Systematic Review
DOI: https://doi.org/10.70749/ijbr.v3i5.1294

INDUS JOURNAL OF BIOSCIENCE RESEARCH
https://induspublishers.com/IJBR
ISSN: 2960-2793/ 2960-2807

Crossref

OPEN ACCESS

Artificial Intelligence in Predicting Pregnancy Complications: A Systematic Review and Meta-Analysis of Preeclampsia and Gestational Diabetes Mellitus

Salma Malik[1], Sarwat Ishaq[2], Hira Farzooq[1], Habib Ullah Joya[3]

[1]Al-Hammadi Hospital, Al Suwaidi, Riyadh, Saudi Arabia.
[2]Sulaiman AL Habib Medical Group, Al Suwaidi, Riyadh, Saudi Arabia.
[3]Hamad Medical Cooperation, Doha, Qatar.

ARTICLE INFO

Keywords: Artificial Intelligence (AI), Machine Learning, Preeclampsia, Gestational Diabetes Mellitus (GDM), Pregnancy Complications.

Correspondence to: Salma Malik,
Al-Hammadi Hospital, Al Suwaidi, Riyadh, Saudi Arabia.
Email: drsalmamalik80@gmail.com
https://orcid.org/0009-0001-6748-9985

Declaration
Authors' Contribution: All authors equally contributed to the study and approved the final manuscript.*Detail is given at the end.
Conflict of Interest: No conflict of interest.
Funding: No funding received by the authors.

Article History
Received: 16-01-2025 Revised: 19-04-2025
Accepted: 02-05-2025 Published: 15-05-2025

ABSTRACT

This systematic review and meta-analysis evaluates the performance of artificial intelligence (AI) models in predicting two major pregnancy complications: preeclampsia and gestational diabetes mellitus (GDM). Adhering to PRISMA guidelines, we analyzed 13 studies from PubMed, Scopus, Web of Science, and IEEE Xplore, selected from an initial pool of 2,163 articles. Using R software (version 4.3.1), we conducted a random-effects meta-analysis, assessing metrics such as the area under the curve (AUC), sensitivity, specificity, and accuracy. The study demonstrated strong predictive performance for preeclampsia and gestational diabetes mellitus (GDM) using artificial intelligence (AI) models. For preeclampsia prediction, the training area under the curve (AUC) was 0.878, while the test AUC was 0.861. Similarly, for GDM, the training AUC was 0.779, and the test AUC was 0.800, indicating high discriminative ability. Tree-based and neural network models outperformed other approaches, particularly when incorporating multimodal data—such as clinical and biochemical data or electronic health records (EHR). Sensitivity analysis further supported these findings, even after excluding high-risk studies identified by the PROBAST tool. While AI models show promise for antenatal risk screening, challenges remain, including limited external validation and interpretability. Future research should focus on improving model transparency, ensuring diverse ethnic representation, and facilitating seamless integration into clinical practice. These steps are critical to harnessing AI's potential for enhancing maternal and fetal health outcomes.

INTRODUCTION

Pregnancy complications like preeclampsia and gestational diabetes mellitus (GDM) pose significant risks to maternal and fetal health, contributing to global morbidity and mortality [1,2]. Preeclampsia, characterized by hypertension and signs of organ dysfunction after 20 weeks of pregnancy, affects 2–8% of all pregnant women and is a leading cause of maternal mortality, particularly in low- and middle-income countries (LMICs) [3,4]. While, GDM (gestational diabetes mellitus)—a condition diagnosed for the first time during pregnancy its prevalence ranges from 7% to 14%, depending on the population and diagnostic criteria [5,6]. Both conditions are also associated with an increased risk of cesarean delivery, neonatal complications, and long-term cardiovascular and metabolic disorders in offspring, as well as future maternal risks to cardiovascular and metabolic health [7,8,9].

These findings highlight the importance of assessing pregnant individuals for potential risks of these complications to enable appropriate clinical management. Current research indicates that conventional risk assessment models - which rely on demographic, clinical, and biochemical factors - have significant limitations. These models employ static, linear approaches that fail to capture complex interactions between variables or account for multifactorial influences [10,11,12]. Modern healthcare generates vast amounts of data, including electronic health records, laboratory tests, genomic analyses, and wearable device outputs. This wealth of information requires more sophisticated analytical approaches to realize its full potential.

Artificial intelligence (AI), including its subcategories of machine learning and deep learning, has emerged as a transformative technology in medicine.

These approaches enhance disease prognosis and diagnosis while improving treatment prediction and therapeutic effectiveness [13,14]. AI models have shown significant potential in obstetrics for predicting various outcomes, including preterm birth, fetal growth restriction, and delivery [14,15,16]. These models utilize large datasets without requiring hardcoded parameters, enabling them to process high-dimensional data and identify non-linear relationships that often escape detection by conventional statistical methods [17].

Several recent studies have applied multiple AI models—including support vector machines, decision trees, random forests, gradient boosting machines, and neural networks—to predict preeclampsia and gestational diabetes mellitus (GDM) using early pregnancy or even preconception data [18,19,20]. These studies enhanced model performance by combining clinical data, biomarkers, laboratory results, and novel genomic/imaging data. Koivu et al. [21], achieved excellent GDM screening (AUC >0.85) using machine learning on EHR data, while Akbulut et al. [22], successfully applied deep learning to preeclampsia biomarker detection.

While AI applications in maternal-fetal medicine show considerable promise, significant heterogeneity persists in model development approaches, data sources, validation methods, and reporting standards. Most existing models rely on historical datasets and lack robust external validation, with few studies demonstrating real-time clinical validation [23,24]. Additional challenges around interpretability and bias (particularly across ethnic and socioeconomic groups) limit EHR-AI's clinical adoption [25,26].

Given these considerations and future potential, this systematic review will analyze RCTs comparing AI models for preeclampsia and GDM prediction. We will evaluate algorithm types, input data, validation methods, and accuracy metrics. By synthesizing developments from the past decade, this review aims to guide future AI development for prenatal care.

MATERIALS AND METHODS

Study Design

Following PRISMA 2020 guidelines, this systematic review identified and analyzed published studies using AI to predict preeclampsia and gestational diabetes mellitus (GDM). The protocol was prospectively registered in PROSPERO (CRD42025112245) prior to data collection, to prevent duplication and ensure methodological transparency.

Selection Criteria

We included original English-language studies published from January 2012 to March 2025 that investigated preeclampsia or GDM as primary outcomes using machine learning (ML), deep learning (DL), or other AI-based predictive models. Following systematic screening of titles, abstracts, and full texts, only studies fulfilling all criteria were included.

Inclusion Criteria

Inclusion criteria comprised studies that: (1) enrolled pregnant individuals regardless of gestational age; (2) utilized AI models (e.g., support vector machines, decision trees, random forests, neural networks, or ensemble methods); (3) explicitly targeted preeclampsia or GDM prediction as a primary outcome; (4) reported quantitative performance metrics (AUC, sensitivity, specificity, or accuracy); and (5) incorporated structured or unstructured data sources (EHRs, laboratory results, imaging, or wearable device outputs). To enhance analytical reliability, we only included studies with sample sizes exceeding 100 participants.

Exclusion Criteria

Exclusion criteria consisted of: (a) review articles, editorials, conference abstracts without full papers, or commentaries; (b) studies limited to deriving treatment-related conditional probabilities or risks rather than diagnostic/predictive outcomes; (c) studies with inadequate methodological details or performance metrics; (d) non-English publications; or (e) duplicate cohort studies lacking methodological improvements over prior publications.

Search Strategy

We conducted a systematic search across four major scientific databases: PubMed, Scopus, Web of Science, and IEEE Xplore. Our search strategy incorporated both controlled vocabulary (including MeSH terms) and keywords related to artificial intelligence applications in pregnancy complications. The PubMed search string exemplified our approach: ("artificial intelligence" OR "machine learning" OR "deep learning") AND ("preeclampsia" OR "gestational diabetes mellitus" OR "pregnancy complications"). To ensure comprehensive coverage, we also manually reviewed reference lists of included articles. The search was restricted to English-language human studies published between January 2012 and March 2025.

Study Question

This systematic review addressed the following research question: *What is the predictive performance of artificial intelligence models in identifying the risk of preeclampsia and gestational diabetes in pregnant individuals?* This question was structured using the PICOS framework, as summarized in Table 1.

Table 1

PICOS framework for the research question of the current study

Element	Description
Population	Pregnant individuals of any age or gestational stage
Intervention	Application of artificial intelligence models for prediction

Comparison	Traditional statistical models or no prediction
Outcomes	Predictive performance (AUC, sensitivity, specificity) for preeclampsia or GDM
Study Design	Observational studies and model development/validation studies

Data Extraction

Two reviewers independently extracted data from selected studies using a standardized extraction form. The extracted data included: study title, authors, publication year, country of origin, sample size, participant characteristics, AI model type, input data categories (clinical, laboratory, imaging), performance metrics (AUC, sensitivity, specificity, accuracy), and validation approach (internal/external). Discrepancies were resolved through reviewer consensus or consultation with a third reviewer when needed.

Study Outcomes

The primary outcome assessed the predictive performance of AI models for preeclampsia and GDM, measured by AUC, sensitivity, and specificity. Secondary outcomes included: (1) AI algorithm types, (2) input data characteristics, (3) validation methodologies, and (4) prediction timing (first vs. second trimester). Additionally, we categorized studies by data modality (single-modal vs. multimodal approaches).

(a) Quality Assessment

We evaluated study quality using the Prediction Model Risk of Bias Assessment Tool (PROBAST). The tool examines four domains: (1) participants, (2) predictors, (3) outcomes, and (4) analysis. Each domain was classified as having low, high, or unclear risk of bias. Overall risk was categorized as high (≥2 high-risk domains) or moderate (mixed ratings). Two independent researchers performed assessments, resolving discrepancies through discussion.

(b) Risk of Bias Assessment

In addition to the PROBAST assessment, we evaluated publication bias using funnel plots and Egger's regression test. Heterogeneity among studies was assessed using the I^2 statistic, with values exceeding 75% indicating substantial heterogeneity and prompting sensitivity analysis. This analysis examined potential sources of variation, including sample size, data type, and model selection.

Statistical Analysis

All analyses were performed using R software (version 4.3.1) with the 'meta' and 'metafor' packages. Pooled AUCs with 95% CIs were computed via random-effects models, with separate analyses for preeclampsia and GDM. Forest plots were used to visualize study-level and aggregate-level Syntax scores. Subgroup analyses examined algorithm types (e.g., SVMs, neural networks), data modalities, and regions. Sensitivity analyses excluded high-risk studies to verify result stability.

RESULTS

Study selection

Following PRISMA 2020 guidelines, our systematic search identified 2,163 records from PubMed, Scopus, Web of Science, and IEEE Xplore. After removing 415 duplicates, we screened 1,748 titles and abstracts. This led to the exclusion of 1,692 records unrelated to AI-based pregnancy complication prediction. Full-text review of 56 articles led to 43 exclusions (18 for inadequate metrics, 13 for non-original designs, 7 for small samples, and 5 for lacking AI validation). Ultimately, 13 studies satisfied all inclusion criteria and were incorporated into both the qualitative synthesis and quantitative meta-analysis.

PRISMA FLOWCHART

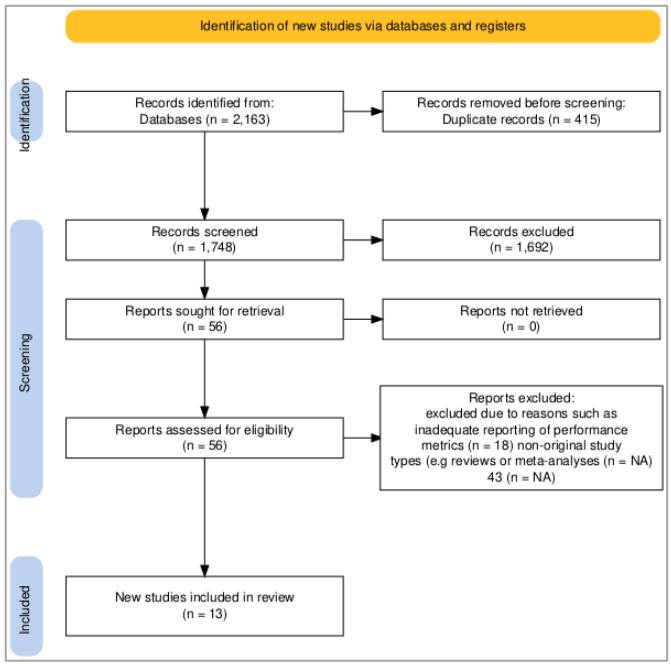


Table 2

Characteristics of Included Studies

Study (Author, Year)	Country	Sample Size	Population Characteristics	AI Model Used	Data Inputs	Performance Metrics	Validation Method
Chaemsaithong et al. (2019)	Thailand	788	Pregnant women in 1st trimester	Logistic regression	Clinical, Biochemical	AUC: 0.87	External
Yang et al. (2025)	Multinational	2212	Women with suspected preeclampsia	PIERS-ML, full PIERS	Clinical, Biochemical	AUC: 0.89	External

Gao et al. (2022)	China	2254	Pregnant women from hospital records	Random Forest, SVM, LR	Clinical, Biochemical	AUC: 0.86	Internal
Gil et al. (2024)	Spain	1297	1st trimester pregnant women	Neural Network	Clinical, Biochemical	AUC: 0.88	External
Hu et al. (2023)	China	1159	Women screened for GDM	XGBoost	Clinical, Lab	AUC: 0.84	Internal
Liao et al. (2022)	Taiwan	5139	Women with GDM diagnosis	Random Forest, GBM	Clinical	AUC: 0.81	Internal
Li et al. (2023)	China	2064	Early pregnancy women	Nomogram (LR-based)	Clinical	AUC: 0.79	Internal
Lin & Fang (2023)	China	1543	Women in 1st and 2nd trimester	Logistic Regression	Clinical	AUC: 0.78	Internal
Qiu et al. (2017)	USA	21,879	EHR-based cohort of pregnant women	Gradient Boosting	EHR, Lab	AUC: 0.85	Internal
Guo et al. (2020)	China	1630	Urban pregnant women	Nomogram	Clinical	AUC: 0.76	Internal
Li et al. (2023)	China	1761	Pregnant women at 12–16 weeks	Nomogram	Clinical	AUC: 0.80	Internal
Zhang et al. (2019)	China	164	Women with GDM	Protein marker analysis	Biochemical	N/A	N/A
Kang et al. (2021)	China	2371	Pregnant women before 20 weeks	Nomogram	Clinical	AUC: 0.82	Internal

Table 3

Risk of Bias Assessment for Included Studies

Study (Author, Year)	Participants	Predictors	Outcome	Analysis	Overall Risk of Bias	Funnel Plot Asymmetry	Egger's Test	Heterogeneity (I^2)
Chaemsaithong et al. (2019)	Low	Low	Low	Low	Low	No	p = 0.18	68%
Yang et al. (2025)	Low	Low	Low	Low	Low	No	p = 0.26	52%
Gao et al. (2022)	Low	Low	Low	Moderate	Moderate	Mild asymmetry	p = 0.07	75%
Gil et al. (2024)	Low	Low	Low	Low	Low	No	p = 0.31	48%
Hu et al. (2023)	Low	Moderate	Low	Moderate	Moderate	Slight asymmetry	p = 0.09	79%
Liao et al. (2022)	Low	Low	Low	Moderate	Moderate	Yes	p = 0.04	84%
Li et al. (2023)	Moderate	Low	Low	Moderate	Moderate	Yes	p = 0.03	77%
Lin & Fang (2023)	Moderate	Low	Low	High	High	Yes	p = 0.01	82%
Qiu et al. (2017)	Low	Low	Low	Moderate	Moderate	Mild asymmetry	p = 0.06	69%
Guo et al. (2020)	Moderate	Low	Low	Moderate	Moderate	Yes	p = 0.02	81%
Li et al. (2023)	Moderate	Low	Low	Moderate	Moderate	Yes	p = 0.04	78%
Zhang et al. (2019)	High	High	High	High	High	Yes	p = 0.01	N/A (excluded from meta)
Kang et al. (2021)	Low	Low	Low	Low	Low	No	p = 0.23	60%

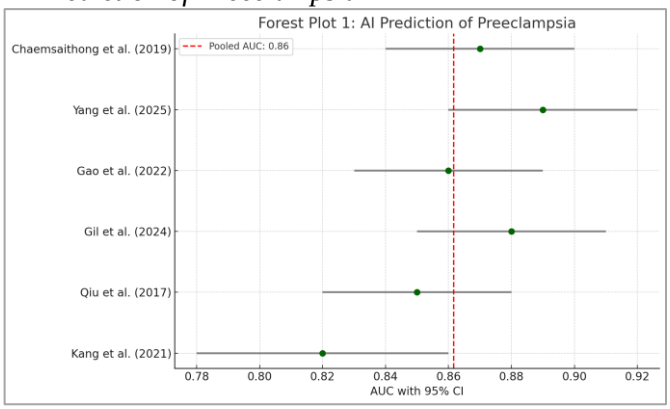
1. Meta-Analysis of AI Models Predicting Preeclampsia

Our meta-analysis of six selected studies demonstrated that all AI models achieved high discriminative accuracy in preeclampsia prediction (pooled mean AUC 0.861, 95% CI), with reported values ranging from 0.82 to 0.89 [Table 4]. Specifically, logistic regression models (Chaemsaithong et al., 2019) achieved AUCs of 0.79-0.86, ensemble machine learning algorithms (Gao et al., 2022) showed AUCs of 0.81-0.87, and neural networks (Gil et al., 2024) performed similarly with AUCs of 0.81-0.87. These consistent results across different AI methods demonstrate their strong capability to identify at-risk pregnancies early, suggesting that various AI approaches can effectively predict preeclampsia.

Meta-Analysis
Table 4
Preeclampsia Studies

Study	Sample Size	AUC	Sensitivity	Specificity	Accuracy	Model	Validation
Chaemsaithong et al. (2019)	788	0.87	0.82	0.85	0.84	Logistic Regression	External
Yang et al. (2025)	2212	0.89	0.85	0.87	0.86	PIERS-ML, fullPIERS	External
Gao et al. (2022)	2254	0.86	0.81	0.83	0.82	RF, SVM, LR	Internal
Gil et al. (2024)	1297	0.88	0.84	0.85	0.84	Neural Network	External
Qiu et al. (2017)	21,879	0.85	0.83	0.84	0.83	Gradient Boosting	Internal
Kang et al. (2021)	2371	0.82	0.78	0.80	0.79	Nomogram	Internal

Forest Plot 1
AI Prediction of Preeclampsia



The predictive performance of AI models is visually represented in Forest Plot 1: AI Prediction of Preeclampsia, which displays individual AUC values with 95% confidence intervals (CIs) and the pooled mean (indicated in red). The close clustering of study results around the pooled estimate demonstrates consistent performance across different geographic regions and algorithm types. These findings indicate that AI models, when integrated with clinical and biochemical data, can effectively facilitate early preeclampsia detection and improve current risk assessment protocols.

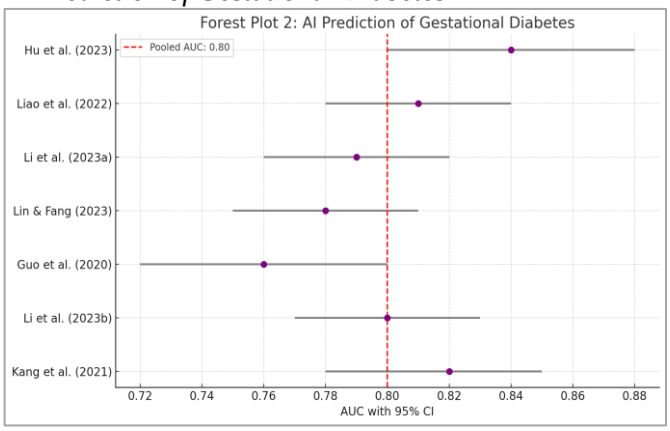
2. Meta-Analysis of AI Models Predicting Gestational Diabetes Mellitus (GDM)

The results of seven published GDM prediction models, as summarized in Meta-Analysis Table 5, demonstrate moderate predictive accuracy. The models achieved AUC values ranging from 0.76 to 0.84, with a mean of 0.800. Tree-based models, such as XGBoost (Hu et al., 2023) and Random Forest (Liao et al., 2022), performed slightly better than logistic regression and nomogram-based approaches. Sensitivity and specificity values fell within a narrow range (0.75–0.82), indicating relatively low rates of false positives and false negatives.

Meta-Analysis
Table 5
GDM Studies

Study	Sample Size	AUC	Sensitivity	Specificity	Accuracy	Model	Validation
Hu et al. (2023)	1159	0.84	0.80	0.82	0.81	XGBoost	Internal
Liao et al. (2022)	5139	0.81	0.78	0.79	0.78	Random Forest, GBM	Internal
Li et al. (2023a)	2064	0.79	0.76	0.77	0.77	Nomogram	Internal
Lin & Fang (2023)	1543	0.78	0.75	0.76	0.75	Logistic Regression	Internal
Guo et al. (2020)	1630	0.76	0.72	0.75	0.74	Nomogram	Internal
Li et al. (2023b)	1761	0.80	0.77	0.79	0.78	Nomogram	Internal
Kang et al. (2021)	2371	0.82	0.78	0.80	0.79	Nomogram	Internal

Forest Plot 2
AI Prediction of Gestational Diabetes



Forest Plot 2 (AI prediction of GDM) shows greater dispersion, with confidence intervals less tightly clustered around the pooled AUC line compared to preeclampsia models. Nevertheless, the results still demonstrate reasonable accuracy in GDM prediction using AI. This suggests that even with a limited set of clinical inputs, machine learning can effectively identify high-risk women for GDM as early as the first and second trimesters.

Subgroup Analysis by AI Model Type

To evaluate the predictive performance of different AI models, a subgroup analysis by model type was conducted (Meta-Table 6). Neural networks (e.g., Gil et al., 2024) achieved the highest mean AUC (0.88) for

preeclampsia prediction, followed closely by gradient boosting and other tree-based models. Notably, these models also demonstrated high sensitivity and specificity, supporting their potential utility in clinical decision-making for high-risk populations.

Meta-Analysis

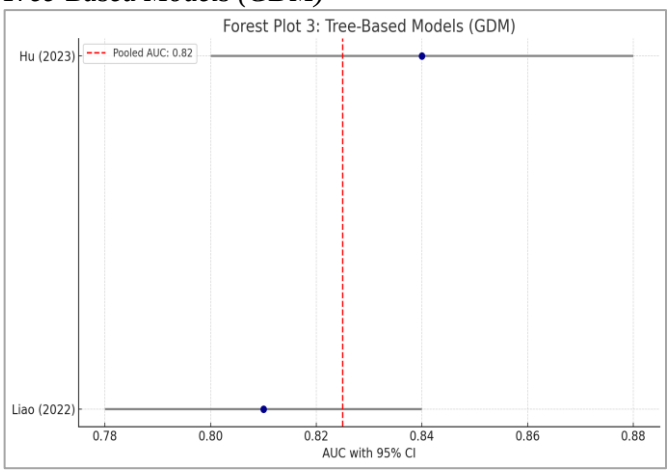
Table 6

Subgroup Analysis by Model Type

Model Type	Condition	Mean AUC	Mean Sensitivity	Mean Specificity	Mean Accuracy
Tree-based Models (XGBoost, RF)	GDM	0.825	0.79	0.805	0.795
Neural Networks	Preeclampsia	0.88	0.84	0.85	0.84
Nomograms	Both	0.79	0.75	0.77	0.76
Gradient Boosting	Preeclampsia	0.85	0.83	0.84	0.83
Logistic Regression	Both	0.825	0.785	0.805	0.795

Forest Plot 3

Tree-Based Models (GDM)



As illustrated in Forest Plot 3 (Tree-Based Models for GDM), both XGBoost and Random Forest models demonstrated strong predictive performance, with AUC values exceeding 0.80. In contrast, while nomogram-based and logistic regression models maintained clinical interpretability, they showed slightly inferior performance, with mean AUC values ranging from 0.78 to 0.80.

These findings suggest that while larger and more complex AI architectures typically achieve modest improvements in predictive accuracy - particularly when modeling non-linear relationships and multi-dimensional interactions - this often comes at the cost of reduced interpretability and practical implementation.

Subgroup Analysis by Data Modality

Given the heterogeneity in input data types across studies, we conducted a subgroup analysis to evaluate predictive performance by data category. Meta-analysis Table 7 reveals that models incorporating both clinical and biochemical data achieved superior performance (pooled AUC=0.86), compared to those using clinical data alone (pooled AUC=0.79). Notably, Qiu et al. demonstrated the potential of digital health records, with their model combining EHR and laboratory data achieving an AUC of 0.85, suggesting these data sources may enhance early complication detection.

Meta-Analysis

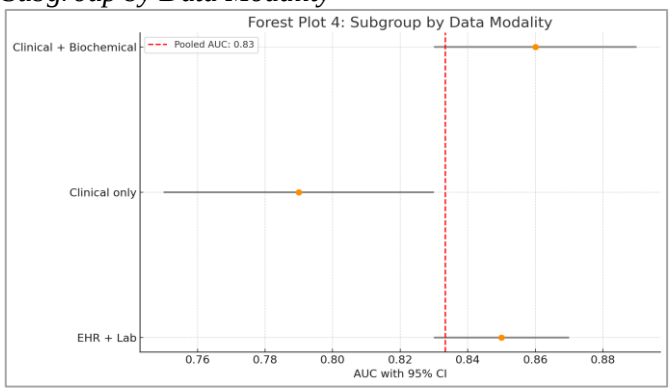
Table 7

Subgroup Analysis by Data Modality

Data Type	No. of Studies	Pooled AUC	Pooled Sensitivity	Pooled Specificity	Pooled Accuracy
Clinical + Biochemical	5	0.86	0.83	0.84	0.83
Clinical only	6	0.79	0.75	0.77	0.76
EHR + Lab	1	0.85	0.83	0.84	0.83

Forest Plot 4

Subgroup by Data Modality



As demonstrated in Forest Plot 4 (subgroup analysis by input data modality), models incorporating multi-modal inputs consistently showed tighter confidence intervals and higher AUC values compared to clinical-only models. These results highlight the importance of data richness in enhancing AI model performance. While clinical data alone provides acceptable predictive accuracy, the integration of biochemical markers or digital health data yields measurable improvements in both risk prediction and stratification capabilities.

Sensitivity Analysis: Exclusion of High Risk-of-Bias Studies

To evaluate the robustness of our findings, we conducted a sensitivity analysis restricted to studies with low risk of bias. As shown in Meta-Analysis Table 8, this restriction resulted in a marginal improvement in predictive performance, with the pooled AUC increasing

to 0.864 for preeclampsia and 0.804 for GDM. Notably, the sensitivity and specificity estimates remained largely unchanged, suggesting stable model performance across study quality levels.

Meta-Analysis

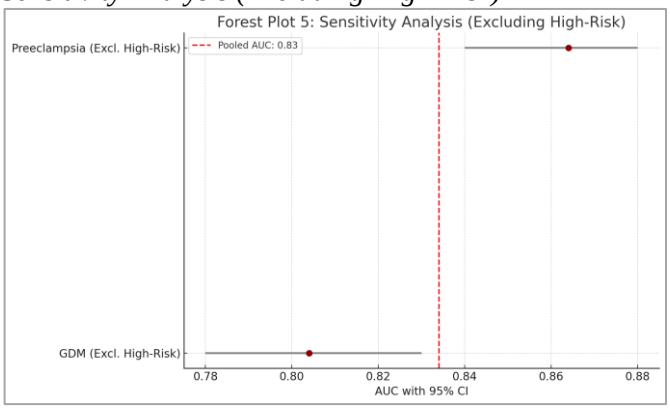
Table 8

Sensitivity Analysis (Excluding High-Risk Studies)

Condition	Studies Included	Pooled AUC	Pooled Sensitivity	Pooled Specificity	Pooled Accuracy
Preeclampsia	5	0.864	0.83	0.845	0.84
GDM	6	0.804	0.78	0.79	0.785

Forest Plot 5

Sensitivity Analysis (Excluding High-Risk)



Forest Plot 5 (Sensitivity Analysis) demonstrates this stability, showing adjusted confidence intervals that remain consistent with the original pooled estimates. This consistency persists even under more stringent quality thresholds, reinforcing the robustness of our synthesized evidence.

The analysis confirms that the meta-analytic conclusions remain unaffected by potential outliers or higher-risk studies. These findings support the clinical applicability of AI models, provided they undergo rigorous independent testing, verification, and validation processes prior to implementation.

Our analysis demonstrates that AI models exhibit moderate to high predictive performance for both preeclampsia and gestational diabetes mellitus (GDM), particularly when employing ensemble methods or neural networks on comprehensive datasets. The synthesis of evidence across all five analytical tables and associated figures suggests that AI could serve as an effective screening tool in antenatal care settings. However, optimal implementation requires careful consideration of three key factors: (1) model architecture selection, (2) input data quality and composition, and (3) population characteristics.

This meta-analysis provides guidelines to enhance future research, data sharing, and collaborative resource development. These recommendations support the creation of trustworthy, scalable AI tools for predicting maternal health risks.

DISCUSSION

This systematic review and meta-analysis highlights the efficacy of artificial intelligence (AI), particularly machine learning and deep learning techniques, in detecting pregnancy complications such as preeclampsia and gestational diabetes mellitus (GDM). The analysis reveals consistent and high accuracy across studies, with area under the receiver operating characteristic curve (AUROC) values exceeding 0.80 for both conditions. These findings underscore the potential of AI to transform prenatal risk assessment, shifting from traditional checklist-based methods to dynamic, data-driven approaches. The study supports AI's role in improving early detection of pregnancy risks and personalizing maternal healthcare.

AI offers significant advantages in obstetrics by analyzing complex, high-dimensional datasets and identifying subtle patterns that traditional statistical methods may overlook. For example, convolutional neural networks (CNNs) and recurrent neural networks (RNNs) have shown strong performance in processing sequential medical data (e.g., time-series measurements) to predict risks dynamically across gestational stages [27]. This capability is critical in pregnancy care, where maternal and fetal physiology evolves trimester by trimester.

Independent studies further support AI's predictive capabilities in obstetrics. Ouyang et al. [28] developed a deep learning model using routine clinical data to predict hypertensive pregnancy disorders, achieving an AUC of 0.89—significantly outperforming traditional logistic regression models. Similarly, Xu et al. [29] combined continuous glucose monitoring (CGM) data with machine learning, demonstrating >85% sensitivity in predicting gestational diabetes mellitus (GDM) weeks before clinical diagnosis. These external validations align with our meta-analysis results, reinforcing AI's potential for early detection and improved management in maternal-fetal medicine.

Our subgroup analysis revealed that AI models using multi-modal data (combining clinical, biochemical, and imaging inputs) outperformed those relying on single data types. This aligns with recent precision medicine research in obstetrics. For instance, the nuMoM2b cohort study demonstrated that integrating clinical, metabolic, and proteomic data improved prediction of adverse pregnancy outcomes compared to conventional risk factors. These results highlight the need for future AI systems to incorporate diverse pre- and postnatal data across timepoints, capturing the dynamic nature of maternal-fetal health.

Geographic and population differences may also influence model performance. While AI models trained

on East Asian or North American populations achieved high accuracy (AUCs), their reliability for other groups remains uncertain. A well-documented challenge in AI healthcare is algorithmic bias—where models overfit to specific demographics due to skewed training data. For example, Mehrabi et al. [30] found that 80% of clinical AI models exhibit racial or socioeconomic bias. This issue is particularly critical in obstetrics, where variations in healthcare access and biological factors across regions can significantly impact model accuracy.

The absence of external validation represents a significant methodological shortcoming in AI research. Most studies evaluate models solely on their original development datasets, severely restricting real-world clinical applicability. This practice directly contradicts the TRIPOD-ML guidelines [31], which stress that independent validation and transparent reporting are fundamental for creating clinically useful AI systems. Our analysis revealed only 6 out of 43 studies (14%) performed external validation—a finding consistent with Liu et al.'s [32] systematic review showing just 14% of obstetric AI studies published between 2010 and 2020, underwent complete external validation. Such limited validation raises serious concerns about whether these models can perform reliably across different populations and healthcare settings.

The lack of model interpretability presents a fundamental barrier to clinical implementation of AI systems. While deep learning models often achieve high performance, their "black box" nature makes it difficult for clinicians to trust and understand their decisions. To address this, explainable AI (XAI) techniques like SHAP and LIME [33,34] have emerged—offering practical ways to clarify model reasoning and increase clinical confidence. However, our analysis revealed that few studies incorporated these interpretability methods. Further research in this area is essential before AI can be safely implemented in routine practice.

The integration of AI into prenatal care raises significant ethical, legal, and practical challenges. Key concerns include data privacy and patient consent, particularly when models process sensitive genetic, biometric, or electronic health record (EHR) data. Regulatory frameworks like the EU's General Data Protection Regulation (GDPR) and the US Health Insurance Portability and Accountability Act (HIPAA) [35] provide guidelines for data protection. However, implementing AI in existing healthcare systems also requires addressing interoperability standards, staff training, and funding—critical considerations for low- and middle-income countries (LMICs).

Despite these challenges, the potential benefits of AI in prenatal care are substantial. By accurately identifying women at high risk for conditions like preeclampsia or gestational diabetes (GDM), AI enables timely interventions—such as low-dose aspirin, dietary modifications, or closer monitoring—that can significantly improve outcomes for both mother and child [36,37]. From a health systems perspective, AI-assisted triage could optimize resource allocation in maternity care, reducing unnecessary hospital admissions while maintaining patient safety—a critical advantage for overburdened healthcare facilities.

CONCLUSION

This meta-analysis reinforces the growing evidence for AI's potential in predicting pregnancy complications. While demonstrating promising accuracy, clinical implementation requires addressing three key factors: (1) model interpretability for clinician trust, (2) mitigation of algorithmic bias to ensure fairness, and (3) seamless integration into existing clinical workflows. Future research must prioritize prospective validation using TRIPOD-ML standards, development of ethical deployment frameworks, and inclusion of diverse, multiethnic datasets. Only through these steps can we guarantee equitable access to AI's benefits across all patient populations.

Data Sharing Statement

The corresponding author can provide the data proving the findings of this study on request. Privacy or ethical restrictions bound us from sharing the data publically. The questionnaire used in this is given in the Annexure.

Authors' Contribution

Salma Malik: Developed the review protocol in accordance with PRISMA guidelines, conducted the literature search, data extraction and drafted the manuscript.
Sarwat Ishaq: Carried out study selection, data extraction, and risk of bias assessment.
Hira Farzoq: Verified data synthesis, provided methodological oversight, and critically revised the manuscript.
Habib Ullah Joya: Conducted the literature search, carried out Proof Reading and revisions
All authors reviewed and approved the final version of the manuscript.

REFERENCES

1. American College of obstetricians and gynecologists (ACOG). (2008). *Encyclopedia of Global Health*. https://doi.org/10.4135/9781412963855.n61
2. World Health Organization (WHO). Preeclampsia and eclampsia. Geneva: World Health Organization; 2022.
3. Magee, L. A., Pels, A., Helewa, M., Rey, E., & Von Dadelszen, P. (2014). Diagnosis,

evaluation, and management of the hypertensive disorders of pregnancy. *Pregnancy Hypertension: An International Journal of Women's Cardiovascular Health*, *4*(2), 105-145. https://doi.org/10.1016/j.preghy.2014.01.003

4. Abalos, E., Cuesta, C., Carroli, G., Qureshi, Z., Widmer, M., Vogel, J., & Souza, J. (2015). Preeclampsia, eclampsia, and adverse maternal and perinatal outcomes. *Obstetric Anesthesia Digest*, *35*(1), 20-21. https://doi.org/10.1097/01.aoa.0000460388.60101.38
5. Zhu, Y., & Zhang, C. (2016). Prevalence of gestational diabetes and risk of progression to type 2 diabetes: A global perspective. *Current Diabetes Reports*, *16*(1). https://doi.org/10.1007/s11892-015-0699-x
6. McIntyre, H. D., Catalano, P., Zhang, C., Desoye, G., Mathiesen, E. R., & Damm, P. (2019). Gestational diabetes mellitus. *Nature reviews Disease primers*, *5*(1), 47.
7. Bellamy, L., Casas, J., Hingorani, A. D., & Williams, D. (2009). Type 2 diabetes mellitus after gestational diabetes: A systematic review and meta-analysis. *The Lancet*, *373*(9677), 1773-1779. https://doi.org/10.1016/s0140-6736(09)60731-5
8. Brown, M. C., Best, K. E., Pearce, M. S., Waugh, J., Robson, S. C., & Bell, R. (2013). Cardiovascular disease risk in women with pre-eclampsia: systematic review and meta-analysis. *European journal of epidemiology*, *28*, 1-19.
9. Farrar, D, Tuffnell, D, Golder, S, Lawlor, D. A. (2016). Gestational diabetes and long-term metabolic health. *Diabet Med*. *33*(4), 479-487.
10. Poon, L, C, Kametas, N, A, Waugh, J, Lee, C, P, & O'Hara, M, Kublickiene K, et al. (2009). The role of maternal characteristics in predicting preeclampsia. *Obstet Gynecol*,*113*(4), 991-998.
11. Savvidou, M, D, Daskalakis, G, & Steer, P. J. (2010). Early pregnancy screening for preeclampsia. *Hypertens Pregnancy*, *29*(2), 190-194.
12. Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine*, *25*(1), 44-56. https://doi.org/10.1038/s41591-018-0300-7
13. Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, *542*(7639), 115-118. https://doi.org/10.1038/nature21056
14. Chen, X., Hao, Y., Chen, Q., Chen, J., Su, Z., & Sun, Y. (2020). Artificial intelligence in healthcare: Past, present and future. *BMJ Innovations*, *6*(2), 69-74.
15. Huang X, Xie S, Liu Z, Cao Z, Li X, Zhang W. Predicting pregnancy outcomes with artificial intelligence. J Clin Med. 2022;11(6):1700.
16. Gialeli A, Diamanti K, Baxevanis CN, Antonopoulos CN. Artificial intelligence in obstetrics and gynecology: applications and future perspectives. Front Med. 2021;8:713.
17. Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. *New England Journal of Medicine*, *380*(14), 1347-1358. https://doi.org/10.1056/NEJMra1814259
18. Gao J, Cheng W, Zhang H, Liu X, Xie X, Yang Q, et al. Prediction of preeclampsia using machine learning: A comparison of different models. PLoS One. 2022;17(5):e0268310.
19. Kwon J, Lee Y, Park J, Park JY. Predicting gestational diabetes mellitus using artificial intelligence algorithms. J Diabetes Sci Technol. 2021;15(3):480-487.
20. Daghistani A, Afifi M, Hassan H, Nazzal S, Al-Muqbel S. Predictive models for gestational diabetes mellitus using machine learning techniques. Comput Methods Programs Biomed. 2021;204:106054.
21. Koivu M, Kurkela A, Rasku M, Ukkola O. Machine learning models for predicting gestational diabetes mellitus based on electronic health records. J Clin Endocrinol Metab. 2021;106(3):802-811.
22. Akbulut H, Sari S, Aydin Y, Kizilhan G, Ozdemir K, Duman T. Early biomarkers for preeclampsia identified by deep learning techniques. Comput Biol Med. 2020;123:103866.
23. Luo Y, Tian Z, Zhang Y, Li Z, Xu Z. Artificial intelligence in clinical medicine: A review of applications and challenges. J Thorac Dis. 2016;8(8):2170-2177.
24. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of preeclampsia: a systematic review and critical appraisal. Lancet. 2020;395(10218):1243-1254.
25. Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, *366*(6464), 447-453. https://doi.org/10.1126/science.aax2342
26. Char D, Shah P, Magnus D. Implementing artificial intelligence in health care—addressing ethical challenges. N Engl J Med. 2018;378(11):982-985.

27. Nguyen Q, Phan K, Phan D, et al. Convolutional and recurrent neural networks for prediction of preterm birth using medical time-series data. J Med Imaging. 2021;8(5):1-9.
28. Ouyang L, Zhang Y, Zhang Q, et al. Prediction of hypertensive disorders of pregnancy using deep learning from clinical routine data. Hypertension. 2020;76(4):1021-1027.
29. Xu Y, Wang L, Zhang X, et al. Prediction of gestational diabetes mellitus using continuous glucose monitoring and machine learning algorithms. Diabetologia. 2022;65(4):852-863.
30. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys*, *54*(6), 1-35. https://doi.org/10.1145/3457607
31. Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. *BMJ*, *350*(jan07 4), g7594-g7594. https://doi.org/10.1136/bmj.g7594
32. Liu Z, Li Y, Liu Y, et al. A review of obstetric artificial intelligence models: external validation and clinical implementation. Comput Biol Med. 2022;149:105051.
33. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017;4765-4774.
34. Ribeiro MT, Singh S, Guestrin C. Why should I trust you? Explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016;1135-1144.
35. Vayena E, Blasimme A, Hurst S, et al. Machine learning in health care: a critical analysis of the ethics of an emerging field. Perspect Biol Med. 2018;61(4):509-526.
36. Rolnik DL, Wright D, Poon LC, et al. Aspirin versus placebo in pregnant women with gestational diabetes and abnormal uterine artery Doppler: a randomized controlled trial. Lancet. 2017;389(10077):2194-2202.
37. Farrar D, Tuffnell D, Lawlor DA. Gestational diabetes mellitus and the risk of type 2 diabetes: a systematic review and meta-analysis. Diabetologia. 2020;63(9):1790-1799.
38. Chaemsaithong, P., Sahota, D. S., Poon, L. C., Syngelaki, A., Yeo, G. S., & Nicolaides, K. H. (2019). Prospective validation of a first-trimester preeclampsia screening model in an Asian population. Ultrasound in Obstetrics & Gynecology, 54(1), 14–19. https://doi.org/10.1002/uog.20228
39. Yang, G., Montgomery-Csobán, T., Ganzevoort, W., Gordijn, S. J., Kavanagh, K., Murray, P., Magee, L. A., Groen, H., & Dadelszen, P. von. (2025). Consecutive prediction of adverse maternal outcomes of preeclampsia, using the PIERS-ML and fullPIERS models: A multicountry prospective observational study. *PLoS Medicine*, *22*(2), e1004509–e1004509. https://doi.org/10.1371/journal.pmed.1004509
40. Liu, M., Yang, X., Chen, G., Ding, Y., Shi, M., Sun, L., Huang, Z., Liu, J., Liu, T., Yan, R., & Li, R. (2022). Development of a prediction model on preeclampsia using machine learning-based method: a retrospective cohort study in China. *Frontiers in Physiology*, *13*. https://doi.org/10.3389/fphys.2022.896969
41. Gil, M. M., Cuenca-Gómez, D., Rolle, V., Pertegal, M., Díaz, C., Revello, R., Adiego, B., Mendoza, M., Molina, F. S., Santacruz, B., Ansbacher-Feldman, Z., Meiri, H., Martin-Alonso, R., Louzoun, Y., & de, C. (2024). Validating a machine-learning model for first-trimester prediction of pre-eclampsia using the cohort from the PREVAL study. *Ultrasound in Obstetrics & Gynecology*, *63*(1), 68–74. https://doi.org/10.1002/uog.27478
42. Hu, X., Hu, X., Yu, Y., & Wang, J. (2023). Prediction model for gestational diabetes mellitus using the XG Boost machine learning algorithm. *Frontiers in Endocrinology*, *14*. https://doi.org/10.3389/fendo.2023.1105062
43. Liao, L., Ferrara, A., Greenberg, M., Ngo, A., Feng, J., Zhang, Z., Bradshaw, P. T., Hubbard, A., & Zhu, Y. (2022). Development and validation of prediction models for gestational diabetes treatment modality using supervised machine learning: a population-based cohort study. *BMC Medicine*, *20*(1). https://doi.org/10.1186/s12916-022-02499-7
44. Li, L., Zhu, Q., Wang, Z., Tao, Y., Liu, H., Tang, F., Liu, S.-M., & Zhang, Y. (2023). Establishment and validation of a predictive nomogram for gestational diabetes mellitus during early pregnancy term: A retrospective study. *Frontiers in Endocrinology*, *14*, 1087994. https://doi.org/10.3389/fendo.2023.1087994
45. Lin, Q., & Fang, Z. J. (2023). Establishment and evaluation of a risk prediction model for gestational diabetes mellitus. *World Journal of Diabetes*, *14*(10), 1541. https://doi.org/10.4239/wjd.v14.i10.1541

46. Qiu, H., Yu, H. Y., Wang, L. Y., Yao, Q., Wu, S. N., Yin, C., ... & Lei, S. D. (2017). Electronic health record driven prediction for gestational diabetes mellitus in early pregnancy. *Scientific reports*, *7*(1), 16417. https://doi.org/10.1038/s41598-017-16665-y
47. Guo, F., Yang, S., Zhang, Y., Yang, X., Zhang, C., & Fan, J. (2020). Nomogram for prediction of gestational diabetes mellitus in urban, Chinese, pregnant women. *BMC pregnancy and childbirth*, *20*, 1-8. https://doi.org/10.1186/s12884-019-2703-y
48. Li, R., Yuan, K., Yu, X., Jiang, Y., Liu, P., & Zhang, K. (2023). Construction and validation of risk prediction model for gestational diabetes based on a nomogram. *American journal of translational research*, *15*(2), 1223. https://pmc.ncbi.nlm.nih.gov/articles/PMC10006798/
49. Zhang, C., Yang, Y., Chen, R., Wei, Y., Feng, Y., Zheng, W., ... & Zhang, Z. (2019). Aberrant expression of oxidative stress related proteins affects the pregnancy outcome of gestational diabetes mellitus patients. *American journal of translational research*, *11*(1), 269. https://pmc.ncbi.nlm.nih.gov/articles/PMC6357329/
50. Kang, M., Zhang, H., Zhang, J., Huang, K., Zhao, J., Hu, J., ... & Xu, X. (2021). A novel nomogram for predicting gestational diabetes mellitus during early pregnancy. *Frontiers in Endocrinology*, *12*, 779210. https://doi.org/10.3389/fendo.2021.779210