Open Access

Predicting preterm birth using electronic medical records from multiple prenatal visits

Chenyan Huang[1,5], Xi Long[2,5*], Myrthe van der Ven[3,4,5], Maurits Kaptein[1], S. Guid Oei[2,4,5] and Edwin van den Heuvel[1,5]

Abstract

This study aimed to predict preterm birth in nulliparous women using machine learning and easily accessible variables from prenatal visits. Elastic net regularized logistic regression models were developed and evaluated using 5-fold cross-validation on data from 8,830 women in the Nulliparous Pregnancy Outcomes Study: New Mothers-to-Be (nuMoM2b) dataset at three prenatal visits: 6^0 - 13^6, 16^0 - 21^6, and 22^0 - 29^6 weeks of gestational age (GA). The models' performance, assessed using Area Under the Curve (AUC), sensitivity, specificity, and accuracy, consistently improved with the incorporation of data from later prenatal visits. AUC scores increased from 0.6161 in the first visit to 0.7087 in the third visit, while sensitivity and specificity also showed notable improvements. The addition of ultrasound measurements, such as cervical length and Pulsatility Index, substantially enhanced the models' predictive ability. Notably, the model achieved a sensitivity of 0.8254 and 0.9295 for predicting very preterm and extreme preterm births, respectively, at the third prenatal visit. These findings highlight the importance of ultrasound measurements and suggest that incorporating machine learning-based risk assessment and routine late-pregnancy ultrasounds into prenatal care could improve maternal and neonatal outcomes by enabling timely interventions for high-risk women.

Keywords Preterm birth, Prediction, Machine learning, Logistics regression

Background

Preterm birth, defined as the delivery at less than 37 completed weeks of pregnancy, is one of the most widely studied obstetrical complications, and remains a leading cause of prenatal morbidity and mortality [1–4]. Preterm birth is associated with profound short-term and long-term consequences in maternal health and neonatal health. A relatively recent study by Henderson et al. [5] has shown that women who had a preterm birth have an increased risk of ill-health, including significantly more anxiety, fatigue and flashbacks, and negative feelings about their baby in the early months. Babies who are born before they get physically ready for the world, often require special care [6] and they face greater risks of severe health problems, including palsy, intellectual impairment, chronic lung disease and vision and hearing loss [7].

The accurate and timely prediction of preterm birth is paramount, serving as the primary motivation for our study. Such predictions are crucial not only for initiating immediate interventions but also for managing the logistical aspects of neonatal care, especially in settings where advanced care facilities are lacking. The essential

*Correspondence:
Xi Long
x.long@tue.nl
[1] Department of Mathematics and Computer Science, Eindhoven University of Technology, Eindhoven 5612 AZ, North Brabant, The Netherlands
[2] Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven 5612 AZ, North Brabant, The Netherlands
[3] Department of Biomedical Engineering, Eindhoven University of Technology, Eindhoven 5612 AZ, North Brabant, The Netherlands
[4] Department of Obstetrics and Gynaecology, Máxima Medical Centre, Dominee Theodor Fliednerstraat 1, 5631 BM Eindhoven, North Brabant, The Netherlands
[5] Eindhoven MedTech Innovation Center, Eindhoven University of Technology, Eindhoven 5612 AZ, North Brabant, The Netherlands

Huang *et al. BMC Pregnancy and Childbirth*   (2024) 24:843

Page 2 of 25

interventions include the administration of steroids to promote lung maturity and magnesium sulfate to protect the developing brain, which are particularly crucial when access to specialized neonatal care is limited. In cases where the necessary care facilities are not immediately available, accurately predicting preterm birth also facilitates the timely transfer of patients either antenatally or postnatally to centers equipped to provide critical care, thus optimizing both maternal and neonatal outcomes [8]. Recently a review study has highlighted the effectiveness of intrapartum and postnatal interventions particularly in low-and-middle-income countries (LMIC) [9]. For instance, prophylactic steroids in preterm labor reduced cerebral hemorrhage by 70%, and neonatal mortality by 37% [10]. Delayed cord clamping has been linked to beneficial effects such as higher circulating blood volume during the first 24 hours of life, less need for blood transfusions, and lower incidence of intraventricular hemorrhage [11, 12]. Evidence supported the use of injectable vitamin K (phytomenadione) shortly after birth to prevent hemorrhage-related morbidity and mortality [13] and hospital-based kangroo mother care (KMC) has a reduction on neonatal mortality of low birth weight babies (normally born preterm) [14, 15]. These studies highlighted the importance of timely clinical actions following accurate preterm birth predictions.

The majority of current work on preterm birth prediction aims to identify risk factors of preterm birth, and many risk factors have been reported to increase the risk of preterm birth such as body mass index (BMI), maternal family history, cervical length, vaginal bleeding, and depression [16–25].

In recent years, several studies have explored the use of Electronic Medical Records (EMRs) for preterm birth prediction. For instance, Tran et al. [26] focused on proper data preparation, addressing issues such as data leakage and class imbalance, and achieved an AUC of 0.79 using a stabilized sparse logistic regression method. Weber et al. [27] used data from over 2 million patients to predict preterm birth in nulliparous women, achieving an AUC of 0.67 by combining racial-ethnic groups. Esty et al. [28] aimed to create a model that surpasses the prediction quality of expensive and invasive fibronectin marker screenings, achieving a sensitivity of 90.9% and specificity of 71.8% using a C5.0 Decision Tree. Gao et al. [29] focused on predicting extreme preterm birth (before 28 weeks) using various machine learning models and achieved an AUC of 0.827 with an ensemble of LSTM models using word2vec embeddings. These studies highlight the growing interest in using EMRs for preterm birth prediction. However, a significant gap in the literature remains: many of these models rely on variables that are difficult to obtain in routine clinical practice or fail to

evaluate the model's performance at multiple time points throughout pregnancy. Our study is to address this gap by focusing on easily accessible variables that can be collected during standard prenatal visits and assessing the model's predictive ability at different stages of pregnancy.

A notable exception is the study by AlSaad et al. which achieved predictive power as evidenced by ROC-AUC scores of 0.82, 0.79, and 0.78 at 1, 3, and 6 months prior to delivery, respectively [30]. Despite these results, their approach used a retrospective design where data collection points were predetermined at 1, 3, 6, and 9 months before a known delivery date. While statistically robust, this method poses significant practical challenges in prospective clinical settings: specifically, this retrospective timing assumes precise knowledge of the delivery date, which is not feasible in real-time clinical practice. Moreover, their model relied on a wide array of complex variables, including diagnoses, medications, procedures, and lab orders, many of which are difficult to collect routinely, especially in resource-limited settings. Abraham et al.'s study also brought up the idea of predicting preterm birth at different time points: 0, 13, and 28 weeks of gestation, and they concluded that the model performance increased from conception (0 weeks) to the highest at 28 weeks with a result of ROC-AUC = 0.75, PR-AUC = 0.40 at 28 weeks of gestation [31]. However, this approach may face practical limitations in real-world applications. Billing codes can vary significantly across different medical practices and healthcare systems, potentially limiting the model's generalization and standardization.

In this study, we aimed to analyze the predictive abilities for preterm birth using "easy-to-acquire" EMR data collected from various prenatal visits in the nuMoM2B database [32], to demonstrate how such predictive models can be effectively integrated into clinical practice. Easy-to-acquire data, which was collected at naturally occurring gestational weeks, including individual and familial medical backgrounds, demographic details, non-intrusive evaluations, and standard diagnostic procedures, from different prenatal visits, is important to help maintain a reasonable cost in predictions and easy-to-handle evaluations. We evaluated models that combine medical history, family medical history, demographic information, non-invasive assessments, and common diagnostic measures to predict preterm birth. We assessed the efficacy of an elastic net regularized logistic regression model for predicting preterm birth, focusing on datasets enriched with data from multiple prenatal visits.

The paper is organized to first introduce the methodology, highlighting our use of the elastic net approach. We then explored the contributions of different variables to model predictions, performed a sensitivity analysis on

Huang *et al. BMC Pregnancy and Childbirth*   (2024) 24:843

Page 3 of 25

subgroups of preterm birth categorized by gestational age, and investigated the impact of updated measurements from later prenatal visits, with a particular focus on ultrasound measurements such as cervical length and Pulsatility Index.

Methodology

Data collection

The Nulliparous Pregnancy Outcomes Study: New Mothers-to-Be (nuMoM2b) dataset, a valuable resource in our study, comprises a comprehensive collection of data. The dataset includes information from 10,038 nulliparous women with singleton pregnancies recruited from 8 clinical centers affiliated with research institutions, detailed in the Appendix A in Fig. 8 and in Table 9. The primary objectives of the nuMoM2b study were to (1) determine maternal characteristics, including genetics, epigenetics, and physiological response to pregnancy and environmental factors that influence or predict adverse pregnancy outcome; (2) identify specific aspects of placental development and function that lead to adverse pregnancy outcome; and (3) characterize genetic, growth, and developmental parameters of the fetus that are associated with adverse pregnancy outcome.

Inclusion criteria for the study were: (1) nulliparous women (defined as a pregnant woman with no prior pregnancy lasting 20 weeks or more), (2) viable singleton gestation (defined as the presence of fetal cardiac activity at the most recent ultrasound before enrollment), (3) gestational age between 6 weeks 0 days and 13 weeks 6 days at recruitment (based on a "project ultrasound" performed by a certified nuMoM2b sonographer), and (4) intention to deliver in a participating hospital of the nuMoM2b Network.

Exclusion criteria included: (1) maternal age less than 13 years, (2) history of three or more spontaneous abortions, (3) fetal malformation evident at or before enrollment that is likely lethal, (4) known fetal aneuploidy, (5) surrogate pregnancy, (6) multifetal pregnancy reduction, (7) participation in an intervention study that is anticipated to influence maternal or fetal morbidities/mortality (unless the study code is made available before enrollment), (8) previous enrollment in the nuMoM2b study, (9) planned pregnancy termination, and (10) inability to provide informed consent.

Data were collected through interviews, self-administered questionnaires, clinical measurements, ultrasounds, and a review of medical records at four scheduled study visits (Table 1). At each visit, specific data collection procedures were followed to ensure data integrity and consistency across clinical sites. All study personnel underwent standardized training on data collection protocols, and regular monitoring visits were conducted to ensure adherence to study procedures.

Table 1 Gestational time points of prenatal visits in the nuMom2B study

Visit	Weeks of Gestation
Visit 1	6^0 - 13^6 (6 weeks and 0 days to 13 weeks and 6 days)
Visit 2	16^0 - 21^6 (16 weeks and 0 days to 21 weeks and 6 days)
Visit 3	22^0 - 29^6 (22 weeks and 0 days to 29 weeks and 6 days)
Visit 4	The time of delivery

After data were received, additional data edits were programmed and run to check for range violations and consistency with skip logic. Site coordinators reviewed items failing these edit checks and made corrections to the database or confirmed that the data were correct as keyed. As data were used for analyses, efforts were made to correct important fields, such as those used to date the pregnancy and calculate the estimated gestational age. The site staff reviewed the edit requests and entered corrections to the data in the database.

Despite these data quality control measures, some errors may have persisted, and some data may remain out of expected ranges or otherwise not make sense. Investigators using the data are responsible for reviewing fields of interest prior to conducting analyses and making decisions on how to handle erroneous data.

Of the 10,038 participants, 9,289 consented to release their anonymized data for research, and 9,127 agreed to release data related to their babies. The information collected encompassed a wide range of variables, including demographics, psychosocial factors, dietary habits, physiological measurements, and pregnancy outcomes. Study visits were scheduled at the following four times during the study as displayed in Table 1:

A full description of the methods employed in the Nulliparous Pregnancy Outcomes Study can be found in the publication by Hass et al. [32].

Sample size considerations

This study utilized the nuMoM2b dataset, which initially included 10,038 nulliparous women. While traditional a priori sample size calculations were not applicable for this secondary analysis, we conducted post-hoc power analyses to ensure adequate statistical power for our machine learning approach.

For logistic regression-based prediction models, a minimum of 10 events per predictor variable (EPV) is generally recommended to ensure model stability and reliability [33]. Our final analytical sample included 8,830 participants with 759 preterm births. The number of predictor variables increased across visits:

Huang *et al. BMC Pregnancy and Childbirth*   (2024) 24:843

Page 4 of 25

- Visit 1: 35 predictors (21.7 events per predictor)
- Visit 2 (Cumulative): 51 predictors (14.9 events per predictor)
- Visit 3 (Cumulative): 73 predictors (10.4 events per predictor)

All models exceeded or met the minimum recommended threshold of 10 EPV, indicating adequate statistical power. Additionally, our sample size of 8,830 participants substantially exceeds the recommended minimum of 2,000 participants suggested by systematic reviews for stable and generalizable machine learning applications in healthcare [34].

The detailed participant selection and variable selection process and reasons for exclusion are presented in the Results section.

Variables

The selection of candidate predictor variables followed a systematic multi-step process. Through comprehensive literature review [8, 26–29, 35–48], we identified established predictors of preterm birth, including biological markers, specialized imaging measurements, and clinical factors. While some recognized predictors (such as inflammatory markers Interleukin-6 in amniotic fluid, interleukin-1 β, and matrix metalloproteinase-8) show potential of strong predictive value, they were excluded due to their absence in the nuMoM2b database and their resource-intensive nature. Our final variable selection focused on clinically accessible predictors available in the database that could be readily obtained during routine prenatal care.

The final variable selection prioritized clinically accessible predictors available within the database that could be easily obtained during routine prenatal care. Variables were evaluated based on multiple criteria: availability during standard prenatal visits, cost-effectiveness of collection, non-invasive nature of measurement, and demonstrated clinical relevance from literature. Selected variables encompassed medical history, family medical history, demographic information, non-invasive assessments, and common diagnostic measures, as detailed in Table 2. The choice of these variables was guided by considerations of feasibility and cost-effectiveness, which can be collected rapidly, affordably, and without the need for invasive or resource-intensive procedures. Consequently, the predictive models can be readily applied in routine clinical practice [49].

Data preprocessing and missing data analysis

Following feature selection, comprehensive data preprocessing was undertaken. A thorough investigation of missing data patterns was conducted, including a detailed examination of data collection procedures and a review of dataset documentation. While many variables exhibited missing data patterns without clear documentation explanation (categorized as missing at random, MAR), certain variables demonstrated systematic missingness attributable to specific clinical protocols.

However, the missingness of certain variables was systematic and could be attributed to specific reasons. For variables such as 'Tobacco Use at Visit 1' and 'Vaginal Bleeding' at all three visits, missing entries were assumed to indicate no tobacco use or no bleeding beyond spotting, respectively, and were imputed accordingly. During data collection, a uterine artery Doppler ultrasound at Visit 3 was mandated only if the findings from the uterine artery Doppler at Visit 2 were abnormal. This protocol is consistent with Hofstaetter et al's study indicating that, in uncomplicated pregnancies, the uterine artery Pulsatility Index (PI) does not significantly vary with gestational age [50]. Hence, for cases with missing data of 'Left uterine artery - Pulsatility Index at Visit 3' and 'Right uterine artery - Pulsatility Index at Visit 3' where Visit 2 measurements were normal, we imputed these missing values by randomly selecting from the normal PI (Pulsatility Index) range appropriate between 22 and 29 weeks, as reported in the study by Cavoretto et al. [51]. This approach was necessary to maintain the integrity of the dataset while accommodating the clinical decision-making process that dictates these measurements.

Table 2 Identified characteristics and variables from NuMom2B dataset

Characteristics	Variables
Basic Preconception Characteristics	maternal age, history of cervical excisional procedures/surgery(Loop Electrosurgical Excision Procedure (LEEP)/conization), height, Body Mass Index(BMI), use of assisted reproductive technologies, family disease history(pregnancy complications and hypertension), and family history of preterm birth
Socio-Demographic Characteristics	socio-economic status, educational attainment, marital status, and ethnicity
Lifestyle Characteristics	tobacco use, alcohol use, mental stress, and maternal Vitamin-D deficiency
Obstetric/Pregnancy Characteristics	gestational diabetes, gestational hypertension, vaginal bleeding, serial transvaginal cervical length, and cervical insufficiency
Medical History Characteristics	pre-pregnancy diabetes history, pre-pregnancy hypertension history and obstetric history

Additionally, we identified that the 'Left uterine artery Pulsatility Index at Visit 1' and 'Right uterine artery - Pulsatility Index at Visit 1' were missing in over 85% of the cases because the uterine artery Doppler assessment at Visit 1 was performed optionally. Given the high degree of missingness and the optional nature of this assessment, these variables were removed from the analysis.

For other variables considered to be MAR and with less than 80% missing data, we employed the K-Nearest Neighbor (KNN) imputation method, a commonly used approach in clinical datasets for handling missing data [52]. In this method, for a given patient with missing records, we applied a k-value of 5, and the KNN methods identified the 5-nearest patients based on Euclidean distance, using the variables that were also used for the prediction. The missing values were replaced using a majority vote for discrete variables and weighted means for continuous features, drawing from the records of these nearest patients [53, 54]. Imputation was conducted once across each training dataset prior to variable extraction for each visit. To validate the integrity of our imputation, we compared the performance of tree-based models on both the original and imputed datasets, observing no significant difference in model performance (AUC score) with a P-value of 0.604. This result confirms the efficacy of our imputation approach, as it maintained consistent predictive performance across datasets.

Model development and reporting

This study follows the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) guidelines [55]. We developed and validated prediction models for preterm birth using three datasets corresponding to different prenatal visits. The model development process included predictor selection, handling of missing data, and assessment of model performance through both discrimination and calibration metrics.

Model training and testing

To establish predictive models, five-fold cross-validation was applied to avoid overfitting. For the five-fold cross-validation, samples (corresponding to pregnant women) were randomly split into five groups of similar size. The model was trained and validated on four parts and tested on the remaining part. This procedure was repeated five times on each set of four groups [27]. The size of the five groups were 1797, 1814, 1726, 1787, and 1707, respectively.

Elastic net regularized logistic regression models were used to predict preterm birth after the standardization of the numerical variables. The elastic net regularized logistic regression model is a practical supervised machine learning algorithm combining least absolute shrinkage and selection operator (LASSO) regression and Ridge regression with a good probabilistic interpretation of variables suitable for disease prediction [56]. The LASSO penalty selected variables by reducing the absolute value of the weight, while the Ridge penalty by further reducing the extremities of weights. Details of the method can be found in the publication by Zou et al. [57].

Two important parameters for the elastic net regression model were used, including α (representing the weight of the penalty) and λ (representing the complexity of the penalty). In this model, α controls the balance between LASSO and Ridge, with $\alpha=1$ corresponding to the LASSO (the default estimator) and $\alpha=0$ corresponding to Ridge regression [58]. To select the proper parameters for the model, we tested α from 0 to 1 with a step size of 0.1, as well as the inverse of the regularization strength from 0.001 to 100. We then used 70%−30% training-validating datasets to select optimal α and λ coefficients for elastic net regression models in seek of maximum of the receiver operating characteristics curve (AUC) scores for validation datasets, and in order to achieve optimal efficiency of the model.

Model performance analysis
Discrimination metrics

In this study, the average AUC scores and the pooled performance metrics over 5 test datasets, including the AUC scores, accuracy, sensitivity, and specificity with the test datasets were used to evaluate the model performance [28, 38, 59]. In pooling, the predictions made in each cross-validation round were pooled into one set and one ROC curve was drawn and one AUC score was calculated from it [60]. This procedure was applied to the datasets at the three visits by training using only features available at each time point, and then the performances of models were compared to assess the predictive capability of models at different visits.

To assess the statistical significance of performance differences between models at different visits, we employed DeLong's test for paired ROC curves [61]. This test was chosen as it accounts for the correlated nature of the ROC curves, given that they were derived from the same population [62]. To control for multiple comparisons across the three visit-specific models, we applied Bonferroni correction (adjusted significance level $\alpha = 0.017$) [63].

All statistical analyses were performed using R version 4.2.3, with the pROC package for ROC analysis and DeLong's test [64], and the caret package for cross-validation procedures [65].

Huang *et al. BMC Pregnancy and Childbirth*   (2024) 24:843

Page 6 of 25

Calibration assessment

To assess the calibration of our prediction models, we examined the agreement between predicted probabilities and observed outcomes. We evaluated calibration through multiple metrics: calibration-in-the-large, calibration slope, and calibration plots. Calibration-in-the-large measures the agreement between the mean predicted probability and the observed event rate. The calibration slope, obtained by regressing the observed outcomes on the predicted probabilities, indicates whether predictions are systematically too extreme (slope < 1) or too modest (slope > 1). We created calibration plots by grouping predictions into deciles and comparing the mean predicted probability with the observed proportion of events in each group, with 95% confidence intervals calculated using the standard error of the observed proportions.

Results

Of the 10,038 women initially enrolled in the nuMoM2b study, 749 were excluded due to missing risk factor data. Among the remaining 9,289 women with available data, 438 were excluded (432 with no record of birth type, 4 with live births but incomplete records, and 2 with stillbirths but incomplete records). An additional 21 participants were excluded due to having more than 80% missing data, which primarily included cases of early pregnancy loss or very early preterm births where subsequent visit data could not be collected. This systematic filtering process resulted in a final analytical sample of 8,830 women (8,071 with full-term births and 759 with preterm births) who had sufficient data through delivery. The flow of data inclusion/exclusion is displayed in Fig. 1.

Among all variables, 35 variables were available at the first visit, 51 variables were available at the second visit, and 73 variables were available at the third visit.

At the first visit, the basic information (dataset B) was collected, including poverty level, ethnicity, pre-gestational diabetes, obesity level, marital status, and educational level. Additionally, we have basic measurement variables (including questionnaires and surveys) such as BMI, vaginal bleeding, smoking, stress level, gestational diabetes diagnosis and hypertension diagnosis measured at all three visits (dataset T1, T2, and T3). Subsequently, at the second and third visits the data was enriched with ultrasound measurements (dataset U2 and U3).

An overview of the datasets is presented in Table 3.

In our study, to answer our research questions, we used several comprehensive datasets composed of
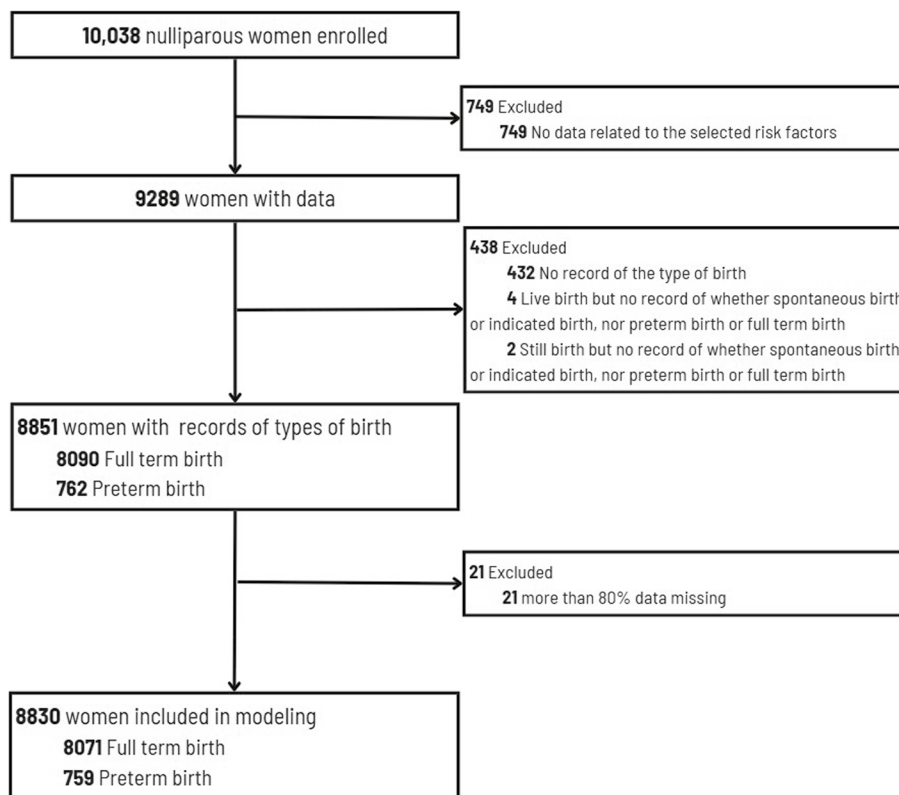

Fig. 1 A flow of data inclusion/exclusion

Huang *et al. BMC Pregnancy and Childbirth*   (2024) 24:843

Page 7 of 25

Table 3 Data collection strategy across three prenatal visits

Dataset	Variables included
B	poverty, race, pre-gestational diabetes diagnosis, marital status, education, obesity level
T1	basic measurements at Visit 1: BMI, vaginal bleeding, smoking habits, stress level, gestational diabetes, hypertension
T2	basic measurements Visit 2: weight, BMI, smoking status, vaginal bleeding, gestational diabetes, hypertension
T3	basic measurements at Visit 3: weight, BMI, smoking status, vaginal bleeding, gestational diabetes, hypertension
U2	Ultrasound measurements at Visit 2: cervical length and Pulsatility Index
U3	Ultrasound measurements at Visit 3: cervical length and Pulsatility Index

Table 4 Dataset composition across prenatal visits

Dataset	Data composition	Dataset composition
Dataset 1	all data available at Visit 1	B + T1
Dataset 2	all data available at Visit 2	B + T1 + T2 + U2
Dataset 3	all data available at Visit 3	B + T1 - 3 + U2 + U3
Dataset V1	basic information + basic measurements at Visit 1	B + T1
Dataset V2	basic information + basic measurements at Visit 2	B + T2
Dataset V3	basic information + basic measurements at Visit 3	B + T3
Dataset U1	basic information + basic measurements at Visit 1	B + T1
Dataset U2	basic information + basic measurements at Visit 2 + ultrasound measurements at Visit 2	B + T2 + U 2
Dataset U3	basic information + basic measurements at Visit 3 + ultrasound measurements at Visit 3	B + T3 + U 3

Dataset V1 and Dataset U1 are actually identical, containing the same set of basic information variables and basic measurement variables from Visit 1

Table 5 AUC scores at varied gestational time points

	Training dataset	Validation dataset	Test dataset
Dataset 1	0.6512 ± 0.0122	0.6142 ± 0.0215	0.6157 ± 0.0217
Dataset 2	0.6731 ± 0.0137	0.6444 ± 0.0202	0.6423 ± 0.0137
Dataset 3	0.7393 ± 0.0094	0.7156 ± 0.0124	0.7084 ± 0.0219

dataset B, T1, T2, T3, U2, and U3, as showcased in Table 4. We first looked into prediction on the three datasets 1, 2, and 3 in Model performance is improved when data from later prenatal visits were added section, studied the most influential variables in the prediction in Analysis of key variable contributions across three prenatal visit datasets section, and investigated the prediction performance for the severity of preterm birth in Sensitivity analysis on subgroups of preterm birth section. In Model prediction on test datasets with only updated information section we investigated the need of updating information during visits where we focused on the other data sets mentioned in Table 4.

We evaluated the predictive performance of models using AUC scores derived from 5-fold cross-validation on datasets that contained data from different prenatal visits. This assessment was followed by a comprehensive analysis of pooled performance metrics, including AUC scores, accuracy, sensitivity, and specificity.

Model performance is improved when data from later prenatal visits were added
Discrimination performance
The summary of model performance across the three gestational periods for the three datasets 1, 2, and 3 (listed in Table 4) is detailed in Table 5, and the pooled ROC curves for these datasets are presented in Fig. 2.

The training dataset performance showed progressive enhancement, with AUC scores increasing from 0.6512 (SD: 0.0122) at Visit 1 to 0.6731 (SD: 0.0137) at Visit 2, and reaching 0.7393 (SD: 0.0094) at Visit 3. This pattern was similarly reflected in the validation datasets, where AUC scores improved from 0.6142 (SD: 0.0215) at Visit 1 to 0.6444 (SD: 0.0202) at Visit 2, and 0.7156 (SD: 0.0124) at Visit 3. Most importantly, the test dataset results confirmed this trend, with AUC scores of 0.6157 (SD: 0.0217) at Visit 1, improving to 0.6423 (SD: 0.0137) at Visit 2, and achieving the highest performance of 0.7084 (SD: 0.0219) at Visit 3. This progression suggests that predictive ability substantially increases as more comprehensive gestational data becomes available, particularly by the information collected in the third visit.
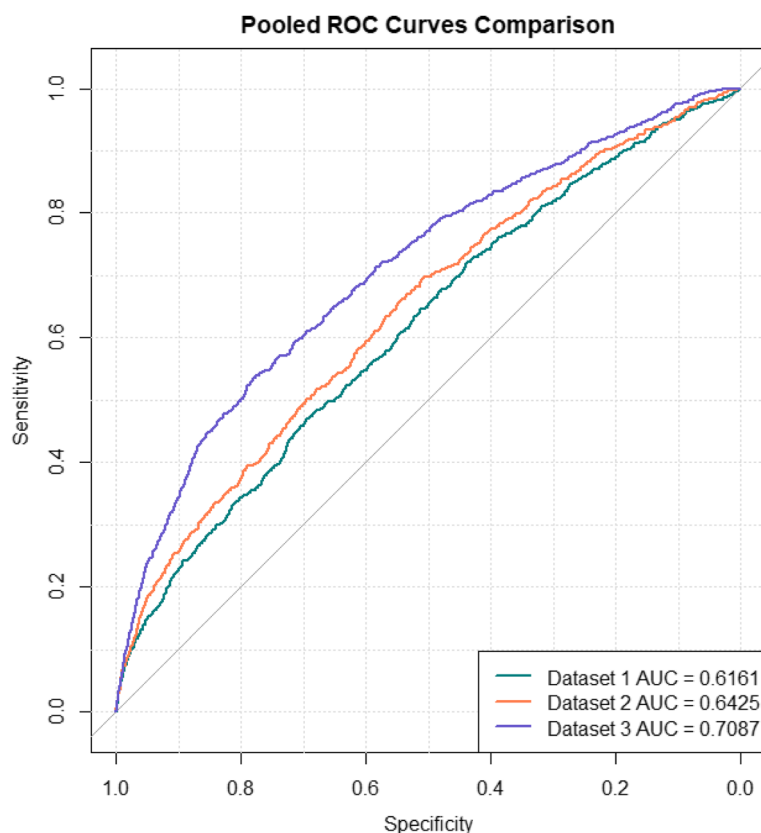
Huang *et al. BMC Pregnancy and Childbirth*   (2024) 24:843

Page 8 of 25

Fig. 2 Pooled ROC curves formed by merging the 5 test folds on 3 datasets

In Fig. 2, we present the pooled ROC curve formed by merging the five test sets from the cross-validation and computing their combined ROC curve for each dataset. It provides a visual representation of these improvements, showing significant increases in AUC from the dataset with information from the first prenatal visit to the dataset with information from the first and the second prenatal visits and subsequently to the dataset with information from all three prenatal visits. The result showed that when data from Visit 2 and Visit 3 were added, the predictive ability substantially improved according to a notable increase in AUC scores. The progression in model performance with the inclusion of later prenatal visit data suggested a direct correlation between the quantity of gestational data and the predictive accuracy of prenatal assessments.

To formally assess the statistical significance of these performance differences between visits, we conducted DeLong's test for paired ROC curves. The analysis confirmed significant improvements in discriminative ability across visits. The improvement from Visit 1 (AUC = 0.6161) to Visit 2 (AUC = 0.6425) was significant (difference = 0.026, 95% CI: 0.009–0.040, Z = −3.059, p = 0.002). The Visit 3 model (AUC = 0.7087) demonstrated

the strongest performance, significantly outperforming both the Visit 1 model (difference = 0.093, 95% CI: 0.071–0.112, Z = −8.624, p < 0.001) and Visit 2 model (difference = 0.066, 95% CI: 0.048–0.086, Z = −6.900, p < 0.001). These progressive improvements in AUC scores were statistically significant even after applying Bonferroni correction for multiple comparisons (adjusted significance level $\alpha = 0.017$), indicating robust enhancement in model performance with the addition of data from later prenatal visits.

A more comprehensive model performance analysis was conducted based on various performance metrics. We optimized the predictive threshold using the Youden Index [66], a widely recognized method for establishing the optimal cutoff for labeling predictions by maximizing the difference between the true positive rate (sensitivity) and the false positive rate (1 - specificity), followed by an assessment of accuracy, sensitivity, and specificity, with 95% confidence interval. Table 6 shows in more detail what change in predictive power occurs as more prenatal data becomes available across later visits.

Further examination of multiple performance metrics revealed important insights into the models' predictive capabilities (Table 6). The overall AUC scores showed

Huang *et al. BMC Pregnancy and Childbirth*   (2024) 24:843

Page 9 of 25

Table 6 Pooled performance metrics on datasets 1, 2, and 3 with 95% confidence intervals

	Dataset 1	Dataset 2	Dataset 3
AUC Scores	0.6161	0.6425	0.7087
Accuracy	0.6314	0.5353	0.6652
	[0.6213, 0.6414]	[0.5249, 0.5457]	[0.6554, 0.6751]
Sensitivity	0.5731	0.6930	0.6324
	[0.5379, 0.6368]	[0.6602, 0.7258]	[0.5981, 0.6667]
Specificity	0.6368	0.5205	0.6683
	[0.6017, 0.6437]	[0.4877, 0.5314]	[0.6340, 0.6786]

steady improvement from 0.6161 at Visit 1 to 0.6425 at Visit 2, and ultimately reaching 0.7087 at Visit 3. The accuracy metrics showed an interesting pattern: starting at 63.14% (95% CI: 62.13–64.14%) in Visit 1, decreasing to 53.53% (95% CI: 52.49–54.57%) in Visit 2, before improving to 66.52% (95% CI: 65.54–67.51%) in Visit 3. This temporary decrease in accuracy at Visit 2 can be explained by examining the sensitivity and specificity trade-offs. The sensitivity improved markedly from Visit 1 (57.31%, 95% CI: 53.79–63.68%) to Visit 2 (69.30%, 95% CI: 66.02–72.58%), though this came at the cost of reduced specificity (from 63.68%, 95% CI: 60.17–64.37% to 52.05%, 95% CI: 48.77–53.14%). By Visit 3, the model achieved a better balance, with sensitivity of 63.24% (95% CI: 59.81–66.67%) and specificity of 66.83% (95% CI: 63.40–67.86%), resulting in the highest overall accuracy.

Model calibration analysis
We assessed the calibration of our prediction models through several metrics. All three models accurately predicted the overall preterm birth rate of 8.6%, with zero calibration-in-the-large, indicating excellent population-level calibration. However, the calibration slopes (Dataset 1: 8.69, Dataset 2: 9.27, Dataset 3: 8.31) suggest that the models are under-confident in their predictions, with true probabilities changing more rapidly than predicted probabilities.

The range of predicted probabilities broadened across visits, with Dataset 3 showing the widest range (1.1% to 85.1%) compared to Dataset 1 (4.1% to 49.4%) and Dataset 2 (1.7% to 64.4%). This pattern suggests that later visits enable more confident risk stratification, consistent with our discrimination analysis. The calibration plots (Fig. 3) demonstrate that while the models effectively rank patients by risk (as evidenced by the positive slopes), they tend to compress predictions toward the population mean, particularly for high-risk cases.

These findings indicate that while our models excel at risk stratification and maintain accurate population-level

predictions, the individual predicted probabilities should be interpreted as relative rather than absolute risk indicators.

Analysis of key variable contributions across three prenatal visit datasets
In Fig. 4, 10 variables with the highest absolute average coefficient values, that contributed the most to the prediction model, are visualized for the datasets 1, 2, and 3. For the categorical variables named 'Assisted reproduction', 'Vaginal bleeding', 'Cervical surgery - LEEP(loop electrosurgical excision procedure)', 'Hypertension condition', 'Diabetes condition'(excluding gestational diabetes), 'Tobacco use', and 'PreGestDM'(Pre-gestational diabetes). The value '1' indicates a 'Yes' response, affirming the presence of the condition or utilization of a service, whereas '2' indicates a 'No' response, signifying the absence of the condition or non-utilization of the service.

Among all the information from the first prenatal visit, diagnosed diabetes at Visit 1(excluding previous gestational diabetes), Pre-gestational diabetes and hypertension at Visit 1 contributed most to the prediction of preterm birth. The history of previous cervical surgery as LEEP and Cone both showed positive associations, as well as the use of assisted reproduction and the use of tobacco in the last month before Visit 1. Also 'Non-Hispanic Black' as a demographic variable with a positive coefficient, suggests a higher risk of preterm birth within this subgroup.

Continuing through dataset 2, consistency showed in pre-gestational diabetes, diagnosed diabetes (excluding previous gestational diabetes), hypertension, LEEP surgery, the use of assisted reproduction, vaginal bleeding at Visit 1, and the use of tobacco in the last month before Visit 1 suggested a continued influence of these conditions throughout the pregnancy. The 'Cervical length at Visit 2', with a negative coefficient, indicated a shorter cervical length related to preterm birth risk. The conditions of diabetes and hypertension at Visit 1 persisted as risk factors.

Upon including the third visit data, the 'Cervical length at Visit 3' presented with a large negative coefficient, and 'Vaginal bleeding at Visit 3' was positively correlated with preterm birth, increasing the predicted risk during the later stages of pregnancy. Pre-gestational diabetes, diagnosed diabetes (excluding previous gestational diabetes), hypertension, LEEP surgery, and the use of assisted reproduction continued to be significant contributors to preterm birth risk.

Sensitivity analysis on subgroups of preterm birth
The World Health Organization defines preterm birth as delivery before 37 completed weeks of gestation and
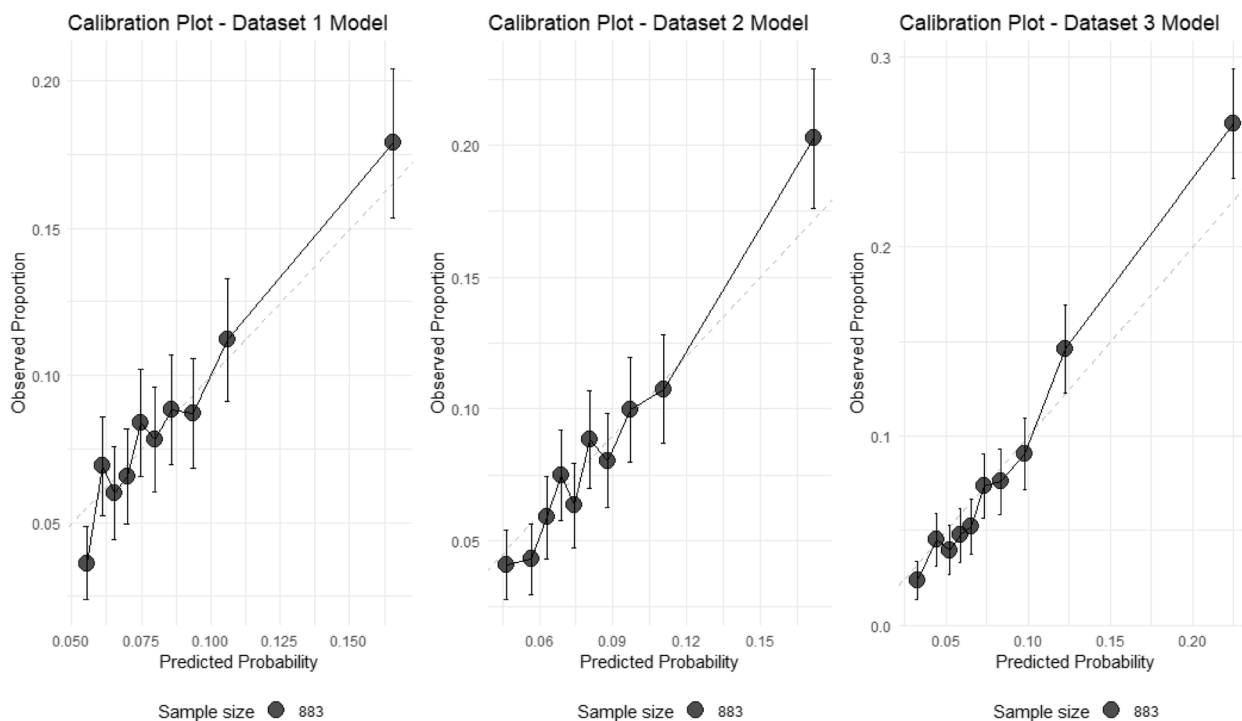
Fig. 3 Calibration plots for preterm birth prediction models across three datasets. The plots show the relationship between predicted probabilities and observed proportions of preterm birth for models based on (**A**) Dataset 1, (**B**) Dataset 2, and (**C**) Dataset 3. The dashed diagonal line represents perfect calibration, a hypothetical scenario where the predicted probabilities of preterm birth exactly match the observed proportions. Points represent deciles of predicted risk, with vertical bars indicating 95% confidence intervals. Point size is proportional to the number of patients in each group. While all models maintain accurate overall prediction of preterm birth rate (8.6%), they tend to compress predictions toward the population mean, particularly for high-risk cases. Dataset 3 model shows the widest range of predictions (1.1% to 85.1%), suggesting improved risk stratification at later visits

further classifies preterm birth into subcategories based on gestational age [67]. These classifications are crucial for understanding the variability in outcomes and tailoring interventions accordingly. Table 7 outlines these subcategories, and Fig. 5 displays the distribution of patients across these groups.

Table 8 provides a detailed examination of the pooled sensitivity with a 95% confidence interval of our predictive models for each preterm birth subgroup across three datasets.

The most striking performance was observed for extreme preterm birth (delivery before 28 weeks), where sensitivity improved progressively across datasets: from 69.57% (95% CI: 64.03–75.11%) in Dataset 1 to 76.81% (95% CI: 71.73–81.89%) in Dataset 2, and reaching its highest at 92.95% (95% CI: 89.83–96.07%) in Dataset 3. This exceptionally high sensitivity in Dataset 3 is particularly meaningful given the critical nature of extreme preterm births. This exceptionally high sensitivity in Dataset 3 can be explained by the proximity of Visit 3 to the delivery date for extreme preterm births; consequently, the data collected during this visit are more indicative of the imminent occurrence of

preterm birth. This high sensitivity is particularly crucial as extreme preterm births require the most urgent and intensive care.

Very preterm births (28–32 weeks) showed similarly robust detection rates, with sensitivity starting at 71.43% (95% CI: 65.74–77.12%) in Dataset 1, reaching its peak of 84.13% (95% CI: 79.53–88.73%) in Dataset 2, and maintaining strong performance at 82.54% (95% CI: 77.76–87.32%) in Dataset 3.

For moderate preterm births (32–34 weeks), the model showed more modest but still meaningful sensitivity. Starting at 53.51% (95% CI: 48.84–58.18%) in Dataset 1, it improved substantially to 69.30% (95% CI: 64.98–73.62%) in Dataset 2, before slightly decreasing to 64.91% (95% CI: 60.44–69.38%) in Dataset 3.

Late preterm births (34–37 weeks) proved the most challenging to predict, with sensitivity ranging from 54.78% (95% CI: 52.58–56.98%) in Dataset 1 to a peak of 66.47% (95% CI: 64.39–68.55%) in Dataset 2, before declining to 56.53% (95% CI: 54.34–58.72%) in Dataset 3.

When considering general preterm birth prediction (all categories combined), the model achieved moderate sensitivity levels: 57.31% (95% CI: 53.79–63.68%) in Dataset
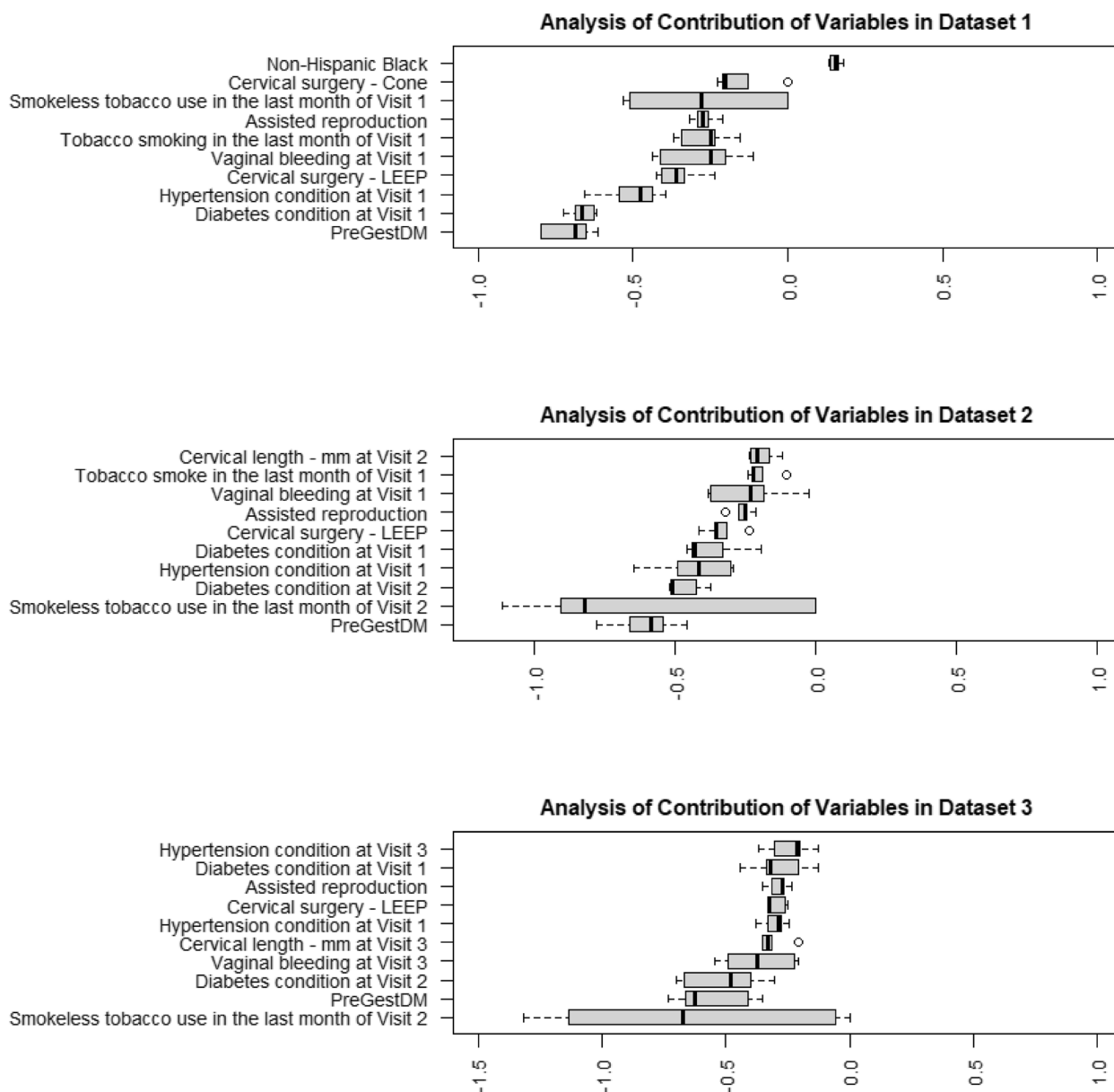
Huang *et al. BMC Pregnancy and Childbirth* (2024) 24:843

Page 11 of 25

Analysis of Contribution of Variables in Dataset 1

Analysis of Contribution of Variables in Dataset 2

Analysis of Contribution of Variables in Dataset 3

Fig. 4 Contribution of the top ten most contributing variables in the prediction models for datasets 1, 2 and 3

Table 7 Sub-categories of preterm birth based on gestational age

Sub-category	Gestational age at delivery
Extreme Preterm Birth	Less than 28 weeks
Very Preterm Birth	28 to 32 weeks
Moderate Preterm Birth	32 to 34 weeks
Late Preterm Birth	34 to 37 weeks

1, improving to 69.30% (95% CI: 66.02–72.58%) in Dataset 2, and maintaining 63.24% (95% CI: 59.81–66.67%) in Dataset 3.

Model prediction on test datasets with only updated information

To determine whether updates to Electronic Health Records (EHR) or the addition of new measurements, such as Pulsatility Index and cervical length, enhance model performance, we analyzed datasets with these updates. Basic information variables including poverty, race, pre-gestational diabetes diagnosis, marital status, education, and obesity level were utilized alongside updated variables from three prenatal visits-BMI, diabetes and hypertension diagnoses, smoking habits, and vaginal bleeding. These analyses were structured

Huang *et al. BMC Pregnancy and Childbirth*   (2024) 24:843

Page 12 of 25

Number of Patients on Subgroups of Preterm Birth
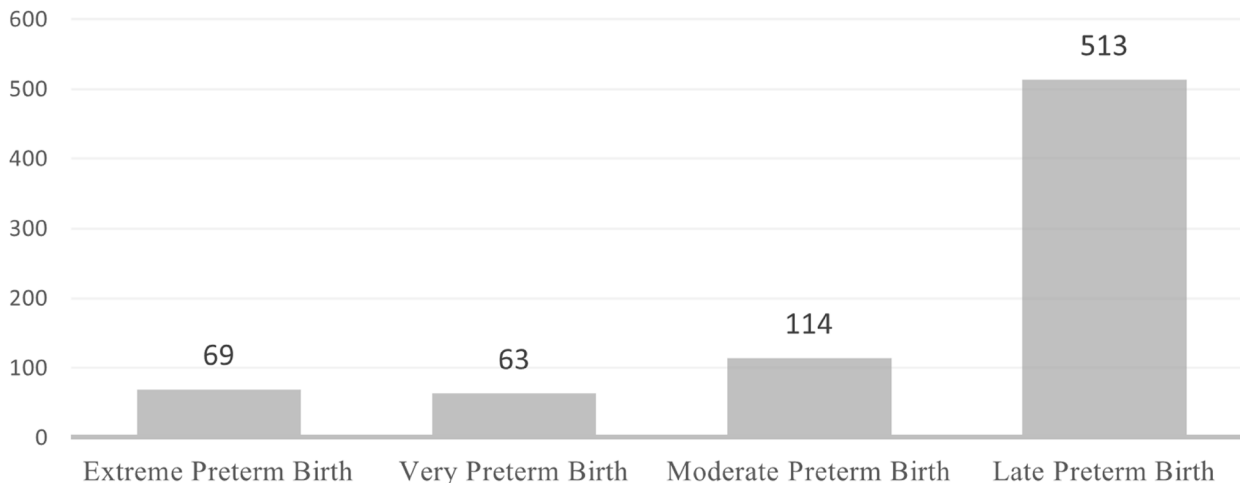

Fig. 5 Number of patients on subgroups of preterm birth

Table 8 Pooled sensitivity of model prediction on subgroups of preterm birth

Subgroup/ Sensitivity	Dataset 1	Dataset 2	Dataset 3
Extreme Preterm Birth	0.6957	0.7681	0.9295
	[0.6403, 0.7511]	[0.7173, 0.8189]	[0.8983, 0.9607]
Very Preterm Birth	0.7143	0.8413	0.8254
	[0.6574, 0.7712]	[0.7953, 0.8873]	[0.7776, 0.8732]
Moderate Preterm Birth	0.5351	0.6930	0.6491
	[0.4884, 0.5818]	[0.6498, 0.7362]	[0.6044, 0.6938]
Late Preterm Birth	0.5478	0.6647	0.5653
	[0.5258, 0.5698]	[0.6439, 0.6855]	[0.5434, 0.5872]
General Preterm Birth	0.5731	0.6930	0.6324
	[0.5379, 0.6368]	[0.6602, 0.7258]	[0.5981, 0.6667]

into two inside-group modeling comparisons: three updated datasets with updated basic measurements (Datasets V1, V2, and V3), and three updated datasets with both basic and ultrasound measurements included at Visit 2 and Visit 3 (Dataset U1, U2, and U3), see Table 4 in Results section, and Tables 10 and 11 in the Appendix C.

We trained the elastic net regularized logistic regression model on these datasets and the pooled ROC curves for 3 test datasets are shown in Fig. 6. The results revealed that mere updates to BMI, smoking habits, and diagnoses alone slightly improved the AUC scores, indicating minimal enhancement in predictive performance by these factors alone.

However, introducing ultrasound measurements in Datasets 2T and 3T, updated from the second and third visits respectively, showed a marked improvement in model performance. As displayed in Fig. 7, the AUC scores for Datasets 2U and 3U closely match those of Datasets 2 and 3. This demonstrates that while updates to general patient information (such as BMI, smoking habits, and diagnoses) provide minimal improvement, the addition of new ultrasound measurements, particularly cervical length and Pulsatility Index, substantially enhances the predictive capability of the model.

These findings suggest that to significantly improve the model's predictive performance, it is crucial to update ultrasound measurements during subsequent visits. In contrast, updating general patient information alone may not be sufficient to meaningfully enhance the model's accuracy in predicting preterm birth.

Discussion

Our study demonstrates that predictive assessments for preterm birth are most effective at Visit 3 (22–29 gestational weeks), with AUC improving from 0.6161 at Visit 1 to 0.7087 at Visit 3. This performance is comparable to or exceeds previous large-scale studies. Weber et al. [27], using multiple machine learning methods on 2+ million patient records, achieved AUCs of 0.62–0.63 for individual racial-ethnic groups and 0.67 for combined groups. Similarly, Koivu et al. [42], employing state-of-the-art machine learning algorithms (including neural networks and gradient boosting) on nearly 16 million observations, reported an AUC of 0.64 for preterm
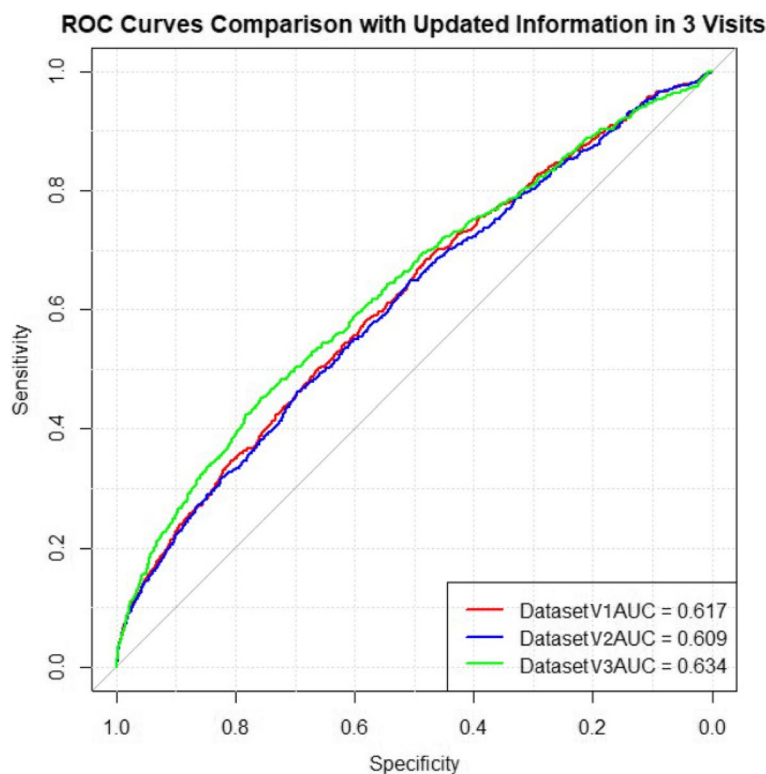
Fig. 6 Pooled ROC curves for model prediction on 3 test datasets with basic measurements updated

birth prediction. While both studies utilized substantially larger datasets and more complex algorithms, our model achieved better performance using only readily available clinical measurements and ultrasound data at Visit 3, suggesting that the timing and type of clinical data may be more crucial than sample size or computational complexity.

Our study's principal finding displays that predictive assessments for preterm birth are most effective when initiated from Visit 3, where prenatal data between 22 and 29 gestational weeks are available. While the AUC scores show improvements, ranging from 0.6161 at the first visit to 0.6425 at the second visit, the most substantial leap is observed from the first to the third visit, reaching 0.7087. Similarly, model accuracy improves from 0.6314 at the first visit to 0.6652 at the third visit, accompanied by a rise in sensitivity from 0.5731 at the first visit to 0.6324 at the third visit and a rise in specificity from 0.6368 at the first visit to 0.6683 in the third visit. These findings suggest that while early prediction attempts at Visit 1 and Visit 2 provide valuable insights, the optimal gains in predictive performance-balancing sensitivity and specificity-are achieved by including data up to Visit 3. Therefore, our results suggest initiating predictive assessments from Visit 3 when prenatal data between 22 and 29 gestational

weeks is available, enhancing both the reliability and accuracy of the predictions.

In the analysis of key variables, those contributed most to the prediction can be explained and aligned with the other research results [68–92]. Throughout the three prenatal visits, tobacco use was a consistent risk factor for preterm birth, despite the low coefficient of the smokeless tobacco use in some fold due to over sparsity – only 3 patients reported their use of smokeless tobacco. Tobacco is by far the most extensively evaluated behavior related influence on preterm birth [68]. Maternal smoking during pregnancy was associated with a 1.27(95%CI 1.21–1.33) times increased risk for preterm birth [69]. Fortunately, studies show that smoking cessation in early pregnancy can reduce this risk, emphasizing the critical need for healthcare interventions and policy support to aid smoking cessation and protect against second-hand smoke [70]. Chronic conditions such as diabetes and hypertension were also linked to increased preterm birth risk [71–75], reinforcing the need for ongoing management in pregnant populations. 'Assisted reproduction' and 'Cervical surgery - LEEP' were also identified as influential factors across the prenatal timeline. Assisted reproductive technologies significantly correlate with preterm birth incidences, even in singleton births [76, 77]. The predictive value of cervical length measurements in forecasting preterm birth aligns
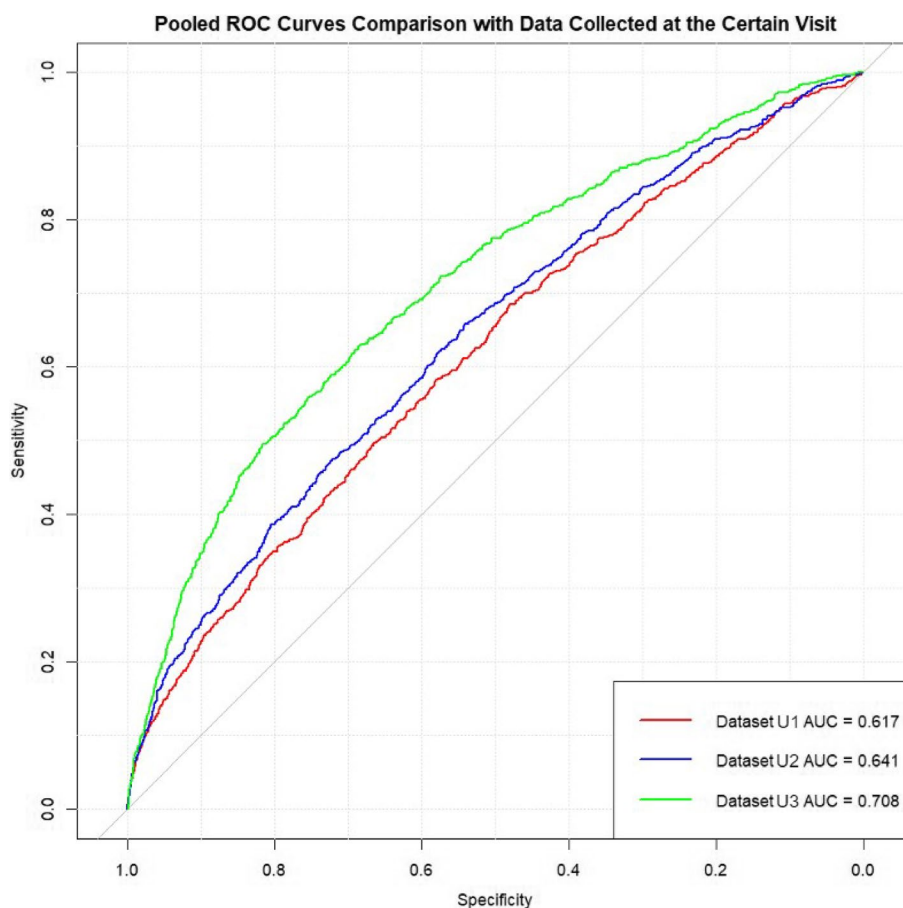
Fig. 7 Pooled ROC curves for model prediction on 3 test datasets with basic and ultrasound measurements updated

closely with existing medical literature. Research demonstrates that changes in cervical length, particularly cervical shortening, can be detected via ultrasound several weeks prior to delivery: for term births, this change is typically observable around 32 weeks of gestation. However, for preterm births, cervical shortening may be detected as early as 16 to 24 weeks [78]. Research indicates that cervical procedures like LEEP can predispose to preterm delivery, potentially due to cervical shortening and local immunological changes post-surgery [79–83]. As for vaginal bleeding, its association with an increased risk of preterm delivery might be due to thrombin-induced uterine contractions or complications from sub-chorionic hematomas leading to placental issues [84–89]. Additionally, the risk associated with being a non-Hispanic Black woman was aligned with extant studies [90–92]. The reasons for this disparity can be complicated, including a complex interaction between genetics, epigenetics, microbiome, and sociodemographic factors contributing to racial disparities in preterm birth rates.

In this sensitivity analysis on subgroups of preterm birth, the high sensitivity for detecting very and extreme preterm births by Visit 3 can be explained by the proximity of Visit 3 to the delivery date for Extreme Preterm Births. Consequently, the data collected during this visit are more indicative of the imminent occurrence of preterm birth. This high sensitivity is crucial as extreme preterm births require the most urgent and intensive care [93]. The varying sensitivity across different preterm birth categories indicates the importance of tailored model development and validation for specific clinical scenarios. The high sensitivity for extreme preterm birth (92.95%, 95% CI: 89.83–96.07%) by Visit 3 is particularly noteworthy, as early detection of these cases carries the most significant clinical implications. This performance characteristic suggests potential utility in prioritizing preventive interventions for the highest-risk cases, though the lower sensitivity for late preterm birth indicates room for improvement in detecting less severe cases.

Comparing the model performance on datasets of successive data and updated data, we find that utilizing basic information variables, basic measurement variables, and ultrasound measurement variables collected during this visit, without the necessity to integrate data from the first two visits, is enough to predict preterm birth. The integration of ultrasound measurements, particularly cervical length and Pulsatility Index, at later visits substantially improved the model's efficacy. This supports our recommendation for routine ultrasound evaluations at later prenatal visits to ensure optimal prediction and preparedness for preterm birth scenarios.

Utilizing logistic regression with elastic net regularization, we combined demographic, lifestyle, obstetric, and medical history data to enhance our model's predictive precision. This approach proved advantageous over more complex models by maintaining interpretability and ease of use, crucial for clinical decision-making [94, 95].

Our study had several strengths. We considered only the variables that are easy to obtain, thus contributing to a reduction in unnecessary resource use. Some studies have used variables with higher cost, for example, Ngo et al. used serum biomarkers [96], and Dabi et al. used acute obstetric changes within days of delivery [97]. We excluded invasive sampling variables, and focused on ones that are easier and cheaper to get. With the performance of our models on the data available at the third visit, we illustrate the potential of a broader practice of cost-efficient variables.

Besides, our model is based on classic logistic regression, which is easier to interpret than deep learning methods. While deep learning models, as explored in Chakoory et al.'s study [98], may offer an efficient preterm birth prediction, they often require complex explanations, such as those provided by SHAP values, to make their outcomes comprehensible. The transparency of our logistic regression models facilitates healthcare delivery by providing interpretable outcomes essential for justifying clinical decisions [94, 95].

We also examined the model performance on the different subgroups of preterm birth, and the varying sensitivity across different preterm birth categories indicates the importance of tailored model development and validation for specific clinical scenarios.

However, important limitations must be acknowledged. NuMom2B dataset not only focused on nulliparous mothers with singleton pregnancy, but also excluded certain cases, for instance, patients with a history of three or more spontaneous abortions and patients with multifetal pregnancy reduction are excluded from the database [32], which means this selection may limit generalizability to broader populations. Future research will aim to broaden the scope and sample size to enhance the generalization of our models.

For now, our model performed well in distinguishing general preterm birth from full-term birth. However, its performance across specific subgroups of preterm birth varies significantly. This variation showcases the necessity for developing personalized strategies tailored to these subgroups [31]. By developing more nuanced predictive models, we can better understand diverse clinical scenarios associated with different stages of preterm birth. Such targeted models not only help in crafting personalized treatment strategies but also provide the essential lead time needed for effective therapeutic interventions. The final models, including detailed coefficients and implementation guidelines, are provided in the Appendix materials for reference and potential application in clinical settings.

From a research perspective, our results open several important avenues for future investigation. The differential performance across preterm birth categories suggests the need for specialized prediction models. Additionally, while our model performed well using only easily obtainable variables, future studies could investigate whether the selective addition of novel biomarkers might improve prediction specifically for those subgroups where our model showed lower sensitivity, while maintaining the overall cost-effectiveness of the approach. Besides, implementation studies examining the model's integration into clinical workflows and its impact on patient outcomes would provide valuable insights into real-world utility.

The implications of our findings extend beyond predictive modeling. First, the model's high sensitivity for extreme preterm birth (92.95% at Visit 3) suggests that implementing this prediction tool during the 22–29 week visit could help clinicians identify patients requiring intensive monitoring and early interventions. This timing aligns with critical clinical decision points, such as the administration of antenatal corticosteroids and the arrangement of maternal transfer to facilities with appropriate neonatal care capabilities. Second, our finding that current visit data alone can achieve strong predictive performance has practical implications for clinical workflow, suggesting that clinicians can make risk assessments using immediately available information rather than requiring extensive historical data compilation. Third, the substantial improvement in model performance with the integration of ultrasound measurements, particularly cervical length and Pulsatility Index, emphasizes the importance of routine ultrasound evaluations during prenatal visits.

This finding supports current clinical guidelines recommending ultrasound screening and suggests that these measurements, when combined with other readily available clinical data, can significantly enhance preterm birth risk assessment without requiring more invasive or costly testing.

Conclusions

Our study applies elastic net regularized logistic regression to predict preterm birth, focusing on easy-to-obtain variables, and emphasizing the significant benefits of incorporating different prenatal data from later visits. Our study demonstrates that predictive assessments for preterm birth are most effective at Visit 3 (22–29 gestational weeks), with AUC improving from 0.6161 at Visit 1 to 0.7087 at Visit 3. This performance is comparable to or exceeds previous large-scale studies. Weber et al. [27], using multiple machine learning methods on 2+ million patient records, achieved AUCs of 0.62–0.63 for individual racial-ethnic groups and 0.67 for combined groups. Similarly, Koivu et al. [42], employing state-of-the-art machine learning algorithms (including neural networks and gradient boosting) on nearly 16 million observations, reported an AUC of 0.64 for preterm birth prediction. While both studies utilized substantially larger datasets and more complex algorithms, our model achieved better performance using only readily available clinical measurements and ultrasound data at Visit 3, suggesting that the timing and type of clinical data may be more crucial than sample size or computational complexity.

By analyzing data across three prenatal periods, we observed that the inclusion of data from the third visit enhances model performance, with stable improvements evident in both sensitivity and specificity metrics, indicating that from Visit 3, we can start predicting preterm birth. Moreover, with data solely from the third visit, our model achieves comparable predictive accuracy to models utilizing data from all three visits. Notably, the integration of ultrasound measurements such as cervical length and the Pulsatility Index significantly bolstered the model's performance, suggesting that we should include ultrasound assessments in late pregnancy stages to ensure optimal prediction and management of preterm births. The model also demonstrates high sensitivity in accurately detecting very and extreme preterm births during the third visit, which means the positive cases of these two subgroups can be correctly predicted, allowing for timely and targeted clinical interventions, such as the administration of necessary medications and the arrangement of appropriate care settings, to help with premature deliveries.

Appendix A: Eight clinical sites

The eight clinical sites are respectively listed as follows: Case Western Reserve University (Site 1); Columbia University (Site 2); Indiana University (Site 3); Magee-Women's Hospital (Site 4); Northwestern University (Site 5); University of California Irvine (Site 6); University of Pennsylvania (Site 7); and University of Utah (Site 8).
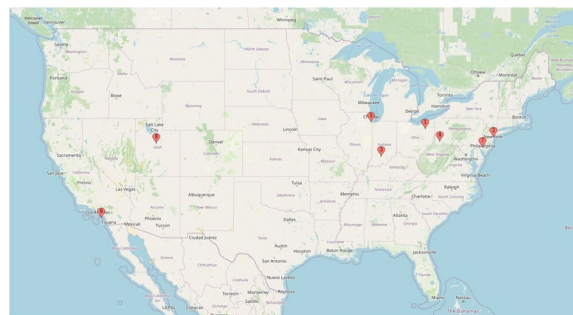
Fig. 8 Geography of 8 clinical centers

Table 9 nuMoM2b clinical site institutional affiliations and subsites

Site 1	**Case Western Reserve University**
	Case Western Reserve University
	The Ohio State University
Site 2	**Columbia University**
	Columbia University
	Christiana Care Health System
Site 3	**Indiana University (no subsites)**
	Indiana University
Site 4	**Magee-Women's Hospital**
	Magee-Women's Hospital
	West Penn Allegheny Health System (quit recruiting in April 2012)
Site 5	**Northwestern University (no subsites)**
	Northwestern University
Site 6	**University of California Irvine**
	University of California Irvine
	Long Beach Memorial Medical Center
	Fountain Valley Regional Medical Center
Site 7	**University of Pennsylvania**
	University of Pennsylvania
	Pennsylvania Hospital
Site 8	**University of Utah**
	University of Utah Health Science Center
	LDS Hospital (quit recruiting in January 2012)
	McKay Dee Hospital
	Utah Valley Regional Medical Center
	Intermountain Medical Center

Appendix B: Bayesian regression tree model

In addition to conventional regression methods, we explored the application of Bayesian Regression Tree(BART) modeling to predict preterm birth. BART combines the robustness of machine learning with the advantages of Bayesian inference, enabling a flexible fitting of regression mean structures through a nonparametric sum-of-trees model. This approach also benefits from the Bayesian framework's capacity for uncertainty quantification and regularization via data-calibrated priors [99].

We employed the bartMachine package in R to construct the models, applying the same 5-fold cross-validation procedure used with logistic regression modeling approaches. The performance of these BART models was evaluated using pooled ROC curves, as depicted in Fig. 9. The results indicate that BART models, similar to the elastic net regularized logistic regression models, show improved performance with the inclusion of data from later prenatal visits. Although the improvement was not superior to that of the logistic regression models, it was consistent with the overall trend observed in predictive enhancements when additional prenatal data were integrated.
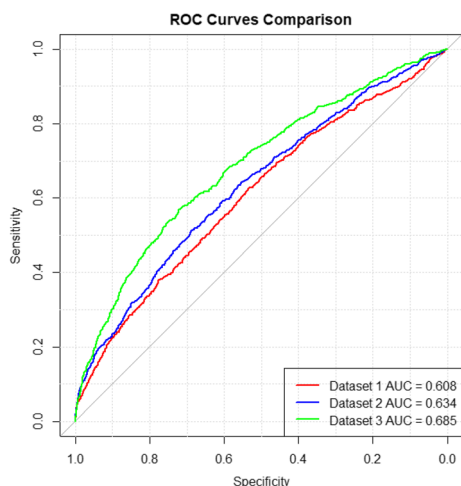
Fig. 9 Pooled ROC curves for model prediction on 3 updated test datasets using BART

Appendix C: Variables used in 3.4

In Tables 10 and 11, we present two comprehensive lists of the variables utilized in the analysis described in Model prediction on test datasets with only updated information section. This table provides a detailed overview of the predictors incorporated into our model, including both demographic factors and ultrasound measurements.

Table 10 Variables in 3 datasets with basic information variables and basic measurement variables updated at each prenatal visit

Dataset V1	Dataset V2	Dataset V3
Height (cm)		
Marital status		
Race Hispanic		
Race Asian		
Race Other		
Race Non-Hispanic White		
Race Non-Hispanic Black		
BMI category at Visit 1		
Number of pregnancies		
Age (years) at visit 1		
Age category (years) at visit 1		
Education status attained		
Income as % of 2013 federal poverty level		
Poverty category based on income and household size		
Vitamin D, mcg		
Vitamin D from supplements		
Pre-gestational diabetes		
Assisted reproduction for this pregnancy		
Previous surgeries - Cervical surgery - Cone		
Previous surgeries - Cervical surgery - LEEP		
Previous surgeries - Cervical surgery - Cryotherapy		
Previous surgeries - Abdominal surgery excluding uterine surgery		
How often do you take Stress-Tabs, B-complex?		
How often do you take Vitamin D, alone, w/ calcium?		
Ever used any tobacco products including cigarettes and smokeless tobacco?		
Smoked tobacco in 3 months prior to pregnancy?		
Used smokeless tobacco in 3 months prior to pregnancy?		
Family history of hypertension		
Family history of Spontaneous preterm delivery (<37 weeks)		
Family history of Early or preterm rupture of membranes		
Family history of Preeclampsia, eclampsia, toxemia or pregnancy-induced hypertension		
BMI at Visit 1	BMI at Visit 2	BMI at Visit 3
Diabetes at Visit 1	Diabetes at Visit 2	Diabetes at Visit 3
Hypertension at Visit 1	Hypertension at Visit 2	Hypertension at Visit 3
Vaginal bleeding until Visit 1	Vaginal bleeding until Visit 2	Vaginal bleeding until Visit 3

Dataset V1	Dataset V2	Dataset V3
Vaginal bleeding days	Vaginal bleeding days since Visit 1	Vaginal bleeding days since Visit 2
Tobacco smoke last month at Visit 1	Tobacco smoke last month at Visit 2	Tobacco use last month at Visit 3
Smokeless tobacco last month at Visit 1	Smokeless tobacco last month at Visit 2	Smokeless tobacco last month at Visit 3

Table 11 Variables in 2 datasets with ultrasound measurements added and updated at visit 2 and visit 3

Dataset U1	Dataset U2	Dataset U3
Height (cm)		
Marital status		
Race Hispanic		
Race Asian		
Race Other		
Race Non-Hispanic White		
Race Non-Hispanic Black		
BMI category at Visit 1		
Number of pregnancies		
Age (years) at visit 1		
Age category (years) at visit 1		
Education status attained		
Income as % of 2013 federal poverty level		
Poverty category based on income and household size		
Vitamin D, mcg		
Vitamin D from supplements		
Pre-gestational diabetes		
Assisted reproduction for this pregnancy		
Previous surgeries - Cervical surgery - Cone		
Previous surgeries - Cervical surgery - LEEP		
Previous surgeries - Cervical surgery - Cryotherapy		
Previous surgeries - Abdominal surgery excluding uterine surgery		
How often do you take Stress-Tabs, B-complex?		
How often do you take Vitamin D, alone, w/ calcium?		
Ever used any tobacco products including cigarettes and smokeless tobacco?		
Smoked tobacco in 3 months prior to pregnancy?		
Used smokeless tobacco in 3 months prior to pregnancy?		
Family history of hypertension		
Family history of Spontaneous preterm delivery (<37 weeks)		
Family history of Early or preterm rupture of membranes		
Family history of Preeclampsia, eclampsia, toxemia or pregnancy-induced hypertension		
BMI at Visit 1	BMI at Visit 2	BMI at Visit 3
Diabetes at Visit 1	Diabetes at Visit 2	Diabetes at Visit 3
Hypertension at Visit 1	Hypertension at Visit 2	Hypertension at Visit 3
Vaginal bleeding until Visit 1	Vaginal bleeding until Visit 2	Vaginal bleeding until Visit 3
Vaginal bleeding days	Vaginal bleeding days since Visit 1	Vaginal bleeding days since Visit 2
Tobacco smoke last month at Visit 1	Tobacco smoke last month at Visit 2	Tobacco use last month at Visit 3

Dataset U1	Dataset U2	Dataset U3
Smokeless tobacco last month at Visit 1	Smokeless tobacco last month at Visit 2	Smokeless tobacco last month at Visit 3
	Cervical length (mm) at Visit 2	Cervical length (mm) at Visit 3
	Left uterine artery PI at Visit 2	Left uterine artery PI at Visit 3
	Right uterine artery PI at Visit 2	Right uterine artery PI at Visit 3

Appendix D: Final predictive models

Table 12 presents a comprehensive list of variable labels and their corresponding meanings used in our study. These variables were employed in the logistic regression models developed using datasets 1, 2, and 3. The coefficients presented in the Tables 13, 14 and 15 represent the average values obtained from these models, providing insight into the relative importance and direction of each predictor's effect on the outcome. Our final logistic regression model uses averaged coefficients from 5-fold cross-validation to predict the probability of preterm birth. The mathematical formula for this model is as follows:

$$P(\text{Preterm Birth}) = \frac{1}{1 + e^{-(\bar{\beta}_0 + \bar{\beta}_1 X_1 + \bar{\beta}_2 X_2 + \ldots + \bar{\beta}_n X_n)}}$$
(1)

where $P(\text{Preterm Birth})$ is the probability of preterm birth, $\bar{\beta}_0$ is the averaged intercept, $\bar{\beta}_1, \bar{\beta}_2, \ldots, \bar{\beta}_n$ are the averaged coefficients corresponding to each predictor variable, and $X_1, X_2, \ldots, X_n$ are the values of the predictor variables. These averaged coefficients are obtained from the 5-fold cross-validation process. This model allows us to estimate the risk of preterm birth based on the combination of demographic factors, medical history, and ultrasound measurements collected across multiple visits. Depending on the extent of data collected up to Visit 1, 2, or 3, the appropriate coefficients can be retrieved from the corresponding coefficient tables (Tables 13, 14, and 15, respectively). This flexibility allows for dynamic risk assessment as more information becomes available throughout the course of prenatal care, potentially enabling more timely and targeted interventions.

Table 12 Variable names and their meanings

Variable name	Variable meaning
PreGestDM	Pre-gestational diabetes
V1AD12b	Previous surgeries - Cervical surgery - LEEP
S02C01	Assisted reproduction for this pregnancy
AgeCat_V1	Age category

Variable name	Variable meaning	Variable name	Variable meaning
Crace_1	Non-Hispanic White	V1AD08a	(At Visit 1) Since you became pregnant, have you had vaginal bleeding more than spotting?
Crace_2	Non-Hispanic Black	V1AG06	(At Visit 1) In the three months prior to this pregnancy did you use smokeless tobacco (chew or snuff)?
Crace_3	Hispanic		
Crace_4	Asian		
Crace_5	Other		
V1AD12e	Previous surgeries - Abdominal surgery excluding uterine surgery	V1AG04	(At Visit 1) Have you ever used any tobacco products including cigarettes and smokeless tobacco?
V1AD12c	Previous surgeries - Cervical surgery - Cryotherapy	V1AH03	(At Visit 1) In the last month, how often have you felt nervous and 'stressed'?
V1AD12a	Previous surgeries - Cervical surgery - Cone		
V2AE06e	Have any of your biological mother, sisters, half-sisters, or female first cousins ever had pregnancy complications - Preeclampsia, eclampsia, toxemia or pregnancy induced hypertension	V2AH05	(At Visit 2) In the last month, did you use smokeless tobacco (chew or snuff)?
		VXXB01ae_V2a	Medical conditions or diagnoses, Diabetes (excluding gestational diabetes in a prior pregnancy) - Condition noted, Visit 2
Education	Education	V2AH02	(At Visit 2) Did you smoke any tobacco products in the last month?
V2AE06d	Have any of your biological mother, sisters, half-sisters, or female first cousins ever had pregnancy complications - Early or preterm rupture of the membranes	U2BB02	(At Visit 2) Cervical length - mm
		V2AD03	(At Visit 2) Since last study visit, have you had vaginal bleeding more than spotting?
Height	Height	BMI_V2	(At Visit 2) BMI
BMI	BMI Category	V2IA07	(At Visit 2) Coping with stress strengthens
V1AG05	Did you smoke any tobacco products in the three months prior to this pregnancy?	V2AE09	Have your father, mother, brother, sister, half-brother or half-sister ever been diagnosed with hypertension (high blood pressure)?
PctFedPoverty	Income as a percentage of 2013 federal poverty level		
Age_at_V1	Age	VXXB01aa_V2a	Medical conditions or diagnoses, High blood pressure (hypertension) - Condition noted, Visit 2
poverty	Poverty category		
VITAMINDAMOUNT	How often do you take Vitamin D, alone, w/ calcium?	V2BA01_KG	(At Visit 2) Weight
V1AF04	What is your current marital status?	V2AE06c	Have any of your biological mother, sisters, half-sisters, or female first cousins ever had pregnancy complications - Spontaneous preterm delivery (<37 weeks)
V1AE01	Including this pregnancy, how many times have you been pregnant - Number of pregnancies		
VITD_MCG	Vitamin D, mcg	U2CD07	(At Visit 2) Right uterine artery - Pulsatility Index (PI)
SUP_VITD	Vitamin D from supplements	V3AD03	(At Visit 3) Since last study visit, have you had vaginal bleeding more than spotting?
VXXB01aa_V1a	Medical conditions or diagnoses, High blood pressure (hypertension) - Condition noted, Visit 1		
VXXB01ae_V1a	Medical conditions or diagnoses, Diabetes (excluding gestational diabetes in a prior pregnancy) - Condition noted, Visit 1	U3BB02	(At Visit 3) Cervical length - mm
		VXXB01aa_V3a	Medical conditions or diagnoses, High blood pressure (hypertension) - Condition noted, Visit 3
V1AG07	(At Visit 1) Did you smoke any tobacco products in the last month?	BMI_V3	(At Visit 3) BMI
V1AD08	(At Visit 1) Since you became pregnant, have you had vaginal bleeding more than spotting?	U3CD07	(At Visit 3) Right uterine artery - Pulsatility Index (PI)
BMI_Cat	BMI Category	V3AF05	(At Visit 3) In the last month, did you use smokeless tobacco (chew or snuff)?
V1AG10	(At Visit 1) In the last month, did you use smokeless tobacco (chew or snuff)?		
		U3CC07	(At Visit 3) Left uterine artery - Pulsatility Index (PI)
STRESSTABSAMOUNT	How often do you take Stress-Tabs, B-complex?		

Huang *et al. BMC Pregnancy and Childbirth*   (2024) 24:843

Page 20 of 25

Variable name	Variable meaning
VXXB01ae_V3a	Medical conditions or diagnoses, Diabetes (excluding gestational diabetes in a prior pregnancy) - Condition noted, Visit 3
V3AG10	(At Visit 3) Perceived Stress Scale - In the last month, how often have you felt difficulties were piling up so high that you could not overcome them?
V3AF02	(At Visit 3) Did you smoke any tobacco products in the last month?
V3AG05	(At Visit 3) Perceived Stress Scale - In the last month, how often have you felt that things were going your way?
V3AG06	(At Visit 3) Perceived Stress Scale - In the last month, how often have you found that you could not cope with all the things that you had to do?
V3AD03a	(At Visit 3) How many days has vaginal bleeding more than spotting happened? - Days
V3AG07	(At Visit 3) Perceived Stress Scale - In the last month, how often have you been able to control irritations in your life?
V3AG09	(At Visit 3) Perceived Stress Scale - In the last month, how often have you been angered because of things that were outside of your control?
V3AG08	(At Visit 3) Perceived Stress Scale - In the last month, how often have you felt that you were on top of things?
V3AG01	(At Visit 3) Perceived Stress Scale - In the last month, how often have you been upset because of something that happened unexpectedly?
U3Dtestcheck	(At Visit 3) Whether uterine artery measurement is done
V3AG04	(At Visit 3) Perceived Stress Scale - In the last month, how often have you felt confident about your ability to handle your personal problems?
V3AG02	(At Visit 3) Perceived Stress Scale - In the last month, how often have you felt that you were unable to control the important things in your life?
V3AG03	(At Visit 3) Perceived Stress Scale - In the last month, how often have you felt nervous and 'stressed'?
V2AD03a	(At Visit 3) How many days has vaginal bleeding more than spotting happened? - Days
U2CC07	(At Visit 2) Left uterine artery - Pulsatility Index (PI)
V3BA01_KG	(At Visit 3) Weight

Table 13 Coefficients for predictors in predictive model for dataset 1

Feature	Coefficient	Odds ratio [95% CI]
PreGestDM	-0.7105	0.4914 [0.4163 - 0.5800]
VXXB01ae_V1a	-0.6642	0.5147 [0.4714 - 0.5619]
VXXB01aa_V1a	-0.5018	0.6054 [0.4943 - 0.7415]
V1AD12b	-0.3531	0.7025 [0.6075 - 0.8123]
V1AD08	-0.2823	0.7540 [0.5735 - 0.9914]
V1AG07	-0.2703	0.7632 [0.6441 - 0.9042]
S02C01	-0.2703	0.7632 [0.7061 - 0.8248]
V1AG10	-0.2642	0.7678 [0.4602 - 1.2811]
V1AD12a	-0.1544	0.8569 [0.7122 - 1.0310]
Crace_2	0.1523	1.1646 [1.1204 - 1.2105]
AgeCat_V1	0.1153	1.1222 [1.0018 - 1.2571]
V1AD12c	-0.1020	0.9030 [0.7427 - 1.0979]
Education	-0.0679	0.9344 [0.8957 - 0.9747]
V1AD12e	-0.0651	0.9370 [0.8535 - 1.0287]
Crace_4	-0.0632	0.9388 [0.7887 - 1.1174]
PctFedPoverty	-0.0420	0.9589 [0.9289 - 0.9899]
BMI_Cat	0.0413	1.0421 [0.9858 - 1.1017]
Height	-0.0388	0.9619 [0.9353 - 0.9893]
Crace_3	-0.0386	0.9622 [0.8837 - 1.0476]
V1AG06	-0.0382	0.9625 [0.8664 - 1.0693]
V1AG05	-0.0295	0.9709 [0.9136 - 1.0318]
Crace_5	0.0285	1.0289 [0.9598 - 1.1030]
STRESSTABSAMOUNT	-0.0211	0.9791 [0.9399 - 1.0200]
poverty	0.0189	1.0191 [0.9687 - 1.0721]
V1AD08a	0.0179	1.0180 [0.9909 - 1.0460]
BMI	0.0154	1.0155 [0.9811 - 1.0511]
V1AH03	0.0137	1.0138 [0.9852 - 1.0433]
V1AE01	0.0123	1.0124 [0.9828 - 1.0429]
Crace_1	-0.0113	0.9888 [0.9523 - 1.0267]
V1AF04	0.0085	1.0086 [0.9771 - 1.0410]
VITAMINDAMOUNT	0.0085	1.0086 [0.9830 - 1.0348]
SUP_VITD	-0.0081	0.9919 [0.9693 - 1.0151]
V1AG04	-0.0065	0.9936 [0.9752 - 1.0123]
Age_at_V1	0.0005	1.0005 [0.9803 - 1.0212]
VITD_MCG	0.0001	1.0001 [0.9873 - 1.0131]

Table 14 Coefficients for predictors in predictive model for dataset 2

Feature	Coefficient	Odds ratio [95% CI]
V2AH05	-0.6355	0.5297 [0.1625 - 1.7264]
PreGestDM	-0.5559	0.5735 [0.4144 - 0.7938]
VXXB01ae_V2a	-0.5111	0.5998 [0.4298 - 0.8371]
V3AD03	-0.3686	0.6917 [0.5139 - 0.9311]
U3BB02	-0.3105	0.7331 [0.6510 - 0.8255]
VXXB01aa_V1a	-0.3030	0.7386 [0.6664 - 0.8186]
V1AD12b	-0.2999	0.7409 [0.6832 - 0.8035]
S02C01	-0.2872	0.7504 [0.6854 - 0.8215]
VXXB01ae_V1a	-0.2859	0.7513 [0.5899 - 0.9569]
VXXB01aa_V3a	-0.2422	0.7849 [0.6514 - 0.9458]

Huang *et al. BMC Pregnancy and Childbirth*   (2024) 24:843

Page 21 of 25

Feature	Coefficient	Odds ratio [95% CI]
BMI_V3	-0.2121	0.8089 [0.7607 - 0.8602]
V1AG07	-0.2059	0.8139 [0.6806 - 0.9732]
V1AD08	-0.2037	0.8157 [0.6213 - 1.0709]
U3CD07	0.1717	1.1873 [1.1351 - 1.2419]
V3AF05	-0.1550	0.8565 [0.6113 - 1.1999]
U3CC07	0.1449	1.1559 [1.1238 - 1.1889]
V2AH02	-0.1273	0.8805 [0.7075 - 1.0957]
AgeCat_V1	0.1182	1.1255 [0.9583 - 1.3217]
U2BB02	-0.1103	0.8956 [0.8684 - 0.9235]
Crace_3	-0.1052	0.9001 [0.7995 - 1.0134]
BMI_Cat	0.0964	1.1012 [0.9674 - 1.2536]
VXXB01ae_V3a	-0.0913	0.9128 [0.7639 - 1.0906]
V2AD03	-0.0790	0.9241 [0.7255 - 1.1769]
Crace_4	-0.0761	0.9267 [0.7477 - 1.1485]
V1AD12e	-0.0717	0.9308 [0.8566 - 1.0116]
V1AD12c	-0.0713	0.9312 [0.8185 - 1.0594]
V1AD12a	-0.0693	0.9330 [0.7810 - 1.1146]
V2AE06e	-0.0693	0.9330 [0.8523 - 1.0215]
Crace_2	0.0675	1.0698 [0.9527 - 1.2014]
U2CC07	0.0669	1.0692 [1.0266 - 1.1135]
Education	-0.0545	0.9469 [0.9216 - 0.9729]
V2AE06d	-0.0480	0.9531 [0.8228 - 1.1041]
Height	-0.0432	0.9578 [0.9279 - 0.9885]
BMI	0.0405	1.0413 [0.9829 - 1.1033]
V3AG10	0.0363	1.0370 [0.9943 - 1.0815]
V1AG10	-0.0347	0.9659 [0.8294 - 1.1247]
V3AF02	-0.0317	0.9688 [0.8433 - 1.1130]
BMI_V2	0.0269	1.0273 [0.9797 - 1.0772]
V3AG05	-0.0237	0.9765 [0.9341 - 1.0209]
V2IA07	0.0196	1.0198 [0.9765 - 1.0651]
STRESSTABSAMOUNT	-0.0189	0.9813 [0.9428 - 1.0213]
V3AG06	-0.0175	0.9827 [0.9387 - 1.0288]
V2AE09	-0.0171	0.9830 [0.9362 - 1.0321]
V3AD03a	0.0168	1.0170 [0.9941 - 1.0403]
V1AG05	-0.0168	0.9833 [0.9317 - 1.0378]
Crace_5	0.0143	1.0144 [0.9335 - 1.1023]
V3AG07	-0.0134	0.9867 [0.9571 - 1.0172]
V1AD08a	0.0126	1.0126 [0.9896 - 1.0362]
V1AG06	-0.0107	0.9893 [0.9439 - 1.0370]
VXXB01aa_V2a	-0.0088	0.9913 [0.9540 - 1.0301]
V3AG09	-0.0086	0.9914 [0.9604 - 1.0235]
PctFedPoverty	-0.0084	0.9917 [0.9560 - 1.0287]
Age_at_V1	0.0069	1.0069 [0.9768 - 1.0380]
V1AG04	-0.0064	0.9936 [0.9661 - 1.0219]
poverty	0.0058	1.0058 [0.9805 - 1.0318]
VITAMINDAMOUNT	0.0053	1.0053 [0.9918 - 1.0190]
V1AF04	0.0051	1.0051 [0.9831 - 1.0276]
V2BA01_KG	0.0047	1.0047 [0.9881 - 1.0216]
Crace_1	-0.0047	0.9953 [0.9752 - 1.0159]
V3AG08	-0.0046	0.9954 [0.9815 - 1.0095]
V1AH03	0.0046	1.0046 [0.9918 - 1.0176]

Feature	Coefficient	Odds ratio [95% CI]
V2AE06c	0.0044	1.0044 [0.9851 - 1.0241]
V1AE01	0.0041	1.0041 [0.9864 - 1.0221]
U2CD07	0.0034	1.0034 [0.9901 - 1.0168]
V3AG01	0.0032	1.0032 [0.9894 - 1.0171]
U3Dtestcheck	-0.0019	0.9981 [0.9900 - 1.0063]
V3AG04	0.0018	1.0018 [0.9941 - 1.0095]
V3AG02	-0.0014	0.9986 [0.9861 - 1.0113]
V3BA01_KG	0.0013	1.0013 [0.9977 - 1.0050]
VITD_MCG	-0.0009	0.9991 [0.9965 - 1.0018]
SUP_VITD	-0.0006	0.9994 [0.9969 - 1.0019]
V3AG03	0.0004	1.0004 [0.9988 - 1.0019]
V2AD03a	0.0003	1.0003 [0.9990 - 1.0016]

Table 15 Coefficients for predictors in predictive model for dataset 3

Feature	Coefficient	Odds ratio [95% CI]
V2AH05	-0.6355	0.5297 [0.1625 - 1.7264]
PreGestDM	-0.5559	0.5735 [0.4144 - 0.7938]
VXXB01ae_V2a	-0.5111	0.5998 [0.4298 - 0.8371]
V3AD03	-0.3686	0.6917 [0.5139 - 0.9311]
U3BB02	-0.3105	0.7331 [0.6510 - 0.8255]
VXXB01aa_V1a	-0.3030	0.7386 [0.6664 - 0.8186]
V1AD12b	-0.2999	0.7409 [0.6832 - 0.8035]
S02C01	-0.2872	0.7504 [0.6854 - 0.8215]
VXXB01ae_V1a	-0.2859	0.7513 [0.5899 - 0.9569]
VXXB01aa_V3a	-0.2422	0.7849 [0.6514 - 0.9458]
BMI_V3	-0.2121	0.8089 [0.7607 - 0.8602]
V1AG07	-0.2059	0.8139 [0.6806 - 0.9732]
V1AD08	-0.2037	0.8157 [0.6213 - 1.0709]
U3CD07	0.1717	1.1873 [1.1351 - 1.2419]
V3AF05	-0.1550	0.8565 [0.6113 - 1.1999]
U3CC07	0.1449	1.1559 [1.1238 - 1.1889]
V2AH02	-0.1273	0.8805 [0.7075 - 1.0957]
AgeCat_V1	0.1182	1.1255 [0.9583 - 1.3217]
U2BB02	-0.1103	0.8956 [0.8684 - 0.9235]
Crace_3	-0.1052	0.9001 [0.7995 - 1.0134]
BMI_Cat	0.0964	1.1012 [0.9674 - 1.2536]
VXXB01ae_V3a	-0.0913	0.9128 [0.7639 - 1.0906]
V2AD03	-0.0790	0.9241 [0.7255 - 1.1769]
Crace_4	-0.0761	0.9267 [0.7477 - 1.1485]
V1AD12e	-0.0717	0.9308 [0.8566 - 1.0116]
V1AD12c	-0.0713	0.9312 [0.8185 - 1.0594]
V1AD12a	-0.0693	0.9330 [0.7810 - 1.1146]
V2AE06e	-0.0693	0.9330 [0.8523 - 1.0215]
Crace_2	0.0675	1.0698 [0.9527 - 1.2014]
U2CC07	0.0669	1.0692 [1.0266 - 1.1135]
Education	-0.0545	0.9469 [0.9216 - 0.9729]
V2AE06d	-0.0480	0.9531 [0.8228 - 1.1041]
Height	-0.0432	0.9578 [0.9279 - 0.9885]
BMI	0.0405	1.0413 [0.9829 - 1.1033]

Huang *et al. BMC Pregnancy and Childbirth*   (2024) 24:843

Page 22 of 25

Feature	Coefficient	Odds ratio [95% CI]
V3AG10	0.0363	1.0370 [0.9943 - 1.0815]
V1AG10	-0.0347	0.9659 [0.8294 - 1.1247]
V3AF02	-0.0317	0.9688 [0.8433 - 1.1130]
BMI_V2	0.0269	1.0273 [0.9797 - 1.0772]
V3AG05	-0.0237	0.9765 [0.9341 - 1.0209]
V2IA07	0.0196	1.0198 [0.9765 - 1.0651]
STRESSTABSAMOUNT	-0.0189	0.9813 [0.9428 - 1.0213]
V3AG06	-0.0175	0.9827 [0.9387 - 1.0288]
V2AE09	-0.0171	0.9830 [0.9362 - 1.0321]
V3AD03a	0.0168	1.0170 [0.9941 - 1.0403]
V1AG05	-0.0168	0.9833 [0.9317 - 1.0378]
Crace_5	0.0143	1.0144 [0.9335 - 1.1023]
V3AG07	-0.0134	0.9867 [0.9571 - 1.0172]
V1AD08a	0.0126	1.0126 [0.9896 - 1.0362]
V1AG06	-0.0107	0.9893 [0.9439 - 1.0370]
VXXB01aa_V2a	-0.0088	0.9913 [0.9540 - 1.0301]
V3AG09	-0.0086	0.9914 [0.9604 - 1.0235]
PctFedPoverty	-0.0084	0.9917 [0.9560 - 1.0287]
Age_at_V1	0.0069	1.0069 [0.9768 - 1.0380]
V1AG04	-0.0064	0.9936 [0.9661 - 1.0219]
poverty	0.0058	1.0058 [0.9805 - 1.0318]
VITAMINDAMOUNT	0.0053	1.0053 [0.9918 - 1.0190]
V1AF04	0.0051	1.0051 [0.9831 - 1.0276]
V2BA01_KG	0.0047	1.0047 [0.9881 - 1.0216]
Crace_1	-0.0047	0.9953 [0.9752 - 1.0159]
V3AG08	-0.0046	0.9954 [0.9815 - 1.0095]
V1AH03	0.0046	1.0046 [0.9918 - 1.0176]
V2AE06c	0.0044	1.0044 [0.9851 - 1.0241]
V1AE01	0.0041	1.0041 [0.9864 - 1.0221]
U2CD07	0.0034	1.0034 [0.9901 - 1.0168]
V3AG01	0.0032	1.0032 [0.9894 - 1.0171]
U3Dtestcheck	-0.0019	0.9981 [0.9900 - 1.0063]
V3AG04	0.0018	1.0018 [0.9941 - 1.0095]
V3AG02	-0.0014	0.9986 [0.9861 - 1.0113]
V3BA01_KG	0.0013	1.0013 [0.9977 - 1.0050]
VITD_MCG	-0.0009	0.9991 [0.9965 - 1.0018]
SUP_VITD	-0.0006	0.9994 [0.9969 - 1.0019]
V3AG03	0.0004	1.0004 [0.9988 - 1.0019]
V2AD03a	0.0003	1.0003 [0.9990 - 1.0016]

Appendix E: TRIPOD (transparent reporting of a multivariable prediction model for individual prognosis or diagnosis) checklist

Section	Item	Checklist item	Page number
Title and Abstract			

Section	Item	Checklist item	Page number
Title	1	Identify the study as developing a prediction model, validation study, or updating a prediction model	Page 1
Abstract	2	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions	Pages 1-2
Methods			
Source of data	3a	Describe the study design or source of data	Data collection section
	3b	Specify the key study dates	Data collection section
Participants	4a	Specify key elements of the study setting	Data collection section
	4b	Describe eligibility criteria for participants	Data collection section
Outcome	5a	Clearly define the outcome that is predicted	Background section
	5b	Report any actions to blind assessment of the outcome	Data collection section
Predictors	6a	Clearly define all predictors	Variables section, Table 2
	6b	Report any actions to blind assessment of predictors	Data collection, Variables sections
Sample size	7	Explain how the study size was arrived at	Sample size considerations section
Missing data	8	Describe how missing data were handled	Data preprocessing and missing data analysis section
Model Development			
Statistical analysis	9a	Describe all steps in model specification	Model training and testing section
	9b	Specify all measures used to assess model performance	Model performance analysis section
Risk groups	10	Describe how risk groups were created	Sensitivity analysis on subgroups of preterm birth section
Results			
Participants	11	Report the number of participants and events	Fig. 1

Huang *et al. BMC Pregnancy and Childbirth*   (2024) 24:843

Page 23 of 25

Section	Item	Checklist item	Page number
Model development	12a	Present the full prediction model	Results section, Appendix D
	12b	Explain how to use the prediction model	Model prediction on test datasets with only updated information section
Model performance	13	Report performance measures with confidence intervals	Model performance is improved when data from later prenatal visits were added section
Discussion			
Limitations	14	Discuss any limitations	Discussion section
Interpretation	15a	Discuss the results	Discussion section
	15b	Give an overall interpretation considering objectives	Discussion section
Implications	16	Discuss clinical usefulness of the model	Discussion section
Other Information			
Supplementary	17	Provide supplementary information	Appendices A-D
Funding	18	Give the source of funding	Funding section

Notes:

- We recommend using the TRIPOD Statement to ensure transparent reporting of the prediction model.
- Items in this checklist have been fulfilled in our manuscript as indicated by the page/section numbers.

Abbreviations

EPV	Events per predictor variable
LMIC	Low-and-middle income country
KMC	Kangaroo mother care
BMI	Body mass index
CPT	Current procedural terminology
EMR	Electronic medical records
EHR	Electronic health records
AUC	Area under the receiver operating characteristics curve
ROC	Receiver operating characteristics
LEEP	Loop electrosurgical excision procedure
KNN	K-Nearest Neighbor
LASSO	East absolute shrinkage and selection operator
ICD-9	Classification of diseases ninth edition
PI	Pulsatility index

Declarations

References
1. Kramer MS, Demissie K, Yang H, Platt RW, Sauvé R, Liston R, et al. The contribution of mild and moderate preterm birth to infant mortality. Jama. 2000;284(7):843–9.
2. Demissie K, Rhoads GG, Ananth CV, Alexander GR, Kramer MS, Kogan MD, et al. Trends in preterm birth and neonatal mortality among blacks and whites in the United States from 1989 to 1997. Am J Epidemiol. 2001;154(4):307–15.
3. Ananth CV, Joseph KS, Oyelese Y, Demissie K, Vintzileos AM. Trends in preterm birth and perinatal mortality among singletons: United States, 1989 through 2000. Obstet Gynecol. 2005;105(5 Part 1):1084–91.
4. Ananth CV, Getahun D, Peltier MR, Salihu HM, Vintzileos AM. Recurrence of spontaneous versus medically indicated preterm birth. Am J Obstet Gynecol. 2006;195(3):643–50.
5. Henderson J, Carson C, Redshaw M. Impact of preterm birth on maternal well-being and women's perceptions of their baby: a population-based survey. BMJ Open. 2016;6(10):e012676.
6. UK NHS. Special care: ill or premature babies. n.d. https://www.nhs.uk/pregnancy/labour-and-birth/after-the-birth/special-care-ill-or-premature-babies/. Accessed 9 Jan 2024.
7. Howson CP, Kinney MV, McDougall L, Lawn JE, Born Too Soon Preterm Birth Action Group. Born too soon: preterm birth matters. Reprod Health. 2013;10:1–9.
8. Rawashdeh H, Awawdeh S, Shannag F, Henawi E, Faris H, Obeid N, et al. Intelligent system based on data mining techniques for prediction of preterm birth for women with cervical cerclage. Comput Biol Chem. 2020;85:107233.
9. Boyle AK, Rinaldi SF, Norman JE, Stock SJ. Preterm birth: Inflammation, fetal injury and treatment strategies. J Reprod Immunol. 2017;119:62–6.
10. Crowley P, Pregnancy C, Group C. Prophylactic corticosteroids for preterm birth. Cochrane Database Syst Rev. 1996;2006(2):CD000065.
11. Rabe H, Reynolds GJ, Diaz-Rosello JL. Early versus delayed umbilical cord clamping in preterm infants. Cochrane Database Syst Rev. 2004;(4).
12. Rabe H, Reynolds G, Diaz-Rosello J. A systematic review and meta-analysis of a brief delay in clamping the umbilical cord of preterm infants. Neonatology. 2008;93(2):138–44.
13. Victora CG, Van Haecke P. Vitamin K prophylaxis in less developed countries: policy issues and relevance to breastfeeding promotion. Am J Public Health. 1998;88(2):203–9.

Huang *et al. BMC Pregnancy and Childbirth*   (2024) 24:843

Page 24 of 25

14. Conde-Agudelo A, Belizán JM, Diaz-Rossello J. Cochrane Review: Kangaroo mother care to reduce morbidity and mortality in low birthweight infants. Evid Based Child Health Cochrane Rev J. 2012;7(2):760–876.

15. Worku B, Kassie A. Kangaroo mother care: a randomized controlled trial on effectiveness of early kangaroo mother care for the low birthweight infants in Addis Ababa, Ethiopia. J Trop Pediatr. 2005;51(2):93–7.

16. Son M, Miller ES. Predicting preterm birth: cervical length and fetal fibronectin, vol. 41. Amsterdam: Elsevier; 2017. p. 445–51.

17. Vivanti AJ, Maraux B, Bornes M, Daraï E, Richard F, Rouzier R. Threatened preterm birth: Validation of a nomogram to predict the individual risk of very preterm delivery in a secondary care center. J Gynecol Obstet Hum Reprod. 2019;48(7):501–7.

18. Koire A, Chu DM, Aagaard K. Family history is a predictor of current preterm birth. Am J Obstet Gynecol MFM. 2021;3(1):100277.

19. Lv M, Chen C, Qiu L, Jin N, Wang M, Zhao B, Chen D, Luo Q. A nomogram to predict extremely preterm birth in women with singleton pregnancies undergoing cervical cerclage. Heliyon. 2022;8(10):e10731.

20. Saroj RK, Anand M. Environmental factors prediction in preterm birth using comparison between logistic regression and decision tree methods: an exploratory analysis. Soc Sci Humanit Open. 2021;4(1):100216.

21. Facco FL, Parker CB, Hunter S, Reid KJ, Zee PP, Silver RM, et al. Later sleep timing is associated with an increased risk of preterm birth in nulliparous women. Am J Obstet Gynecol MFM. 2019;1(4):100040.

22. Esplin MS, Elovitz MA, Iams JD, Parker CB, Wapner RJ, Grobman WA, et al. Predictive accuracy of serial transvaginal cervical lengths and quantitative vaginal fetal fibronectin levels for spontaneous preterm birth among nulliparous women. Jama. 2017;317(10):1047–56.

23. Arabi Belaghi R, Beyene J, McDonald SD. Prediction of preterm birth in nulliparous women using logistic regression and machine learning. PLoS ONE. 2021;16(6):e0252025.

24. Doktorchik C, Premji S, Slater D, Williamson T, Tough S, Patten S. Patterns of change in anxiety and depression during pregnancy predict preterm birth. J Affect Disord. 2018;227:71–8.

25. Jesse DE, Seaver W, Wallace DC. Maternal psychosocial risks predict preterm birth in a group of women from Appalachia. Midwifery. 2003;19(3):191–202.

26. Tran T, Luo W, Phung D, Morris J, Rickard K, Venkatesh S. Preterm birth prediction: Deriving stable and interpretable rules from high dimensional data. Stat. 2016;1050:28. https://doi.org/10.48550/arXiv.1607.08310.

27. Weber A, Darmstadt GL, Gruber S, Foeller ME, Carmichael SL, Stevenson DK, et al. Application of machine-learning to predict early spontaneous preterm birth among nulliparous non-Hispanic black and white women. Ann Epidemiol. 2018;28(11):783–9.

28. Esty A, Frize M, Gilchrist J, Bariciak E. Applying data preprocessing methods to predict premature birth. In: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). New York: IEEE; 2018. p. 6096–9.

29. Gao C, Osmundson S, Edwards DRV, Jackson GP, Malin BA, Chen Y. Deep learning predicts extreme preterm birth from electronic health records. J Biomed Inform. 2019;100:103334.

30. AlSaad R, Malluhi Q, Boughorbel S. PredictPTB: an interpretable preterm birth prediction model using attention-based recurrent neural networks. BioData Min. 2022;15(1):6.

31. Abraham A, Le B, Kosti I, Straub P, Velez-Edwards DR, Davis LK, et al. Dense phenotyping from electronic health records enables machine learning-based prediction of preterm birth. BMC Med. 2022;20(1):333.

32. Haas DM, Parker CB, Wing DA, Parry S, Grobman WA, Mercer BM, et al. A description of the methods of the Nulliparous Pregnancy Outcomes Study: monitoring mothers-to-be (nuMoM2b). Am J Obstet Gynecol. 2015;212(4):539-e1.

33. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol. 1996;49(12):1373–9.

34. Riley RD, Ensor J, Snell KI, Harrell FE, Martin GP, Reitsma JB, et al. Calculating the sample size required for developing a clinical prediction model. Bmj. 2020;368. https://doi.org/10.1136/bmj.m441.

35. Sharifi-Heris Z, Laitala J, Airola A, Rahmani AM, Bender M, et al. Machine learning approach for preterm birth prediction using health records: systematic review. JMIR Med Inform. 2022;10(4):e33875.

36. Lee KS, Ahn KH. Artificial neural network analysis of spontaneous preterm labor and birth and its major determinants. J Korean Med Sci. 2019;34(16):e128. https://doi.org/10.3346/jkms.2019.34.e128.

37. Woolery LK, Grzymala-Busse J. Machine learning for an expert system to predict preterm birth risk. J Am Med Inform Assoc. 1994;1(6):439–46.

38. Grzymala-Busse JW, Woolery LK. Improving prediction of preterm birth using a new classification scheme and rule induction. In: Proceedings of the Annual Symposium on Computer Application in Medical Care. Washington: American Medical Informatics Association (AMIA); 1994. p. 730.

39. Vovsha I, Rajan A, Salleb-Aouissi A, Raja A, Radeva A, Diab H, et al. Predicting preterm birth is not elusive: Machine learning paves the way to individual wellness. In: 2014 AAAI Spring Symposium Series. Menlo Park: AAAI Press; 2014.

40. Frize M, Yu N, Weyand S. Effectiveness of a hybrid pattern classifier for medical applications. Int J Hybrid Intell Syst. 2011;8(2):71–9.

41. Goodwin L, Maher S. Data Mining for Preterm Birth Prediction. vol. 1. 2000. pp. 46–51. https://doi.org/10.1145/335603.335680.

42. Koivu A, Sairanen M. Predicting risk of stillbirth and preterm pregnancies with machine learning. Health Inf Sci Syst. 2020;8:1–12.

43. Khatibi T, Kheyrikoochaksarayee N, Sepehri MM. Analysis of big data for prediction of provider-initiated preterm birth and spontaneous premature deliveries and ranking the predictive features. Arch Gynecol Obstet. 2019;300:1565–82.

44. Oskovi Kaplan ZA, Ozgu-Erdinc AS. Prediction of Preterm Birth: Maternal Characteristics, Ultrasound Markers, and Biomarkers: An Updated Overview. J Pregnancy. 2018;2018:8. https://doi.org/10.1155/2018/8367571.

45. Joseph K, Fahey J, Shankardass K, Allen VM, O'Campo P, Dodds L, et al. Effects of socioeconomic position and clinical risk factors on spontaneous and iatrogenic preterm birth. BMC Pregnancy Childbirth. 2014;14:1–9.

46. Gurung A, Wrammert J, Sunny AK, Gurung R, Rana N, Basaula YN, et al. Incidence, risk factors and consequences of preterm birth-findings from a multi-centric observational study for 14 months in Nepal. Arch Public Health. 2020;78:1–9.

47. Wong TTC, Yong X, Tung JSZ, Lee BJY, Chan JMX, Du R, et al. Prediction of labour onset in women who present with symptoms of preterm labour using cervical length. BMC Pregnancy Childbirth. 2021;21(1):359.

48. Parker MG, Ouyang F, Pearson C, Gillman MW, Belfort MB, Hong X, et al. Prepregnancy body mass index and risk of preterm birth: association heterogeneity by preterm subgroups. BMC Pregnancy Childbirth. 2014;14:1–10.

49. Wouter B, Schut MC, Abu-Hanna A, van Baal JG, van Netten JJ, Bus SA. Development of a prediction model for foot ulcer recurrence in people with diabetes using easy-to-obtain clinical variables. BMJ Open Diabetes Res Care. 2021;9(1):e002257.

50. Hofstaetter C, Dubiel M, Gudmundsson S, Marsal K. Uterine artery color Doppler assisted veloeimetry and perinatal outcome. Acta Obstet Gynecol Scand. 1996;75(7):612–9.

51. Cavoretto P, Salmeri N, Candiani M, Farina A. Reference ranges of uterine artery pulsatility index from first to third trimester based on serial Doppler measurements: longitudinal cohort study. Ultrasound Obstet Gynecol. 2023;61(4):474–80.

52. Rahman MM, Davis DN. Machine learning-based missing value imputation method for clinical datasets. In: IAENG Transactions on Engineering Technologies: Special Volume of the World Congress on Engineering 2012. Cham: Springer Nature; 2013. p. 245–57.

53. Batista GE, Monard MC, et al. A study of K-nearest neighbour as an imputation method. His. 2002;87(251–260):48.

54. Ye C, Fu T, Hao S, Zhang Y, Wang O, Jin B, et al. Prediction of incident hypertension within the next year: prospective study using statewide electronic health records and machine learning. J Med Internet Res. 2018;20(1):e22.

55. 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;350:211–9.

56. Zhang X, Hu W, Wang Y, Wang W, Liao H, Zhang X, et al. Plasma metabolomic profiles of dementia: a prospective study of 110,655 participants in the UK Biobank. BMC Med. 2022;20(1):1–10.

57. Zou H, Hastie T. Regularization and variable selection via the elastic net. J R Stat Soc Ser B Stat Methodol. 2005;67(2):301–20.

58. Ahrens A, Hansen CB, Schaffer ME. lassopack: Model selection and prediction with regularized regression in Stata. Stata J. 2020;20(1):176–235.
59. Goodwin L, Maher S. Data mining for preterm birth prediction. In: Proceedings of the 2000 ACM symposium on Applied computing-Volume 1. New York: Association for Computing Machinery (ACM); 2000. p. 46–51.
60. Airola A, Pahikkala T, Waegeman W, De Baets B, Salakoski T. An experimental comparison of cross-validation techniques for estimating the area under the ROC curve. Comput Stat Data Anal. 2011;55(4):1828–44.
61. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988;44(3):837–45.
62. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982;143(1):29–36.
63. Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. Bmj. 1995;310(6973):170.
64. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011;12:1–8.
65. Kuhn M. Building predictive models in R using the caret package. J Stat Softw. 2008;28:1–26.
66. Youden WJ. Index for rating diagnostic tests. Cancer. 1950;3(1):32–5.
67. World Health Organization. Preterm birth. 2023. https://www.who.int/news-room/fact-sheets/detail/preterm-birth. Accessed 11 Aug 2024
68. Savitz DA, Murnane P. Behavioral influences on preterm birth: a review. Epidemiology. 2010;21(3):291–9.
69. Shah NR, Bracken MB. A systematic review and meta-analysis of prospective studies on the association between maternal cigarette smoking and preterm delivery. Am J Obstet Gynecol. 2000;182(2):465–72.
70. Wagijo Ma, Sheikh A, Duijts L, Been JV. Reducing tobacco smoking and smoke exposure to prevent preterm birth and its complications. Paediatr Respir Rev. 2017;22:3–10.
71. Berger H, Melamed N, Davis BM, Hasan H, Mawjee K, Barrett J, et al. Impact of diabetes, obesity and hypertension on preterm birth: Population-based study. PLoS ONE. 2020;15(3):e0228743.
72. Hedderson MM, Ferrara A, Sacks DA. Gestational diabetes mellitus and lesser degrees of pregnancy hyperglycemia: association with increased risk of spontaneous preterm birth. Obstet Gynecol. 2003;102(4):850–6.
73. Chatzi L, Plana E, Pappas A, Alegkakis D, Karakosta P, Daraki V, et al. The metabolic syndrome in early pregnancy and risk of gestational diabetes mellitus. Diabetes Metab. 2009;35(6):490–4.
74. Yanit K, Cheng YW, Snowden J, Caughey AB. 781: The impact of chronic hypertension and diabetes mellitus on pregnancy outcomes. Am J Obstet Gynecol. 2012;206(1):S344.
75. Bodnar LM, Catov JM, Klebanoff MA, Ness RB, Roberts JM. Prepregnancy body mass index and the occurrence of severe hypertensive disorders of pregnancy. Epidemiology. 2007;18(2):234–9.
76. Blickstein I. Does assisted reproduction technology, per se, increase the risk of preterm birth? BJOG Int J Obstet Gynaecol. 2006;113:68–71.
77. Bu Z, Zhang J, Hu L, Sun Y. Preterm birth in assisted reproductive technology: an analysis of more than 20,000 singleton newborns. Front Endocrinol. 2020;11:558819.
78. Iams JD. Prediction and early detection of preterm labor. Obstet Gynecol. 2003;101(2):402–12.
79. Jakobsson M, Gissler M, Paavonen J, Tapper AM. Loop electrosurgical excision procedure and the risk for preterm birth. Obstet Gynecol. 2009;114(3):504–10.
80. Jin G, LanLan Z, Li C, Dan Z. Pregnancy outcome following loop electrosurgical excision procedure (LEEP) a systematic review and meta-analysis. Arch Gynecol Obstet. 2014;289:85–99.
81. Crane JM, Delaney T, Hutchens D. Transvaginal ultrasonography in the prediction of preterm birth after treatment for cervical intraepithelial neoplasia. Obstet Gynecol. 2006;107(1):37–44.
82. Parikh R, Horne H, Feinstein SJ, Anasti JN. Cervical length screening in patients who have undergone loop electrosurgical excision procedure. J Reprod Med. 2008;53(12):909–13.
83. Sadler L, Saftlas A, Wang W, Exeter M, Whittaker J, McCowan L. Treatment for cervical intraepithelial neoplasia and risk of preterm delivery. Jama. 2004;291(17):2100–6.
84. La Verde M, Riemma G, Torella M, Torre C, Cianci S, Conte A, et al. Impact of Braxton-Hicks contractions on fetal wellbeing; a prospective analysis through computerised cardiotocography. J Obstet Gynaecol. 2022;42(4):569–73.
85. Gibb D, Arulkumaran S. ASSESSMENT OF UTERINE CONTRACTIONS. In: Gibb D, Arulkumaran S, editors. Fetal Monitoring in Practice - E-Book: Fetal Monitoring in Practice - E-Book. 5th ed. Elsevier Health Sciences; 2023. p. 106–11.
86. Hossain R, Harris T, Lohsoonthorn V, Williams MA. Risk of preterm delivery in relation to vaginal bleeding in early pregnancy. Eur J Obstet Gynecol Reprod Biol. 2007;135(2):158–63.
87. Rosen T, Kuczynski E, O'Neill L, Funai E, Lockwood C. Plasma levels of thrombin-antithrombin complexes predict preterm premature rupture of the fetal membranes. J Matern Fetal Med. 2001;10(5):297–300.
88. Saraswat L, Bhattacharya S, Maheshwari A, Bhattacharya S. Maternal and perinatal outcome in women with threatened miscarriage in the first trimester: a systematic review. BJOG Int J Obstet Gynaecol. 2010;117(3):245–57.
89. Weiss JL, Malone FD, Vidaver J, Ball RH, Nyberg DA, Comstock CH, et al. Threatened abortion: a risk factor for poor pregnancy outcome, a population-based screening study. Am J Obstet Gynecol. 2004;190(3):745–50.
90. Egbe TI, Montoya-Williams D, Wallis K, Passarella M, Lorch SA. Risk of extreme, moderate, and late preterm birth by maternal race, ethnicity, and nativity. J Pediatr. 2022;240:24–30.
91. Schaaf JM, Liem SM, Mol BWJ, Abu-Hanna A, Ravelli AC. Ethnic and racial disparities in the risk of preterm birth: a systematic review and meta-analysis. Am J Perinatol. 2013;30(06):433–50.
92. Manuck TA. Racial and ethnic differences in preterm birth: a complex, multifactorial problem. In: Seminars in perinatology, vol. 41. Amsterdam: Elsevier; 2017. p. 511–8.
93. Baron IS, Rey-Casserly C. Extremely preterm birth outcome: a review of four decades of cognitive research. Neuropsychol Rev. 2010;20:430–52.
94. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. Nat Med. 2019;25(1):30–6.
95. Genther-Yoshida P, Casassa M, Shull R, Pomrenke G, Thomas I, Price R, et al. National Science and Technology Council Committee on Technology The Interagency Working Group on Nanoscience, Engineering and Technology September 1999, Washington, DC About the National Science and Technology Council. Washington, DC: National Science and Technology Council; 1999.
96. Ngo TT, Moufarrej MN, Rasmussen MLH, Camunas-Soler J, Pan W, Okamoto J, et al. Noninvasive blood tests for fetal development predict gestational age and preterm delivery. Science. 2018;360(6393):1133–6.
97. Dabi Y, Nedellec S, Bonneau C, Trouchard B, Rouzier R, Benachi A. Clinical validation of a model predicting the risk of preterm delivery. PLoS ONE. 2017;12(2):e0171801.
98. Chakoory O, Barra V, Rochette E, Blanchon L, Sapin V, Merlin E, et al. DeepMPTB: a vaginal microbiome-based deep neural network as artificial intelligence strategy for efficient preterm birth prediction. Biomark Res. 2024;12(1):25.
99. Hill J, Linero A, Murray J. Bayesian additive regression trees: A review and look forward. Ann Rev Stat Appl. 2020;7:251–78.

Publisher's Note