ARTIFICIAL INTELLIGENCE PREDICTION MODELS FOR POSTPARTUM HEMORRHAGE: A SYSTEMATIC REVIEW AND META-ANALYSIS

Kridsada Sirichaisit, Kiraphol Kaladee **, Warangkana Chankong **, Walisa Romsaiyud ****

***Doctor of Philosophy Program in Public Health, School of Health Science, Sukhothai Thammathirat Open University, Nonthaburi, Thailand**
****School of Health Science, Sukhothai Thammathirat Open University, Nonthaburi, Thailand**
*****School of Science and Technology, Sukhothai Thammathirat Open University, Nonthaburi, Thailand**

Abstract

Background: Postpartum hemorrhage (PPH) is a leading cause of maternal mortality globally, with the highest burden in low- and middle-income countries, including regions of Southeast Asia. Given the limited predictive accuracy of traditional risk assessment models, artificial intelligence (AI)-based predictive models have emerged as a promising approach to enhance early detection and prevention.
Objectives: To evaluate the effectiveness of AI-based predictive models for PPH through a systematic review and meta-analysis.
Methods: Following Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines, we conducted a systematic search of multiple databases (EMBASE, MEDLINE, ScienceDirect, CINAHL, Google Scholar, and Thai-specific resources) for studies published from 2015 to 2025. Conference abstracts, reviews, and studies without specific PPH outcomes were excluded. Two independent reviewers screened studies, extracted data using the Checklist for Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies (CHARMS) and PROBLAST. Statistical analysis was performed using R version 4.4.3.
Results: Eleven studies were included, employing algorithms such as XGBoost, Logistic Regression, Random Forest, and Gradient Boosting. The pooled AUROC was 0.850 (95% CI: 0.789–0.912), indicating good predictive performance. However, there was substantial heterogeneity (I^2 = 99.6%), primarily due to differences in populations, PPH definitions, and modeling approaches. Most studies relied on internal validation and did not originate from Southeast Asia, highlighting a significant regional evidence gap. The risk of bias was largely unclear due to inadequate reporting on blinded predictor assessment and validation methods. Furthermore, a funnel plot analysis suggested potential publication bias, especially among smaller studies.
Conclusion: AI-based models show promise for predicting PPH but require external validation to confirm generalizability. The absence of studies from Southeast Asia underscores the need for region-specific research, including in Thailand, to develop and validate context-appropriate models for clinical use.

Keywords: postpartum hemorrhage, risk prediction, artificial intelligence, machine learning, systematic review

J Southeast Asian Med Res 2025: 9: e0240
https://doi.org/10.55374/jseamed.v9.240

Corresponding to:
Kaladee K, School of Health Science, Sukhothai Thammathirat Open University, Nonthaburi 11120, Thailand
Email: Kiraphol.kal@stou.ac.th, ORCID ID: 0000-0001-6522-3262

Received: 22 April 2025
Revised: 13 August 2025
Accepted: 19 August 2025

Introduction

Postpartum hemorrhage (PPH) remains one of the most serious complications of childbirth and is a leading cause of maternal mortality worldwide. Despite being largely preventable, PPH continues to contribute significantly to maternal deaths.[1] Traditionally, the prevention of PPH has relied on risk factor-based assessment, which stratifies women into low, moderate, or high-risk categories based on known predictors such as multiple gestation, macrosomia, or placenta previa. However, recent guidelines have shifted from this traditional approach toward prioritizing early detection due to the limited predictive accuracy of conventional risk assessment tools.[2–5] Evidence shows that PPH can occur in women with no apparent risk factors and, conversely, may not occur in those classified as high risk. Additionally, some relevant risk factors, such as maternal diabetes, hypertension, and ethnicity, are not consistently integrated into standard risk assessment protocols.[6] These limitations underscore the unpredictable and multifactorial nature of PPH and highlight the need for more dynamic and individualized approaches. The transition toward early detection reflects a growing recognition that accurate prediction based solely on pre-existing risk factors is insufficient. Early identification of bleeding through real-time monitoring enables timely clinical Intervention, potentially reducing morbidity, improving maternal outcomes, and increasing survival rates.[5,7–9]

Several tools have been developed to predict PPH risk, including the California Maternal Quality Care Collaborative (CMQCC) tool, introduced in 2010.[10] This tool stratifies patients by risk level (low, moderate, high) and has been implemented in various settings, including hospitals in Thailand, such as Buriram Hospital[11] and Charoenkrung Pracharak Hospital.[12] However, these evaluations reported no significant improvement in the prediction of antepartum hemorrhage. Other established tools include those developed by the Association of Women's Health, Obstetric and Neonatal Nurses (AWHONN)[13] and the American College of Obstetricians and Gynecologists (ACOG) Safe Motherhood Initiative.[14] While CMQCC emphasizes hospital-level implementation guidelines, AWHONN focuses on nursing-led clinical assessment tools, and ACOG provides broader obstetric practice guidelines within its Safe Motherhood Initiative. Despite their widespread use, these tools show varying performance. Reported sensitivity for predicting transfusion of ≥1 unit of blood ranges from 0.57 to 0.83 (CMQCC), 0.96 (AWHONN), and 0.54 to 0.88 (ACOG). Specificity values are similarly limited, ranging from 0.26 to 0.72 (CMQCC), 0.09 (AWHONN), and 0.25 to 0.70 (ACOG).[1] These figures illustrate that the predictive capabilities of current tools remain suboptimal, further emphasizing the need for more effective early detection strategies.

These limitations in traditional scoring systems have prompted exploration into advanced computational methods capable of integrating diverse clinical and demographic variables. Advancements in Machine Learning (ML) and Big Data analytics have introduced transformative opportunities for developing more accurate and adaptive medical risk prediction models. This approach aligns with the concept of P4 Medicine, which encompasses Predictive, Preventive, Personalized, and Participatory medicine, a significant direction in modern healthcare[15] i.e., hospitals and clinics alone are not capable to cope with this situation. One of the major technology that aids contemporary healthcare solutions is the smart and connected wearables. The advancement in Internet of Things (IoT. Compared to traditional risk assessment tools, ML-based models have demonstrated significantly improved predictive performance across a range of clinical applications. However, prior studies on AI-based PPH prediction models vary widely in terms of model selection, sample size, feature inclusion, and study design, reflecting both the complexity and adaptability of this emerging field.[16–18]

A systematic review and meta-analysis are essential to assess the predictive accuracy, methodological rigor, and clinical applicability of AI-driven postpartum hemorrhage (PPH)

prediction models. This review will address the research question: "What is the effectiveness of AI-based predictive models for postpartum hemorrhage prediction?" The primary objectives are to synthesize existing evidence, evaluate methodological quality, and identify research gaps. A significant gap exists in Southeast Asia, including Thailand, where PPH-related maternal mortality is a significant concern, limiting the applicability of existing global models. The findings will contribute to the development of AI-powered risk prediction tools and inform strategic recommendations for their implementation within Thailand's maternal healthcare system.

Methods

This study is a systematic literature review and meta-analysis. A comprehensive search was conducted across multiple electronic databases, including EMBASE, MEDLINE, ScienceDirect, CINAHL, Google Scholar, Thai Journals Online, ThaiLIS Digital Collection, and the Thai Thesis & Research Databases, covering the period from 2015 to February 2025. The review adhered to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines[19] and followed the TRIPOD-SRMA (Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis Systematic Reviews and Meta-Analyses) reporting standards.[20] Two independent reviewers screened the titles, abstracts, and full-text articles based on predefined inclusion and exclusion criteria. Any discrepancies were resolved through discussion, with consultation from a third reviewer when consensus could not be reached.

Eligibility criteria for study selection

Studies were included if they involved the development of prediction models using artificial intelligence (AI) for PPH. Eligible studies must be original research published between 2015 and 2025 and must employ machine learning, deep learning, or other statistical modeling techniques to predict PPH. Studies will be excluded if they are conference abstracts, review articles, or if PPH is not evaluated as a distinct outcome (i.e., included as part of a composite with other obstetric complications). Articles for which the full text is not accessible will also be excluded from the analysis.

Search strategy

The search strategy was guided by the PICO framework. For the **Participant (P)** component, no additional search terms were required, as postpartum hemorrhage (PPH) is a condition that exclusively affects pregnant women. The **Intervention (I)** component incorporated keywords related to predictive modeling, including "prediction model," "statistical model," "machine learning," "deep learning," and "artificial intelligence." The **Comparison (C)** component was not applicable, as the study did not involve a comparative intervention. The **Outcome (O)** component was defined as "postpartum hemorrhage." The complete search strategies for each database, including keywords, Boolean operators, and filters to ensure reproducibility. To manage and organize the search results efficiently, bibliographic management software was employed throughout the literature review process.

Study selection

All search results were screened for relevance based on the predefined eligibility criteria. Title and abstract screening, followed by full-text review, was independently performed by two reviewers. Discrepancies at any stage were resolved through discussion; if consensus was not reached, a third reviewer was consulted to make the final decision.

Risk of bias assessment

The risk of bias in included studies were assessed using the Prediction Model Risk of Bias Assessment Tool (PROBAST).[21] PROBAST is specifically designed for evaluating the methodological quality of prediction model studies. Two reviewers performed the risk of bias assessments independently. Any disagreements was discussed and resolved by consensus, with the involvement of a third reviewer if necessary.

Data extraction

Researchers utilized a data extraction form that integrates elements from both the Checklist for Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies (CHARMS) and PROBAST frameworks, as recommended by Fernandez-Felix et al.[22], to facilitate simultaneous data extraction and quality assessment. The extracted data included key study characteristics (e.g., first author, publication year, country of origin, study design, sample size, and study period), model characteristics (e.g., type of machine learning algorithm, input features, prediction time point, and outcome definition), and model development details (e.g., training and validation methods, sample size used for model development, and approaches to handling missing data). Additionally, information on model performance was collected, including discrimination metrics (e.g., area under the receiver operating characteristic curve (AUROC)), calibration metrics, and any other reported performance measures. Missing data were addressed through two distinct approaches. First, for studies with incomplete data, missingness was recorded as 'not specified' if no method was reported. Second, for studies that did not report a 95% confidence interval (CI) for the AUROC, the CI was input to enable quantitative synthesis.

Data analysis

Data analysis was conducted using R software (version 4.4.3), employing the meta and metafor packages for statistical synthesis. Descriptive statistics were used to summarize the general characteristics of the included studies and prediction models. The primary measure of predictive performance was the C-statistic (area under the receiver operating characteristic curve, AUROC), reported with a 95% confidence interval (CI). Where studies did not report the 95% CI of the AUROC, it was calculated using the formula: 95% CI = Estimate ± (1.96 × Standard Error). This approach provided approximations for the lower and upper bounds of the CI. A random-effects meta-analysis was performed to pool AUROC values across studies, accounting for between-study variability. Results were visually presented using forest plots. Heterogeneity was assessed using the I^2 statistic, with thresholds of 25%, 50%, and 75% representing low, moderate, and high heterogeneity, respectively. Publication bias was evaluated through visual inspection of funnel plots. To assess the robustness of the meta-analysis results, a sensitivity analysis was conducted by excluding studies in which the 95% CI of the AUROC was derived via calculation rather than directly reported.

Subgroup analyses were conducted to explore the high heterogeneity observed in the pooled analysis. The first analysis compared models based on their validation strategy, categorizing them into those with internal validation only and those with both internal and external validation; this aimed to evaluate generalizability and robustness. The second subgroup analysis was performed by country to assess the impact of national context, including population characteristics and healthcare systems, on predictive performance. These analyses focused on comparing pooled AUROC values and heterogeneity statistics[12] for each subgroup to gain deeper insights into the sources of variability.

Results

Study selection

The study selection process began with a comprehensive search of six electronic databases: EMBASE(607records),MEDLINE(246records), ScienceDirect (739 records), CINAHL (122 records), Google Scholar (160 records), and Thai databases (0 records), resulting in a total of 1,874 records. After removing 376 duplicate records, 1,498 unique records remained for screening. During the initial screening phase based on titles and abstracts, 1,395 records were excluded for not meeting the inclusion criteria, and an additional 36 records were removed for being conference abstracts; this resulted in 67 articles subjected to full-text review. After a detailed assessment, 56 articles were excluded for the following reasons: 24 did not involve the development of a prediction model, 14 were not conducted in a general postpartum population, six did not use machine learning techniques, 4 were image-based model studies, 4 were nar-

rative reviews, and 4 were systematic reviews. Ultimately, 11 studies met all inclusion criteria and were included in the final analysis **(Figure 1)**.

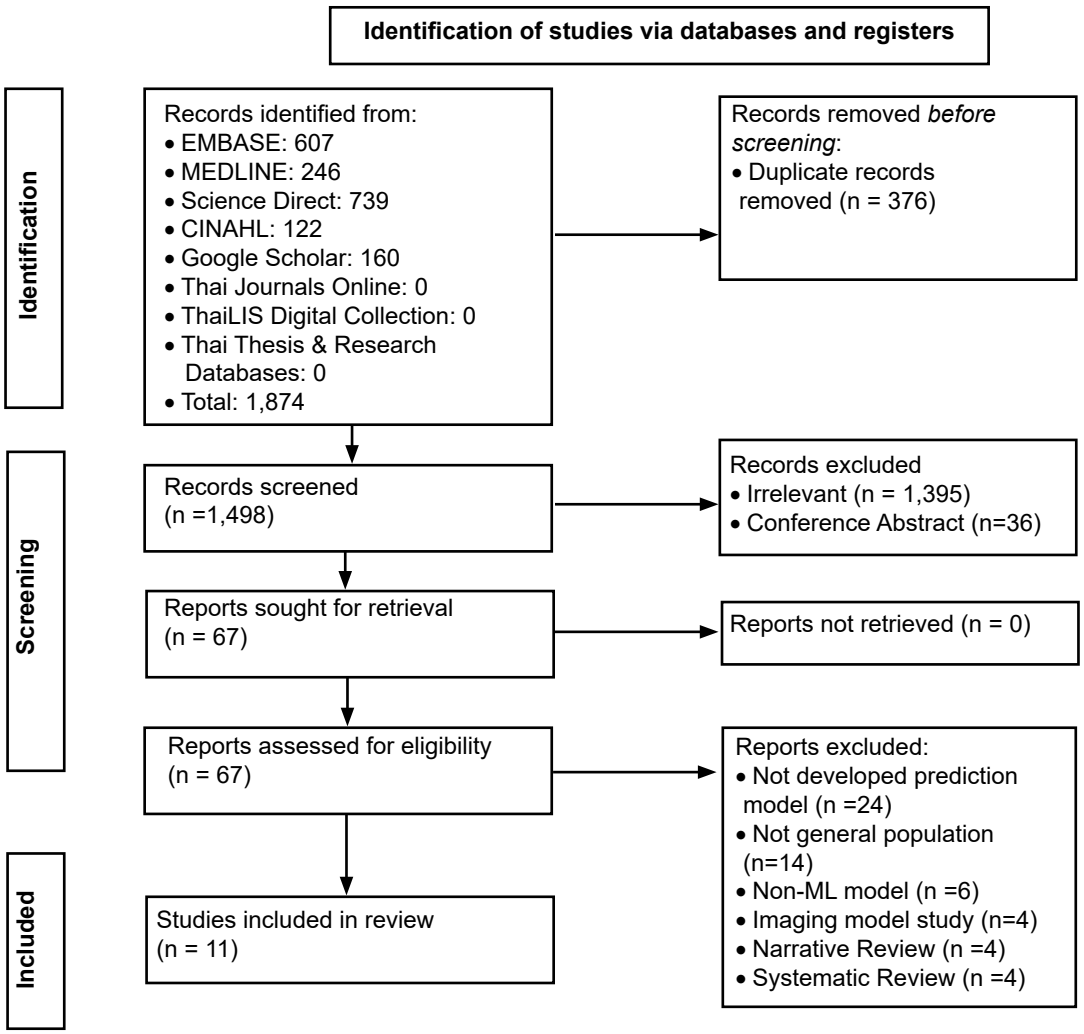

Figure 1. Results of the literature search and selection of included studies

Study and model characteristics

Among the 11 studies included in this review on PPH prediction using machine learning, publications spanned from 2020 to 2024 and were conducted across various countries, including the United States, China, Iran, Kenya, Rwanda, and Nigeria. Most studies employed a retrospective design (n = 9), while two were prospective and one was a case-control study. Study periods ranged from 2002 to 2023, with data sourced from electronic medical records, maternal and infant health surveillance systems, and structured interviews. Sample sizes varied significantly, ranging from 430 to 152,279 participants. Reported PPH incidence ranged from 1.4% to 37.1%. The study settings were diverse, including community health centers, single hospitals, multiple hospitals, tertiary care facilities, and quaternary medical centers. Definitions of PPH were not uniform; most studies defined it as an estimated blood loss (EBL) of ≥1,000 milliliters, while others used thresholds of >500 milliliters for vaginal delivery or >1,000 milliliters for cesarean delivery. Additional criteria included a hematocrit reduction of >10% or the requirement for blood transfusion **(Table 1)**.

The machine learning models applied in these studies varied in algorithm selection and development methods. XGBoost, a tree-based gradient boosting algorithm, was the most frequently used, appearing in five studies and identified as the best-performing model in two. Other com-

monly employed algorithms included Logistic Regression (LR), Random Forest (RF), and various gradient boosting methods. Feature selection techniques differed across studies, incorporating approaches such as LASSO regularization, backward stepwise elimination, and recursive feature elimination. All models aimed to predict PPH during the peripartum period, encompassing both intrapartum and immediate postpartum phases. Approaches to handling missing data also varied, with complete-case analysis, multiple imputation, and median imputation being the most prevalent. For internal validation, studies commonly utilized random data splitting, cross-validation, or bootstrapping methods (**Table 2**).

Risk of bias and applicability

The results of the risk of bias assessment for machine learning-based PPH prediction models indicated that, in the Participants domain, most studies demonstrated a low risk of bias. In contrast, the Predictors and Outcome domains were predominantly rated as having an unclear risk, reflecting limited reporting or variability in definitions and measurement. In the analysis domain, studies generally showed either low or unclear risk of bias, depending on the transparency and robustness of the modeling and validation approaches employed (**Table 3**). In terms of applicability, all included studies were assessed as having low concern across all domains; this suggests that, despite some uncertainty in methodological reporting, the models have strong potential for application in relevant clinical settings.

Results of Syntheses

A total of 11 studies were included in the meta-analysis using a forest plot with imputed confidence intervals (CIs). The pooled area under the receiver operating characteristic curve (AUROC) was 0.908 (95% CI: 0.904–0.911) under a fixed-effect (common-effect) model, and 0.850 (95% CI: 0.789–0.912) under a random-effects model. Among the studies, the highest AUROC was observed in Mehmoush et al. (2023), with a value of 0.990 (95% CI: 0.978–1.002), followed by Krishnamoorthy et al. (2022) and Westcott et al. (2022), reporting AUROCs of 0.980 (95% CI: 0.965–0.995) and 0.979 (95% CI: 0.972–0.986), respectively. Conversely, the lowest AUROC values were reported by Shah et al. (2023) at 0.760 (95% CI: 0.745–0.775) and Zheutlin et al. (2022) at 0.710 (95% CI: 0.695–0.725) (**Figure 2**).

To assess the robustness of the findings, a sensitivity analysis was conducted, excluding studies with imputed confidence intervals. In this restricted analysis of the five studies that reported complete AUROC CIs, the pooled AUROC remained consistent under a fixed-effect model (0.907; 95% CI: 0.903–0.912) and slightly decreased under a random-effects model (0.830; 95% CI: 0.723–0.937). Heterogeneity was found to be statistically significant in both analyses ($p < 0.001$), with I^2 values of 99.8% and 99.6%, respectively, indicating substantial variability across studies **(Figure 3)**.

Subgroup analyses were conducted to investigate the high heterogeneity among studies. A comparison of models based on validation strategy showed that those with external validation had a slightly lower pooled AUROC (0.834) than those with only internal validation (0.856) **(Figure 4)**. While this difference was significant in the common-effect model (p=0.01), it was not in the random-effects model (p=0.72), suggesting external validation provides more conservative and realistic performance estimates, which was crucial for assessing generalizability. A separate subgroup analysis by country revealed significant between-country differences in pooled AUROC values, which ranged from 0.990 in Iran to 0.720 in Nigeria (p<0.001) **(Figure 5)**; this indicates that the national context encompasses variations in population and healthcare systems.

Visual inspection of the funnel plot revealed notable asymmetry in the distribution of study effect sizes, suggesting potential publication bias **(Figure 6).** Specifically, studies with higher standard errors, typically corresponding to smaller sample sizes, exhibited greater variability in reported AUROC values, ranging from low to high, whereas studies with lower standard errors, usually larger studies, demonstrated a more symmetrical and consistent distribution of AUROC

Table 1. Characteristics of studies in the systematic review and meta-analysis

Author, Year	Country	Study Design	Study Period	Data Source	Sample Size	PPH Case (%)	Setting	PPH Definition
Venkatesh et al., 2020[17]	USA	Retrospective study	2002-2008	Electronic Medical Records	152,279	7,279 (4.8%)	Multiple hospitals	Estimated blood loss (EBL) ≥1,000 mL
Goad et al., 2021[23]	USA	Retrospective study	2016-2019	Electronic Medical Records	9,774	618 (6.3%)	Single tertiary care hospital	EBL ≥1,000 mL at delivery
Krishnamoorthy et al., 2022[24]	China	Retrospective study	2020-2021	Electronic Medical Records	11,000	1,042 (9.5%)	Single hospital	Blood loss >500 mL within 24 hours
Westcott et al., 2022[25]	USA	Retrospective study	2013-2018	Electronic Medical Records	30,867	2,179 (7.1%)	Single tertiary care hospital	Blood loss ≥1,000 mL (regardless of delivery method)
Zheutlin et al., 2022[26]	USA	Retrospective study	2011–2019	Electronic Medical Records	70,948	6,639 (9%)	Single hospital	EBL ≥1,000 mL or hematocrit drop/lab shifts or hemorrhage-related interventions/ICD codes.
Mehrnoush et al., 2023[27]	Iran	Retrospective study	2020–2022	Electronic Medical Records	8,888	163 (1.8%)	Single hospital	Blood loss >500 mL (vaginal) or >1000 mL (cesarean), or 10% hemoglobin drop.
Shah et al., 2023[28]	Kenya	Prospective study	2020–2022	Maternal and Newborn Health Monitoring Report	1,576	40 (2.5%)	Community and healthcare facilities	EBL >500 mL (vaginal) or >1,000 mL (cesarean)
Ende et al., 2024[29]	USA	Retrospective study	2018-2022	Electronic Medical Records	21,108	235 (1.4%)	Single quaternary medical center	Blood loss ≥1,000 mL with postpartum transfusion
Holcroft et al., 2024[30]	Rwanda	Case-control study	JAN-JUN 2020	Medical Records and Interviews	430	108 (25.5%)	Multiple hospitals	Blood loss >500 mL within the first hour or requiring transfusion
Wang et al., 2024[31]	China	Retrospective study	2021-2022	Electronic Medical Records	10,803	337 (5.4%)	Single hospital	Vaginal: ≥500 mL; Cesarean: ≥1,000 mL within 24 hours
Okunade et al., 2024[32]	Nigeria	Prospective study	JAN–JUN 2023	Electronic Medical Records	1,222	441 (37.1%)	Multiple hospitals	Blood loss >500 mL within 24 hours (measured by calibrated V-drape)

Table 2. Characteristics of prediction models in the systematic review and meta-analysis

Author, Year	ML Model(s)	Best Model	Feature Selection Method	Time of Prediction	Missing Data Handling Method	Internal Validation	External Validation
Venkatesh et al., 2020	LR, RF, XGBoost	XGBoost	Logistic Regression with Lasso regularization	Peripartum	Multiple Imputed Chained Equations (MICE)	Bootstrapping, Cross-validation	Temporal and Geographical
Goad et al., 2021	LR, GLM, XGBoost	LR	Backward stepwise elimination	At delivery	Complete-case analysis	Bootstrapping	Completely independent
Krishnamoorthy et al., 2022	OBCSA-OSAE, RF, XGBoost, GBDT, SVM, LR	OBCSA-OSAE	Optimal Stacked Auto Encoder	Peripartum	Not specified	Cross-validation	None
Westcott et al., 2022	LR, RF, GBDT, SVM	GBDT	Feature selection algorithms and domain knowledge	Peripartum	Not specified	Random split data	None
Zheutlin et al., 2022	IML	IML	SHapley Additive exPlanations (SHAP) values	Peripartum	Complete-case analysis	Random split data, Cross-validation	None
Mehrnoush et al., 2023	XGBoost, LightGBM, RF, LR, DT, PF, DL, SVM	XGBoost	Pre-specified model	Peripartum	Not specified	Random split data	None
Shah et al., 2023	LR, NB, DT, RF	NB	Extra trees classifier	Peripartum	Not specified	K-fold cross-validation	None
Ende et al., 2024	LR	LR	Stepwise selection	Peripartum	Median imputation	Random split data	Temporal
Holcroft et al., 2024	LR. RF. ERT, XGBoost	RF	Elastic-Net Regularisation	Peripartum	Not specified	Random split data, Cross-validation	None
Wang et al., 2024	RF, AdaBoost, GaussianNB, GradientBoosting, HistGradientBoosting, MLPClassifier, LR	RF	Recursive feature elimination (RFE), recursive feature elimination with cross-validation (RFECV), and SelectKBest	Peripartum	Complete-case analysis	Random split data, Cross-validation	None
Okunade et al., 2024	LR	LR	Backward stepwise elimination	Peripartum	Multiple Imputation	Apparent validation	None

LR: Logistic Regression, RF: Random Forest, XGBoost: Extreme Gradient Boosting, GLM: Generalized Linear Model, GBDT: Gradient Boosting Decision Tree, SVM: Support Vector Machine, OBCSA-OSAE: Optimal Binary Classifier with Stacked Autoencoder - Optimized Sparse Autoencoder, IML: Interpretable Machine Learning, LightGBM: Light Gradient Boosting Machine, DT: Decision Tree, PF: Permutation feature, DL: Deep Learning, NB: Naive Bayes, ERT: Extremely Randomized Trees, AdaBoost: Adaptive Boosting, GaussianNB: Gaussian Naive Bayes, HistGradientBoosting: Histogram-based Gradient Boosting, MLPClassifier: Multi-Layer Perceptron Classifier

Table 3. Risk of bias and applicability assessment

Author, Year	Risk of Bias				Applicability			Overall	
	1. Participants	2. Predictors	3. Outcome	4. Analysis	1. Participants	2. Predictors	3. Outcome	Risk of Bias	Applicability
Venkatesh et al., 2020	+	?	?	?	+	+	+	?	+
Goad et al., 2021	+	?	?	+	+	+	+	?	+
Krishnamoorthy et al., 2022	+	?	?	?	+	+	+	?	+
Westcott et al., 2022	+	?	?	?	+	+	+	?	+
Zheutlin et al., 2022	+	+	+	+	+	+	+	+	+
Mehrnoush et al., 2023	+	?	?	?	+	+	+	?	+
Shah et al., 2023	+	+	?	-	+	+	+	-	+
Ende et al., 2024	+	?	?	+	+	+	+	?	+
Holcroft et al., 2024	+	?	?	?	+	+	+	?	+
Wang et al., 2024	+	?	?	?	+	+	+	?	+
Okunade et al., 2024	+	+	?	+	+	+	+	?	+

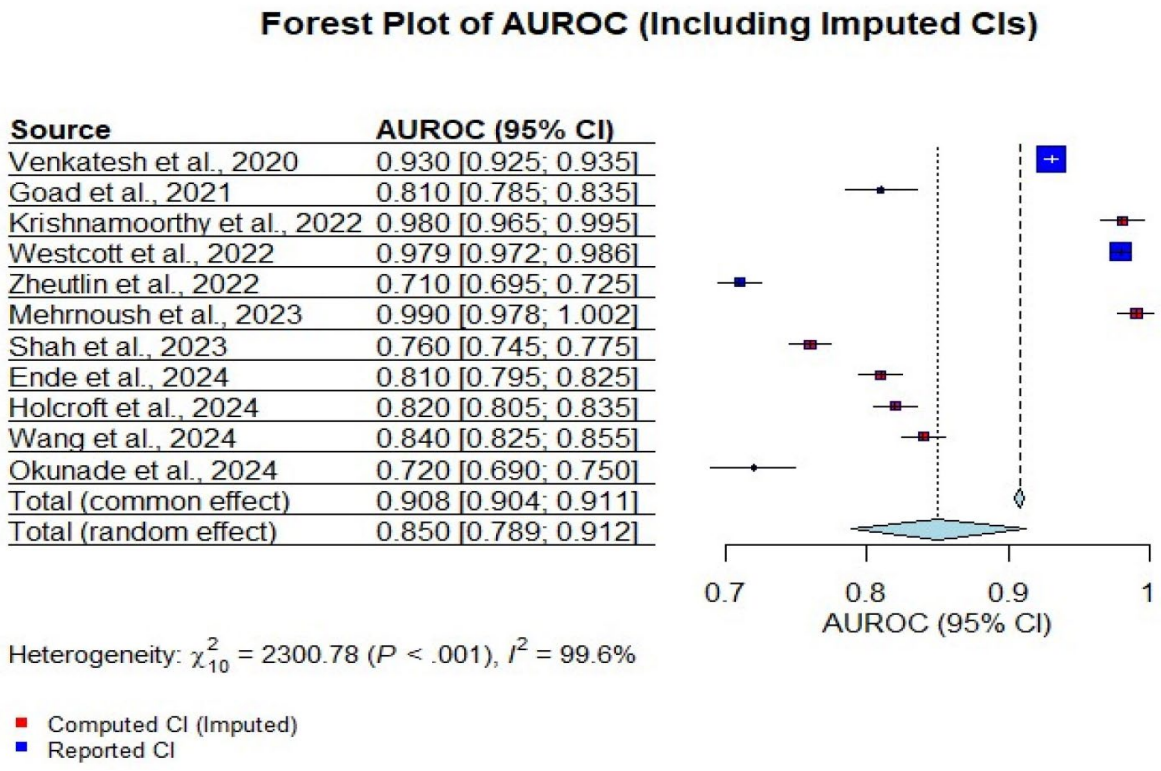

Figure 2. Forest Plot of AUROC for Studies Including Imputed CIs

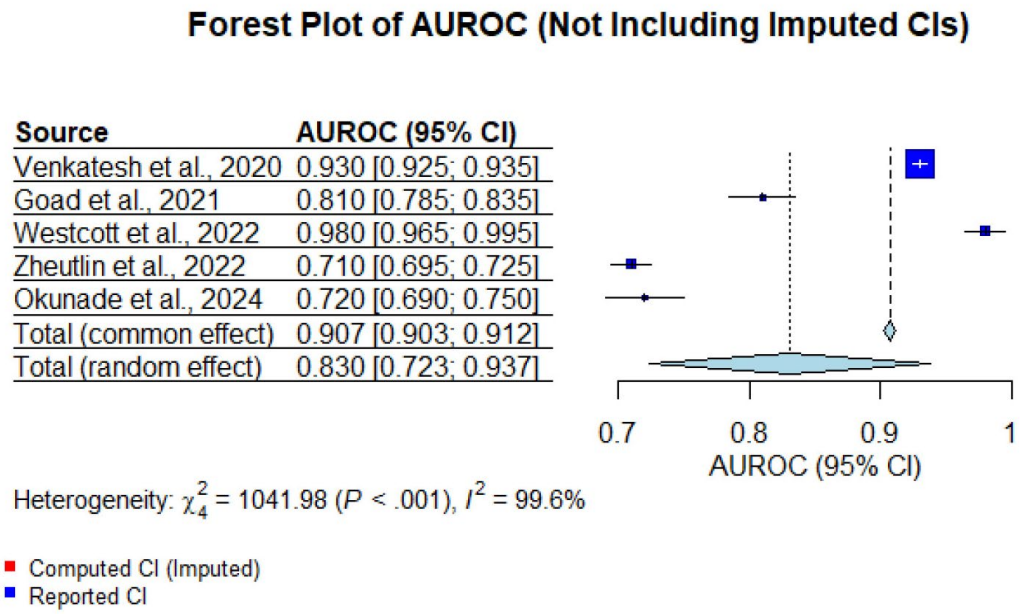

Figure 3. Forest plot of AUROC for studies with reported Cis

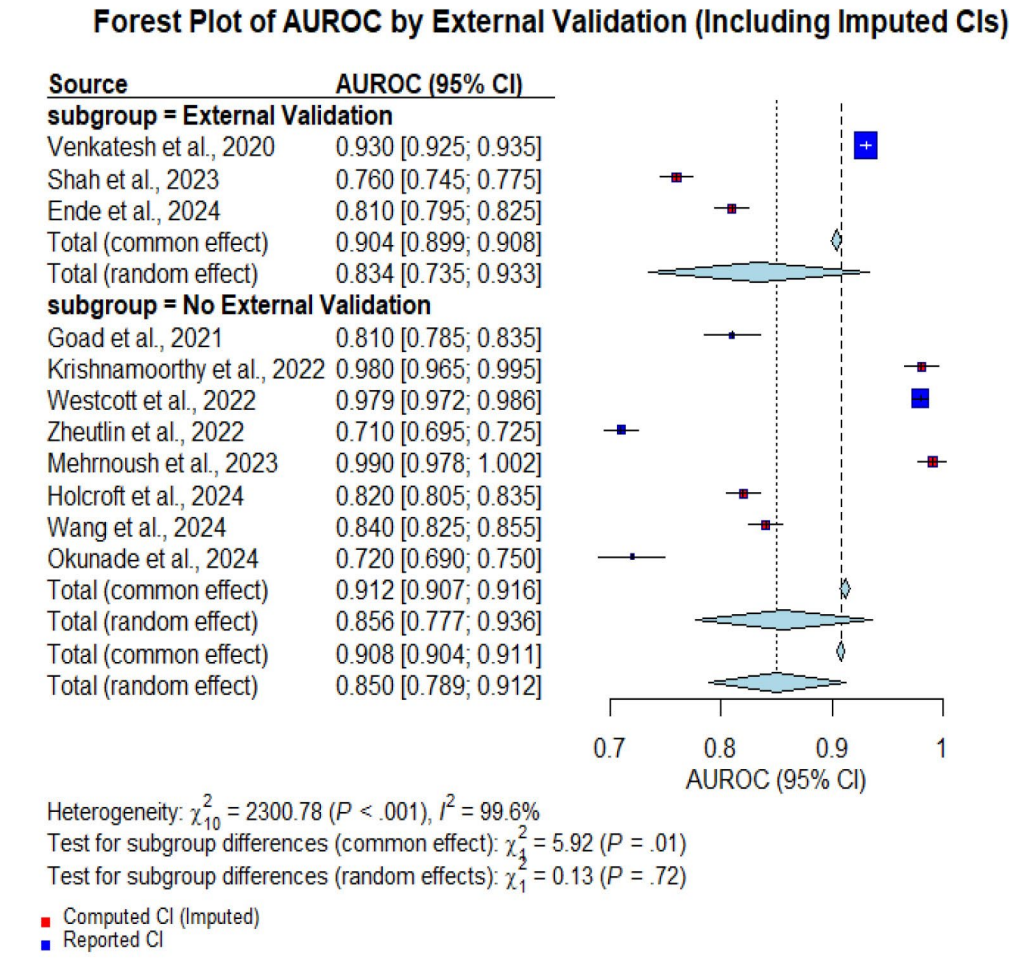

Figure 4. Forest plot of AUROC by external validation status (including imputed CIs).

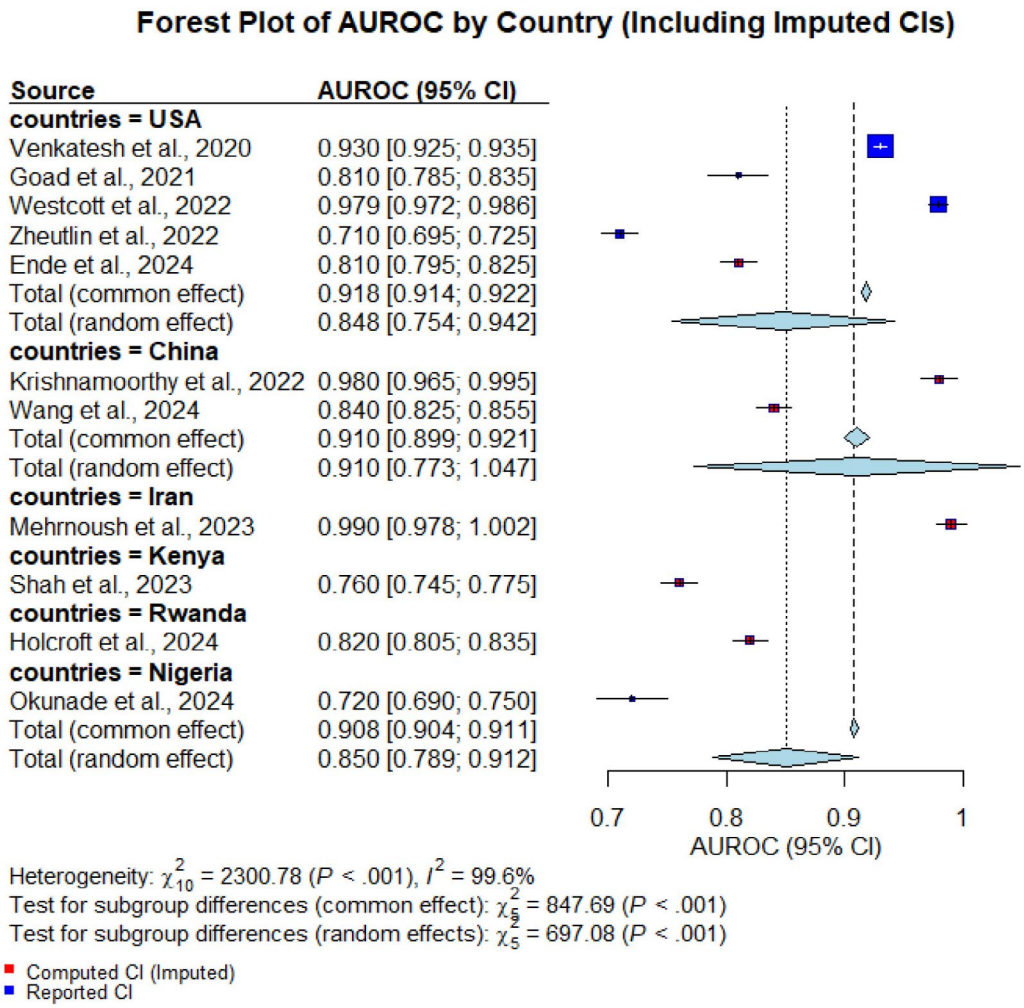

Figure 5. Forest plot of AUROC by country (including imputed CIs).

estimates. This pattern is indicative of small-study effects and may reflect selective reporting or publication bias. Egger's regression test supported this observation, yielding an intercept of -18.52 ($p = 0.041$), which indicated statistically significant asymmetry.

Discussion

This systematic review and meta-analysis included 11 studies that employed various AI/ML algorithms for PPH prediction. The pooled AUROC of 0.850 (95% CI: 0.789–0.912) suggests good predictive performance. However, this finding is significantly tempered by substantial heterogeneity (I^2>99%), which was driven by differences in study populations, PPH definitions, and modeling approaches. Models with exceptionally high AUROC values (e.g., 0.98 - 0.99) should be interpreted cautiously as they may overfit, especially without external validation.

Subgroup analyses revealed significant heterogeneity. Country-level comparisons showed pooled AUROC values ranging from 0.990 in Iran to 0.720 in Nigeria (p< 0.001). Models with external validation yielded slightly lower pooled AUROC, indicating that external validation offers a more conservative estimate of generalizability.

The risk of bias assessment was complicated by poor reporting on key methodological domains, such as the blinding of predictor assessment and details of validation. This lack of transparency raises concerns about the reproducibility of reported performance metrics. Furthermore, a funnel plot analysis suggested potential publication bias, where smaller studies may have reported more extreme positive results, which could inflate the overall pooled performance estimates.

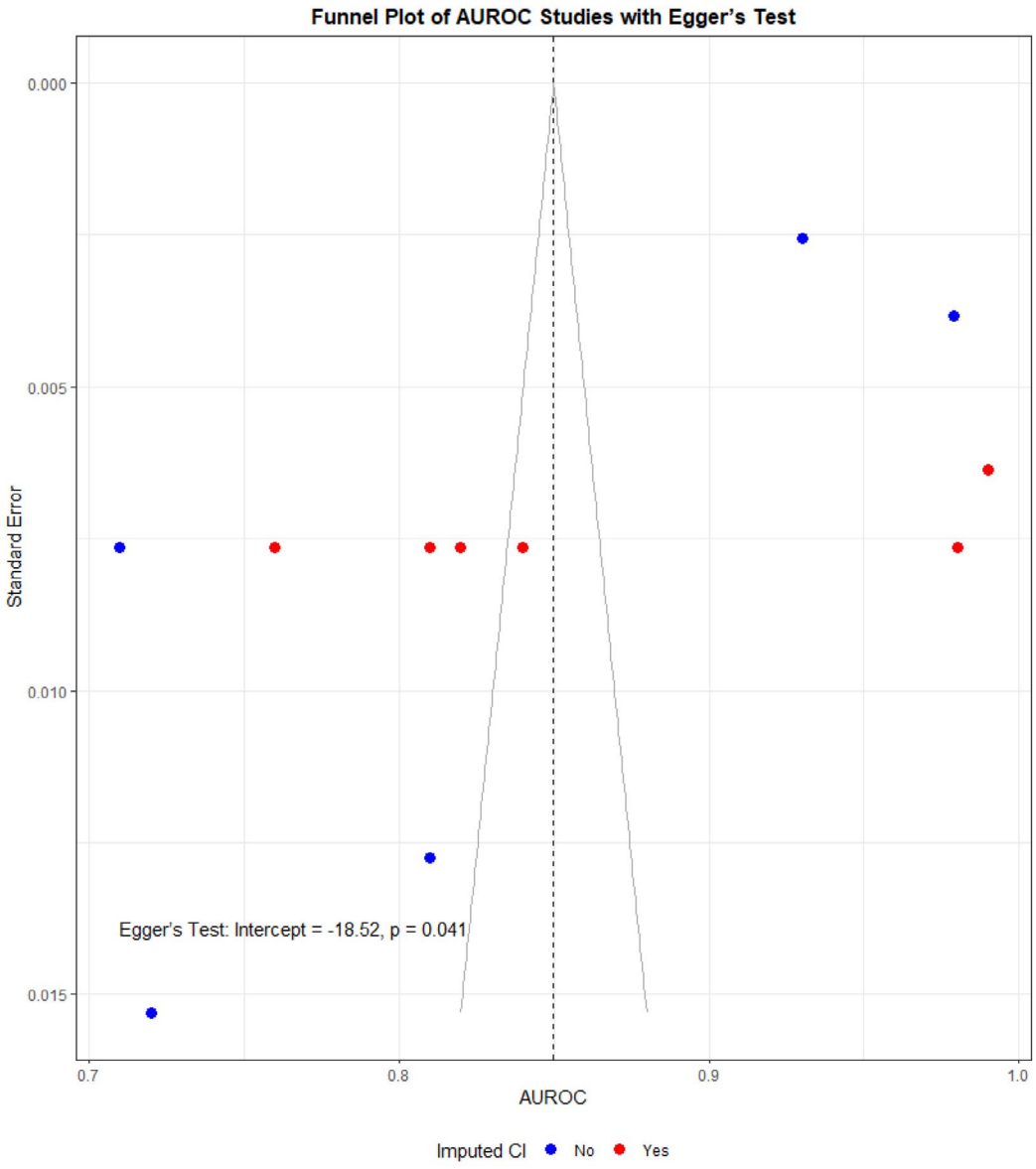

Figure 6. Funnel Plot

A critical limitation is the complete absence of studies from Southeast Asia, including Thailand, a region where PPH remains a significant cause of maternal mortality. This gap underscores the urgent need for region-specific research to develop and validate context-appropriate models. In conclusion, while AI-based PPH prediction models hold considerable promise, their current evidence base is limited by high heterogeneity, potential publication bias, a lack of external validation, and regional gaps. Addressing these issues is crucial for translating predictive performance into tangible clinical benefits and improving maternal health outcomes.

This systematic review and meta-analysis have several notable strengths. First, it incorporates data from studies conducted in multiple countries, contributing to demographic and geographic diversity that enhances the generalizability of the findings. The use of varied machine learning models across the studies allowed for a broad evaluation of predictive approaches, many of which demonstrated strong performance. Notably, most studies reported AUROC values greater than 0.8, and the pooled AUROC values of 0.908 under the fixed-effect model and 0.850 under the random-effects model indicate that the models generally performed well in predicting PPH. Furthermore, the risk of bias assessments revealed that the majority of studies had a

low risk of bias in the participant domain and demonstrated good applicability across all evaluated domains.

Key limitations include substantial heterogeneity ($I^2 > 99\%$), driven by variability in populations, PPH definitions, and methodologies. Although subgroup analyses showed significant AUROC differences across countries, persistent within-country heterogeneity indicates that national context alone does not explain performance variation. Funnel plot asymmetry suggests potential publication bias, while poor reporting hindered risk of bias assessment. Predominant reliance on internal validation limits generalizability, and the paucity of external validation undermines confidence in real-world applicability. Despite promising performance, the evidence base is constrained by these factors and by limited interpretability. Region-specific model development and validation, particularly in Southeast Asia, are essential to enhance translational impact on maternal health.

Interpretability is essential for clinical adoption of AI models. Complex ensemble methods often function as "black boxes," undermining clinician trust and integration into decision making. Few studies employed explainability techniques like SHAP to elucidate feature importance. Future research should prioritize interpretable models or incorporate such tools to enhance transparency and clinician acceptance.

Recommendations for future research and studies in Thailand

Future research should focus on comparing different machine learning models using standardized datasets to enable more reliable comparisons of model performance. Developing prediction models that can assess PPH risk during the antenatal period is also essential, as this would allow for timely preventive interventions and resource planning. Additionally, there is a need to broaden the geographic scope of research to include more studies from developing countries, especially in Southeast Asia. Thailand, in particular, would benefit from local research efforts that account for population-specific risk factors, healthcare infrastructure, and clinical practices.

Policy recommendations

To facilitate the development of accurate and context-appropriate PPH prediction models, Thailand should consider establishing a centralized national database that aggregates data from hospitals nationwide. Such a system would support large-scale model training and validation efforts. Investment in health information technology infrastructure is also critical, particularly systems capable of real-time data collection and analysis to support clinical decision making. Furthermore, national treatment guidelines should be updated to incorporate the use of validated ML-based risk prediction tools as part of routine maternal care. Integrating these models into standard clinical workflows could enhance the timely identification and management of high-risk cases, ultimately improving maternal health outcomes.

Conclusion

Machine learning models, particularly XGBoost, have shown strong potential for predicting PPH, with pooled AUROC values indicating good to excellent performance. However, most models remain in the early stages of development, with limited external validation, and are therefore not yet ready for routine clinical use. To move toward clinical readiness, future work should focus on validating these models in diverse, real-world populations and integrating them into clinical decision-support systems. Equally critical is ensuring model interpretability. Clinicians must be able to understand the key predictors and decision logic to build trust and enable effective use in time-sensitive obstetric care. Advancing both the robustness and transparency of these tools, while adapting them to local healthcare contexts, will be essential for their successful adoption and impact on maternal outcomes.

References

1. Ende HB. Risk assessment tools to predict postpartum hemorrhage. Best Pract Res Clin Anaesthesiol 2022; 36: 341–8.
2. Ende HB, Lozada MJ, Chestnut DH, Osmundson SS, Walden RL, Shotwell MS, et al. Risk factors for atonic postpartum hemorrhage. Obstet Gynecol 2021; 137: 305–23.

3. Huang C, Xue B, Gao Y, Yue S, Redding SR, Wang R, et al. Incidence and risk factors for postpartum hemorrhage after vaginal delivery: A systematic review and meta-analysis. J Obstet Gynaecol Res 2023; 49: 1663–76.
4. Liu CN, Yu FB, Xu YZ, Li JS, Guan ZH, Sun MN et al. Prevalence and risk factors of severe postpartum hemorrhage: a retrospective cohort study. BMC Pregnancy Childbirth 2021; 21: 332.
5. Post SE, Rood KM, Kiefer MK. Interventions of postpartum hemorrhage. Clin Obstet Gynecol 2023; 66: 367–83.
6. Simon A, Pratt M, Hutton B, Skidmore B, Fakhraei R, Rybak N, et al. Guidelines for the management of pregnant women with obesity: A systematic review. Obes Rev 2020; 21: e12972.
7. Boatin AA, Ngonzi J. Early detection and bundled treatment for postpartum hemorrhage. N Engl J Med 2023; 389: 79–80.
8. Williams EV, Goranitis I, Oppong R, Perry SJ, Devall AJ, Martin JT, et al. A cost-effectiveness analysis of early detection and bundled treatment of postpartum hemorrhage alongside the E-MOTIVE trial. Nat Med 2024; 30: 2343–8.
9. Wongthirakun T. Development of participation model in family and community for increase rate of antenatal care before 12 weeks in Ngaeng Health Promoting Hospital, Pua District, Nan Province. J Health Sci Thail 2021; 30(Supplement 2): S283–91.
10. Bingham D, Melsop K, Main E. CMQCC Obstetric hemorrhage toolkit hospital level implementation guide. The California Maternal Quality Care Collaborative (CMQCC). Stanford University School of Medicine; 2010.
11. Litmarg J, Luangkwan S, TaTong N, Tuphimai B. Effectiveness of using risk assessment and application in nursing process for the prevention of early postpartum haemorrhage in buriram hospital. Med J Sissaket Surin Buriram Hosp 2020; 35: 111–24.
12. Wattanaluangarun R, Chantrapitak W, Kajang N, Onlaor R. Effectiveness of using risk assessment and clinical nursing practice guideline for the prevention of early postpartum haemorrhage from uterine atony in Charoenkrung Pracharak Hospital. J Charoenkrung Pr Hosp. 2017; 13: 28–42.
13. Association of Women's Health, Obstetric and Neonatal Nurses. Postpartum hemorrhage (PPH) risk assessment Table 1 [Internet]. 2017 [cited 2024 Dec 14]. Available from: https://cdn-links.lww.com/permalink/aog/b/aog_134_6_2019_10_06_kawakita_19-1065_sdc2.pdf
14. American College of Obstetricians and Gynecologists. Obstetric hemorrhage [Internet].2024 [cited 2024 Dec 14]. Available from: https://www.acog.org/community/districts-and-sections/district-ii/programs-and-resources/safe-motherhood-initiative/obstetric-hemorrhage
15. Li W, Chai Y, Khan F, Jan SRU, Verma S, Menon VG, et al. A Comprehensive survey on machine learning-based big data analytics for IoT-enabled smart healthcare system. Mob Netw Appl ; 26: 234–52.
16. Stafford IS, Gosink MM, Mossotto E, Ennis S, Hauben M. A systematic review of artificial intelligence and machine learning applications to inflammatory bowel disease, with practical guidelines for interpretation. Inflamm Bowel Dis 2022; 28: 1573–83.
17. Venkatesh KK, Strauss RA, Grotegut CA, Heine RP, Chescheir NC, Stringer JSA, et al. Machine learning and statistical models to predict postpartum hemorrhage. Obstet Gynecol 2020; 135: 935–44.
18. Zahrt AF, Henle JJ, Denmark SE. Cautionary guidelines for machine learning studies with combinatorial datasets. ACS Comb Sci 2020; 22: 586–91.
19. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021 Mar 29;n71.
20. Snell KIE, Levis B, Damen JAA, Dhiman P, Debray TPA, Hooft L, et al. Transparent reporting of multivariable prediction models for individual prognosis or diagnosis: checklist for systematic reviews and meta-analyses (TRIPOD-SRMA). BMJ. 2023; 381: e073538.
21. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al.

PROBAST: A tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med 2019; 170: 51–8.

22. Fernandez-Felix BM, López-Alcalde J, Roqué M, Muriel A, Zamora J. CHARMS and PROBAST at your fingertips: a template for data extraction and risk of bias assessment in systematic reviews of predictive models. BMC Med Res Methodol 2023; 23: 44.
23. Goad L, Rockhill K, Schwarz J, Heyborne K, Fabbri S. Development and validation of a prediction model for postpartum hemorrhage at a single safety net tertiary care center. Am J Obstet Gynecol MFM [Internet]. 2021; 3(5). Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L2013388888&from=export
24. Krishnamoorthy S, Liu Y, Liu K. A novel oppositional binary crow search algorithm with optimal machine learning based postpartum hemorrhage prediction model. BMC Pregnancy Childbirth [Internet]. 2022; 22(1). Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L2018285395&from=export
25. Westcott JM, Hughes F, Liu W, Grivainis M, Hoskins I, Fenyo D. Prediction of maternal hemorrhage using machine learning: retrospective cohort study. J Med Internet Res [Internet]. 2022; 24(7). Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L2019631114&from=export
26. Zheutlin AB, Vieira L, Shewcraft RA, Li S, Wang Z, Schadt E, et al. Improving postpartum hemorrhage risk prediction using longitudinal electronic medical records. J Am Med Inform Assoc 2022; 29: 296–305.
27. Mehrnoush V, Ranjbar A, Farashah MV, Darsareh F, Shekari M, Jahromi MS. Prediction of postpartum hemorrhage using traditional statistical analysis and a machine learning approach. AJOG Glob Rep 2023; 3: 100185.
28. Shah SY, Saxena S, Rani SP, Nelaturi N, Gill S, Tippett Barr B, et al. Prediction of postpartum hemorrhage (PPH) using machine learning algorithms in a Kenyan population. Front Glob Womens Health 2023; 4: 1161157.
29. Ende HB, Domenico HJ, Polic A, Wesoloski A, Zuckerwise LC, Mccoy AB, et al. Development and validation of an automated, real-time predictive model for postpartum hemorrhage. Obstet Gynecol [Internet]. 2024;((Ende H.B.) Departments of Anesthesiology, Biostatistics, Obstetrics and Gynecology, and Biomedical Informatics, Vanderbilt University Medical Center, and Vanderbilt University School of Medicine, Nashville, Tennessee; and the Department of Anesthesiology, Boston Children's Hospital, Boston, Massachusetts). Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L644226072&from=export
30. Holcroft S, Karangwa I, Little F, Behoor J, Bazirete O. Predictive modelling of postpartum haemorrhage using early risk factors: A comparative analysis of statistical and machine learning models. Int J Environ Res Public Health 2024; 21(5). Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L2029986187&from=export
31. Wang W, Liao C, Zhang H, Hu Y. Postpartum haemorrhage risk prediction model developed by machine learning algorithms: A single-centre retrospective analysis of clinical data. Clin Exp Obstet Gynecol [Internet]. 2024;51(3). Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L2030924266&from=export
32. Okunade KS, Ugwu AO, Adenekan MA, Olumodeji A, Oshodi YA, Ojo T, et al. Development of antepartum risk prediction model for postpartum hemorrhage in Lagos, Nigeria: A prospective cohort study (Predict-PPH study). Int J Gynecol Obstet [Internet]. 2024. Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L2027946470&from=export