


# Evaluating the performance of an AI-powered VBAC prediction system within a decision-aid birth choice platform for shared decision-making

DIGITAL HEALTH  
Volume 10: 1–10  
© The Author(s) 2024  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/20552076241257014  
journals.sagepub.com/home/dhj



Cherng Chia Yang<sup>1</sup>, Ching Fu Wang<sup>2</sup>, Wei Ming Lin<sup>3</sup>, Shu Wen Chen<sup>4</sup>   
and Hsiang Wei Hu<sup>5</sup>

## Abstract

**Background:** Vaginal birth after cesarean (VBAC) is generally regarded as a safe and viable birthing option for most women with prior cesarean delivery. Nonetheless, concerns about heightened risks of adverse maternal and perinatal outcomes have often dissuaded women from considering VBAC. This study aimed to assess the performance of an artificial intelligence (AI)-powered VBAC prediction system integrated into a decision-aid birth choice platform for shared decision-making (SDM).

**Materials and Methods:** Employing a retrospective design, we collected medical records from a regional hospital in northern Taiwan from January 2019 to May 2023. To explore a suitable model for tabular data, we compared two prevailing modeling approaches: tree-based models and logistic regression models. We subjected the tree-based algorithm, CatBoost, to binary classification.

**Results:** Forty pregnant women with 347 records were included. The CatBoost model demonstrated a robust performance, boasting an accuracy rate of 0.91 (95% confidence interval (CI): 0.86–0.94) and an area under the curve of 0.89 (95% CI: 0.86–0.93), surpassing both regression models and other boosting techniques. CatBoost captured the data characteristics on the significant impact of gravidity and the positive influence of previous vaginal birth, reinforcing established clinical guidelines, as substantiated by the SHapley Additive exPlanations analysis.

**Conclusion:** Using AI techniques offers a more accurate assessment of VBAC risks, boosting women's confidence in selecting VBAC as a viable birthing option. The seamless integration of AI prediction systems with SDM platforms holds a promising potential for enhancing the effectiveness of clinical applications in the domain of women's healthcare.

## Keywords

Vaginal birth after cesarean, elective repeat cesarean delivery, shared decision making, artificial intelligence prediction, pregnant women

Submission date: 8 October 2023; Acceptance date: 8 May 2024

<sup>1</sup>Department of Obstetrics and Gynecology, Saint Paul's Hospital, Taoyuan

<sup>2</sup>Biomedical Engineering Research and Development Center, National Yang Ming Chiao Tung University, Taipei

<sup>3</sup>Department of Information Management, I-Shou University, Chiayi

<sup>4</sup>School of Nursing, National Taipei University of Nursing and Health Sciences, Taipei

<sup>5</sup>Biomedical Technology and Device Research Laboratories, Industrial Technology Research Institute, Hsinchu

## Corresponding authors:

Shu Wen Chen, School of Nursing, National Taipei University of Nursing and Health Sciences, No. 365, Ming-Te Road, Peitou District, Taipei City, 11219.

Email: shuwen@ntunhs.edu.tw

Hsiang Wei Hu, Biomedical Technology and Device Research Laboratories, Industrial Technology Research Institute, No. 195, Chung Hsing Road, Chutung District, Hsinchu City 31057.

Email: itriB20131@itri.org.tw



## Introduction

Vaginal birth after cesarean (VBAC) is a safe and viable birthing option for the majority of women following a previous caesarean delivery (CD).<sup>1</sup> This choice not only offers potential benefits for the women, but also plays a crucial role in reducing unnecessary elective repeat cesarean delivery (ERCD), which is associated with an increased risk of adverse maternal and perinatal mortality and complications.<sup>2,3</sup> Nevertheless, the decision-making processes regarding VBAC are complex and multifaceted, necessitating a careful consideration of various medical, obstetric, and individual factors.<sup>4,5</sup>

Artificial intelligence (AI) has emerged as a powerful tool in healthcare, offering the potential to support medical professionals and patients in making informed decisions.<sup>4,5</sup> Drawing on previous research findings,<sup>6-9</sup> we created an innovative web-based platform serving as the basis for a decision-aid (DA) system for birth choices aimed at fostering shared decision-making (SDM).<sup>10</sup> Through this DA platform, we strive to enhance understanding, boost confidence, and improve communication between medical professionals and pregnant women. The DA platform could be easily operated for pregnant women and healthcare professionals through mobile devices, such as smartphones, tablets, and computers.<sup>10</sup> The DA platform is equipped with essential features, including a video to introduce SDM, an overview of the functions and features of the birth DA, and comprehensive information on the risks and benefits of VBAC and ERCD.<sup>10</sup> Most importantly, the platform incorporates an AI calculator to empower SDM between healthcare providers and pregnant women by providing personalized, data-driven insights and risk assessments for VBAC.<sup>10</sup>

Although several VBAC prediction models have been developed, their generalizability and applicability remain unclear<sup>4,5</sup>; thus, high-quality internal and external validation studies are needed. Most importantly, none of these VBAC prediction models intergrade the AI technique into the DA platform. Thus, this study aims to evaluate the performance of an AI-powered VBAC prediction system with a DA birth choice platform for SDM.

## Materials and methods

### Study design

We conducted a retrospective study with a quantitative and descriptive design. The DA platform was implemented on a cloud server with a responsive web design (RWD) to ensure the stability and reliability of the system.<sup>11,12</sup> This platform consisted of four key components: an interactive web service (IWS)<sup>13</sup> for SDM; a responsive web design (RWD) for data collection and visualization; a cloud-based VBAC database for collaborative use among multiple users;

and a VBAC prediction system enhanced with AI techniques (Figure 1). The VBAC prediction system leverages cloud-based AI to compute and employ a predictive model, which assesses the likelihood of a successful vaginal birth during pregnancy.

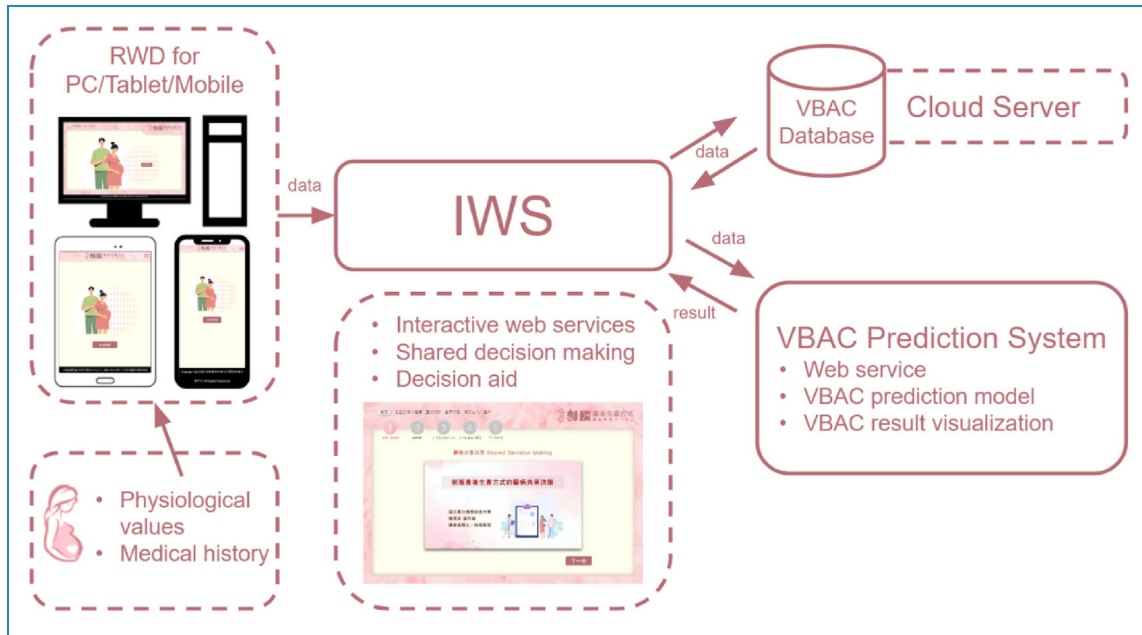
The interactive web service (IWS) was tailored for AI-powered predictions, utilizing medical record data and a machine learning model,<sup>14</sup> notably CatBoost, with finely tuned hyperparameters. Its primary function is to anticipate the probability of a successful VBAC and identify potential birth-related risks for pregnant women. The seamless integration of AI and web services has empowered healthcare professionals to provide informed decision-making support to pregnant women (Figure 2).

### Data sources and collection

Prior to the commencement of the study, ethical approval was obtained from the Institutional Review Board of Fu Jen Catholic University (No. C111042). To conduct the training and testing verification of the AI model, this study collected and used retrospective medical record data from the regional hospital in northern Taiwan between Jan 2019 and May 2023. A total of 44 pregnant with 386 record data met the study criteria. The inclusion criteria were (1) pregnant women, (2) aged between 20 and 45 years, and (3) had a previous CD. The exclusion criteria were (1) multiple pregnancies and (2) previous classical CD or myomectomy surgery. These criteria were considered based on their potential impact on the outcomes of the study. After excluding cases with major fetal anomalies, intrauterine fetal demise, gestational weeks less than 12, and those with missing data, our final analysis included a total of 40 pregnant women with 347 recorded data points constituting 89.89% of the originally considered dataset (Figure 3). Among the trial of labor after cesarean (TOLAC) attempts, 117 (33.72%) led to a successful vaginal birth referred to as the 'VBAC Success' group, while 230 (66.28%) resulted in a failed TOLAC attempt designated as the 'VBAC Failure' group. The clinical information with statistically significant differences between these two groups can be found in Table 1. The variables in this study can be categorized into two main types: continuous variables and categorical variables. Continuous variables encompass maternal age, gestational week, and BMI, while categorical variables encompass gravidity, previous vaginal birth, prolonged labor, gestational diabetes, and chronic hypertension.

### Development of machine learning models with hyperparameter design and architectural comparison

The model's input comprised seven variables, as referenced in a previous study.<sup>30</sup> These variables included maternal



**Figure 1.** Web-based decision-aid platform comprising four components: The **Interactive Web Service (IWS)** facilitated shares decision-making between doctors and pregnant women and provides decision aids. The platform uses **Responsive Web Design (RWD)** to ensure optimal performance on personal computers, tablets, and mobile devices. The RWD also supports the storage of pregnant women's medical history and physiological data, which are achieved through integration with the **VBAC database system** on a cloud server. The **VBAC prediction system** utilizes cloud-based AI technology to compute and employ a predictive model that determines the probability of a successful vaginal birth during pregnancy.

age, gravidity, BMI, previous vaginal birth, prolonged labor, gestational diabetes, and chronic hypertension. Furthermore, recognizing that the BMI naturally increases as the gestational weeks progress, a necessary adjustment was made to account for this variable. This adjustment involved dividing the BMI by the gestational week, ensuring that this variable did not disproportionately influence the model predictions across different gestational weeks and periods.

This research primarily focused on developing and refining machine learning models designed to predict outcomes using tabular data sources, as indicated in previous studies.<sup>15,16</sup> The dataset analyzed displayed a tabular structure and had previously shown enhancements when utilized with tree-based models well-known for their ability to effectively classify data with diverse features. Drawing upon insights from prior studies,<sup>17</sup> we incorporated the boosting technique along with the GridSearch method. This strategic approach allowed us to finely tune the hyperparameters, ultimately optimizing the performance. As an example, we made adjustments to various hyperparameters, including depth, (e.g., CatBoost with choices [6, 8, 10]), learning rate (options: [0.01, 0.05, 0.1]), l2\_leaf\_reg (values: [2, 3, 4]), iterations (alternatives: [500, 1000]), and loss function (selections: ['Logloss']). This approach enabled us to incorporate non-linear parameter adjustments, which could potentially result in improved training outcomes. Additionally, the attributes within the tabular

dataset used in this study were recognized for their relatively distinct and independent characteristics. Consequently, we conducted a comparative analysis between two widely adopted model categories: tree-based models and logistic regression. Specifically, we evaluated the performance of six distinct algorithms, namely, logistic regression and five tree-based models, that is, decision tree, random forest,<sup>18</sup> eXtreme gradient boosting (XGBoost), light gradient boosting machine (LightGBM),<sup>19</sup> and CatBoost,<sup>20</sup> employing them for the binary classification tasks.<sup>21</sup>

### *Evaluation of model performance and interpretation with SHapley Additive exPlanations-based feature selection*

In this phase, we assessed the performance of the VBAC models utilizing a range of machine learning algorithms, including logistic regression, decision tree, random forest, XGBoost, LightGBM, and CatBoost. We relied on two primary metrics, namely, accuracy and the area under the curve (AUC), as the key indicators to gauge the effectiveness of our optimized models. To enhance the interpretability of these models, we employed SHapley Additive exPlanations (SHAP)-based feature selection. This method was instrumental in identifying the most influential features contributing to the models' predictions.<sup>22</sup> The SHAP values offer a unified measure of feature importance that can be consistently applied across different model

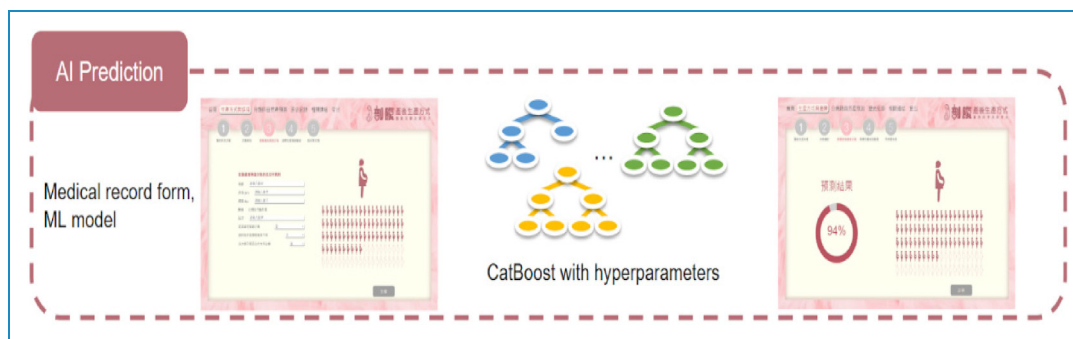


Figure 2. AI prediction with medical record form and ML model.

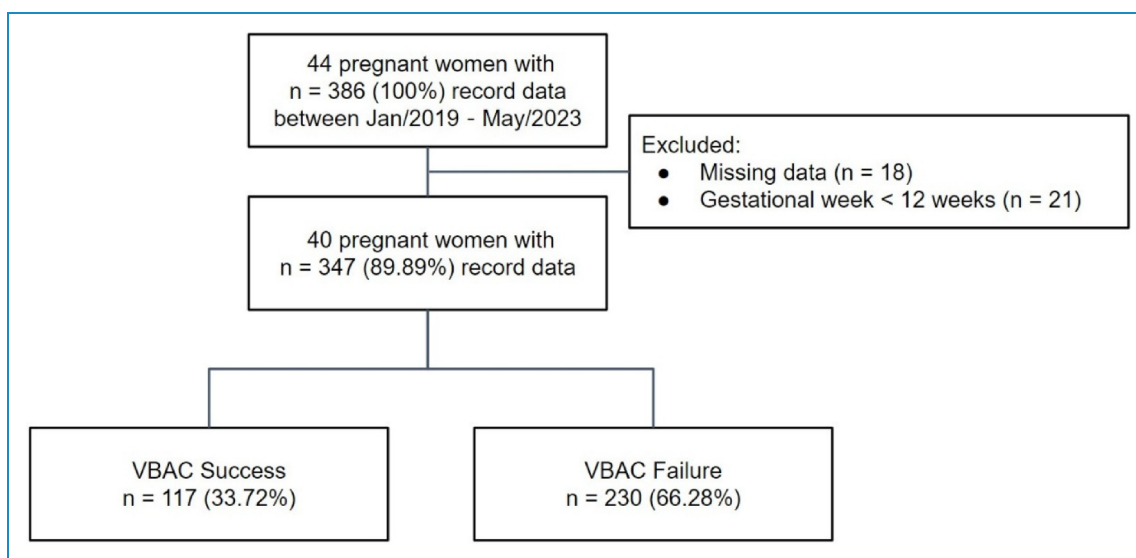


Figure 3. Recruitment flow chart for the AI model.

types.<sup>23</sup> The feature selection procedure was carried out using the SHAP library in Python. To comprehensively evaluate the models' performance, we employed a range of metrics, including accuracy, sensitivity, specificity, precision, positive likelihood ratio (LR+), negative likelihood ratio (LR-), and AUC.<sup>24</sup> To assess the robustness and generalizability of the models, we conducted a bootstrapping testing procedure involving 1000 resamples of the testing dataset. This process enabled us to estimate the confidence intervals (CIs) related to the accuracy and AUC metrics, providing insights into the models' stability and their ability to perform effectively on new and unseen data.

### Statistical analysis

Statistical analysis was conducted using SciPy modules version 1.11.2 in Python. To establish statistical significance, a significance level of  $p < 0.05$  was applied for the two-tailed test. The evaluation of continuous variables, such as maternal age, gravidity, and BMI, was carried out using the Wilcoxon rank-sum test. Nominal categorical

variables like previous vaginal birth, prolonged labor, pre-gestational diabetes, and chronic hypertension were assessed using the Fisher's exact test, and two-sided  $p$ -values were computed.

## Results

### Performance validation comparison between ML models

A total of 40 subjects with 347 data were included in the testing cohort (40, 89.89%). Seven variables were utilized as the input factors for the six ML methods of logistic regression, random forest, XGBoost, LightGBM, decision tree, and CatBoost. In Table 2, the CatBoost model demonstrated the utmost efficacy within the experimental framework, yielding a commendable accuracy of 0.91 (95% CIs: 0.86–0.94). In contrast, the alternative models of logistic regression, random forest, XGBoost, LightGBM, and decision tree achieved comparatively modest accuracy values of 0.69 (95% CIs: 0.65–0.74), 0.88 (95% CIs:

**Table 1.** Clinical information with statistic difference between VBAC success and failure group.

	VBAC success ( <i>n</i> = 117)	VBAC failure ( <i>n</i> = 230)	<i>p</i> -value	Total ( <i>n</i> = 347)
Maternal age [y, [min-max]]	34.6 ± 3.7 [29-42]	34.9 ± 3.6 [28-45]	0.44	34.8 ± 3.7 [28-45]
Gestational week [w, [min-max]]	27.5 ± 8.2 [12-40]	28.6 ± 8.0 [12-39]	0.22	28.3 ± 8.1 [12-40]
Gravidity [ <i>n</i> , [min-max]]	3.2 ± 2.0 [1-8]	2.7 ± 0.9 [2-5]	0.74	2.9 ± 1.4 [1-8]
1 [ <i>n</i> , (%)]	10 (8.5)	0 (0)		10 (2.9)
2 [ <i>n</i> , (%)]	53 (45.3)	114 (49.6)		167 (48.1)
≥3 [ <i>n</i> , (%)]	54 (46.2)	116 (50.4)		170 (49.0)
BMI [kg/m <sup>2</sup> , [min-max]]	24.9 ± 3.0 [18.4-35.3]	26.5 ± 4.0 [18.7-38.6]	<0.001	26.0 ± 3.8 [18.4-38.6]
Previous vaginal birth [ <i>n</i> , (%)]	28 (23.93)	40 (17.39)	0.15	68 (19.60)
Prolonged labor [ <i>n</i> , (%)]	12 (10.26)	53 (23.04)	0.004	65 (18.73)
Pregestational diabetes [ <i>n</i> , (%)]	10 (8.55)	0 (0.00)	<0.001	10 (2.88)
Chronic hypertension [ <i>n</i> , (%)]	3 (2.56)	16 (6.96)	0.13	19 (5.48)

Continuous variables are expressed as mean ± standard deviation and range, including maternal age, gestational week, and BMI.

Categorical variables are presented as number *n* (%), including gravidity, previous vaginal birth, prolonged labor, pregestational diabetes, and chronic hypertension.

CI: confidence interval; VBAC: vaginal birth after cesarean.

**Table 2.** Performance comparison for the VBAC classification based on different ML models. The ML models were six kinds of architecture, namely, logistic regression, random forest, XGBoost, LightGBM, decision tree, and CatBoost.

Model	Accuracy	Sensitivity	Specificity	Precision	LR +	LR–	AUC
Logistic regression	0.69 (0.65–0.74)	0.14	1.00	1.00	NaN	0.86	0.57 (0.54–0.6)
Random forest	0.88 (0.85–0.92)	0.73	0.97	0.93	24.45	0.28	0.85 (0.81–0.89)
XGBoost	0.82 (0.78–0.85)	0.76	0.85	0.74	5.07	0.29	0.80 (0.76–0.85)
LightGBM	0.90 (0.87–0.93)	0.84	0.94	0.89	14.03	0.17	0.89 (0.85–0.93)
Decision tree	0.83 (0.79–0.86)	0.65	0.93	0.83	8.69	0.38	0.79 (0.74–0.83)
CatBoost	0.91 (0.86–0.94)	0.81	0.97	0.94	27.16	0.20	0.89 (0.86–0.93)

tn: true negative was expressed in the model predict number for the VBAC failure group.

tp: true positive was expressed in the model predict number for the VBAC success group.

fn: false negative was expressed in the model predict number for the VBAC failure group.

fp: false positive was expressed in the model predict number for the VBAC success group.

Accuracy is (tn + tp) / (fn + fp). Sensitivity is tp / (tp + fp). Specificity is tn / (tn + fp). Positive likelihood ratio (LR +) is sensitivity / (1 – specificity). Negative likelihood ratio (LR–) is (1 – sensitivity) / specificity.

0.85–0.92), 0.82 (95% CIs: 0.78–0.85), 0.90 (95% CIs: 0.87–0.93), and 0.83 (95% CIs: 0.79–0.86), respectively, as discerned from the outcomes obtained within the testing cohort of the study and the AUC values of 0.57

(95% CIs: 0.54–0.6), 0.85 (95% CIs: 0.81–0.89), 0.80 (95% CIs: 0.76–0.85), 0.89 (95% CIs: 0.85–0.93), and 0.79 (95% CIs: 0.74–0.83). Confusion matrices for each ML model and ROC curves for the performance evaluation



of each ML model were plotted in Figures 4 and 5, respectively. Furthermore, the sensitivity, specificity, precision, LR+ and LR− were calculated to further evaluate the performance of the six models. The results are also presented in Table 2. Despite not achieving the highest sensitivity among the various models examined, it is noteworthy that the CatBoost model's enhanced performance can be attributed to the intricate interplay of hyperparameters associated with the boosting technique, a phenomenon similarly observed in the case of XGBoost and LightGBM. To encapsulate the findings, when evaluating pivotal metrics, such as accuracy and AUC, the CatBoost model emerged as the discerning choice for the incorporation into our innovative decision-aid platform as the designated AI predictive model.

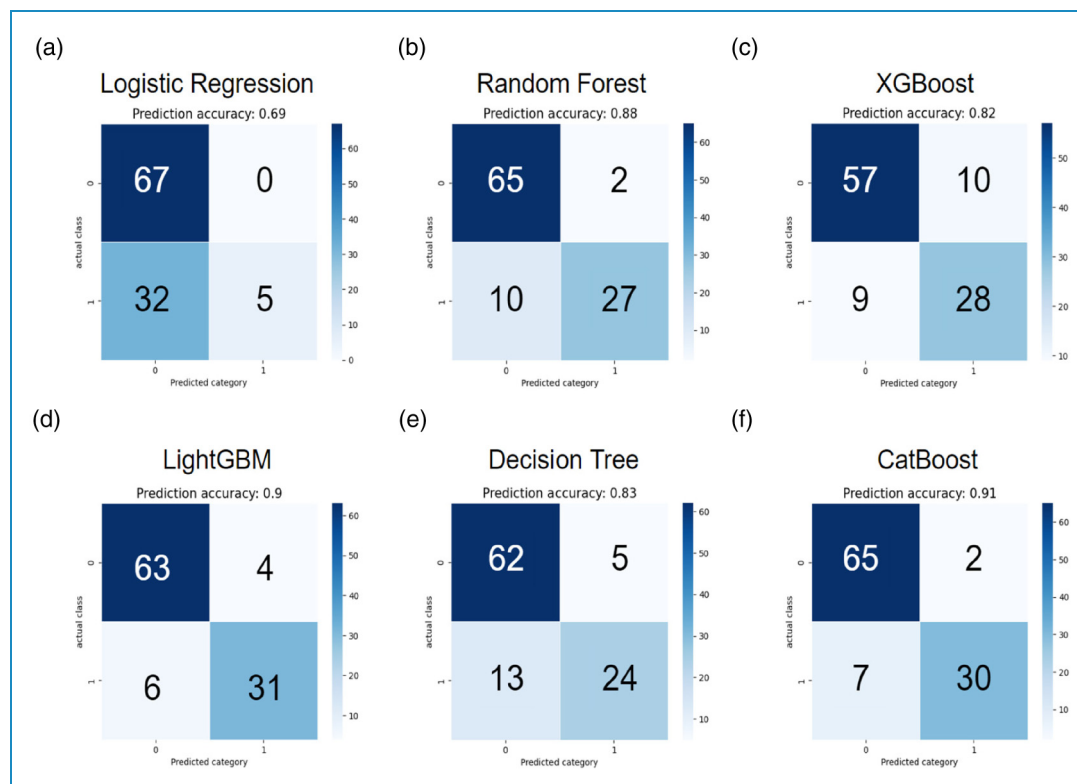
### Model explainability comparison for each ML model

In this study, we aimed to provide a deeper understanding of the ML models used for predicting mortality by interpreting its black box using SHAP values. The SHAP summary plot ranks feature importance and is presented in Figure 6. On the right side of Figure 6, the top four most significant variables contributing to the model were identified as maternal age, gravidity, BMI, and previous vaginal birth in random forest, XGBoost, LightGBM, and CatBoost. Additionally, gravidity emerged as the most significant

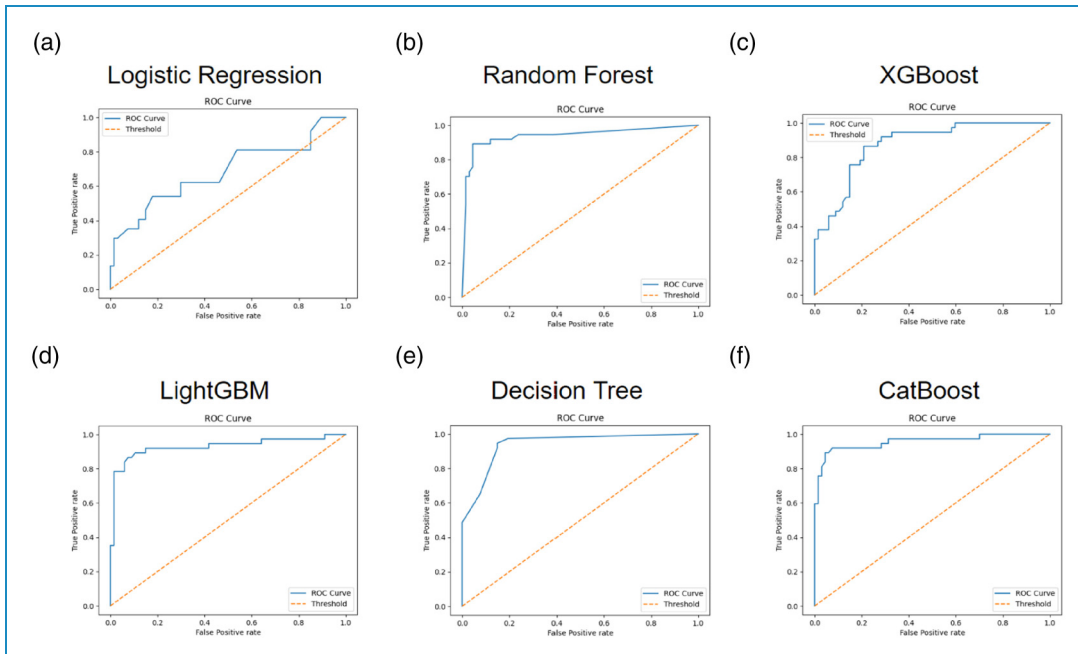
variable in CatBoost, while maternal age held this distinction in random forest, XGBoost, and LightGBM. Notably, features like gestational diabetes were among the top four most significant variables in the logistic regression and decision tree models. As depicted on the left side of Figure 6, we present more detailed results regarding the top four most important clinical features categorized based on their positive and negative impact factors, which influence the predictive output of the ML models. Previous vaginal birth was the most positively significant variable in random forest, LightGBM, decision tree, and CatBoost. Conversely, gravidity had the most negative impact on logistic regression, while maternal age held the most positive significance in XGBoost.

### Discussion

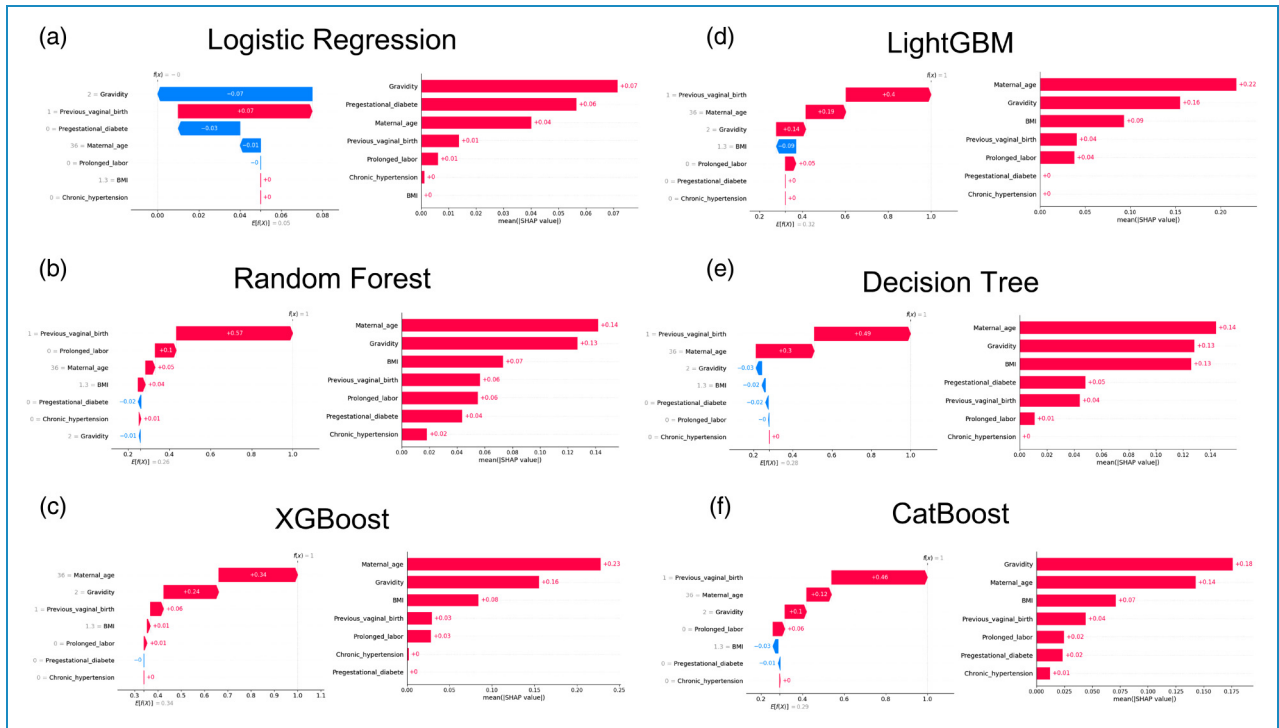
We developed an AI predictive model for VBAC, which was integrated into an innovative DA birth choice platform for SDM.<sup>10</sup> In contrast to conventional DA like booklets, video CDs, pamphlets, or computer-based tools commonly used to aid pregnant women in making birth choices after previous cesarean delivery,<sup>25–29</sup> this platform integrates an AI predictive model. Through this model, women can input seven crucial factors—maternal age, gravidity, BMI, gestational week, history of previous



**Figure 4.** Confusion matrix for every ML model, including logistic regression, random forest, XGBoost, LightGBM, decision tree, and CatBoost. 0: VBAC failure group. 1: VBAC success group.



**Figure 5.** ROC curve for the performance evaluation at every ML model, including logistic regression, random forest, XGBoost, LightGBM, decision tree, and CatBoost.



**Figure 6.** Feature importance based on SHAP for every ML model, including logistic regression, random forest, XGBoost, LightGBM, decision tree, and CatBoost.

vaginal birth, prolonged labor, gestational diabetes, and pregnancy-induced hypertension—to promptly ascertain their likelihood of achieving a successful VBAC. This groundbreaking advancement proves particularly advantageous in regions where hospital access is limited. While prior research predominantly focused on diverse variables concerning pregnant women and fetuses to validate models, often relying on datasets from later gestational stages (i.e., gestational week > 27 weeks), our model exhibited robust performance and strong validity across a broad spectrum of gestational weeks ranging from 12 to 40 weeks.

Addressing the common challenge of limited generalizability in AI-based healthcare applications, often stemming from inadequate training procedures and evaluation protocols,<sup>30</sup> we adopted a comprehensive approach. Our study distinguishes itself by achieving a comparable predictive performance with only seven parameters, a significant reduction compared to that in existing research.<sup>4,5</sup> Our strategy consisted of a two-fold approach: during the training phase, we employed boosting techniques and applied bootstrapping for rigorous testing.<sup>31,32</sup> Our results demonstrated that when analyzing tabular data, the tree-based models consistently outperformed the regression models. We assessed six different models and fine-tuned their hyperparameters through the boosting technique, leading us to identify CatBoost as the top-performing model in alignment with the findings of the study by Li et al.<sup>19</sup>

We incorporated six variables (i.e., maternal age, pregnancy weight, height, arrest disorder, previous vaginal birth, and chronic hypertension) drawn from prior research by Grobman et al.<sup>33</sup> To enhance our model's predictive capability, we introduced two critical factors: gravidity and gestational diabetic status. Nonetheless, the influence of obstetric disease, such as chronic hypertension and gestational diabetes, remains intricate. While the incidence of obstetric disease in VBAC failed cases appears higher compared to VBAC success cases,<sup>34</sup> another study suggests that within the obstetric disease subgroup, both incidences of the chronic hypertension and pregestational diabetes of cases favor VBAC success over VBAC failed.<sup>35</sup> Moreover, both studies fail to establish significant statistical differences in obstetric disease. Recent efforts have sought to consolidate critical parameters through review and synthesis of studies. Surprisingly, amidst these endeavors, the significance of hypertension within obstetric disease appears less pronounced and evident.<sup>4</sup> To ensure the effective utilization of these influencing factors, our study employed a machine learning methodology rather than conventional statistical approaches, such as linear regression. Our findings indeed demonstrate that this addition significantly improved the accuracy of our model. Unlike previous studies that mainly focused on data beyond 37 weeks of gestation,<sup>33</sup> our research not only adjusted for BMI, but also extended its applicability to as early as 12 weeks into term gestation. Our substantial improvement in

predictive performance was a result of transitioning from conventional regression to tree-based models. These tree-based models effectively managed both categorical and continuous variables, capturing intricate relationships. Consequently, our model exhibited a notable performance boost, achieving an AUC of 0.89 (95% CI: 0.86–0.93) and surpassing the AUC of 0.75 (95% CI: 0.74–0.77) reported in previous studies.<sup>33</sup> This advancement underscores the potential significance of our research in the realm of tabular data analysis for pregnancy-related outcomes.

The SHAP analysis provided valuable insights into the key factors driving prediction outcomes, making it a valuable tool for interpreting machine learning models in the medical field.<sup>22</sup> In this study, we demonstrated the efficacy of the SHAP method in interpreting the VBAC prediction models within the medical domain. By utilizing a game-theoretic approach, SHAP consistently and reliably estimated the importance of features, making it an ideal choice for analyzing medical data that combine both continuous and categorical variables. Previous research primarily relied on linear *p*-values to assess parameters in VBAC prediction.<sup>33</sup> However, our research uncovered that the VBAC prediction model trained with SHAP consistently upheld a coherent hierarchy of variable importance, closely mirroring the factors that influence clinical decision-making, similar to previous investigations.<sup>36,37</sup> In the context of the SHAP analysis for the CatBoost model, the significance of gravidity as the most influential factor can be illuminated from multiple perspectives. Gravidity may intricately correlate with physiological elements, such as uterine status, uterine musculature elasticity, and ligamentous tension, all of which collectively influence the likelihood of a successful vaginal birth.<sup>38</sup> Furthermore, our current study's SHAP analysis demonstrated that a history of successful vaginal births significantly impacts the VBAC success. Women with such a history may possess more favorable birth canals, thereby increasing the likelihood of a successful current vaginal delivery.<sup>39</sup>

### Limitation and recommendation

This study has notable limitations that warrant acknowledgment. First, its single-hospital focus restricts its generalizability. To bolster external validity, future research should encompass multiple hospitals, particularly those with diverse geographical locations and healthcare systems. This broader approach would render the research more applicable and representative of real-world scenarios, facilitating a deeper understanding of its impact on clinical practices and outcomes. Second, the current sample may not adequately capture the population's diversity, potentially leading to oversights in comprehending the dynamics of decision-making regarding birth choices across various racial backgrounds. Expanding the inclusion of diverse



racial and ethnic groups would contribute to a more comprehensive understanding of the phenomenon. Third, the timing of ERCD significantly influences birth outcomes, and it is vital to consider when women opt for cesarean sections. Integrating the timing of childbirth into the study parameters would offer a more accurate reflection of the decision-making processes during actual ERCD procedures. The timing of birth can substantially impact various aspects of maternal and neonatal outcomes, including complications and interventions. Incorporating data on childbirth timing would provide additional insights for a comprehensive analysis and inform evidence-based guidelines for optimizing obstetric care. Despite these limitations, implementing the study's findings into existing healthcare systems can enhance the effectiveness of encouraging pregnant women to attempt VBAC, providing valuable insights for clinical decision-making and ultimately improving the quality of care.

## Conclusion

The integration of AI prediction systems into DA birth choice platform for SDM holds great potential for clinical practice. Our study successfully designed and deployed such a platform in clinical settings, which has been well-received by pregnant women. It has demonstrated a remarkable accuracy when using local hospital data, with tree-based models, particularly CatBoost, outperforming traditional regression models. The SHAP analysis further validated the model's effectiveness in grasping data intricacies and emphasized the importance of key variables like gravidity and vaginal birth history. This AI-powered VBAC prediction system substantially enhances the accuracy of AI models and is adaptable for use in the early stages of pregnancy. Its primary objective is to improve decision-making processes related to birth choice and enhance the overall pregnancy experience. This holds particular significance in addressing the extremely low VBAC rates in Taiwan, aiming to empower pregnant women to make well-informed decisions concerning their mode of birth.

**Acknowledgements:** We would like to express our sincere gratitude to Saint Paul's Hospital for providing the valuable data necessary for this research.

**Contributorship:** CCY contributed to conceptualization, methodology, data curation, funding acquisition, resource, and formal analysis. SWC contributed to conceptualization, methodology, formal analysis, supervision, project administration, and writing of the original draft, review, and editing. HWH contributed to conceptualization, methodology, formal analysis, and writing of the original draft. CFW contributed to conceptualization, methodology, formal analysis, and writing of the original draft. WML contributed to conceptualization, data curation, and formal analysis. All authors reviewed and edited the manuscript and approved the final version of the manuscript.

**Declaration of conflicting interests:** The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**Funding:** The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Ministry of Science and Technology (grant number MOST 110-2314-B-227-006-MY2).

**Guarantor:** SWC.

**Research ethics and patient consent:** This study was conducted under the approval of the Ethics Committee of Fu Jen Catholic University (REC number: C111042), adhering strictly to the Helsinki Declaration guidelines and regulations prior to commencement. As this was retrospective medical record data, the requirement for obtaining written informed consent from all participants was waived by the Ethics Committee of Fu Jen Catholic University.

**ORCID ID:** Shu Wen Chen  <https://orcid.org/0000-0002-0044-5750>

## References

1. Joseph Nggada B. Vaginal birth after caesarean (VBAC) [Internet]. In: *New aspects in cesarean sections*. USA: IntechOpen, 2023, pp.1–19.
2. İzbudak G, Tozkır E, Cogendez E, et al. Comparison of maternal–neonatal results of vaginal birth after cesarean and elective repeat cesarean delivery. *Ginekol Pol* 2021; 92: 306–311.
3. Osterman MJ. Changes in primary and repeat cesarean delivery: United States, 2016–2021 Vital statistics rapid release; No 21. National Center for Health Statistics, 2022.
4. Deng B, Li Y, Chen J-Y, et al. Prediction models of vaginal birth after cesarean delivery: a systematic review. *Int J Nurs Stud* 2022; 135: 1–11.
5. Black N, Henderson I, Al Wattar BH, et al. Predictive models for estimating the probability of successful vaginal birth after cesarean delivery: a systematic review. *Obstet Gynecol* 2022; 140: 821–841.
6. Chen SW. Birth choice decision aids in women planning vaginal birth after cesarean delivery. *Formosan J Med* 2021; 25: 460–473.
7. Chen SW, Hutchinson AM, Nagle C, et al. Women's decision-making processes and the influences on their mode of birth following a previous caesarean. *BMC Pregnancy Childbirth* 2018; 18: 1–17.
8. Chen SW, Yang CC, Te JC, et al. Birth choices after caesarean in Taiwan: a mixed methods pilot study of a decision aid for shared decision making. *Midwifery* 2021; 95: 1–8.
9. Chen SW and Cheng PJ. Mode of birth following a primary caesarean section: Taiwanese obstetricians' decision-making strategies. *Taiwan Midwives J* 2015; 57: 55–69.
10. Chen SW, Shorten A, Yeh CC, et al. An innovative web-based decision-aid about birth after cesarean for shared decision making in Taiwan: study protocol for a randomized control trial. *Trials* 2023; 24: 1–12.

11. Qureshey E, Rochon M, Hesham H, et al. Patient compliance and satisfaction using web-based glucose monitoring for the management of pregnant women with pregestational diabetes. *J Matern Fetal Neonatal Med* 2022; 35: 5943–5948.
12. West BT, Axinn WG, Couper MP, et al. A web-based event history calendar approach for measuring contraceptive use behavior. *Field Methods* 2022; 34: 3–19.
13. Nguyen VH and Harley KG. Prenatal cannabis use and infant birth outcomes in the pregnancy risk assessment monitoring system. *J Pediatr* 2022; 240: 87–93.
14. Marzouk R, Alluhaidan AS and El\_Rahman SA. An analytical predictive models and secure web-based personalized diabetes monitoring system. *IEEE Access* 2022; 10: 105657–73.
15. Shwartz-Ziv R and Armon A. Tabular data: deep learning is not all you need. *Inf Fusion* 2022; 81: 84–90.
16. Seedat N, Crabbé J, Bica I, et al. Data-IQ: characterizing subgroups with heterogeneous outcomes in tabular data. *Adv Neural Inf Process Syst* 2022; 35: 23660–23674.
17. Ahmad GN, Fatima H, Ullah S, et al. Efficient medical diagnosis of human heart diseases using machine learning techniques with and without GridSearchCV. *IEEE Access* 2022; 10: 80151–80173.
18. Kadambi A, Wen T, Nguyen E, et al. Random forests for accurate prediction of the risk of hypertensive disorders of pregnancy at term [A208]. *Obstet Gynecol* 2022; 139: 60S–61S.
19. Li Y, Duan H and Wang S. An XGBoost predictive model of ongoing pregnancy in patients following hysteroscopic adhesiolysis. *Reprod Biomed Online* 2023; 46: 965–972.
20. Prokhorenkova L, Gusev G, Vorobev A, et al. CatBoost: unbiased boosting with categorical features. *Adv Neural Inf Process Syst* 2018; 31: 1–11.
21. Xue Y, Chen S, Zhang M, et al. The prediction models for high-risk population of stroke based on logistic regressive analysis and LightGBM algorithm separately. *Iran J Public Health* 2022; 51: 999–1009.
22. Baptista ML, Goebel K and Henriques EM. Relation between prognostics predictor evaluation metrics and local interpretability SHAP values. *Artif Intell* 2022; 306: 1–22.
23. Hu C, Li L, Huang W, et al. Interpretable machine learning for early prediction of prognosis in sepsis: a discovery and validation study. *Infect Dis Ther* 2022; 11: 1117–1132.
24. Buckley A, Sestito S, Ogundipe T, et al. Racial and ethnic disparities among women undergoing a trial of labor after cesarean delivery: performance of the VBAC calculator with and without patients' race/ethnicity. *Reprod Sci* 2022; 29: 2030–2038.
25. Montgomery AA, Emmett C, Fahey T, et al. Two decision aids for mode of delivery among women with previous cesarean section: randomised controlled trial. *Br Med J* 2007; 334: 1305.
26. Torigoe I and Shorten A. Using a pregnancy decision support program for women choosing birth after a previous cesarean in Japan: a mixed methods study. *Women Birth* 2018; 31: e9–e19.
27. Vankan E, Schoorel E, van Kuijk S, et al. The effect of the use of a decision aid with individual risk estimation on the mode of delivery after a caesarean section: a prospective cohort study. *PLoS One* 2019; 14: 1–15.
28. Wise MR, Sadler L, Shorten B, et al. Birth choices for women in a 'positive birth after caesarean' clinic: randomised trial of alternative shared decision support strategies. *Aust N Z J Obstet Gynaecol* 2019; 59: 684–692.
29. Kuppermann M, Kaimal AJ, Blat C, et al. Effect of a patient-centered decision support tool on rates of trial of labor after previous cesarean delivery: the PROCEED randomized clinical trial. *Jama* 2020; 323: 2151–2159.
30. Kelly CJ, Karthikesalingam A, Suleyman M, et al. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med* 2019; 17: 1–9.
31. Plaia A, Buscemi S, Fürnkranz J, et al. Comparing boosting and bagging for decision trees of rankings. *J Classif* 2022; 39: 78–99.
32. Egbert J and Plonsky L. Bootstrapping techniques. In: *A practical handbook of corpus linguistics*. Germany: Springer, 2021, pp.593–610.
33. Grobman WA, Sandoval G, Rice MM, et al. Prediction of vaginal birth after cesarean delivery in term gestations: a calculator without race and ethnicity. *Am J Obstet Gynecol* 2021; 225: 664.e1–664.e7.
34. Chen YT, Hsieh Y-C, Shen H, et al. Vaginal birth after cesarean section: experience from a regional hospital. *Taiwan J Obstet Gynecol* 2022; 61: 422–426.
35. Tsai H-T and Wu C-H. Vaginal birth after cesarean section—the world trend and local experience in Taiwan. *Taiwan J Obstet Gynecol* 2017; 56: 41–45.
36. Liao Q, Luo J, Zheng L, et al. Establishment of an antepartum predictive scoring model to identify candidates for vaginal birth after cesarean. *BMC Pregnancy Childbirth* 2020; 20: 1–7.
37. Rodríguez-Pérez R and Bajorath J. Interpretation of machine learning models using SHapley values: application to compound potency and multi-target activity predictions. *J Comput-Aided Mol Des* 2020; 34: 1013–1026.
38. Fidalgo D, Pouca MV, Oliveira D, et al. Mechanical effects of a Maylard scar during a vaginal birth after a previous cesarean. *Ann Biomed Eng* 2021; 49: 3593–3608.
39. Mekonnen BD and Asfaw AA. Predictors of successful vaginal birth after a cesarean section in Ethiopia: a systematic review and meta-analysis. *BMC Pregnancy Childbirth* 2023; 23: 1–12.