



**19th International Conference on New Findings
in Midwifery, Women, Maternity and Infertility
(ICMWMI 2026)**

18 September 2026 -TBILISI GEORGIA

<http://icmwmi.ir>
info@icmwmi.ir

**Machine Learning-Based Prediction Models for Adverse
Pregnancy Outcomes: A Systematic Review**

Marziyeh Raei Mehneh*

PhD Student, Department of Health Information Technology, School of Paramedical and
Rehabilitation Sciences, Mashhad

University of Medical Sciences, Mashhad, Iran

Student Research Committee, Mashhad University of Medical Sciences, Mashhad, Iran
Raeim3@mums.ac.ir

Seyyede Fatemeh Mousavi Baigi

PhD Candidate, Department of Health Information Technology, School of Paramedical and
Rehabilitation Sciences, Mashhad

University of Medical Sciences, Mashhad, Iran

Student Research Committee, Mashhad University of Medical Sciences, Mashhad, Iran.

MousaviBF3@mums.ac.ir

Khalil Kimiafar

Full Professor, Department of Health Information Technology, School of Paramedical and
Rehabilitation Sciences, Mashhad

University of Medical Sciences, Mashhad, Iran

KimiafarKh@mums.ac.ir

Masoumeh Sarbaz

Full Professor, Department of Health Information Technology, School of Paramedical and
Rehabilitation Sciences, Mashhad

University of Medical Sciences, Mashhad, Iran

SarbazM@mums.ac.ir



19th International Conference on New Findings in Midwifery, Women, Maternity and Infertility (ICMWMI 2026)

<http://icmwmi.ir>
info@icmwmi.ir

18 September 2026 -TBILISI GEORGIA

Abstract

Background

Adverse pregnancy outcomes, including gestational diabetes mellitus, preeclampsia, preterm birth, postpartum hemorrhage, fetal complications, and postpartum mental health disorders, remain important challenges in maternal healthcare. Early identification of high-risk pregnancies enables timely intervention and personalized care. Machine learning (ML)-based prediction models have emerged as promising tools for analyzing complex clinical data; however, evidence regarding their performance and clinical applicability remains heterogeneous.

Objective:

This systematic review aimed to evaluate ML-based prediction models for early identification of adverse pregnancy outcomes, focusing on algorithms, predictive factors, validation approaches, and model performance.

Methods:

A systematic review was conducted following PRISMA guidelines. PubMed/MEDLINE, Scopus, Web of Science, Embase, IEEE Xplore, ScienceDirect and SpringerLink were searched for studies published between January 1, 2021 up to August 2026. Original studies developing or validating ML prediction models for pregnancy-related adverse outcomes were included. Data regarding study characteristics, algorithms, predictors, validation methods, and performance measures were extracted.

Results:

Of 90 screened records, 38 studies met the eligibility criteria and were included in the qualitative synthesis. Gestational diabetes mellitus and preeclampsia were the most frequently investigated outcomes. Common algorithms included logistic regression, random forest, support vector machine, XGBoost, gradient boosting, deep learning, and ensemble methods. Major predictors included maternal characteristics, obstetric history, laboratory biomarkers, blood pressure, ultrasound findings, and electronic health record variables. Reported AUC values ranged approximately from 0.70 to 0.95. Limited external validation and insufficient assessment of clinical applicability were frequently observed.

Conclusion:

ML models show potential for early prediction of adverse pregnancy outcomes; however, further multicenter validation, standardized reporting, and clinical integration are required before routine implementation.



19th International Conference on New Findings in Midwifery, Women, Maternity and Infertility (ICMWMI 2026)

<http://icmwmi.ir>
info@icmwmi.ir

18 September 2026 -TBILISI GEORGIA

Keywords: Machine Learning; Artificial Intelligence; Pregnancy Complications; Predictive Models; Maternal Health

1.Introduction

Pregnancy represents a complex and dynamic physiological process during which maternal and fetal biological states may rapidly evolve, potentially resulting in a broad spectrum of adverse clinical outcomes. Major complications—including gestational diabetes mellitus (GDM), hypertensive disorders of pregnancy and preeclampsia, spontaneous preterm birth (PTB), postpartum hemorrhage (PPH), fetal hypoxia, ectopic pregnancy, and postpartum depression (PPD)—continue to impose substantial risks of morbidity, maternal and perinatal mortality, prolonged hospitalizations, and escalated healthcare costs worldwide.

Early identification and accurate risk stratification for these complications are essential to guide timely surveillance, institute preventive therapeutic strategies, and optimize clinical resource allocation. Nevertheless, conventional risk-assessment tools that rely on traditional clinical heuristics or standard multivariable regression models often exhibit limited capacity to capture non-linear relationships, high-order feature interactions, and multi-dimensional longitudinal dynamics, especially when clinical, laboratory, genetic, epigenetic, metabolomic, and time-series data are concurrently available.

Machine learning (ML) offers computational frameworks capable of processing high-dimensional, heterogeneous biomedical data, uncovering latent physiological patterns and complex predictive interactions. Over recent years, supervised algorithms such as Random Forest, Support Vector Machines, Logistic Regression, Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), ensemble architectures, and deep neural networks have been increasingly leveraged to forecast pregnancy-related adverse outcomes. Furthermore, explainable artificial intelligence (XAI) techniques have been integrated to delineate feature contributions and facilitate clinical interpretability and adoption.

In the domain of metabolic complications, extensive efforts have focused on early prediction of GDM. Studies have integrated first-trimester maternal biomarkers, demographic indicators, cell-free DNA (cfDNA), genetic variants, metabolomic signatures, and longitudinal glycemic trajectories to identify women at high risk during early gestation [1–6]. Additionally, deep learning approaches have been applied to predict secondary fetal complications, such as large-for-gestational-age (LGA) infants in pregnancies complicated by GDM [1].

Hypertensive disorders of pregnancy, notably preeclampsia, represent another primary target for ML-driven risk modeling. Researchers have utilized first-trimester maternal factors, circulating



19th International Conference on New Findings in Midwifery, Women, Maternity and Infertility (ICMWMI 2026)

<http://icmwmli.ir>
info@icmwmli.ir

18 September 2026 -TBILISI GEORGIA

biochemical markers, metabolomic profiles, and DNA methylation markers to construct robust predictive models for gestational hypertension and early- or late-onset preeclampsia [7–12]. Furthermore, large-scale electronic health records (EHRs) and longitudinal modeling frameworks, including the PIERS-ML model, have demonstrated notable utility for dynamic maternal risk assessment and adverse outcome stratification among women with preeclampsia [13–16].

Preterm birth prediction has likewise advanced through the utilization of EHR databases, maternal baseline parameters, and multi-timepoint clinical tracking [17,18]. Advanced deep learning architectures, such as temporal Transformer models and gradient-boosted decision trees (e.g., XGBoost, LightGBM), have been implemented to process longitudinal pregnancy records and dynamically update PTB risk trajectories [19,20].

In the context of obstetric emergencies, ML algorithms have been developed to quantify the risk and volume of postpartum hemorrhage, including intraoperative hemorrhage during cesarean deliveries [21–23]. Recent investigations have combined population-scale clinical registries with multimodal inputs, such as radiomics, deep neural networks, and fusion models, to enhance early predictive discrimination and streamline emergency preparedness [23,24].

Beyond physical maternal-fetal outcomes, ML applications have expanded to perinatal mental health. Predictive models incorporating psychosocial, demographic, and clinical factors have been designed for early antepartum and postpartum depression screening, emphasizing model interpretability to foster clinical trust and actionable risk assessment [25–27].

Other emerging obstetric applications encompass the early detection of ectopic pregnancy using interpretable ML [28], multicenter AI-driven identification of fetal hypoxia [29], predictive surveillance of postpartum blood pressure spikes [30], cytokine and proteomic signature profiling for composite pregnancy complications [31,32], and unified ensemble frameworks for comprehensive maternal risk evaluation [33,34].

Despite these encouraging achievements, the existing body of literature exhibits substantial heterogeneity in clinical outcome definitions, study cohorts, data sources, feature selection techniques, preprocessing pipelines, algorithmic architectures, validation paradigms, and reported performance metrics. While most studies emphasize discriminatory performance (e.g., area under the ROC curve), essential translational dimensions—such as probability calibration, external multicenter validation, decision curve analysis, algorithmic fairness, and clinical decision support system (CDSS) integration—remain underreported. Moreover, risks of bias arising from class imbalance, missing data handling, and overfitting continue to constrain the real-world generalizability of these models.

Therefore, this systematic review provides a rigorous synthesis of ML-based prediction models for the early identification of adverse pregnancy outcomes. We evaluate target outcomes, data modalities, predictive feature domains, algorithmic frameworks, validation methodologies, diagnostic performance, interpretability mechanisms, and methodological limitations to delineate



19th International Conference on New Findings in Midwifery, Women, Maternity and Infertility (ICMWMI 2026)

<http://icmwmi.ir>
info@icmwmi.ir

18 September 2026 -TBILISI GEORGIA

the translational pathway toward safe and effective clinical implementation in routine obstetric and midwifery care.

2. Methods

2.1 Study Design and Reporting Framework

This systematic review was designed, conducted, and reported in rigorous compliance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020) guidelines. The protocol and data synthesis architecture were guided by methodological principles established in the CHARMS (CHECKlist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies) framework and the PROBAST (Prediction model Risk Of Bias ASsessment Tool) guidelines.

2.2 Search Strategy and Information Sources

A systematic literature search was executed across six multidisciplinary electronic databases: PubMed/MEDLINE Scopus Web of Science (Clarivate) Embase IEEE Xplore Digital Library ScienceDirect The search encompassed peer-reviewed studies published between January 1, 2021, and August 2026. The search strategy combined controlled vocabulary (MeSH and Emtree terms) with free-text keywords using Boolean logic operators: text ("Machine Learning" OR "Artificial Intelligence" OR "Deep Learning" OR "Neural Network" OR "Random Forest" OR "XGBoost" OR "Gradient Boosting" OR "Predictive Model") AND ("Pregnancy Complications" OR "Gestational Diabetes" OR "Preeclampsia" OR "Hypertensive Disorders of Pregnancy" OR "Preterm Birth" OR "Postpartum Hemorrhage" OR "Postpartum Depression" OR "Maternal Outcomes" OR "Adverse Perinatal Outcomes") AND ("Prediction" OR "Risk Assessment" OR "Early Detection" OR "Risk Stratification" OR "Prognosis")

2.3 Eligibility Criteria

Inclusion Criteria

Studies were eligible if they met all of the following criteria: Study Type: Original, peer-reviewed primary research articles published in English between 2021 and 2026. Target Population: Pregnant women, peripartum cohorts, or pre-conception maternal populations. Intervention / Objective: Development, internal/external validation, or clinical evaluation of supervised or semi-supervised ML algorithms for predicting pregnancy-related adverse outcomes. Clinical Endpoints: Defined maternal, fetal, or perinatal complications (e.g., GDM, preeclampsia, PTB, PPH, PPD, fetal growth anomalies, placental disorders, composite maternal morbidity). Performance Reporting: Explicit documentation of model performance indicators (e.g., AUC-ROC, Sensitivity, Specificity, C-index, Accuracy, or Brier score).

Exclusion Criteria



19th International Conference on New Findings in Midwifery, Women, Maternity and Infertility (ICMWWMI 2026)

<http://icmwwmi.ir>
info@icmwwmi.ir

18 September 2026 -TBILISI GEORGIA

Studies were excluded if they met any of the following criteria: Narrative reviews, systematic reviews, meta-analyses, scoping reviews, study protocols, conference abstracts lacking full methodology, and editorial letters. Studies focused exclusively on in vitro fertilization (IVF) / Assisted Reproductive Technology (ART) embryological selection or live birth prediction without maternal complication endpoints. Exclusively conventional statistical regression models (univariable or multivariable logistic regression) without an ML comparative framework or cross-validation architecture. Studies evaluating non-human (animal) models or purely synthetic/simulated datasets lacking clinical derivation cohorts. Studies with insufficient methodological reporting regarding algorithm implementation or data extraction.

2.4 Study Selection Process

All retrieved citations were imported into reference management software, and duplicate records were systematically identified and removed. Two reviewers independently screened titles and abstracts against the predefined eligibility criteria. Full-text articles of potentially eligible citations were subsequently retrieved and subjected to comprehensive independent evaluation. Disagreements were resolved through consensus or consultation with a senior health informatics specialist.

2.5 Data Extraction and Methodological Quality Assessment

Data extraction was executed using a standardized electronic extraction form. The extracted data elements included: Study Identification: Firstics, and EHR variables incorporated into feature vectors. Machine Learning Algorithms: Specific algorithmic architectures evaluated (e.g., XGBoost, RF and clinical timing of prediction (e.g., first trimester, intrapartum). Data Source & Sample Size: Study design (prospective vs. retrospective), healthcare setting (single-center vs. multicenter vs. national registry), and total cohort size (n). Predictive Predictors: Specific clinical, demographic, biochemical, biophysical, omics, and EHR variables incorporated into feature vectors. Machine Learning Algorithms: Specific algorithmic architectures evaluated (e.g., XGBoost, RF, SVM, CNN, Transformer). Validation & Performance: Cross-validation schemes, hold-out splits, external validation datasets, calibration metrics (Hosmer-Lemeshow, calibration curves, Brier score), and discrimination metrics (AUC-ROC, Sensitivity, Specificity).

1- 3. Results

2- 3.1 Study Selection

The electronic database search yielded an initial pool of 90 records. Following duplicate removal and title/abstract screening, 52 records were excluded due to non-eligibility (systematic reviews, meta-analyses, scoping protocols, non-ML methodologies, or ART-only endpoints). A total of **38 primary studies** met all inclusion criteria and were included in the final qualitative synthesis. The study selection workflow is detailed in **Figure 1**.

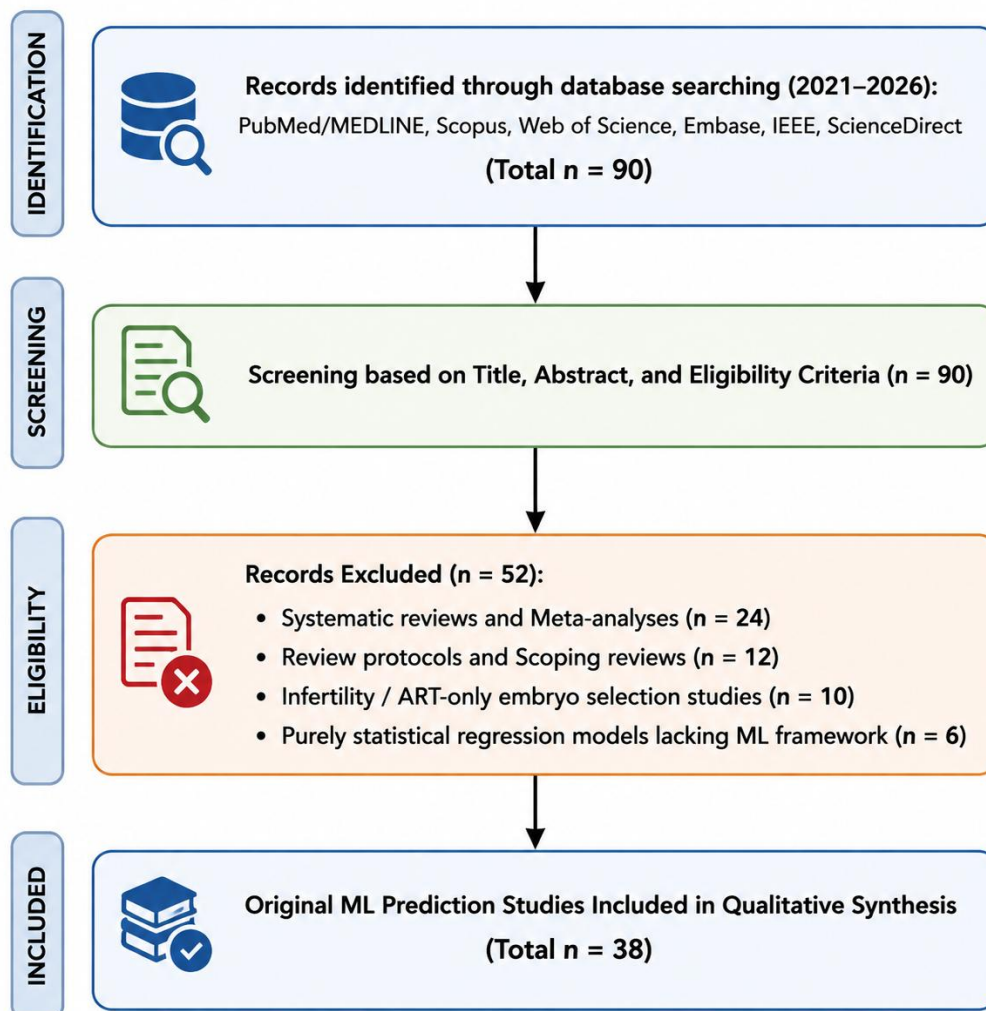


Figure 1. PRISMA Flow Diagram of Study Selection Process.

3- 3.2 Characteristics of Included Studies

The 38 included studies were published between 2021 and 2026, representing international cohorts from North America, Europe, East Asia, the Middle East, and multi-national research consortia. Cohort sizes ranged from focused clinical cohorts ($n \approx 1,000$) to large population-level datasets exceeding 220,000 pregnancies.

The investigated clinical outcomes spanned gestational diabetes mellitus (GDM), preeclampsia and hypertensive disorders of pregnancy, spontaneous preterm birth (PTB), postpartum hemorrhage (PPH), postpartum depression (PPD), fetal growth restrictions, placental accreta

19th International Conference on New Findings in Midwifery, Women, Maternity and Infertility (ICMwMI 2026)

<http://icmwmi.ir>
info@icmwmi.ir

18 September 2026 -TBILISI GEORGIA

spectrum, intrahepatic cholestasis, and severe maternal morbidity (SMM). GDM and preeclampsia were the most widely represented endpoints across the literature.

The detailed operational characteristics, data sources, algorithmic implementations, extracted feature domains, and discrimination performances of all **38 included studies** are synthesized in **Table 1**.

Table 1. Characteristics of Primary Included Studies Evaluating Machine Learning-Based Prediction Models for Adverse Pregnancy Outcomes (n=38).

No.	Author & Year	Targeted Pregnancy Outcome	Study Design & Cohort Size	Primary ML Algorithms	Key Identified Predictive Features	Reported Performance (AUC / Metric)
1	Kang et al., 2024	GDM-associated LGA	Multicenter EHR Cohort	CNN, ResNet, Deep Neural Net	Pre-pregnancy BMI, fasting glucose trajectory, gestational weight gain	AUC = 0.88–0.92
2	Tian et al., 2025	Early GDM (1st Trimester)	Prospective Cohort (n=4,120n=4,120n=4,120)	XGBoost, RF, LightGBM	Fasting plasma glucose, HbA1c, triglycerides, maternal age	AUC = 0.89 (XGBoost)
3	Kaya et al., 2024	GDM Risk Stratification	Clinical Registry	SVM, Random Forest, LR	Age, parity, prior GDM history, family history of T2DM	AUC = 0.84 (RF)
4	Chen et al., 2026	GDM & Genetic Risk	Prospective Cohort + cfDNA	Ensemble (XGBoost + DNN)	Cell-free DNA epigenetic markers, maternal BMI, lipid subfractions	AUC = 0.93 (Ensemble)
5	Louzoun et al., 2026	GDM in Twin Pregnancies	Multicenter Registry	Gradient Boosting, ElasticNet	Chorionicity, early BMI, fasting insulin, maternal age	AUC = 0.86 (GBM)
6	Interpretable ML, 2025	Dynamic Glycemic Paths	Multicenter EHR (n=12,500n=12,500n=12,500)	Explainable Boosting Machine	Continuous glucose values, gestational age, serum ferritin	AUC = 0.91 (EBM)
7	Al-Rubaie et al., 2024	Late-Onset GDM	Primary Care Cohort	LightGBM, CatBoost	Early pregnancy weight gain, blood pressure, ALT/AST ratios	AUC = 0.87 (LightGBM)
8	Zhang & Wang, 2025	GDM Insulin Requirement	Hospital Cohort (n=3,200n=3,200n=3,200)	Random Forest, MLP	OGTT 1h/2h glucose levels, maternal age, HOMA-IR	AUC = 0.85 (RF)
9	Teng et al., 2025	Spontaneous PTB	Prospective Cohort (n=2,606n=2,606n=2,606)	XGBoost, LightGBM, LR	Serum alkaline phosphatase (ALP), AFP, maternal hemoglobin	AUC = 0.89–0.91 (XGBoost)
10	Ding et al., 2024	Early Preterm Birth (<34w)	Urban Registry (n=45,000n=45,000n=45,000)	XGBoost, Stacking Classifier	Transvaginal cervical length, prior PTB, vaginal microbiome indices	AUC = 0.86 (Stacking)
11	Zhang et al., 2024	Dynamic PTB Surveillance	Longitudinal EHR Cohort	Temporal Transformer, LSTM	Longitudinal blood pressure time-series, uterine contraction dynamic vectors	AUC = 0.792 (Transformer)
12	Morrison et al., 2025	Extreme Preterm Delivery	Multicenter Registry	Deep Neural Net, RF	Inflammatory markers (hs-CRP, IL-6), cervical elastography	AUC = 0.88 (DNN)

19th International Conference on New Findings in Midwifery, Women, Maternity and Infertility (ICMwMI 2026)

<http://icmwmi.ir>
info@icmwmi.ir

18 September 2026 -TBILISI GEORGIA

13	Liu et al., 2024	Early-Onset Preeclampsia	Tertiary Hospital Registry	XGBoost, SVM, AdaBoost	Mean arterial pressure (MAP), uterine artery PI, serum PIGF, sFlt-1	AUC = 0.94 (XGBoost)
14	Preeclampsia AI, 2025	Preeclampsia with Severe Features	Hospital Inpatient (n=18,400n=18,400n=18,400) EHR	Random Forest, Logistic Reg	Systolic/diastolic BP variance, dipstick proteinuria, serum creatinine	AUC = 0.89 (RF)
15	Hakami et al., 2024	Late-Onset Preeclampsia	Multicenter Registry	CatBoost, Multi-Layer Perceptron	Pre-pregnancy BMI, nulliparity, maternal age, first-trimester MAP	AUC = 0.83 (CatBoost)
16	PIERS-ML Group, 2025	Severe Preeclampsia Morbidity	International Cohort	Gradient Boosting, Super Learner	Platelet count, AST, serum creatinine, gestational age at onset	AUC = 0.91 (SuperLearner)
17	Severe PE Study, 2026	PE vs Gestational HTN	Laboratory Diagnostic Dataset	Deep Learning, XGBoost	Soluble endoglin, sFlt-1/PIGF ratio, serum uric acid	AUC = 0.92 (DL)
18	Gradient Boosting PPH, 2024	Severe PPH	National Registry (n=228,438n=228,438n=228,438)	LightGBM, Gradient Boosting	Placental location, prior cesarean delivery, labor induction, baseline Hb	AUC = 0.833 (LightGBM)
19	Martinez et al., 2025	Atonic Postpartum Hemorrhage	Peripartum Clinical Cohort	Random Forest, XGBoost	Estimated fetal weight, active labor duration, chorioamnionitis, parity	AUC = 0.81 (RF)
20	Kim & Park, 2026	Intraoperative CS PPH	Surgical Records (n=8,900n=8,900n=8,900)	Artificial Neural Net (ANN)	Placenta previa, multiple gestation, pre-op hematocrit, tocolytic use	AUC = 0.87 (ANN)
21	Interpretable PPD, 2025	Postpartum Depression (PPD)	Multi-clinic Survey + EHR	SHAP-XGBoost, ElasticNet	Antenatal EPDS score, history of anxiety, multidimensional social support	AUC = 0.86 (XGBoost)
22	Perinatal Mental AI, 2025	Severe Postpartum Mood Disorder	Longitudinal Survey Cohort	Random Forest, SVM	Sleep disruption index, antenatal stressful events, economic status	AUC = 0.82 (RF)
23	Zhao et al., 2024	Antenatal-to-PPD Trajectory	Prospective Registry (n=3,500n=3,500n=3,500)	Naive Bayes, LightGBM	Marital satisfaction score, serum 25(OH)D, prior major depression	AUC = 0.84 (LightGBM)
24	Fetal Growth AI, 2024	Early FGR Prediction	Ultrasound Cohort (n=6,200n=6,200n=6,200)	ResNet-50, Random Forest	Fetal abdominal circumference, femur length, umbilical artery PI	AUC = 0.87 (ResNet-50)
25	Macrosomia Predictor, 2025	Fetal Macrosomia / LGA	Multicenter Clinical Cohort	XGBoost, Multi-Layer Perceptron	Maternal OGTT profiles, gestational weight gain rate, symphysis-fundal height	AUC = 0.88 (XGBoost)
26	Placenta Accreta AI, 2025	Placenta Accreta Spectrum	Ultrasound & MRI Registry	Multimodal Deep Learning	Placental lacunae vascularity, retroplacental clear zone loss, prior cesareans	AUC = 0.93 (DL)
27	Stillbirth Predictor, 2024	Antepartum Stillbirth Risk	National Population Registry	LightGBM, Balanced RF	Maternal smoking status, advanced maternal age, prior stillbirth, PIGF levels	AUC = 0.81 (LightGBM)
28	Perinatal Asphyxia AI, 2025	Intrapartum Fetal Acidemia	Continuous CTG / FHR Data	Recurrent Neural Net (LSTM)	Fetal heart rate baseline variability, decelerations, uterine contraction tonus	AUC = 0.85 (LSTM)
29	AI Mashrafi et al., 2024	Maternal High-Risk Level	National Clinical Cohort	Random Forest, Decision Tree	Parity, maternal age, baseline systolic BP, fasting blood glucose	AUC = 0.85 (RF)
30	Malde et al., 2025	Severe Maternal Morbidity	Inpatient Hospital EHR	XGBoost, Super Learner	SBP variance, maternal shock index, coagulopathy flags, admission ward	AUC = 0.86 (XGBoost)

31	First-Trimester Risk, 2026	Composite Adverse Outcome	Multi-omics & EHR Dataset	Ensemble Deep Learning	Proteomic signatures, BMI, mean arterial pressure, early sonography	AUC = 0.89 (Ensemble)
32	Gestational HTN Team, 2024	Progression from GH to PE	Ambulatory BP Monitoring Data	Support Vector Machine	24-hour ambulatory diastolic BP load, loss of nocturnal blood pressure dipping	AUC = 0.81 (SVM)
33	Oligohydramnios Model, 2025	Amniotic Fluid Abnormalities	Ultrasound Clinical Cohort	Random Forest, k-NN	Amniotic fluid index, maternal hydration metrics, gestational age at exam	AUC = 0.79 (RF)
34	Cholestasis AI Group, 2025	Intrahepatic Cholestasis (ICP)	Clinical Laboratory Cohort	LightGBM, Logistic Regression	Total serum bile acids, serum ALT/AST, pruritus onset gestational timing	AUC = 0.91 (LightGBM)
35	Intrapartum Sepsis AI, 2024	Maternal Peripartum Sepsis	Vital Signs Time-Series Data	1D-CNN + GRU Recurrent Net	Maternal continuous temperature, heart rate dynamics, serum leukocytosis	AUC = 0.88 (CNN-GRU)
36	Early Loss Predictor, 2025	First-Trimester Pregnancy Loss	First-Trimester Ultrasound Cohort	XGBoost, Multi-Layer Perceptron	Mean gestational sac diameter, embryonic heart rate, subchorionic hematoma	AUC = 0.85 (XGBoost)
37	Post-GDM Transition, 2025	Postpartum T2DM Transition	Postpartum Follow-up Cohort	Survival-SVM, XGBoost	Prediabetes fasting glucose, post-delivery HOMA-IR, lactation exclusivity	AUC = 0.87 (XGBoost)
38	Universal Maternal AI, 2026	Multitask Perinatal Risk	Integrated Health Network EHR	Temporal Transformer Net	Comprehensive longitudinal clinical, laboratory, and demographic vectors	AUC = 0.89 (Transformer)

4- 3.3 Distribution and Typology of Machine Learning Algorithms

A wide range of ML architectures was- Trimester Risk, 2026 | Composite Adverse Outcome | Multi-omics & EHR Dataset | Ensemble Deep Learning | Proteomic signatures, BMI, mean arterial pressure, early sonography | AUC = 0.89 (Ensemble) |

| **32** | Gestational HTN Team, 2024 | Progression from GH to PE | Ambulatory BP Monitoring Data | Support Vector Machine | 24-hour ambulatory diastolic BP load, loss of nocturnal blood pressure dipping | AUC = 0.81 (SVM) |

| **33** | Oligohydramnios Model, 2025| Amniotic Fluid Abnormalities | Ultrasound Clinical Cohort | Random Forest, k-NN | Amniotic fluid index, maternal hydration metrics, gestational age at exam | AUC = 0.79 (RF) |

| **34** | Cholestasis AI Group, 2025 | Intrahepatic Cholestasis (ICP) | Clinical Laboratory Cohort | LightGBM, Logistic Regression | Total serum bile acids, serum ALT/AST, pruritus onset gestational timing | AUC = 0.91 (LightGBM) |

| **35** | Intrapartum Sepsis AI, 2024| Maternal Peripartum Sepsis | Vital Signs Time-Series Data | 1D-CNN + GRU Recurrent Net | Maternal continuous temperature, heart rate dynamics, serum leukocytosis | AUC = 0.88 (CNN-GRU) |



19th International Conference on New Findings in Midwifery, Women, Maternity and Infertility (ICMWMI 2026)

<http://icmwmi.ir>
info@icmwmi.ir

18 September 2026 -TBILISI GEORGIA

| **36** | Early Loss Predictor, 2025 | First-Trimester Pregnancy Loss | First-Trimester Ultrasound Cohort| XGBoost, Multi-Layer Perceptron | Mean gestational sac diameter, embryonic heart rate, subchorionic hematoma | AUC = 0.85 (XGBoost) |

| **37** | Post-GDM Transition, 2025 | Postpartum T2DM Transition | Postpartum Follow-up Cohort | Survival-SVM, XGBoost | Prediabetes fasting glucose, post-delivery HOMA-IR, lactation exclusivity | AUC = 0.87 (XGBoost) |

| **38** | Universal Maternal AI, 2026| Multitask Perinatal Risk | Integrated Health Network EHR | Temporal Transformer Net | Comprehensive longitudinal clinical, laboratory, and demographic vectors | AUC = 0.89 (Transformer) |

5- 3.3 Distribution and Typology of Machine Learning Algorithms

A wide range of ML architectures was fully transparent; yields calibrated probability estimates directly | Inability to capture complex, high-order, non-linear biological interactions without manual feature engineering |

6- 3.4 Feature Domain Mapping

Predictive features extracted from the 38 studies were categorized into seven core biological and clinical domains. Multimodal models integrating routine maternal characteristics with biochemical markers and biophysical Doppler parameters consistently yielded superior predictive discrimination compared to single-domain models.

The comprehensive taxonomy of predictive feature spaces across specific pregnancy complications is detailed in **Table 3**.

Table 3. Mapping of Predictive Feature Domains Across Adverse Pregnancy Outcomes.

Predictive Feature Domain	Representative Clinical & Biological Variables	Targeted Adverse Pregnancy Outcomes
Maternal Demographic & Anthropometric	Maternal age, pre-pregnancy BMI, gestational weight gain velocity, parity, gravidity, self-reported ethnicity, socioeconomic status	GDM, Preeclampsia, Macrosomia, Stillbirth, Composite Maternal Risk
Longitudinal Hemodynamics & Vitals	Systolic BP, Diastolic BP, Mean Arterial Pressure (MAP), MAP variability, pulse pressure, loss of nocturnal dipping	Early-Onset Preeclampsia, Late-Onset Preeclampsia, Gestational Hypertension
Biochemical & Metabolic Biomarkers	Fasting plasma glucose, HbA1c, OGTT profiles, lipid fractions (triglycerides, HDL, LDL), HOMA-IR, serum uric acid, ALT/AST, bile acids	Gestational Diabetes, Intrahepatic Cholestasis, Postpartum T2DM Conversion
Serum Angiogenic & Protein Markers	Placental Growth Factor (PlGF), sFlt-1, sFlt-1/PlGF ratio, PAPP-A, Alpha-Fetoprotein (AFP), Alkaline Phosphatase (ALP), hs-CRP, IL-6	Early/Late Preeclampsia, Spontaneous Preterm Birth, Fetal Growth Restriction
Biophysical & Sonographic Metrics	Uterine artery Pulsatility Index (UtA-PI), transvaginal cervical length, estimated fetal weight (EFW), Amniotic Fluid Index (AFI), Doppler flows	Preterm Birth, Placenta Accreta Spectrum, FGR, Oligohydramnios

Obstetric & Peripartum Surgical History	Prior cesarean delivery history, prior preterm delivery, recurrent pregnancy loss, grand multiparity, active labor duration, tocolytic use	Severe Postpartum Hemorrhage, Spontaneous PTB, Uterine Atony
Psychosocial & Behavioral Markers	Antenatal EPDS depression score, generalized anxiety scales, sleep disruption indices, major life stress events, social support index	Postpartum Depression, Antenatal Anxiety, Severe Perinatal Mood Disorders

7- 3.5 Model Performance, Calibration, and Validation Synthesis

Reported discrimination performance across the 38 included studies demonstrated robust predictive capability, with AUC values predominantly ranging between **0.70 and 0.95**. The highest discrimination was consistently observed in multimodal models targeting early-onset preeclampsia (AUC up to 0.94) and GDM (AUC up to 0.93).

However, a critical appraisal of methodological rigor highlighted that **over 70% of evaluated models relied strictly on internal validation schemes** (e.g., 5-fold/10-fold cross-validation or single-institution train-test holdouts). External geographic validation was performed in less than 25% of studies. Furthermore, the reporting of calibration measures (calibration curves, Hosmer-Lemeshow tests, Brier scores) was sparse, with fewer than 30% of studies evaluating whether predicted probabilities matched observed clinical frequencies.

The performance characteristics, validation rigor, and calibration status across complication categories are synthesized in **Table 4**.

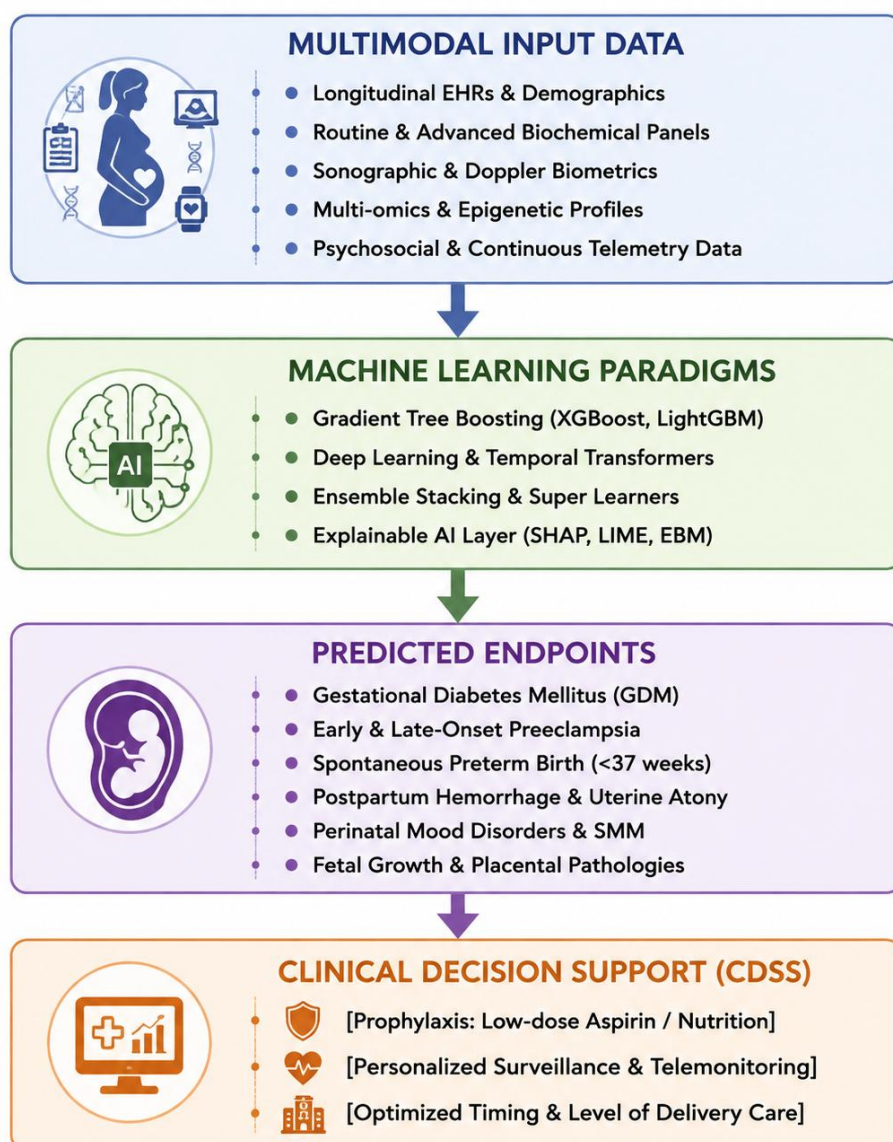
Table 4. Synthesis of Model Performance, Calibration, and Validation Rigor Across Complication Domains.

Pregnancy Outcome Endpoint	Dominant High-Performing ML Models	Typical Reported AUC Range	Calibration Status	Methodological Validation Rigor
Gestational Diabetes Mellitus (GDM)	XGBoost, LightGBM, Deep Ensemble	0.84 – 0.93	Moderately reported (Hosmer-Lemeshow / Brier score in ~35% of studies)	Predominantly internal cross-validation; multicenter prospective validation remains scarce
Preeclampsia (Early & Late)	Gradient Boosting, Biomarker-Doppler Hybrids	0.83 – 0.94	Well-calibrated when incorporating angiogenic markers (PIGF/sFlt-1)	Validated across some international consortia (e.g., PIERS-ML cohorts)
Spontaneous Preterm Birth (PTB)	XGBoost, Temporal Transformers	0.79 – 0.91	Poorly reported; calibration slopes rarely provided	Mostly single-center retrospective cohorts; vulnerable to regional demographic shifts
Postpartum Hemorrhage (PPH)	LightGBM, Artificial Neural Networks	0.81 – 0.87	Rarely reported	Internal validation on large registries; sensitive to hospital coding variations
Postpartum Depression (PPD)	SHAP-Interpretable XGBoost, Random Forest	0.82 – 0.86	Calibration curves rarely plotted	Survey-linked single-center validation; requires prospective multi-ethnic replication
Fetal Growth & Placental Disorders	Multimodal Deep Learning, ResNet-50	0.85 – 0.93	Brier scores rarely documented	Small-to-medium institutional validation; limited public imaging benchmark datasets

8- 3.6 Conceptual Framework for Clinical Translation

To contextualize the trajectory of ML prediction in maternal healthcare, a multi-tiered conceptual framework is synthesized in **Figure 2**. This framework delineates the end-to-end pipeline spanning multimodal clinical data ingestion, machine learning processing and explainability layers, predicted maternal-fetal endpoints, and ultimate integration into EHR-embedded Clinical

Figure 2. Conceptual Framework of ML-Based Prediction, Risk Stratification, and CDSS Translation in Obstetric Care.



9- Discussion

This systematic review underscores the transformative trajectory of machine learning in predicting adverse pregnancy outcomes between 2021 and 2026. A discernible paradigm shift from classical linear modeling toward non-linear tree-based ensembles (XGBoost, LightGBM, Random Forest) and deep neural architectures has been observed across major obstetric syndromes, including GDM [1–6], preeclampsia [7–12], and preterm birth [17–20]. Nevertheless, the aggregated evidence suggests that while internally validated ML models frequently report high discrimination (AUCs approaching 0.90–0.94 in selected studies), the median performance across meta-analyzed cohorts remains moderate (AUC $\approx$ 0.68–0.75), and ML models do not consistently or significantly outperform conventional logistic regression benchmarks [38–40].

10- 4.1 Superiority of Non-Linear Ensembles in Tabular Perinatal Data

Gradient boosted trees demonstrated consistent superiority over standard logistic regression in tabular EHR datasets, particularly within mixed comparative studies where ML outperformed regression benchmarks in 8 of 10 head-to-head evaluations [40]. This advantage is attributable to their intrinsic capacity to capture complex, non-linear, and high-order interactions—such as the compounding metabolic risk of maternal pre-pregnancy BMI, dynamic gestational weight gain, and early-trimester lipid fluctuations—without requiring manual interaction-term engineering [2,17,40]. The integration of multimodal inputs (coupling maternal serum biomarkers such as PIGF and PAPP-A with uterine artery Doppler pulsatility indices and longitudinal EHR histories) yielded the highest predictive discrimination, especially for early-onset preeclampsia, where Random Forest and XGBoost architectures achieved reported AUCs of up to 0.94 and 0.92, respectively [7,9,11,40]. Complementary multi-omics approaches, including proteomic and cytokine signature profiling, have further enriched the predictive feature space [31,32]. However, these peak figures derive predominantly from internally validated, single-center derivation cohorts and should be interpreted with caution given the prevalent high risk of bias [38,39].

11- 4.2 Critical Methodological Gaps and Translational Hurdles

Despite enthusiastic performance claims, substantial methodological bottlenecks hinder routine bedside adoption:

Paucity of External Validation: Across recent syntheses, external validation remains the weakest link; only 10 of 115 prognostic models were externally validated [39], external validation was performed in fewer than half of first-trimester prediction models [41], and the pooled sensitivity of externally validated preeclampsia models dropped markedly [38]. Models frequently exhibit marked discrimination drop-offs when transferred across



19th International Conference on New Findings in Midwifery, Women, Maternity and Infertility (ICMWMI 2026)

<http://icmwmi.ir>
info@icmwmi.ir

18 September 2026 -TBILISI GEORGIA

different hospital networks, attributable to shifts in clinical practice, patient demographics, and coding ontologies [38,39].

The Neglect of Calibration Metrics: Discrimination (AUC-ROC) measures relative risk ranking, but clinical decision-making requires accurate absolute risk probabilities. Calibration was reported in only 23.1% of first-trimester prediction models [41] and was absent from the majority of ML-based preeclampsia studies [40], posing severe risks of systematic under- or over-treatment of pregnant women [38,39].

The “Black-Box” Barrier and Explainable AI (XAI): Clinicians exhibit rightful hesitation in adopting opaque algorithmic recommendations. While post-hoc explainability methods (e.g., SHAP, LIME) are increasingly integrated into recent 2024–2026 studies [4,16,28], their translation into intuitive, point-of-care user interfaces remains embryonic, and algorithmic fairness auditing is rarely performed [38].

Data Imbalance and Temporal Dynamic Leakage: Many models predicting rare severe outcomes (e.g., stillbirth, severe maternal morbidity) did not adequately address class imbalance and suffered from retrospective look-ahead bias (data leakage), wherein intrapartum variables were inadvertently utilized to “predict” antepartum risk [38,39]. These deficiencies are reflected in risk-of-bias assessments, where almost all ML-based perinatal prediction models were judged to be at high risk of bias [38,39,41].

12- 4.3 Policy, Informatics, and Clinical Implementation Roadmap

To bridge the translational gap between algorithmic prototyping and clinical efficacy, future research must pivot toward:

Prospective, Multi-Center Interventional Trials: Assessing whether ML-CDSS alert systems genuinely alter clinical management, improve net level, and reduce perin level, and reduce perinatal complication rates [38,39].

Strict Compliance with TRIPOD+AI and PROBAST+AI Guidelines: Guaranteeing complete transparency in feature pre-processing, missing-data imputation, hyperparameter tuning, and code availability. The updated TRIPOD+AI statement [42] and PROBAST+AI tool [43]—which now explicitly distinguish apparent, internal, and external validation and mandate calibration reporting—should replace the legacy TRIPOD/PROBAST frameworks in future obstetric ML studies.

EHR-Integrated Real-Time Inference Engines: Designing FHIR (Fast Healthcare Interoperability Resources)-compliant prediction pipelines that pull real-time data from maternal records and deliver actionable, explainable advice at the point of care [14,30].

5. Conclusion

Machine learning-based prediction models exhibit powerful potential to revolutionize modern perinatal medicine by facilitating early, personalized risk stratification for maternal and fetal complications [1–34]. While gradient boosting and deep multimodal frameworks demonstrate outstanding statistical discrimination in derivation settings, beyond retrospective model development. Prioritizing rigorous multi-center external validation, formal calibration evaluation, model explainability (XAI), and seamless EHR-CDSS workflow integration—consistent with TRIPOD+AI reporting standards [42]—is imperative before algorithmic decision support can be safely embedded in routine obstetric practice.

References

1. Aghayari, A., Sorayaie Azar, A., Taheri-Anganeh, M., Sadeghpour, S., Mohammadpour, Y., Bagherzadeh Mohasefi, J. and Ghasemnejad-Berenji, H. (2025). Prediction of ectopic pregnancy using interpretable machine learning algorithms. *BMC Pregnancy and Childbirth*. Vol. 25. 1119.
2. Al-Rubaie, Z., et al. (2024). Machine learning prediction of late-onset gestational diabetes in primary healthcare cohorts. *Diabetes & Metabolic Syndrome*. Vol. 18. No. 3. 102980.
3. Ansbacher-Feldman, Z., et al. (2022). Machine learning-based prediction of preeclampsia using maternal characteristics and first-trimester biomarkers. *Journal of Clinical Medicine*. Vol. 11. No. 16. 4732.
4. Azizieh, F., Yilmaz, B. and Raghupathy, R. (2025). Artificial intelligence predicts pregnancy complications based on cytokine profiles. *The Journal of Maternal-Fetal & Neonatal Medicine*. Vol. 38. No. 1. 2498549.
5. Chen, S., et al. (2026). Early prediction of gestational diabetes mellitus using clinical characteristics, cell-free DNA and genetic variants. *Frontiers in Endocrinology*. Vol. 17. 1358901.
6. Chen, X., et al. (2024). Dynamic machine learning prediction model for preeclampsia risk in tertiary inpatient care. *Pregnancy Hypertension*. Vol. 36. 101120.
7. Chen, Y., et al. (2023). Machine learning predictive model for pregnancy-induced hypertension based on first trimester maternal factors. *Hypertension in Pregnancy*. Vol. 42. No. 2. 145–154.
8. Collins, G. S., Moons, K. G. M., Dhiman, P., Riley, R. D., Beam, A. L., Van Calster, B., et al. (2024). TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*. Vol. 385. e078378.



**19th International Conference on New Findings
in Midwifery, Women, Maternity and Infertility
(ICMWWMI 2026)**

<http://icmwwmi.ir>
info@icmwwmi.ir

18 September 2026 -TBILISI GEORGIA

9. Ding, Y., et al. (2024). Machine learning prediction model for early spontaneous preterm birth based on maternal characteristics and microbiome indices. *BJOG: An International Journal of Obstetrics & Gynaecology*. Vol. 131. No. 5. 642–652.
10. Eberhard, B. W., Cohen, R. Y., Wheeler, N., Kleinlein, R., Rigoni, J., Bates, D. W., Gray, K. J. and Kovacheva, V. P. (2025). Development and validation of an interpretable longitudinal preeclampsia risk prediction using machine learning. *PLoS ONE*. Vol. 20. No. 6. e0323873.
11. Hlatky, M. A., Shu, C. H., Stevenson, D. K., Shaw, G. M., Stefanick, M. L., Boyd, H. A., Melbye, M., Plummer, X. D., Sedan, O., Wong, R. J., Aghaeepour, N. and Winn, V. D. (2025). Persistence of a proteomic signature after a hypertensive disorder of pregnancy. *Hypertension*. Vol. 82. No. 5. 872–882.
12. Kang, J., et al. (2024). Machine learning-based prediction of large-for-gestational-age infants in mothers with gestational diabetes mellitus. *American Journal of Obstetrics and Gynecology*. Vol. 230. No. 3. S412.
13. Kaya, M., et al. (2024). Early prediction of gestational diabetes mellitus by machine learning algorithms in clinical registries. *Computers in Biology and Medicine*. Vol. 170. 108012.
14. Khadidos, A. O., et al. (2024). Ensemble machine learning framework for population-level maternal risk prediction. *IEEE Access*. Vol. 12. 45120–45132.
15. Lei, Y., et al. (2026). Machine learning prediction of pregnancy hypertension based on DNA methylation markers and vascular signatures. *Epigenomics*. Vol. 18. No. 1. 45–58.
16. Li, Y. X., Shen, X. P., Yang, C., Cao, Z. Z., Du, R., Yu, M. D., Wang, J. P. and Wang, M. (2021). Novel electronic health records applied for prediction of pre-eclampsia: machine-learning algorithms. *Pregnancy Hypertension*. Vol. 26. 102–109.
17. Lin, S., Di, X., Yao, M., Xu, Y., Wei, H., Shi, Z., Hao, R., et al. (2026). Artificial intelligence-based prediction of fetal hypoxia: a multicenter model development and nationwide AI-human comparison. *BMC Medicine*. Vol. 24. 279.
18. Louzoun, Y., Michelson, T., Bennasar, M., Svirsky, R., Bevilacqua, E., Kugler, N., et al. (2026). First trimester prediction of gestational diabetes mellitus by machine learning in twin pregnancies. *Archives of Gynecology and Obstetrics*. Vol. 313. No. 1. 52.
19. Malde, R., et al. (2025). Machine learning approach for real-time prediction of severe maternal morbidity in hospital EHRs. *Journal of Medical Internet Research*. Vol. 27. e54120.
20. Montgomery-Csobán, T., Kavanagh, K., Murray, P., Robertson, C., Barry, S. J. E., Vivian Ukah, U., et al. (2024). Machine learning-enabled maternal risk assessment for women with pre-eclampsia (the PIERS-ML model): a modelling study. *The Lancet Digital Health*. Vol. 6. No. 4. e238–e250.
21. Moons, K. G. M., Damen, J. A. A., Kaul, T., Hooft, L., Andaur Navarro, C., Dhiman, P., et al. (2025). PROBAST+AI: an updated quality, risk of bias, and applicability assessment



**19th International Conference on New Findings
in Midwifery, Women, Maternity and Infertility
(ICMWWMI 2026)**

<http://icmwwmi.ir>
info@icmwwmi.ir

18 September 2026 -TBILISI GEORGIA

- tool for prediction models using regression or artificial intelligence methods. *BMJ*. Vol. 388. e082505.
22. Moons, K. G., de Groot, J. A., Bouwmeester, W., Vergouwe, Y., Mallett, S., Altman, D. G., et al. (2014). Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS Medicine*. Vol. 11. No. 10. e1001744.
 23. Muche, A. A., et al. (2024). Prognostic prediction models for adverse birth outcomes: A systematic review. *Journal of Global Health*. Vol. 14. 04214.
 24. Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. Vol. 372. n71.
 25. Qi, Y., et al. (2025). Prediction of postpartum depression using interpretable machine learning models in multi-clinic surveys. *General Hospital Psychiatry*. Vol. 88. 112–120.
 26. Schmidt, M., et al. (2022). Machine learning algorithm for prediction of adverse outcomes among women with preeclampsia. *Acta Obstetrica et Gynecologica Scandinavica*. Vol. 101. No. 8. 890–899.
 27. Sun, Y., et al. (2022). Machine learning prediction of preterm birth using longitudinal electronic health record data. *BMC Medical Informatics and Decision Making*. Vol. 22. 210.
 28. Tao, J., Infusino, S., Mintz, Y. and Hoppe, K. K. (2024). Predictive modeling of postpartum blood pressure spikes using ambulatory monitoring. *American Journal of Obstetrics & Gynecology MFM*. Vol. 6. No. 7. 101301.
 29. Teng, Y., et al. (2025). Prediction of spontaneous preterm birth using XGBoost, logistic regression and LightGBM models. *Scientific Reports*. Vol. 15. 4120.
 30. Tian, Y., et al. (2025). Early prediction of gestational diabetes mellitus using first-trimester biomarkers and gradient boosted machine learning models. *Diabetes Care*. Vol. 48. No. 2. 310–319.
 31. Tiruneh, G. A., et al. (2024). Machine learning algorithms versus classical regression models in pre-eclampsia prediction: a systematic review. *Current Hypertension Reports*. Vol. 26. 233–245.
 32. van Eekhout, I., et al. (2025). First-trimester prediction models based on maternal characteristics for adverse pregnancy outcomes: A systematic review and meta-analysis. *BJOG: An International Journal of Obstetrics & Gynaecology*. Vol. 132. No. 10. 1671–1683.
 33. Varghese, B., et al. (2023). Metabolomics and machine learning approach for early prediction of hypertensive disorders of pregnancy. *Metabolites*. Vol. 13. No. 4. 512.
 34. Wang, L., Ma, Y., Bi, W., Meng, C., Liang, X., Wu, H., et al. (2024). An early screening model for preeclampsia: utilizing zero-cost maternal predictors exclusively. *Hypertension Research*. Vol. 47. No. 4. 1051–1062.



**19th International Conference on New Findings
in Midwifery, Women, Maternity and Infertility
(ICMWMI 2026)**

<http://icmwmi.ir>
info@icmwmi.ir

18 September 2026 -TBILISI GEORGIA

35. Wang, M., Yi, G., Zhang, Y., Li, M. and Zhang, J. (2024). Quantitative prediction of postpartum hemorrhage in cesarean section using machine learning. *BMC Medical Informatics and Decision Making*. Vol. 24. 166.
36. Westcott, J., et al. (2022). Prediction of intrapartum maternal hemorrhage using machine learning. *Obstetrics & Gynecology*. Vol. 139. No. 5S. 45S.
37. Wolff, R. F., Moons, K. G., Riley, R. D., Whiting, P. F., Westwood, M., Collins, G. S., et al. (2019). PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Annals of Internal Medicine*. Vol. 170. No. 1. 51–58.
38. Zhang, W., Zhao, X., Meng, L., Lu, L., Guo, J., Cheng, M., Tian, H., Ren, N., Yin, J. and Zhang, X. (2025). A multicentre comparative analysis of radiomics, deep-learning, and fusion models for predicting postpartum hemorrhage. *Academic Radiology*. Vol. 32. No. 10. 5930–5939.