INTERNATIONAL JOURNAL OF INTELLIGENT DATA AND MACHINE LEARNING

# MACHINE LEARNING MODEL IMPLEMENTATION STRATEGIES AND PREDICTIVE FACTORS FOR PREECLAMPSIA FORECASTING: A REVIEW

**Dr. Elena Petrova**
**Institute of Biomedical Engineering and Informatics, Sechenov University, Moscow, Russia**

**Prof. David J. Hernandez**
**Department of Computer Science, University of California, San Diego, USA**

**Published Date: 19 December 2024 // Page no.:- 17-22**

## ABSTRACT

Preeclampsia remains a leading cause of maternal and perinatal morbidity and mortality globally. Accurate and early prediction is crucial for timely intervention and improved outcomes. Machine learning (ML) has emerged as a promising approach for identifying individuals at high risk of developing preeclampsia by leveraging complex patterns within diverse datasets. However, translating promising ML research models into effective, reliable, and scalable clinical deployment presents significant challenges. This article reviews the current landscape of machine learning applications in preeclampsia prediction, focusing on identified deployment patterns and key predictive features. We synthesize findings from recent literature, discussing commonly employed ML algorithms, the types of data and features utilized (including maternal characteristics, biomarkers, and clinical history), and the reported predictive performance. Crucially, we examine the challenges and considerations related to the practical implementation of these models within healthcare systems, including data quality, model interpretability, integration into clinical workflows, and the necessity of robust MLOps practices. This review highlights the critical need to address deployment-related aspects to ensure that ML models for preeclampsia prediction can move beyond research settings and achieve real-world clinical impact, ultimately contributing to improved maternal health outcomes.

**Keywords:** Preeclampsia, Machine Learning, Prediction, Deployment, Implementation, Predictive Factors, Features, Review, MLOps, Healthcare AI.

## INTRODUCTION

Preeclampsia, a hypertensive disorder of pregnancy, is a major global health concern, contributing significantly to adverse maternal and perinatal outcomes, including preterm birth, fetal growth restriction, placental abruption, and maternal organ failure [1, 2]. Affecting approximately 2-8% of pregnancies worldwide, its unpredictable onset and rapid progression necessitate effective strategies for early identification and risk stratification [1, 2]. Current clinical approaches to preeclampsia risk assessment often rely on maternal history and basic clinical factors, which have limited predictive accuracy, particularly for late-onset preeclampsia [4, 30]. While interventions like low-dose aspirin have shown efficacy in preventing preterm preeclampsia in high-risk women, identifying these individuals accurately remains a challenge [3].

The increasing availability of large datasets in healthcare, including electronic health records, -omics data, and wearable device information, coupled with advancements in computational power, has fueled interest in applying machine learning (ML) techniques to complex medical prediction tasks [5, 9]. Machine learning models have the capacity to analyze intricate relationships within high-dimensional data, potentially uncovering subtle patterns that are not apparent through traditional statistical methods [34]. Consequently, numerous studies have explored the use of ML for preeclampsia prediction, employing a variety of algorithms and data sources [7, 9, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32].

While research has demonstrated promising predictive performance in controlled settings, the successful translation of these ML models into routine clinical practice remains a significant hurdle [5, 9]. The challenges extend beyond model development to encompass critical aspects of deployment and implementation within complex healthcare environments. These include ensuring data quality and accessibility [33], integrating models seamlessly into existing clinical workflows, maintaining model performance over time (considering potential concept drift), ensuring model interpretability and trustworthiness for clinicians, and addressing ethical and regulatory considerations [55, 56, 57]. The gap between developing a high-performing model in a research environment and deploying it reliably and effectively in a clinical setting is substantial [40, 44, 45].

This article aims to provide a review of machine learning applications in preeclampsia prediction with a specific focus on the patterns observed in model deployment and the key features identified as predictive. By synthesizing findings from recent systematic reviews and individual studies, we seek to highlight the state of the art in model development and critically examine the practical considerations and challenges associated with implementing these models in real-world clinical settings. Understanding these deployment patterns and key features is crucial for guiding future research and development efforts towards creating ML-based preeclampsia prediction tools that are not only accurate but also clinically viable and impactful.

## 2. METHODS

This review synthesizes information from existing literature on machine learning applications for preeclampsia prediction, with a specific focus on identifying common deployment patterns and key predictive features. The methodology employed for this review involved searching and analyzing relevant publications to extract pertinent information regarding model types, features used, reported performance, and discussions related to implementation challenges.

### 2.1. Literature Search and Selection

A targeted search of academic databases (such as PubMed, IEEE Xplore, and Google Scholar) was conducted using keywords such as "preeclampsia," "machine learning," "prediction," "forecasting," "artificial intelligence," "deployment," "implementation," "features," and "biomarkers." The search focused on publications from recent years to capture the most up-to-date advancements in the field. While a formal systematic review with strict inclusion/exclusion criteria and quality assessment tools like PROBAST [12] was not performed for this article, the selection process prioritized systematic reviews [7, 8, 9], meta-analyses [3], and individual studies that provided detailed information on model development, features, and, ideally, discussions on implementation or validation in different cohorts [15, 16, 18, 19, 20, 24, 25, 26, 27, 28, 31, 32]. Publications focusing solely on biological mechanisms without predictive modeling were generally excluded.

### 2.2. Data Extraction

Information extracted from the selected literature included:

- Type of Machine Learning Model(s) used (e.g., logistic regression, support vector machines, random forests, neural networks, Bayesian networks).
- Types of Features used for prediction (e.g., maternal demographics, medical history, clinical measurements like blood pressure [38, 39], laboratory biomarkers, ultrasound parameters).
- Timing of Prediction (e.g., first trimester, second trimester, dynamic prediction throughout pregnancy).
- Reported Predictive Performance Metrics (e.g., Accuracy, Area Under the Receiver Operating Characteristic Curve (AUC), sensitivity, specificity).
- Discussions on Model Validation (internal, external) and Implementation Challenges.
- Mentions of Deployment Strategies or Infrastructure (though expected to be limited in research papers).

### 2.3. Synthesis and Analysis

The extracted data were synthesized to identify recurring patterns in the types of ML models and features commonly employed for preeclampsia prediction. The reported performance metrics were reviewed to understand the general state of predictive accuracy in the field. A critical analysis was conducted to identify common themes and challenges related to the deployment and implementation of these models in clinical settings, drawing upon discussions within the selected papers and general knowledge of deploying ML systems in healthcare [40, 41, 42, 44, 45, 46, 47, 52]. The review also considered the importance of data quality [33], model interpretability [16, 58, 59], and ethical considerations [55, 56, 57] as crucial aspects influencing deployment.

### 2.4. Focus on Deployment Patterns

Given the title's emphasis, particular attention was paid to any mention of how models were intended to be used clinically, challenges encountered during real-world testing or validation, or discussions about the infrastructure and processes required for operationalizing these models (even if these mentions were brief or conceptual). This included looking for discussions related to:

- Integration with Electronic Health Records (EHRs).
- Requirements for real-time or near real-time prediction.
- Strategies for model updates and maintenance (MLOps concepts) [41, 42, 44, 45, 46, 47, 52].
- Challenges in diverse patient populations or different healthcare settings.

This methodological approach, while not a full systematic review, allowed for a focused synthesis of the literature to address the specific objectives of identifying deployment patterns and key features in ML-based preeclampsia prediction.

## 3. RESULTS

The review of the literature on machine learning for preeclampsia prediction revealed several common patterns in model development and identified key features frequently associated with predictive performance. While comprehensive details on clinical deployment infrastructure were often limited in research publications, recurring themes regarding implementation challenges and necessary considerations emerged.

### 3.1. Commonly Employed Machine Learning Models

A variety of machine learning algorithms have been applied to preeclampsia prediction. Frequently encountered models include:

- Logistic Regression: Often used as a baseline due to its interpretability and simplicity [34, 37].
- Support Vector Machines (SVM): Popular for their ability to handle high-dimensional data.
- Random Forests and Gradient Boosting Machines: Ensemble methods known for their robustness and good performance [11]. Adaptive variants of these models have also been explored in data stream contexts, which could be relevant for continuous monitoring [9, 10, 11].
- Neural Networks (including Deep Learning): Increasingly used, particularly for complex datasets, though often considered "black box" models [19, 21, 22, 23, 25]. Imbalance-aware neural networks have been developed to address the class imbalance inherent in preeclampsia data [22].
- Bayesian Networks: Provide probabilistic predictions and can offer some level of interpretability [13].
- Decision Trees and Model Trees: Offer interpretability and can be adapted for streaming data [16, 17, 18].

Many studies compare the performance of multiple algorithms on the same dataset [7, 9, 27, 28], and ensemble techniques are often found to achieve state-of-the-art results [9, 11].

### 3.2. Key Predictive Features

The literature consistently highlights several types of features as important predictors of preeclampsia:

- Maternal Characteristics: Age, parity (number of previous pregnancies), body mass index (BMI), ethnicity [21], and medical history (e.g., chronic hypertension, diabetes, previous preeclampsia) are fundamental predictors included in almost all models [4, 6, 15, 26, 29, 30, 31, 32].
- Biomarkers: Biochemical markers, particularly those measured in the first trimester, such as Placental Growth Factor (PlGF) and soluble fms-like tyrosine kinase-1 (sFlt-1), are strong predictors, especially when combined with other factors [4, 6, 26]. Other biomarkers related to angiogenesis and inflammation are also explored.
- Clinical Measurements: Mean arterial pressure (MAP) in early pregnancy is a crucial predictor [4, 6, 26]. Changes in blood pressure throughout pregnancy are also highly informative [38, 39]. Uterine artery Doppler pulsatility index (UtAD-PI) is another important clinical measurement [4, 6, 26].
- Clinical History: Specific details from previous pregnancies, family history of preeclampsia, and pre-existing medical conditions contribute significantly to risk assessment [4, 6, 15, 26, 29, 30, 31, 32].

Combinations of these features, particularly maternal factors, biomarkers, and UtAD-PI in the first trimester, form the basis of many high-performing prediction models [4, 6, 26]. Some studies explore the use of "zero-cost" predictors derived solely from maternal characteristics to improve accessibility [29].

### 3.3. Reported Predictive Performance

Reported predictive performance varies widely across studies due to differences in datasets, features used, algorithms, and evaluation methodologies [7, 8, 9]. However, many recent ML models, particularly those using combinations of maternal factors and biomarkers in the first trimester, report AUC values ranging from approximately 0.8 to over 0.9 for predicting preterm preeclampsia [4, 6, 15, 26, 27]. Prediction of late-onset preeclampsia generally shows lower performance [31]. External validation on independent cohorts is crucial but often reveals a drop in performance compared to internal validation, highlighting challenges in generalizability [20, 24, 25]. Leaderboards and systematic evaluation frameworks are needed to provide a clearer picture of comparative performance across different models and datasets [35, 36].

### 3.4. Deployment Patterns and Challenges

While explicit descriptions of clinical deployment infrastructure are rare in the reviewed literature, several patterns and challenges related to implementation are consistently discussed:

- Research Prototype Focus: The vast majority of studies present models as research prototypes validated on historical datasets. The focus is primarily on demonstrating predictive accuracy rather than addressing the complexities of real-world deployment.
- Data Quality and Accessibility: A major challenge is obtaining high-quality, standardized data from diverse clinical sources [33]. Data heterogeneity, missing values, and the lack of interoperability between different electronic health record (EHR) systems hinder model development and deployment.
- Integration into Clinical Workflow: Seamless integration of ML prediction tools into existing clinical workflows is critical for adoption. Clinicians need user-friendly interfaces that provide timely and actionable risk assessments without disrupting their routines. This requires careful consideration of the human-computer interaction aspects [5].
- Model Interpretability: Clinicians often require interpretable models to understand the rationale behind a

prediction and build trust in the system [16, 58, 59]. "Black box" models like deep neural networks can face resistance in clinical settings. Efforts towards explainable AI (XAI) are relevant here [58, 59].

- Scalability and Infrastructure: Deploying ML models to handle predictions for a large number of pregnancies in real-time requires robust and scalable infrastructure [43, 50, 51]. Cloud-based platforms and MLOps practices are essential for managing the model lifecycle, including training, deployment, monitoring, and updates [41, 42, 44, 45, 46, 47, 52]. Challenges related to technical debt in ML systems are significant in this context [40].

- Model Maintenance and Concept Drift: Preeclampsia risk factors or their relationships might subtly change over time or vary across different populations [8]. Deployed models require continuous monitoring of performance and periodic retraining or updates to adapt to potential concept drift [41, 42, 44, 45, 46, 47, 52].

- Ethical and Regulatory Considerations: Deploying AI in healthcare raises significant ethical concerns, including bias in predictions across different demographic groups, data privacy (e.g., GDPR compliance) [49, 55], accountability for errors, and ensuring equitable access to the technology [55, 56, 57]. Regulatory frameworks for medical AI are still evolving.

- Validation in Diverse Populations: Models developed and validated on specific populations may not generalize well to others [20, 24, 25]. Prospective external validation on diverse cohorts is crucial before widespread deployment.

These results indicate that while ML research in preeclampsia prediction is advancing, significant work is needed to address the practicalities of deploying these models effectively and responsibly in clinical practice.

4. Discussion and Conclusion

The application of machine learning to preeclampsia prediction holds immense promise for improving maternal and neonatal outcomes by enabling earlier and more accurate risk stratification [5, 7, 9]. The review of the literature highlights the diversity of ML models explored and the consistent identification of key predictive features, including maternal characteristics, biomarkers, and clinical measurements [4, 6, 15, 26, 29, 30, 31, 38, 39]. While reported predictive performance in research settings is often encouraging, the translation of these models into widespread clinical use is hampered by significant deployment and implementation challenges.

A major finding of this review is the prevalent focus on research-stage model development, with limited detailed reporting or analysis of real-world deployment strategies. This gap between model development and clinical implementation is a critical barrier to realizing the full potential of ML in preeclampsia prediction. Challenges related to data quality and accessibility within heterogeneous healthcare systems are fundamental and require standardized data collection methodologies and interoperability solutions [33, 54].

Furthermore, the successful integration of ML models into existing clinical workflows necessitates user-centric design and a focus on providing interpretable insights to clinicians [5, 16, 58, 59]. Black-box models, despite potentially high accuracy, may face resistance if clinicians cannot understand the basis for a prediction. The ethical implications, including algorithmic bias and data privacy, must be proactively addressed throughout the development and deployment lifecycle to ensure equitable and trustworthy AI systems [49, 55, 56, 57]. Frameworks for responsible AI in health are crucial in this regard [55, 56].

The need for robust MLOps practices is paramount for maintaining the performance and reliability of deployed models over time [41, 42, 44, 45, 46, 47, 52]. Continuous monitoring of model performance, mechanisms for detecting and adapting to concept drift [8], and streamlined processes for model updates are essential for ensuring that predictions remain accurate as underlying patterns evolve or new data becomes available. Addressing technical debt early in the development process is also vital for long-term maintainability and scalability [40].

Future research should shift focus from solely demonstrating predictive accuracy on historical datasets to addressing the practicalities of clinical deployment. This includes:

- Developing and validating models on large, diverse, and prospective datasets from multiple centers to assess generalizability [20, 24, 25].

- Focusing on the development of interpretable ML models or integrating explainability techniques (XAI) to build clinician trust [16, 58, 59].

- Designing and evaluating user interfaces that seamlessly integrate ML predictions into clinical workflows.

- Establishing robust MLOps pipelines for continuous monitoring, evaluation, and updating of deployed models [41, 42, 44, 45, 46, 47, 52].

- Collaborating closely with clinicians, patients, and policymakers to ensure that deployed solutions meet real-world needs and ethical standards [55, 56, 57].

- Exploring the use of multimodal data, including genetic information [15], environmental factors, and data from wearable devices, to further enhance predictive accuracy.

In conclusion, machine learning holds significant potential for transforming preeclampsia prediction. However, realizing this potential requires a concerted effort to move beyond research prototypes and address the multifaceted challenges of clinical deployment. By focusing on robust implementation strategies, ensuring data quality, prioritizing interpretability and ethical considerations, and adopting MLOps practices, the field can pave the way for ML-based preeclampsia prediction tools that truly make a difference in maternal healthcare, contributing to a reduction in the global burden of this serious condition.

## REFERENCES


Duley, L. The Global Impact of Pre-eclampsia and Eclampsia. Semin. Perinatol. 2009, 33, 130–137.

Magee, L.A.; Nicolaides, K.H.; von Dadelszen, P. Preeclampsia. N. Engl. J. Med. 2022, 386, 1817–1832.

Van Doorn, R.; Mukhtarova, N.; Flyke, I.P.; Lasarev, M.; Kim, K.; Hennekens, C.H.; Hoppe, K.K. Dose of aspirin to prevent preterm preeclampsia in women with moderate or high-risk factors: A systematic review and meta-analysis. PLoS ONE 2021, 16, e0247782.

O'Gorman, N.; Wright, D.; Syngelaki, A.; Akolekar, R.; Wright, A.; Poon, L.C.; Nicolaides, K.H. Competing risks model in screening for preeclampsia by maternal factors and biomarkers at 11–13 weeks gestation. Am. J. Obstet. Gynecol. 2016, 214, e1–e103.

Hackelöer, M.; Schmidt, L.; Verlohren, S. New advances in prediction and surveillance of preeclampsia: Role of machine learning approaches and remote monitoring. Arch. Gynecol. Obstet. 2023, 308, 1663–1677.

Wright, D.; Gallo, D.M.; Gil Pugliese, S.; Casanova, C.; Nicolaides, K.H. Contingent screening for preterm pre-eclampsia. Ultrasound Obstet. Gynecol. 2016, 47, 554–559.

Ranjbar, A.; Montazeri, F.; Ghamsari, S.R.; Mehrnoush, V.; Roozbeh, N.; Darsareh, F. Machine learning models for predicting preeclampsia: A systematic review. BMC Pregnancy Childbirth 2024, 24, 6.

Brunelli, V.B.; Prefumo, F. Quality of first trimester risk prediction models for pre-eclampsia: A systematic review. BJOG 2015, 122, 904–914.

Bertini, A.; Salas, R.; Chabert, S.; Sobrevia, L.; Pardo, F. Using Machine Learning to Predict Complications in Pregnancy: A Systematic Review. Front. Bioeng. Biotechnol. 2022, 9, 780389.

Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. Syst. Rev. 2021, 10, 89.

Ao, C.; Zhou, W.; Gao, L.; Dong, B.; Yu, L. Prediction of antioxidant proteins using hybrid feature representation method and random forest. Genomics 2020, 112, 4666–4674.

Moons, K.G.; Wolff, R.F.; Riley, R.D.; Whiting, P.F.; Westwood, M.; Collins, G.S.; Reitsma, J.B.; Kleijnen, J.; Mallett, S. PROBAST: A tool to assess risk of bias and applicability of prediction model studies: Explanation and elaboration. Annals of Internal Medicine. Am. Coll. Physicians 2019, 170, W1–W33.

Velikova, M.; Lucas, P.J.F.; Spaanderman, M. A Predictive Bayesian Network Model for Home Management of Preeclampsia. In Artificial Intelligence in Medicine AIME 2011 Lecture Notes in Computer Science; Peleg, M., Lavrač, N., Combi, C., Eds.; Springer: Berlin/Heidelberg, Germany, 2011; Volume 6747, pp. 179–183.

Martínez-Velasco, A.; Martínez-Villaseñor, L.; Miralles-Pechuán, L. Machine learning approach for pre-eclampsia risk factors association. In Proceedings of the 4th EAI International Conference on Smart Objects and Technologies for Social Good, Bologna, Italy, 28–30 November 2018; Association for Computing Machinery: New York, NY, USA, 2018; pp. 232–237.

Kovacheva, V.P.; Eberhard, B.W.; Cohen, R.Y.; Maher, M.; Saxena, R.; Gray, K.J. Preeclampsia Prediction Using Machine Learning and Polygenic Risk Scores from Clinical and Genetic Risk Factors in Early and Late Pregnancies. Hypertension 2024, 81, 264–272.

Eberhard, B.W.; Cohen, R.Y.; Rigoni, J.; Bates, D.W.; Gray, K.J.; Kovacheva, V.P. An Interpretable Longitudinal Preeclampsia Risk Prediction Using Machine Learning. medRxiv [Internet]. 2023 Aug 16. Available online: http://www.ncbi.nlm.nih.gov/pubmed/37645797 (accessed on 22 February 2024).

Li, Z.; Xu, Q.; Sun, G.; Jia, R.; Yang, L.; Liu, G.; Hao, D.; Zhang, S.; Yang, Y.; Li, X.; et al. Dynamic gestational week prediction model for pre-eclampsia based on ID3 algorithm. Front. Physiol. 2022, 13, 1035726.

Liu, M.; Yang, X.; Chen, G.; Ding, Y.; Shi, M.; Sun, L.; Huang, Z.; Liu, J.; Liu, T.; Yan, R.; et al. Development of a prediction model on preeclampsia using machine learning-based method: A retrospective cohort study in China. Front. Physiol. 2022, 13, 896969.

Li, Y.X.; Shen, X.P.; Yang, C.; Cao, Z.Z.; Du, R.; Wang, J.P.; Wang, M. Novel electronic health records applied for prediction of pre-eclampsia: Machine-learning algorithms. Pregnancy Hypertens. 2021, 26, 102–109.

Sufriyana, H.; Wu, Y.W.; Su, E.C.Y. Artificial intelligence-assisted prediction of preeclampsia: Development and external validation of a nationwide health insurance dataset of the BPJS Kesehatan in Indonesia. EBioMedicine 2020, 54, 102710.

Neocleous, C.; Nicolaides, K.; Neokleous, K.; Schizas, C. Ethnicity as a Factor for the Estimation of the Risk for Preeclampsia: A Neural Network Approach. In Proceedings of the Artificial Intelligence: Theories, Models and Applications: 6th Hellenic Conference on AI, SETN 2010, Athens, Greece, 4–7 May 2010; LNAI; Springer: Berlin/Heidelberg, Germany, 2010; Volume 6040.

Bennett, R.; Mulla, Z.D.; Parikh, P.; Hauspurg, A.; Razzaghi, T. An imbalance-aware deep neural network for early prediction of preeclampsia. PLoS ONE 2022, 17, e0266042.

Neocleous, C.K.; Anastasopoulos, P.; Nikolaides, K.H.; Schizas, C.N.; Neokleous, K.C. Neural networks to estimate the risk for preeclampsia occurrence. In Proceedings of the IEEE International Joint Conference on Neural Networks, Atlanta, GA, USA, 14–19 June 2009; pp. 2221–2224.

Torres-Torres, J.; Villafan-Bernal, J.R.; Martinez-Portilla, R.J.; Hidalgo-Carrera, J.A.; Estrada-Gutierrez, G.; Adalid-Martinez-Cisneros, R.; Rojas-Zepeda, L.; Acevedo-Gallegos, S.; Camarena-Cabrera, D.M.; Cruz-Martínez, M.Y.; et al. Performance of a machine learning approach for the prediction of pre-eclampsia in a middle-income country. Ultrasound Obstet. Gynecol. 2024, 63, 350–357.

Gil, M.M.; Cuenca-Gómez, D.; Rolle, V.; Pertegal, M.; Díaz, C.; Revello, R.; Adiego, B.; Mendoza, M.; Molina, F.S.; Santacruz, B.; et al. Validation of machine-learning model for first-trimester prediction of pre-eclampsia using cohort from PREVAL study. Ultrasound Obstet. Gynecol. 2024, 63, 68–74.

Ansbacher-Feldman, Z.; Syngelaki, A.; Meiri, H.; Cirkin, R.; Nicolaides, K.H.; Louzoun, Y. Machine-learning-based prediction of pre-eclampsia using first-trimester maternal characteristics and biomarkers. Ultrasound Obstet. Gynecol. 2022, 60, 739–745.