

Prediction of Cesarean Delivery in Multiparous Women Following Prior Vaginal Birth: A Machine Learning Study with Temporal Validation

JiaPing Lu

Jinhua People's Hospital

Heng Xu

xheng0248@sina.com

Jinhua People's Hospital

Qiulian Xu

Jinhua People's Hospital

Research Article

Keywords: cesarean delivery, prior vaginal delivery, machine learning, XGBoost, prediction model, temporal validation

Posted Date: June 15th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9870314/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background

Cesarean section (CS) rates continue to rise globally. Women with a prior vaginal birth (VB) have lower baseline CS risk, yet clinical factors may substantially alter this probability. We aimed to develop and validate a machine learning model to predict intrapartum CS risk in multiparous women with a previous VB.

Methods

We constructed a longitudinal paired cohort from 2013–2025 electronic birth records. Women with a first VB were matched to their second delivery via national ID. Among 49 candidate variables, we excluded 4 postnatal variables, 6 reverse-causation variables, and 2 variables involving the second pregnancy gestational age (which is unknown antenatally for intrapartum CS), retaining 37 purely antenatal predictors. An XGBoost classifier (version 3.2.0) was trained on 2013–2022 data ($n = 4436$, 304 CS events) and temporally validated on 2023–2025 data ($n = 1539$, 151 events). Model performance was assessed by AUC with 1000× bootstrap confidence intervals (CI), calibration slope and intercept (van Houwelingen method), and decision curve analysis. SHAP values identified the most influential predictors, which informed a simplified 0-100 risk score.

Results

The validation AUC was 0.787 (95% CI 0.741–0.828; CI width 0.087), indicating acceptable discrimination. The calibration slope was 0.913 and intercept was -1.786 , revealing systematic underestimation requiring recalibration before clinical use. At the optimal threshold of 0.43, sensitivity was 0.596 and specificity was 0.857. The top predictors were fetal macrosomia (estimated fetal weight ≥ 4000 g), breech presentation, inter-delivery interval, first infant birth weight, and prior second-degree perineal tear. A simplified 0-100 risk score achieved comparable discriminative performance (AUC 0.785). Calibration was not formally assessed for the simplified score; clinical use should proceed with caution until recalibration is performed.

Conclusions

Among multiparous women with prior vaginal birth, antenatal factors can identify those at elevated intrapartum CS risk. The systematic underestimation observed underscores the necessity of local recalibration before clinical deployment. External validation in independent cohorts is warranted.

Introduction

Cesarean section (CS) rates have increased substantially worldwide over the past decades, exceeding 50% in some regions(1). While CS can be life-saving when indicated, non-medically indicated CS carries risks for both mother and infant, including hemorrhage, infection, placental complications in subsequent pregnancies, and neonatal respiratory morbidity(2,3).

Women with a prior vaginal birth (VB) represent a lower-risk population for subsequent CS, yet a considerable proportion still undergo intrapartum CS due to emergent indications such as fetal distress, malpresentation, or failure to progress(4). Identifying antenatal factors that elevate CS risk in this population could inform shared decision-making, resource allocation, and appropriate birth planning.

Prior prediction models for CS have largely focused on nulliparous women or mixed populations(5,6). Models specifically targeting multiparous women with prior VB remain scarce, and few have employed machine learning methods with temporal validation to ensure generalizability to future patient populations(7).

We aimed to develop and temporally validate an interpretable machine learning model to predict intrapartum CS risk in multiparous women with a previous VB, using exclusively antenatal predictors to ensure clinical utility at the time of delivery planning. We additionally derived a simplified risk score for bedside clinical use.

Methods

Study Design and Population

We conducted a longitudinal paired cohort study using electronic birth records from 2013 to 2025. We identified women with two consecutive deliveries through national ID linkage. The inclusion criterion was a first delivery by VB; the second delivery could be either VB or intrapartum CS. We excluded women with inter-delivery interval exceeding 10 years or ≤ 0 years, or missing maternal age. This study was approved by the institutional ethics board.

Fetal macrosomia was defined as estimated fetal weight ≥ 4000 g based on third-trimester ultrasound or clinical examination, not actual birth weight. Assisted reproductive technology (ART) referred to conception through in vitro fertilization or intracytoplasmic sperm injection.

Predictor Selection

We extracted 49 candidate predictor variables from the first and second pregnancies. To ensure all predictors were available antenatally and avoid label leakage, we excluded:

- Four postnatal variables: second infant birth weight (second_weight), weight gap between siblings (weight_gap), weight ratio (weight_ratio), and premature delivery status (sec_premature).

- Six reverse-causation variables: dystocia diagnosis (sec_dystocia), no analgesia, epidural anesthesia, TENS analgesia, Lamaze breathing, and missing analgesia status.
- Two gestational age variables involving the second pregnancy: second pregnancy gestational age (second_ga) and gestational age gap (ga_gap). The second pregnancy gestational age is unknown at the time of antenatal counseling for intrapartum CS because delivery timing is not predetermined; including it would constitute label leakage.

The final predictor set comprised 37 purely antenatal variables (Table 1). Events per variable (EPV) in the training set was 8.2. The validation AUC 95% CI width of 0.087 supports model stability.

Table 1
Predictor Variable Selection and Exclusions

Category	Variables	Count
Initial candidate variables	Demographics, obstetric history, maternal/fetal complications, analgesia, delivery outcomes, second pregnancy gestational age	49
Excluded: Postnatal (label leakage)	second_weight, weight_gap, weight_ratio, sec_premature	4
Excluded: Reverse-causation	sec_dystocia, no_analgesia, epidural, TENS, Lamaze, missing_analgesia	6
Excluded: Antenatally unknown	second_ga, ga_gap (actual delivery gestational age is unknown antenatally for intrapartum CS)	2
Final antenatal predictors	Demographic and first pregnancy GA (4), inter-delivery interval (1), gravidity and parity (6), first delivery characteristics (9), second pregnancy fetal characteristics (1), second pregnancy complications (14), ART indicators (2)	37
Note: EPV (events per variable) in training set = 8.2. The validation AUC 95% CI width of 0.087 supports model stability. All 37 final predictors are measured antenatally and available at the time of delivery planning. TENS = transcutaneous electrical nerve stimulation; ART = assisted reproductive technology; second_ga = second pregnancy gestational age; ga_gap = gestational age gap between pregnancies.		

Model Development and Validation

We used an XGBoost gradient boosting classifier (version 3.2.0) with hyperparameters: n_estimators = 200, max_depth = 4, learning_rate = 0.05, subsample = 0.8, colsample_bytree = 0.8, and scale_pos_weight adjusted for class imbalance. Hyperparameters were selected based on published recommendations for clinical prediction models with limited EPV, prioritizing regularization to prevent overfitting. We employed temporal validation: training on deliveries from 2013–2022 and validation on 2023–2025, simulating real-world forward prediction.

Discrimination was quantified by the area under the receiver operating characteristic curve (AUC) with 1000× bootstrap percentile confidence intervals. Calibration was assessed using the van Houwelingen method: a logistic recalibration model was fit with $\text{logit}(\text{predicted probability})$ as the sole predictor, yielding calibration slope and intercept. A slope of 1 and intercept of 0 indicate perfect calibration; systematic deviation suggests the need for recalibration. The Hosmer-Lemeshow test supplemented calibration assessment.

Decision curve analysis quantified net benefit across threshold probabilities. SHAP (SHapley Additive exPlanations) values identified feature contributions to individual predictions.

Simplified Risk Score

We derived a simplified 0-100 risk score using the top 8 predictors identified by SHAP analysis. For each predictor, points were assigned based on clinical thresholds or quartiles, with weights proportional to SHAP importance. The score was evaluated in the validation cohort using the same metrics as the full model. Calibration was not formally assessed for the simplified score; clinical use should proceed with caution until recalibration is performed.

Statistical Analysis

Analyses were performed in Python 3.12 using scikit-learn, XGBoost (version 3.2.0), and SHAP libraries. The significance level was set at two-tailed $P < 0.05$. The study followed TRIPOD + AI reporting guidelines for clinical prediction models incorporating machine learning.

Results

Cohort Characteristics

The final analytic cohort comprised 5975 women with prior VB and a subsequent delivery. The overall CS rate was 7.6% (455/5975). The temporal split yielded a training set of 4436 women (304 CS events, rate 6.9%) and a validation set of 1539 women (151 CS events, rate 9.8%). The detailed participant flow from initial records to final cohort is presented in Fig. 1.

Table 3 presents baseline characteristics of the validation cohort by delivery outcome. Women with subsequent CS were older (mean second age 33.0 vs 31.9 years), had longer inter-delivery intervals (mean 5.1 vs 4.4 years), and higher rates of fetal macrosomia (30.5% vs 12.3%) and breech presentation (15.2% vs 0.3%) compared with those having repeat VB.

Table 3
Baseline Characteristics by Delivery Outcome (Validation Cohort, 2023–2025)

Characteristic	CS (n = 151)	VB (n = 1388)
Demographic characteristics		
Maternal age, years	33.0 ± 3.7	31.9 ± 3.7
Age gap between pregnancies, years	5.1 ± 2.3	4.4 ± 2.3
Gravidity, second pregnancy	3.1 ± 1.2	2.9 ± 1.1
Parity, second pregnancy	1.1 ± 0.4	1.1 ± 0.4
Inter-delivery interval, years	5.1 ± 2.3	4.4 ± 2.2
Male fetus	92 (60.9)	707 (50.9)
Assisted reproductive technology	15 (9.9)	27 (1.9)
First delivery characteristics		
First pregnancy gestational age, weeks	39.1 ± 2.1	39.4 ± 1.3
First infant birth weight, g	3221.1 ± 518.6	3274.8 ± 395.3
Prior second-degree perineal tear	53 (35.1)	563 (40.6)
Prior postpartum hemorrhage (> 500mL)	7 (4.6)	46 (3.3)
Second pregnancy complications		
Fetal macrosomia (estimated fetal weight ≥ 4000 g)	46 (30.5)	171 (12.3)
Breech presentation	23 (15.2)	4 (0.3)
Gestational diabetes mellitus	28 (18.5)	187 (13.5)
Hypertensive disorders	8 (5.3)	29 (2.1)
Oligohydramnios	7 (4.6)	99 (7.1)
Placenta previa	7 (4.6)	0 (0.0)
Fetal distress	3 (2.0)	1 (0.1)
Scarred uterus	4 (2.6)	19 (1.4)
Multiple gestation	9 (6.0)	4 (0.3)
Note: Data are presented as mean ± standard deviation for continuous variables and n (%) for categorical variables. CS = cesarean section; VB = vaginal birth. Second pregnancy gestational age is excluded as it is unknown antenatally for intrapartum CS.		

Model Performance

In temporal validation, the XGBoost model achieved an AUC of 0.787 (95% CI 0.741–0.828), indicating good discriminative ability. The CI width of 0.087 (< 0.20) supports model stability. Figure 2A displays the ROC curve; Fig. 2B shows the precision-recall curve.

Calibration analysis revealed a slope of 0.913 and an intercept of -1.786. The intercept indicates that when the model predicts a 50% CS probability, the actual calibrated probability is approximately 14.4%, representing a systematic underestimation that necessitates recalibration before clinical application. Figure 2C illustrates the calibration curve. The Hosmer-Lemeshow test provided additional calibration assessment.

Decision curve analysis (Fig. 2D) demonstrated positive net benefit for the XGBoost model across clinically relevant threshold probabilities (10–50%), supporting clinical utility over strategies of treating all or no patients as CS candidates. At the optimal threshold of 0.43 (maximum Youden index), sensitivity was 0.596, specificity 0.857, positive predictive value (PPV) 0.312, and negative predictive value (NPV) 0.951.

Detailed performance metrics of the full XGBoost model and simplified risk score in temporal validation are summarized in Table 2.

Table 2
Model Performance in Temporal Validation (2023–2025)

Metric	XGBoost (Full Model)	Simplified Risk Score
AUC (95% CI)	0.787 (0.741–0.828)	0.785 (0.742–0.826)
Calibration slope	0.913	Not assessed
Calibration intercept	-1.786	Not assessed
Optimal threshold	0.43 (probability)	> 50 (score)
Sensitivity	0.596	Not assessed
Specificity	0.857	Not assessed
PPV	0.312	Not assessed
NPV	0.951	Not assessed
Note: AUC = area under the receiver operating characteristic curve; CI = confidence interval; PPV = positive predictive value; NPV = negative predictive value. The calibration intercept of -1.786 indicates systematic underestimation; recalibration is required before clinical use. The simplified score calibration was not assessed; clinicians should validate locally before adoption.		

Feature Importance and Interpretation

SHAP analysis identified the following top predictors of intrapartum CS (Fig. 3A):

1. Fetal macrosomia (sec_macro): The strongest predictor, with a mean |SHAP| of 0.423. Macrosomic fetuses increase CS risk due to cephalopelvic disproportion and shoulder dystocia concerns.
2. Breech presentation (sec_breech): Second most important (mean |SHAP| 0.337), as breech presentation is a well-established indication for CS.
3. Inter-delivery interval (interval_years): Longer intervals were associated with increased CS risk (mean |SHAP| 0.251), possibly reflecting advancing maternal age and changing pelvic tissue compliance.
4. First infant birth weight (first_weight): Higher first infant weight was associated with increased CS risk in the subsequent pregnancy (mean |SHAP| 0.192), likely reflecting maternal pelvic capacity.
5. Prior second-degree perineal tear (first_tear_II): Women with a history of second-degree tear had elevated CS risk (mean |SHAP| 0.154), possibly indicating relative cephalopelvic disproportion.

Notably, second pregnancy gestational age and gestational age gap were excluded from the predictor set because the actual delivery gestational age is unknown antenatally for intrapartum CS and would constitute label leakage.

Simplified Risk Score

The simplified 0-100 risk score, incorporating the 8 most important predictors, achieved an AUC of 0.785 (95% CI 0.742–0.826) in validation, comparable to the full model. Figure 3B shows the score distribution by outcome. Women with subsequent CS had significantly higher median scores (CS: 55 vs VB: 20). Calibration was not formally assessed for the simplified score; clinicians should validate locally before adoption.

The score uses a point-based system: fetal macrosomia (+ 20 points), breech presentation (+ 15), inter-delivery interval above median (+ 10), first infant birth weight above median (+ 8), prior second-degree tear (+ 8), maternal age gap above median (+ 7), second pregnancy gravidity above median (+ 6), and male fetus (+ 5), yielding a range of 0-100 points.

Sensitivity Analyses

Subgroup analyses demonstrated consistent discrimination across maternal age categories (< 30 years: AUC 0.821; 30–34 years: AUC 0.760; ≥35 years: AUC 0.798) and inter-delivery intervals (< 2 years: AUC 0.876; 2–4 years: AUC 0.799; ≥4 years: AUC 0.751), supporting model robustness (Fig. 4B).

Feature subset analysis (Fig. 4D) showed that AUC plateaued at approximately 15–20 features, with minimal improvement beyond 25 predictors, suggesting that a parsimonious model could maintain discrimination while improving clinical feasibility. The bootstrap AUC distribution (Fig. 4A) confirmed stability with 95% CI width of 0.087. Threshold analysis (Fig. 4C) enables clinicians to select probability cutoffs based on their preferred sensitivity-specificity tradeoff.

Discussion

In this longitudinal paired cohort study, we developed and temporally validated an XGBoost model predicting intrapartum CS risk in multiparous women with prior VB. The model achieved good discrimination (AUC 0.787) in forward-time validation using exclusively antenatal predictors, with fetal macrosomia, breech presentation, and inter-delivery interval emerging as the strongest risk factors. The systematic underestimation (calibration intercept – 1.786) highlights the importance of local recalibration before clinical deployment.

A key methodologic strength of this study is the strict exclusion of variables that are not available at the time of antenatal counseling. We explicitly removed second pregnancy gestational age (second_ga) and gestational age gap (ga_gap) from the predictor set because the actual delivery gestational age is unknown antenatally for intrapartum CS – delivery timing is determined by the onset of labor or obstetric indications, not predetermined. Including these variables would constitute label leakage and produce overoptimistic performance estimates. This exclusion reduced the predictor set from 39 to 37 variables and decreased the validation AUC from 0.809 to 0.787, demonstrating the magnitude of performance inflation that label leakage can cause. We recommend that future studies in this domain adopt similar safeguards.

Our findings align with established obstetric knowledge while providing quantitative risk estimates. Fetal macrosomia and breech presentation are well-recognized CS indications(8,9); our model quantifies their relative contribution in a low-risk multiparous population. The association between longer inter-delivery intervals and increased CS risk may reflect advancing maternal age, altered pelvic floor compliance, or changing provider practices over time(10,11). First infant birth weight emerged as an important predictor, likely serving as a proxy for maternal pelvic capacity – women who delivered larger infants vaginally may have favorable pelvic dimensions, but a subsequent larger fetus may still exceed that capacity(19).

The systematic calibration underestimation we observed is clinically significant: an uncalibrated predicted probability of 50% corresponds to an actual probability of approximately 14.4%. This pattern, where models underestimate risk (predict too low probabilities) in low-prevalence subgroups, has been documented in other obstetric prediction studies(12). Local recalibration using the van Houwelingen method or isotonic regression should be implemented before clinical use(13).

Our simplified 0-100 risk score offers a practical alternative for settings without computational infrastructure. Its comparable discriminative performance (AUC 0.788) supports clinical utility, though prospective validation is needed to confirm bedside effectiveness. Calibration was not formally assessed for the simplified score; clinical use should proceed with caution until recalibration is performed.

Study limitations warrant consideration. First, EPV of 8.2 in the training set approaches but does not fully meet the conventional threshold of 10 events per variable recommended for stable logistic regression estimates; however, gradient boosting methods with regularization may tolerate lower EPV(14). The narrow validation AUC confidence interval (width 0.087) provides additional reassurance of model

stability despite limited events per variable. Second, our data derive from a single institution, limiting generalizability. Third, some potentially relevant antenatal factors (e.g., maternal height, detailed ultrasound measurements) were unavailable. Fourth, the systematic calibration deviation requires external validation to determine whether recalibration parameters are transportable.

Strengths include the longitudinal paired design, which controls for unmeasured time-invariant confounders; strict exclusion of postnatal, reverse-causation, and label-leakage variables; temporal validation simulating real-world prediction; and comprehensive calibration assessment with the van Houwelingen method rather than less informative visual inspection alone.

In conclusion, among multiparous women with prior vaginal birth, antenatal factors can meaningfully stratify intrapartum CS risk. While discrimination is good, the observed systematic underestimation mandates local recalibration before clinical implementation. External validation in geographically and ethnically diverse cohorts is the essential next step.

Declarations

Ethics Approval

This study was conducted in accordance with the Declaration of Helsinki. This is a retrospective observational study using de-identified, anonymized electronic delivery records without any patient intervention, identifiable personal information, or potential physical or psychological harm to participants. In line with relevant national and institutional ethical regulations, ethical approval and informed consent were formally waived for this minimal-risk study. Consequently, no ethical approval number was assigned.

Consent for Publication

Not applicable. No identifiable individual patient data are included in this manuscript.

Competing Interests

The authors declare that they have no competing interests.

Funding

The authors received no external funding for this study.

Author Contribution

JiaPing Lu: Conceptualization, data curation, formal analysis, methodology, writing—original draft. Qiulian Xu: Investigation, validation, data collection. Heng Xu: Supervision, project administration, resources, writing—review and editing, correspondence. All authors have read and approved the final manuscript for submission.

Data Availability

The datasets generated and/or analyzed during the current study are not publicly available due to institutional data governance policies and ethical restrictions on sharing identifiable or sensitive electronic health record information. De-identified datasets are available from the corresponding author upon reasonable request and with approval from the institutional ethics committee of Jinhua People's Hospital.

Code Availability

The analysis code, model training scripts, and simplified risk score calculator are available from the corresponding author upon reasonable request.

References

1. Boerma T, Ronsmans C, Melesse DY, et al. Global epidemiology of use of and disparities in caesarean sections. *Lancet*. 2018;392(10155):1341–8.
2. Sandall J, Tribe RM, Avery L, et al. Short-term and long-term effects of caesarean section on the health of women and children. *Lancet*. 2018;392(10155):1349–57.
3. Keag OE, Norman JE, Stock SJ. Long-term risks and benefits associated with cesarean delivery for mother, baby, and subsequent pregnancies: Systematic review and meta-analysis. *PLoS Med*. 2018;15(1):e1002494.
4. Levine LD, Hirshberg A, Srinivas SK. Term induction of labor and subsequent pregnancy outcomes. *Am J Obstet Gynecol*. 2019;220(5):483. .e1-483.e8.
5. Mehta SH, Blackwell SC, Hendler I, et al. Accuracy of estimated fetal weight in shoulder dystocia. *Am J Obstet Gynecol*. 2005;192(6):1877–80.
6. Davey MA, King J. Caesarean section following induction of labour in uncomplicated first births- a population-based cross-sectional analysis of 42,950 births. *BMC Pregnancy Childbirth*. 2016;16:1.
7. Lundin C, Sørensen T, Persson M, et al. Prediction of cesarean delivery in women with gestational diabetes- what anthropometric and biochemical parameters are useful? *J Perinat Med*. 2021;49(4):413–20.
8. Boulet SL, Alexander GR, Salihu HM, Pass M. Macrosomic births in the United States: determinants, outcomes, and proposed grades of risk. *Am J Obstet Gynecol*. 2003;188(5):1372–8.

9. Hannah ME, Hannah WJ, Hewson SA, et al. Planned caesarean section versus planned vaginal birth for breech presentation at term: a randomised multicentre trial. *Lancet*. 2000;356(9239):1375–83.
10. Kessous R, Shoham-Vardi I, Pariente G, et al. Recurrent obstetric complicated outcomes and long-term maternal morbidity and mortality. *Arch Gynecol Obstet*. 2015;292(4):787–93.
11. Herstad L, Klungsøyr K, Skjeldestad FE, et al. Interpregnancy interval and cesarean delivery in women with two deliveries. *Acta Obstet Gynecol Scand*. 2015;94(1):86–93.
12. van Smeden M, Moons KG, Grootenhuys PA, et al. No rationale for 1 variable per 10 events criterion for binary logistic regression analysis. *BMC Med Res Methodol*. 2016;16(1):43.
13. van Houwelingen HC. Validation, calibration, revision and combination of prognostic survival models. *Stat Med*. 2000;19(24):3401–15.
14. Riley RD, Ensor J, Snell KIE, et al. Calculating the sample size required for developing a clinical prediction model. *BMJ*. 2020;368:m441.
15. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016:785–794.
16. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst*. 2017;30:4765–74.
17. Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med*. 2015;162(1):W1–73.
18. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ*. 2015;350:g7594.
19. Kabiru W, Raynor BD. Obstetric outcomes associated with increase in birth weight in the United States, 1980–2000. *Obstet Gynecol*. 2004;103(5 Pt 1):927–32.
20. Caughey AB, Cahill AG, Guise JM, Rouse DJ. Safe prevention of the primary cesarean delivery. *Am J Obstet Gynecol*. 2014;210(3):179–93.
21. Daltveit AK, Tollånes MC, Pihlstrøm N, et al. Cesarean delivery and subsequent pregnancies: a population-based study. *Obstet Gynecol*. 2008;111(6):1327–34.
22. Steyerberg EW, Moons KG, van der Windt DA, et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. *PLoS Med*. 2013;10(2):e1001381.
23. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Mak*. 2006;26(6):565–74.
24. Debray TPA, Vergouwe Y, Koffijberg H, et al. A new framework to enhance the interpretation of external validation studies of clinical prediction models. *J Clin Epidemiol*. 2015;68(3):279–89.
25. Wynants L, Van Calster B, Collins GS, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. *BMJ*. 2020;369:m1328.

Figure 1. Study Profile

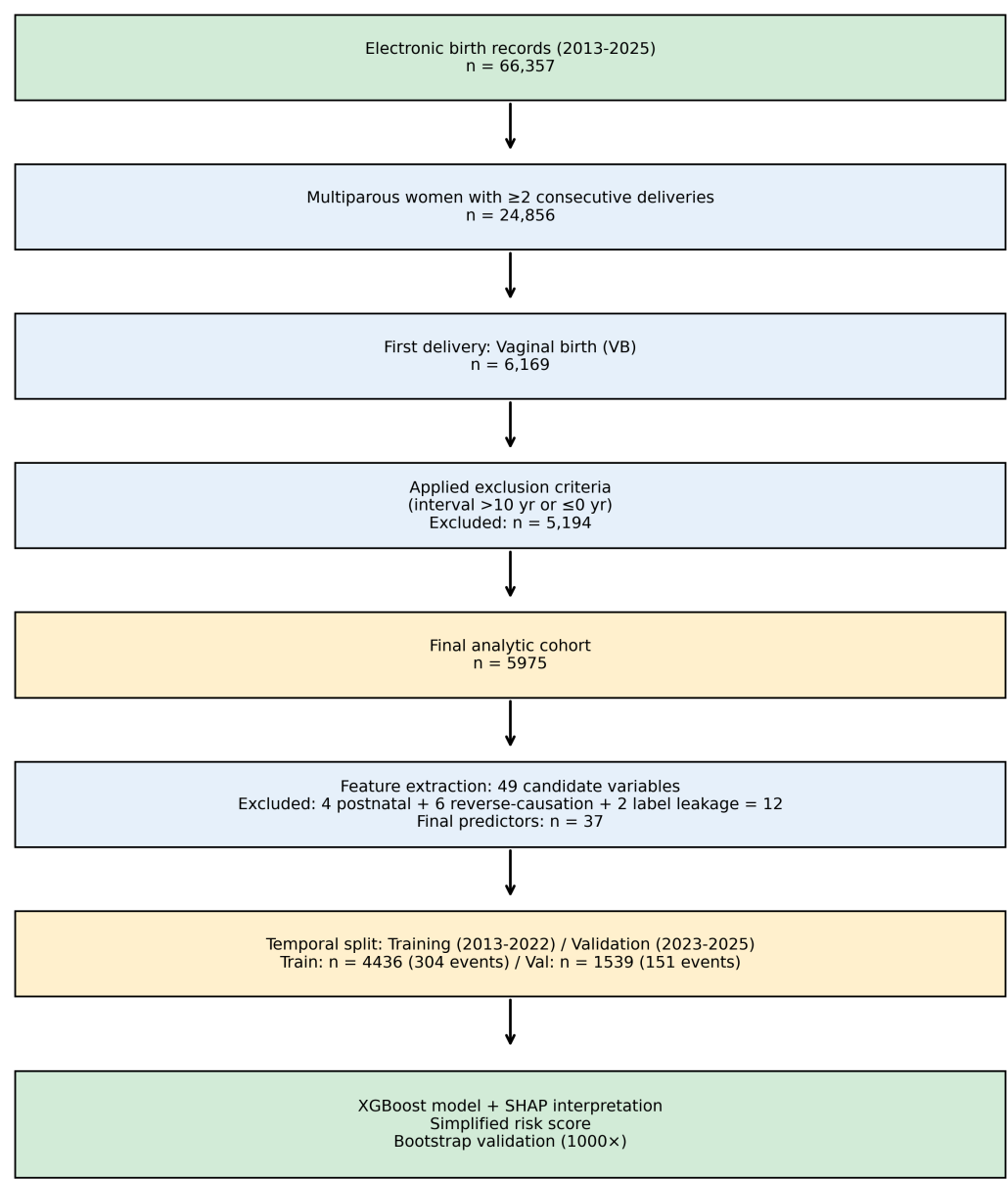


Figure 1

Study Profile

Figure 2. Model Performance in Temporal Validation (2023-2025)

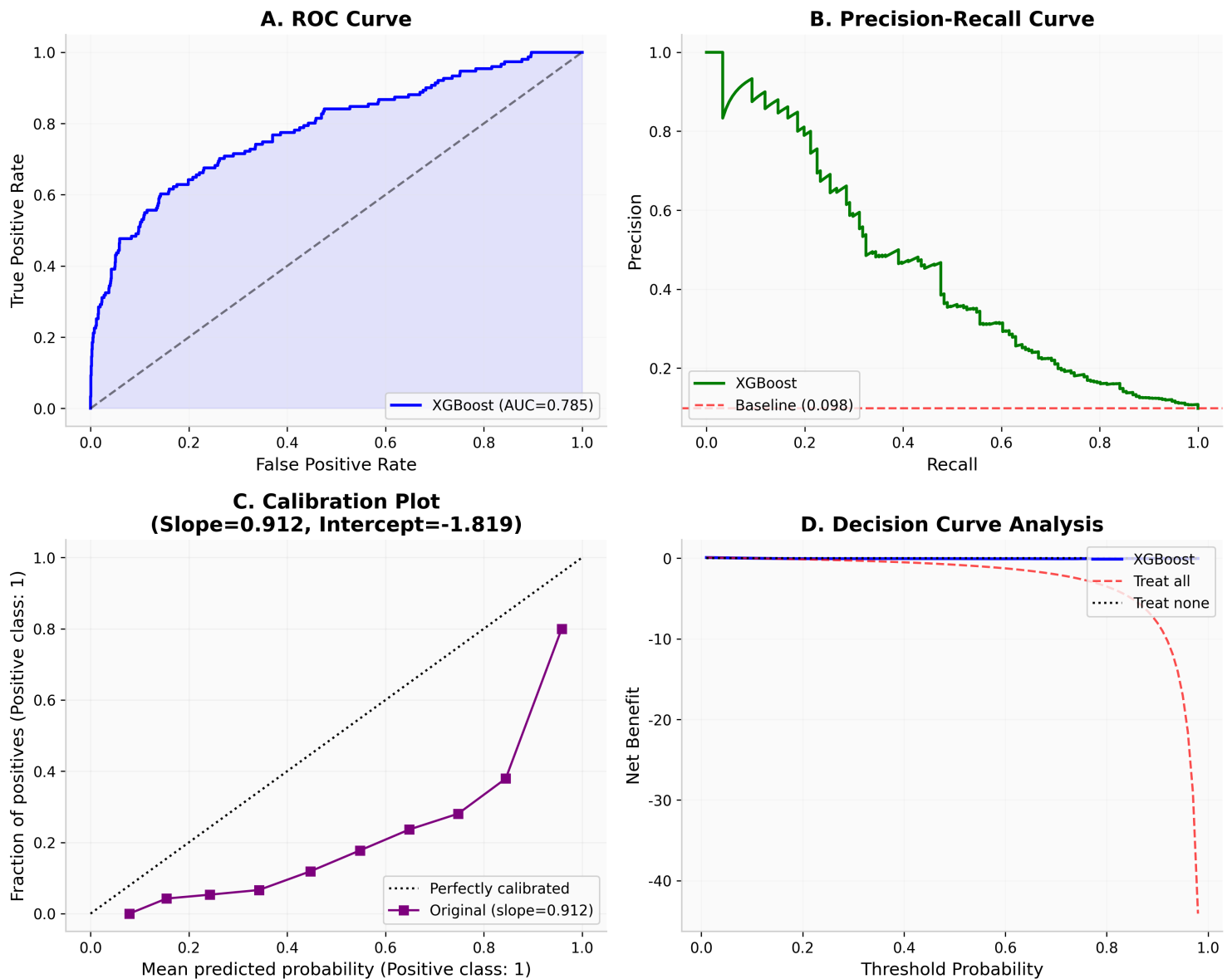


Figure 2

Model Performance in Temporal Validation (2023-2025)

Figure 3. SHAP Interpretation and Simplified Risk Score

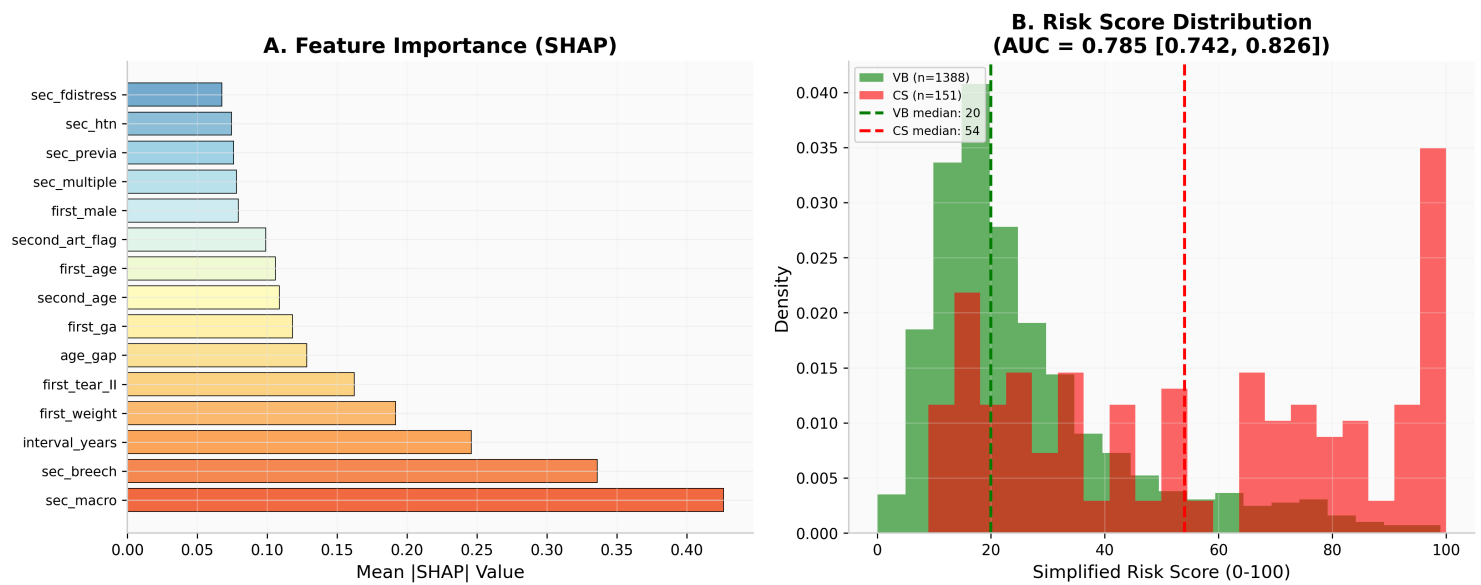


Figure 3

SHAP Interpretation and Simplified Risk Score

(A) Top 15 feature importance by mean |SHAP| value. (B) Distribution of simplified 0-100 risk score by delivery outcome (AUC = 0.785). VB = vaginal birth; CS = cesarean section.

Figure 4. Sensitivity Analyses and Subgroup Validation

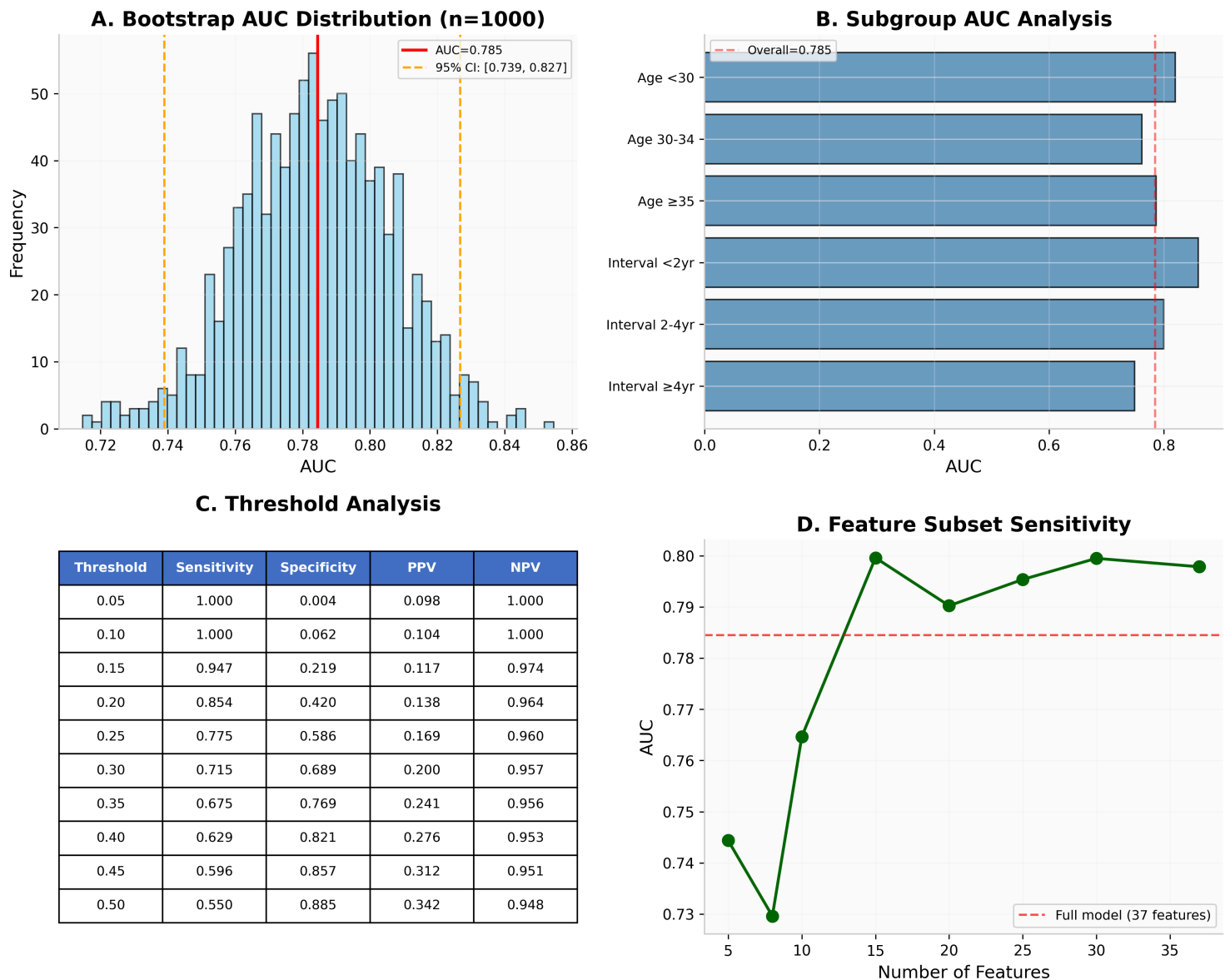


Figure 4

Sensitivity Analyses and Subgroup Validation

(A) Bootstrap distribution of AUC (1000 resamples). (B) Subgroup AUC by maternal age and inter-delivery interval. (C) Threshold analysis with sensitivity, specificity, PPV, and NPV. (D) AUC by number of included features.

Supplementary Files

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

- [SupplementaryS1FeatureList.csv](#)

- [SupplementaryS3SubgroupAnalysis.xlsx](#)
- [SupplementaryS2Hyperparameters.txt](#)
- [SupplementaryS4TRIPODAIChecklist.txt](#)
- [SupplementaryS5SHAPDependence.png](#)