









Systematic Review

Artificial Intelligence for Predicting Delivery Modes: A Systematic Review of Applications, Challenges, and Future Directions

Yu Zhou¹, Jing Li¹, Xiaomei Hou¹, Zhen Li¹, Yanan Xu¹, Yan Wang¹,
Mingze Sun¹, Fumin Zheng¹, Enhui Guo², Jun Zhou^{1,*}

¹Department of Obstetrics, The Affiliated Hospital of Qingdao University, 266003 Qingdao, Shandong, China

²Department of Gynecology, The Affiliated Hospital of Qingdao University, 266003 Qingdao, Shandong, China

*Correspondence: zhoujun@qdu.edu.cn (Jun Zhou)

Academic Editors: Ugo Indraccolo and Michael H. Dahan

Submitted: 9 February 2025 Revised: 10 May 2025 Accepted: 28 May 2025 Published: 25 July 2025

Abstract

Background: The application of artificial intelligence (AI) in medicine has advanced significantly, particularly in obstetrics, where it plays an increasingly prominent role in predicting modes of delivery and assessment of maternal risks. AI-assisted prediction of delivery modes, a cutting-edge field at the intersection of medicine and computer science, aims to support clinicians in making more accurate and safer delivery decisions by utilizing advanced AI technologies and big data analytics. With increasing individual variability among pregnant women, traditional clinical experience is often insufficient to meet the requirements of personalized medicine; therefore, establishing a scientific prediction model is particularly crucial. This systematic review aims to evaluate the current state of research on AI-assisted prediction of delivery modes, compare AI predictions and traditional statistical methods, and propose future research directions. **Methods:** A comprehensive literature search was conducted in the PubMed, Web of Science, and ScienceDirect databases, encompassing publications up to November 2024. **Results:** Analysis of existing studies demonstrates that AI models outperform conventional statistical methods in predicting delivery modes, highlighting their potential as valuable tools in obstetric diagnosis and clinical decision-making. However, several critical limitations persist in current research, including: (a) the absence of real-time decision support during dynamic labor progression; (b) insufficient multi-center collaboration and a lack of external validation frameworks; and (c) inadequate standardization of clinical parameters (e.g. inconsistent definitions of cervical dilation thresholds and fetal descent metrics). These methodological gaps limit the clinical applicability and generalizability of AI-driven predictive systems across diverse obstetric populations and care settings. **Conclusions:** Future research should prioritize data standardization and sharing, enhance the generalizability of prediction models, address ethical considerations, and ensure the fairness and transparency of AI algorithms to improve clinical trust and applicability. **Registration:** The study has been registered on <https://www.crd.york.ac.uk/prospero/> (registration number: CRD420251068005).

Keywords: artificial intelligence; mode of delivery prediction; cesarean section; natural childbirth; obstetrics; machine learning

1. Introduction

Artificial intelligence (AI) is a broad and all-encompassing term. Machine learning (ML), deep learning, and natural language processing (NLP) are its subtypes [1]. The widespread use of AI has completely transformed various fields of life, including business and trade, social and electronic media, education and learning, manufacturing, and medicine [2]. Particularly in the medical field, the application of AI in predicting and assisting in guiding diagnosis and treatment, providing personalized medical services for patients, and other aspects is becoming increasingly common. With the continuous development and improvement of AI, its status in the medical field is also rising. Currently, AI often assists in medical decision-making by extracting features from large and complex data sets. Current research proposed that AI has provided assistance and convenience in multiple medical fields, such as disease prediction, decision-making based on extracted medical features, and patient management [2]. Addition-

ally, with the continuous improvement of electronic health records and the increase in available data, AI is increasingly being used to establish predictive models to assist in medical decision-making and clinical consultations [3]. These predictive models are the foundation of biomedical research and are used as an indispensable part of the clinical decision-making process [4].

As the application of AI in the medical field deepens, especially in obstetrics, its contribution to predicting delivery methods and assessing maternal risks is becoming increasingly prominent [5]. For example, in the prediction of preterm birth, Chen H-Y *et al.* [6] employed neural networks and decision tree algorithms to identify factors associated with preterm delivery. Rawashdeh H *et al.* [7] utilized random forest (RF), decision trees (DT), K-nearest neighbors (KNN), and neural networks (NN) to assess the risk of preterm birth. In the context of shoulder dystocia prediction, Tsur A *et al.* [8] developed and externally validated a machine learning model integrating maternal risk



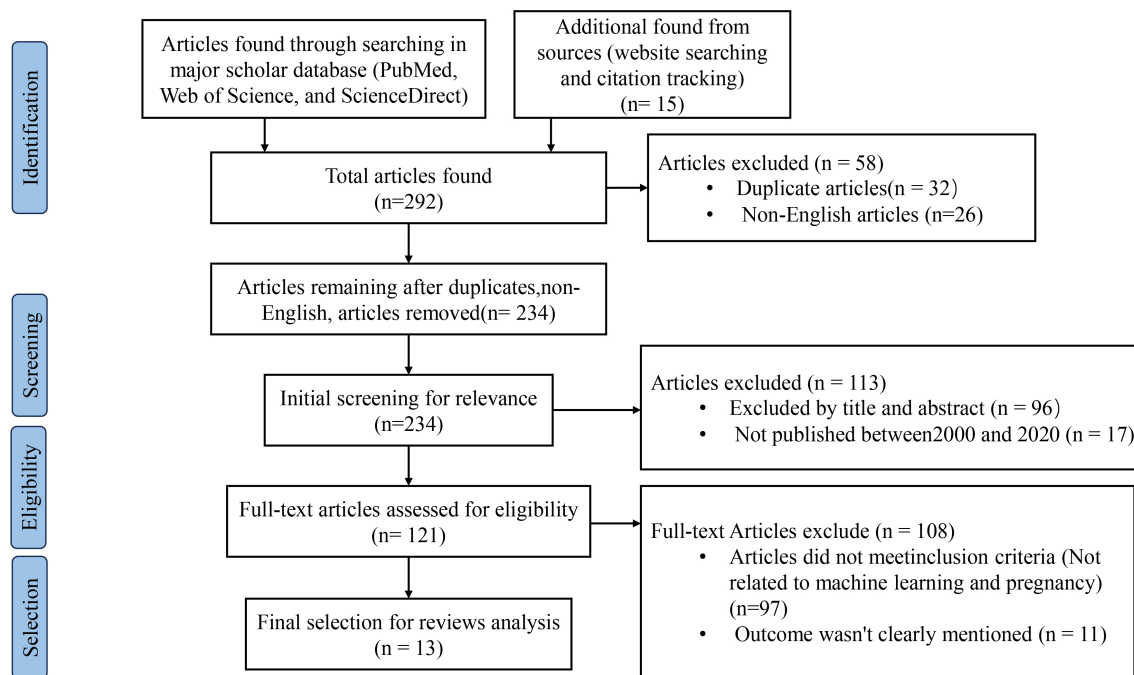


Fig. 1. Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) flow diagram for the selection of articles.

factors with fetal biometric parameters through biostatistical methods to forecast shoulder dystocia. Furthermore, AI has been increasingly applied in predicting the mode of delivery. Currently, the main delivery methods are vaginal delivery and cesarean section. When complications occur during vaginal delivery, vacuum extractions and obstetric pincers can be chosen [9]. How to assist pregnant women in choosing the most appropriate delivery method is a fundamental capability that obstetricians must possess. The choice of delivery method mainly relies on the experience of obstetricians and the results of auxiliary examinations, which require high experience from the physicians and are subject to subjectivity and uncertainty, lacking certain objective data support. In addition, the subjective feelings of pregnant women also affect the choice of delivery method. Usually, due to factors such as unbearable pain and fear, the rate of cesarean section increases. Therefore, for obstetricians, accurately predicting the delivery method remains a challenge [10]. Accurate prediction of vaginal delivery and cesarean section can reduce unnecessary medical intervention, optimize maternal and neonatal outcomes, improve delivery prognosis, and lower medical costs.

Currently, the prediction of delivery methods mainly relies on clinical judgment and traditional statistical methods, which have limitations. Traditional statistical methods can only include a limited number of variables and may not fully capture the complex interactions among various risk factors, making them susceptible to subjective biases. Limited data also restricts the ability to conduct a comprehensive assessment. Moreover, the application of traditional statistical methods in predicting delivery methods is also

limited, which leads to restrictions on the amount of data analyzed in a single analysis and the ability to analyze complex data sets. In contrast, AI, as a tool with the ability to analyze complex data sets and identify complex patterns, provides a more powerful and accurate solution for assisting in predicting delivery methods. Recent studies have shown that ML algorithms and other AI technologies can effectively predict delivery methods by analyzing various factors, including maternal age, body mass index, fetal position, and previous pregnancy and childbirth history [11]. AI-assisted prediction of obstetric delivery methods, as a frontier field of intersection between medicine and computer science, aims to assist clinicians in making more accurate and safe delivery decisions by applying advanced AI technologies and big data analysis methods [1]. With the increasing individual differences among pregnant women, traditional clinical experience makes it difficult to meet the needs of personalized medical care, making the establishment of scientific predictive models particularly important. With the rapid advancement and extensive application of AI, AI-assisted prediction of delivery methods has emerged as a crucial tool for facilitating the professional growth of obstetricians and aiding in their judgment. The existing research mainly concentrates on predicting delivery methods, vaginal delivery, cesarean section, and vaginal delivery after cesarean section.

This review aims to analyze the current research status of AI applications in assisting the prediction of delivery methods, identify the shortcomings in current research, and propose future research directions to improve the application of AI in assisting the prediction of delivery methods.

Ultimately, this will help improve the health outcomes of pregnant women and fetuses and enhance the health levels of mothers and infants.

2. Method

2.1 Search Strategy

A comprehensive search strategy was employed to identify relevant studies on the application of AI in predicting the mode of delivery. The databases searched included PubMed, Web of Science, and ScienceDirect. The search terms used were combinations of artificial intelligence, machine learning, mode of delivery, cesarean section, vaginal delivery, and obstetrics. Boolean operators (AND, OR) were utilized to refine the search. The search was limited to articles published from inception up until November 2024 to ensure the inclusion of the most recent and relevant studies. In addition, reference lists of identified articles were further reviewed for inclusion. This study was previously registered with International Prospective Register of Systematic Reviews (PROSPERO) (CRD: 420251068005) and followed Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.

2.2 Inclusion and Exclusion

In order to ensure that only suitable articles are being selected for this study, some eligibility criteria were considered. Out of the 292 studies, only 13 studies were considered for the systematic review. The few selected articles were chosen using some inclusion-exclusion criteria. A study was eligible for reviewing if it met all the following criteria: (a) investigations utilizing AI-based techniques for predicting mode of delivery; (b) full-text research articles; and (c) publications dated between 2000 and 2020. Exclusion criteria comprised: (a) abstract-only studies; (b) duplicate publications; and (c) non-English language articles. If an article undeniably met one or more of these criteria, it was ruled out from later review. The summary of the search and selection of final articles are illustrated in Fig. 1. The papers were selected by focusing on the abstract and introduction mainly. 292 research works were discovered as primary materials during the preliminary search. 58 articles were chosen after duplicates, non-English articles were removed. After evaluating articles' titles and abstracts, the first level of screening yielded 121 articles excluding 113 articles. Following that, after reading the abstract and introduction, and methodology, the next level of screening was carried out, yielding a list of 13 articles that were selected for the final review analysis. The study characteristics (e.g., AI/ML models used, sample size, outcome measures, performance metrics) and quality assessment results of the reviewed articles included in this systematic review are summarized in Table 1 (Ref. [2,3,11–21]).

2.3 Quality Assessment

The Newcastle-Ottawa Scale (NOS) was applied to evaluate study quality (Table 1), with scores ≤ 5 indicating high risk of bias, 6–7 moderate, and ≥ 8 low risk. Through comparison, it can be found that high-quality studies (NOS ≥ 8) usually have reliable external validation (Guedalia J *et al.* [15]) or large-scale multicohort designs (Meyer R *et al.* [18]); However, low-quality studies (NOS ≤ 5) are generally limited by small samples or absent sensitivity analyses. This systematic review employed the NOS to assess the methodological quality of the 13 reviewed articles, revealing significant heterogeneity in study quality. High-quality studies (scores 8–9, $n = 5$) were characterized by the use of multicenter large-scale cohorts, rigorous data preprocessing, external validation to ensure model robustness, and the integration of critical clinical variables (e.g., induction indications, obstetric history) with adjustment for confounding variables. Moderate-quality studies (scores 5–7, $n = 8$) were constrained by single-center small-sample designs, overreliance on synthetic data, or inadequate validation protocols. While these studies demonstrated methodological innovations, the generalizability of their findings remains questionable. Future research should prioritize expanding multicenter data collaboration, standardizing dynamic feature extraction protocols, and strengthening model calibration and clinical translational validation to enhance the practical utility of predictive tools.

3. Results

3.1 Predicting the Mode of Delivery

A systematic analysis of three pivotal studies [2,11,12] reveals distinct methodological approaches and performance outcomes in AI-driven delivery mode prediction (Table 2, Ref. [2,11,12]). Data were categorized into three dimensions: algorithmic architecture, dataset characteristics, and validation rigor.

3.1.1 Algorithmic Diversity and Performance

(a) Traditional ML Models: De Ramón Fernández A *et al.* [11] employed Support Vector Machine (SVM), Multilayer Perceptron (MLP), and Random Forest (RF) on a large retrospective cohort ($n = 25,038$), achieving $\geq 90\%$ accuracy in distinguishing cesarean versus vaginal delivery and 86–87% accuracy for instrumental versus eutocic vaginal delivery. Notably, RF demonstrated marginal superiority (accuracy 91% vs. SVM 90%, MLP 90%), suggesting ensemble methods' robustness in handling heterogeneous obstetric data.

Ullah Z *et al.* [2] compared five ML algorithms (DT, RF, AdaBoostM1, Bagging, k-NN) on a small enriched dataset ((synthetic minority oversampling technique) SMOTE-augmented $n = 80$). K-NN achieved the highest accuracy (84.38%), while AdaBoostM1 performed poorest (80.63%). Data augmentation improved model performance by 3–5%, highlighting its utility in addressing class imbalance.

Table 1. Study characteristics and the quality assessment results of the reviewed articles.

Authors	Reference	Year	AI/ML models used	Sample size	Outcome measures	Performance metrics	NOS score
De Ramón Fernández A <i>et al.</i>	[11]	2022	SVM, MLP, RF	25,038 records	Cesarean section, eutocic vaginal delivery, instrumental vaginal delivery	SVM: accuracy >90% (cesarean vs. vaginal), 87% (instrumental vs. eutocic); MLP: accuracy 90% (cesarean vs. vaginal), 86% (instrumental vs. eutocic); RF: accuracy 91% (cesarean vs. vaginal), 87% (instrumental vs. eutocic)	9
Ullah Z <i>et al.</i>	[2]	2021	DT, RF, AdaBoostM1, Bagging, and K-NN	80 records	Predict the mode of delivery (cesarean section, vaginal delivery)	k-NN: accuracy 84.38%; bagging: accuracy 83.75%; RF: accuracy 83.13%; DT: accuracy 81.25%; AdaBoostM1: accuracy 80.63%	5
Kuanar A <i>et al.</i>	[12]	2024	DNN	101 records	Predict the mode of delivery (cesarean section, vaginal delivery)	Train set: AUC 0.99, KS score 0.98; The prediction error rates: cesarean section 0.02, vaginal delivery 0.00	5
Ferreira I <i>et al.</i>	[13]	2025	LR, MLP, RF, SVM, XGBoost, and AdaBoost classifiers	2434 records	Predict vaginal delivery after labor induction	LR: AUC 0.794, sensitivity 0.766, specificity 0.910; RF: AUC 0.777, sensitivity 0.756, specificity 0.904; SVM: AUC 0.774, sensitivity 0.747, specificity 0.954; AdaBoost: AUC 0.767, sensitivity 0.753, specificity 0.890; XGBoost: AUC 0.754, sensitivity 0.738, specificity 0.863; MLP: AUC 0.744, sensitivity 0.737, specificity 0.855	8
Wong MS <i>et al.</i>	[14]	2024	Automated ML (Partometer)	37,932 records	Predict vaginal delivery	Partometer accuracy: 87.1%, AUC: 0.82	6
Guedalia J <i>et al.</i>	[15]	2021	Gradient boosting (CatBoost)	94,480 records	Predict cesarean section	Admission data only: AUC of 0.817; Real-time cervical examination data: Initial AUC of 0.819, increasing to 0.917; Real-time FHR data: Initial AUC of 0.824, increasing to 0.928; All-inclusive real-time data: Initial AUC of 0.833, increasing to 0.932	9
Fergus P <i>et al.</i>	[16]	2017	Deep learning classifiers, Fisher's linear discriminant analysis classifiers, and random forest classifiers	506 controls and 46 cases	Predict cesarean section	Deep learning classification: sensitivity = 94%, specificity = 91%, Area under the curve = 99%, F-score = 100%, and mean square error = 1%	5

Table 1. Continued.

Authors	Reference	Year	AI/ML models used	Sample size	Outcome measures	Performance metrics	NOS score
Nagayasu Y <i>et al.</i>	[17]	2022	Continuous recursive rule extraction (Re-RX) algorithm with J48graft	1513 singleton deliveries	Predict an emergency cesarean section	Average accuracy: 81.90%; AUC: 71.46%	5
Meyer R <i>et al.</i>	[18]	2023	XGBoost, DRF, GBM, XRT	73,667 records	Predict unplanned cesarean delivery	Training data set: AUC 0.874; Validation data set: AUC 0.839; Test data set: AUC 0.84 (XGBoost)	8
Islam MS <i>et al.</i>	[19]	2022	GNB, LDA, KNN, GBC, LR	15,409 records	Predict cesarean section	HGSORF: accuracy 98.34%; GBC: accuracy 93.20%; GNB: accuracy 87.36%; KNN: accuracy 88.32%; LDA: accuracy 91.90%; LR: accuracy 92.24%	5
Lindblad Wollmann C <i>et al.</i>	[3]	2021	Conditional inference tree, Conditional RF, Lasso binary regression	3116 records	Predict vaginal birth after previous cesarean	AUC ranged from 0.61 to 0.69, with sensitivity (probability of correctly identifying a VBAC for second delivery) above 91% and specificity (probability of correctly identifying a repeat CD for second delivery) below 22% for all models	9
Meyer R <i>et al.</i>	[20]	2022	RF, GLM, XGBoost	989 records	Predict successful VBAC or failed TOLAC	RF: AUC-PR 0.351, XGBoost: AUC-PR 0.350, GLM: AUC-PR 0.336; MFMU: AUC-PR 0.325	7
Lipschuetz M <i>et al.</i>	[21]	2020	Gradient boosting	9888 records	Predict VBAC	First-trimester model: AUC of 0.745; Pre-labor model: AUC of 0.793; stratification into risk groups with VBAC success rates of 97.3% (low), 90.9% (medium), and 73.3% (high)	5

AI, artificial intelligence; ML, Machine Learning; NOS, Newcastle-Ottawa Scale; SVM, support vector machine; RF, random forest; DT, decision trees; k-NN, k-nearest neighbor; GNB, Gaussian Naive Bayes; LR, logistic regression; FHR, fetal heart rate; AUC, Area Under the ROC Curve; AUROC, areas under the receiver-operating-characteristics curve; AdaBoostM1, Adaptive Boosting version M1; TOLAC, trial of labor after cesarean delivery; MFMU, Maternal-Fetal Medicine Units; XGBoost, eXtreme Gradient Boosting; DNN, Deep Neural Networks; LDA, Linear Discriminant Analysis; GBC, Gradient Boosting Classifier; GLM, Generalized Linear Model; CD, Cesarean Delivery; VBAC, Vaginal Birth After Cesarean; PR, Precision-Recall; KS, Kolmogorov-Smirnov statistic; HGSORF, Henry Gas Solubility Optimization-based Random Forest; MLP, Multilayer Perceptron.

Table 2. Synthesis of key findings (predicting the mode of delivery).

Metric	De Ramón Fernández A <i>et al.</i> [11]	Ullah Z <i>et al.</i> [2]	Kuanar A <i>et al.</i> [12]
Sample size	25,038 (retrospective)	80 (SMOTE-augmented)	101 (single-center)
Top algorithm	Random Forest (91% Acc)	k-NN (84.38% Acc)	DNN (AUC 0.99)
Strengths	Large-scale validation	SMOTE efficacy proven	High theoretical AUC
Limitations	Static features only	High risk of overfitting	Minimal external validity

Acc, accuracy; SMOTE, synthetic minority oversampling technique.

Table 3. Synthesis of key findings (predicting vaginal delivery).

Metric	Ferreira I <i>et al.</i> [13]	Wong MS <i>et al.</i> [14]
Sample size	2434 (retrospective)	37,932 (real-time intrapartum)
Top algorithm	Logistic Regression (AUC 0.79)	AutoML (AUC 0.82)
Key predictors	Bishop score, maternal height	Cervical dilation rate
Strengths	High interpretability	Dynamic data integration
Limitations	Static features only	Limited model transparency

AutoML, automated machine learning.

(b) Deep Learning (DL) Innovations: Kuanar A *et al.* [12] pioneered Deep Neural Networks (DNN) adoption ($n = 101$), reporting exceptional training metrics (Area Under the ROC Curve (AUC) = 0.99, Kolmogorov-Smirnov statistic (KS) = 0.98) but limited external validity due to minimal sample size. Prediction error rates for cesarean and vaginal delivery were 0.02 and 0.00, respectively, though these results may reflect overfitting.

3.1.2 Dataset Characteristics and Limitations

(a) Scale Disparity: Large-scale studies [11] ($n = 25,038$) demonstrated stable performance (AUC 0.87–0.91), whereas small cohorts [2,12] ($n = 80$ – 101) exhibited inflated metrics (AUC up to 0.99), likely due to limited generalizability.

(b) Feature Engineering: Maternal age, Body Mass Index (BMI), and parity (Static Parameters) dominated input features across studies [2,11]. None incorporated real-time intrapartum progression metrics (e.g., cervical dilation rate), constraining clinical utility [11].

3.1.3 Validation and Clinical Applicability

(a) Internal Validation: All studies used random splits, with only De Ramón Fernández A *et al.* [11] achieving NOS = 9 through rigorous sensitivity analysis.

(b) Temporal/External Gaps: No study implemented prospective temporal validation, and inter-hospital transportability tests were absent.

3.2 Predicting Vaginal Delivery

A synthesis of two seminal studies [13,14] demonstrates advancements and limitations in AI-driven vaginal delivery prediction, categorized into algorithmic innovation, dynamic data integration, and clinical validation (Table 3, Ref. [13,14]).

3.2.1 Algorithmic Approaches and Performance

(a) Traditional ML Models: Ferreira I *et al.* [13] developed a multivariable logistic regression (LR) model using retrospective data from singleton term pregnancies ($n = 2434$). The LR model achieved an area under the receiver-operating-characteristics curve (AUROC) of 0.794, with Bishop score, maternal height, and inter-delivery interval identified as top predictors through SHAP analysis. Despite moderate discrimination, the model prioritized clinical interpretability over complex architectures. This study excluded real-time intrapartum parameters, relying solely on static admission features, which constrained its utility in dynamic labor management.

(b) Automated Machine Learning (AutoML) Innovations: Wong MS *et al.* [14] pioneered AutoML (Partometer) using real-time intrapartum data ($n = 37,932$). The model achieved 87.1% accuracy and AUC = 0.82 in predicting vaginal delivery within 4 hours of admission. Key dynamic predictors included cervical dilation rate and fetal head descent, underscoring the value of temporal feature integration. While AutoML streamlined model development, the “black-box” nature of feature selection reduced clinician interpretability, a critical barrier to adoption.

3.2.2 Dataset Characteristics and Clinical Relevance

(a) Scale and Diversity: Large-Scale Dynamic Data [14]: The inclusion of 37,932 records with real-time monitoring metrics provided robust statistical power, though the cohort was limited to primiparous women in high-resource settings.

Static Data Limitations [13]: Despite a moderate sample size ($n = 2434$), reliance on retrospective, single-center data impeded generalizability to diverse obstetric populations.

Table 4. Synthesis of key findings (predicting cesarean section).

Metric	Guedalia J <i>et al.</i> [15]	Fergus P <i>et al.</i> [16]	Nagayasu Y <i>et al.</i> [17]	Meyer R <i>et al.</i> [18]	Islam MS <i>et al.</i> [19]
Sample size	989 (inter-hospital)	552 (FHR signals)	1513 (single-center)	73,667 (multi-center)	15,409 (balanced)
Top algorithm	RF (AUC-PR 0.351)	Deep Learning (AUC 0.99)	Re-RX (Acc 81.9%)	XGBoost (AUC 0.84)	HGSORF (Acc 98.34%)
Key predictors	Fetal head position	FHR variability	Bishop score, parity	Maternal BMI	Placental markers
Strengths	Transportability focus	High sensitivity	Rule-based clarity	Large-scale validation	XAI interpretability
Limitations	Performance variability	Small sample size	Low AUC	Static features	Limited external tests

XAI, explainable AI; Re-RX, recursive-rule eXtraction.

3.2.3 Validation Rigor and Clinical Utility

(a) Internal Validation: Both studies employed cross-validation [13,14], with Wong MS *et al.* [14] achieving superior discrimination (AUC = 0.82) due to dynamic feature inclusion.

(b) Temporal and External Gaps: No External Validation: Neither study tested models across institutions or regions, risking overfitting to local practice patterns.

Real-World Implementation: While Wong MS *et al.* [14] demonstrated the feasibility of real-time prediction, the absence of clinician-AI interaction protocols limited practical utility.

3.3 Predicting Cesarean Section

A synthesis of five pivotal studies [15–19] reveals diverse algorithmic strategies and challenges in AI-driven cesarean section (CS) prediction, categorized into model architecture, dataset dynamics, and validation rigor (Table 4, Ref. [15–19]).

3.3.1 Algorithmic Approaches and Performance

(a) Traditional ML Models: Meyer R *et al.* [18] employed XGBoost on a large cohort ($n = 73,667$), achieving AUC = 0.84 for unplanned CS prediction. Feature importance analysis identified maternal BMI and labor progression rate as top predictors, aligning with clinical intuition.

Islam MS *et al.* [19] proposed the Henry Gas Solubility Optimization-based Random Forest (HGSORF) algorithm (optimized RF), reporting 98.34% accuracy on a balanced dataset ($n = 15,409$). Explainable AI (XAI) analysis revealed placental insufficiency and uterine contractility patterns as critical drivers, enhancing model interpretability.

(b) Deep Learning (DL) and Hybrid Models: Fergus P *et al.* [16] applied deep learning classifiers to fetal heart rate (FHR) signals ($n = 552$), achieving 94% sensitivity and 91% specificity. AUC = 0.99 underscored DL's potential but raised concerns about overfitting in small samples.

Nagayasu Y *et al.* [17] utilized the recursive-rule eXtraction (Re-RX) rule extraction method ($n = 1513$), yielding 81.9% accuracy and AUC = 0.71. While interpretable, the model's lower discrimination highlighted trade-offs between simplicity and predictive power.

(c) Cross-Institutional Validation: Guedalia J *et al.* [15] tested model transportability between hospitals, finding performance drops (Δ AUC: -0.12) due to interfacility

measurement variability. Adjustments to fetal head position metrics restored parity, emphasizing standardized feature protocols.

3.3.2 Dataset Characteristics and Limitations

(a) Scale and Diversity: Large-Scale Cohorts [18,19]: Studies with >15 k records demonstrated stable performance (AUC 0.84–0.98), whereas smaller datasets [16,17] ($n = 552$ – 1513) exhibited inflated metrics (AUC up to 0.99).

(b) Dynamic Data Integration: Only Fergus P *et al.* [16] incorporated real-time FHR signals, while others relied on static admission parameters (e.g., parity, BMI) [17–19].

3.3.3 Validation and Clinical Utility

(a) Internal Validation: All studies used cross-validation [15–19], with Islam MS *et al.* [19] achieving the highest accuracy (98.34%) through Adaptive Synthetic Sampling Approach (ADASYN)-balanced data.

(b) External and Temporal Gaps

Limited Generalizability: Only Guedalia J *et al.* [15] addressed inter-hospital variability, revealing institutional bias as a critical barrier.

Real-Time Application: Despite high accuracy, no study has implemented prospective real-time prediction in clinical workflows.

3.4 Predicting Vaginal Delivery After Cesarean Section

A synthesis of three pivotal studies [3,20,21] highlights advancements and persistent challenges in AI-driven VBAC prediction, categorized into algorithmic strategies, dataset robustness, and clinical validation (Table 5, Ref. [3,20,21]).

3.4.1 Algorithmic Approaches and Performance

(a) Traditional ML Models: Lindblad Wollmann C *et al.* [3] compared ML models (conditional RF, lasso regression) with existing clinical scores (Swedish cohort, $n = 3116$). All models achieved AUROC 0.61–0.69, with sensitivity $>91\%$ but specificity $<22\%$, indicating strong ability to identify VBAC candidates but poor discrimination for repeat cesarean risks.

Meyer R *et al.* [20] implemented RF and XGBoost ($n = 989$), reporting AUC-PR 0.351 (RF) vs. 0.325 (MFMU model). The XGBoost model required only 8 variables, emphasizing parsimony but sacrificing nuanced risk stratification.

Table 5. Synthesis of key findings (predicting vaginal delivery after cesarean section).

Metric	Lindblad Wollmann C <i>et al.</i> [3]	Meyer R <i>et al.</i> [20]	Lipschuetz M <i>et al.</i> [21]
Sample size	3116 (population-based)	989 (single-center)	9888 (multi-stage)
Top algorithm	Conditional RF (AUC 0.69)	XGBoost (AUC-PR 0.351)	Gradient Boosting (AUC 0.79)
Key predictors	Prior vaginal delivery	Maternal age, BMI	Gestational age, parity
Strengths	Population diversity	Model simplicity	Dynamic risk stratification
Limitations	Low specificity	Limited feature depth	Retrospective data

(b) Dynamic Risk Stratification: Lipschuetz M *et al.* [21] developed gradient boosting models using first-trimester and pre-labor data ($n = 9888$). The pre-labor model achieved AUC = 0.793, significantly outperforming the first-trimester model (AUC = 0.745). Risk stratification categorized 42.4% of women as low-risk (VBAC success 97.3%, demonstrating clinical utility for personalized counseling).

3.4.2 Dataset Characteristics and Limitations

(a) Scale and Diversity: Large National Cohorts [3]: Population-based data ($n = 3116$) enhanced generalizability but lacked granular intrapartum metrics (e.g., cervical dilation trends).

Real-Time Data Gaps: All studies relied on retrospective, static parameters (e.g., prior vaginal delivery, maternal BMI), neglecting dynamic labor progression [20,21].

Geographic Bias: Studies focused on high-income populations (Sweden [3], Israel [21]), limiting applicability to low-resource settings.

3.4.3 Validation and Clinical Relevance

(a) Internal Validation: Lindblad Wollmann C *et al.* [3] used cross-validation but reported low specificity (22%), reducing clinical confidence in avoiding unnecessary cesareans.

Lipschuetz M *et al.* [21] demonstrated temporal validity with pre-labor data integration, aligning closer to clinical workflows.

(b) External and Practical Gaps: No Inter-Hospital Testing: Despite Meyer R *et al.* [20]’s multi-algorithm comparison, no study validated models across institutions, risking practice pattern overfitting.

Patient-Clinician Discordance: Meyer R *et al.* [20] noted 28% patient refusal of VBAC attempts due to anxiety, a psychological factor absent in AI frameworks.

4. Discussion

4.1 Advantages and Challenges

4.1.1 Advantages

AI is a transformative technology that aims to simulate, extend, and augment human intelligence through the development of advanced algorithms and data analysis techniques. It can concurrently handle and analyze a vast amount of clinical data. In facilitating the predic-

tion of delivery methods, it can fully exploit the advantages of big data to enhance the accuracy and reliability of predictions. Moreover, AI enables individualized clinical decision-making. In the current era marked by the progressive improvement of electronic health records, AI can leverage its strengths to comprehensively analyze historical big data. Based on the specific circumstances of each pregnant woman and the current pregnancy examination data, it can offer more personalized guidance. Simultaneously, it can provide more objective support for the clinical diagnosis and treatment work of obstetricians, optimize delivery outcomes, and continuously elevate the health status of mothers and infants.

4.1.2 Shortcomings and Challenges

4.1.2.1 Data Quality and Security. Delivery is a dynamic process. Some unpredictable variables may appear during labour thereby affecting the final outcome. During the labor process, the selection of delivery mode and maternal-fetal outcomes are influenced by a constellation of factors, including objective maternal-fetal parameters, environmental variables, and maternal subjective perceptions. Investigations must comprehensively account for the influence of these multifaceted variables on predictive outcomes. However, inherent methodological limitations inevitably arise in such studies. Current research has predominantly focused on static data parameters, while neglecting the monitoring of dynamic physiological indicators such as fetal heart rate variability and cervical dilation progression. Future studies incorporating these time-varying parameters could significantly enhance the predictive accuracy of delivery mode outcomes.

In addition, the NOS quality assessment reveals that while single-center studies may control for confounding factors influencing delivery mode prediction during labor, multicenter study designs enhance methodological rigor. However, current healthcare data quality exhibits significant heterogeneity across medical institutions, with multimodal data (e.g., electronic health records, imaging, and monitoring signals) suffering from inconsistent acquisition standards and insufficient structuralization, thereby constraining the generalizability of AI-based predictive models. Future research should prioritize establishing a tripartite framework to address these limitations: (1) Standardized Data Acquisition Protocol Development: Implement lifecycle-wide standardized protocols aligned with Health

Level Seven Fast Healthcare Interoperability Resources (HL7 FHIR) specifications to unify perinatal data element definitions (e.g., gestational age measurement rules, delivery mode coding systems). Leverage natural language processing (NLP) techniques to extract structured insights from unstructured labor progression narratives. Concurrently deploy intelligent validation engines for real-time monitoring of data completeness and logical consistency. (2) Privacy-Enhanced Data Sharing Mechanisms: Enable cross-institutional collaborative modeling through federated learning frameworks. Integrate homomorphic encryption and secure multi-party computation (SMPC) to ensure “data usability without visibility” of raw datasets. Implement dynamic de-identification for high-risk pregnancy data. Establish data consortia with governance rules for contribution assessment and ethical oversight. (3) Source-Level Data Quality Control: Directly interface Internet of things (IoT)-enabled devices (e.g., fetal monitors, ultrasound systems) to automate time-series data acquisition and real-time calibration. Systematically reduce manual entry errors through sensor-to-database pipelines. Collectively, these interventions would substantially enhance data utility and lay critical foundations for developing generalizable predictive models.

4.1.2.2 Generalization Ability of Prediction Models. The global obstetric field currently faces significant regional disparities in healthcare resource allocation: variations in medical standards, professional competencies of healthcare providers, and region-specific care delivery models persist across nations and institutions. Maternal and neonatal healthcare outcomes in high-income regions markedly surpass those in low-resource settings, reflecting both technological gradients and strong correlations with regional economic development. Existing perinatal health prediction models predominantly derive from single-center datasets, exhibiting critical limitations in generalizability and cross-institutional interoperability that constrain clinical translation efficacy.

To address these challenges, we propose a “multi-center collaboration × technological empowerment” framework. (1) Establishment of Transnational Perinatal Data Networks: Develop unified perinatal data standards (e.g., diagnostic coding for pregnancy complications, neonatal outcome metrics) through international consensus. Aggregate multi-center clinical data spanning diverse economic contexts to construct standardized datasets covering preconception, antenatal, and postpartum phases. (2) Implementation of Privacy-Preserving AI Architectures: Deploy federated learning systems for distributed model training without raw data transfer. Integrate homomorphic encryption and differential privacy mechanisms to safeguard patient confidentiality. (3) Blockchain-Enhanced Data Equity Solutions: Create blockchain-driven data sharing platforms with contributor recognition protocols.

Implement contribution-weighted benefit allocation to ensure equitable representation of low-resource regions in model development.

This integrated approach effectively mitigates single-center study biases, enhances model adaptability across heterogeneous healthcare environments, and provides scalable technical infrastructure for global maternal-neonatal health optimization.

4.1.2.3 Ethical and Social Implications. In the clinical implementation of AI-assisted delivery mode prediction, robust data sharing mechanisms and standardization frameworks serve as foundational prerequisites for algorithmic optimization and model development. Current obstetric data systems are plagued by core challenges including multi-source heterogeneity, inconsistent standards, and cross-institutional sharing barriers. Without standardized data governance, algorithmic fairness and model generalizability remain fundamentally compromised. We propose implementing a comprehensive data stewardship framework through the following steps. (1) Standardized Perinatal Data Protocols: Develop unified data collection guidelines specifying critical delivery-related metrics (e.g., pelvimetry parameters, labor progression staging), aligned with international standards such as HL7 FHIR for structured interoperability. Establish definitive coding schemas for obstetric indicators through multidisciplinary consensus. (2) Secure Cross-Institutional Collaboration Infrastructure: Deploy federated learning architectures for distributed model training without raw data transfer. Integrate homomorphic encryption and dynamic de-identification techniques to preserve patient confidentiality. Implement quantifiable data contribution assessment mechanisms to incentivize multi-center participation.

Building upon this technical foundation, addressing ethical and societal implications requires prioritized attention. (1) Algorithmic Transparency: Ensure interpretability of feature weights in delivery prediction models through XAI frameworks. (2) Medicolegal Accountability: Formalize legal liability delineation protocols for human-AI decision conflicts. (3) Patient-Centric Governance: Establish dynamic consent management systems for continuous data usage authorization.

Concurrent interdisciplinary collaboration among medical ethics boards, AI developers, and clinical practitioners is imperative to formulate obstetric-specific AI governance guidelines. This tripartite synergy enhances clinician-patient acceptance of predictive systems while balancing technological innovation with ethical imperatives, ultimately fostering responsible integration of AI in maternal care.

4.2 Prospects and Outlook

Through a review of existing research, it becomes evident that AI-assisted prediction of delivery methods,

as a cutting-edge domain at the intersection of medicine and computer science, still demands intensified collaboration between obstetricians and computer scientists. In the ensuing research, efforts should be directed towards further optimizing algorithms and prediction models. On this foundation, data sharing and standardization efforts among various medical centers should be strengthened. We should persistently undertake multi-center studies in collaboration with the global community to ultimately facilitate the widespread application of AI technology in clinical decision-making, achieving individualized and precise medical care. Naturally, while conducting in-depth research, it is imperative to augment the exploration of the ethical and social implications of AI technology.

We anticipate that with the unwavering efforts of all obstetricians and computer scientists, a comprehensive AI-assisted prediction model for delivery methods can be established, furnishing clinical practitioners with more accurate predictions and decision support for delivery methods, enabling personalized clinical decision-making and real-time monitoring and early warning, and providing more all-encompassing and effective safeguards for the health of mothers and infants.

5. Conclusions

This review highlights the significant potential of AI in predicting delivery modes, demonstrating its superiority over traditional statistical methods in terms of accuracy and reliability. However, several challenges remain, including data standardization, model generalizability, and ethical concerns. Future research should prioritize multi-center collaborations to enhance the generalizability of AI models, develop standardized protocols for data collection and sharing, and address the ethical implications of AI in obstetrics. By addressing these challenges, AI can be effectively integrated into clinical practice, ultimately improving maternal and neonatal outcomes.

Availability of Data and Materials

All relevant data are within the manuscript and its supporting information files.

Author Contributions

JZ and YZ designed the research study. YZ was responsible for manuscript writing. The table was conducted by JL and FMZ, while the graphic figures were created by ZL and EHG, the data from studies were compiled by XMH and YNX. YW and MZS have been involved in drafting the manuscript and provided help and advice on the search of reference. All authors contributed to editorial changes in the manuscript. All authors read and approved the final manuscript. All authors have participated sufficiently in the work and agreed to be accountable for all aspects of the work.

Ethics Approval and Consent to Participate

Not applicable.

Acknowledgment

Not applicable.

Funding

This research was funded by Qingdao Outstanding Health Professional Development Fund. This research was also funded by the Clinical Medicine +X Scientific Research Project of the Affiliated Hospital of the Affiliated Hospital of Qingdao University, grant number QDFY+X2024111.

Conflict of Interest

The authors declare no conflict of interest.

Supplementary Material

Supplementary material associated with this article can be found, in the online version, at <https://doi.org/10.31083/CEOG37807>.

References

- [1] Medjedovic E, Stanojevic M, Jonuzovic-Prosic S, Ribic E, Begic Z, Cerovac A, *et al.* Artificial intelligence as a new answer to old challenges in maternal-fetal medicine and obstetrics. *Technology and Health Care: Official Journal of the European Society for Engineering and Medicine*. 2024; 32: 1273–1287. <https://doi.org/10.3233/THC-231482>.
- [2] Ullah Z, Saleem F, Jamjoom M, Fakieh B. Reliable Prediction Models Based on Enriched Data for Identifying the Mode of Childbirth by Using Machine Learning Methods: Development Study. *Journal of Medical Internet Research*. 2021; 23: e28856. <https://doi.org/10.2196/28856>.
- [3] Lindblad Wollmann C, Hart KD, Liu C, Caughey AB, Stephansson O, Snowden JM. Predicting vaginal birth after previous cesarean: Using machine-learning models and a population-based cohort in Sweden. *Acta Obstetrica et Gynecologica Scandinavica*. 2021; 100: 513–520. <https://doi.org/10.1111/aogs.14020>.
- [4] Fergus P, Selvaraj M, Chalmers C. Machine learning ensemble modelling to classify caesarean section and vaginal delivery types using Cardiotocography traces. *Computers in Biology and Medicine*. 2018; 93: 7–16. <https://doi.org/10.1016/j.compbiomed.2017.12.002>.
- [5] Islam MN, Mustafina SN, Mahmud T, Khan NI. Machine learning to predict pregnancy outcomes: a systematic review, synthesizing framework and future research agenda. *BMC Pregnancy and Childbirth*. 2022; 22: 348. <https://doi.org/10.1186/s12884-022-04594-2>.
- [6] Chen H-Y, Chuang C-H, Yang Y-J, Wu T-P. Exploring the risk factors of preterm birth using data mining. *Expert Systems With Applications*. 2011; 38: 5384–5387. <https://doi.org/10.1016/j.eswa.2010.10.017>.
- [7] 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. *Computational Biology and Chemistry*. 2020; 85: 107233. <https://doi.org/10.1016/j.compbiolchem.2020.107233>.
- [8] Tsur A, Batsry L, Toussia-Cohen S, Rosenstein MG, Barak O, Brezinov Y, *et al.* Development and validation of a machine-

- learning model for prediction of shoulder dystocia. *Ultrasound in Obstetrics & Gynecology: the Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology*. 2020; 56: 588–596. <https://doi.org/10.1002/uog.21878>.
- [9] Pereira S, Portela F, Santos MF, Machado J, Abelha A. Predicting Type of Delivery by Identification of Obstetric Risk Factors through Data Mining. *Procedia Computer Science*. 2015; 64: 601–609. <https://doi.org/10.1016/j.procs.2015.08.573>.
 - [10] Islam MN, Mahmud T, Khan NI, Mustafina SN, Islam AKMN. Exploring Machine Learning Algorithms to Find the Best Features for Predicting Modes of Childbirth. *IEEE Access*. 2021; 9: 1680–1692. <https://doi.org/10.1109/access.2020.3045469>.
 - [11] De Ramón Fernández A, Ruiz Fernández D, Prieto Sánchez MT. Prediction of the mode of delivery using artificial intelligence algorithms. *Computer Methods and Programs in Biomedicine*. 2022; 219: 106740. <https://doi.org/10.1016/j.cmpb.2022.106740>.
 - [12] Kuanar A, Akbar A, Sujata P, Kar D. Deep neural network (DNN) modelling for prediction of the mode of delivery. *European Journal of Obstetrics, Gynecology, and Reproductive Biology*. 2024; 297: 241–248. <https://doi.org/10.1016/j.ejogrb.2024.04.012>.
 - [13] Ferreira I, Simões J, Correia J, Areia AL. Predicting vaginal delivery after labor induction using machine learning: Development of a multivariable prediction model. *Acta Obstetrica et Gynecologica Scandinavica*. 2025; 104: 164–173. <https://doi.org/10.1111/aogs.14953>.
 - [14] Wong MS, Wells M, Zamanzadeh D, Akre S, Pevnick JM, Bui AAT, *et al.* Applying Automated Machine Learning to Predict Mode of Delivery Using Ongoing Intrapartum Data in Laboring Patients. *American Journal of Perinatology*. 2024; 41: e412–e419. <https://doi.org/10.1055/a-1885-1697>.
 - [15] Guedalia J, Lipschuetz M, Cohen SM, Sompolinsky Y, Walfisch A, Sheiner E, *et al.* Transporting an Artificial Intelligence Model to Predict Emergency Cesarean Delivery: Overcoming Challenges Posed by Interfacility Variation. *Journal of Medical Internet Research*. 2021; 23: e28120. <https://doi.org/10.2196/28120>.
 - [16] Fergus P, Hussain A, Al-Jumeily D, Huang DS, Bouguila N. Classification of caesarean section and normal vaginal deliveries using foetal heart rate signals and advanced machine learning algorithms. *Biomedical Engineering Online*. 2017; 16: 89. <https://doi.org/10.1186/s12938-017-0378-z>.
 - [17] Nagayasu Y, Fujita D, Ohmichi M, Hayashi Y. Use of an artificial intelligence-based rule extraction approach to predict an emergency cesarean section. *International Journal of Gynaecology and Obstetrics: the Official Organ of the International Federation of Gynaecology and Obstetrics*. 2022; 157: 654–662. <https://doi.org/10.1002/ijgo.13888>.
 - [18] Meyer R, Weisz B, Eilenberg R, Tsadok MA, Uziel M, Sivan E, *et al.* Utilizing machine learning to predict unplanned cesarean delivery. *International Journal of Gynaecology and Obstetrics: the Official Organ of the International Federation of Gynaecology and Obstetrics*. 2023; 161: 255–263. <https://doi.org/10.1002/ijgo.14433>.
 - [19] Islam MS, Awal MA, Laboni JN, Pinki FT, Karmokar S, Mumenin KM, *et al.* HGSORF: Henry Gas Solubility Optimization-based Random Forest for C-Section prediction and XAI-based cause analysis. *Computers in Biology and Medicine*. 2022; 147: 105671. <https://doi.org/10.1016/j.compbiomed.2022.105671>.
 - [20] Meyer R, Hendin N, Zamir M, Mor N, Levin G, Sivan E, *et al.* Implementation of machine learning models for the prediction of vaginal birth after cesarean delivery. *The Journal of Maternal-fetal & Neonatal Medicine: the Official Journal of the European Association of Perinatal Medicine, the Federation of Asia and Oceania Perinatal Societies, the International Society of Perinatal Obstetricians*. 2022; 35: 3677–3683. <https://doi.org/10.1080/14767058.2020.1837769>.
 - [21] Lipschuetz M, Guedalia J, Rottenstreich A, Novoselsky Persky M, Cohen SM, Kabiri D, *et al.* Prediction of vaginal birth after cesarean deliveries using machine learning. *American Journal of Obstetrics and Gynecology*. 2020; 222: 613.e1–613.e12. <https://doi.org/10.1016/j.ajog.2019.12.267>.