BMJ Open

Accuracy of machine learning and traditional statistical models in the prediction of postpartum haemorrhage: a systematic review

Thais Baeta, Ana Luiza Lunardi Rocha, Juliana Almeida Oliveira, Ana Paula Couto da Silva, Zilma Silveira Nogueira Reis

To cite: Baeta T, Rocha ALL, Oliveira JA, *et al*. Accuracy of machine learning and traditional statistical models in the prediction of postpartum haemorrhage: a systematic review. *BMJ Open* 2025;**15**:e094455. doi:10.1136/bmjopen-2024-094455

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (https://doi.org/10.1136/bmjopen-2024-094455).

Received 01 October 2024
Accepted 03 February 2025

Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

Correspondence to
Dr Thais Baeta;
thaisbaeta3@gmail.com

ABSTRACT

Objectives To evaluate whether postpartum haemorrhage (PPH) can be predicted using both machine learning (ML) and traditional statistical models.

Design Diagnostic systematic review and meta-analysis of observational and clinical studies, prospectively registered on PROSPERO, performed accordingly to the Preferred Reporting Items for Systematic Reviews and Meta-analysis and Prediction model risk of bias assessment tool for studies developing, validating or updating prediction models, with the use of an independent analysis by a large language model (GPT-4 Open AI).

Data sources MEDLINE/PubMed, LILACS-BVS, Cochrane Library, Scopus-Elsevier, Embase-Elsevier and Web of Science.

Eligibility criteria for selected studies The literature search was conducted on 4 January 2024 and included observational studies and clinical trials published in the past 10 years that assessed early PPH and PPH prediction and that applied accuracy metrics for outcomes evaluation. We excluded studies that did not define PPH or had exclusive PPH subgroups evaluation.

Primary and secondary outcome measures The primary outcome is the accuracy of PPH prediction using both ML and conventional statistical models. A secondary outcome is to describe the strongest risk factors of PPH identified by ML and traditional statistical models.

Results Of 551 citations screened, 35 studies were eligible for inclusion. The synthesis gathered 383 648 patients in 24 studies conducted with conventional statistics (CS), 9 studies using ML models and 2 studies using both methods. Multivariate regression was a preferred modelling approach to predict PPH in CS studies, while ML approaches used multiple models and a myriad of features. ML comparison to CS was only performed in two studies, and ML models demonstrated a 95% higher likelihood of PPH prediction compared with CS when applied to the same dataset (OR 1.95, 95% CI 1.88 to 2.01, p<0.001). The I² had a value of 54%, p=0.14, indicating moderate heterogeneity between the studies.

Conclusions ML models are promising for predicting PPH. Nevertheless, they often require a large number of predictors, which may limit their applicability or necessitate automation through digital systems. This poses challenges in resource-scarce settings where the majority of PPH complications occur.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ We followed a rigorous article screening and data extraction process to present a comprehensive analysis of the role of machine learning (ML) compared with conventional statistics (CS) in the prediction of postpartum haemorrhage (PPH).
⇒ An independent quality assessment was performed using a large language model (GPT-4) to analyse Prediction model risk of bias assessment tool for studies developing, validating or updating prediction model items.
⇒ The high variability across studies, stemming from differences in methodology, included subgroups, PPH definitions and predictors, may limit comparison among studies and increase heterogeneity.
⇒ Only two studies directly compared ML models to CS models, which points to the need for more comparative research.

PROSPERO registration number CRD42024521059.

BACKGROUND

According to the WHO, postpartum haemorrhage (PPH) is responsible for more than a quarter of all maternal deaths worldwide. PPH is the leading cause of maternal mortality and accounts for more than 100 000 maternal deaths each year in developing countries.[1]

PPH can be defined as a blood loss (BL) greater than 500 mL after a vaginal delivery or than 1000 mL after a caesarean delivery or as any postpartum vaginal bleeding leading to haemodynamic instability.[2] Its prevention could be achieved through identification of women at highest risk, allowing adherence to prophylactic measures to avoid maternal morbidity and mortality.[3–5]

Although traditional predictive methods have been consistently studied, their predictive results are not satisfactory. A significant portion of haemorrhages (up to 43%) occur in women deemed to be at low risk.

Unfortunately, only 60% of women at high risk for PPH can be identified in advance.[3–5]

Risk factors for PPH have been established, including previous episodes of PPH, multiple gestation, pre-eclampsia, augmented labour, foetal macrosomia, operative vaginal delivery and complex lacerations, among others.[6–9] Other individual risk factors for PPH were identified, but these do not reliably identify women at greatest risk by combining multiple risk factors. A combination of risk factors is common in clinical practice, but quantifying the associated risk without the aid of a clinical prediction model is challenging.[10]

Previous studies suggested that PPH can be predicted using both traditional statistical analyses and machine learning (ML) models.[11] Artificial intelligence can generate rules and patterns based on the input and output data and has the advantage of processing non-additive relationships and incorporating complex interactions between factors that do not need to be pre-specified.[11]

ML approaches might accurately identify women at highest risk of PPH and improve obstetric decision-making and clinical outcomes. This systematic review aims to confirm whether PPH can be predicted with similar discriminative ability using both ML and conventional statistical (CS) models. An additional outcome is to describe the strongest risk factors of PPH identified by ML and traditional statistical models.

METHODS

Protocol and registry

This review adhered to principles outlined in guidance published by Preferred Reporting Items for Systematic Reviews and Meta-analysis[12] and Prediction model risk of bias assessment tool for studies developing, validating or updating prediction models (PROBAST).[13] The protocol for this review has been registered in PROSPERO (ID CRD42024521059).

Search strategy

The literature search was conducted on 4 January 2024 in the following databases: PubMed (MEDLINE), LILACS-BVS, Cochrane Library, Scopus-Elsevier, Embase-Elsevier and Web of Science. The search was restricted to studies published in the last 10 years. A high sensitivity search strategy was developed with the assistance of an experienced librarian for each database, online supplemental table S1.

Eligibility criteria

We included studies that fit the following criteria: (1) assessed early PPH, independently of the definition followed; (2) evaluated PPH prediction; (3) applied accuracy metrics for outcomes evaluation; and (4) clinical trials or observational studies (cross-sectional, cohort, database analysis and case–control). We excluded studies that (1) did not classify excessive bleeding according to a widely accepted definition of PPH or studies that evaluated only severe PPH, excluding non-severe PPH, according to the definition used in which particular study. We included papers in all idioms with abstracts available in English.

Study selection, data extraction and quality of studies evaluation

We used the State of the Art software through Systematic Review[14] to import articles, identify duplicates and select studies according to inclusion and exclusion criteria. Titles and abstracts were independently screened by two reviewers, and any disagreements were resolved by a third reviewer. Data extraction was conducted independently by two reviewers.

We assessed the risk of bias and applicability of each study according to PROBAST tools. The assessment was performed by a reviewer, followed by the use of the large language model (LLM) GPT-4 Open AI[15] support to independently analyse PROBAST items. The GPT-4 analysis was conducted as reported by a prompt design developed by the authors. A senior reviewer resolved conflicts between human and machine judgments. The LLM prompt and evaluation criteria are available in the online supplemental tables S2 and S3.

We evaluated CS models, ML approaches and accuracy metrics (sensitivity, specificity, negative and positive predictive values, Area under the Curve (AUC), F1, precision and recall and other accuracy rates) for PPH prediction. We considered a method as ML according to the author's description. CS was considered as a method of conventional statistics (CS) using the entire database.

The predictors used in the best model of each study were extracted and used to construct a word cloud. To allow for counting the frequency of predictors, terms with similar clinical meaning, such as age and maternal age, or neonatal weight and macrosomia, were homogenised. In these cases, we chose the most appropriate term to represent the predictor. Predictors from models whose studies did not mention them were not included.

Statistical analysis

ORs with 95% CIs were used to compare the odds of PPH prediction using ML models to the odds of PPH prediction using CS approaches when both approaches were evaluated in the same sample population. A pooled OR of 1 would suggest that there is no difference in the odds of predicting PPH between the two methods. The Cochran Q test and I^2 statistics were used to assess for heterogeneity, with a p value less than 0.10 providing evidence of heterogeneity of intervention effects.[16] We performed a random-effects meta-analysis to incorporate heterogeneity among studies. Review Manager V.5.4.1 (Nordic Cochrane Centre, The Cochrane Collaboration, Copenhagen, Denmark) was used for the meta-analysis.[17]

Patients and public involvement

None.

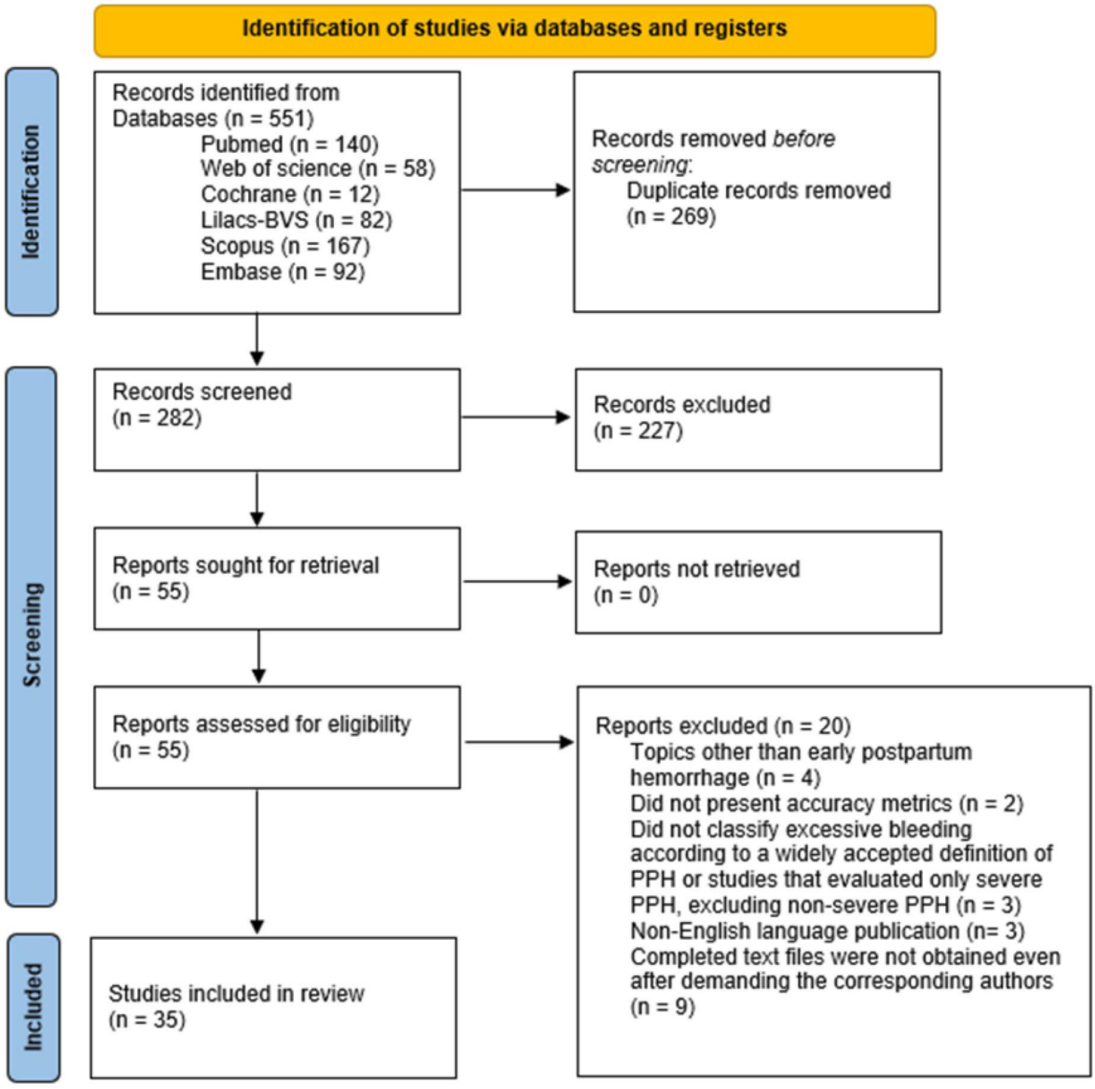

Figure 1 Flowchart with detailed research data for the identified studies for each phase, according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses. PPH, postpartum haemorrhage.

RESULTS

The search yielded 551 papers, and 55 papers were assessed for eligibility after duplicates removal and screening (figure 1). 35 studies fitted inclusion and exclusion criteria.[7 11 18–50]

The synthesis of data gathered 383 648 patients in 24 studies conducting PPH prediction with CS,[18–26 28–30 32 35 36 39 42–44 46–50] 9 studies using ML models[7 11 31 33 34 37 38 41 45] and 2 studies using both methods.[27 40]

Most of the selected articles had a retrospective design, while only five used prospective data collection.[18 20 22 30 50] There were expressive differences in the sampled populations across studies. 11 studies included the general population,[7 11 17 19 24 29–31 40 45 50] while the remaining studies included specific subgroups of analysis as vaginal birth (VB)[20 22 33 34 38 39 47 48] only or specific diseases.[18 23 28 36 41 42 44 46 49] The number of participants in the studies ranged from 72 to 152 279. 118 917 (31%) women were included in studies assessing ML models exclusively, while 264 731 (69%) used only CS. The studies comparing both methods had 51 772 (8.8%) assessed with ML models versus 223 227 (81.2%) with CS.

Outcomes description and measurement also varied among studies. Most studies considered the traditional definition of PPH as an estimated BL of more than 500 mL for a VB or more than 1000 mL for a CS. Some studies considered PPH as cumulative BL greater than 1000 mL with signs and symptoms of hypovolaemia within 24 hours of delivery, regardless of the mode of delivery,[7 18 25 27–29 34 35 41 43] which is in accordance with the 2017 American College of Obstetrics and Gynecologists redefinition.[51] For studies that evaluated the prediction models for both PPH (BL≥500 mL) and severe PPH (BL≥1000 mL) or need for transfusion, we described data only for early PPH in order to allow the comparison between models with lower heterogeneity among findings.[19 21 30] Characteristics of included studies are presented in table 1.

Table 1 Characteristics of included studies

Study	Country	Main objective	Study design	Women (n)	Subgroup	PPH diagnosis criteria
Koopmans *et al*[18]	The Netherlands	PPH prediction	Multicentre clinical trial database	1153	PIH or mild PE>37 weeks	BL>1000 mL 24 hours after delivery
Chi *et al*[19]	China	PPH risk factor screening	Case–control	923	General population	BL>500 mL after delivery
Niepraschk-von *et al*[20]	Germany	Predelivery FIB level for PPH prediction	Prospective cohort	689	VB >37 weeks	BL≥500 mL
Sittiparn and Siwadune[21]	Thailand	Antepartum risk score for PPH prediction	Retrospective cohort	650	VB	BL≥500 mL
Bamberg *et al*[22]	Germany	Antenatal FXIII and FIB levels for PPH prediction	Prospective cohort	548	VB Singleton >37 weeks	Moderate PPH: BL≥500 mL Severe PPH: BL≥1000 mL
Cui *et al*[23]	China	Coagulation markers for PPH prediction	Retrospective cohort	206	HELLP syndrome	BL≥500 mL after VB BL≥1000 mL after CS
Miyoshi and Khondowe[24]	Zambia	Parity values for PPH prediction	Retrospective cohort and survey	1555	General population	BL≥500 mL after VB BL≥1000 mL after CS
Shinohara *et al*[25]	Japan	Recovery time after oxytocin pre treatment for PPH prediction	Retrospective cohort	103	CS following oxytocin	BL>1000 mL
Privitera *et al*[26]	Italy	Predelivery Na+, K+ and Na*K product for PPH prediction	Case–control	130	>37 weeks	BL≥500 mL after VB BL≥1000 mL after CS
Venkatesh *et al*[27]	USA	ML and statistical models for PPH prediction at admission	Retrospective cohort	152 279	≥23 weeks	BL≥1000 mL
Huang *et al*[28]	China	PPH prediction model	Multicentre database	432	ITP	BL>1000 mL or symptoms or signs of hypovolaemia within 24 hours after the birth regardless of route of delivery
Goad *et al*[29]	USA	Antenatal and intrapartum prediction model for PPH	Retrospective cohort study	9774	>20 weeks	BL≥1000 mL
Dodge *et al*[30]	USA	Predelivery FIB for PPH prediction	Nested case–control in cohort	1225	General population	BL>500 mL after VB BL>1000 mL after CS or additional BL of>500 mL in the postpartum period
Pressly *et al*[31]	USA	PPH prediction model	Database	51 000	General population	Based on the International Classification of Diseases codes (9 or 10)

Continued

Table 1 Continued

Study	Country	Main objective	Study design	Women (n)	Subgroup	PPH diagnosis criteria
Pubu *et al*[32]	China	PPH prediction model	Retrospective	4796	>28 weeks	BL>500 mL 24 hours after VB BL>1000 mL after CS
Zhang *et al*[33]	China	Ensemble learning-based PPH prediction model	Retrospective	3842	VB	BL>500 mL after VB
Akazawa *et al*[34]	Japan	ML models for PPH prediction	Database	9894	VB >22 weeks	BL of>1000 mL
Maher *et al*[35]	Ireland	PPH prediction model	Cross-sectional	5807	Singleton	BL≥1000 mL
Westcott *et al*[7]	USA	ML models for PPH prediction	Retrospective cohort	30 867	Age 18–55 years	BL≥1000 mL, regardless of mode of delivery
Zhou *et al*[36]	China	Nomogram for PPH prediction	Retrospective cohort	246	PPP >28 semanas	BL>1000 mL during CS
Krishnamoorthy *et al*[37]	China	ML OBCSA-OSAE model for PPH prediction	Cross-sectional	11 000	Not specified	BL>500 mL 24 hours after VB
Liu *et al*[38]	China	ML model for PPH prediction	Cohort	10 520	VB	BL>500 mL after VB
Jiang *et al*[39]	China	Plasma cytokines model for atonic PPH prediction	Case–control	72	Atonic PPH VB Singleton >37 weeks	BL>500 mL 24 hours after VB
Zheutlin *et al*[40]	USA	Comparative analysis of ML model for PPH prediction with traditional statistical models and existing clinical tools	Retrospective	70 948	General population	BL≥500 mL 24 hours after birth Haemodynamic compromise, use of uterotonics, PPH-related procedures, laboratory changes indicating PPH
Huang *et al*[41]	China	ANN based on maternal blood parameters and clinical indicators for perinatal outcomes prediction	Case–control	270	Singleton PE complicated by FGR	BL>1000 mL
Chu *et al*[42]	China	MRI-based radiomics model for intraoperative CS BL prediction	Cohort	131	PAS CS	BL≥1000 mL
Qi and Fu[43]	China	PPH prediction model	Cohort	2045	Twin pregnancy >28 weeks	BL>1000 mL 24 hours after delivery
Zhong and Zhang[44]	China	PPH prediction model using line graph	Cohort	809	PIH >28 weeks	BL≥500 mL 24 hours after delivery
Shah *et al*[45]	Kenya	ML models to predict PPH	Database	1576	General population	BL>500 mL after VB BL>1000 mL after CS

Continued

Table 1 Continued

Study	Country	Main objective	Study design	Women (n)	Subgroup	PPH diagnosis criteria
Mehrnoush *et al*[11]	Iran	Traditional analytical approach and ML model to predict PPH	Retrospective cohort	8888	General population	BL>500 mL after VB BL>1000 mL after CS or any BL that causes a 10% decrease in haemoglobin levels or a change in vital signs
Lu *et al*[46]	China	MRI parameters (DWI, IVIM and DKI) in PPH prediction	Cross-sectional	109	PAS suspected	BL>1000 mL
Liu *et al*[47]	China	Models for PPH prediction	Retrospective cohort	274	Twin VB	BL≥500 mL after VB BL ≥1000 mL after CS within 24 hours post birth
Yang and Wu[48]	China	PPH prediction model	Retrospective	351	VB Age>35 years	BL>500 mL 24 hours of delivery
Song *et al*[49]	China	Clinical and imaging-based model for PPH prediction	Retrospective cohort	158	PPP >28 weeks	BL≥1000 mL 24 hours after CS
Zhong *et al*[50]	China	Thickness of the uterine muscle layer at placenta attachment for prediction of PPH within 2 hours after delivery	Prospective cohort	378	General population	BL>500 mL after VB BL>1000 mL after CS

ANN, artificial neural network; BL, blood loss; CS, caesarean section; DKI, diffusion kurtosis imaging; DWI, diffusion-weighted imaging; FGR, foetal growth restriction; FIB, fibrinogen; FXIII, factor XIII; HELLP, hemolysis, elevated liver enzyme levels, and low platelet levels; ITP, immune thrombocytopenia; IVIM, intravoxel incoherent motion; K+, serum potassium; ML, machine learning; Na+, serum sodium; OBCSA-OSAE, opposition binary crow search algorithm with an optimal stacked auto encoder; PAS, placenta accreta spectrum; PE, pre-eclampsia; PIH, pregnancy-induced hypertension; PPH, postpartum haemorrhage; PPP, pernicious placenta previa; VB, vaginal birth.

The evaluation of extracted data from studies using only CS (available in the online supplemental table S4) pointed to multivariate regression as the preferred modelling approach to predict PPH occurrence. The number of predictors ranged from 1 to 21 per study. 8 studies evaluated a single predictor as predelivery fibrinogen, factor XIII, parity or use of oxytocin before CS in the prediction of PPH, while 16 studies used combined predictors in a model to predict PPH. There was no association between accuracy and number of predictors or subgroups of analysis. Song *et al*[49] found the highest accuracy (AUC 0.930, 95% CI 0.893 to 0.967) in the prediction of PPH in a subgroup of pernicious placenta previa, and Chi *et al*[19] developed the model with the highest accuracy (AUC 0.868, 95% CI 0.844 to 0.892) for the prediction of PPH in the general population.

The characteristics of studies using ML approaches for PPH prediction are presented in the online supplemental table S5. Most studies developed multiple models using different ML approaches. For the purpose of comparison of predictive model accuracy, we presented in online supplemental table S5 only the model with the best AUC or accuracy metrics. The ML models exhibited varying accuracies, with the AUC ranging from 0.633 to 0.98. The number of features does not seem to be directly related to the accuracy of PPH prediction, but we are unable to run analysis on this matter for concrete results. Our assumption can be noticed on Westcott *et al*,[7] who developed a 212-features initial model with an accuracy of 98.1%, while Mehrnoush *et al*[11] found a similar accuracy using only 11 features.[35 45]

Of the 35 studies included in this review, we identified only 2 studies that compared ML models with CS models for the prediction of PPH. Both studies considered the general population for analysis and compared the accuracy of the methods within the same sampled population. We presented in the online supplemental table S6 a comparative analysis of these studies, with the accuracy

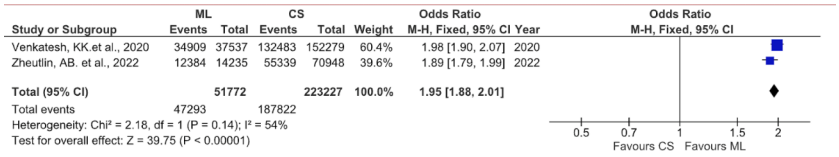

Figure 2 Forest plot of postpartum haemorrhage accuracy prediction by machine learning (ML) and conventional statistics (CS) models.

metrics for the prediction of an estimated BL>1000 mL. Zheutlin *et al*[40] described the accuracy metrics for BL>500 mL, but also presented them for BL>1000 mL in the online supplemental material. Mehrnoush *et al*[11] also performed data analysis with ML and CS but presented the methodology of PPH prediction only for ML models (online supplemental table S5), which is the reason why it was not included in online supplemental table S6).

The pooled OR combining Venkatesh *et al*[27] and Zheutlin *et al*[40] studies resulted in OR 1.95 (95% CI 1.88 to 2.01), figure 2. It means that ML models presented a chance 95% higher of PPH prediction in relation to CS when used in the same database. The I^2 had a value of 54%, p=0.14, indicating moderate heterogeneity between the studies.[15]

Risk factors of PPH identified by ML and traditional statistical models

The most commonly identified predictors were maternal age (10.4%, n=13), hypertensive disorders (8%), body mass index (6.4%), abnormal placentation (6.4%), caesarean section (6.4%), gestational age (5.6%), neonatal weight (5.6%) and platelet count (5.6%). Less frequently assessed were anaemia, parity, haemoglobin and fibrinogen levels (4.8% each), as well as multiple gestation, amniotic fluid and oxytocin (4% each). The 30 most frequent predictors of PPH can be visualised in the word cloud, where its size is associated with its frequency among studies (figure 3) and its frequency can be found in online supplemental table 7.

Quality analysis of included articles

The evaluation of PROBAST items is presented in the online supplemental table S8. On average, the risk of bias seems to be greater among studies using CS (11 of 24 studies presented a high risk of bias vs 3 of 11 studies using ML). However, applicability appears to be lower among studies using ML approaches (5 of 11 studies showed low applicability vs 7 of 24 studies using CS). Among studies included in the meta-analysis, both Venkatesh *et al*[27] and Zheutlin *et al*[40] were deemed to have low risk of bias and high applicability.

All studies included in this review reported the development of prediction models. We did not identify studies validating previously developed models. Of the nine studies using only ML approaches, all reported only internal validation with resampling methods.

The presence and handling of missing data were frequently omitted from analysis, and some ML studies reported using multiple imputation methods to deal with missing data. As GPT-4 was not able to analyse the studies,[48 50] these evaluations were carried out by two independent human reviewers.

The risk of publication bias was not assessed due to the limited number of studies included in the meta-analysis.

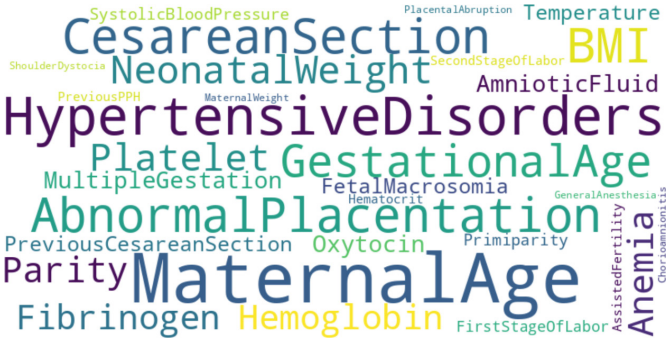

Figure 3 Word Cloud showing the 30 most frequent predictors of postpartum haemorrhage (PPH) among studies. BMI, body mass index.

DISCUSSION

PPH is one of the leading causes of maternal mortality worldwide, accounting for 27.1% and 16.3% of maternal deaths, respectively, in developing and developed regions.[51] Nevertheless, the accurate prediction of PPH remains a challenge.

The main strength of this systematic review is to present the first comprehensive analysis of the accuracy of ML compared with CS in the prediction of PPH. It was noted that PPH can be predicted using both methods, and the findings from this review supported the possibility of using ML models to enhance the accuracy of PPH prediction. No previous reviews were found to compare ML models and traditional statistical models for PPH prediction. Carr *et al*[52] conducted a systematic review of 16 studies using CS models to predict PPH. None of the included studies were considered by the authors to be ready for clinical application, especially due to the absence of external validation and use of variables that are known after birth.[53] In a previous review, Neary *et al* evaluated 14 CS models for PPH prediction and reported some potential for their use only for specific subgroups as CS, placenta accreta spectrum and pernicious placenta previa.[10]

The results obtained from ML models for predicting PPH are promising, since ML approaches appear to be able to predict PPH with greater accuracy than CS. However, the possibility of using a large number of

predictors might limit the use of ML models in the daily hospital routine. Models with a large number of predictors may require automation for their practical implementation, which is particularly limiting in low-middle and lower-middle-income countries.

The main limitation of our review was the high variability among the studies. There were expressive differences in the subgroup of analysis, number of women, definition of PPH, number of predictors and study design, resulting in clinical and methodological diversity that makes it difficult to compare accuracy across studies. Nevertheless, we identified two studies with sufficiently methodological homogeneity and low ROB comparing ML models and CS models in the same sample population and using the same group of analysis and the same definition of PPH. The analysis of these two studies pointed to a 95% higher chance of PPH prediction using ML in relation to CS. We addressed moderate heterogeneity with the random-effect meta-analysis, considering the mean effect across studies. This way, we alert caution for the OR 1.95 interpretation as a limitation.

Venkatesh *et al* assessed 152 279 births to develop predictive models using both CS and ML and reported the extreme gradient boosting model to have the best discriminative ability to predict PPH, after temporal and site validation.[27] ML models were shown to perform the best but at the cost of possibly increased complexity and minimal clinical significance, as CS models had lower, but also good discriminative ability. Zheutlin *et al* compared their ML model with three clinical risk-assessment tools and one previously published model and found the 24-feature ML model to be superior to all the other currently used tools.[40]

Implications for future research

Martins *et al* found vaginal delivery to be a protective factor for PPH, while multiple pregnancy, active bleeding on admission, non-cephalic presentation, retained placenta and placenta abruption were shown to be the strongest risk factors of PPH.[54] These risk factors are in line with those used as predictors described in this review. However, the identification of isolated risk factors has been proven to be insufficient to identify women at high risk of PPH. The use of ML approaches might enhance PPH prediction by analysing complex interactions between combined risk factors and identifying risk patterns associated with PPH.

CONCLUSIONS

The enlarging use of ML models has shown promising results and potential for enhancing the accuracy of PPH prediction. Nevertheless, its applicability or requirement of automation through digital systems still challenges its practical implementation, as well as the limited literature. Additional comparative studies are needed to confirm the potential superiority of ML over CS in predicting PPH. Furthermore, risk factors for PPH have been widely described by multiple previous studies. Despite this, this knowledge is still insufficient to identify women at highest risk for PPH. ML approaches might contribute to improving the prediction of PPH through the analysis of complex nonlinear interactions and identification of new risk patterns.

X Zilma Silveira Nogueira Reis @zilmamedicina

Contributors TB contributed to the conception and design, acquisition of data, analysis and interpretation, manuscript drafting and revision and approval of the final version. JAO contributed to the analysis and interpretation, manuscript drafting and revision and approval of the final version. ALLR contributed to the conception, manuscript revision and approval of the final version. APCdS contributed to the data analysis and interpretation, manuscript revision and approval of the final version. ZSNR contributed to the conception and design, acquisition of data, analysis and interpretation, manuscript drafting and revision and approval of the final version. ZSNR is the guarantor. We assessed the risk of bias and applicability of each study according to PROBAST tools. The assessment was performed by a reviewer, followed by the use of the large language model (LLM) GPT-4 support to independently analyse PROBAST items. The GPT-4 analysis was conducted as reported by a prompt design developed by the authors. A senior reviewer resolved conflicts between human and machine judgements. The LLM prompt and evaluation criteria are available in the online supplemental tables S2 and S3.

Funding This study was funded by the Center for Innovation and Artificial Intelligence in Health (CI-IA Saúde), partially supported by the São Paulo Research Foundation (FAPESP), Grant No. 2020/09866-4, the Minas Gerais Research Foundation (FAPEMIG), Grant No. PPE-00030-21, and UNIMED Belo Horizonte. ZSNR is a researcher with a grant from the Conselho Nacional de Pesquisa (CNPq 305837/2021-4), Brazil. APCdS is a researcher with a grant from the Conselho Nacional de Pesquisa (CNPq 310668/2020-4), Brazil.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as online supplemental information.

ORCID iDs
Thais Baeta http://orcid.org/0000-0002-0486-1911
Zilma Silveira Nogueira Reis http://orcid.org/0000-0001-6374-9295

REFERENCES

1 WHO recommendations: uterotonics for the prevention of postpartum haemorrhage. Geneva World Health Organization; 2018.
2 Committee on Practice Bulletins-Obstetrics. Practice Bulletin No. 183: Postpartum Hemorrhage. *Obstet Gynecol* 2017;130:e168–86.
3 Vogel JP, Williams M, Gallos I, *et al*. WHO recommendations on uterotonics for postpartum haemorrhage prevention: what works, and which one? *BMJ Glob Health* 2019;4:e001466.
4 Ruppel H, Liu VX, Gupta NR, *et al*. Validation of Postpartum Hemorrhage Admission Risk Factor Stratification in a Large Obstetrics Population. *Am J Perinatol* 2021;38:1192–200.
5 Kramer MS, Berg C, Abenhaim H, *et al*. Incidence, risk factors, and temporal trends in severe postpartum hemorrhage. *Am J Obstet Gynecol* 2013;209:449.

6 Oyelese Y, Ananth CV. Postpartum hemorrhage: epidemiology, risk factors, and causes. *Clin Obstet Gynecol* 2010;53:147–56.
7 Westcott JM, Hughes F, Liu W, *et al*. Prediction of Maternal Hemorrhage Using Machine Learning: Retrospective Cohort Study. *J Med Internet Res* 2022;24:e34108.
8 Nyfløt LT, Sandven I, Stray-Pedersen B, *et al*. Risk factors for severe postpartum hemorrhage: a case-control study. *BMC Pregnancy Childbirth* 2017;17:17.
9 Sheldon WR, Blum J, Vogel JP, *et al*. Postpartum haemorrhage management, risks, and maternal outcomes: findings from the World Health Organization Multicountry Survey on Maternal and Newborn Health. *BJOG* 2014;121 Suppl 1:5–13.
10 Neary C, Naheed S, McLernon DJ, *et al*. Predicting risk of postpartum haemorrhage: a systematic review. *BJOG* 2021;128:46–53.
11 Mehrnoush V, Ranjbar A, Farashah MV, *et al*. Prediction of postpartum hemorrhage using traditional statistical analysis and a machine learning approach. *AJOG Glob Rep* 2023;3:100185.
12 Moher D, Liberati A, Tetzlaff J, *et al*. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009;6:e1000097.
13 Moons KGM, Wolff RF, Riley RD, *et al*. PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration. *Ann Intern Med* 2019;170:W1–33.
14 LaPES. StArt (state of the art through systematic review) [software] (laboratório de pesquisa em engenharia de software; 3.0.3 beta) [portuguese; laboratório de pesquisa em engenharia de software]. Universidade Federal de São Carlos (DC/UFSCar); 2020.
15 Wang SH. OpenAI — explain why some countries are excluded from ChatGPT. *Nature New Biol* 2023;615:34.
16 Deeks JJ, Higgins JPT, Altman DG, *et al*. Chapter 10: analysing data and undertaking meta-analyses. In: Higgins JPT, ThomasJ, Chandler J, et al, eds. *Cochrane Handbook for Systematic Reviews of Interventions version 6.4*. Cochrane, Available: www.training.cochrane.org/handbook
17 RevMan whenever its output is used: review manager (revman) [computer program]. The Cochrane Collaboration; 2020.
18 Koopmans CM, van der Tuuk K, Groen H, *et al*. Prediction of postpartum hemorrhage in women with gestational hypertension or mild preeclampsia at term. *Acta Obstet Gynecol Scand* 2014;93:399–407.
19 Chi Z, Zhang S, Wang Y, *et al*. Research of the assessable method of postpartum hemorrhage. *Technol Health Care* 2016;24 Suppl 2:S465–9.
20 Niepraschk-von Dollen K, Bamberg C, Henkelmann A, *et al*. Predelivery maternal fibrinogen as a predictor of blood loss after vaginal delivery. *Arch Gynecol Obstet* 2016;294:745–51.
21 Sittiparn W, Siwadune T. Risk Score for Prediction of Postpartum Hemorrhages in Normal Labor at Chonburi Hospital. *J Med Assoc Thai* 2017;100:382–8.
22 Bamberg C, Mickley L, Henkelmann A, *et al*. The impact of antenatal factor XIII levels on postpartum haemorrhage: a prospective observational study. *Arch Gynecol Obstet* 2019;299:421–30.
23 Cui C, Ma S, Qiao R. Prenatal Plasma Fibrinogen Level Predicts Postpartum Hemorrhage of Patients With HELLP Syndrome. *Clin Appl Thromb Hemost* 2020;26.
24 Miyoshi Y, Khondowe S. Optimal parity cut-off values for predicting postpartum hemorrhage in vaginal deliveries and cesarean sections. *Pan Afr Med J* 2020;37:336.
25 Shinohara S, Okuda Y, Hirata S, *et al*. Association between time from cessation of oxytocin infusion for labor to delivery and intraoperative severe blood loss during cesarean section: a retrospective cohort study. *J Matern Fetal Neonatal Med* 2020;33:1532–7.
26 Privitera AA, Fiore M, Valenti G, *et al*. The role of serum potassium and sodium levels in the development of postpartum hemorrhage. A retrospective study. *It Journ Gyn Obs* 2020;32:126.
27 Venkatesh KK, Strauss RA, Grotegut CA, *et al*. Machine Learning and Statistical Models to Predict Postpartum Hemorrhage. *Obstet Gynecol* 2020;135:935–44.
28 Huang Q-S, Zhu X-L, Qu Q-Y, *et al*. Prediction of postpartum hemorrhage in pregnant women with immune thrombocytopenia: Development and validation of the MONITOR model in a nationwide multicenter study. *Am J Hematol* 2021;96:561–70.
29 Goad L, Rockhill K, Schwarz J, *et al*. Development and validation of a prediction model for postpartum hemorrhage at a single safety net tertiary care center. *Am J Obstet Gynecol MFM* 2021;3:100404.
30 Dodge LE, Carterson AJ, Hacker MR, *et al*. Antepartum fibrinogen concentration as a predictor of bleeding complications. *J Matern Fetal Neonatal Med* 2021;34:3586–90.
31 Pressly MA, Parker RS, Waters JH, *et al*. Improvements and limitations in developing multivariate models of hemorrhage and transfusion risk for the obstetric population. *Transfusion* 2021;61:423–34.
32 Pubu Z-M, Bianba Z-M, Yang G, *et al*. Factors Affecting the Risk of Postpartum Hemorrhage in Pregnant Women in Tibet Health Facilities. *Med Sci Monit* 2021;27:e928568.
33 Zhang Y, Wang X, Han N, *et al*. Ensemble Learning Based Postpartum Hemorrhage Diagnosis for 5G Remote Healthcare. *IEEE Access* 2021;9:18538–48.
34 Akazawa M, Hashimoto K, Katsuhiko N, *et al*. Machine learning approach for the prediction of postpartum hemorrhage in vaginal birth. *Sci Rep* 2021;11:22620.
35 Maher GM, McKernan J, O'Byrne L, *et al*. Predicting risk of postpartum haemorrhage during the intrapartum period in a general obstetric population. *Eur J Obstet Gynecol Reprod Biol* 2022;276:168–73.
36 Zhou Y, Song Z, Wang X, *et al*. Ultrasound-based nomogram for postpartum hemorrhage prediction in pernicious placenta previa. *Front Physiol* 2022;13:982080.
37 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* 2022;22:560.
38 Liu J, Wang C, Yan R, *et al*. Machine learning-based prediction of postpartum hemorrhage after vaginal delivery: combining bleeding high risk factors and uterine contraction curve. *Arch Gynecol Obstet* 2022;306:1015–25.
39 Jiang H, Shi H, Chen L, *et al*. Is there a relationship between plasma, cytokine concentrations, and the subsequent risk of postpartum hemorrhage? *Am J Obstet Gynecol* 2022;226:835.
40 Zheutlin AB, Vieira L, Shewcraft RA, *et al*. Improving postpartum hemorrhage risk prediction using longitudinal electronic medical records. *J Am Med Inform Assoc* 2022;29:296–305.
41 Huang K-H, Chen F-Y, Liu Z-Z, *et al*. Prediction of pre-eclampsia complicated by fetal growth restriction and its perinatal outcome based on an artificial neural network model. *Front Physiol* 2022;13:992040.
42 Chu C, Liu M, Zhang Y, *et al*. MRI-Based Radiomics Analysis for Intraoperative Risk Assessment in Gravid Patients at High Risk with Placenta Accreta Spectrum. *Diagnostics (Basel)* 2022;12:485.
43 Qi S, Fu X. Establishment of a predictive model for postpartum hemorrhage in twins: a retrospective study. *BMC Pregnancy Childbirth* 2023;23:644.
44 Zhong X, Zhang P. Analysis of risk factors associated with different degrees of postpartum hemorrhage in patients with pregnancy-induced hypertension and construction of a prediction model using line graph. *J Matern Fetal Neonatal Med* 2023;36:2239983.
45 Shah SY, Saxena S, Rani SP, *et al*. Prediction of postpartum hemorrhage (PPH) using machine learning algorithms in a Kenyan population. *Front Glob Womens Health* 2023;4:1161157.
46 Lu T, Li M, Li H, *et al*. Diffusion kurtosis and intravoxel incoherent motion in predicting postpartum hemorrhage in patients at high risk for placenta accreta spectrum disorders. *Quant Imaging Med Surg* 2023;13:5921–33.
47 Liu Z, Chen R, Huang H, *et al*. Predicting risk of postpartum hemorrhage associated with vaginal delivery of twins: A retrospective study. *Medicine (Baltimore)* 2023;102:e36307.
48 Yang C, Wu H. Establishment and Validation of Risk Prediction Model for Postpartum Hemorrhage for Pregnant Women ≥35 Years of Age in Natural Delivery. *Altern Ther Health Med* 2023;29:AT9338:876–81:.
49 Song Z, Wang P, Zou L, *et al*. Enhancing postpartum hemorrhage prediction in pernicious placenta previa: a comparative study of magnetic resonance imaging and ultrasound nomogram. *Front Physiol* 2023;14.
50 Zhong H, Zu M, Xie Y, *et al*. The Effect of the Thickness of Uterine Muscle at Placenta Attachment on Postpartum Blood Loss. *Altern Ther Health Med* 2023;10.
51 Hemorrhage P. Postpartum Hemorrhage ACOG PRACTICE BULLET IN Clinical Management Guidelines for Obstetrician–Gynecologists. *Replaces Practice Bulletin Number* 2017;130:76.
52 Carr BL, Jahangirifar M, Nicholson AE, *et al*. Predicting postpartum haemorrhage: A systematic review of prognostic models. *Aust N Z J Obstet Gynaecol* 2022;62:813–25.
53 Say L, Chou D, Gemmill A, *et al*. Global causes of maternal death: a WHO systematic analysis. *Lancet Glob Health* 2014;2:e323–33.
54 Martins RIL, Novais J de SM, Reis ZSN. Postpartum hemorrhage in electronic health records: risk factors at admission and in-hospital occurrence. *Rev Bras Ginecol Obstet* 2024;46:e-rbgo14.