Available online at www.medicinescience.org

ORIGINAL ARTICLE

Medicine Science International Medical Journal

Medicine Science 2026;15(3):1389-95

Machine learning-based multiclass classification of preeclampsia risk using routine clinical data: An XGBoost approach

Umran Karabulut Dogan[1], Seyma Yasar[2], Erhan Huseyin Comert[3], Pinar Kadirogullari[4], Ozan Dogan[5]

[1]Acıbadem Kayseri Hospital, Department of Obstetrics and Gynecology, Kayseri, Türkiye
[2]İnönü University, Faculty of Medicine, Department of Biostatistics and Medical Informatics, Malatya, Türkiye
[3]Private Clinic, Obstetrics and Gynecology, İstanbul, Türkiye
[4]Acıbadem University, Faculty of Medicine, Department of Obstetrics and Gynecology, İstanbul, Türkiye
[5]İstanbul Nişantaşı University, Faculty of Medicine, Department of Obstetrics and Gynecology, İstanbul, Türkiye

Received 03 April 2026; Accepted 06 August 2026
Available online 25 August 2026 with doi: 10.5455/medscience.2026.04.072

Abstract

Preeclampsia is a major cause of maternal and perinatal morbidity and mortality worldwide, and early identification of high-risk pregnancies remains a clinical challenge, particularly in settings where advanced biomarkers and imaging techniques are not readily available. This study aimed to develop and evaluate a machine learning model for multiclass classification of preeclampsia risk levels using routinely available maternal, clinical, and fetal parameters. A publicly available dataset including 203 pregnant women was analyzed. Variables encompassed obstetric characteristics, maternal demographics, medical history, clinical measurements, and fetal parameters. An Extreme Gradient Boosting (XGBoost) algorithm was applied within a multiclass classification framework to categorize preeclampsia risk as low, mid, or high. Model performance was evaluated using accuracy, balanced accuracy, sensitivity, specificity, positive predictive value, negative predictive value, and F1-score. The model achieved an accuracy of 0.932 in the training dataset and 0.829 in the independent test dataset, with F1-scores of 0.926 and 0.784, respectively. Feature importance analysis identified history of hypertension, proteinuria, systolic blood pressure, and hemoglobin as the most influential predictors, whereas maternal age, parity, and fetal weight showed comparatively lower importance. These findings suggest that an XGBoost-based model using routinely available clinical variables can provide an effective approach for preeclampsia risk classification. However, external validation is required before clinical implementation.

Keywords: Preeclampsia, artificial intelligence, risk prediction, extreme gradient boosting, maternal health, pregnancy complications

Introduction

Preeclampsia (PE) is a hypertensive disorder unique to pregnancy, typically manifesting after the 20th gestational week and characterized by proteinuria and/or signs of maternal end-organ damage and compromised uteroplacental circulation. It complicates approximately 3–8% of all pregnancies worldwide and constitutes one of the foremost causes of maternal and neonatal morbidity and mortality [1,2]. The World Health Organization has documented that hypertensive complications during pregnancy are responsible for a significant proportion of maternal fatalities globally, emphasizing the critical need for reliable early detection and systematic risk assessment [1].

The pathophysiological basis of preeclampsia is multifactorial, encompassing dysregulated placentation, endothelial dysfunction, and heightened systemic inflammation. Multiple maternal risk factors have been associated with its development, including advancing maternal age, nulliparity, elevated body mass index (BMI), pre-existing diabetes mellitus, and chronic hypertension [2-4]. Arterial blood pressure-particularly systolic and diastolic readings-remains a cornerstone in both the diagnosis and ongoing

CITATION
Karabulut Dogan U, Yasar S, Comert EH, et al. Machine learning-based multiclass classification of preeclampsia risk using routine clinical data: An XGBoost approach. Med Science. 2026;15(3):1389-95.

Corresponding Author: Umran Karabulut Dogan, Acıbadem Kayseri Hospital, Department of Obstetrics and Gynecology, Kayseri, Türkiye
Email: dr.umrankarabulutdogan@gmail.com

DOI: 10.5455/medscience.2026.04.072

clinical management of the condition [3]. Established clinical guidelines from the American College of Obstetricians and Gynecologists (ACOG) and the National Institute for Health and Care Excellence (NICE) underscore the value of these parameters in stratifying women at elevated risk [3,4].

While a substantial number of predictive models have been developed for preeclampsia, many rely on specialized laboratory biomarkers such as placental growth factor (PlGF) or uterine artery Doppler velocimetry, resources that are often unavailable in routine obstetric care. Systematic analyses have demonstrated that the most widely adopted predictors in existing models-including maternal age, parity, BMI, blood pressure, and clinical history—are generally accessible in standard antenatal settings [5]. Recent investigations applying machine learning algorithms further suggest that models constructed from routine clinical variables can attain acceptable predictive accuracy, though limitations in external generalizability remain a recurring concern [6,7]. Beyond maternal factors, fetal and placental biomarkers also hold potential for refining risk estimates; proteinuria serves as a hallmark diagnostic criterion, whereas restricted fetal growth and amniotic fluid abnormalities frequently signal placental insufficiency [8,9]. Accordingly, an integrated approach incorporating gravida, parity, gestational age, maternal age, BMI, diabetes, hypertension history, blood pressure values, hemoglobin, fetal weight estimates, proteinuria, and amniotic fluid indices may yield a more thorough preeclampsia risk evaluation.

Building on these premises, the current study sought to construct and assess a multiclass risk prediction model for preeclampsia using routinely obtainable obstetric, clinical, and fetal parameters. Such a model could facilitate timely identification of high-risk pregnancies, support evidence-based clinical decision-making, and contribute to improved maternal and fetal outcomes-particularly in healthcare environments with restricted access to advanced diagnostics [1,5]. To address potential nonlinear dependencies among the predictor variables, an extreme gradient boosting (XGBoost) algorithm was selected for the classification task.

Material and Methods

Dataset

This study utilized a publicly available open-access dataset from the Kaggle platform, designated as the "Preeclampsia in Pregnant Women Dataset." The dataset contains de-identified clinical records from 203 pregnant women and encompasses a broad spectrum of obstetric, maternal, and fetal variables pertinent to preeclampsia assessment. Predictor variables include obstetric parameters (gravida, parity, gestational age in weeks), maternal demographic information (age, BMI), medical background (diabetes status, hypertension history), clinical measurements (systolic and diastolic blood pressure, hemoglobin concentrations), and fetal/biochemical indicators (estimated fetal weight, proteinuria, amniotic fluid index).

The outcome variable, termed Risk_level, reflects a categorized preeclampsia risk designation and was operationalized as a three-class variable: low, mid, and high. This served as the dependent variable throughout the modeling process. Table 1 provides a comprehensive summary of all variables included in the analysis.

Table 1. Clinical and demographic variables included in the dataset used for preeclampsia risk prediction

Variable	Type	Description	Unit
Gravida	**Integer**	Total number of pregnancies	–
Parity	**Integer**	Number of births beyond viability	–
Gestational age	**Continuous**	Gestational age at assessment	Weeks
Age	**Continuous**	Maternal age	Years
BMI	**Continuous**	Body mass index	kg/m²
Diabetes	**Categorical (Binary)**	Presence of diabetes (Yes/No)	–
History of hypertension	**Categorical (Binary)**	Previous diagnosis of hypertension (Yes/No)	–
Systolic BP	**Continuous**	Systolic blood pressure	mmHg
Diastolic BP	**Continuous**	Diastolic blood pressure	mmHg
Hemoglobin (HB)	**Continuous**	Maternal hemoglobin level	g/dL
Fetal weight	**Continuous**	Estimated fetal weight	kg
Proteinuria	**Categorical**	Presence of protein in urine (0 = absent, 1 = present)	–
Amniotic fluid levels	**Continuous**	Amniotic fluid index	cm
Risk_level	**Categorical (Multiclass)**	Preeclampsia risk classification (low, mid, high)	–

BMI: body mass index, BP: blood pressure

DOI: 10.5455/medscience.2026.04.072

Model Development

Risk level classification was performed using XGBoost, an ensemble tree-based algorithm. XGBoost operates through an additive sequential learning procedure in which each new decision tree corrects the residual prediction errors generated by its predecessors, enabling the algorithm to capture intricate nonlinear patterns and variable interactions. Due to the structured format of the dataset and minimal preprocessing demands, all predictor variables were incorporated into the model directly without prior transformation. XGBoost is inherently capable of handling mixed variable types and demonstrates robustness to feature scale differences, rendering it well-suited for datasets that combine continuous clinical measurements with categorical variables.

The classification task was formulated as a multiclass problem targeting three clinically meaningful preeclampsia risk strata: low, mid, and high. To obtain an unbiased performance estimate, the dataset was partitioned into separate training and test subsets, ensuring that model evaluation was conducted on data independent of the training process. The model was fitted on the training subset and subsequently assessed on the held-out test set. Initial runs employed default hyperparameters configurations, with model behavior monitored for robustness and generalization capacity.

Model Evaluation

Model performance was quantified using multiple classification metrics that are routinely employed in the medical informatics literature, including overall accuracy, balanced accuracy, sensitivity, specificity, and F1-score. Overall accuracy captures the proportion of correctly classified observations, while balanced accuracy adjusts for class imbalance by computing the mean of per-class recall values. Sensitivity and specificity quantify the model's capacity to correctly detect true positives and true negatives, respectively. Positive predictive value (PPV) and negative predictive value (NPV) were also computed to characterize the reliability of the model's affirmative and negative predictions. The F1-score, representing the harmonic mean of precision and recall, was employed as a composite measure to reflect overall classification balance.

Feature Importance Analysis

To determine which predictor variables exerted the greatest influence on model output, variable importance scores were extracted from the trained XGBoost model. These scores quantify the proportional contribution of each feature to the overall preeclampsia risk classification decisions.

Statistical Analysis

Continuous variables were characterized using median values alongside interquartile ranges (IQR), reflecting the non-normal distributional profiles of the data. Categorical variables were expressed as absolute frequencies and proportions. Normality of continuous variable distributions was formally evaluated with the Shapiro–Wilk test. Given that no variable satisfied the normality criterion, nonparametric statistical procedures were applied for between-group comparisons. Kruskal-Wallis tests were used to detect overall group-level differences. When a statistically significant global difference was identified, post-hoc pairwise testing was performed using Mann-Whitney U tests, with Bonferroni correction applied to adjust for the multiplicity of comparisons. All hypothesis tests were two-tailed, with a significance threshold set at $p < 0.05$. Statistical computations were conducted using IBM SPSS Statistics software.

Ethics Statement

This study utilized a publicly available, de-identified dataset. No new patient recruitment or data collection was performed. As the dataset was anonymized and freely accessible for research purposes, ethics committee approval was not required in accordance with institutional policies governing secondary analysis of publicly available datasets. Individual patient consent was not obtained as all data were de-identified prior to public release.

Results

Descriptive characteristics and group comparisons across preeclampsia risk strata are presented in Table 2.

Group comparisons across the three risk categories revealed statistically significant differences for several clinical and obstetric variables. Both gravida and parity differed across strata, with the highest median values observed among mid-risk women ($p < 0.001$). Gestational age at the time of assessment progressively increased across risk groups, reaching its highest values in the high-risk category ($p < 0.001$). Maternal age, however, did not significantly differ between groups ($p = 0.128$). BMI exhibited a pronounced elevation in the high-risk group relative to the low- and mid-risk categories ($p < 0.001$). A consistent upward trend was also observed for both systolic and diastolic blood pressure, with significantly higher readings in the mid- and high-risk groups ($p < 0.001$). Hemoglobin concentrations were highest in the high-risk group ($p < 0.001$). No statistically meaningful differences were observed for estimated fetal weight ($p = 0.872$) or amniotic fluid index ($p = 0.344$). With respect to categorical variables, diabetes, hypertension history, and proteinuria each differed significantly across the three groups ($p < 0.001$). Diabetes was entirely absent among low-risk women yet was frequently present in the mid- and high-risk groups. Both prior hypertension and proteinuria were disproportionately prevalent in the high-risk category.

DOI: 10.5455/medscience.2026.04.072

Table 2. Comparison of clinical and demographic variables across preeclampsia risk groups

Variables		Preeclampsia risk classification			p-value
		Low	Mid	High	
		Median (IQR)	Median (IQR)	Median (IQR)	
Total number of pregnancies		3[a] (1)	4[b] (2)	1[c] (2)	**< 0.001**
Number of births beyond viability		1[a] (1)	2[b] (1)	0[a] (2)	**< 0.001**
Gestational age at assessment		22.2[a] (6)	26[a] (0.2)	31[b] (4.9)	**< 0.001**
Age (yrs)		25 (3)	22 (5)	26 (10)	0.128
BMI (kg/m²)		21.2[a] (7.4)	21.3[a] (12)	29.75[b] (2.7)	**< 0.001**
Systolic BP		110[a] (10)	120[b] (20)	130[b] (10)	**< 0.001**
Diastolic BP		70[a] (0)	80[b] (10)	80[b] (10)	**< 0.001**
Maternal hemoglobin level		9.3[a] (0.9)	8.5[b] (1.9)	10.1[c] (0.6)	**< 0.001**
Estimated fetal weight		0.5[a] (0.37)	0.63[a] (0.62)	0.6[a] (1.45)	0.872
Amniotic fluid index		10[a] (3)	10[a] (3.6)	12.6[a] (4.9)	0.344
		Count (Percent)	Count (Percent)	Count (Percent)	
Presence of diabetes	**No**	107 (100.00%)	30[a] (42.86%)	12[a] (46.15%)	**< 0.001**
	Yes	0 (0.00%)	40[a] (57.14%)	14[a] (53.85%)	
History of hypertension	**No**	105[a] (98.13%)	54[b] (77.14%)	7[c] (26.92%)	**< 0.001**
	Yes	2[a] (1.87%)	16[b] (22.86%)	19[c] (73.08%)	
Proteinuria	**Absent**	107 (100.00%)	60[a] (85.71%)	6[b] (23.08%)	**< 0.001**
	Present	0 (0.00%)	10[a] (14.29%)	20[b] (76.92%)	

Values are presented as median (IQR) or n (%); Different superscript letters (a, b, c) indicate statistically significant differences between groups

Model Performance

The classification performance metrics of the XGBoost model across both training and test phases are summarized in Table 3.

Table 3. Performance of the XGBoost model for preeclampsia risk classification in training and test datasets

Metric	Training dataset	Test dataset
Accuracy	0.932	0.829
Balanced accuracy	0.965	0.839
Sensitivity (Recall)	0.946	0.839
Specificity	0.946	0.918
Positive predictive value (PPV)	0.911	0.761
Negative predictive value (NPV)	0.961	0.901
F1-score	0.926	0.784

Values represent performance metrics of the XGBoost model evaluated on training and independent test datasets; Sensitivity refers to recall; PPV: positive predictive value, NPV: negative predictive value

In the training phase, the model demonstrated strong discriminative ability, achieving an accuracy of 0.932 and a balanced accuracy of 0.946. Macro-averaged sensitivity and specificity were 0.946 and 0.965, with PPV 0.911 and NPV 0.961, and a macro F1-score of 0.926 (macro one-vs-rest AUC 0.993). On the predefined independent test set, the model achieved an accuracy of 0.829 and a balanced accuracy of 0.839 (macro AUC 0.965), with macro-averaged sensitivity 0.839, specificity 0.918, PPV 0.761, and NPV 0.901, and a macro F1-score of 0.784 (weighted F1 0.832). Under stratified 5-fold cross-validation the tuned, class-weighted XGBoost achieved a mean balanced accuracy of 0.872 (95% CI 0.805 – 0.939) and mean macro AUC of 0.967, comparable to logistic regression (0.860; 0.812 – 0.908) and random forest (0.851; 0.783 – 0.919). The perfect sensitivity obtained in the earlier single-split analysis was not reproduced under cross-validation. The relative feature importance of the predictor variables derived from the XGBoost model is presented in Figure 1. Per-class results and the confusion matrix are shown in Figure 2, and ROC curves in Figure 3.

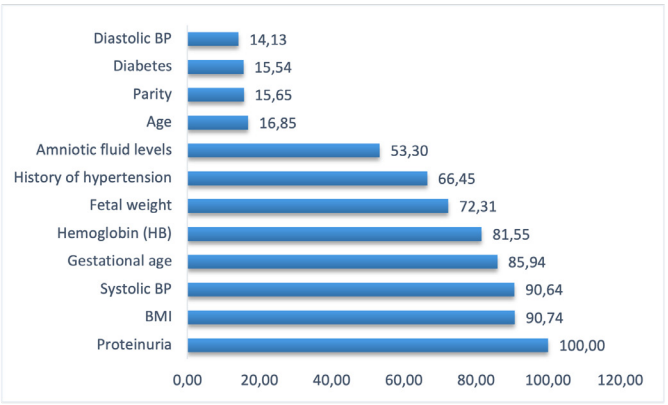

Figure 1. Relative feature importance of variables in the XGBoost model for preeclampsia risk classification

DOI: 10.5455/medscience.2026.04.072

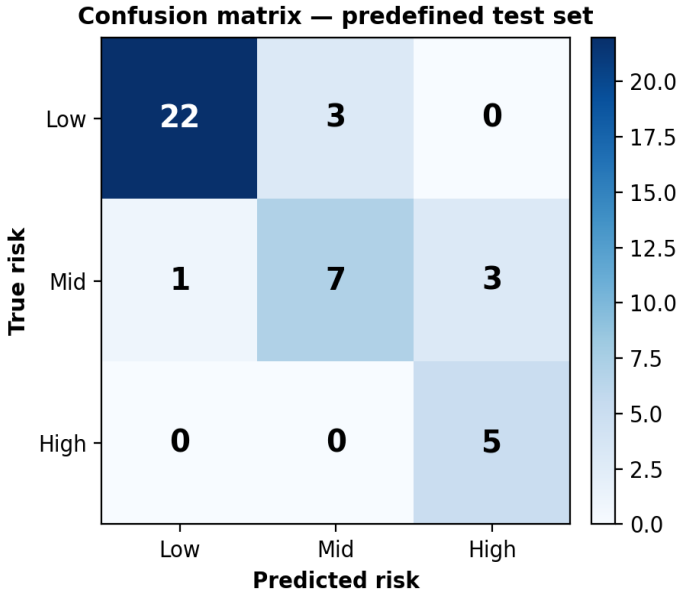

Figure 2. Confusion matrix for the tuned XGBoost model on the predefined test set

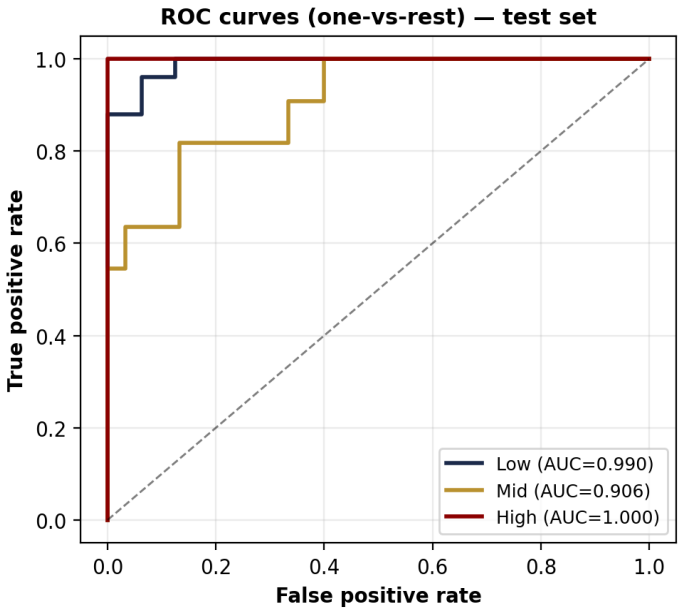

Figure 3. One-vs-rest ROC curves and AUCs on the predefined test set

Variable importance analysis (SHAP; Figure 4) indicated that history of hypertension and proteinuria ranked as the most decisive predictors in the model, followed closely by systolic blood pressure and hemoglobin, all of which made substantial contributions to risk classification. Diabetes status, BMI, and gestational age also demonstrated notable predictive relevance, whereas maternal age, parity, and fetal weight exerted comparatively limited influence. Gain-based importance showed a consistent pattern, with systolic blood pressure, proteinuria, and hemoglobin among the highest-ranked features.

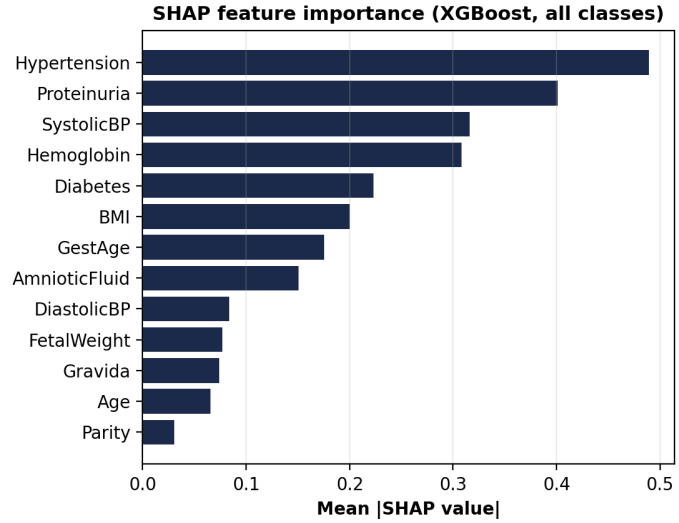

Figure 4. SHAP-based feature importance (mean absolute SHAP value across classes)

Discussion

The present investigation developed an XGBoost-based multiclass algorithm to classify preeclampsia risk into three discrete severity levels-low, mid, and high-using a set of routinely collected antenatal clinical variables. The tuned, class-weighted model achieved an accuracy of 0.932 during training and 0.829 on the predefined independent test set, with corresponding macro F1-scores of 0.926 and 0.784 and a test macro AUC of 0.965; under stratified 5-fold cross-validation the mean balanced accuracy was 0.872 (95% CI 0.805 – 0.939). These results indicate that machine learning frameworks applied to standard antenatal surveillance data can effectively stratify preeclampsia risk without necessitating specialized biomarkers or imaging-based assessment tools. This observation is consistent with a growing body of literature suggesting that ensemble learning models can deliver satisfactory classification performance while leveraging widely available clinical parameters and capturing complex, nonlinear interrelationships among predictors [10,11].

Among the predictor variables, history of hypertension and proteinuria emerged as the dominant contributors to model predictions, trailed by systolic blood pressure and hemoglobin. This ranking carries biological plausibility, as proteinuria and elevated blood pressure are integral to the clinical definition and severity grading of preeclampsia, while elevated BMI and maternal characteristics represent established antecedent risk factors. Prior machine learning investigations have similarly found that maternal hemodynamic and metabolic parameters assume a central role in classification models for hypertensive gestational disorders [11,12].

The leading predictive weight assigned to systolic blood pressure and BMI in the current model aligns with evidence from recent machine learning research demonstrating that routinely obtainable clinical features can yield clinically meaningful predictive

DOI: 10.5455/medscience.2026.04.072

power. Gradient boosting frameworks have consistently demonstrated that blood pressure metrics rank among the most informative predictors of hypertensive pregnancy complications [12]. Furthermore, a multicenter investigation reported in 2026 confirmed that machine learning models constructed from routine clinical inputs can generalize preeclampsia risk assessment across diverse population cohorts, though predictive accuracy may fluctuate with dataset heterogeneity [13].

Conversely, maternal age, parity, and diabetes status contributed minimally to the current model's output. This should not be construed as an absence of association between these variables and preeclampsia; rather, it reflects that their incremental contribution was smaller relative to markers more proximally linked to the active disease process, such as proteinuria and blood pressure. Such variability in predictor importance has been documented in recent systematic syntheses, which attribute differences to methodological factors including model architecture, feature selection strategies, and population characteristics [11].

Nonetheless, these results warrant cautious interpretation. A comprehensive systematic review and meta-analysis noted that while machine learning models for preeclampsia tend to achieve favorable internal validation metrics, considerable cross-study heterogeneity exists and external generalizability remains an unresolved challenge [11,14].

In the current study, although model performance was satisfactory on the held-out test data, the analytical cohort was limited to 203 observations sourced from a single open-access repository. External replication in independent, prospectively collected datasets is therefore a prerequisite for clinical translation [5].

Several methodological strengths distinguish this study. The exclusive reliance on routinely available obstetric and clinical measurements enhances its potential for practical implementation across varied healthcare settings. Adoption of a three-class classification architecture permits a more granular characterization of risk heterogeneity than conventional binary frameworks. In addition, transparent feature importance reporting enhances the interpretability of the model, a factor widely regarded as foundational to responsible integration of artificial intelligence in clinical workflows [10].

Certain limitations also merit acknowledgment. The relatively modest sample size constrains the statistical precision of the estimates and may reduce the generalizability of the findings to broader populations. Moreover, the reliance on pre-labeled risk categories as the outcome means that the model captures existing classification patterns rather than independently forecasting incident preeclampsia in a prospective sense. Finally, the dataset lacked advanced biomarker data, uterine artery Doppler indices, or longitudinal follow-up measurements, all of which have been linked to enhanced predictive accuracy in previously published work [12,14].

In summary, the current study demonstrates that an XGBoost model employing routine clinical and obstetric variables can classify preeclampsia risk with high accuracy. Proteinuria, BMI, and blood pressure-related features were identified as the primary predictive determinants, lending clinical credibility to the model's outputs. Future research should prioritize external validation efforts, integration of supplementary biomarkers, and head-to-head comparisons with alternative modeling approaches to more comprehensively establish the robustness and clinical utility of this methodology.

Conclusion

The findings of this study indicate that an XGBoost-based machine learning algorithm can reproduce a routine-data preeclampsia risk-stratification scheme with moderate accuracy on the predefined test set, using routinely obtainable maternal, clinical, and fetal variables. The identification of history of hypertension, proteinuria, systolic blood pressure, and hemoglobin as dominant predictive features is broadly consistent with the clinical determinants of preeclampsia, although the pre-assigned nature of the outcome label limits any prognostic interpretation. The results suggest that machine learning approaches drawing on accessible antenatal data may hold preliminary, exploratory value for risk stratification, particularly in clinical contexts where advanced diagnostic infrastructure is limited, but should not be regarded as clinically validated tools at this stage. Nevertheless, studies in larger and more heterogeneous populations, combined with rigorous external validation, will be necessary before these findings can be translated into routine clinical practice.

Ethical Approval
Since the dataset used in this study is publicly available and de-identified, ethics committee approval was not required.

Patient Consent
Individual patient consent was not obtained as all data were de-identified prior to public release.

Conflict of Interest
The authors confirm the absence of any conflicts associated with this work.

Funding
The authors declare that they have received no financial support for the study.

References

1. Dimitriadis E, Rolnik DL, Zhou W, et al. Pre-eclampsia. Nat Rev Dis Primers. 2023;9:8.

2. Johnson S, Gordijn S, Damhuis S, et al. Diagnosis and monitoring of white coat hypertension in pregnancy: an ISSHP consensus Delphi procedure. Hypertension. 2022;79:993-1005.

3. Horgan R, Diab YH, Waller J, et al. Low-dose aspirin therapy for the prevention of preeclampsia: time to reconsider our recommendations? Am J Obstet Gynecol. 2023;229:410-8.

4. Cífková R. Hypertension in pregnancy: a diagnostic and therapeutic overview. High Blood Press Cardiovasc Prev. 2023;30:289-303.

DOI: 10.5455/medscience.2026.04.072

5. De Kat AC, Hirst J, Woodward M, et al. Prediction models for preeclampsia: a systematic review. Pregnancy Hypertens. 2019;16:48-66.

6. Ranjbar A, Montazeri F, Rezaei Ghamsari S, et al. Machine learning models for predicting preeclampsia: a systematic review. BMC Pregnancy Childbirth. 2024;24:6.

7. Tiruneh SA, Vu TTT, Moran LJ, et al. Externally validated prediction models for pre-eclampsia: systematic review and meta-analysis. Ultrasound Obstet Gynecol. 2024;63:592-604.

8. Fox R, Kitt J, Leeson P, et al. Preeclampsia: risk factors, diagnosis, management, and the cardiovascular impact on the offspring. J Clin Med. 2019;8:1625.

9. Özgen G, Dincgez Cakmak B, Özgen L, et al. The role of oligohydramnios and fetal growth restriction in adverse pregnancy outcomes in preeclamptic patients. Ginekol Pol. 2022;93:235-41.

10. Layton AT. Artificial intelligence and machine learning in preeclampsia. Arterioscler Thromb Vasc Biol. 2025;45:165-71.

11. Özcan M, Peker S. Preeclampsia prediction via machine learning: a systematic literature review. Health Syst (Basingstoke). 2025;14:208-22.

12. Shyu IL, Liu CF, Tsai YC, et al. Machine learning predictive system to predict the risk of developing pre-eclampsia. BMJ Health Care Inform. 2025;32:e101151.

13. Li H, Li Y, Zang C, et al. Machine learning for dynamic and short-term prediction of preeclampsia using routine clinical data. JAMA Netw Open. 2026;9:e260359.

14. Liu L, Zhu Q, Zong Y, et al. Machine learning prediction models for preeclampsia: systematic review and meta-analysis. J Med Internet Res. 2026;28:e78714.