10/93, 10.8%) and other designs (2/93, 2.2%) (Tables 1–2 and S3).

Table 1
Study characteristics and geographic distribution (n = 93)

Characteristic	SSA (n = 26)	Non-SSA (n = 67)	Total (n = 93)
Study design			
Retrospective	18 (69.2%)	50 (74.6%)	68 (73.1%)
Prospective	5 (19.2%)	8 (11.9%)	13 (14.0%)
Secondary analysis	3 (11.5%)	7 (10.4%)	10 (10.8%)
Other	0 (0.0%)	2 (3.0%)	2 (2.2%)
Outcome studied			
Stillbirth only	9 (34.6%)	34 (50.7%)	43 (46.2%)
Neonatal death only	9 (34.6%)	21 (31.3%)	30 (32.3%)
Both stillbirth and neonatal death	8 (30.8%)	12 (17.9%)	20 (21.5%)
Data source			
Health facility data	17 (65.4%)	46 (68.7%)	63 (67.7%)
Registry / population	7 (26.9%)	17 (25.4%)	24 (25.8%)
International/ multi-country pooled database	2 (7.7%)	3 (4.5%)	5 (5.4%)
Online data	0 (0.0%)	1 (1.5%)	1 (1.1%)
Model family			
Classical only	11 (42.3%)	33 (49.3%)	44 (47.3%)
Classical + ML	4 (15.4%)	12 (17.9%)	16 (17.2%)
Classical + ML + AI	4 (15.4%)	10 (14.9%)	14 (15.1%)
ML only	5 (19.2%)	10 (14.9%)	15 (16.1%)
AI only	1 (3.8%)	1 (1.5%)	2 (2.2%)
ML + AI	1 (3.8%)	0 (0.0%)	1 (1.1%)
Classical + AI	0 (0.0%)	1 (1.5%)	1 (1.1%)
Validation approach			
Internal only	22 (84.6%)	43 (64.2%)	65 (69.9%)
External only	0 (0.0%)	3 (4.5%)	3 (3.2%)
Internal + external	4 (15.4%)	12 (17.9%)	16 (17.2%)
None reported	0 (0.0%)	9 (13.4%)	9 (9.7%)
Geographic distribution			
SSA single-country studies			
Ethiopia	5 (19.2%)	–	5 (5.4%)
Ghana	3 (11.5%)	–	3 (3.2%)
Summary of included studies stratified into two groups: SSA stratum (n = 26) comprising 20 single-country sub-Saharan African studies, 3 SSA multi-country (Africa-only) studies, and 3 multi-region studies that include SSA sites (Shukla 2020 Silva 2025 and Payne 2015); and Non-SSA stratum (n = 67) comprising 65 non-SSA single-country studies plus 2 multi-region studies without SSA sites (Allotey 2022 and Zhang 2023). All percentages use the stratum column total as denominator.			

Characteristic	SSA (n = 26)	Non-SSA (n = 67)	Total (n = 93)
Study design			
Rwanda	3 (11.5%)	–	3 (3.2%)
Tanzania	3 (11.5%)	–	3 (3.2%)
Malawi	1 (3.8%)	–	1 (1.1%)
Nigeria	2 (7.7%)	–	2 (2.2%)
Zimbabwe	1 (3.8%)	–	1 (1.1%)
South Africa	1 (3.8%)	–	1 (1.1%)
Uganda	1 (3.8%)	–	1 (1.1%)
SSA multi-country (Africa only)			
26 SSA countries (DHS pooled)	1 (3.8%)	–	1 (1.1%)
Tanzania and Liberia	1 (3.8%)	–	1 (1.1%)
Sub-Saharan Africa	1 (3.8%)	–	1 (1.1%)
Multi-region including SSA sites			
South Asia (India and Pakistan), Africa (Democratic Republic of Congo, Zambia, and Kenya) and Central America (Guatemala)	1 (3.8%)	–	1 (1.1%)
Democratic Republic of Congo, Guatemala, Zambia, India-Belagavi, India-Nagpur, Pakistan, Kenya and Argentina	1 (3.8%)	–	1 (1.1%)
Brazil, Fiji, Pakistan and South Africa (miniPIERS)	1 (3.8%)	–	1 (1.1%)
Non-SSA single country (by World Bank region)			
East Asia & Pacific	–	19 (28.4%)	19 (20.4%)
North America	–	14 (20.9%)	14 (15.1%)
Europe & Central Asia	–	17 (25.4%)	17 (18.3%)
Middle East & North Africa	–	7 (10.4%)	7 (7.5%)
Latin America & Caribbean	–	4 (6.0%)	4 (4.3%)
South Asia	–	4 (6.0%)	4 (4.3%)
Multi-region without SSA			
UK and USA (Allotey 2022)	–	1 (1.5%)	1 (1.1%)
United States, Puerto Rico, US Virgin Islands, Guam (Zhang 2023)	–	1 (1.5%)	1 (1.1%)
Summary of included studies stratified into two groups: SSA stratum (n = 26) comprising 20 single-country sub-Saharan African studies, 3 SSA multi-country (Africa-only) studies, and 3 multi-region studies that include SSA sites (Shukla 2020 Silva 2025 and Payne 2015); and Non-SSA stratum (n = 67) comprising 65 non-SSA single-country studies plus 2 multi-region studies without SSA sites (Allotey 2022 and Zhang 2023). All percentages use the stratum column total as denominator.			

Table 2
Detailed characteristics of all included studies (n = 93)

Author / Year	Design	Location	Outcome	Database / source	Classic	ML	AI	Model validation	Ref
Adane, et al. 2025	Retrospective study	Ethiopia	Neonatal mortality	Health facility-based data	✓	X	X	Internal validation only	[30]
Adebanji, et al. 2024	Retrospective study	Ghana	Stillbirth	Health facility-based data	✓	✓	✓	Internal validation only	[31]
Aga 2025	Retrospective study	Ethiopia	Stillbirth	Health facility-based data	X	✓	X	Internal validation only	[32]
Al Mashrafi et al. 2025	Retrospective study	Oman	Neonatal mortality	Population/ Registry	X	✓	X	Internal validation only	[33]
Al-Fattah, et al. 2024	Retrospective study	Indonesia	Stillbirth	Health facility-based data	✓	X	X	Internal validation only	[34]
Allotey, et al. 2022	Meta-analysis	UK and USA	Stillbirth	International database	✓	X	X	External validation only	[35]
Alzakari, et al. 2024	Secondary analysis	Saudi Arabia	Stillbirth	Health facility-based data	✓	✓	✓	Internal validation only	[36]
Åmarkl, et al. 2018	Retrospective study	Sweden	Stillbirth	Population/ Registry	✓	X	X	Internal validation only	[37]
Baker, et al. 2021	Retrospective study	United States	Neonatal mortality	Health facility-based data	X	X	✓	Internal validation only	[38]
Batista, et al. 2021	Retrospective study	Brazil	Neonatal mortality	Population/ Registry	✓	✓	✓	Both internal and external validation	[39]
Bogale D.S., et al. 2022	Secondary analysis	Ethiopia	Stillbirth	Population/ Registry	X	✓	X	Internal validation only	[40]
Bosschieter, et al 2023	Retrospective study	United States	Stillbirth	Health facility-based data	✓	✓	X	Both internal and external validation	[41]
Cavallaro, et al 2017	Prospective study	United Kingdom	Stillbirth	Health facility-based data	✓	X	X	Neither internal nor external validation	[42]
Cersonkey, et al. 2023	Secondary analysis	United States	Stillbirth	Population/ Registry	✓	✓	✓	Internal validation only	[43]
Chen, et al. 2022	Retrospective study	China	Neonatal mortality	Health facility-based data	X	✓	X	Internal validation only	[44]
Chiweza, et al. 2024	Retrospective study	Malawi	Stillbirth and early	Health facility-	X	X	✓	Internal validation	[45]

Author / Year	Design	Location	Outcome	Database / source	Classic	ML	AI	Model validation	Ref
			neonatal death	based data				only	
Choi, et al 2022	Retrospective study	South Korea	Infant mortality	Health facility-based data	✓	X	X	Internal validation only	[46]
Conte, et al. 2025	Retrospective study	Italy	Neonatal mortality	Health facility-based data	X	✓	X	Internal validation only	[47]
Cooper, et al 2018	Retrospective study	USA and Canada	Neonatal mortality	Health facility-based data	✓	✓	X	Both internal and external validation	[48]
Dalton, et al 2023	Retrospective study	United States	Perinatal death (stillbirth + early neonatal death)	Health facility-based data	✓	X	X	Neither internal nor external validation	[49]
Espinola-Sánchez, et al. 2023	Retrospective study	Peru	Neonatal mortality	Health facility-based data	X	✓	X	Internal validation only	[50]
Fazzari, et al. 2022	Secondary analysis	United States	Foetal death	Health facility-based data	✓	✓	X	Internal validation only	[51]
Flatley, et al. 2019	Prospective study	Australia	Perinatal death (stillbirth + early neonatal death)	Health facility-based data	✓	X	X	Internal validation only	[52]
Gatabazi, et al. 2020	Retrospective study	Rwanda	Neonatal mortality	Health facility-based data	✓	X	X	Internal validation only	[53]
Giorgione, et al. 2025	Retrospective study	UK, Italy, Belgium	Stillbirth	Health facility-based data	✓	X	X	Internal validation only	[54]
Gunenc, et al. 2025	Retrospective study	Turkey	Stillbirth	Health facility-based data	✓	✓	✓	Internal validation only	[55]
Hauptman, et al. 2025	A case-only study	USA	Neonatal mortality	Online data	✓	X	X	Neither internal nor external validation	[56]
Houweling, et al. 2019	Secondary analysis	Nepal	Neonatal mortality	Population/ Registry	✓	X	X	Internal validation only	[57]
Hromadnikova, et al 2024	Retrospective study	Czech Republic	Stillbirth	Health facility-based data	✓	X	X	Internal validation only	[58]
Hromadnikova, et al. 2023	Retrospective study	Czech Republic	Stillbirth	Health facility-based data	✓	X	X	Neither internal nor external validation	[59]

Author / Year	Design	Location	Outcome	Database / source	Classic	ML	AI	Model validation	Ref
Huang, et al. 2021	Retrospective study	USA	Neonatal mortality	Health facility-based data	✓	X	X	Internal validation only	[60]
Huang, et al. 2025	Retrospective study	China	Stillbirth	Health facility-based data	✓	X	X	Internal validation only	[61]
Inyang, et al 2021	Retrospective study	Nigeria	Stillbirth	Health facility-based data	X	✓	X	Internal validation only	[62]
Kaliappan, et al. 2023	Retrospective study	India	Stillbirth	Health facility-based data	X	✓	X	Internal validation only	[63]
Khatibi, et al. 2021	Retrospective study	Iran	Stillbirth	Population/ Registry	✓	✓	X	Internal validation only	[64]
Koivu, et al. 2020	Retrospective study	USA	Stillbirth	Population/ Registry	✓	✓	✓	Both internal and external validation	[65]
Kovacs, et al 2021	Prospective study	Tanzania	Neonatal mortality	Health facility-based data	✓	✓	X	Internal validation only	[66]
Kuanakewinyoo, et al 2025	Retrospective study	Thailand	Stillbirth	Population/ Registry	✓	✓	X	Internal validation only	[67]
Kubahoniyesu, et al. 2024	Retrospective study	Rwanda	Stillbirth	Population/ Registry	✓	✓	X	Internal validation only	[68]
Lee, et al. 2021	Retrospective study	USA	Infant mortality	Health facility-based data	X	✓	X	Internal validation only	[69]
Li, et al. 2025	Retrospective study	United Kingdom	Stillbirth	Health facility-based data	✓	✓	X	Both internal and external validation	[70]
Lin, et al. 2022	Retrospective study	China	Neonatal mortality	Health facility-based data	✓	X	X	Internal validation only	[71]
Olde Loohuis, et al 2025	Prospective study	Ghana	Perinatal death (stillbirth + early neonatal death)	Health facility-based data	✓	X	X	Both internal and external validation	[72]
Malacov, et al. 2020	Retrospective study	Australia	Stillbirth	Health facility-based data	✓	✓	✓	Internal validation only	[73]
Mastrodima, et al. 2016	Prospective study	United Kingdom	Stillbirth	Health facility-based data	✓	X	X	Neither internal nor external validation	[74]

Author / Year	Design	Location	Outcome	Database / source	Classic	ML	AI	Model validation	Ref
Mboya I.B., et al. 2020	Retrospective study	Tanzania	Perinatal death (stillbirth + early neonatal death)	Population/ Registry	✓	✓	✓	Internal validation only	[75]
McLachlan, et al. 2025	Retrospective study	United Kingdom	Stillbirth and neonatal mortality	Population/ Registry	X	✓	X	Internal validation only	[76]
Mello, et al. 2025	Retrospective study	Brazil	Stillbirth	Population/ Registry	X	✓	X	Internal validation only	[77]
Meng, et al. 2025	Retrospective study	China	Perinatal death (stillbirth + early neonatal death)	Health facility-based data	✓	✓	X	Both internal and external validation	[78]
Mengying, et al. 2024	Retrospective study	China	Early death (within 30 days of operation) in neonates with gastric perforation (NGP)	Health facility-based data	✓	X	X	Internal validation only	[79]
Mieghem, et al. 2022	Retrospective study	Canada	Foetal death	International database	✓	X	X	Internal validation only	[80]
Monfared, et al. 2023	Retrospective study	Iran	Neonatal mortality	Health facility-based data	✓	✓	X	Internal validation only	[81]
Narayan, et al. 2025	Case-control study	India	Foetal death	Health facility-based data	✓	✓	✓	Internal validation only	[82]
Ngusie, et al. 2024	Retrospective study	26 Sub-Saharan African countries	Stillbirth	Population/ Registry	✓	✓	✓	Internal validation only	[83]
Ngwenya, et al. 2021	Retrospective study	Zimbabwe	Stillbirth and neonatal mortality	Health facility-based data	✓	X	X	Both internal and external validation	[84]
Ognean, et al 2023	Retrospective study	Romania	Neonatal mortality	Health facility-based data	✓	X	X	Neither internal nor external validation	[85]
Oh, et al. 2025	Retrospective study	South Korea	Neonatal mortality	Health facility-based data	X	✓	X	Internal validation only	[86]
Paydar, et al. 2025	Retrospective study	Iran	Spontaneous abortion	Health facility-based data	✓	X	X	Internal validation only	[87]

Author / Year	Design	Location	Outcome	Database / source	Classic	ML	AI	Model validation	Ref
Pepler, et al. 2012	Retrospective study	South Africa	Neonatal mortality	Health facility-based data	✓	X	X	Both internal and external validation	[88]
Prins, et al. 2025	Secondary analysis	Netherlands	Stillbirth and neonatal mortality	Health facility-based data	✓	X	X	Internal validation only	[89]
Punyapet, et al. 2023	Retrospective study	Thailand	Perinatal death (stillbirth + early neonatal death)	Health facility-based data	✓	X	X	Neither internal nor external validation	[90]
Ramos, et al. 2025	Retrospective study	Brazil	Neonatal mortality	Population/ Registry	✓	✓	✓	Both internal and external validation	[91]
Rees, et al 2025	Secondary analysis	Tanzania and Liberia	Neonatal mortality	Health facility-based data	X	✓	✓	Internal validation only	[92]
Sayeed, et al. 2024	Retrospective study	Bangladesh	Neonatal mortality	Population/ Registry	✓	✓	X	Internal validation only	[93]
Shao, et al. 2024	Retrospective cohort study	China	Stillbirth	Health facility-based data	✓	X	X	Internal validation only	[94]
Sheikhtaheri, at al. 2021	Prospective study	Iran	Neonatal mortality	Population/ Registry	✓	✓	✓	Both internal and external validation	[95]
Shukla, et al. 2020	Prospective study	South Asia (India and Pakistan), Africa (Democratic Republic of Congo, Zambia, and Kenya), and Latin America (Guatemala)	Stillbirth and neonatal mortality	International database	X	✓	X	Internal validation only	[96]
Silva, et al 2025	Retrospective study	Democratic Republic of Congo, Guatemala, Zambia, India-Belagavi, India-Nagpur, Pakistan, Kenya and Argentina	Neonatal mortality	International database	✓	✓	✓	Internal validation only	[97]
Sylvain, et al. 2025	Retrospective study	Rwanda	Stillbirth	Health facility-based data	✓	✓	X	Internal validation only	[98]

Author / Year	Design	Location	Outcome	Database / source	Classic	ML	AI	Model validation	Ref
Takramah, et al. 2021	Retrospective study	Ghana	Neonatal mortality	Population/ Registry	✓	X	X	Internal validation only	[99]
Tesfie, et al. 2024	Retrospective study	Ethiopia	Neonatal mortality	Health facility-based data	✓	X	X	Internal validation only	[100]
Thai, et al. 2025	Retrospective study	United States	Gestational age at stillbirth	Population/ Registry	✓	✓	X	Both internal and external validation	[101]
Triggs, et al. 2025	Retrospective study	Australia	Stillbirth	Population/ Registry	✓	X	X	Internal validation only	[102]
Wang, et al. 2024	Retrospective study	China	Stillbirth	Health facility-based data	✓	✓	X	Internal validation only	[103]
Wu, et al. 2019	Retrospective study	Shanghai	Stillbirth and neonatal mortality	Health facility-based data	✓	X	X	Internal validation only	[104]
Yehuala, et al. 2025	Retrospective study	Sub-Saharan Africa countries	Stillbirth	Population/ Registry	X	✓	X	Internal validation only	[105]
Yerlikya, et al. 2016	Prospective study	United Kingdom	Stillbirth	Health facility-based data	✓	X	X	Neither internal nor external validation	[106]
Yin, et al. 2024	Retrospective study	China	Intrauterine Foetal Demise	Health facility-based data	✓	X	X	Both internal and external validation	[107]
Young, et al. 2023	Retrospective study	Canada, Austria and Germany	Stillbirth	International database	X	✓	X	Internal validation only	[108]
Zhang, et al. 2023	Retrospective study	states of the United States, Puerto Rico, the Virgin Islands, and Guam	Neonatal mortality	Population/ Registry	✓	✓	✓	Both internal and external validation	[109]
Zhao, et al. 2025	Prospective study	Shanghai	Perinatal death (stillbirth + early neonatal death)	Health facility-based data	✓	X	X	Internal validation only	[110]
Zheng, et al. 2022	Retrospective study	China	Stillbirth and neonatal mortality	Health facility-based data	✓	✓	X	Internal validation only	[111]

Author / Year	Design	Location	Outcome	Database / source	Classic	ML	AI	Model validation	Ref
Zimmerman, et al. 2025	Secondary analysis	United States	Perinatal death (stillbirth + early neonatal death)	Health facility-based data	✓	X	✓	Both internal and external validation	[112]
Mengstie, et al 2025	Retrospective study	Ethiopia	Neonatal mortality	Single country, population level data	✓	✓	X	Internal validation only	[113]
Akolekar, et al. 2011	Prospective screening study	United Kingdom	Stillbirth (and miscarriage)	Health facility-based data	✓	X	X	Internal validation only	[114]
Akolekar, et al. 2016	Prospective screening study	United Kingdom	Stillbirth	Health facility-based data	✓	X	X	Neither internal nor external validation	[115]
Awor, et al. 2022	Prospective cohort study	Uganda	Stillbirth	Health facility-based data	✓	X	X	Internal validation only	[116]
Trudell, et al. 2017	Retrospective cohort study	United States	Stillbirth	Health facility-based data	✓	X	X	Internal validation only	[117]
Housseine, et al. 2022	Prospective cohort study	Tanzania	Perinatal death (stillbirth + neonatal death)	Health facility-based data	✓	X	X	Both internal and external validation	[118]
Kayode, et al. 2016	Retrospective cohort study	Nigeria	Stillbirth	Health facility-based data	✓	X	X	Internal validation only	[119]
Muin, et al. 2022	Retrospective case-control study	Austria	Stillbirth (intrauterine foetal death)	Health facility-based data	✓	X	X	External validation only	[120]
Payne, et al. 2015	Secondary analysis	Brazil, Fiji, Pakistan and South Africa (miniPIERS, five low-resourced countries)	Perinatal death (stillbirth + neonatal death)	Health facility-based data	✓	X	X	Internal validation only	[121]
Schuit, et al. 2012	Retrospective registry-based study	Netherlands	Stillbirth and neonatal death (very preterm infants)	Population / Registry	✓	X	X	External validation only	[122]

Of the 26 (28.0%) studies from SSA, 20 were conducted wholly within a single SSA country, three were SSA multi-country analyses, and three were multi-region analyses that included SSA sites [96, 97, 121]. The 67 (72.0%) non-SSA studies comprised 65 single-country studies and two multi-region studies (without SSA) (Table 1; Figure S1). Within SSA, Ethiopia contributed the largest single-country share (5 studies) [30, 32, 40, 100, 113], with smaller contributions from Ghana [31, 72, 99], Rwanda [53, 68, 98], Tanzania [66, 75, 118], Malawi [45], Nigeria [62, 119], South Africa [88], Uganda [116] and Zimbabwe [84] (Table 1). Three studies pooled data across multiple African countries: two Demographic and Health Survey (DHS) studies [83, 105], and a facility cohort in Tanzania and Liberia [92] ; the three multi-region analyses combined African, South Asian and Latin American cohorts [96, 97, 121].

Among the 67 non-SSA studies, the largest clusters were East Asia and Pacific (19/67, 28.4%), Europe and Central Asia (17/67, 25.4%) and North America (14/67, 20.9%) (Table 1).

Stillbirth alone was the primary outcome in 43 studies (46.2%), neonatal death alone in 30 (32.3%), and a composite of stillbirth and neonatal death in 20 (21.5%) (Table 1).

Findings are presented across the six analytical domains in turn: predictors and risk factors (Domain I), data sources (II), modelling approaches (III), evaluation metrics (IV), model performance (V), and validation practices (VI). Figure 3 provides the cross-domain evidence map across World Bank regions.

Domain I: Predictors and risk factors

The frequency of reported predictors is summarised by category in Table 3 and by individual named predictor in Table S4. Regional patterns of predictor–model linkage are shown in Fig. 4.

Table 3
Predictor category reporting across outcomes (n = 93 studies)

Predictor category	All studies (n = 93)			Stillbirth only (n = 43)			Neonatal death only (n = 30)			Both SB + NND (n = 20)		
	SSA	Non-SSA	Total n (%)	SSA	Non-SSA	Total n (%)	SSA	Non-SSA	Total n (%)	SSA	Non-SSA	Total n (%)
Studies included (n)	26	67	93	9	34	43	9	21	30	8	12	20
Maternal	20	43	63 (67.7%)	9	27	36 (83.7%)	5	9	14 (46.7%)	6	7	13 (65.0%)
Foetal	15	39	54 (58.1%)	3	15	18 (41.9%)	8	15	23 (76.7%)	4	9	13 (65.0%)
Clinical	6	16	22 (23.7%)	2	11	13 (30.2%)	2	4	6 (20.0%)	2	1	3 (15.0%)
Sociodemographic	8	13	21 (22.6%)	3	10	13 (30.2%)	3	2	5 (16.7%)	2	1	3 (15.0%)
A study is counted under category C if and only if Table 2 displays at least one named predictor for that category in that study. Each study contributes at most one count per category. n (%) columns use the column denominator.												

Maternal factors were the most frequently reported predictor category (63/93, 67.7%), followed by fetal (54/93, 58.1%), clinical (22/93, 23.7%) and sociodemographic factors (21/93, 22.6%) (Table 3). Stillbirth-only models predominantly used maternal predictors (36/43, 83.7%), while neonatal-death-only models predominantly used fetal predictors (23/30, 76.7%) (Table 3). In composite stillbirth and neonatal death models, maternal and fetal predictors were equally common (13/20, 65.0% each). Clinical predictors were less common across outcome groups (13/43 stillbirth-only, 6/30 neonatal-death-only, 3/20 composite), and sociodemographic predictors were most frequent in stillbirth-only models (13/43, 30.2%). No study reported environmental predictors (such as air pollution, ambient temperature or seasonality), the fifth pre-specified category, which is therefore omitted from Table 3.

At the level of individual named predictors, gestational age and maternal age were the most frequently reported, in 31 of 93 studies each (Table S4). Birth weight was the next most frequent (n = 26 studies), followed by hypertensive disorders of pregnancy (n = 19), maternal body mass index (BMI; n = 18), parity (n = 16) and Apgar score (n = 12).

Predictors differed between strata (SSA and non-SSA). Maternal age was reported more than twice as often in SSA studies (14/26, 53.8%) as in non-SSA studies (17/67, 25.4%), and anaemia or haemoglobin in 6 of 26 (23.1%) SSA studies compared with 2 of 67 (3.0%) non-SSA studies (Table S4). The pattern differed for some predictors: maternal BMI was reported in a single SSA study (1/26, 3.8%) versus 17 non-SSA (25.4%), and hypertensive disorders of pregnancy in three SSA studies versus 16 non-SSA. Umbilical artery Doppler was not reported in any SSA study but featured in eight non-SSA.

DHS-based analyses drew on sociodemographic predictors such as wealth, residence, marital status and antenatal-care attendance [40, 83, 99, 105]. Facility cohorts contributed more clinical and foetal predictors, including gestational age, birth weight, Apgar score and perinatal asphyxia [30, 53, 66, 88, 92, 98, 100]. The single pragmatic intervention study in the region [45] used intrapartum cardiotocographic features rather than the demographic or antenatal predictors. Across the 26 SSA-stratum studies, 8 (30.8%) reported at least one sociodemographic predictor [30, 40, 68, 75, 99, 105, 113, 119], compared with 13 of 67 non-SSA studies (19.4%) (Table 3).

Domain II: Databases and data sources

Hospital and clinic records provided the primary data for 63 of the 93 studies (67.7%). Population and civil-registry sources accounted for 24 studies (25.8%), multi-country pooled databases for 5 (5.4%), and web-scraped online data for a single study (1.1%) (Table 1 and Fig. 3).

Facility-based designs predominated in both strata: 17 of 26 SSA studies (65.4%) and 46 of 67 non-SSA studies (68.7%) (Table 1). Registry/population use was similar in frequency (7/26 SSA versus 17/67 non-SSA) but differed in kind. Six of the seven SSA registry/population studies drew on Demographic and Health Survey (DHS) rounds, in single-country and pooled analyses [40, 68, 83, 99, 105, 113]. The seventh used a long-running Tanzanian medical birth registry at the Kilimanjaro Christian Medical Centre (KCMC)[75]. Non-SSA registry studies were built on national perinatal or civil vital-statistics systems; the largest, a United States analysis by Zhang and colleagues [109], assembled 22.7 million live births from 2016 to 2021.

The five international-database studies were heterogeneous. Two were the multi-region pooled cohorts with SSA sites [96, 97], mixing African, South Asian and Latin American populations. The remaining three linked individual-participant data without SSA coverage: the IPPIC (International Prediction of Pregnancy Complications) Network meta-analysis of 19 stillbirth cohorts [35], a Canadian multi-hospital foetal-medicine consortium [80] and a German perinatal database [108]. The single online-data study [56] scraped a United States paediatric mortality registry. No SSA study drew on online-only data.

DHS rounds provided nationally representative survey data re-analysed for stillbirth and pregnancy loss outcomes [40, 68, 83, 105], while facility cohorts contributed across all three outcome groups (Table 2).

Domain III: Modelling approaches

Algorithm frequencies by model family, outcome and region are summarised in Figure S2. Classical statistical methods appeared in 75 of 93 studies (80.6%), machine learning in 46 (49.5%) and artificial intelligence in 18 (19.4%) (Table 1). These three counts are not mutually exclusive: many studies applied more than one family within the same analysis.

The seven model family combinations are tabulated in Table 1. Classical methods alone accounted for the largest share at 44 studies (47.3%), typically logistic, Poisson or penalised regression with stepwise or lasso-based variable selection. Examples include Ethiopian facility cohorts [30, 100], a Zimbabwean pre-eclampsia model [84], a South African neonatal-mortality model with temporal validation [88], a Ghanaian DHS logistic-regression study[99] and earlier non-SSA work from the United Kingdom [42, 74], Sweden [37], Nepal [57] and Australia [52]. A further 31 studies (33.3%) combined classical methods with at least one newer approach: Classical + ML in 16 studies (17.2%), Classical + ML + AI in 14 (15.1%), and Classical + AI in one (1.1%). The remaining 18 studies (19.4%) used ML or AI methods without a classical arm: ML only in 15, AI only in two and ML combined with AI in one.

The distribution differed between SSA and non-SSA studies. The 26 SSA studies were distributed across the same three groups (classical-only, hybrid, ML or AI-only), with 11 (42.3%) classical-only [30, 53, 72, 84, 88, 99, 100, 116, 118, 119, 121], eight (30.8%) hybrid [31, 66, 68, 75, 83, 97, 98, 113] and seven (26.9%) using ML or AI without a classical arm [32, 40, 45, 62, 92, 96, 105]. The 67 non-SSA studies leaned more heavily on classical, with 33 (49.3%) classical-only, 23 (34.3%) hybrid and 11 (16.4%) ML or AI only.

Random forest and XGBoost dominated the SSA machine-learning work, across studies from Ethiopia[32, 40], Ghana [31], Nigeria [62] and Rwanda [68, 98], the 26-country DHS analysis [83], the pan-SSA DHS study[105] and the Tanzania–Liberia facility cohort[92]. Head-to-head benchmarking across six or more algorithms was reported in three SSA analyses [68, 83, 92]. Neural networks, support vector machines and boosting variants supplemented these ensembles in several studies [31, 62, 68, 83, 96, 98]. The only pragmatic SSA deployment was the Malawian PeriWatch Vigilance AI cardiotocograph system, which was embedded in a labour ward rather than developed as a research model [45].

The move away from classical-only modelling occurred almost entirely from 2020 onwards (Fig. 2). Of the fifteen pre-2020 studies, fourteen used classical methods alone [37, 42, 52, 57, 74, 88, 104, 106, 114, 115, 117, 119, 121, 122] and one combined classical with ML [48]. Among the 78 studies published from 2020 onwards, 30 were classical-only, 30 hybrid and 18 ML or AI only, so the classical-only share fell from 93% before 2020 to 38% afterwards.

Domain IV: Evaluation metrics reported

The distribution of performance metrics reported across the 93 included studies is summarised in Figure S3, with regional and outcome-specific patterns shown in Figure S4.

Discrimination metrics (AUC, AUROC, C-statistic or C-index) were the main focus of reporting, present in 77 of 93 studies (82.8%), 21 of 26 (80.8%) in the SSA stratum and 56 of 67 (83.6%) in the non-SSA stratum (Table S5). Reporting was uniform across outcome groups, at around 83% in each (stillbirth-only 36/43, neonatal-death-only 25/30, composite 16/20). Five of the SSA studies did not report a comparable discrimination metric; the underlying reasons are addressed in the Discussion [45, 53, 66, 72, 75].

Classification metrics were the next most commonly reported with sensitivity or recall in 37 studies (39.8%), accuracy in 29 (31.2%), specificity in 21 (22.6%), precision or positive predictive value in 14 (15.1%), and F1 score in seven (7.5%). Reporting was broadly similar across strata, with one exception: accuracy was reported by 10 of 26 SSA studies (38.5%) compared with 19 of 67 non-SSA studies (28.4%).

Calibration related metrics, assessed using calibration plots, calibration slopes, intercepts, Hosmer–Lemeshow tests or Brier score, were reported in 12 of 93 studies (12.9%): four SSA studies and eight non-SSA studies. Among SSA studies, three reported internal calibration evidence (Hosmer–Lemeshow or Brier score, [100]; calibration slope, [118]; calibration assessment, [119]) and one reported external calibration evidence [72]. The remaining eight studies that reported calibration were non-SSA [33, 35, 41, 46, 48, 104, 110, 122]. Area under the precision-recall curve was reported in two non-SSA studies [33, 38], the Matthews correlation coefficient in one [36], and decision-curve analysis or net-benefit in one further non-SSA study [110]. No SSA study reported AUPRC, Matthews correlation coefficient or decision-curve analysis.

Domain V: Model performance

Full distribution statistics of model discrimination are given in Table S5, with a summary by model family, region and outcome in Table S6, and visualisation in Fig. 5. A discrimination metric (AUC, AUROC, or C-statistic or C-index) was reported for 77 of 93 studies (82.8%); the remaining 16 studies reported only other metrics (AUPRC, calibration, hazard ratios, pre- and post-intervention incidence, or classification thresholds) and are not represented in Fig. 5 or Table S6. Across these studies, AUCs ranged from 0.610 to 0.996, with a median of 0.860 (IQR 0.800 to 0.950). The highest AUCs were concentrated in single-cohort facility studies (Table S6).

Performance was broadly comparable between strata. Of the 77 studies reporting discrimination, 21 were in the SSA stratum (median 0.851, IQR 0.816 to 0.950, range 0.750 to 0.996) and 56 in the non-SSA stratum (median 0.863, IQR 0.794 to 0.935, range 0.610 to 0.992) (Table S5). The lowest values came from a non-SSA external-validation study of three published stillbirth models [35], where summary C-statistics ranged from 0.530 to 0.650 and the authors concluded the models had little to no clinical utility. The highest values, clustered at or just below 0.996, came from a mix of SSA and non-SSA cohorts described below.

A modest gradient across model families is visible in Table S6. Across the 77 discrimination-reporting studies, 39 used classical methods only, 25 combined classical methods with ML or AI (Classical + ML, Classical + ML+AI or Classical + AI), and 13 used ML, AI or ML + AI without a classical arm. Median AUCs were 0.828 (IQR 0.749 to 0.893) in classical-only studies, 0.900 (IQR 0.830 to 0.967) in hybrid studies, and 0.940 in studies without a classical arm. Distributions overlapped substantially. Notable classical exemplars included the Ethiopian preterm-sepsis nomogram (AUC 0.967) [100], Hromadnikova's combined early-stillbirth and miscarriage screen (0.992) [59], and Dalton's composite-outcome model (0.990) [49].

Several SSA analyses sat at the upper end of the AUC distribution, reaching AUROCs of 0.950 to 0.996: Rees's Tanzania-Liberia XGBoost neonatal model (0.990) [92], Nigerian and Ethiopian machine-learning stillbirth models (0.950 to 0.982) [32, 40, 62, 83], a 26-country DHS-pooled model (0.950) [32, 83], and Mengstie's Ethiopian CatBoost neonatal-mortality model reported the highest

SSA value (0.996) [113]. A parsimonious Zimbabwean logistic-regression model performed comparably 0.902 (95% CI 0.876 to 0.927) [84]. The non-SSA best models were XGBoost models from Brazil (n = 1.1 million) r(0.971) [39], and the United States (0.980) [109]. Individual values are given in Table S5 and Fig. 5.

Five SSA studies could not be placed on the discrimination axis: the Malawian PeriWatch Vigilance cardiotocograph deployment reported absolute reductions in intrapartum stillbirth (0.39% to 0.02%) and early neonatal death (0.75% to 0.17%) rather than a discrimination curve [45]; a Rwandan multi-risk survival model reported hazard ratios only [53]; a Tanzanian NICU analysis noted that generalised linear models outperformed decision trees without a numeric AUC [66]; Olde Loohuis and colleagues reported calibration metrics alone for their Ghanaian external validation [72]; and the Tanzanian KCMC registry analysis noted no significant AUC differences between five candidate algorithms without stating the values [75].

Domain VI: Model validation practices and transportability

Validation practices across the corpus are summarised in Table 1, with per-study transfer routes for the 19 transportability-relevant studies (n = 6 involving SSA, n = 13 within non-SSA settings) detailed in Table S7 and visualised in Fig. 6.

Internal validation alone was the dominant practice, reported in 65 of 93 studies (69.9%); internal plus external validation was reported in 16 (17.2%), external validation only in three (3.2%), and no validation of any kind in nine (9.7%). External validation of any form was therefore documented in 19 studies (20.4%), while 74 studies (79.6%) relied on internal validation alone or did not validate.

The two strata diverged. Every SSA study was validated at least internally: 22 reported internal validation only (84.6%) and four reported both internal and external validation (15.4%). External-only and unvalidated studies were only found in the non-SSA stratum. Among 67 non-SSA studies, 43 reported internal validation only (64.2%), 12 both internal and external (17.9%), three external validations only (4.5%), and nine no validation (13.4%). The nine unvalidated studies were all non-SSA and all used classical methods only [42, 49, 56, 59, 74, 85, 90, 106, 115].

External validation was concentrated in classical and hybrid studies. Of the 18 studies using ML, AI, or ML + AI without a classical arm, none reported external validation; all 19 external validations came from studies using classical methods either alone (n = 8) or in combination with ML or AI (n = 11).

What counted as "external" varied widely in design. Among the 19 transportability-relevant studies (Table S7), four were temporal splits within the same cohort [39, 88, 91, 109]; five were within-country hold-out samples [41, 65, 78, 107, 112]; seven were cross-country validations against an external or international dataset [35, 70, 72, 84, 118, 120, 122]; one was a prospective validation in the same facility [95]; and two were multi-region pooled or temporal validations incorporating SSA sites [96, 97]. Cross-country validation was therefore reported in only seven of 93 studies (7.5%).

Transportability of predictive models to, from and across sub-Saharan Africa

Six of the 19 transportability-relevant studies in Table S7 (visualised in Fig. 6 Panel A) had itemised SSA development or validation cohorts; the other 13 transferred models within or between non-SSA settings (Fig. 6 Panel B). The SSA-involving studies showed a consistent pattern. High-income models transported into SSA calibrated and discriminated poorly: the Payne perinatal-death model in the Ghana SPOT cohort (calibration slope 0.08, c-statistic 0.52) [72] and the Kayode stillbirth model in Zanzibar (c-statistic 0.57), the latter re-derived locally into a 15-predictor model with an optimism-adjusted c-statistic of 0.78 [118]. A Zimbabwean pre-eclampsia model validated against the high-income PIERS dataset reached AUC 0.90 [84], and a South African and Namibian NICU model reached AUC 0.85 for mortality but with low positive predictive value[88] Two multi-region pooled analyses including SSA sites reported AUROCs of about 0.71 to 0.82, with country-specific models rarely outperforming the pooled model [96, 97].

Allotey 2022 [35], externally validated three published stillbirth prediction models (Smith 2007, Yerlikaya 2016 and Trudell 2017) using individual-participant data from 19 IPPIC cohorts (491,201 pregnant women); none of the 19 cohorts was based in sub-Saharan Africa, and summary C-statistics ranged from 0.53 to 0.65, summary calibration slopes from 0.40 to 0.88, and decision-curve analysis showed no net benefit over treat-all or treat-none strategies. The analysis is mapped in Fig. 6 Panel B as a UK evidence-hub loop.

Methodological quality and risk of bias

Per-study TRIPOD + AI reporting completeness and PROBAST + AI risk-of-bias assessments for all 93 included studies are presented in Tables S2-S3. Reporting completeness was moderate, with a median TRIPOD + AI score of 55.3% of applicable items fully reported (IQR 50.5 to 62.2%). Risk of bias was high: under PROBAST + AI, 73 of 93 studies (78.5%) were at high overall risk, 14 unclear and six low. This was almost always driven by the analysis domain (high in 66 of 93), reflecting small samples, limited handling of overfitting, and sparse calibration and external validation; the participant domain was low risk in most studies.

DISCUSSION

SSA carries roughly half the global stillbirth and neonatal-death burden [1–3], yet contributes just over a quarter of the predictive-model evidence: 26 of 93 included studies (Table 1). Of those 26 SSA studies, 22 (84.6%) reported only internal validation and four reported internal and external validation. This imbalance, rather than the absolute volume of models, is the central finding of the review: where SSA-developed models exist, they have rarely been tested outside the cohort in which they were built.

Predictor selection was shaped by data availability. Gestational age, birth weight, maternal age, body mass index, hypertensive disorders of pregnancy, parity, and Apgar score were the most prominent predictors. Umbilical artery Doppler was reported in eight non-SSA studies and in none of the SSA studies, mirroring the regional distribution of routine obstetric ultrasound.

Sociodemographic predictors appeared in eight of 26 SSA studies (Table 3); four of these drew on DHS surveys [40, 83, 99, 105], reflecting the prominence of DHS as a major data source in SSA.

The methodological shift since 2020 was marked: 78 of 93 included studies were published from 2020 onwards, 30 of which combined classical methods with ML or AI and 18 of which used ML or AI methods without a classical arm. More complex methods did not yield consistently better performance. The highest-AUCs were comparable across model families: 0.992 in a classical-only study (Hromadnikova 2023, Czech Republic [59]), 0.967 in another classical-only study (Tesfie 2024, Ethiopia [100]), 0.996 in a Classical + ML study (Mengstie 2025, Ethiopia [113]) and 0.990 in an ML + AI study (Rees 2025, Tanzania and Liberia [92]). Well-specified classical models matched or exceeded the performance of the most complex hybrids in several studies. This is consistent with Christodoulou and colleagues' finding that machine learning offers no consistent discrimination advantage over logistic regression once predictor structure, sample size and events-per-variable are accounted for [123].

Discrimination reporting was broadly similar across regions: 21 of 26 SSA studies (80.8%) and 56 of 67 non-SSA studies (83.6%) reported a discrimination metric. The five SSA studies that did not report a comparable discrimination metric had specific methodological reasons: the Malawian PeriWatch Vigilance deployment used pre- and post-intervention incidence [45], a Rwandan multi-risk survival model reported hazard ratios only [53], a Tanzanian NICU analysis compared models qualitatively without numeric AUCs [66], Olde Loohuis 2025 reported calibration metrics alone [72], and a Tanzanian KCMC registry analysis noted no statistically significant AUC differences without stating values [75]. The gap is therefore not in whether models report discrimination but in whether they report it together with calibration and external-validation evidence.

Calibration evidence was reported in only 12 of 93 studies (12.9%). Four of these were SSA studies, and only one [72] provided external calibration evidence; the other three reported internal calibration evidence only: a Hosmer–Lemeshow and Brier-score assessment of an Ethiopian neonatal-sepsis nomogram [100], an optimism-adjusted calibration slope in a Tanzanian model [118], and internal calibration of a Nigerian stillbirth model [119]. Without a calibration slope, intercept or plot from an external cohort, a prediction is not safely actionable in routine clinical use [28, 124, 125].

External validation was documented in only 19 of 93 (20.4%) studies (Table 1). The wider transportability evidence points in the same direction. The validation of a high-income hypertensive-pregnancy model in the Ghana SPOT cohort returned a calibration slope of 0.08 and a C-statistic of 0.52 [72], the kind of miscalibration that routine calibration reporting is designed to detect. The largest external validation of stillbirth models published to date (Allotey 2022) [35], using individual-patient data from 19 IPPIC cohorts (491,201 women), validated three published high-income models (Smith 2007 [126], Yerlikaya 2016 [106] and Trudell 2017 [117]) and reported summary C-statistics of 0.53 to 0.65 and no net benefit on decision-curve analysis. None of the 19 IPPIC cohorts was SSA-based. Models developed and validated entirely in high-income settings cannot be assumed to transfer cleanly to SSA populations, where prevalence, predictor distributions and care pathways differ.

The methodological infrastructure already exists. Wynants and colleagues' systematic review of COVID-19 prediction models found that the overwhelming majority were at high risk of bias, with poor reporting and weak external validation; this review documents the same pattern in the perinatal mortality field [127]. TRIPOD + AI (2024) requires calibration and external validation reporting alongside discrimination [28] and PROBAST + AI (2025) penalises single-cohort development without held-out validation [29].

Data infrastructure for developing and externally validating perinatal-mortality prediction models in SSA exists including DHS surveys at national scale [128, 129], multi-country pregnancy cohorts, trials and population-based studies such as PRECISE pregnancy cohort (n = 6,932 women) across The Gambia, Kenya and Mozambique [130]; the Alliance for Maternal and Newborn Health Improvement (n = 278,186 pregnancies) across 11 sites (eight countries in SSA and three South Asia) [131]; the Every Newborn-INDEPTH study (n = 69,176 women) across five INDEPTH sites [132]; and the ALERT trial perinatal e-registry (n = 130,000) births across 16 hospitals in Benin, Malawi, Tanzania and Uganda [133]. Many have not been used to develop and externally validate a stillbirth or neonatal-death prediction modelling that meets the TRIPOD + AI standards, and we are now using PRECISE, ALERT and EN-INDEPTH data to begin closing that gap.

Strengths and limitations

This review is, to our knowledge, the first to systematically stratify perinatal prediction models by SSA versus non-SSA across all six domains of predictors, data sources, modelling approaches, evaluation metrics, model performance and validation practices. The review searches five major databases, extracts to a six-domain framework agreed in advance, applies PRISMA-ScR reporting, and reports per-study transportability routes for the 19 transportability-relevant studies (Fig. 6 and Table S7). The review also presents per-study TRIPOD + AI reporting completeness and PROBAST + AI risk-of-bias assessments for all 93 included studies, which to our knowledge has not previously been done at this scale for perinatal prediction models.

Four limitations should be noted. First, head-to-head comparison of AUCs across model families and outcomes was constrained by small cell sizes once studies were stratified by region, family and outcome; AUPRC, calibration and decision-curve metrics were reported too sparsely to allow a parallel performance analysis. Second, the review reflects publication patterns rather than implementation. Only one study (the Malawian PeriWatch Vigilance deployment [45]) reported pragmatic deployment in an SSA labour ward, so the gap between published models and clinical use in SSA is one this review can frame but not answer. Third, definitions of stillbirth and neonatal death varied across the corpus (gestational-age cut-offs of ≥ 20 , ≥ 24 , ≥ 28 weeks and others), and outcome counts in this review are based on each study's own definition rather than a harmonised threshold. Fourth, although we conducted TRIPOD + AI and PROBAST + AI assessments, we did not pool risk-of-bias quantitatively. Pooled synthesis was outside the scope of this scoping review and should be the subject of a future systematic or scoping review.

Implications and recommendations

Three priorities follow for strengthening predictive modelling for stillbirth and neonatal death in SSA.

Priority 1

Model development in SSA should adopt TRIPOD + AI [28] reporting and report calibration alongside discrimination as a default. Routine perinatal data systems should also be strengthened, since data completeness and quality currently constrain model development in the region. The current calibration reporting rate of four of 26 SSA studies, only one of which used an external dataset remains a major barrier to clinical adoption.

Priority 2

Multi-site external validation of promising SSA-developed and SSA-validated models should be pre-registered, conducted on populations independent of those used for development, and use shared protocols across SSA networks (PRECISE, AMANHI, EN-INDEPTH, ALERT) so that results are comparable; validated models should then be embedded in existing platforms and workflows, such as perinatal e-registries and cardiotocography monitoring, so that they change decisions at the point of care.

Priority 3

Decision-curve and net-benefit reporting, currently absent in the SSA studies we reviewed, should be standard for any model intended to inform clinical practice; without it, the question of whether a model improves care cannot be answered. Priority should also go to models built on predictors routinely available in SSA facilities, so that risk tools remain usable where ultrasound and laboratory markers are scarce, which is a question of equity as much as of method.

CONCLUSION

Prediction models for stillbirth and neonatal death in SSA remain under-represented and largely internally validated. Methodological standards and SSA-relevant data platforms to address this gap are already in place; the bottleneck is application. Future SSA work should be pre-registered on existing perinatal data platforms, externally validated across countries, and reported in line with TRIPOD + AI guidelines.

Abbreviations

Abbreviation	Definition
AI	Artificial intelligence
ALERT	Action Leveraging Evidence to Reduce perinatal mortality and morbidity (trial / perinatal e-registry)
AMANHI	Alliance for Maternal and Newborn Health Improvement
ANC	Antenatal care
ANN	Artificial neural network
APGAR	Appearance, Pulse, Grimace, Activity and Respiration score
AUC	Area under the (receiver operating characteristic) curve
AUC-PR (AUPRC)	Area under the precision-recall curve
AUROC	Area under the receiver operating characteristic curve
BMI	Body mass index
BW	Birth weight
C-index / C-statistic	Concordance index / concordance statistic
CART	Classification and regression tree
CatBoost	Categorical boosting
CDH	Congenital diaphragmatic hernia
CI	Confidence interval
CNN	Convolutional neural network
CTG	Cardiotocography
CV	Cross-validation
DCA	Decision-curve analysis
DHS	Demographic and Health Survey
DT	Decision tree
EBM	Explainable boosting machine
EDHS	Ethiopian Demographic and Health Survey
EFW	Estimated foetal weight
EN-INDEPTH	Every Newborn-INDEPTH study
ENND	Early neonatal death
FGR	Foetal growth restriction
GA	Gestational age
GAM	Generalised additive model
GB	Gradient boosting
GDM	Gestational diabetes mellitus
GLM	Generalised linear model
HDP	Hypertensive disorders of pregnancy
HELLP	Haemolysis, elevated liver enzymes and low platelets

HR	Hazard ratio
INDEPTH	International Network for the Demographic Evaluation of Populations and their Health
IPPIC	International Prediction of Pregnancy Complications (network)
IQR	Interquartile range
ISS	Injury Severity Score
IUFD	Intrauterine foetal death (demise)
IVH	Intraventricular haemorrhage
KCMC	Kilimanjaro Christian Medical Centre
KNN	K-nearest neighbours
LASSO	Least absolute shrinkage and selection operator
LDA	Linear discriminant analysis
LightGBM (LGBM)	Light gradient-boosting machine
LR	Logistic regression
LSTM	Long short-term memory
MCC	Matthews correlation coefficient
MICE	Multiple imputation by chained equations
ML	Machine learning
MLP	Multilayer perceptron
MNHR	Maternal and Newborn Health Registry
MSE	Mean squared error
NB	Naive Bayes
NEC	Necrotising enterocolitis
NICU	Neonatal intensive-care unit
NND	Neonatal death
NPV	Negative predictive value
NSQIP-P	National Surgical Quality Improvement Program - Pediatric
nuMoM2b	Nulliparous Pregnancy Outcomes Study: Monitoring Mothers-to-be
PAPP-A	Pregnancy-associated plasma protein-A
PCC	Population, Concept, Context
PE	Pre-eclampsia
PGM	Probabilistic graphical model
pp	Percentage points
PPHN	Persistent pulmonary hypertension of the newborn
PPV	Positive predictive value
PRECISE	PRECISE Network pregnancy cohort (The Gambia, Kenya and Mozambique)
PRISMA-ScR	Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews

PROBAST+AI	Prediction model Risk Of Bias ASsessment Tool (artificial intelligence extension)
RDS	Respiratory distress syndrome
RF	Random forest
ROC	Receiver operating characteristic
SB	Stillbirth
SE	Standard error
SGA	Small for gestational age
SHAP	SHapley Additive exPlanations
SLE	Systemic lupus erythematosus
SMOTE	Synthetic minority oversampling technique
SSA	Sub-Saharan Africa
SVM	Support vector machine
TOP	Termination of pregnancy
TRIPOD+AI	Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (artificial intelligence extension)
UtA-PI	Uterine artery pulsatility index
VLBW	Very low birth weight
VON	Vermont Oxford Network
WHO	World Health Organization
XGBoost	Extreme gradient boosting

Declarations

Acknowledgements

Not applicable

Authors' contributions

JA conceived the scoping review and is guarantor. JA, EOO, HB and TC designed the study. TC, PW, AV, MB, HB, EOO and JA developed the protocol registered on the Open Science Framework. TC, AV and PW conducted the database searches, duplicate title and abstract screening, duplicate full-text screening, duplicate data extraction and independent narrative synthesis, with disagreements resolved by JA. AV and MB contributed to the analytical framework, methodological quality assessment using TRIPOD+AI and PROBAST-AI, and the visual evidence-map design. JA, PW, HB and EOO provided senior methodological and substantive oversight on maternal and newborn health and prediction-model reporting. TC drafted the first manuscript. All authors critically revised successive drafts for important intellectual content, approved the final version for submission and agree to be accountable for all aspects of the work. Authorship follows the International Committee of Medical Journal Editors (ICMJE) recommendations.

Funding

This research was made possible with funding from the Wellcome Trust Accelerator Grant 314747/Z/24/Z.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data availability

All data extracted and analysed in this scoping review are included in the published article and its supplementary materials. The extraction workbook and analytic code are available from the corresponding author on reasonable request and at the Open Science Framework registration for the protocol.

Declarations

Patient and public involvement

Patients and the public were not directly involved in the design or conduct of this scoping review, as it used only published literature.

Ethics approval and consent to participate

This scoping review synthesised published evidence and did not involve primary data collection from human participants. Ethical approval was obtained at the London School of Hygiene & Tropical Medicine (ref: 32451)

Consent for publication: Not applicable

Competing interests: The authors declare no competing interests.

References

1. UN IGME. Child Mortality, Stillbirth, and Causes of Death Estimates. 2024.
2. UNICEF. Standing Up for Stillbirth: Current estimates and key interventions. UNICEF DATA. 2024. <https://data.unicef.org/resources/standing-up-for-stillbirth-report/>. Accessed 26 Sept 2025.
3. Maugeri A, Barchitta M, Schillaci G, Agodi A. Spatial patterns and temporal trends in stillbirth, neonatal, and infant mortality: an exploration of country-level data from 2000 to 2021. *J Glob Health*. 2025;15:04034. <https://doi.org/10.7189/jogh.15.04034>.
4. United Nations. The Sustainable Development Goals Report 2023: Special Edition.
5. World Health Organization. Every Woman Every Newborn Everywhere (EWENE). Every Woman Every Newborn Everywhere (EWENE). 2024. <https://ewene.org/>. Accessed 26 Nov 2025.
6. Mukherjee A, Di Stefano L, Blencowe H, Mee P. Determinants of stillbirths in sub-Saharan Africa: A systematic review. *BJOG Int J Obstet Gynaecol*. 2024;131:140–50. <https://doi.org/10.1111/1471-0528.17562>.
7. Rahman A, Lee KH, Arifeen SE, Hossain MZ, Chowdhury AI, Alam M et al. Causes of stillbirth in sub-Saharan Africa and South Asia: Findings from Child Health and Mortality Prevention Surveillance, 2016–2023. 2025;2025.09.09.25335469. <https://doi.org/10.1101/2025.09.09.25335469>
8. Jena BH, Jaldo MM. Determinants of neonatal mortality in sub-Saharan Africa: systematic review and meta-analysis of adverse newborn conditions. *BMC Public Health*. 2025;25:3058. <https://doi.org/10.1186/s12889-025-24419-z>.
9. Wariri O, Dotse-Gborgbortsi W, Agbla SC, Jah H, Cham M, Jawara BF, et al. Beyond proximity: an observational study of stillbirth rates and emergency obstetric and newborn care accessibility in The Gambia. *BMJ Glob Health*. 2025;10:e016579. <https://doi.org/10.1136/bmjgh-2024-016579>.
10. Lawn JE, Blencowe H, Oza S, You D, Lee AC, Waiswa P, et al. Every Newborn: progress, priorities, and potential beyond survival. *Lancet*. 2014;384:189–205. [https://doi.org/10.1016/S0140-6736\(14\)60496-7](https://doi.org/10.1016/S0140-6736(14)60496-7).
11. World Health Organization. Standards for Improving the Quality of Care for Small and Sick Newborns in Health Facilities. 1st ed. Geneva: World Health Organization; 2020.

12. Akuze J, Otieno GP, Amare SY, Ngatia B, Phillip W, Kirakoya-Samadoulougou F, et al. Data science and artificial intelligence for maternal, newborn and child health: scoping review and thematic analysis. *BMC Public Health*. 2025. <https://doi.org/https://doi.org/10.1186/s12889-025-25430-0>.
13. Sufriyana H, Husnayain A, Chen Y-L, Kuo C-Y, Singh O, Yeh T-Y, et al. Comparison of Multivariable Logistic Regression and Other Machine Learning Algorithms for Prognostic Prediction Studies in Pregnancy Care: Systematic Review and Meta-Analysis. *JMIR Med Inf*. 2020;8:e16503. <https://doi.org/10.2196/16503>.
14. Bajwa J, Munir U, Nori A, Williams B. Artificial intelligence in healthcare: transforming the practice of medicine. *Future Healthc J*. 2021;8:e188–94. <https://doi.org/10.7861/fhj.2021-0095>.
15. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A guide to deep learning in healthcare. *Nat Med*. 2019;25:24–9. <https://doi.org/10.1038/s41591-018-0316-z>.
16. Mangold C, Zoretic S, Thallapureddy K, Moreira A, Chorath K, Moreira A. Machine Learning Models for Predicting Neonatal Mortality: A Systematic Review. *Neonatology*. 2021;118:394–405. <https://doi.org/10.1159/000516891>.
17. Stringer JSA, Pokaprakarn T, Prieto JC, Vwalika B, Chari SV, Sindano N, et al. Diagnostic Accuracy of an Integrated AI Tool to Estimate Gestational Age From Blind Ultrasound Sweeps. *JAMA*. 2024;332:649–57. <https://doi.org/10.1001/jama.2024.10770>.
18. Victor A. The role of artificial intelligence in maternal and child health: Progress, controversies, and future directions. *PLOS Digit Health*. 2025;4:e0000938. <https://doi.org/10.1371/journal.pdig.0000938>.
19. El Arab RA, Al Moosa OA, Albahrani Z, Alkhalil I, Somerville J, Abuadas F. Integrating Artificial Intelligence into Perinatal Care Pathways: A Scoping Review of Reviews of Applications, Outcomes, and Equity. *Nurs Rep*. 2025;15:281. <https://doi.org/10.3390/nursrep15080281>.
20. Haniffa R, Isaam I, De Silva AP, Dondorp AM, De Keizer NF. Performance of critical care prognostic scoring systems in low and middle-income countries: a systematic review. *Crit Care*. 2018;22:18. <https://doi.org/10.1186/s13054-017-1930-8>.
21. Yang J, Dung NT, Thach PN, Phong NT, Phu VD, Phu KD, et al. Generalizability assessment of AI models across hospitals in a low-middle and high income country. *Nat Commun*. 2024;15:8270. <https://doi.org/10.1038/s41467-024-52618-6>.
22. Wong A, Otles E, Donnelly JP, Krumm A, McCullough J, DeTroyer-Cooley O, et al. External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients. *JAMA Intern Med*. 2021;181:1065–70. <https://doi.org/10.1001/jamainternmed.2021.2626>.
23. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *Int J Soc Res Methodol*. 2005;19–32.
24. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Ann Intern Med*. 2018;169:467–73.
25. Chaweewong T, Wanduru P, Victor A, Bourke M, Blencowe H, Ohuma EO, et al. Predictive models for stillbirths and neonatal deaths: Protocol for a global scoping review with focus on sub-Saharan Africa. *Wellcome Open Res*. 2026;11:90. <https://doi.org/10.12688/wellcomeopenres.25574.1>.
26. Blencowe H, Hug L, Moller A-B, You D, Moran AC. Definitions, terminology and standards for reporting of births and deaths in the perinatal period: International Classification of Diseases (ICD-11). *Int J Gynecol Obstet*. 2025;168:1–9. <https://doi.org/10.1002/ijgo.15794>.
27. Rayyan – Intelligent Systematic Review - Rayyan. 2021. <https://www.rayyan.ai/>. Accessed 3 Sept 2024.
28. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Calster BV, et al. TRIPOD + AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*. 2024;385:e078378. <https://doi.org/10.1136/bmj-2023-078378>.
29. Moons KGM, Damen JAA, Kaul T, Hooft L, Navarro CA, Dhiman P et al. PROBAST + AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. 2025. <https://doi.org/10.1136/bmj-2024-082505>
30. Adane SM, Teshale AB, Belay DG, Nigatu SG. Development and validation of a risk score to predict neonatal mortality among NICU admissions in Southern Ethiopia: a retrospective follow-up study. *Front Pediatr*. 2025;13. <https://doi.org/10.3389/fped.2025.1496019>.

31. Adebajji AO, Asare C, Gyamerah SA. Predictive analysis on the factors associated with birth Outcomes: A machine learning perspective. *Int J Med Inf.* 2024;189:105529. <https://doi.org/10.1016/j.ijmedinf.2024.105529>.
32. Aga MA. Predicting stillbirth and identifying key maternal risk factors using machine learning. *BMJ Paediatr Open.* 2025;9. <https://doi.org/10.1136/bmjpo-2025-004000>.
33. Al Mashrafi SS, Tafakori L, Abdollahian M. Predicting early and late neonatal mortality using machine learning models in Oman. *BMC Public Health.* 2025;26:47. <https://doi.org/10.1186/s12889-025-25796-1>.
34. Al-Fattah AN, Mahindra MP, Yusrika MU, Mapindra MP, Marizni S, Putri VP, et al. A prediction model for stillbirth based on first trimester pre-eclampsia combined screening. *Int J Gynaecol Obstet Off Organ Int Fed Gynaecol Obstet.* 2024;167:1101–8. <https://doi.org/10.1002/ijgo.15755>.
35. Allotey J, Whittle R, Snell KIE, Smuk M, Townsend R, von Dadelszen P, et al. External validation of prognostic models to predict stillbirth using International Prediction of Pregnancy Complications (IPPIC) Network database: individual participant data meta-analysis. *Ultrasound Obstet Gynecol Off J Int Soc Ultrasound Obstet Gynecol.* 2022;59:209–19. <https://doi.org/10.1002/uog.23757>.
36. Alzakari SA, Aldrees A, Umer M, Cascone L, Innab N, Ashraf I. Artificial intelligence-driven predictive framework for early detection of still birth. *SLAS Technol.* 2024;29:100203. <https://doi.org/10.1016/j.slast.2024.100203>.
37. Åmark H, Westgren M, Persson M. Prediction of stillbirth in women with overweight or obesity—A register-based cohort study. *PLoS ONE.* 2018;13:e0206940. <https://doi.org/10.1371/journal.pone.0206940>.
38. Baker S, Xiang W, Atkinson I. Hybridized neural networks for non-invasive and continuous mortality risk assessment in neonates. *Comput Biol Med.* 2021;134:104521. <https://doi.org/10.1016/j.combiomed.2021.104521>.
39. Batista AFM, Diniz CSG, Bonilha EA, Kawachi I, Chiavegatto Filho ADP. Neonatal mortality prediction with routinely collected data: a machine learning approach. *BMC Pediatr.* 2021;21:322. <https://doi.org/10.1186/s12887-021-02788-9>.
40. Bogale DS, Abuhay TM, Dejene BE. Predicting perinatal mortality based on maternal health status and health insurance service using homogeneous ensemble machine learning methods. *BMC Med Inf Decis Mak.* 2022;22:341. <https://doi.org/10.1186/s12911-022-02084-1>.
41. Bosschieter TM, Xu Z, Lan H, Lengerich BJ, Nori H, Painter I, et al. Interpretable Predictive Models to Understand Risk Factors for Maternal and Fetal Outcomes. *J Healthc Inf Res.* 2024;8:65–87. <https://doi.org/10.1007/s41666-023-00151-4>.
42. Cavallaro A, Veglia M, Svirko E, Vannuccini S, Volpe G, Impey L. Using fetal abdominal circumference growth velocity in the prediction of adverse outcome in near-term small-for-gestational-age fetuses. *Ultrasound Obstet Gynecol Off J Int Soc Ultrasound Obstet Gynecol.* 2018;52:494–500. <https://doi.org/10.1002/uog.18988>.
43. Cersonsky TEK, Ayala NK, Pinar H, Dudley DJ, Saade GR, Silver RM, et al. Identifying risk of stillbirth using machine learning. *Am J Obstet Gynecol.* 2023;229. <https://doi.org/10.1016/j.ajog.2023.06.017>. :327.e1-327.e16.
44. Chen J, Ji Y, Su T, Jin M, Yuan Z, Peng Y, et al. Prediction of Adverse Outcomes in De Novo Hypertensive Disorders of Pregnancy: Development and Validation of Maternal and Neonatal Prognostic Models. *Healthcare.* 2022;10:2307. <https://doi.org/10.3390/healthcare10112307>.
45. Chiweza C, Iwuh I, Hasan A, Malata A, Belfort M, Wilkinson J. Can artificial intelligence-augmented fetal monitoring prevent intrapartum stillbirth and neonatal death in a low-income setting: An observational study? *BJOG Int J Obstet Gynaecol.* 2024;131:109–11. <https://doi.org/10.1111/1471-0528.17321>.
46. Choi HJ, Kim I, Lee H-J, Oh HJ, Ahn MK, Baek WI, et al. Clinical characteristics of neonatal cholestasis in a tertiary hospital and the development of a novel prediction model for mortality. *eBioMedicine.* 2022;77. <https://doi.org/10.1016/j.ebiom.2022.103890>.
47. Conte L, Amodeo I, De Nunzio G, Raffaeli G, Borzani I, Persico N, et al. A machine learning approach to predict mortality and neonatal persistent pulmonary hypertension in newborns with congenital diaphragmatic hernia. A retrospective observational cohort study. *Eur J Pediatr.* 2025;184. <https://doi.org/10.1007/s00431-025-06073-0>.
48. Cooper JN, Minneci PC, Deans KJ. Postoperative neonatal mortality prediction using superlearning. *J Surg Res.* 2018;221:311–9. <https://doi.org/10.1016/j.jss.2017.09.002>.
49. Dalton SE, Sakowicz A, Charles AG, Stamilio DM. Major trauma in pregnancy: prediction of maternal and perinatal adverse outcomes. *Am J Obstet Gynecol MFM.* 2023;5:101069. <https://doi.org/10.1016/j.ajogmf.2023.101069>.

50. Espinola-Sánchez M, Sanca-Valeriano S, Campaña-Acuña A, Caballero-Alvarado J. Prediction of neonatal death in pregnant women in an intensive care unit: Application of machine learning models. *Heliyon*. 2023;9. <https://doi.org/10.1016/j.heliyon.2023.e20693>.
51. Fazzari MJ, Guerra MM, Salmon J, Kim MY. Adverse pregnancy outcomes in women with systemic lupus erythematosus: can we improve predictions with machine learning? *Lupus Sci Med*. 2022;9. <https://doi.org/10.1136/lupus-2022-000769>.
52. Flatley C, Gibbons K, Hurst C, Flenady V, Kumar S. Cross-validated prediction model for severe adverse neonatal outcomes in a term, non-anomalous, singleton cohort. *BMJ Paediatr Open*. 2019;3. <https://doi.org/10.1136/bmjpo-2018-000424>.
53. Gatabazi P, Melesse SF, Ramroop S. Comparison of three classes of Marginal Risk Set Model in predicting infant mortality among newborn babies at Kigali University Teaching Hospital, Rwanda, 2016. *BMC Pediatr*. 2020;20:62. <https://doi.org/10.1186/s12887-020-1945-1>.
54. Giorgione V, Trapani M, Lopian M, Brutto M, Ferrante MG, Bhide A, et al. Predicting Adverse Perinatal Outcomes in Dichorionic Twin Pregnancies: A Multicentre Cohort Study. *BJOG Int J Obstet Gynaecol*. 2025;132:983–90. <https://doi.org/10.1111/1471-0528.18125>.
55. Gunenc O, Dogru S, Yaman FK, Ezveci H, Metin US, Acar A. The Application of Machine Learning Models to Predict Stillbirths. *Med (Mex)*. 2025;61:472. <https://doi.org/10.3390/medicina61030472>.
56. Hauptman I, Gill KS, Lim T, Mack WJ, Wilson ML. Prediction of neonatal outcomes using gestational age vs ACOG definitions of maternal disease severity in hypertensive disorders of pregnancy. *Arch Gynecol Obstet*. 2025;311:639–48. <https://doi.org/10.1007/s00404-024-07684-y>.
57. Houweling TAJ, van Klaveren D, Das S, Azad K, Tripathy P, Manandhar D, et al. A prediction model for neonatal mortality in low- and middle-income countries: an analysis of data from population surveillance sites in India, Nepal and Bangladesh. *Int J Epidemiol*. 2019;48:186–98. <https://doi.org/10.1093/ije/dyy194>.
58. Hromadnikova I, Kotlabova K, Krofta L. First-trimester predictive models for adverse pregnancy outcomes—a base for implementation of strategies to prevent cardiovascular disease development. *Front Cell Dev Biol*. 2024;12. <https://doi.org/10.3389/fcell.2024.1461547>.
59. Hromadnikova I, Kotlabova K, Krofta L. First-Trimester Screening for Miscarriage or Stillbirth—Prediction Model Based on MicroRNA Biomarkers. *Int J Mol Sci*. 2023;24:10137. <https://doi.org/10.3390/ijms241210137>.
60. Huang X, Liang Z, Li T, Lingna Y, Zhu W, Li H. A nomogram to predict in-hospital mortality of neonates admitted to the intensive care unit. *Int Health*. 2021;13:633–9. <https://doi.org/10.1093/inthealth/ihab012>.
61. Huang Y, Li Z, Zhu J, Xiao L, Huang Q, Li W, et al. Construction of a risk prediction model for adverse pregnancy outcomes in primipara with gestational diabetes mellitus combined with pregnancy-induced hypertension syndrome. *Clin Exp Hypertens*. 2025;47:2492621. <https://doi.org/10.1080/10641963.2025.2492621>.
62. Predictive Decision Support Analytic Model for Intelligent Obstetric Risks Management. | Springer Nature Link. https://link.springer.com/chapter/10.1007/978-3-030-80216-5_8. Accessed 23 Feb 2026.
63. Kaliappan J, Bagepalli AR, Almal S, Mishra R, Hu Y-C, Srinivasan K. Impact of Cross-Validation on Machine Learning Models for Early Detection of Intrauterine Fetal Demise. *Diagnostics*. 2023;13:1692. <https://doi.org/10.3390/diagnostics13101692>.
64. Khatibi T, Hanifi E, Sepehri MM, Allahqoli L. Proposing a machine-learning based method to predict stillbirth before and during delivery and ranking the features: nationwide retrospective cross-sectional study. *BMC Pregnancy Childbirth*. 2021;21:202. <https://doi.org/10.1186/s12884-021-03658-z>.
65. Koivu A, Sairanen M. Predicting risk of stillbirth and preterm pregnancies with machine learning. *Health Inf Sci Syst*. 2020;8:14. <https://doi.org/10.1007/s13755-020-00105-9>.
66. Kovacs D, Msanga DR, Mshana SE, Bilal M, Oravcova K, Matthews L. Developing practical clinical tools for predicting neonatal mortality at a neonatal intensive care unit in Tanzania. *BMC Pediatr*. 2021;21:537. <https://doi.org/10.1186/s12887-021-03012-4>.
67. Kuanakewinyoo M, Ntarasit AI, Saelee J, Muangprathub J. Developing a predictive model for stillbirth risk using machine learning algorithms. *Int J Innov Res Sci Stud*. 2025;8:1897–907. <https://doi.org/10.53894/ijirss.v8i4.8259>.
68. Kubahoniyesu T, Kabano IH. Predicting adverse pregnancy outcome in Rwanda using machine learning techniques. *PLoS ONE*. 2024;19. <https://doi.org/10.1371/journal.pone.0312447>. 12 December.

69. Lee J, Cai J, Li F, Vesoulis ZA. Predicting mortality risk for preterm infants using random forest. *Sci Rep.* 2021;11:7308. <https://doi.org/10.1038/s41598-021-86748-4>.
70. Li J, Liu X, He S, Ren Y. Interpretable Machine Learning for Predicting Adverse Pregnancy Outcomes in Gestational Diabetes: Retrospective Cohort Study. *JMIR Med Inf.* 2025;13:e71539. <https://doi.org/10.2196/71539>.
71. Lin C, Mi J, Zhang Y, Duan S, Wu J, Li Y. A nomogram prediction model for early death in patients with persistent pulmonary hypertension of the newborn. *Front Cardiovasc Med.* 2022;9. <https://doi.org/10.3389/fcvm.2022.1077339>.
72. Olde Loohuis KM, Luijken K, Brown Amoakoh H, Adu-Bonsaffoh K, Grobbee DE, Klipstein-Grobusch K, et al. Predicting complications in hypertensive disorders of pregnancy: external validation of a prognostic model for adverse perinatal outcomes. *AJOG Glob Rep.* 2025;5:100455. <https://doi.org/10.1016/j.xagr.2025.100455>.
73. Malacova E, Tippaya S, Bailey HD, Chai K, Farrant BM, Gebremedhin AT, et al. Stillbirth risk prediction using machine learning for a large cohort of births from Western Australia, 1980–2015. *Sci Rep.* 2020;10. <https://doi.org/10.1038/s41598-020-62210-9>.
74. Mastrodima S, Akolekar R, Yerlikaya G, Tzelepis T, Nicolaides KH. Prediction of stillbirth from biochemical and biophysical markers at 11–13 weeks. *Ultrasound Obstet Gynecol Off J Int Soc Ultrasound Obstet Gynecol.* 2016;48:613–7. <https://doi.org/10.1002/uog.17289>.
75. Mboya IB, Mahande MJ, Mohammed M, Obure J, Mwambi HG. Prediction of perinatal death using machine learning models: a birth registry-based cohort study in northern Tanzania. 2020. <https://doi.org/10.1136/bmjopen-2020-040132>
76. McLachlan S, Daley BJ, Saidi S, Kyrimi E, Dube K, Grosan C, et al. A Bayesian Network model of pregnancy outcomes for England and Wales. *Comput Biol Med.* 2025;189:110026. <https://doi.org/10.1016/j.compbimed.2025.110026>.
77. Ferro de Mello ME, da Silva Rocha É, Endo PT. Machine learning for preventing stillbirths: is it possible to transform data into life-saving insights? *BMC Pregnancy Childbirth.* 2025;25:906. <https://doi.org/10.1186/s12884-025-08028-7>.
78. Meng X, Wang L, Wu M, Zhang N, Li X, Wu Q. Construction of a Prediction Model for Adverse Perinatal Outcomes in Foetal Growth Restriction Based on a Machine Learning Algorithm: A Retrospective Study. *BJOG Int J Obstet Gynaecol.* 2025;132:1514–23. <https://doi.org/10.1111/1471-0528.18226>.
79. Mengying C, Pengfei C, Jinfeng H, Yi W, Wei L, Zhenhua G. Visualization of Risk Factors and Predictive Models for Early Death of Neonatal Gastric Perforation. *Clin Pediatr (Phila).* 2024;63:986–94. <https://doi.org/10.1177/00099228231201315>.
80. Van Mieghem T, Lewi L, Slaghekke F, Lopriore E, Yinon Y, Raio L, et al. Prediction of fetal death in monochorionic twin pregnancies complicated by Type-III selective fetal growth restriction. *Ultrasound Obstet Gynecol Off J Int Soc Ultrasound Obstet Gynecol.* 2022;59:756–62. <https://doi.org/10.1002/uog.24896>.
81. Monfared V, Hashemi A. Prediction Analysis of Preterm Neonates Mortality using Machine Learning Algorithms via Python Programming. 2024;:2023.01.20.524905. <https://doi.org/10.1101/2023.01.20.524905>
82. Narayan Y, Jindal DN, Ramya DC, Sathiya ADM. The Role of Artificial Intelligence in Predicting Intrauterine Fetal Demise: A Case-Control Study. *Vasc Endovasc Rev.* 2025;8:58–65.
83. Ngusie HS, Mengiste SA, Zemariam AB, Molla B, Tesfa GA, Seboka BT, et al. Predicting adverse birth outcome among childbearing women in Sub-Saharan Africa: employing innovative machine learning techniques. *BMC Public Health.* 2024;24. <https://doi.org/10.1186/s12889-024-19566-8>.
84. Ngwenya S, Jones B, Mwembe D, Nare H, Heazell AEP. Development and validation of risk prediction models for adverse maternal and neonatal outcomes in severe preeclampsia in a low-resource setting, Mpilo Central Hospital, Bulawayo, Zimbabwe. *Pregnancy Hypertens.* 2021;23:18–26. <https://doi.org/10.1016/j.preghy.2020.10.011>.
85. Ognean ML, Coțovanu B, Teacoe DA, Radu IA, Todor SB, Ichim C, et al. Identification of the Best Predictive Model for Mortality in Outborn Neonates—Retrospective Cohort Study. *Healthcare.* 2023;11:3131. <https://doi.org/10.3390/healthcare11243131>.
86. Oh M-Y, Kim S, Kim M, Seo YM, Yum SK. Machine-learning-based evaluation of the usefulness of lactate for predicting neonatal mortality in preterm infants. *Pediatr Neonatol.* 2025;66:310–5. <https://doi.org/10.1016/j.pedneo.2024.09.003>.
87. Paydar K, Sheikhtaheri A. Prediction of pregnancy outcomes in women with systemic lupus erythematosus before pregnancy: Application of machine learning models. *Heliyon.* 2025;11:e42679. <https://doi.org/10.1016/j.heliyon.2025.e42679>.
88. Pepler FT, Uys DW, Nel DG. Predicting mortality and length-of-stay for neonatal admissions to private hospital neonatal intensive care units: a Southern African retrospective study. *Afr Health Sci.* 2012;12:166–73.

<https://doi.org/10.4314/ahs.v12i2.14>.

89. Prins LI, Bruin CM, Kornaat EMN, Pels A, Gordijn SJ, Naaktgeboren CA, et al. Prediction of perinatal mortality in early-onset fetal growth restriction: A post hoc analysis of the Dutch STRIDER trial to predict perinatal mortality in early-onset fetal growth restriction. *Eur J Obstet Gynecol Reprod Biol.* 2025;304:23–9. <https://doi.org/10.1016/j.ejogrb.2024.11.008>.
90. Punyapet P, Suwanrath C, Chainarong N, Sawaddisan R, Vichitkunakorn P. Predictors of adverse perinatal outcomes in fetal growth restriction using a combination of maternal clinical factors and simple ultrasound parameters. *Int J Gynecol Obstet.* 2023;162:669–75. <https://doi.org/10.1002/ijgo.14721>.
91. Ramos AP, Caldieraro F, Nascimento ML, Saldanha R. Reducing inequalities using an unbiased machine learning approach to identify births with the highest risk of preventable neonatal deaths. *Popul Health Metr.* 2025;23:59. <https://doi.org/10.1186/s12963-025-00420-x>.
92. Rees CA, Kisenge R, Godfrey E, Ideh RC, Kamara J, Coleman-Nekar Y-JG, et al. Machine learning approaches to identify neonates and young children at risk for postdischarge mortality in Dar es Salaam, Tanzania and Monrovia, Liberia. *BMJ Paediatr Open.* 2025;9. <https://doi.org/10.1136/bmjpo-2025-003547>.
93. Sayeed MA, Rahman A, Rahman A, Rois R. On the interpretability of the SVM model for predicting infant mortality in Bangladesh. *J Health Popul Nutr.* 2024;43:170. <https://doi.org/10.1186/s41043-024-00646-9>.
94. Shao C, Chen Q, Tang S, Wang C, Sun R. Development, Validation and Clinical Utility of a Risk Prediction Model for Maternal and Neonatal Adverse Outcomes in Pregnant Women with Hypothyroidism. *J Multidiscip Healthc.* 2024;17:1953–69. <https://doi.org/10.2147/JMDH.S457818>.
95. Sheikhtaheri A, Zarkesh MR, Moradi R, Kermani F. Prediction of neonatal deaths in NICUs: development and validation of machine learning models. *BMC Med Inf Decis Mak.* 2021;21:131. <https://doi.org/10.1186/s12911-021-01497-8>.
96. Shukla VV, Eggleston B, Ambalavanan N, McClure EM, Mwenechanya M, Chomba E, et al. Predictive Modeling for Perinatal Mortality in Resource-Limited Settings. *JAMA Netw Open.* 2020;3:e2026750. <https://doi.org/10.1001/jamanetworkopen.2020.26750>.
97. Development and evaluation of machine learning training strategies for neonatal mortality prediction using multicountry data | Scientific Reports. <https://www.nature.com/articles/s41598-025-04066-5>. Accessed 20 Feb 2026.
98. Sylvain MH, Nyabyenda EC, Uwase M, Komezusenge I, Ndikumana F, Ngaruye I. Prediction of adverse pregnancy outcomes using machine learning techniques: evidence from analysis of electronic medical records data in Rwanda. *BMC Med Inf Decis Mak.* 2025;25. <https://doi.org/10.1186/s12911-025-02921-z>.
99. Takramah WK, Aheto JMK. Predictive model and determinants of odds of neonates dying within 28 days of life in Ghana. *Health Sci Rep.* 2021;4:e248. <https://doi.org/10.1002/hsr2.248>.
100. Tesfie TK, Anlay DZ, Abie B, Chekol YM, Gelaw NB, Tebeje TM, et al. Nomogram to predict risk of neonatal mortality among preterm neonates admitted with sepsis at University of Gondar Comprehensive Specialized Hospital: risk prediction model development and validation. *BMC Pregnancy Childbirth.* 2024;24:139. <https://doi.org/10.1186/s12884-024-06306-4>.
101. Thai TN, Smolinski NE, Nduaguba S, Zhu Y, Bird S, Straub L, et al. Development and validation of an algorithm to predict stillbirth gestational age in Medicaid billing records. *Am J Epidemiol.* 2025;194:2295–303. <https://doi.org/10.1093/aje/kwae369>.
102. Triggs T, Crawford K, Flenady V, Kumar S. Predicting the Risk of Stillbirth for Small Infants Using Maternal and Pregnancy Characteristics. *Paediatr Perinat Epidemiol.* 2026;40:4–16. <https://doi.org/10.1111/ppe.70097>.
103. Wang L, Qin A, Yang Y, Jin Y, Huang Q, Huang X, et al. Evaluation of Pregnancy Risks in Women with Subchorionic Hematoma Using Machine Learning Models. *Med Sci Monit.* 2024;30. <https://doi.org/10.12659/MSM.945472>.
104. Wu J, Zhang W-H, Ma J, Bao C, Liu J, Di W. Prediction of fetal loss in Chinese pregnant patients with systemic lupus erythematosus: a retrospective cohort study. 2019. <https://doi.org/10.1136/bmjopen-2018-023849>
105. Yehuala TZ, Mengesha SB, Baykemagn ND. Predicting pregnancy loss and its determinants among reproductive-aged women using supervised machine learning algorithms in Sub-Saharan Africa. *Front Glob Womens Health.* 2025;6. <https://doi.org/10.3389/fgwh.2025.1456238>.
106. Yerlikaya G, Akolekar R, McPherson K, Syngelaki A, Nicolaides KH. Prediction of stillbirth from maternal demographic and pregnancy characteristics. *Ultrasound Obstet Gynecol Off J Int Soc Ultrasound Obstet Gynecol.* 2016;48:607–12.

<https://doi.org/10.1002/uog.17290>.

107. Yin F, Jin M, Li Y, Li Y, Yin X, Xie J, et al. Externally validated nomogram for predicting short-term pregnancy outcome of singleton pregnancies with fetal growth restriction (FGR). *Gynecol Obstet Clin Med*. 2024;4. <https://doi.org/10.1136/gocm-2024-000009>.
108. Young D, Houshmand B, Tan CC, Kirubarajan A, Parbhakar A, Dada J, et al. Predicting adverse outcomes in pregnant patients positive for SARS-CoV-2: a machine learning approach- a retrospective cohort study. *BMC Pregnancy Childbirth*. 2023;23:553. <https://doi.org/10.1186/s12884-023-05679-2>.
109. Zhang Z, Xiao Q, Luo J. Infant death prediction using machine learning: A population-based retrospective study. *Comput Biol Med*. 2023;165:107423. <https://doi.org/10.1016/j.compbiomed.2023.107423>.
110. Zhao Y, Xu L, An P, Zhou J, Zhu J, Liu S, et al. A nomogram for predicting adverse perinatal outcome with fetal growth restriction: a prospective observational study. *BMC Pregnancy Childbirth*. 2025;25:132. <https://doi.org/10.1186/s12884-025-07252-5>.
111. Zheng D, Hao X, Khan M, Wang L, Li F, Xiang N, et al. Comparison of machine learning and logistic regression as predictive models for adverse maternal and neonatal outcomes of preeclampsia: A retrospective study. *Front Cardiovasc Med*. 2022;9. <https://doi.org/10.3389/fcvm.2022.959649>.
112. Zimmerman RM, Hernandez EJ, Yandell M, Tristani-Firouzi M, Silver RM, Grobman W, et al. AI-based analysis of fetal growth restriction in a prospective obstetric cohort quantifies compound risks for perinatal morbidity and mortality and identifies previously unrecognized high risk clinical scenarios. *BMC Pregnancy Childbirth*. 2025;25:80. <https://doi.org/10.1186/s12884-024-07095-6>.
113. Mengstie MA, Telele MT. Predicting neonatal mortality using ensemble machine learning algorithms in the case of Ethiopian Rural Areas. *Discov Artif Intell*. 2025;5:220. <https://doi.org/10.1007/s44163-025-00305-w>.
114. Akolekar R, Bower S, Flack N, Bilardo CM, Nicolaides KH. Prediction of miscarriage and stillbirth at 11–13 weeks and the contribution of chorionic villus sampling. *Prenat Diagn*. 2011;31:38–45. <https://doi.org/10.1002/pd.2644>.
115. Akolekar R, Tokunaka M, Ortega N, Syngelaki A, Nicolaides KH. Prediction of stillbirth from maternal factors, fetal biometry and uterine artery Doppler at 19–24 weeks. *Ultrasound Obstet Gynecol*. 2016;48:624–30. <https://doi.org/10.1002/uog.17295>.
116. Awor S, Byanyima R, Abola B, Kiondo P, Orach CG, Ogwal-Okeng J, et al. Prediction of stillbirth low resource setting in Northern Uganda. *BMC Pregnancy Childbirth*. 2022;22:855. <https://doi.org/10.1186/s12884-022-05198-6>.
117. Trudell AS, Tuuli MG, Colditz GA, Macones GA, Odibo AO. A stillbirth calculator: Development and internal validation of a clinical prediction model to quantify stillbirth risk. *PLoS ONE*. 2017;12:e0173461. <https://doi.org/10.1371/journal.pone.0173461>.
118. Housseine N, Rijken MJ, Weller K, Nassor NH, Gbenga K, Dodd C et al. Development of a clinical prediction model for perinatal deaths in low resource settings. *EClinicalMedicine*. 2022;44. <https://doi.org/10.1016/j.eclinm.2022.101288>.
119. Kayode GA, Grobbee DE, Amoakoh-Coleman M, Adeleke IT, Ansah E, de Groot JAH, et al. Predicting stillbirth in a low resource setting. *BMC Pregnancy Childbirth*. 2016;16:274. <https://doi.org/10.1186/s12884-016-1061-2>.
120. Muin DA, Windsperger K, Attia N, Kiss H. Predicting singleton antepartum stillbirth by the demographic Fetal Medicine Foundation Risk Calculator—A retrospective case-control study. *PLoS ONE*. 2022;17:e0260964. <https://doi.org/10.1371/journal.pone.0260964>.
121. Payne BA, Groen H, Ukah UV, Ansermino JM, Bhutta Z, Grobman W, et al. Development and internal validation of a multivariable model to predict perinatal death in pregnancy hypertension. *Pregnancy Hypertens Int J Womens Cardiovasc Health*. 2015;5:315–21. <https://doi.org/10.1016/j.preghy.2015.08.006>.
122. Schuit E, Hukkelhoven CWPM, Manktelow BN, Papatsonis DNM, de Kleine MJK, Draper ES, et al. Prognostic Models for Stillbirth and Neonatal Death in Very Preterm Birth: A Validation Study. *Pediatrics*. 2012;129:e120–7. <https://doi.org/10.1542/peds.2011-0803>.
123. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol*. 2019;110:12–22. <https://doi.org/10.1016/j.jclinepi.2019.02.004>.

124. Saito T, Rehmsmeier M. The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. PLoS ONE. 2015;10:e0118432. <https://doi.org/10.1371/journal.pone.0118432>.

125. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Mak Int J Soc Med Decis Mak. 2006;26:565–74. <https://doi.org/10.1177/0272989X06295361>.

126. Smith GCS, Yu CKH, Papageorgiou AT, Cacho AM, Nicolaides KH. Group for the FMFSTS. Maternal Uterine Artery Doppler Flow Velocimetry and the Risk of Stillbirth. Obstet Gynecol. 2007;109:144. <https://doi.org/10.1097/01.AOG.0000248536.94919.e3>.

127. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. BMJ. 2020;369:m1328. <https://doi.org/10.1136/bmj.m1328>.

128. Croft TN, Marshall AM, Allen CK, Arnold F, Assaf S, Balian S. Guide to DHS Statistics. Rockville, MD ICF. - References - Scientific Research Publishing. Rockville, Maryland, USA: ICF; 2018.

129. Akuze J, Cousens S, Lawn JE, Waiswa P, Gordeev VS, Arnold F, et al. Four decades of measuring stillbirths and neonatal deaths in Demographic and Health Surveys: historical review. Popul Health Metr. 2021;19(Suppl 1):8. <https://doi.org/10.1186/s12963-020-00225-0>.

130. Craik R, Joseph Akuze, Volvert M-L, Blencowe H, Mukhanya M, Makanga PT, Tchavana C et al. PREgnancy Care Integrating translational Science, Everywhere (PRECISE): a prospective cohort study of African pregnant and non-pregnant women to investigate placental disorders – cohort profile. 2025. <https://doi.org/10.1136/bmjopen-2024-091831>

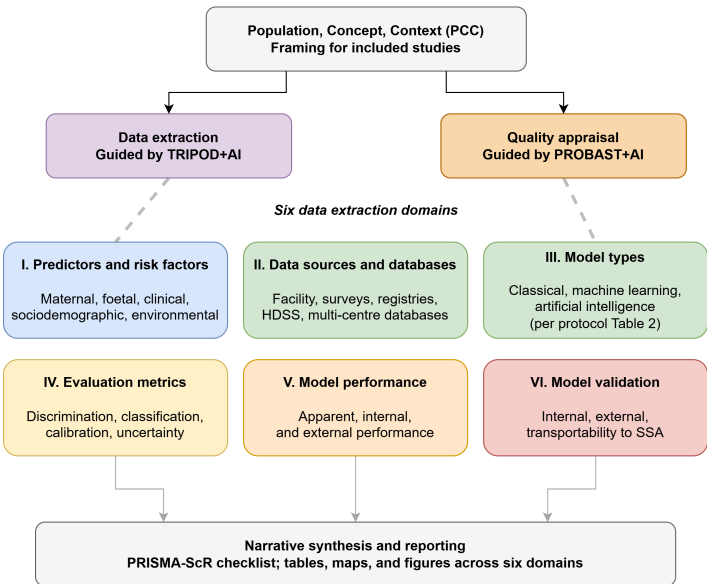
131. The Alliance for Maternal and Newborn Health Improvement (AMANHI) mortality study group, Ahmed I, Ali SM, Amenga-Etego S, Ariff S, Bahl R, et al. Population-based rates, timing, and causes of maternal deaths, stillbirths, and neonatal deaths in south Asia and sub-Saharan Africa: a multi-country prospective cohort study. Lancet Glob Health. 2018;6:e1297–308. [https://doi.org/10.1016/S2214-109X\(18\)30385-1](https://doi.org/10.1016/S2214-109X(18)30385-1).

132. Akuze J, Blencowe H, Waiswa P, Baschieri A, Gordeev VS, Kwesiga D, et al. Randomised comparison of two household survey modules for measuring stillbirths and neonatal deaths in five countries: the Every Newborn-INDEPTH study. Lancet Glob Health. 2020;8:e555–66. [https://doi.org/10.1016/S2214-109X\(20\)30044-9](https://doi.org/10.1016/S2214-109X(20)30044-9).

133. Akuze J, Annerstedt KS, Benova L, Chipeta E, Dossou J-P, Gross MM, et al. Action leveraging evidence to reduce perinatal mortality and morbidity (ALERT): study protocol for a stepped-wedge cluster-randomised trial in Benin, Malawi, Tanzania and Uganda. BMC Health Serv Res. 2021;21:1–14. <https://doi.org/10.1186/s12913-021-07155-z>.

Figures

A. Conceptual framework for data extraction



Adapted from: Chawee Wong et al., Wellcome Open Res 2026;11:90

B. PRISMA-ScR flow diagram

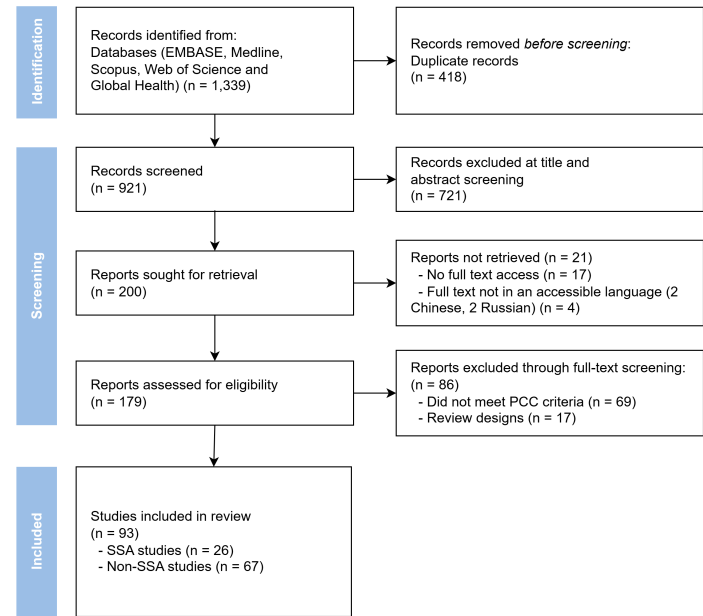


Figure 1

Conceptual framework and study flow

Conceptual framework and study flow. (A) Conceptual framework structured around the Population, Concept, Context (PCC) approach, with data extraction guided by TRIPOD+AI and quality appraisal guided by PROBAST+AI across six predefined domains. (B) PRISMA-ScR flow diagram showing study identification, screening, and inclusion.

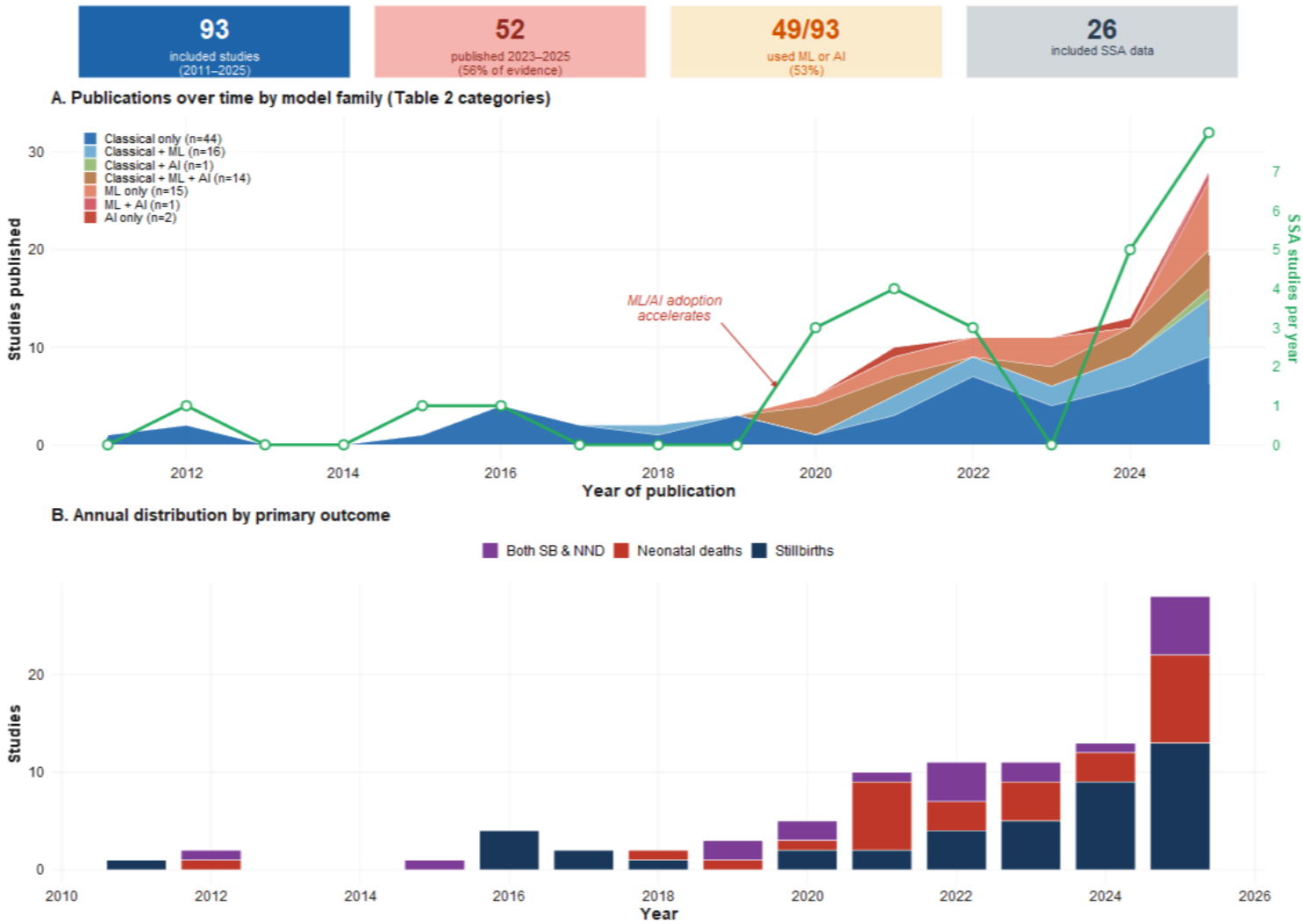


Figure 2

Temporal dynamics of predictive modelling research for stillbirths and neonatal deaths, 2011-2024.

Evidence Map: Prediction Models for Stillbirths and Neonatal Deaths **Systematic Scoping Review (N = 94 studies)**

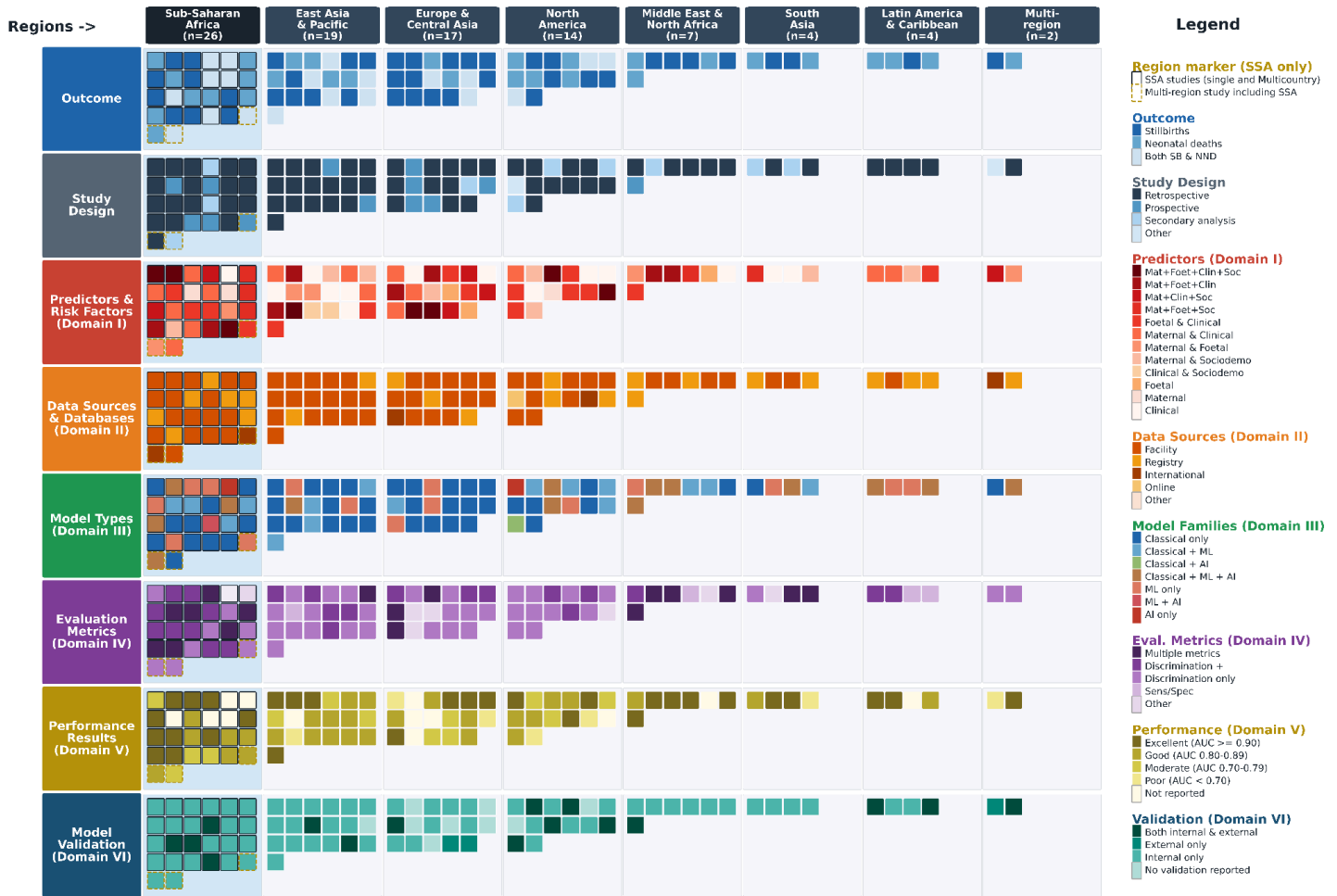
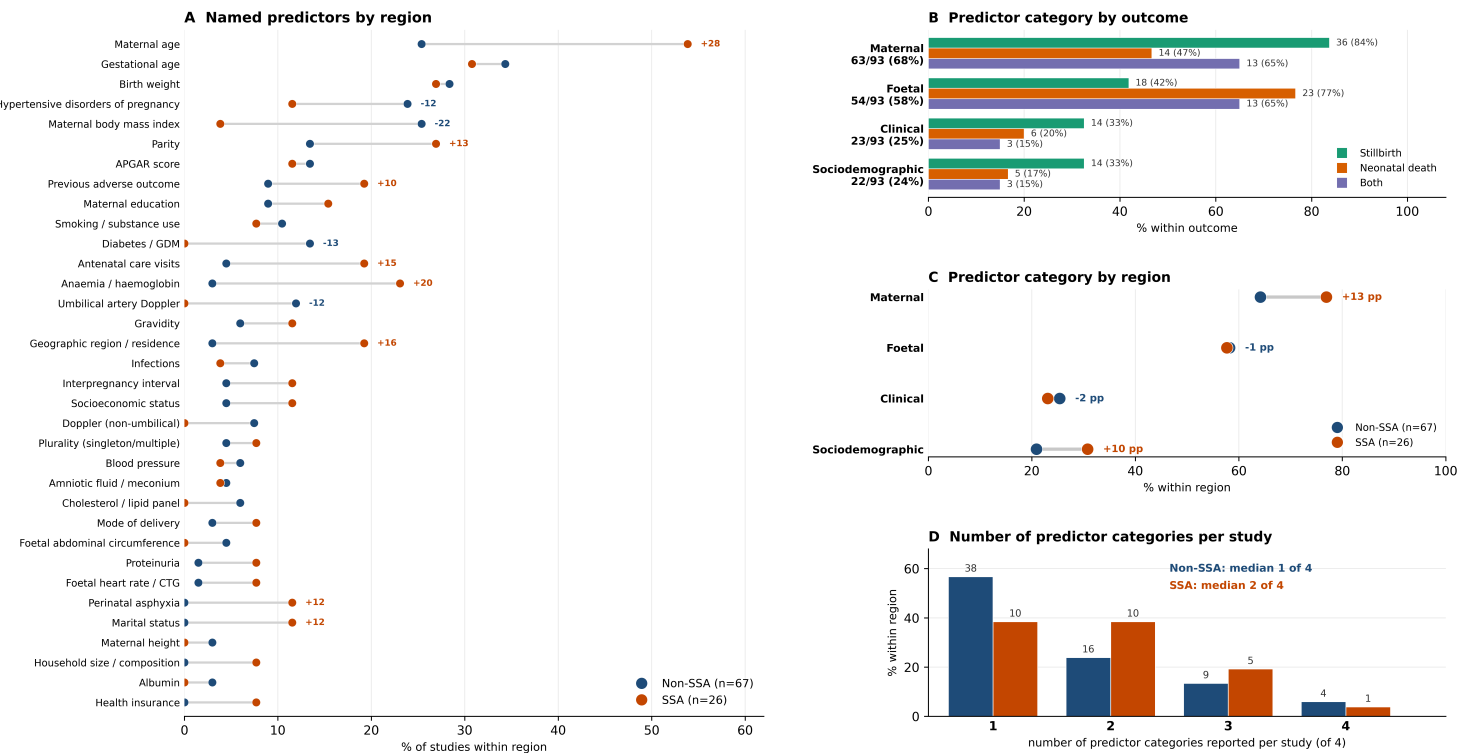


Figure 3

Evidence map of predictive models for stillbirths and neonatal deaths: a systematic summary across six analytical domains and World Bank regions (N = 93 studies).

SB refers to stillbirths, NND refers to neonatal deaths and Both SB & NND refers to stillbirths and neonatal deaths and/or perinatal deaths measured in the same study

Predictor landscape across the 93 included studies (categories: Maternal | Foetal | Clinical | Sociodemographic)



Panel A: labels show the SSA minus Non-SSA gap in percentage points, displayed only where the gap is at least 10.

Figure 4

Predictor landscape across the 93 studies

Predictor categories include maternal, foetal, clinical and sociodemographic. Panel A labels show the SSA minus non-SSA gap in percentage points displaying only where the gap is at least 10%.

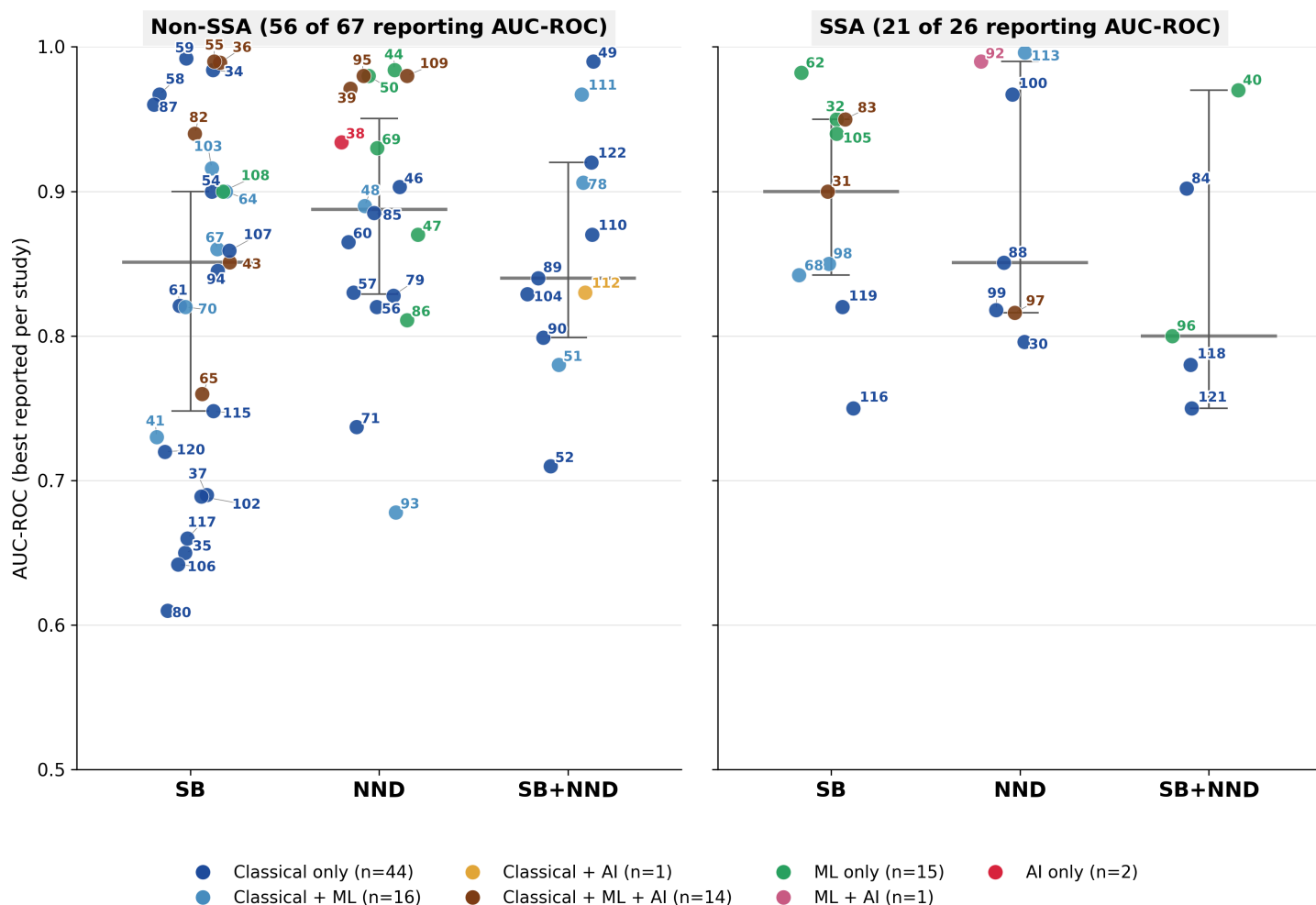


Figure 5

Best reported AUC-ROC by model family, outcome and region

One dot per study (paper's highest AUC-ROC from performance text; AUPRC/AUC-PR excluded), coloured by seven mutually-exclusive model-family combinations in Table 1. Grey crossbar = median across studies in each model family. SB refers to stillbirths, NND refers to neonatal deaths and Both SB & NND refers to stillbirths and neonatal deaths and/or perinatal deaths measured in the same study

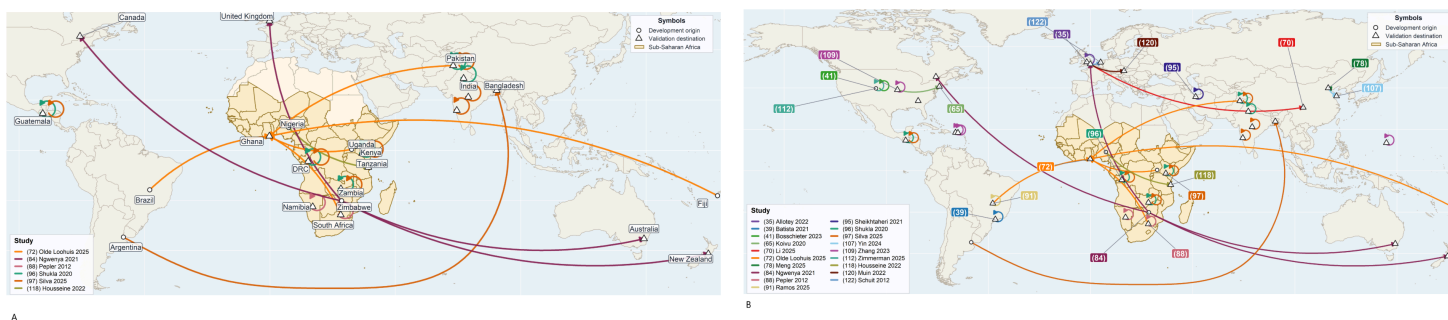


Figure 6

External validation and transportability of predictive models.

Panel A: Predictive models involving sub-Saharan African countries (n=6).

Panel B: External validation and transportability involving SSA and non-SSA sites (n=19).

Panel A: Predictive models with itemised sub-Saharan African development or validation cohorts (n=6: Olde Loohuis 2025 [72], Ngwenya 2021 [84], Pepler 2012 [88], Shukla 2020 [96], Silva 2025 [97], Housseine 2022 [118]). Each colour is one study; multi-country studies draw one arrow or self-loop per mapped country route.

Panel B: External validation and transportability across the 19 studies reporting external validation. Each colour represents one study; multi-country studies draw one arrow or self-loop per mapped country route. Five validation-design modes are shown: (i) temporal validation within the same country (n=4: refs 39, 88, 91, 109); (ii) geographic or hold-out validation within the same country (n=5: refs 41, 65, 78, 107, 112); (iii) cross-country external validation (n=7: refs 35, 70, 72, 84, 118, 120, 122); (iv) prospective external validation in the same facility (n=1: ref 95); and (v) multi-region pooled or temporal validation including SSA and non-SSA sites (n=2: refs 96, 97). Within mode (iii), one transport is within sub-Saharan Africa (Nigeria to Tanzania; ref 118), two are intra-European (United Kingdom to Austria, ref 120; United Kingdom to the Netherlands, ref 122), one is a United Kingdom to China transport (ref 70), two involve SSA as origin or destination (refs 72, 84), and one bypasses sub-Saharan Africa entirely (ref 35). Allotey 2022 (ref 35) is shown as a United Kingdom evidence-hub loop because the IPPIC validation cohort countries are not itemised in the source paper.

Base map: Natural Earth medium scale and country boundaries. Circles denote development origins and triangles validation destinations; self-loops indicate validation in the same location as development.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [PRISMA ScR checklist.docx](#)
- [Supplementary tables 20260705.docx](#)
- [S2 Quality assessment TRIPOD PROBAST.xlsx](#)
- [Supplementary figures 20260705.docx](#)