


A multimodal model in the prediction of the delivery mode using data from a digital twin-empowered labor monitoring system

DIGITAL HEALTH
Volume 10: 1–10
© The Author(s) 2024
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20552076241304934
journals.sagepub.com/home/dhj



Jieyun Bai^{1,2,*} , Xue Kang^{1,*}, Weishan Wang^{1,*}, Ziduo Yang¹,
Weiguang Ou³, Yuxin Huang^{4,*} and Yaosheng Lu¹

Abstract

Objective: This study aims to address the limitations of current clinical methods in predicting delivery mode by constructing a multimodal neural network-based model. The model utilizes data from a digital twin-empowered labor monitoring system, including computerized cardiotocography (cCTG), ultrasound (US) examination data, and electronic health records (EHRs) of pregnant women.

Methods: The model integrates three modalities of data from 105 pregnant women (76 vaginal deliveries and 29 cesarean deliveries) at the Department of Obstetrics and Gynecology of The First Affiliated Hospital of Jinan University, Guangzhou, China. It employs a hybrid architecture of a convolutional neural network (CNN) and bi-directional long short-term memory (BiLSTM) to compress the data into a single feature vector for each patient.

Results: The designed model achieves a cross-validation accuracy of 93.33%, an F1-score of 86.26%, an area under the receiver operating characteristic curve of 97.10%, and a Brier Score of 6.67%. Importantly, while cCTG and EHRs are crucial for labor management, the integration of US imaging data significantly enhances prediction accuracy.

Conclusion: The findings of this study suggest that the developed multimodal model is a promising tool for predicting delivery mode and provides a comprehensive approach to intrapartum maternal and fetal health monitoring. The integration of multi-source data, including real-time information, holds potential for further improving the algorithm's predictive accuracy as the volume of analyzed data increases. This could be highly beneficial for dynamically fusing data from different sources throughout the maternal and fetal health lifecycle, from pregnancy to delivery.

Keywords

Digital twin, multimodal model, cesarean section, delivery mode, cardiotocography, intrapartum ultrasound, electronic health record

Submission date: 27 June 2024; Acceptance date: 19 November 2024

¹Guangdong Provincial Key Laboratory of Traditional Chinese Medicine Information Technology, Jinan University, Guangzhou, China

²Auckland Bioengineering Institute, University of Auckland, Auckland, New Zealand

³The First Affiliated Hospital of Jinan University, Jinan University, Guangzhou, China

⁴Department of Obstetrics and Gynecology, Zhujiang Hospital, Southern Medical University, Guangzhou, China

*Contributed equally.

House, Level, 6/70 Symonds Street, Grafton, Auckland 1010, New Zealand.
Email: jbai996@aucklanduni.ac.nz

Yuxin Huang, Department of Obstetrics and Gynecology, Zhujiang Hospital, Southern Medical University, Guangzhou 510280, China.
Email: 15521280986@163.com

Yaosheng Lu, Guangdong Provincial Key Laboratory of Traditional Chinese Medicine Information Technology, Jinan University, Guangzhou 510632, China.
Email: tluy@s@jnu.edu.cn

Corresponding authors:

Jieyun Bai, Auckland Bioengineering Institute, University of Auckland, UniServices



Introduction

Globally, an alarming number of maternal fatalities (approximately 287,000), neonatal deaths (2.4 million), and stillbirths (1.9 million) are reported annually, with the bulk occurring in low- and middle-income countries.¹ A significant proportion (up to 45%) of these adverse outcomes transpire during labor, birth, and the immediate post-natal period.² Although cesarean deliveries hold life-saving potential, their increasing rate poses a substantial global health challenge. According to the Guangzhou Women and Children's Medical Center birth cohort study in China, the cesarean section rate (i.e. 32%) in Guangzhou is still much higher than the ideal cesarean section rate (10–15%) recommended by the World Health Organization.³ By 2030, it is projected that around 28.5% of all global births will involve cesarean sections, which equates to approximately 38 million women annually.⁴ Optimizing the use of cesarean section is a matter of global concern as underutilization can lead to heightened maternal and perinatal mortality and morbidity, while overutilization is linked to non-reduced mortality rates and can cause complications in subsequent pregnancies.^{5,6}

The process of childbirth is influenced by the “three Ps” model: the passageway (maternal bony pelvis and soft tissues), the passenger (fetus), and the power (uterine contractions and voluntary maternal efforts during delivery).⁷ Effective decision-making necessitates the integration of a vast array of data collected during antenatal visits, upon admission to the delivery unit, and throughout labor.⁸ This emphasizes the need for a dynamic, individualized assessment approach, which is at the core of precision medicine—tailoring medical treatments to the unique characteristics of each patient.^{9,10}

Technological progressions, such as big data analytics, cloud computing, virtual reality, and the Internet of Things and Artificial Intelligence, play a crucial role in advancing this personalized approach to medicine.^{11,12} These technologies facilitate the development of digital twins (DTs), sophisticated simulations that amalgamate real-time data from diverse sources like electronic health records (EHRs), medical devices, and genomic information.¹³ DT technology is transforming healthcare systems by leveraging real-time data integration, advanced analytics, and virtual simulations to enhance patient care, enable predictive analytics, optimize clinical operations, and facilitate training and simulations.^{14,15} With their ability to analyze extensive patient data from various sources, DTs can offer personalized treatment plans and facilitate predictive analytics and preventive interventions through machine learning algorithms.^{16,17} The implementation of DT technology in healthcare holds the potential to significantly improve patient outcomes, enhance patient safety, and drive innovation in the healthcare industry. While DTs have been applied in various healthcare

domains such as asthma,¹⁸ diabetes,¹⁹ cancer,²⁰ and cardiovascular disease,²¹ their potential in labor care remains largely unexplored.²²

In this brief communication, we described the digital twin-empowered labor monitoring system (DTLMS), developed by Guangzhou Lianyin Medical Co., Ltd (Lian-Med, China). This system integrates IoT devices, virtual reality, and artificial intelligence to provide personalized labor care support (Figure 1). The DTLMS can display patient characteristics in EHRs, real-time electronic fetal monitoring signal, and three-dimensional (3D) virtual labor progress based on intrapartum ultrasound imaging.^{23–28} Here, we utilize these heterogeneous data obtained by the DTLMS to develop and assess an automated assessment tool capable of predicting the delivery mode.

Methods

Proposed digital twin framework

The DTLMS is developed by Guangzhou Lianyin Medical Co., Ltd (Lian-Med, China) for comprehensive maternal and fetal safety monitoring, and also offers decision support to optimize the use of cesarean section. Consequently, it does not specifically address the challenges of women starting antenatal care (ANC) late or not attending the recommended number of visits. The proposed framework consists of four parts: the physical space, the DT data module, the digital space, and the DT intelligent module. The DT data module encompasses dynamic multi-source recordings of maternal and fetal health, including EHR, ultrasound (US) images, and physiological signals (Supplementary Figure S1). The digital space represents a 3D visualization of maternal and fetal structures during labor, based on intrapartum ultrasound examinations (Supplementary Video S1). The DT intelligent module, central to this framework, is a predictive model designed to identify risk factors and predict the likelihood of adverse events or specific delivery modes, thus enhancing clinical decision-making capabilities. Note: this device is designed to handle a variety of data sources and formats. In China, a mix of electronic systems and paper registers is observed, but this system has the capability to automatically recognize and extract information from unstructured paper-based medical reports.²⁹ This functionality ensures that even in areas with limited technological infrastructure, valuable medical data can be utilized for accurate monitoring and prediction.

Multimodal dataset collection

This retrospective study analyzed data from January 2012 to December 2020 from the Department of Obstetrics and Gynecology at The First Affiliated Hospital of Jinan

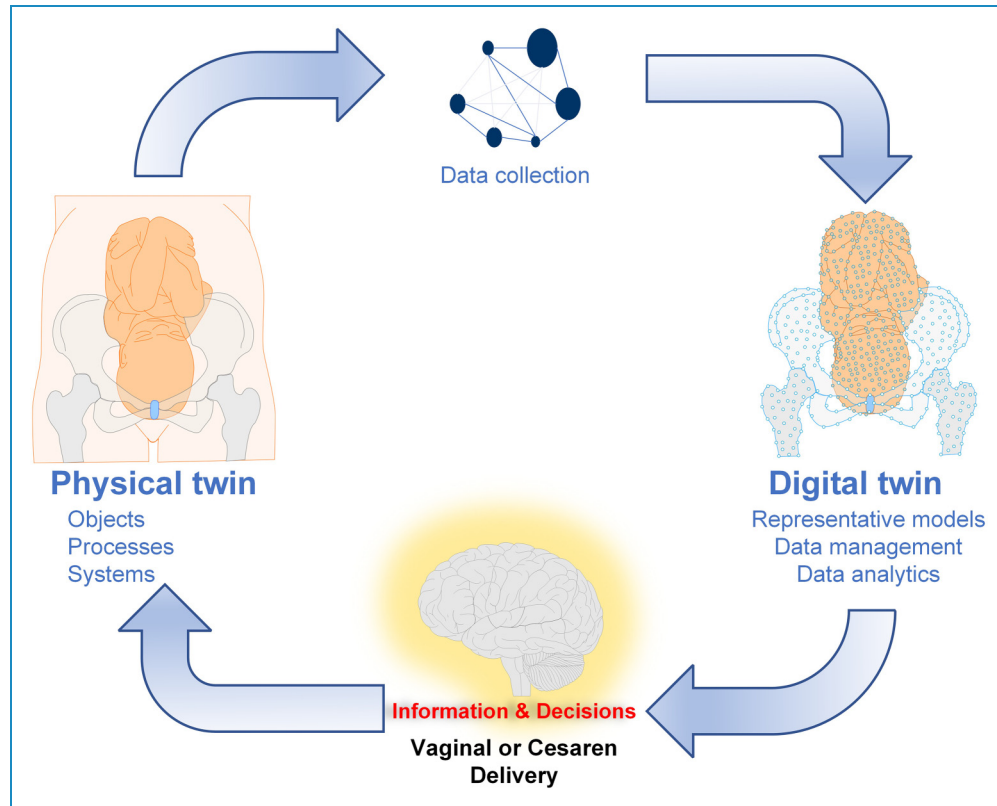


Figure 1. A digital twin framework for maternal and fetal health. This work focuses on its application in assisting decision-making.

University, Guangzhou, China. Exclusions were multifetal gestations, planned cesarean deliveries, antepartum fetal death, major fetal anomalies, and preterm deliveries. Additionally, cases lacking comprehensive electronic data on the labor process were excluded. The dataset included EHR, US examination data, and raw 30-minutes computerized cardiotocography (cCTG) records (Supplementary Figure 2).³⁰ Post-delivery data such as infant gender and actual birth weight were not utilized in the prediction analysis. Institutional ethical review board waived the need for the consent (JNUKY-2022-018).

Multimodal deep learning model

We employed a hybrid model to predict the delivery mode, integrating discrete data (EHR and US) with continuous signal data (cCTG, which provides a continuous graphic record of fetal heart rate (FHR) and uterine contractions (UCs)). The hybrid model combines a convolutional neural network (CNN) and bi-directional long short-term memory (BiLSTM) to capture complex nonlinear spatial and temporal features of the cCTG, while features from EHR and US were digitized and normalized. These data types were then fused to perform classification. The model's architecture is detailed in Supplementary Figure 3.

Model training utilized the TensorFlow library, taking advantage of the parallel-processing capabilities of Graphical Processing Units. Stochastic gradient descent with momentum (momentum=0.9) was the chosen optimization algorithm. Cross-entropy loss function was used for loss computation. The model underwent 120 training epochs, with a learning rate initially set at 0.01. To optimize learning dynamics, we employed a decay factor of 10 after 15 and 90 epochs, maintaining a stable learning pace throughout training. To mitigate overfitting, L2 regularization with a factor of 0.0001 was applied.

Evaluation metrics

Model performance was assessed through a 5-fold cross-validation process, evaluating each model on accuracy, sensitivity, specificity, area under the receiver operating curve (AUC), F1 Score, and Brier Score. The 95% confidence intervals for these metrics were calculated using a z -value of 1.95996, based on the Gaussian distribution. Here are brief definitions of each metric: *accuracy*: The ratio of correctly predicted instances to the total number of instances. It measures the overall performance of a model but can be misleading when classes are imbalanced. *Sensitivity*: The proportion of actual positives that are correctly identified by the model. It reflects the model's ability to identify

positive cases. *Specificity*: The proportion of actual negatives that are correctly identified by the model. It measures how well the model can avoid false positives. *Area under the receiver operating characteristic curve*: The area under the ROC curve, which plots the true positive rate (sensitivity) against the false positive rate (1 - Specificity) at different threshold levels. AUC-ROC provides a measure of how well a model distinguishes between classes. An AUC of 1 indicates a perfect classifier, while 0.5 represents random guessing. *F1 Score*: The harmonic mean of precision and recall (sensitivity), balancing the trade-off between false positives and false negatives. It's useful when class distribution is uneven or when false positives and false negatives have different costs. *Brier Score*: Measures the mean squared difference between predicted probabilities and the actual outcomes (0 or 1). It evaluates the accuracy of probabilistic predictions, where a lower score indicates better model performance. These metrics are commonly used to evaluate the performance of machine learning models, especially in classification tasks.

Statistical analysis

Statistical analysis was performed with Python 3.7.12, scikit-learn library 1.0.2, scipy library 1.4.1. Dichotomous features were compared using Welch's *t*-test and Chi-squared test in cases of small numbers, as appropriate.

Results

Table 1 presents the clinical profiles of 105 patients undergoing vaginal and cesarean deliveries as recorded in the EHRs. Demographically and physically, no significant differences were observed between the vaginal ($N=76$) and cesarean ($N=29$) delivery groups concerning age, height, gestational age, pregnancy number, and parity. However, the body mass index (BMI) was significantly higher in the cesarean delivery group ($26.74 \pm 3.19 \text{ kg/m}^2$) compared to the vaginal delivery group ($24.82 \pm 2.27 \text{ kg/m}^2$), with a *p*-value of 0.0057. Maternal complications such as anemia, gestational diabetes, liver disease, thyroid disorders, and pregnancy-induced hypertension showed no significant differences between the groups. Yet, the incidences of premature rupture of membranes and amniotic fluid contamination were notably higher in the cesarean delivery group (*p*-values of 0.0435 and 0.0075, respectively). For neonatal outcomes, a higher proportion of male infants was noted in the cesarean group ($p=0.0309$). Birth weights were comparable across both groups ($p=0.3480$), as were the Apgar scores at 1, 5, and 10 minutes post-delivery ($p>0.05$). A significant distinction was found in the umbilical cord pH levels, which were higher in the cesarean delivery group (7.28 ± 0.05) compared to the vaginal delivery group (7.20 ± 0.1), with a *p*-value of less than 0.0001.

In conjunction with these EHRs, US examination data and 30-minute cCTG were also collected using the DTLMS. US examination revealed no significant differences between the groups in measurements of biparietal diameter, abdominal circumference, femur length, and head circumference. However, fetal positioning displayed a highly significant difference; the occiput anterior position was prevalent in 97.37% of vaginal deliveries, compared to just 3.45% in cesarean deliveries. Conversely, the occiput transverse and occiput posterior positions were predominantly observed in the cesarean delivery group (75.86% and 20.69%, respectively) (Supplementary Table S1). Changes in FHR and their temporal relationship to UCs in cCTG records demonstrated marked fluctuations in some cases of cesarean delivery compared to vaginal delivery (Supplementary Figure S2). However, when examining specific cCTG features such as baseline, accelerations, decelerations, variability, and the nature of uterine contractions, no significant differences were observed between the groups (Supplementary Table S2). Using this limited dataset with cCTGs, EHRs and US examination data, a deep learning model (Supplementary Figure S3) was trained with five-fold cross-validation to assess the predictive power of different data combinations on delivery mode. Employing solely cCTGs, the model yielded an AUC of 0.4642 ± 0.0320 (95% confidence interval: 0.2509–0.6775). By integrating EHRs and cCTG signals, the classification performance improved, achieving an AUC of 0.6447 ± 0.1014 (95% confidence interval: 0.4400–0.8494). The combination of all data sources (i.e. cCTG, EHRs, and US) enabled the model to achieve its best performance, with an AUC of 0.9710 ± 0.0211 (95% confidence interval: 0.8992–1.0428) (Supplementary Figure S4). This trend was similarly reflected in other evaluation metrics, such as accuracy (ACC), sensitivity (SEN), specificity (SPE), and F1 and Brier scores (Table 2). Training and validation accuracy (Acc) curve and precision-recall curves of the deep learning model are shown in Supplementary Figures S5 and S6.

Discussion

These results demonstrate the feasibility of DTLMS to predict the mode of delivery, showing an increasing level of prediction accuracy as the volume of data analyzed by the model increases. Prior studies have explored various features in recordings from patients undergoing vaginal and cesarean deliveries.^{31–54} While these studies have significantly enhanced the characterization of delivery modes, they often relied on specific features extracted at different points in antepartum and intrapartum care. These studies primarily concentrated on specific features extracted from data collected at various stages, either during antepartum^{40,50} or intrapartum care.^{31–39,41–54}

Table 1. Clinical profile of patients in electrical health recordings (EHR).

Features	Vaginal delivery (N= 76)	Cesarean delivery (N= 29)	p-Value
Age (year)	29.37 ± 3.95	29.31 ± 4.42	0.9516
Height (cm)	158.21 ± 4.62	159.17 ± 5.21	0.3946
BMI (kg/m ²)	24.82 ± 2.27	26.74 ± 3.19	0.0057
Gestational age (weeks)	39.00 ± 0.95	39.17 ± 0.91	0.4021
Pregnancy			
1	38 (50.00%)	17 (58.62%)	0.6616
2	25 (32.89%)	7 (24.14%)	
≥3	13 (17.11%)	5 (17.24%)	
Parity			
1	49 (64.47%)	23 (79.31%)	0.2190
2	25 (32.89%)	7 (24.14%)	
Complications			
Anemia	7 (9.21%)	3 (10.34%)	0.8456
Gestational diabetes	7 (9.21%)	5 (17.24%)	0.4160
Liver disease	7 (9.21%)	0 (0%)	0.2097
Pregnancy-induced hypertension	0 (0%)	0 (0%)	ns
Thyroid disorders	6 (7.89%)	3 (10.34%)	0.9911
Preeclampsia	0 (0%)	2 (6.90%)	0.1302
Oligohydramnios	0 (0%)	1 (3.45%)	0.6150
Premature rupture of membranes	11 (14.47%)	10 (34.48%)	0.0435
Amniotic fluid contamination	1 (1.32%)	5 (17.24%)	0.0075
Fetal distress	0 (0%)	1 (3.45%)	0.6150
Fetal growth restriction	0 (0%)	0 (0%)	ns
Infant gender (male)	41 (53.95%)	23 (46.05%)	0.0309
Birth weight (g)	3169.47 ± 343.62	3224.48 ± 362.35	0.3480
Low birth weight	1 (1.32%)	0 (0%)	0.6149
Macrosomia	1 (1.32%)	1 (3.45%)	0.9333

(continued)

Table 1. Continued.

Features	Vaginal delivery (N= 76)	Cesarean delivery (N= 29)	p-Value
Apgar score at 1 minute	8.7237 ± 0.6996	8.7586 ± 0.5666	0.7954
Apgar score at 5 minutes	9.8816 ± 0.3615	9.8621 ± 0.4334	0.8329
Apgar score at 10 minutes	9.9079 ± 0.3316	9.9655 ± 0.1825	0.2664
Umbilical cord pH	7.20 ± 0.1	7.28 ± 0.05	<0.0001

Note. ns, not significant.

Table 2. The performance of the deep learning model using different data combinations.

Conditions	ACC (%)	SEN (%)	SPE (%)	F1 (%)	AUC	Brier
cCTG	56.19 ± 6.32	32.00 ± 21.77	66.08 ± 12.92	25.97 ± 15.01	0.4642 ± 0.0320	0.4381 ± 0.0632
cCTG + EHR	61.90 ± 11.66	59.33 ± 26.53	63.50 ± 17.89	45.09 ± 13.14	0.6447 ± 0.1014	0.3810 ± 0.1166
cCTG + EHR + US	93.33 ± 4.86	85.33 ± 18.09	96.08 ± 3.20	86.26 ± 11.70	0.9710 ± 0.0211	0.0667 ± 0.0486
EHR	60.95 ± 1.33	62.00 ± 22.07	60.83 ± 16.14	46.91 ± 14.72	0.6375 ± 0.0721	0.3905 ± 0.1327
US	83.81 ± 4.86	43.33 ± 24.94	98.67 ± 2.67	53.88 ± 24.94	0.9883 ± 0.0145	0.1619 ± 0.0486
cCTG + US	87.62 ± 4.86	57.33 ± 23.89	98.67 ± 2.67	68.18 ± 19.82	0.9889 ± 0.0172	0.1238 ± 0.0486

Note. CTG, cardiotocography; EHR, electronic health records; US, ultrasound examination results; Acc, Accuracy; SEN, Sensitivity; SPE, Specificity; F1, F1 Score; Brier, Brier Score.

It is important to note that certain features, like those from cCTG—including baseline, accelerations, and decelerations^{31,55–58}—are subjective and heavily reliant on the interpretation by caregivers. In this study, we followed an alternative approach, to predict the delivery mode using a broad spectrum of raw data gathered during the delivery process, rather than characterizing it. Notably, while intrapartum cCTG and US imaging remain crucial for labor management, our findings underscore the enhanced prediction accuracy provided by integrating supplementary information from EHRs.^{53,59} In recent years, fetal monitoring has advanced with the integration of novel technologies. Wearable devices with multi-modal sensors offer a promising solution for continuous and non-invasive monitoring. For example, Ghosh et al.^{60,61} and Mesbah et al.⁶² have proposed approaches that leverage sensor fusion techniques and machine learning algorithms to improve fetal movement detection and automate the process. These advancements highlight the growing trend towards more sophisticated sensor-based fetal monitoring systems.

Our study serves as a proof-of-concept, demonstrating the utility of computer-assisted analysis in predicting delivery modes. To improve prediction accuracy further, it is essential to adhere to comprehensive data collection

guidelines such as those recommended by the WHO Labour Care Guide (2020).⁶³ Furthermore, optimizing data gathering and labeling are critical to create databases that allow the model to learn and analyze effectively. For the model's application to be considered reliable across various clinical settings, conducting independent validation studies is necessary. Additionally, randomized prospective trials should be undertaken to assess the model's efficacy and its impact on clinical decision-making processes. Looking ahead, future research should aim to develop models that utilize real-time data, enhancing the continuous improvement of the model by integrating various levels of data recorded throughout labor.⁵⁹ This will not only improve the model's applicability but also its reliability in clinical practice.

Once fully developed, the DTLMS holds enormous potential for routine clinical use, observational studies, and interventional trials. Its capability to dynamically integrate data from multiple sources—including cCTG, US examination data,⁶⁴ and EHRs—throughout the various stages of maternal and fetal health from pregnancy to delivery, is particularly valuable. This study underscores the DTLMS's potential, supported by artificial intelligence,^{65–68} to reassure patients about the feasibility of different

delivery modes and enable early preparation for necessary interventions. As the system collects data in real-time during the delivery process, its algorithm's ability to accurately predict the likelihood of vaginal delivery improves significantly. This functionality not only enhances the decision-making process regarding delivery modes but also contributes to the overall aim of reducing unnecessary cesarean deliveries, thereby improving maternal and fetal health outcomes. It will be particularly beneficial for dynamically fusing data from multiple sources across different stages of the maternal and fetal health lifecycle, from pregnancy through to delivery. This study highlights the potential of the DTLMS and Artificial Intelligence to provide reassurance to patients about the feasibility of delivery modes and to facilitate early preparations for interventions.^{69–73} As data is continuously acquired in real time during the delivery process, the algorithm's ability to predict the likelihood of vaginal delivery significantly increases.

However, there are limitations to this study. One limitation is the uncertainty regarding the device's ability to reduce the number of cesarean sections. Although the system can collect information on the entire pregnancy-to-delivery lifecycle and the number of cesarean sections in the study is known, it is not clear if the device can actually make a difference. The system's role in reducing cesarean sections or simply predicting delivery mode to improve efficiency remains undetermined. The device's role in addressing specific causes of cesarean sections in different contexts, especially in low- and middle-income countries and sub-Saharan Africa, is yet to be determined. Further research is needed to establish the relationship between the device and existing evidence on reducing cesarean sections. Another limitation is the relatively small size of the dataset compared to some other studies. This may limit the generalizability and robustness of the deep learning model. A smaller dataset might lead to overfitting or reduced accuracy when applied to larger and more diverse populations. It may also not fully capture the complexity and variability of real-world scenarios. Future research should involve collecting more data from multiple centers to increase the size and diversity of the dataset and improve the model's generalizability and performance.

It is worth noting that while our current study focused on predicting the delivery mode, the aspect of birth weight adequacy is an important consideration in the overall assessment of pregnancy outcomes. Birth weight adequacy can have significant implications for the health of the newborn and may potentially be related to the mode of delivery as well. For example, macrosomia (a condition where the fetus is significantly larger than average) may increase the likelihood of a cesarean delivery due to potential difficulties during vaginal birth. Although we do not have specific results regarding the prediction of birth weight adequacy in this study, it presents an interesting avenue for future

research. Incorporating factors related to birth weight into our model could potentially enhance its predictive ability not only for delivery mode but also for a more comprehensive understanding of the pregnancy process and outcomes. Future studies could explore the collection of more detailed data related to fetal growth and development throughout pregnancy, such as serial ultrasound measurements of fetal size and growth velocity, in combination with other clinical and demographic factors. This could help in building a more integrated model that can better predict both delivery mode and birth weight adequacy, providing clinicians with a more powerful tool for prenatal assessment and decision-making. Furthermore, understanding the relationship between birth weight adequacy and the digital twin-empowered labor monitoring system could offer new insights. The real-time data collected by the system, including cCTG, US examination data, and EHRs, may contain valuable information that could be utilized to develop predictive models for birth weight adequacy. For instance, changes in fetal growth patterns detected by serial ultrasounds or alterations in cCTG patterns that might be associated with fetal growth could be incorporated into future models.

Conclusion

In conclusion, this study presents a multimodal model for predicting delivery mode using data from a digital twin-empowered labor monitoring system. The model integrates cCTG, US examination data, and EHRs to offer a comprehensive approach to intrapartum maternal and fetal health monitoring. While this model shows great promise, further research is needed to validate its efficacy in different clinical settings and explore its impact on clinical decision-making. As the healthcare industry continues to advance, technologies like DTs and artificial intelligence hold immense potential for improving patient outcomes. This multimodal model represents a significant step in the right direction, offering a tool for personalized and predictive care during labor and delivery. Future research should focus on refining the model, expanding its application to other areas of maternal and fetal health, and investigating its potential in reducing unnecessary cesarean deliveries and enhancing overall health outcomes.

Acknowledgements: The authors thank the technical staff of Guangzhou Lianyin Medical Co., Ltd (Lian-Med, China) and researchers from The First Affiliated Hospital of Jinan University for obtaining data.

Code availability: The code is available at <https://zenodo.org/records/13927255>.

Contributorship: Jieyun Bai was involved in conception, organization and execution of research project and writing of manuscript. Weishan Wang and Xue Kang were involved in

visualization, software, methodology, data curation and review and editing. Ziduo Yang and Yuxin Huang were involved in clinical assistance, data collection, data analysis and review and editing. Yaosheng Lu and Weiguang Ou were involved in organization and execution of research project and review of manuscript.

Data availability: The data cannot be publicly available due to privacy and ethical restrictions. However, we provided the US dataset used for the PSFHS challenge of MICCAI 2023 (<https://ps-fh-aop-2023.grand-challenge.org/>). This dataset comprises two parts: one is the PSFHS dataset (<https://doi.org/10.5281/zenodo.10969427>) and the other is from the JNU-IFM dataset (<https://doi.org/10.6084/m9.figshare.14371652>). The images from the PSFHS dataset also was utilized for the Intrapartum Ultrasound Grand Challenge (IUGC) 2024 of MICCAI 2024 (<https://codalab.lisn.upsaclay.fr/competitions/18413>).

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: The work is funded by Guangzhou Municipal Science and Technology Bureau Guangzhou Key Research and Development Program (2024B03J1289 and 2024B03J1283), Key Research and Development Program of Guangxi Province under Project No. 2023AB22074, Natural Science Foundation of Guangzhou under Project No. 202201010544, Natural Science Foundation of Guangdong Province under Project No. 2024A1515011886 and 2023A1515012833, Guangzhou Science and Technology Planning Project under Grant 2023B03J1297, National Key Research and Development Program of China under Project No. 2019YFC0120100 and 2019YFC0121907, National Natural Science Foundation of China under Project No. 61901192, and China Scholarship Council under Project No. 202206785002.

Guarantor: Jieyun Bai.

ORCID iD: Jieyun Bai  <https://orcid.org/0000-0002-2847-350X>

Supplemental material: Supplemental material for this article is available online.

References

- Hug L, You D, Blencowe H, et al. Global, regional, and national estimates and trends in stillbirths from 2000 to 2019: a systematic assessment. *Lancet* 2021; 398: 772–785.
- Ahmed I, Ali SM, Amenga-Etego S, et al. Population-based rates, timing, and causes of maternal deaths, stillbirths, and neonatal deaths in south Asia and sub-Saharan Africa: a multi-country prospective cohort study. *Lancet Glob Health* 2018; 6: e1297–e1308.
- Xia X, Zhou Z, Shen S, et al. Effect of a two-stage intervention package on the cesarean section rate in Guangzhou, China: a before-and-after study. *PLoS Med* 2019; 16: e1002846.
- Betran AP, Ye J, Moller A-B, et al. Trends and projections of caesarean section rates: global and regional estimates. *BMJ Glob Health* 2021; 6: e005671.
- Hehir MP, Ananth CV, Siddiq Z, et al. Cesarean delivery in the United States 2005 through 2014: a population-based analysis using the Robson 10-Group classification system. *Am J Obstet Gynecol* 2018; 219: 105.e1–105.e11.
- Betrán AP, Temmerman M, Kingdon C, et al. Interventions to reduce unnecessary caesarean sections in healthy women and babies. *Lancet* 2018; 392: 1358–1368.
- Li X, Kruger JA, Nash MP, et al. Modeling childbirth: elucidating the mechanisms of labor. *Wiley Interdiscip Rev Syst Biol Med* 2010; 2: 460–470.
- Clark SL., Garite TJ, Hamilton EF, et al. “Doing something” about the cesarean delivery rate. *Am J Obstet Gynecol* 2018; 219: 267–271.
- Liu LY, Miller ES and Yee LM. Association between time of day and performance, indications, and outcomes of obstetric interventions among nulliparous women delivering at term. *J Perinatol* 2019; 39: 808–813.
- Stevenson DK., Wong RJ, Aghaepour N, et al. Towards personalized medicine in maternal and child health: integrating biologic and social determinants. *Pediatr Res* 2021; 89: 252–258.
- Hodson R. Precision medicine. *Nature* 2016; 537: S49–S49.
- Bai J, Lu Y, Liu H, et al. *New technologies improve maternal and newborn safety*. 2024. Lausanne, Switzerland: Frontiers Media SA, p. 1372358.
- National Academies of Sciences, Engineering and Medicine. *Foundational research gaps and future directions for digital twins*. Washington, DC, 2023.
- Katsoulakis E, Wang Q, Wu H, et al. Digital twins for health: a scoping review. *NPJ Digital Med* 2024; 7: 77.
- Zhang S, Lu W, Yang F, et al. Computational analysis of arrhythmogenesis in KCNH2 T618I mutation-associated short QT syndrome and the pharmacological effects of quinidine and sotalol. *NPJ Syst Biol Appl* 2022; 8: 43.
- Hu W, Zhang W, Zhang K, et al. Evaluating pro-arrhythmic effects of the T634S-hERG mutation: insights from a simulation study. *Interface Focus* 2023; 13: 20230035.
- Kang X, Zeng R, Yi H, et al. Prediction of delivery mode from fetal heart rate and electronic medical records using machine learning. In: 2022 computing in cardiology (CinC), 2022. IEEE.
- Gonsard A, AbouTaam R, Prévost B, et al. Children’s views on artificial intelligence and digital twins for the daily management of their asthma: a mixed-method study. *Eur J Pediatr* 2023; 182: 877–888.
- Cappon G, Vettoretti M, Sparacino G, et al. ReplayBG: a digital twin-based methodology to identify a personalized model from type 1 diabetes data and simulate glucose concentrations to assess alternative therapies. *IEEE Trans Biomed Eng* 2023; 70: 3227–3238.
- Stahlberg EA, Abdel-Rahman M, Aguilar B, et al. Exploring approaches for predictive cancer patient digital twins: opportunities for collaboration and innovation. *Front Digit Health* 2022; 4: 1007784.

21. Coorey G, Figtree GA, Fletcher DF, et al. The health digital twin to tackle cardiovascular disease—a review of an emerging interdisciplinary field. *NPJ Digital Med* 2022; 5: 126.
22. Calcaterra V, Pagani V and Zuccotti G. Maternal and fetal health in the digital twin era. *Front Pediatr* 2023; 11: 1251427.
23. Jiang J, Wang H, Bai J, et al. Intrapartum ultrasound image segmentation of pubic symphysis and fetal head using dual student–teacher framework with CNN-ViT Collaborative Learning. In: International conference on medical image computing and computer-assisted intervention, 2024. Springer.
24. Bai J, Zhou Z, Ou Z, et al. PSFHS challenge report: pubic symphysis and fetal head segmentation from intrapartum ultrasound images. *Med Image Anal* 2024; 99: 103353.
25. Qiu R, Zhou M, Bai J, et al. PSFHSP-Net: an efficient lightweight network for identifying pubic symphysis-fetal head standard plane from intrapartum ultrasound images. *Med Biol Eng Comput* 2024; 62(10): 2975–2986. <https://doi.org/10.1007/s11517-024-03111-1>
26. Chen G, Bai J, Ou Z, et al. PSFHS: intrapartum ultrasound image dataset for AI-based segmentation of pubic symphysis and fetal head. *Sci Data* 2024; 11: 436.
27. Chen Z, Lu Y, Long S, et al. Fetal head and pubic symphysis segmentation in intrapartum ultrasound image using a dual-path boundary-guided residual network. *IEEE J Biomed Health Inform* 2024; 28: 4648–4659.
28. Zhou Z, Lu Y, Bai J, et al. Segment anything model for fetal head-pubic symphysis segmentation in intrapartum ultrasound image analysis. *Expert Syst Appl* 2024; 245: 125699.
29. Liu H, Wang H, Bai J, et al. DeepSSR: a deep learning system for structured recognition of text images from unstructured paper-based medical reports. *Ann Transl Med* 2022; 10: 740.
30. Bai J, Wang W, Kang X, et al. Machine learning-based prediction of fetal health using cardiotocography. *ESS Open Arch Eprints* 2024; 560: 56010945.
31. Al Housseini A, Newman T, Cox A, et al. Prediction of risk for cesarean delivery in term nulliparas: a comparison of neural network and multiple logistic regression models. *Am J Obstet Gynecol* 2009; 201: 113.e1–113.e6.
32. Boulmedais M, Monperrus M, Corbel E, et al. Predictive value of head-perineum distance measured at the initiation of the active second stage of labor on the mode of delivery: a prospective cohort study. *Eur J Obstet Gynecol Reprod Biol* 2023; 280: 132–137.
33. Coutinho-Almeida J, Cardoso A, Cruz-Correia R, et al. Fast healthcare interoperability resources-based support system for predicting delivery type: model development and evaluation study. *JMIR Form Res* 2024; 8: e54109.
34. Dall’Asta A, Ramirez Zegarra R, Corno E, et al. Role of fetal head-circumference-to-maternal-height ratio in predicting cesarean section for labor dystocia: prospective multicenter study. *Ultrasound Obstet Gynecol* 2023; 61: 93–98.
35. De Ramón Fernández A, Ruiz Fernández D and Prieto Sánchez MT. Prediction of the mode of delivery using artificial intelligence algorithms. *Comput Methods Programs Biomed* 2022; 219: 106740.
36. Eggebø TM, Hassan WA, Salvesen KÅ, et al. Prediction of delivery mode by ultrasound-assessed fetal position in nulliparous women with prolonged first stage of labor. *Ultrasound Obstet Gynecol* 2015; 46: 606–610.
37. Fergus P, Hussain A, Al-Jumeily D, et al. Classification of caesarean section and normal vaginal deliveries using foetal heart rate signals and advanced machine learning algorithms. *Biomed Eng Online* 2017; 16: 89.
38. Fergus P, Selvaraj M and Chalmers C. Machine learning ensemble modelling to classify caesarean section and vaginal delivery types using cardiotocography traces. *Comput Biol Med* 2018; 93: 7–16.
39. Gimovsky AC, Zhuo D, Levine JT, et al. Benchmarking cesarean delivery rates using machine learning-derived optimal classification trees. *Health Serv Res* 2022; 57: 796–805.
40. Grobman WA, Lai Y, Landon MB, et al. Development of a nomogram for prediction of vaginal birth after cesarean delivery. *Obstet Gynecol* 2007; 109: 806–812.
41. Guedalia J, Lipschuetz M, Cohen SM, et al. Transporting an artificial intelligence model to predict emergency cesarean delivery: overcoming challenges posed by interfacility variation. *J Med Internet Res* 2021; 23: e28120.
42. Hadad S, Oberman M, Ben-Arie A, et al. Intrapartum ultrasound at the initiation of the active second stage of labor predicts spontaneous vaginal delivery. *Am J Obstet Gynecol MFM* 2021; 3: 100249.
43. Kuanar A, Akbar A, Sujata P, et al. Deep neural network (DNN) modelling for prediction of the mode of delivery. *Eur J Obstet Gynecol Reprod Biol* 2024; 297: 241–248.
44. Lipschuetz M, Guedalia J, Rottenstreich A, et al. Prediction of vaginal birth after cesarean deliveries using machine learning. *Am J Obstet Gynecol* 2020; 222: 613.e1–613.e12.
45. Manchu M and Redla V. Prediction of mode of delivery by an ultrasound score similar to Bishop score and performance of independent predictors. *J Ultrasound* 2023; 26: 619–626.
46. Meyer R, Hendin N, Zamir M, et al. Implementation of machine learning models for the prediction of vaginal birth after cesarean delivery. *J Matern Fetal Neonatal Med* 2022; 35: 3677–3683.
47. Meyer R, Weisz B, Eilenberg R, et al. Utilizing machine learning to predict unplanned cesarean delivery. *Int J Gynaecol Obstet* 2023; 161: 255–263.
48. Rao J, Fan D, Li P, et al. Prediction of vaginal birth after cesarean in China. *Int J Gynaecol Obstet* 2023; 163: 202–210.
49. Ricciardi C, Improta G, Amato F, et al. Classifying the type of delivery from cardiotocographic signals: a machine learning approach. *Comput Methods Programs Biomed* 2020; 196: 105712.
50. Rossi RM, Requarth E, Warshak CR, et al. Risk calculator to predict cesarean delivery among women undergoing induction of labor. *Obstet Gynecol* 2020; 135: 559–568.
51. Saleem S, Naqvi SS, Manzoor T, et al. A strategy for classification of “vaginal vs. cesarean section” delivery: bivariate empirical mode decomposition of cardiotocographic recordings. *Front Physiol* 2019; 10: 246.
52. Schulz KW., Gaither K, Zigler C, et al. Optimal mode of delivery in pregnancy: individualized predictions using national vital statistics data. *PLOS Digit Health* 2022; 1: e0000166.
53. Ullah Z, Saleem F, Jamjoom M, et al. Reliable prediction models based on enriched data for identifying the mode of childbirth by using machine learning methods: development study. *J Med Internet Res* 2021; 23: e28856.

54. Wie JH, Lee SJ, Choi SK, et al. Prediction of emergency cesarean section using machine learning methods: development and external validation of a Nationwide Multicenter Dataset in Republic of Korea. *Life (Basel)* 2022; 12: 604.
 55. Wu Q, Lu Y, Kang X, et al. ETCNN: an ensemble transformer-convolutional neural network for automatic analysis of fetal heart rate. *Biomed Signal Process Control* 2024; 96: 106629.
 56. Liu M, Zeng R, Xiao Y, et al. Automated fetal heart rate analysis for baseline determination using EMAU-Net. *Inf Sci (NY)* 2023; 644: 119281.
 57. Bai J, Pan X, Lu Y, et al. Comparison of fetal heart rate baseline estimation by the cardiotocograph network and clinicians: a multidatabase retrospective assessment study. *Front Cardiovasc Med* 2023; 10: 1059211.
 58. Liu M, Zeng R, Xiao Y, et al. Baseline/acceleration/deceleration determination of fetal heart rate signals using a novel ensemble LCResU-net. *Expert Syst Appl* 2023; 218: 119610.
 59. Guedalia J, Lipschuetz M, Novoselsky-Persky M, et al. Real-time data analysis using a machine learning model significantly improves prediction of successful vaginal deliveries. *Am J Obstet Gynecol* 2020; 223: 437.e1–437.e15.
 60. Ghosh AK., Catelli DS, Wilson S, et al. Multi-modal detection of fetal movements using a wearable monitor. *Inf Fusion* 2024; 103: 102124.
 61. Ghosh AK, Balasubramanian S, Devasahayam S, et al. Detection and analysis of fetal movements using an acoustic sensor-based wearable monitor. In: 2020 7th International Conference on Information Science and Control Engineering (ICISCE), 2020. IEEE.
 62. Mesbah M, Khlif MS, Layeghy S, et al. Automatic fetal movement recognition from multi-channel accelerometry data. *Comput Methods Programs Biomed* 2021; 210: 106377.
 63. Vogel JP, Pujar Y, Vernekar SS, et al. Effects of the WHO Labour Care Guide on cesarean section in India: a pragmatic, stepped-wedge, cluster-randomized pilot trial. *Nat Med* 2024; 30: 463–469.
 64. Lu Y, Zhou M, Zhi D, et al. The JNU-IFM dataset for segmenting pubic symphysis-fetal head. *Data Brief* 2022; 41: 107904.
 65. Chen Z, Ou Z, Lu Y, et al. Direction-guided and multi-scale feature screening for fetal head–pubic symphysis segmentation and angle of progression calculation. *Expert Syst Appl* 2024; 245: 123096.
 66. Ou Z, Bai J, Chen Z, et al. RTSeg-Net: a lightweight network for real-time segmentation of fetal head and pubic symphysis from intrapartum ultrasound images. *Comput Biol Med* 2024; 175: 108501.
 67. Zhou M, Wang C, Lu Y, et al. The segmentation effect of style transfer on fetal head ultrasound image: a study of multi-source data. *Med Biol Eng Comput* 2023; 61: 1017–1031.
 68. Bai J, Sun Z, Yu S, et al. A framework for computing angle of progression from transperineal ultrasound images for evaluating fetal head descent using a novel double branch network. *Front Physiol* 2022; 13: 940150.
 69. Zhong M, Yi H, Lai F, et al. CTGNet: automatic analysis of fetal heart rate from cardiotocograph using artificial intelligence. *Matern Fetal Med* 2022; 4: 103–112.
 70. Xiao Y, Lu Y, Liu M, et al. A deep feature fusion network for fetal state assessment. *Front Physiol* 2022; 13: 969052.
 71. Lu Y, Zhi D, Zhou M, et al. Multitask deep neural network for the fully automatic measurement of the angle of progression. *Comput Math Methods Med* 2022; 2022: 5192338.
 72. Liu M, Lu Y, Long S, et al. An attention-based CNN-BiLSTM hybrid neural network enhanced with features of discrete wavelet transformation for fetal acidosis classification. *Expert Syst Appl* 2021; 186: 115714.
 73. Zeng R, Lu Y, Long S, et al. Cardiotocography signal abnormality classification using time-frequency features and ensemble cost-sensitive SVM classifier. *Comput Biol Med* 2021; 130: 104218.
-