

ORIGINAL RESEARCH

Open access

# Explainable Gradient Boosting Machine for Predicting Postpartum Hemorrhage Risk Using Intrapartum Electronic Fetal Monitoring, Maternal Vital Signs, and Labor Progression Data

Fatima Al-Zahra<sup>1\*</sup>, Amina El Idrissi<sup>1</sup>

## Abstract

Postpartum hemorrhage (PPH) is the leading cause of maternal mortality worldwide, accounting for 25–30% of deaths, particularly in low-resource settings, and early identification of high-risk patients during labor could enable timely interventions such as uterotonic administration, blood preparation, and escalation of care; however, current risk stratification models rely mainly on static antepartum factors and fail to incorporate dynamic intrapartum physiological changes. Existing tools, including those from the California Maternal Quality Care Collaborative, use baseline maternal characteristics such as prior PPH, BMI, parity, and comorbidities, but do not capture continuously evolving labor data, despite intrapartum signals like fetal heart rate patterns, maternal vital sign trends, and labor progression metrics containing rich predictive information that remains underused in real-time decision-making, while clinical judgment is limited by inter-observer variability and inability to integrate complex temporal trends. To address this gap, we propose an explainable gradient boosting machine framework for real-time PPH risk prediction that integrates electronic fetal monitoring parameters (baseline rate, variability, decelerations), maternal vital signs (heart rate, blood pressure, temperature, oxygen saturation), and labor progression features (cervical dilation, contraction frequency, stage duration, and oxytocin use), producing continuously updated risk scores throughout labor. The system combines a gradient boosting model (XGBoost or LightGBM), a SHAP-based explainability module, a real-time feature extraction pipeline, and a clinician-facing dashboard that displays risk scores and key contributing factors, where SHAP provides both global and patient-specific interpretability by identifying how features such as tachysystole or prolonged labor stages influence predictions, thereby improving transparency and clinical trust. Overall, this framework enables dynamic, interpretable PPH risk assessment using routinely collected intrapartum data, combining predictive accuracy with explainability to support earlier detection of hemorrhage risk and more timely, targeted interventions.

**Keywords** Postpartum hemorrhage, Explainable artificial intelligence, Gradient boosting machine, SHAP, Electronic fetal monitoring, Intrapartum prediction

\*Correspondence:

Fatima Al-Zahra  
fatima.zahra@gmail.com

<sup>1</sup> Department of Healthcare Analytics and AI Systems, University of Casablanca, Casablanca, Morocco

## Introduction

Postpartum hemorrhage, defined as cumulative blood loss exceeding 500 milliliters following vaginal delivery or 1,000

milliliters after cesarean section—or any blood loss causing hemodynamic instability—represents the foremost preventable cause of maternal mortality, responsible for one-quarter to one-third of maternal deaths globally [1, 2].

The condition complicates approximately 1–5% of deliveries in high-resource settings, with substantially higher rates in low- and middle-income countries where access to uterotonic agents, blood products, and surgical intervention may be limited [3]. Beyond mortality, PPH contributes to severe maternal morbidity including hypovolemic shock, disseminated intravascular coagulation, acute renal failure, and Sheehan syndrome, underscoring the imperative for early recognition and intervention [4]. Despite decades of clinical awareness and protocolized management, hemorrhage-related deaths have not declined proportionally to other causes of maternal mortality, highlighting persistent gaps in risk prediction and timely response.

The etiological framework of PPH is classically organized around the "Four Ts": Tone, accounting for 70–80% of cases through uterine atony; Trauma, encompassing genital tract lacerations, uterine rupture, and uterine inversion; Tissue, including retained placenta and intrauterine clots; and Thrombin, referring to acquired or congenital coagulopathies [5, 6]. Crucially, many of these underlying mechanisms manifest or intensify during labor rather than being identifiable prenatally, rendering static admission-time risk assessments inherently incomplete [7]. Uterine atony risk increases with prolonged labor, chorioamnionitis, oxytocin use patterns, and uterine overdistension—factors that evolve dynamically throughout the intrapartum period [8]. Similarly, traumatic hemorrhage risk escalates with operative vaginal delivery, episiotomy extension, and prolonged second stage, none of which can be predicted with certainty at admission. This temporal evolution of risk necessitates predictive approaches that update continuously as labor progresses.

Current clinical risk assessment paradigms, exemplified by the California Maternal Quality Care Collaborative toolkit and similar instruments, stratify patients based predominantly on static prenatal characteristics: prior PPH history, parity, multiple gestation, hypertensive disorders, body mass index, and hematocrit at admission [9, 10]. While these tools identify patients with pre-existing risk factors, they exhibit limited discrimination for the majority of PPH cases occurring in patients without traditional risk factors [11]. Moreover, they cannot incorporate the rich temporal data generated during labor—electronic fetal monitoring tracings, maternal heart rate and blood pressure trends, cervical dilation trajectories, and oxytocin infusion parameters—that collectively signal developing pathophysiology [12]. The resulting clinical gap—subjective

assessment supplemented by static risk scores—fails to leverage the full information content of modern intrapartum monitoring.

We propose an explainable gradient boosting machine framework for real-time PPH risk prediction that integrates three continuously updated data streams: electronic fetal monitoring parameters capturing uterine activity and fetal response, maternal vital sign trajectories reflecting hemodynamic compensation or decompensation, and labor progression metrics quantifying the mechanical and pharmacological aspects of parturition [13]. By embedding SHAP-based explainability directly into the prediction pipeline, the framework generates not only risk scores but also clinician-interpretable explanations specifying which specific features drove each prediction [14].

## Background

### Postpartum hemorrhage etiology

The "Four Ts" mnemonic—Tone, Trauma, Tissue, and Thrombin—provides a comprehensive etiological framework for postpartum hemorrhage that guides both clinical management and predictive modeling [15, 16]. Uterine atony (Tone) represents the predominant mechanism, occurring when the myometrium fails to contract adequately after placental separation, leaving uterine spiral arteries unoccluded and permitting continued blood loss [1]. Risk factors for atony include uterine overdistension from multiple gestation or polyhydramnios, prolonged labor with myometrial fatigue, chorioamnionitis impairing contractile responsiveness, and pharmacological agents including halogenated anesthetics and tocolytics [2]. Genital tract trauma (Trauma) encompasses cervical and vaginal lacerations, episiotomy extensions, and the rare but catastrophic uterine rupture, with risk amplified by operative vaginal delivery and macrosomia [17]. Retained placental tissue (Tissue) prevents complete uterine contraction and is associated with placenta accreta spectrum, succenturiate lobes, and manual placental extraction [18]. Coagulation disorders (Thrombin), whether pre-existing as in von Willebrand disease or acquired through dilutional coagulopathy and disseminated intravascular coagulation, impair hemostatic competence and exacerbate bleeding from any source [3].

### Intrapartum electronic fetal monitoring

Electronic fetal monitoring (EFM) continuously records fetal heart rate patterns and uterine contraction activity, generating data streams that reflect both fetal well-being and uterine function relevant to PPH prediction [19, 20]. Standard EFM parameters include baseline fetal heart rate (normal range 110–160 beats per minute), baseline variability (absent, minimal 0–5 bpm, moderate 6–25 bpm, or marked >25 bpm), and periodic decelerations classified as early (mirroring contractions), variable (abrupt onset, V-shaped), or late (onset after contraction peak) [21]. While EFM interpretation traditionally focuses on fetal acid-base status prediction, uterine contraction parameters—frequency, duration, resting tone, and the presence of tachysystole defined as more than five contractions in a ten-minute window—carry direct relevance to postpartum hemorrhage risk through their relationship with uterine muscle fatigue and atony development [10, 11]. Prolonged decelerations and recurrent late decelerations may signal uterine hyperstimulation, particularly in the context of oxytocin augmentation, which independently increases atony risk through receptor desensitization [22].

## Maternal vital signs during labor

Serial maternal vital sign measurements during labor provide a window into hemodynamic status, sympathetic nervous system activation, and developing pathophysiological processes that precede clinically overt hemorrhage [13, 23]. Maternal heart rate elevation (tachycardia, typically >100 beats per minute) represents one of the earliest compensatory responses to hypovolemia, preceding measurable blood pressure changes in the spectrum of hemorrhagic shock [24]. Conversely, progressive blood pressure decline signals decompensation from compensated shock, with systolic blood pressure below 90 mmHg or mean arterial pressure below 65 mmHg indicating clinically significant volume loss [12]. Temperature elevation above 38°C raises suspicion for chorioamnionitis, which independently increases uterine atony risk through inflammatory mediator effects on myometrial contractility [14]. The shock index—calculated as heart rate divided by systolic blood pressure—integrates these two parameters into a single metric; values exceeding 0.9 in the obstetric population correlate with increased transfusion requirements and adverse outcomes, with higher sensitivity than either vital sign alone [5].

## Labor progression

Labor progression metrics quantify the mechanical and temporal dimensions of parturition that directly influence uterine muscle physiology and PPH risk [8, 25]. Cervical dilation rate, measured in centimeters per hour during the active phase, reflects the efficiency of uterine contractions and the resistance of cervical tissue; protracted labor with dilation below established thresholds (1.2 cm/hour for nulliparas, 1.5 cm/hour for multiparas) indicates dysfunctional labor potentially requiring prolonged oxytocin exposure [26]. The duration of the second stage of labor—from complete cervical dilation to delivery—carries particular significance, with nulliparous patients exceeding two hours and multiparous patients exceeding one hour demonstrating increased rates of uterine atony, operative delivery, and postpartum hemorrhage [15]. Contraction frequency and pattern, including tachysystole and hypertonus, quantify uterine work and predict subsequent atony through mechanisms of muscle fatigue and receptor downregulation [6]. Oxytocin administration parameters—total dose, maximum infusion rate, and duration of exposure—modulate uterine responsiveness and independently predict hemorrhage risk, with prolonged high-dose infusion associated with oxytocin receptor desensitization and reduced endogenous contractile capacity after delivery [9].

## Framework Overview

### High-level architecture

The proposed framework implements a streaming prediction pipeline that continuously ingests intrapartum monitoring data, extracts clinically meaningful features, generates PPH risk scores, and provides SHAP-based explanations through a clinician-facing interface [16, 27]. Data acquisition modules interface with electronic fetal monitors via standard output protocols (e.g., Philips IntelliSpace Perinatal, GE Centricity Perinatal) to capture fetal heart rate and uterine contraction waveforms, while maternal vital signs flow from automated non-invasive blood pressure cuffs, pulse oximeters, and temperature probes at configurable intervals [28]. Labor progression events—cervical examination findings, membrane rupture status, oxytocin infusion rates, and delivery details—are extracted from structured electronic health record fields or entered directly by clinicians [17]. The feature extraction engine transforms raw waveforms and discrete events into the structured feature vectors described in Section 5, which then pass to a trained gradient boosting model (XGBoost or

LightGBM) for risk score computation. SHAP explainers process each prediction in parallel, decomposing the risk score into positive and negative feature contributions displayed on a real-time dashboard updated every 15–30 minutes [23]. This architecture ensures that predictions and explanations remain temporally synchronized with the patient's evolving clinical status, enabling truly dynamic risk assessment.

Figure 1 presents the proposed explainable intrapartum machine-learning architecture linking streaming fetal monitoring, maternal vital signs, and labor progression data to SHAP-interpretable postpartum hemorrhage risk prediction and clinical action readiness.

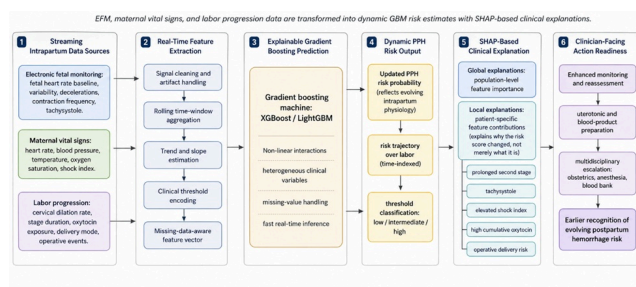


Figure 1. Explainable real-time intrapartum gradient boosting framework for dynamic postpartum hemorrhage risk prediction.

### Core assumptions

The framework operates under several foundational assumptions regarding data availability, quality, and clinical context that constrain its applicability and inform implementation requirements [4, 7]. First, we assume continuous or near-continuous electronic fetal monitoring throughout active labor, consistent with guidelines from the American College of Obstetricians and Gynecologists for high-risk patients and common practice during oxytocin-augmented labor; settings where intermittent auscultation predominates would require protocol modifications [22]. Second, maternal vital signs must be recorded at regular intervals—typically every 15–30 minutes during active labor—with electronic capture rather than manual charting to minimize latency and transcription errors [18]. Third, the framework requires a training dataset containing linked intrapartum monitoring data with verified postpartum hemorrhage outcomes, applying standardized blood loss quantification methods (quantitative blood loss measurement rather than visual estimation) to ensure label quality [1]. Fourth, we assume integration with an electronic

health record system capable of receiving and displaying real-time risk scores within existing clinical workflows without introducing excessive cognitive burden or alert fatigue [24].

### Design principles

The framework design is governed by four principles that prioritize clinical utility and patient safety while maintaining technical rigor [21, 25]. Real-time operation demands that feature extraction, prediction, and explanation computation complete within the 15-minute update cycle, imposing computational efficiency constraints on model complexity and feature dimensionality [18]. Interpretability constitutes the central design requirement, mandating that every risk prediction be accompanied by human-understandable explanations specifying which physiological features contributed most to the score and in which direction—thus enabling clinicians to verify predictions against their clinical judgment rather than accepting algorithmic outputs blindly [14]. Non-invasive data acquisition leverages sensors already standard in modern labor and delivery units, avoiding additional patient burden, infection risk, or workflow disruption that would impede adoption [26]. Finally, actionability requires that risk scores and explanations map clearly to established clinical protocols: a high-risk alert must suggest specific interventions (uterotonic preparation, blood product activation, senior clinician notification) rather than providing vague warnings that clinicians cannot operationalize [13].

Table 1 links each intrapartum data modality to the physiological signals and hemorrhage mechanisms it contributes to the proposed prediction framework.

Table 1. Conceptual Mapping between Intrapartum Data Modalities, Pathophysiological Signals, and PPH Risk Mechanisms

Intrapartum modality	Representative features	Physiological or clinical signal captured	PPH mechanism informed
Electronic fetal monitoring	Contraction frequency, tachysystole, contraction duration, resting tone	Uterine workload and possible myometrial fatigue	Tone: uterine atony

Electronic fetal monitoring	Late decelerations, prolonged decelerations, reduced variability	Possible uteroplacental stress, hyperstimulation, or maternal hemodynamic compromise	Tone: uterine activity system
Maternal vital signs	Heart rate, systolic blood pressure, mean arterial pressure	Compensatory or decompensating hemodynamic state	Thrombotic system instability early signs
Maternal vital signs	Shock index	Integrated tachycardia–hypotension relationship	Severe hemorrhage preparation
Maternal vital signs	Temperature elevation	Possible chorioamnionitis or inflammatory stress	Tone: impaired myometrial contraction
Labor progression	Cervical dilation rate, active-phase duration, second-stage duration	Labor efficiency, cumulative uterine work, prolonged exertion	Tone: Trauma

Labor progression	Oxytocin dose, maximum infusion rate, exposure duration	Pharmacological stimulation and possible receptor desensitization	Tone: uterine activity
Labor progression	Operative delivery, episiotomy, laceration risk, macrosomia	Mechanical trauma and delivery complexity	Trauma

## Gradient Boosting Machine Algorithm selection

Gradient boosting machines offer particular advantages for clinical prediction tasks that align with the requirements of real-time PPH risk assessment: they naturally handle heterogeneous feature types (continuous vital signs, categorical deceleration patterns, ordinal variability grades) without requiring extensive preprocessing or normalization [18, 19]. XGBoost and LightGBM, the two leading implementations, both provide native handling of missing values through sparse-aware split finding, addressing the reality that intrapartum data collection includes inevitable gaps from intermittent monitoring, technical artifacts, and clinical workflow interruptions [28]. These algorithms automatically learn non-linear relationships and high-order feature interactions, capturing phenomena such as the synergistic effect of tachysystole combined with prolonged second stage on uterine atony risk, without requiring explicit specification of interaction terms [29]. Fast inference—sub-millisecond per prediction—ensures that risk score computation imposes negligible latency even when processing multiple concurrent laboring patients, while built-in feature importance metrics provide a foundation for subsequent SHAP-based explainability [27]. The regularized objective functions of both XGBoost and LightGBM, incorporating L1 and L2 penalties, mitigate overfitting when applied to moderate-sized clinical datasets with numerous candidate features [20].

## Training setup

Model training requires a curated dataset of historical deliveries with complete intrapartum monitoring records and verified PPH outcomes, preferably sourced from large multicenter databases such as the Consortium on Safe Labor or institutional data warehouses [5, 15]. The binary outcome is defined as postpartum hemorrhage—blood loss exceeding 500 mL for vaginal delivery or 1,000 mL for cesarean, or clinician-diagnosed hemorrhage with intervention—with quantification preferably based on calibrated drape measurement or gravimetric methods rather than visual estimation [11]. Given the low PPH incidence (approximately 1–5% of deliveries), class imbalance poses a significant challenge requiring explicit mitigation: we apply `scale_pos_weight` parameter adjustment in XGBoost or `class_weight` in LightGBM to penalize false negatives proportionally to prevalence, and utilize area under the precision-recall curve (AUPRC) rather than accuracy as the primary training metric [3]. Hyperparameter optimization employs Bayesian search (e.g., Hyperopt or Optuna) over learning rate (0.01–0.3), maximum tree depth (3–8), subsample ratio (0.6–1.0), and regularization parameters, with five-fold cross-validation stratified by PPH outcome and temporal ordering to prevent data leakage [24]. The final model is trained on the full development dataset with optimized hyperparameters, then evaluated on held-out temporal validation data as described in Section 8 [6].

## Feature Engineering

### EFM features

Electronic fetal monitoring features quantify both fetal condition and uterine activity patterns relevant to hemorrhage prediction, extracted from raw cardiocotographic signals using signal processing techniques applied to 30-minute windows preceding each prediction update [10, 22]. Baseline fetal heart rate is computed as the rounded mean heart rate excluding accelerations and decelerations over the preceding 10-minute segment, with categories designating normal (110–160 bpm), bradycardic (<110 bpm), or tachycardic (>160 bpm) states [21]. Baseline variability is classified using NICHD criteria into absent (undetectable amplitude), minimal ( $\leq 5$  bpm), moderate (6–25 bpm), or marked (>25 bpm), with minimal or absent variability potentially signaling fetal compromise from uteroplacental insufficiency or maternal hypotension [23]. Deceleration features include

the count of late decelerations (onset after contraction peak, gradual return), variable decelerations (abrupt onset,  $\geq 15$  bpm drop lasting  $\geq 15$  seconds), and prolonged decelerations ( $\geq 2$  minutes but <10 minutes) within the window, each extracted using automated detection algorithms validated against expert annotation [27]. Uterine contraction parameters—frequency measured in contractions per 10 minutes, mean duration, resting tone between contractions, and the presence of tachysystole (exceeding 5 contractions in 10 minutes averaged over 30 minutes)—capture uterine workload that directly relates to subsequent atony risk through myometrial fatigue mechanisms [15].

### Maternal vital signs features

Maternal vital sign features capture hemodynamic trends and physiological compensation that may signal developing hemorrhage or predisposing conditions before clinically apparent blood loss occurs [12, 13]. Heart rate features include the most recent single measurement, the mean over the preceding 60 minutes, and the trend direction (increasing, stable, or decreasing) computed through linear regression slope, with tachycardia exceeding the 95th percentile for laboring patients (approximately 100–110 bpm) flagged as an elevated-risk indicator [24]. Systolic and diastolic blood pressure are processed analogously, with particular attention to decreasing trajectories that may reflect progressive volume loss; mean arterial pressure (calculated as diastolic + one-third pulse pressure) is included as an integrated perfusion metric [5]. The shock index—heart rate divided by systolic blood pressure—with established obstetric thresholds of  $\geq 0.9$  for concern and  $\geq 1.0$  for severe risk, consolidates compensatory tachycardia and hypotensive trends into a single interpretable feature shown to predict transfusion requirements and massive hemorrhage in obstetric populations [14]. Temperature features flag current fever ( $\geq 38.0^\circ\text{C}$ ) or temperature increases  $\geq 0.5^\circ\text{C}$  over the preceding two hours, both associated with chorioamnionitis and consequent uterine atony, while oxygen saturation trends below 95% may indicate respiratory compensation for metabolic acidosis from hypoperfusion [9].

### Labor progression features

Labor progression features quantify the mechanical progress of parturition and pharmacological interventions that modulate uterine physiology, extracted from structured electronic health record entries [8, 26]. Cervical dilation

metrics include the most recent examination value, the rate of dilation over the preceding two hours (cm/hour), and binary indicators for protracted active phase or arrest of dilation per established labor curves, with slower progression associated with prolonged oxytocin exposure and increased atony risk [18]. Stage durations—time since onset of active phase (6 cm dilation), duration of second stage from complete dilation stratified by parity, and third stage duration from delivery to placental expulsion—provide temporal context reflecting cumulative uterine work [26]. Oxytocin features capture the total cumulative dose administered since labor onset (milliunits), current infusion rate (milliunits/minute), maximum infusion rate reached, and duration of oxytocin exposure in hours, all contributing to receptor desensitization and atony risk. Delivery-related features include mode of delivery (spontaneous vaginal, vacuum-assisted, forceps-assisted, or cesarean), presence and degree of perineal laceration or episiotomy, and birth weight (grams), with operative delivery and macrosomia (>4,000 g) increasing both traumatic and atonic hemorrhage risk [7]. The combination of prolonged second stage with high cumulative oxytocin and operative delivery creates particularly elevated risk profiles that gradient boosting models detect through their intrinsic interaction learning [18].

## Explainability (SHAP)

### Global interpretability

SHAP (SHapley Additive exPlanations) values decompose gradient boosting predictions into additive feature contributions using Shapley values from cooperative game theory, providing a unified interpretability framework [23, 24]. Global feature importance is derived by averaging absolute SHAP values across all patients, ranking intrapartum parameters by their consistent influence on PPH predictions—unlike traditional split-count metrics, SHAP captures both effect magnitude and directionality [27, 28]. Clinically, the top-ranked features should align with established risk factors: prolonged second stage, tachysystole, cumulative oxytocin dose, shock index elevation, and operative delivery should dominate if the model has learned meaningful relationships [15, 16]. Any discrepancy between SHAP rankings and known predictors signals potential data quality issues, confounding, or novel associations warranting investigation, making global interpretability both a validation tool and a discovery mechanism [29].

### Local (Per-Patient) explanations

Local SHAP explanations address the essential clinical question of why a specific patient received a particular risk score by quantifying each feature's contribution relative to the population baseline [14, 23]. The SHAP force plot visualizes this decomposition: for a nulliparous patient with prolonged second stage, tachysystole, and elevated shock index, the plot might show baseline risk of 5% plus 12% from prolonged second stage, plus 8% from tachysystole, plus 5% from shock index, minus contributions from protective features, yielding a final risk of 25% [1, 24]. This granular explanation enables clinicians to verify model reasoning against their own assessment and target modifiable contributing factors [25]. SHAP respects local accuracy—the sum of all contributions plus baseline exactly equals the prediction—ensuring explanations faithfully represent the model's decision process rather than offering post-hoc rationalizations, while temporal evolution of explanations across sequential predictions illuminates changing risk profiles during labor [6, 27].

## Clinical Decision Support

### Real-time risk dashboard

The clinician-facing dashboard translates predictions and explanations into an actionable interface for the high-cognitive-load labor environment, displaying the current PPH risk score with color-coded bands: green (<10%), yellow (10–30%), and red (>30%), configurable to institutional protocols [5, 13, 17, 26]. The top contributing features appear alongside directional arrows and SHAP contribution magnitudes, while a trend panel plots risk trajectory over the preceding four hours to distinguish escalating from stable or decreasing risk [22, 23]. The dashboard refreshes automatically every 15–30 minutes and supports on-demand updates after clinical events such as membrane rupture, epidural placement, or arrest of dilation [9]. Integration with existing fetal monitoring displays and electronic health records minimizes new screen space requirements, embedding risk assessment within tools clinicians already use [15].

### Alerting and action pathways

Risk thresholds trigger escalating responses mapped to established obstetric emergency protocols: intermediate risk (10–30%) prompts increased vital sign monitoring to every 15 minutes, verification of uterotonic availability, and

type-and-screen confirmation [3, 7, 11]. High risk (>30%) triggers simultaneous alerts to obstetrics, anesthesiology, and blood bank, activating massive transfusion protocol and preparing second-line uterotonics [12]. Both tiers include SHAP-identified contributing features, enabling clinicians to address modifiable drivers—reducing oxytocin for tachysystole, considering operative delivery for prolonged second stage, or administering antibiotics for fever [2]. A clinician acknowledgment mechanism ensures alerts are received, while audit logs create feedback loops for system improvement and documentation [24]. Thresholds are customizable through institutional governance, balancing sensitivity against specificity to prevent alarm fatigue, with initial recommendations derived from validation dataset receiver operating characteristic analysis [21].

**Table 2** demonstrates how SHAP-derived feature contributions can be translated into clinically interpretable explanations and proportionate bedside responses.

**Table 2.** Clinical Interpretability and Actionability Matrix for SHAP-Guided PPH Risk Prediction

SHAP-identified driver	Direction of risk contribution	Clinical interpretation	Possible bedside response
Prolonged second stage	Increases risk	Cumulative uterine fatigue and higher probability of operative delivery	Senior obstetric review; prepare uterotonics; reassess delivery plan
Tachysystole	Increases risk	Excessive uterine activity may indicate hyperstimulation and later atony vulnerability	Reassess oxytocin infusion; increase monitoring; evaluate fetal and maternal status
Elevated shock index	Increases risk	Maternal hemodynamic strain or early	Repeat vitals; ensure IV access;

		compensatory instability	prepare blood products if risk remains high
High cumulative oxytocin exposure	Increases risk	Possible receptor desensitization and impaired postpartum contractility	Review dose history; anticipate uterine atony; prepare second-line uterotonic
Fever or rising temperature	Increases risk	Possible chorioamnionitis-associated atony risk	Evaluate infection; consider antibiotics; increase postpartum vigilance
Stable maternal vitals	Decreases risk	No current evidence of hemodynamic compromise	Continue routine monitoring unless other risk drivers escalate
Normal labor progression	Decreases risk	Efficient labor without prolonged uterine workload	Maintain standard readiness while monitoring trend changes
Absence of tachysystole with moderate variability	Decreases risk	No evidence of excessive contraction burden or fetal compromise	Continue standard EFM surveillance

## Evaluation Strategy

### Prediction metrics

Model evaluation employs metrics suited to PPH's imbalanced nature, where standard accuracy would mislead [6, 19]. Primary discrimination metrics include AUROC and AUPRC, the latter penalizing false alarms more appropriately when positive cases are rare [28]. At clinically selected operating points, we report sensitivity, specificity, positive predictive value, and negative predictive value [27]. Calibration is assessed via Brier score and calibration plots, ensuring predicted probabilities match observed frequencies—essential when predictions inform threshold-based decisions [18]. The F1 score balances false negatives and false positives, guiding selection of an operating point that minimizes both missed hemorrhages and unnecessary interventions [20].

## Feature importance validation

SHAP-derived rankings undergo clinical validation to confirm alignment with obstetric knowledge, constituting "explainability-based validation" beyond statistical metrics [23, 29]. An obstetric panel reviews the top 20 features for biological plausibility, consistency with PPH literature, and potential confounding [16]. Concordance—prolonged second stage, tachysystole, oxytocin dose, and shock index ranking prominently—indicates clinically meaningful learning rather than spurious correlation [5, 8]. Discordant features trigger investigation for data artifacts such as proxies for cesarean delivery, encoding errors, or novel pathophysiological associations [13]. Stability analysis across temporal splits, parity subgroups, delivery modes, and hyperparameter perturbations identifies unreliable rankings suggesting overfitting [15].

## Temporal validation

Temporal validation assesses generalization by training on earlier data (e.g., 2017–2022) and testing on later data (2023–2024), simulating real-world deployment where historical data must predict future outcomes [15, 17, 26]. Performance degradation relative to cross-validation quantifies temporal dataset shift from evolving clinical protocols, demographics, documentation patterns, or PPH definitions [3]. Subgroup analyses across parity, gestational age, delivery mode, and oxytocin exposure identify populations requiring recalibration or exclusion [9]. The framework also yields realistic estimates of alert frequency and positive predictive value under contemporary conditions, informing acceptable false-alarm rates and resource allocation for triggered interventions [21].

# Limitations

## Technical limitations

Technical constraints affecting reliability include EFM interpretation variability—inter-observer kappa values of 0.4–0.6 for deceleration and variability assessment introduce label noise when training on clinician-annotated patterns [10, 19, 22]. Missing data from irregular vital sign recording, EFM signal dropout, and episodic cervical examinations requires imputation strategies whose assumptions about missingness mechanisms may not universally hold [27]. The low PPH base rate (1–5%) fundamentally limits positive predictive value; even excellent discrimination (AUROC 0.85) yields modest precision given low prevalence, potentially generating more false alarms than true positives [6]. These issues compound the single-institution development paradigm; performance may degrade substantially across different populations, equipment, or protocols without external validation [16].

## Clinical limitations

Clinical implementation risks unintended consequences requiring prospective evaluation [13, 26]. Elevated risk scores may paradoxically increase interventions if clinicians respond with preemptive operative deliveries or oxytocin escalation, creating self-fulfilling prophecies where predicted risk leads to actions causing the outcome [17]. Alarm fatigue from excessive false alarms desensitizes clinicians, necessitating rigorous threshold calibration balancing sensitivity against alert frequency [11, 24]. Continuous EFM reliance excludes settings using intermittent auscultation, low-resource environments, and home births [21]. Workflow integration demands substantial investment in interoperability, training, and quality assurance, raising cost-effectiveness questions relative to simpler interventions like standardized blood loss measurement [4]. Medicolegal implications—liability for missed hemorrhages despite low predicted risk, documentation standards for algorithm-generated scores—require professional society guidance before adoption [7].

# Conclusion

This framework presents an explainable gradient boosting machine approach for real-time postpartum hemorrhage risk prediction integrating three continuously updated intrapartum data modalities: electronic fetal monitoring

parameters, serial maternal vital signs, and labor progression metrics. By embedding SHAP-based explainability directly into the prediction pipeline, the framework generates dynamic risk scores alongside patient-specific explanations decomposing each prediction into individual feature contributions. The approach leverages routinely collected monitoring data without additional invasive sensors, and the computational efficiency of gradient boosting machines enables 15–30 minute updates throughout active labor.

The key advantages include dynamic risk updating reflecting evolving physiology rather than static admission assessment, transparent and quantifiable explanations clinicians can verify against their own judgment, and real-time actionability through integration with established obstetric emergency protocols. SHAP-based global interpretability reveals which intrapartum features most consistently drive predictions, serving as both validation mechanism and potential discovery tool. Per-patient local explanations answer why a specific risk score was generated, supporting shared decision-making and targeted interventions. The design principles—real-time operation, interpretability, non-invasive data acquisition, and clinical actionability—reflect the requirements of high-stakes environments where opaque predictions are neither acceptable nor safe.

Important limitations temper immediate deployment enthusiasm. Data quality depends on EFM interpretation accuracy, vital sign documentation completeness, and labor progression recording consistency—all variable in routine practice. The low PPH base rate creates inherent sensitivity–positive predictive value trade-offs, with even well-calibrated models generating substantial false-positive alerts risking clinician desensitization. External validation across diverse populations, equipment, and protocols remains essential, as single-institution datasets may not capture full physiological and practice variability. Unintended consequences—increased intervention rates,

alarm fatigue, and medicolegal complexity—necessitate prospective evaluation addressing clinical outcomes beyond prediction accuracy.

Validation on large multicenter obstetric databases such as the Consortium on Safe Labor or the California Perinatal Quality Care Collaborative represents the essential next step toward clinical translation. Such validation should assess statistical discrimination and calibration alongside clinical utility: whether risk predictions and explanations change clinician behavior to reduce hemorrhage severity, transfusion requirements, and maternal morbidity without increasing unnecessary interventions. Ultimately, the goal is augmenting obstetric clinicians with interpretable, evidence-based risk information enabling earlier recognition and more timely response to developing postpartum hemorrhage, not replacing clinical judgment with algorithmic decision-making.

## Acknowledgements

None

## Conflict of interest

None

## Financial support

None

## Ethics statement

None

Received: 26 Feb 2025 Revised: 13 May 2025 Accepted: 04 Jul 2025

Published online: 20 January 2026

## Rights and permissions

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Venkatesh KK, Strauss RA, Grotegut CA, Heine RP, Chescheir NC, Stringer JS, et al. Machine learning and statistical models to predict postpartum hemorrhage. *Obstet Gynecol*. 2020;135(4):935-44.
- Katsuhiko N, Akazawa M, Hashimoto K, Kaname Y. Machine learning approach for the prediction of postpartum hemorrhage in vaginal birth. *Sci Rep*. 2021;11(1):22620.
- 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(1):560.
- 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(2):100185.
- Boujarzadeh B, Ranjbar A, Banihashemi F, Mehrnoush V, Darsareh F, Saffari M. Machine learning approach to predict postpartum haemorrhage: a systematic review protocol. *BMJ Open*. 2023;13(1):e067661.
- Ranjbar A, Ghamsari SR, Boujarzadeh B, Mehrnoush V, Darsareh F. Predicting risk of postpartum hemorrhage using machine learning approach: a systematic review. *Gynecol Obstet Clin Med*. 2023;3(3):170-4.
- Wen T. Interpretable machine learning predicts postpartum hemorrhage at time of admission. *Am J Obstet Gynecol MFM*. 2024;6(8):101396.
- Lengerich BJ, Caruana R, Painter I, Weeks WB, Sitcov K, Souter V. Interpretable machine learning predicts postpartum hemorrhage with severe maternal morbidity in a lower-risk laboring obstetric population. *Am J Obstet Gynecol MFM*. 2024;6(8):101391.
- Ahmadzia HK, Dzienny AC, Bopf M, Phillips JM, Federspiel JJ, Amdur R, et al. Machine learning models for prediction of maternal hemorrhage and transfusion: model development study. *JMIR Bioinform Biotechnol*. 2024;5(1):e52059.
- 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*. 2024;51(3):60.
- Song Z, Lin H, Shao M, Wang X, Chen X, Zhou Y, et al. Integrating SHAP analysis with machine learning to predict postpartum hemorrhage in vaginal births. *BMC Pregnancy Childbirth*. 2025;25(1):529.
- Peleg D, Nofal MA, Shachar IB. Postpartum hemorrhage after vaginal delivery [12E]. *Obstet Gynecol*. 2019;133(Suppl 1):54S.
- Lérias-Cambeiro M, Mugeiro-Silva R, Rodrigues A, Dias-Domingues T, Lança F, Vaz Carneiro A. Enhancing postpartum haemorrhage prediction through the integration of classical logistic regression and machine learning algorithms. *Mathematics*. 2025;13(21):3376.
- Mathewlynn SJ, Soltaninejad M, Collins SL. Artificial intelligence and postpartum hemorrhage. *Matern Fetal Med*. 2025;7(1):22-8.
- Hoodbhoy Z, Noman M, Shafique A, Nasim A, Chowdhury D, Hasan B. Use of machine learning algorithms for prediction of fetal risk using cardiotocographic data. *Int J Appl Basic Med Res*. 2019;9(4):226-30.
- Cömert Z, Şengür A, Budak Ü, Kocamaz AF. Prediction of intrapartum fetal hypoxia considering feature selection algorithms and machine learning models. *Health Inf Sci Syst*. 2019;7(1):17.
- Esteban-Escano J, Castan B, Castan S, Cholz-Ezquerro M, Asensio C, Laliena AR, et al. Machine learning algorithm to predict acidemia using electronic fetal monitoring recording parameters. *Entropy*. 2021;24(1):68.
- Francis F, Luz S, Wu H, Stock SJ, Townsend R. Machine learning on cardiotocography data to classify fetal outcomes: a scoping review. *Comput Biol Med*. 2024;172:108220.
- Gunaratne SA, Panditharatne SD, Chandrarahan E. Prediction of neonatal acidosis based on the type of fetal hypoxia observed on the cardiotocograph (CTG). *Eur J Med Health Sci*. 2022;4(2):8-18.
- O'Sullivan ME, Considine EC, O'Riordan M, Marnane WP, Rennie JM, Boylan GB. Challenges of developing robust AI for intrapartum fetal heart rate monitoring. *Front Artif Intell*. 2021;4:765210.
- Ben M'Barek I, Jauvion G, Vitrou J, Holmström E, Koskas M, Ceccaldi PF. DeepCTG® 1.0: an interpretable model to detect fetal hypoxia from cardiotocography data during labor and delivery. *Front Pediatr*. 2023;11:1190441.
- Kuzu A, Santur Y. Early diagnosis and classification of fetal health status from a fetal cardiotocography dataset using

ensemble learning. *Diagnostics (Basel)*. 2023;13(15):2471.

Melaet R, de Vries IR, Kok RD, Oei SG, Huijben IA, van Sloun RJ, et al. Artificial intelligence based cardiotocogram assessment during labor. *Eur J Obstet Gynecol Reprod Biol*. 2024;295:75-85.

Lovers A, Daumer M, Frasch MG, Ugwumadu A, Warrick P, Vullings R, et al. Advancements in fetal heart rate monitoring: a report on opportunities and strategic initiatives for better intrapartum care. *BJOG*. 2025;132(7):853-66.

Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell*. 2020;2(1):56-67.

Lundberg SM, Lee SI. A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst*. 2017;30:4765-74.

Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. LightGBM: a highly efficient gradient boosting decision tree. *Adv Neural Inf Process Syst*. 2017;30:3146-54.

Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A. CatBoost: unbiased boosting with categorical features. *Adv Neural Inf Process Syst*. 2018;31:6638-48.

McCoy JA, Levine LD, Wan G, Chivers C, Teel J, La Cava WG, et al. Intrapartum electronic fetal heart rate monitoring to predict acidemia at birth with the use of deep learning. *Am J Obstet Gynecol*. 2025;232(1):116.e1-116.e12.