



## OPEN ACCESS

## EDITED BY

Stefano Restaino,  
Ospedale Santa Maria della Misericordia di  
Udine, Italy

## REVIEWED BY

Guglielmo Stabile,  
Institute for Maternal and Child Health Burlo  
Garofolo (IRCCS), Italy  
Sathya Selvarajan,  
MGM Healthcare, India

## \*CORRESPONDENCE

Zhou Zhou  
✉ evanzhou\_0826@163.com  
Jianjun Zou  
✉ zoujianjun100@126.com  
Hongjie Ruan  
✉ hongjie\_ruan@126.com

†These authors have contributed equally to  
this work and share first authorship

RECEIVED 16 October 2025

REVISED 23 November 2025

ACCEPTED 25 November 2025

PUBLISHED 09 December 2025

## CITATION

Du X, Chen Q, Lu M, Hu J, Chen C, Huang K,  
Ji C, Zhou Z, Zou J and Ruan H (2025) An  
individualized risk prediction tool for ectopic  
pregnancy within the first 10 weeks of  
gestation based on machine learning  
algorithms.  
*Front. Med.* 12:1726606.  
doi: 10.3389/fmed.2025.1726606

## COPYRIGHT

© 2025 Du, Chen, Lu, Hu, Chen, Huang, Ji,  
Zhou, Zou and Ruan. This is an open-access  
article distributed under the terms of the  
[Creative Commons Attribution License  
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction  
in other forums is permitted, provided the  
original author(s) and the copyright owner(s)  
are credited and that the original publication  
in this journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted  
which does not comply with these terms.

# An individualized risk prediction tool for ectopic pregnancy within the first 10 weeks of gestation based on machine learning algorithms

Xin Du<sup>1†</sup>, Qianping Chen<sup>2,3†</sup>, Mengmeng Lu<sup>2,3†</sup>, Jing Hu<sup>4†</sup>,  
Chen Chen<sup>3</sup>, Kaizong Huang<sup>3</sup>, Chunya Ji<sup>1</sup>, Zhou Zhou<sup>3\*</sup>,  
Jianjun Zou<sup>3\*</sup> and Hongjie Ruan<sup>1\*</sup>

<sup>1</sup>Department of Obstetrics and Gynecology, Women's Hospital of Nanjing Medical University, Nanjing Women and Children's Healthcare Hospital, Nanjing, China, <sup>2</sup>School of Basic Medicine and Clinical Pharmacy, China Pharmaceutical University, Nanjing, China, <sup>3</sup>Department of Pharmacy, Nanjing First Hospital, Nanjing Medical University, Nanjing, China, <sup>4</sup>Department of Medical Imaging, Women's Hospital of Nanjing Medical University, Nanjing Women and Children's Healthcare Hospital, Nanjing, China

**Background:** As the main cause of maternal deaths in early pregnancy, delayed diagnosis of ectopic pregnancy (EP) may lead to severe consequences. Patients with pregnancy of unknown location (PUL) exhibit a significantly higher incidence of EP and associated risks compared to the general population. Therefore, this study aims to construct an early prediction model to identify EP risk among patients with PUL and provide a valuable direction for clinical intervention.

**Methods:** Retrospectively recruited 1896 patients with PUL within 10 weeks of gestation. Feature selection was done using the least absolute shrinkage and selection operator (LASSO). Logistic Regression (LR), Extreme Gradient Boosting (XGB), Random Forest (RFC), Support Vector Machine (SVM), and CatBoost were used to construct the early risk prediction model of EP. The model's performance was evaluated by the area under the receiver operating characteristic curve (AUROC), the area under the precision-recall curve (AUPRC), and the F1 score. SHapley Additive exPlanations (SHAP) algorithms ranked the feature importance for model output interpretation.

**Results:** Among the PUL patients included in this study, 66 (4.08%) were diagnosed with EP. Key predictors selected for model construction included vaginal bleeding, progesterone, homogeneous adnexal mass, gravidity, hCG levels, history of cesarean section, abdominal tenderness, and history of pelvic surgery. Among the five models, the CatBoost algorithm demonstrated the best performance, achieving an AUROC of 0.930 (95% CI, 0.876–0.984) and an AUPRC of 0.685 (95% CI, 0.464–0.845). A user-friendly web-based platform was developed for EP risk assessment based on this model. According to SHAP analysis, the three most important clinical predictors were vaginal bleeding, progesterone levels, and the presence of a homogeneous adnexal mass.

**Conclusion:** This study employed the CatBoost algorithm to develop an individualized risk prediction model by integrating multiple features from the initial visit. This model enhances the detection rate of EP in patients with PUL during early pregnancy. Additionally, we created a web-based tool, offering potential for future clinical applications.

## KEYWORDS

first trimester, pregnancy of unknown location, ectopic pregnancy, machine learning, prediction model

## 1 Introduction

Pregnancy of unknown location (PUL) is characterized by a positive pregnancy test without a definitive embryonic location determined by ultrasound (1). Approximately 4–27% of PUL cases are eventually ultimately confirmed as ectopic pregnancy (EP) (2), a rate significantly higher than the rate compared to the 2% incidence in the general population (3). EP refers to the implantation of an embryo outside the uterine cavity and accounts for approximately 75% of early pregnancy-related deaths (4). Diagnosis typically relies on serial human chorionic gonadotropin (hCG) monitoring and transvaginal ultrasound (TVS) (5–7). However, delayed diagnosis during follow-up can lead to postponed treatment, potentially resulting in severe consequences such as tubal rupture and hemorrhagic shock. While advanced diagnostic techniques such as contrast-enhanced (CE)—magnetic resonance imaging (MRI) (8), intrauterine genomic classifier (9), and biomarkers like miR-519d and sFLT-1 (10, 11) demonstrate promising potential but are restricted by inaccessibility and invasiveness. Consequently, developing convenient, non-invasive, and cost-effective methods for early EP detection remains an urgent priority.

The Society of Obstetricians and Gynaecologists of Canada (SOGC) guidelines on managing PUL recommend utilizing prediction models to facilitate clinical decisions (1). The M4 and M6 are the most extensive models for predicting risk. Both are logistic regression (LR) models that consist of initial serum progesterone, hCG level, and the serum hCG ratio (hCG 48 h/hCG 0 h) (12–15). However, linear models struggle to capture nonlinear relationships among complex clinical factors, impacting risk stratification accuracy. Additionally, low patient compliance and extended follow-up periods increase the risk of adverse outcomes. Machine learning (ML) has proved to be effective for more accurate and personalized medical diagnoses by leveraging large datasets and pattern recognition (16–18). Recent investigations have developed various EP prediction models using ML techniques, exhibiting superior diagnostic performance compared to conventional approaches (19–21). Nonetheless, they are limited by factors such as reliance on single indicators and difficulty in balancing promptness and precision, which restricts their clinical utility.

This study developed an early risk prediction model for PUL by integrating initial clinical assessment data to facilitate triage of high-risk patients. The findings offer a theoretical and methodological foundation for future clinical decision-support tools.

## 2 Methods

### 2.1 Study population

This study strictly followed the TRIPOD guidelines (22). The study population consisted of 1,896 patients diagnosed with PUL at the Women's Hospital of Nanjing Medical University from January 1, 2020 to July 19, 2020. All patients included in the study had a gestational age of less than 10 weeks, presented with biochemically

confirmed pregnancy, but lacked an identifiable gestational sac location, which are clinically defined as PUL. Patients were excluded if they had (1) heterotopic pregnancy, (2) hemodynamic instability, (3) test results from other medical institutions, (4) previous surgical or medical treatment, or (5) identification of heterogeneous adnexal mass, which was highly suspicious of EP or adnexal mass of unknown nature. Ethical approval for this study (ID: 2023KY-150) was provided on January 3, 2024, by the Medical Ethics Committee of the Women's Hospital of Nanjing Medical University (Chair: Yanjing Kan). As the study was retrospective, the ethics committee waived the requirement for informed consent. The flow chart of this study is shown in Figure 1.

### 2.2 Data collection

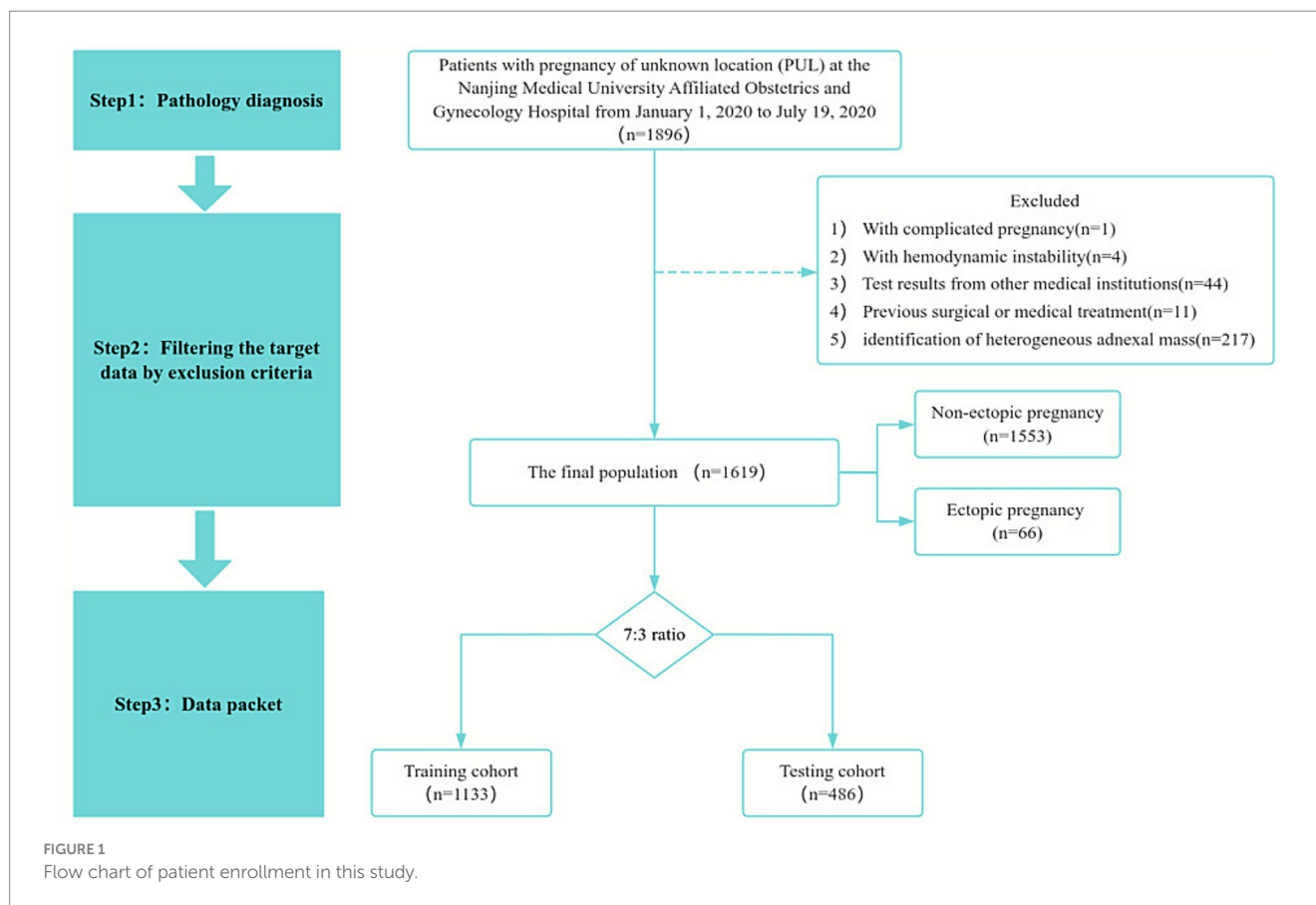
Clinical data were collected retrospectively from electronic health records in a structured format. In this study, the variables included were baseline characteristics (age, gravidity, parity, abortion), medical history (chronic diseases, gynecological diseases, EP, surgical history, emergency contraceptive pills (ECPs), assisted reproductive technology (ART)), clinical symptoms (abdominal pain, vertigo, diarrhea, abdominal tenderness, cervical motion tenderness, vaginal bleeding), ultrasound findings (homogeneous adnexal mass, pelvic effusion), and serologic marker tests (hCG, progesterone). A homogeneous adnexal mass was defined as an adnexal mass that exhibits uniform internal echogenicity, e.g., cystic or solid masses (23). The serum levels of hCG and progesterone were assessed during the patient's initial visit. The measurements of hCG and progesterone were conducted in a designated laboratory (Medical Laboratory of Women's Hospital of Nanjing Medical University) using a fully automated chemiluminescent analyzer and its corresponding reagents (Beckman Coulter, Inc.; Brea, CA, United States). Additionally, in this study, the definition of vaginal bleeding was based on the amount of bleeding compared to menstrual flow, therefore patients were categorized into three groups: no bleeding, bleeding less than, and equivalent menstrual flow.

### 2.3 Sample size calculation

In this research, the sample size of the binary outcome prediction model was calculated according to the following formula:

$$N = \exp\left(\frac{-0.508 + 0.259 \ln(\varphi) + 0.504 \ln(P) - \ln(\text{MAPE})}{0.544}\right) \quad (24). \text{ In}$$

this formula,  $\varphi$  is the anticipated outcome proportion ( $\varphi = 0.0408$ ),  $P$  is the number of candidate predictor parameters ( $p = 8$ ), and MAPE is the average absolute error between the observed and true outcome probability (MAPE = 0.05). According to calculations, the minimum sample size required for the training set is 145. The total data were randomly divided into the training set and the testing set in a ratio of 7:3, the minimum total effective sample size required is 208.



## 2.4 Outcome

After the entire period of pregnancy follow-up, the final diagnosis was categorized as either EP or non-EP (including normal pregnancy, threatened abortion, spontaneous abortion, missed abortion or inevitable abortion). The outcome of this study was EP, and evidence for diagnosis of EP included (1) ultrasound suggesting the presence of a gestational sac containing a yolk sac or fetal pole (with or without heartbeat) outside the uterine cavity. (2) postoperative pathology confirmation of chorionic villi outside the uterine cavity (1, 4).

## 2.5 Data preprocessing and feature selection

Proper data preprocessing is essential before analysis. In this study, all variables had a completion rate of 100%, except for progesterone (88.2%). Missing values for progesterone were imputed using the k-nearest neighbors (KNN) algorithm (25). KNN fills in missing values by calculating the optimal number of neighbors. To ensure consistency across features, all continuous variables were standardized using z-score normalization, while categorical variables were transformed via one-hot encoding (26). Subsequently, all patients were randomly divided into training and testing cohorts in a 7:3 ratio, ensuring a similar incidence of EP in each cohort. Univariate analysis was performed to identify variables significantly associated with EP ( $p < 0.05$ ). To reduce multicollinearity and improve model performance, the Least Absolute Shrinkage and Selection Operator

(LASSO) was employed for feature selection (27). The variance inflation factor (VIF) was subsequently calculated for the selected variables, with  $VIF < 5$  indicating no significant multicollinearity (28).

## 2.6 Model development and evaluation

Five ML algorithms were employed to develop predictive models for EP among patients with PUL: Logistic Regression (LR), Extreme Gradient Boosting (XGB), Random Forest Classifier (RFC), Support Vector Machine (SVM), and CatBoost. Hyperparameters for each model were optimized on the training set using stratified 10-fold cross-validation with a combination of grid search and manual tuning, with the goal of maximizing the area under the precision-recall curve (AUPRC). To preserve the real-world clinical distribution, no oversampling methods (e.g., SMOTE) were applied to balance the outcome classes. Instead, automatic class weight adjustment was utilized during model training. Model performance was evaluated on the testing cohort using primary metrics, including AUPRC, the area under the receiver operating characteristic curve (AUROC), F1 score, sensitivity and specificity. The optimal classification threshold for each model was determined using the Youden index (Youden index = Sensitivity + Specificity - 1). AUROC point estimates, and 95% confidence intervals (CI) were computed using DeLong's nonparametric method. For AUPRC, sensitivity, specificity, precision and F1-score, 95% CIs were estimated by non-parametric bootstrap resampling (2,000 replicates) using the percentile method. Given the class imbalance in the dataset, AUPRC was emphasized for its sensitivity to

TABLE 1 Demographic characteristics of patients in the whole cohort.

Variables	Overall (n = 1,619)	Non-EP n = 1,553, 95.92%	EP n = 66, 4.08%	P
<b>Demographics</b>				
Age, year	30.00 [27.00, 33.00]	30.00 [27.00, 33.00]	31.50 [28.00, 34.00]	0.144
Gravidity, number of pregnancies	1.00 [0.00, 2.00]	1.00 [0.00, 2.00]	2.00 [0.00, 3.00]	<0.001
Parity, number of deliveries	0.00 [0.00, 1.00]	0.00 [0.00, 1.00]	0.00 [0.00, 1.00]	0.029
Abortion, number of abortions	0.00 [0.00, 1.00]	0.00 [0.00, 1.00]	1.00 [0.00, 2.00]	<0.001
<b>Comorbidities</b>				
History of EP, n (%)	39 (2.4)	33 (2.1)	6 (9.1)	0.001
History of laparotomy, n (%)	7 (0.4)	7 (0.5)	0 (0.0)	1.000
History of pelvic surgery, n (%)	50 (3.1)	41 (2.6)	9 (13.6)	<0.001
History of cesarean section, n (%)	171 (10.6)	155 (10.0)	16 (24.2)	<0.001
History of uterine surgery, n (%)	14 (0.9)	13 (0.8)	1 (1.5)	1.000
ECPs, n (%)	16 (1.0)	14 (0.9)	2 (3.0)	0.281
ART, n (%)	29 (1.8)	28 (1.8)	1 (1.5)	1.000
Uterine fibroid, n (%)	148 (9.1)	139 (9.0)	9 (13.6)	0.282
Endometriosis, n (%)	10 (0.6)	9 (0.6)	1 (1.5)	0.882
Polycystic ovary syndrome, n (%)	11 (0.7)	11 (0.7)	0 (0.0)	1.000
Cesarean scar diverticulum, n (%)	4 (0.2)	4 (0.3)	0 (0.0)	1.000
Vaginitis, n (%)	114 (7.0)	112 (7.2)	2 (3.0)	0.291
PID, n (%)	10 (0.6)	7 (0.5)	3 (4.5)	0.001
IUA, n (%)	3 (0.2)	2 (0.1)	1 (1.5)	0.270
CUA, n (%)	14 (0.9)	12 (0.8)	2 (3.0)	0.207
Cervical polyp, n (%)	10 (0.6)	9 (0.6)	1 (1.5)	0.882
Hypertension, n (%)	2 (0.1)	2 (0.1)	0 (0.0)	1.000
Diabetes, n (%)	3 (0.2)	2 (0.1)	1 (1.5)	0.270
Thyroid diseases, n (%)	24 (1.5)	22 (1.4)	2 (3.0)	0.587
<b>Symptoms</b>				
Abdominal pain, n (%)	593 (36.6)	565 (36.4)	28 (42.4)	0.386
Vertigo, n (%)	3 (0.2)	2 (0.1)	1 (1.5)	0.270
Diarrhea, n (%)	5 (0.3)	5 (0.3)	0 (0.0)	1.000
Abdominal tenderness, n (%)	24 (1.5)	13 (0.8)	11 (16.7)	<0.001
Cervical motion tenderness, n (%)	6 (0.4)	3 (0.2)	3 (4.5)	<0.001
Vaginal bleeding (compare with menstrual flow), n (%)				<0.001
None	887 (54.8)	873 (56.2)	14 (21.2)	
Less	657 (40.6)	609 (39.2)	48 (72.7)	
Equivalent	75 (4.6)	71 (4.6)	4 (6.1)	
<b>Ultrasound findings</b>				
Homogeneous adnexal mass, n (%)	205 (12.7)	173 (11.1)	32 (48.5)	<0.001
Pelvic effusion, cm	0.00 [0.00, 0.00]	0.00 [0.00, 0.00]	0.00 [0.00, 0.84]	<0.001
Intrauterine echoes, cm	0.58 [0.00, 1.04]	0.60 [0.00, 1.04]	0.00 [0.00, 1.15]	0.354
<b>Serum marker</b>				
hCG, n (%)				<0.001
hCG < 1,000, mIU/ml	472 (29.2)	446 (28.7)	26 (39.4)	

(Continued)

TABLE 1 (Continued)

Variables	Overall	Non-EP	EP	P
	(n = 1,619)	n = 1,553, 95.92%	n = 66, 4.08%	
1,000 ≤ hCG < 2000, mIU/ml	163 (10.1)	145 (9.3)	18 (27.3)	
2,000 ≤ hCG < 3,000, mIU/ml	124 (7.7)	119 (7.7)	5 (7.6)	
3,000 ≤ hCG < 4,000, mIU/ml	92 (5.7)	91 (5.9)	1 (1.5)	
4,000 ≤ hCG < 5,000, mIU/ml	70 (4.3)	65 (4.2)	5 (7.6)	
hCG ≥ 5,000, mIU/ml	698 (43.1)	687 (44.2)	11 (16.7)	
Progesterone, ng/ml	16.16 [10.60, 21.82]	16.36 [10.97, 21.98]	9.21 [4.82, 14.70]	<0.001

Continuous variables were summarized as median (interquartile range, IQR) and categorical variables were summarized as frequencies and percentages. EP, ectopic pregnancy; ECPs, emergency contraceptive pills; ART, assisted reproductive technology; PID, pelvic inflammatory disease; IUA, intrauterine adhesion; CUA, congenital uterine anomaly; hCG, human chorionic gonadotrophin.

positive class performance (29, 30). The F1 score, representing the harmonic mean of precision and recall, was used to summarize model performance at the cut-off. Sensitivity measures the proportion of actual EP cases correctly identified by the model, reflecting its ability to detect high-risk patients. Specificity indicates the proportion of non-EP cases correctly classified, reflecting the model’s ability to avoid unnecessary interventions in low-risk patients. These two metrics are particularly important in imbalanced datasets, where missing a true EP case (false negative) can have severe clinical consequences, while excessive false positives may lead to unnecessary anxiety or diagnostic procedures. Model calibration was measured by the calibration curve and brier score. Calibration curve was plotted to assess how well predicted probabilities matched observed outcomes. The Brier score is the average squared distance between the predicted probability of the outcome and the true label. Decision curve analysis (DCA) was performed to evaluate the clinical usefulness of each prediction model by quantifying the net benefit across a range of threshold probabilities. To improve the stability of estimates, bootstrap resampling (n = 1,000) was applied to obtain 95% CIs for each model’s net benefit curve. Based on these comprehensive evaluations, the best-performing model was selected for further interpretation. All models were implemented using

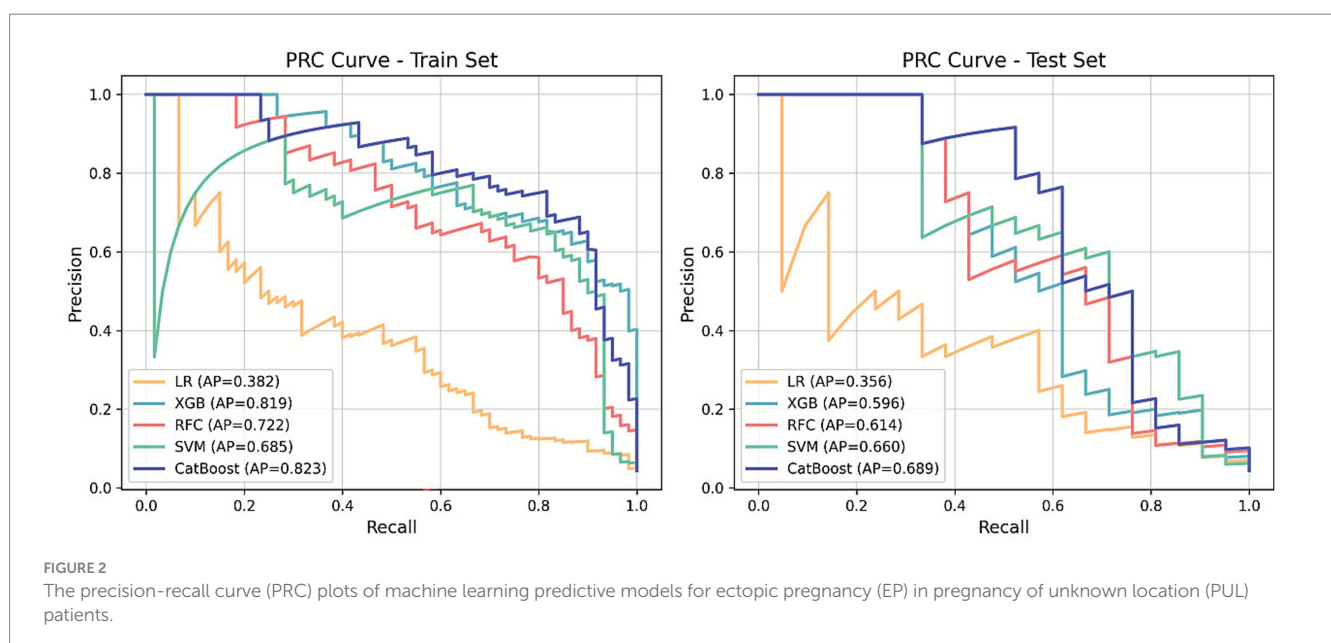
Python version 3.12.4, with libraries including Scikit-learn version 1.1.2, XGBoost version 1.0.3, and Keras version 3.5.0.

### 2.7 Model interpretation

The SHapley Additive exPlanations (SHAP) algorithm was employed to enhance the interpretability of the best-performing model, addressing the longstanding “black box” nature of ML models (29). SHAP summary plots were used to display the global importance of each feature based on average absolute SHAP values. Each point on the SHAP scatter plot represents the impact of a given feature on the model’s output for an individual patient, indicating whether the feature increases or decreases the predicted risk of EP. SHAP analysis was performed in Python 3.12.4 using the SHAP package (v0.46.0).

### 2.8 Statistical analysis

Continuous variables were assessed for normality using the Shapiro–Wilk test. Normally distributed data were presented as mean



± standard deviation (SD) and compared using the Student's *t*-test. Non-normally distributed variables were expressed as median (interquartile range, IQR) and analyzed using the Mann–Whitney U test. Categorical variables were summarized as frequencies and percentages, and comparisons were made using the Chi-square test or Fisher's exact test, as appropriate. All statistical analyses were performed using R version 4.4.0. Two-tailed *p*-values <0.05 were considered statistically significant.

median age was 30 years (IQR: 27–33), and 39 patients (2.4%) reported a prior history of EP. Patients were randomly assigned to a training cohort (1,133 patients, 70%) and a testing cohort (486 patients, 30%). No significant differences in baseline characteristics were observed between the two groups (Supplementary Table S1). The overall incidence of EP was 4.08% (66/1,619), with 4.0% (45/1,133) in the training cohort and 4.3% (21/486) in the testing cohort.

### 3 Results

#### 3.1 Baseline characteristics and outcome

A total of 1,619 patients met the inclusion and exclusion criteria (Figure 1). Baseline characteristics are summarized in Table 1. The

#### 3.2 Feature selection and model development

Missing values were imputed using the KNN algorithm, with detailed imputation statistics presented in Supplementary Table S1. Fourteen variables (Table 1) were initially included in the LASSO

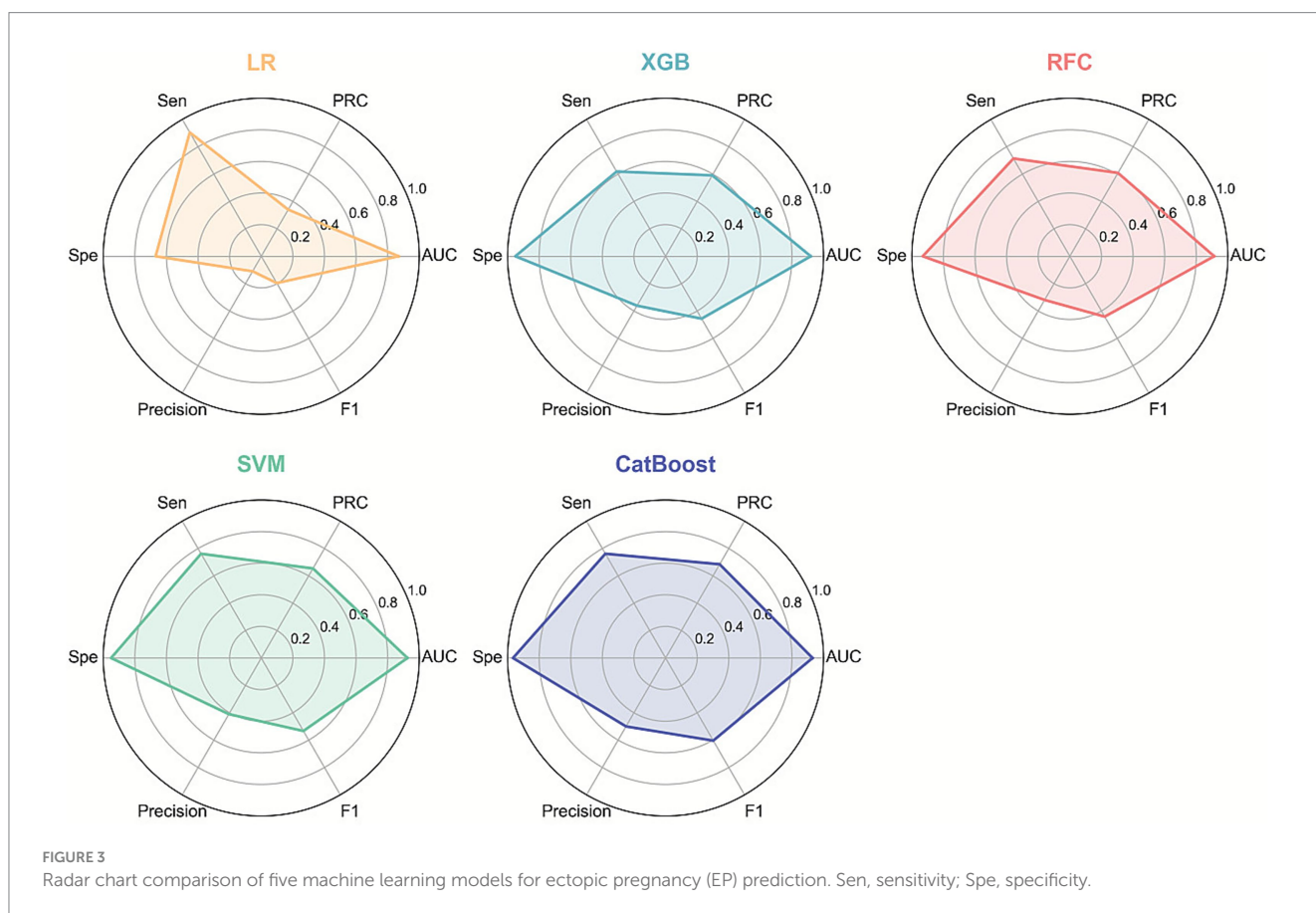


TABLE 2 Performance of five machine learning models in the testing cohort.

Model	Cut-off	AUROC (95%CI)	AUPRC (95%CI)	F1 score (95%CI)	Sensitivity (95%CI)	Specificity (95%CI)	Precision (95%CI)	Brier score
LR	0.356	0.869 (0.794–0.944)	0.341 (0.173–0.560)	0.196 (0.123–0.275)	0.905 (0.762–1.000)	0.669 (0.627–0.711)	0.110 (0.066–0.160)	0.137
XGB	0.583	0.919 (0.862–0.976)	0.591 (0.377–0.775)	0.456 (0.281–0.603)	0.620 (0.409–0.824)	0.951 (0.930–0.970)	0.361 (0.205–0.526)	0.057
RFC	0.455	0.913 (0.852–0.975)	0.609 (0.393–0.789)	0.441 (0.281–0.571)	0.714 (0.524–0.895)	0.931 (0.909–0.952)	0.319 (0.186–0.453)	0.063
SVM	0.136	0.925 (0.850–1.000)	0.654 (0.435–0.823)	0.533 (0.367–0.678)	0.762 (0.571–0.944)	0.951 (0.930–0.970)	0.410 (0.256–0.575)	0.025
CatBoost	0.611	0.930 (0.829–1.000)	0.685 (0.493–0.863)	0.604 (0.432–0.750)	0.762 (0.571–0.944)	0.966 (0.949–0.981)	0.500 (0.323–0.680)	0.064

LR, logistic regression; XGB, extreme gradient boosting; RFC, random forest classifier; SVM, support vector machine.

regression, which ultimately selected eight key predictors: gravidity, vaginal bleeding, hCG, progesterone, homogeneous adnexal mass, history of cesarean section, history of pelvic surgery, and abdominal tenderness. The corresponding non-zero coefficients are provided in [Supplementary Table S2](#). VIF analysis indicated no significant multicollinearity among these predictors ([Supplementary Table S3](#)).

### 3.3 Model evaluation and comparison

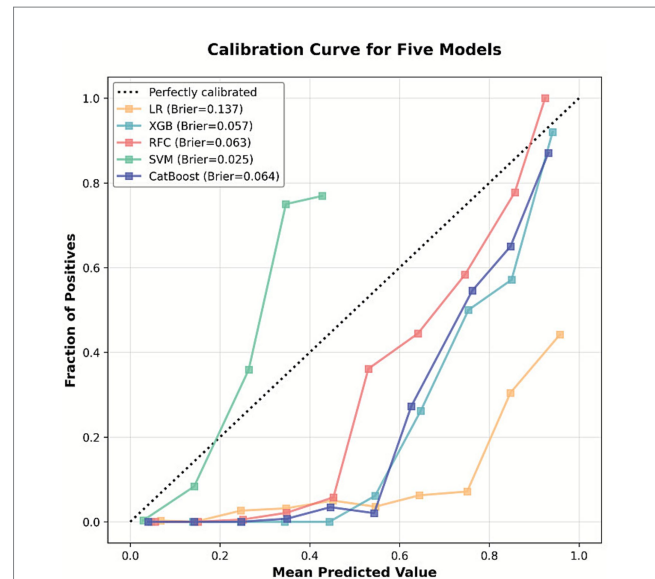
Five ML algorithms—LR, XGB, RFC, SVM, and CatBoost—were trained to predict EP in patients with PUL. Optimal hyperparameter configurations are listed in [Supplementary Table S4](#). Performance metrics are visualized in PRC plots ([Figure 2](#)) and radar charts ([Figure 3](#)). The cut-off calculated via the Youden Index and the corresponding performance metrics for all models in the testing cohort are presented in [Table 2](#). Among the models, CatBoost achieved the highest performance, with AUPRC of 0.685 (0.493–0.863), AUPRC of 0.930 (0.829–1.000), F1 score of 0.604 (0.432–0.750), sensitivity of 0.762 (0.571–0.944), specificity of 0.966 (0.949–0.981), precision of 0.500 (0.323–0.680), and brier score of 0.064. The calibration curve ([Figure 4](#)) showed an acceptable level of calibration for most models, albeit with a tendency to overestimate risk across the predicted probability spectrum. The DCA ([Supplementary Figure S1](#)) curve revealed that compared with the “Treat All” and “Treat None” strategies, the CatBoost model provided more clinical utility for EP risk stratification ([Supplementary Figure S1](#)). At the cut-off of 0.611, the confusion matrix ([Figure 5](#)) on the testing cohort ( $n = 486$ ; EP cases = 21) included 16 true positives, 5 false negatives, 427 true negatives, and 38 false positives ([Figure 5](#)). The CatBoost model correctly identified 76.2% of EP cases (16/21). Given the importance of maintaining high sensitivity while limiting false positives in imbalanced datasets, CatBoost demonstrated an effective balance and was selected as the final predictive model.

### 3.4 Sensitivity analyses

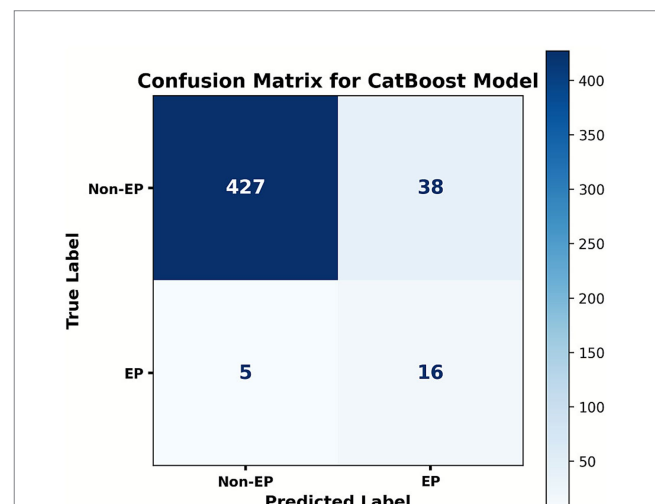
To evaluate the robustness of our findings to different missing value handling strategies, we repeated model development using two alternative imputation methods (median imputation and multivariate imputation by chained equations [MICE]), in addition to the primary k-nearest neighbors (KNN) imputation. Model performance under these three approaches is summarized in [Supplementary Tables S5, S6](#). Results showed no statistically significant differences across imputation strategies (all  $p > 0.05$ ), indicating that our conclusions were not sensitive to the choice of missing value imputation method.

### 3.5 Feature importance visualization

Based on model evaluation, the CatBoost model was confirmed as the optimal predictor. SHAP summary plots (bar and dot) were generated to illustrate feature contributions ([Figure 6](#)). Vaginal bleeding, progesterone, and homogeneous



**FIGURE 4**  
Calibration curve for five machine learning models. The calibration curves depict the agreement between predicted probabilities and actual outcomes in predicting ectopic pregnancy (EP). The diagonal dashed line represents perfect calibration.



**FIGURE 5**  
Confusion matrix of the CatBoost model at the cut-off of 0.611. Ordinates represent the actual results, and the abscissa represents the model's predictive results. The matrix summarizes the case counts of TP (true positive), FN (false negative), FP (false positive), and TN (true negative).

adnexal mass were the top three predictors based on SHAP values. To further explore potential non-linear effects of individual predictors, we plotted univariable SHAP dependence plots ([Supplementary Figure S2](#)). The SHAP dependence plot indicated that less vaginal bleeding (relative to menstrual flow), lower progesterone levels and presence of a homogeneous adnexal mass were associated with increased EP risk. Additionally, a history of cesarean section or pelvic surgery further elevated EP risk in the model's predictions.

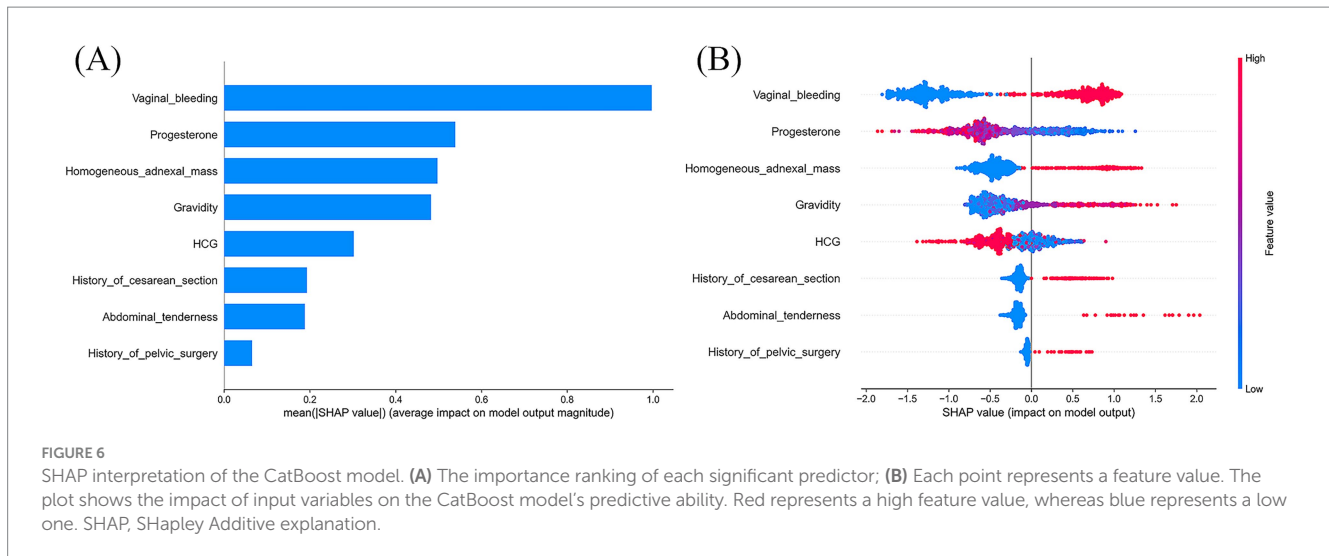


FIGURE 6

SHAP interpretation of the CatBoost model. **(A)** The importance ranking of each significant predictor; **(B)** Each point represents a feature value. The plot shows the impact of input variables on the CatBoost model's predictive ability. Red represents a high feature value, whereas blue represents a low one. SHAP, SHapley Additive explanation.

### 3.6 Web-based prediction platform

To facilitate clinical application, a user-friendly web platform (Figure 7) was developed for EP risk prediction: <https://kclba5qc7vr7gyoa8kjakj.streamlit.app/>. Users can input relevant clinical data to receive a personalized EP risk prediction, categorized as low (0) or high (1) risk. The platform also provides SHAP-based visual explanations, including force plots, to highlight key factors influencing individual predictions.

## 4 Discussion

In this study, we established an early prediction model to identify EP risk among patients with PUL, utilizing five ML algorithms, including LR, XGB, RFC, SVM, and CatBoost. Among them, the CatBoost model had the highest performance, with a PRC of 0.685 (95% CI, 0.464–0.845), a sensitivity of 76.2%, a specificity of 96.6%, and an F1 score of 0.604. Feature interpretation was conducted using SHAP, which enabled ranking the importance of eight predictors. The top five clinical factors contributing to EP risk were: the amount of vaginal bleeding, serum progesterone level, presence of adnexal mass, gravidity, and hCG concentration.

Innovatively employing baseline clinical data, the CatBoost model helped clinicians complete early EP risk stratification at the patient's initial visit. Conventionally, previous models (30, 31) relied on serial hCG monitoring over 48 h to 7 days, which might delay clinical decisions and compromise patient compliance. From a pathophysiological perspective, trophoblast dysfunction in EP patients stemmed from abnormal embryo implantation, leading to significantly lower early hormone levels than normal intrauterine pregnancies (32, 33). Therefore, baseline hCG and progesterone, as direct indicators of trophoblast development, were valuable for initial EP risk assessment (1). To some extent, the model aided in preventing delayed intervention in high-risk patients and mitigated the risk of fatal complications, such as tubal rupture and hemorrhagic shock, in regions with constrained healthcare resources or high patient mobility, this model had the potential to minimize the probability of missed

diagnosis and loss to follow-up, thereby improving diagnostic timeliness and clinical accessibility.

Notably, this study focused on PUL patients within the first 10 weeks of gestation, rather than the traditional 14-week definition (34). During early pregnancy, hCG levels followed a characteristic trajectory: serum  $\beta$ -hCG became detectable approximately 10 days post-fertilization, rose exponentially around week 5, peaked at 9–10 weeks, and then gradually declined to a plateau in the second and third trimesters (35, 36). Early hCG levels were critical for umbilical cord development, inhibition of uterine contractions, fetal organogenesis, angiogenesis, and immune tolerance regulation (37). Therefore, we selected a predictive time window within the first 10 weeks to capture subtle fluctuations in early hCG dynamics and evaluate relevant risk factors.

Through the integration of multiple features, this study investigated the synergistic predictive value of baseline serum biomarkers and clinical symptoms to better understand disease progression and optimize clinical management in the PUL population. We found that higher initial hCG and progesterone levels did not indicate an increased risk of EP. Conversely, EP patients generally exhibited lower levels compared to non-EP individuals. From a pathophysiological perspective, implantation in EP occurred outside the uterus, predominantly in the fallopian tube. When the trophoblast was in a suboptimal environment, hCG and progesterone synthesis and secretion might reduce, leading to hormonal imbalance. The hormonal imbalance impaired endometrial maintenance, this could cause decidual tissue breakdown and shedding, which typically manifested clinically as vaginal bleeding or spotting (38, 39). This was why irregular vaginal bleeding, especially when bleeding volume was less than normal menstrual blood loss, might suggest a higher risk of EP in PUL patients. Interestingly, the SHAP revealed that the contribution of vaginal bleeding to EP risk prediction was higher than that of hCG and progesterone levels. Meanwhile, several previous studies corroborated our findings (40–42). Moreover, while gravidity is a well-established risk factor, its confounding associations with other characteristics must also be considered, as they may significantly influence outcome prediction.

# Ectopic Pregnancy Risk Assessment

Gravidity:

6 - +

History of pelvic surgery:

Yes v

History of cesarean section:

No v

Abdominal tenderness:

Yes v

Homogeneous adnexal mass:

Yes v

Vaginal bleeding (compare with menstrual flow):

Less v

hCG(mIU/ml):

hCG < 1000 v

Progesterone(ng/ml):

15.00 - +

Predict

**Predicted Probability: 78.8%**

**Predicted Class: 1 (1: High risk of EP, 0: Low risk of EP)**

## Model Explanation

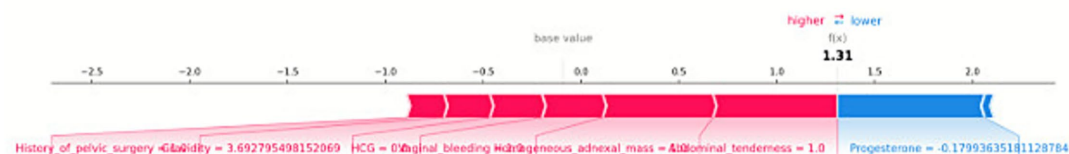


FIGURE 7 Web-based prediction platform for ectopic pregnancy (EP) risk assessment.

Despite the low prevalence of EP in the PUL population (4.08%), relying on the multidimensional optimization of sensitivity, specificity, and precision, the model achieved good predictive performance. Specifically, the 96.6% specificity effectively excluded non-EP cases, thereby minimizing unnecessary examinations, conserving medical resources, and alleviating patient anxiety. Although the model's sensitivity (76.2%) was lower than that of the M4 (80.0–86.4%) and M6NP (90.6–95.0%) models (12–15, 43, 44), it still captured over 75% of potential EP cases, providing a critical time window for early intervention. Additionally, our CatBoost model achieved a precision of 50%, which was higher than that of the M6NP model (18.7%) and the low-accuracy subgroup of the M4 model (37.8%). Notably, although the LR model demonstrated superior sensitivity, the CatBoost model achieved a more reasonable balance between sensitivity and precision. In EP prediction, calibration curves evaluating the CatBoost model revealed systematic overestimation. Given that the primary objective of this study was to achieve early identification of potential EP cases and thereby prevent missed diagnoses that might lead to life-threatening complications. Thus, prioritization of higher predicted risk was considered preferable to underestimation. This approach aids in identifying high-risk patients and provides valuable guidance for clinical intervention. Notably, DCA demonstrated that the model's net benefit within the threshold probability range commonly used in clinical practice was significantly superior to both the “Treat All” and “Treat None” strategies.

False-negative and false-positive results were used to assist clinicians in identifying the causes of model prediction bias, thereby improving the accuracy and safety of EP diagnosis. In the CatBoost model's testing set, 5 false-negative cases were associated with hCG and progesterone levels at diagnostic cutoffs, absence of typical ultrasonographic findings, and atypical early EP symptoms. Meanwhile, 38 false-positive cases were attributed to three core factors: highly analogous clinical phenotypes, comorbidity interference, and predictive feature interactions. For instance, concurrent benign adnexal masses or pelvic inflammatory disease were erroneously captured as positive signals, resulting in false positives. Therefore, given the complexity of EP pathogenesis and the limitations of early clinical detection methods, prediction errors were inevitable. To address this issue, the AUPRC was chosen as a more suitable evaluation metric (45). By balancing precision and recall, the AUPRC minimized the risk of missed diagnoses and false positives, ensuring the model's predictive performance was both reliable and stable. Importantly, the model was intended as a clinical decision-support tool, not as a replacement for clinicians' professional judgment. It provided objective risk assessment references to aid clinicians in diagnosing and making decisions about EP. Despite unavoidable misclassifications, the model demonstrated strong overall performance (AUROC: 0.930, AUPRC: 0.685) and remained viable for widespread clinical use.

Nevertheless, this study has several limitations. Firstly, as a single-center retrospective analysis, the demographic characteristics of the study population were relatively homogeneous. Previous studies had shown that populations from different regions and social backgrounds demonstrate substantial variations in the incidence and prognosis of EP (46, 47). Secondly, the number of positive cases was limited. However, methodological optimizations were applied to partially mitigate its potential impact. Thirdly, as this study centered on initial-visit assessments, serial  $\beta$ -hCG measurements were not incorporated.

Finally, certain characteristics were absent, including the lack of EP cases occurring in rare locations such as the interstitial segment or ovary, which could result in severe outcomes when diagnosis was delayed (48, 49). Furthermore, due to sample size limitations, the model incorporated neither unobserved nor undocumented symptoms from our patient cohort. Future multi-center prospective studies are planned to have the sample size expanded, a more diverse patient cohort enrolled, and comprehensive baseline clinical data collected, including dynamic serological markers and multidimensional clinical symptoms, thereby enhancing the external validity and robustness of the developed model.

## 5 Conclusion

We have devised a personalized risk prediction model utilizing the CatBoost algorithm to facilitate the early detection of EP within the PUL population. This model incorporates eight crucial predictors including vaginal bleeding, progesterone level, and adnexal mass. It differentiates non-EP cases and identifies potential EP cases. The importance of these predictors was illustrated through SHAP summary plots. Furthermore, a web platform was created as a proof-of-concept tool, setting the stage for the rigorous external validation and prospective studies prior to clinical application.

## Data availability statement

The original contributions presented in the study are included in the article/[Supplementary material](#), further inquiries can be directed to the corresponding authors.

## Ethics statement

Ethical approval for this study (ID: 2023KY-150) was provided on January 3, 2024, by the Medical Ethics Committee of the Women's Hospital of Nanjing Medical University. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

## Author contributions

XD: Writing – review & editing, Investigation, Conceptualization, Writing – original draft, Data curation, Methodology. QC: Validation, Writing – original draft, Visualization, Formal analysis, Software, Writing – review & editing. ML: Formal analysis, Writing – original draft, Writing – review & editing. JH: Writing – review & editing, Writing – original draft, Investigation. CC: Writing – review & editing. KH: Writing – review & editing, Formal analysis. CJ: Writing – review & editing, Data curation. ZZ: Writing – review & editing, Methodology. JZ: Conceptualization, Writing – review & editing, Funding acquisition. HR: Writing – review & editing, Methodology, Conceptualization, Funding acquisition.

## Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by the National Natural Science Foundation of China (81972435, 82173899), Nanjing Medical Science and Technical Development Foundation (ZKX22030), Jiangsu Pharmaceutical Association (H202108, A2021024, Q202202, JY202207, Z04JKM2023E040, and A202309).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

## References

- Po, L, Thomas, J, Mills, K, Zakhari, A, Tulandi, T, Shuman, M, et al. Guideline no. 414: Management of Pregnancy of unknown location and tubal and nontubal ectopic pregnancies. *J Obstet Gynaecol Can.* (2021) 43:614–630.e1. doi: 10.1016/j.jogc.2021.01.002
- Nippita, S, Cansino, C, Goldberg, AB, Qasba, N, White, K, Goyal, V, et al. Society of Family Planning Clinical Recommendation: management of undesired pregnancy of unknown location and abortion at less than 42 days of gestation. *Contraception.* (2025) 150:110865. doi: 10.1016/j.contraception.2025.110865
- Obeagu, EI, Faduma, MH, Obeagu, GU, Agu, CC, and Kazibwe, S. Ectopic pregnancy: a review. USA: International Journal of current Research in Chemistry and Pharmaceutical Sciences. (2023)
- Barnhart, KT, Hansen, KR, Stephenson, MD, Usadi, R, Steiner, AZ, Cedars, MI, et al. Effect of an active vs expectant management strategy on successful resolution of pregnancy among patients with a persisting pregnancy of unknown location: the ACT or NOT randomized clinical trial. *JAMA.* (2021) 326:390–400. doi: 10.1001/jama.2021.10767
- Webster, K, Eadon, H, Fishburn, S, and Kumar, G. Ectopic pregnancy and miscarriage: diagnosis and initial management: summary of updated NICE guidance. *BMJ.* (2019) 367:l6283. doi: 10.1136/bmj.l6283
- Kirk, E, Ankum, P, Jakab, A, Le Clef, N, Ludwin, A, Small, R, et al. Terminology for describing normally sited and ectopic pregnancies on ultrasound: ESHRE recommendations for good practice. *Hum Reprod Open.* (2020) 2020:hoaa055. doi: 10.1093/hropen/hoaa055
- Al Naimi, A, Moore, P, Brüggmann, D, Krysa, L, Louwen, F, and Bahlmann, F. Ectopic pregnancy: a single-center experience over ten years. *Reprod Biol Endocrinol.* (2021) 19:79. doi: 10.1186/s12958-021-00761-w
- Nishio, N, Kido, A, Kurata, Y, Minami, M, Tokunaga, K, Honda, M, et al. Investigation of clinical utility of contrast-enhanced MRI in the diagnosis of ectopic pregnancy. *Clin Radiol.* (2020) 75:543–51. doi: 10.1016/j.crad.2020.02.013
- Lentscher, JA, Colburn, ZT, Ortogero, N, Gillette, L, Leonard, GT, Burney, RO, et al. An intrauterine genomic classifier reliably delineates the location of nonviable pregnancies. *Fertil Steril.* (2021) 116:138–46. doi: 10.1016/j.fertnstert.2021.02.005
- Selvarajan, S, Ramalingam, JM, and Kumar, DS. Advancing diagnostics: integrating microRNA profiling and protein markers in ectopic pregnancy detection. *Gynecol Obstet Clin Med.* (2024) 4:e000034. doi: 10.1136/gocm-2024-000034
- Romero-Ruiz, A, Avendaño, M S, Dominguez, F, Lozoya, T, Molina-Abril, H, Sangiao-Alvarellos, S, et al. Deregulation of miR-324/KISS1/kisspeptin in early ectopic pregnancy: mechanistic findings with clinical and diagnostic implications. *Am J Obstet Gynecol.* (2019) 220:480.e1–e17. doi: 10.1016/j.ajog.2019.01.228
- Bobdiwala, S, Saso, S, Verbakel, JY, Al-Memar, M, Van Calster, B, Timmerman, D, et al. Diagnostic protocols for the management of pregnancy of unknown location: a systematic review and meta-analysis. *BJOG Int J Obstet Gynaecol.* (2019) 126:190–8. doi: 10.1111/1471-0528.15442
- Fistouris, J, Bergh, C, and Strandell, A. Pregnancy of unknown location: external validation of the hCG-based M6NP and M4 prediction models in an emergency gynaecology unit. *BMJ Open.* (2022) 12:e058454. doi: 10.1136/bmjopen-2021-058454
- Maheut, C, Panjo, H, and Capmas, P. Diagnostic accuracy validation study of the M6 model without initial serum progesterone (M6NP) in triage of pregnancy of unknown location. *Eur J Obstet Gynecol Reprod Biol.* (2024) 296:360–5. doi: 10.1016/j.ejogrb.2024.03.010
- Bobdiwala, S, Christodoulou, E, Farren, J, Mitchell-Jones, N, Kyriacou, C, Al-Memar, M, et al. Triaging women with pregnancy of unknown location using two-step protocol including M6 model: clinical implementation study. *Ultrasound Obstet Gynecol.* (2020) 55:105–14. doi: 10.1002/uog.20420
- Handelman, GS, Kok, HK, Chandra, RV, Razavi, AH, Lee, MJ, and Asadi, H. Ed octor: machine learning and the future of medicine. *J Intern Med.* (2018) 284:603–19. doi: 10.1111/joim.12822
- Wang, L, Wang, C, Li, C, Murai, T, Bai, Y, Song, Z, et al. AI-assisted multi-modal information for the screening of depression: a systematic review and meta-analysis. *NPJ Digit Med.* (2025) 8:523. doi: 10.1038/s41746-025-01933-3
- Zhang, Y, Xu, D, Gao, J, Wang, R, Yan, K, Liang, H, et al. Development and validation of a real-time prediction model for acute kidney injury in hospitalized patients. *Nat Commun.* (2025) 16:68. doi: 10.1038/s41467-024-55629-5
- Rueangkiet, P, Rittiluechai, K, and Prayote, A. Predictive analytical model for ectopic pregnancy diagnosis: statistics vs. machine learning. *Front Med.* (2022) 9:976829. doi: 10.3389/fmed.2022.976829
- Link, CA, Maissiat, J, Mol, BW, Barnhart, KT, and Savaris, RF. Diagnosing ectopic pregnancy using Bayes theorem: a retrospective cohort study. *Fertil Steril.* (2023) 119:78–86. doi: 10.1016/j.fertnstert.2022.09.016
- Barnhart, KT, Bollig, KJ, Senapati, S, Takacs, P, Robins, JC, Haisenleder, DJ, et al. Multiplexed serum biomarkers to discriminate nonviable and ectopic pregnancy. *Fertil Steril.* (2024) 122:482–93. doi: 10.1016/j.fertnstert.2024.04.028
- Collins, GS, Moons, KG, Dhiman, P, Riley, RD, Beam, AL, Van Calster, B, et al. Tripod+ AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ.* (2024) 385:e078378. doi: 10.1136/bmj-2023-078378
- Suh-Burgmann, EJ, Flanagan, T, Lee, N, Osinski, T, Sweet, C, Lynch, M, et al. Large-scale implementation of structured reporting of adnexal masses on ultrasound. *J Am Coll Radiol.* (2018) 15:755–61. doi: 10.1016/j.jacr.2018.01.026
- Riley, RD, Ensor, J, Snell, KI, Harrell, FE, Martin, GP, Reitsma, JB, et al. Calculating the sample size required for developing a clinical prediction model. *BMJ.* (2020) 368:m441. doi: 10.1136/bmj.m441
- Regression N. An introduction to kernel and nearest-neighbor. *Am Stat.* (1992) 46:175–85.
- Okada, S, Ohzeki, M, and Taguchi, S. Efficient partition of integer optimization problems with one-hot encoding. *Sci Rep.* (2019) 9:13036. doi: 10.1038/s41598-019-49539-6
- Vasquez, MM, Hu, C, Roe, DJ, Chen, Z, Halonen, M, and Guerra, S. Least absolute shrinkage and selection operator type methods for the identification of serum

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmed.2025.1726606/full#supplementary-material>

- biomarkers of overweight and obesity: simulation and application. *BMC Med Res Methodol.* (2016) 16:154. doi: 10.1186/s12874-016-0254-8
28. Slinker, B, and Glantz, S. Multiple regression for physiological data analysis: the problem of multicollinearity. *Am J Phys Regul Integr Comp Phys.* (1985) 249:R1–R12. doi: 10.1152/ajpregu.1985.249.1.R1
29. Park, SH, Goo, JM, and Jo, C-H. Receiver operating characteristic (ROC) curve: practical review for radiologists. *Korean J Radiol.* (2004) 5:11–8. doi: 10.3348/kjr.2004.5.1.11
30. Davis, J, and Goadrich, M (2006) The relationship between precision-recall and ROC curves. In *Proceedings of the 23rd international conference on Machine learning.* New York, NY: Association for Computing Machinery. pp. 233–40
31. Lundberg, SM, and Lee, S-I. A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst.* (2017) 30:6785–95.
32. Zee, J, Sammel, M, Chung, K, Takacs, P, Bourne, T, and Barnhart, K. Ectopic pregnancy prediction in women with a pregnancy of unknown location: data beyond 48 h are necessary. *Hum Reprod.* (2014) 29:441–7. doi: 10.1093/humrep/det450
33. Van Mello, NM, Mol, F, Ankum, WM, Mol, BW, Van Der Veen, F, and Hajenius, PJ. Ectopic pregnancy: how the diagnostic and therapeutic management has changed. *Obstet Gynecol Surv.* (2013) 68:110–2. doi: 10.1097/01.ogx.0000427624.57640.ab
34. d'Hauterive, SP, Close, R, Grیدهlet, V, Mawet, M, Nisolle, M, and Geenen, V. Human chorionic gonadotropin and early embryogenesis. *Int J Mol Sci.* (2022) 23:1380. doi: 10.3390/ijms23031380
35. McGlade, EA, Miyamoto, A, and Winuthayanon, W. Progesterone and inflammatory response in the oviduct during physiological and pathological conditions. *Cells.* (2022) 11:1075. doi: 10.3390/cells11071075
36. Salomon, L, Alfirevic, Z, Bilardo, C, Chalouhi, G, Ghi, T, Kagan, K, et al. ISUOG practice guidelines: performance of first-trimester fetal ultrasound scan. *Ultrasound Obstet Gynecol.* (2013) 41:102–13. doi: 10.1002/uog.12342
37. Korevaar, TI, Steegers, EA, de Rijke, YB, Schalekamp-Timmermans, S, Visser, WE, Hofman, A, et al. Reference ranges and determinants of total hCG levels during pregnancy: the generation R study. *Eur J Epidemiol.* (2015) 30:1057–66. doi: 10.1007/s10654-015-0039-0
38. Wang, Z, Gao, Y, Zhang, D, Li, Y, Luo, L, and Xu, Y. Predictive value of serum  $\beta$ -human chorionic gonadotropin for early pregnancy outcomes. *Arch Gynecol Obstet.* (2020) 301:295–302. doi: 10.1007/s00404-019-05388-2
39. Cole, LA. Biological functions of hCG and hCG-related molecules. *Reprod Biol Endocrinol.* (2010) 8:102. doi: 10.1186/1477-7827-8-102
40. Ng, S-W, Norwitz, GA, Pavlicev, M, Tilburgs, T, Simón, C, and Norwitz, ER. Endometrial decidualization: the primary driver of pregnancy health. *Int J Mol Sci.* (2020) 21:4092. doi: 10.3390/ijms21114092
41. Hendriks, E, Rosenberg, R, and Prine, L. Ectopic pregnancy: diagnosis and management. *Am Fam Physician.* (2020) 101:599–606.
42. Zhang, Y, Li, Z, Ren, B, Wu, W, Liu, Y, Wang, X, et al. Diagnostic value of a single  $\beta$ -hCG test in predicting reproductive outcomes in women undergoing cleavage embryo transfer: a retrospective analysis from a single center. *Reprod Health.* (2022) 19:145. doi: 10.1186/s12978-022-01455-1
43. Kyriacou, C, Ledger, A, Bobdiwala, S, Ayim, F, Kirk, E, Abughazza, O, et al. Updating M6 pregnancy of unknown location risk-prediction model including evaluation of clinical factors. *Ultrasound Obstet Gynecol.* (2024) 63:408–18. doi: 10.1002/uog.27515
44. Nadim, B, Leonardi, M, Stamatopoulos, N, Reid, S, and Condous, G. External validation of risk prediction model M4 in an Australian population: Rationalising the management of pregnancies of unknown location. *Aust N Z J Obstet Gynaecol.* (2020) 60:928–34. doi: 10.1111/ajo.13201
45. Saito, T, and Rehmsmeier, M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS One.* (2015) 10:e0118432. doi: 10.1371/journal.pone.0118432
46. Huang, J, Man, Y, Shi, Z, Fu, X, Shi, W, and Liang, X. Global, regional, and national burden of maternal disorders, 1990–2021: a systematic analysis from the global burden of disease study 2021. *BMC Public Health.* (2025) 25:2576. doi: 10.1186/s12889-025-23814-w
47. Zhang, S, Liu, J, Yang, L, Li, H, Tang, J, and Hong, L. Global burden and trends of ectopic pregnancy: an observational trend study from 1990 to 2019. *PLoS One.* (2023) 18:e0291316. doi: 10.1371/journal.pone.0291316
48. Stabile, G, Cracco, F, Zinicola, G, Carlucci, S, Mangino, FP, Stampalija, T, et al. Subserosal pregnancy: systematic review with proposal of new diagnostic criteria and ectopic pregnancy classification. *Eur J Obstet Gynecol Reprod Biol.* (2024) 297:254–9. doi: 10.1016/j.ejogrb.2024.04.037
49. Stabile, G, Vona, L, Carlucci, S, Zullo, F, Lagana, AS, Etrusco, A, et al. Conservative treatment of cesarean scar pregnancy with the combination of methotrexate and mifepristone: a systematic review. *Womens Health (Lond).* (2024) 20:17455057241290424. doi: 10.1177/17455057241290424