Proceedings of the
13th International Conference on Applied Informatics
Eger, Hungary, February 19–21, 2026
DOI: 10.17048/icai.2026.146
pp. 146–160

An Interpretable Machine Learning Framework for Early Ectopic Pregnancy Prediction Using Routine Clinical Variables

Dhafer G. Honi[ab*], **Ahmed Abed Mohammed**[cd], **Laszlo Szathmary**[e]

[a]Doctoral School of Informatics, University of Debrecen, Debrecen, Hungary
dhafer.honi@inf.unideb.hu

[b]Department of Computer Science, College of Education for Pure Sciences, University of Basrah, Basrah, Iraq

[c]College of Computer Science and Information Technology, University of Al-Qadisiyah, Al Diwaniyah, Iraq
a.alsherbe@qu.edu.iq

[d]College of Technical Engineering, Islamic University, Najaf, Iraq

[e]Department of IT, University of Debrecen, Debrecen, Hungary
szathmary.laszlo@inf.unideb.hu

Abstract. Ectopic pregnancy (EP) remains a major cause of first-trimester maternal morbidity and mortality, while early diagnosis is challenging because clinical symptoms frequently overlap with those of other gynaecological conditions. This study proposes an interpretable machine-learning framework for early EP risk stratification using routinely available clinical variables. The proposed framework was evaluated on a cohort of 2,060 patients comprising 1,030 EP cases and 1,030 matched controls described by age and 12 binary clinical variables.

To ensure unbiased evaluation, the data were divided into independent training and test sets before model development. Synthetic samples were generated exclusively within the training partition using a class-conditional Gaussian copula model to augment model training while preserving the sta-

*Corresponding author: dhafer.honi@inf.unideb.hu

tistical characteristics of the original data. All validation, calibration, statistical analyses, and final performance evaluation were performed exclusively on previously unseen real patient data.

Six machine-learning classifiers, including Logistic Regression, Random Forest, Gradient Boosting, Support Vector Machine, Multilayer Perceptron, and a soft-voting ensemble, were evaluated using discrimination, calibration, clinical utility, and explainability analyses. Performance was assessed using ROC-AUC, PR-AUC, calibration metrics, decision-curve analysis, SHAP interpretation, and statistical comparisons based on the DeLong, McNemar, and Friedman tests.

The soft-voting ensemble achieved the highest performance on the independent test set, with an ROC-AUC of 0.609 (95% CI: 0.563–0.653). Multivariable logistic regression and SHAP consistently identified prior ectopic pregnancy as the strongest predictor (adjusted OR = 27.8, 95% CI: 8.7–88.8), followed by psychiatric disease and previous genital surgery. Although predictive performance from routine clinical variables alone was modest, the proposed framework provides transparent, reproducible, and well-calibrated risk estimation. The findings further demonstrate that synthetic data are effective for training augmentation while preserving the independence and validity of real-world clinical evaluation.

Keywords: ectopic pregnancy, interpretable machine learning, synthetic data augmentation, Gaussian copula, SHAP, calibration, decision curve analysis, clinical decision support

1. Introduction

Ectopic pregnancy (EP) is the implantation of a fertilised ovum outside the endometrial cavity, most often in the Fallopian tube [4], and is a leading cause of first-trimester maternal mortality [8]. Presentation is non-specific: lower abdominal pain, vaginal bleeding, and menstrual irregularity are shared with other gynaecological emergencies, so early discrimination is difficult [14]. Despite transvaginal ultrasound and serial hCG testing, about half of EP cases are undiagnosed at first presentation [13, 15], motivating data-driven decision support.

Prior machine-learning studies report encouraging performance. Aghayari et al. [1] obtained AUC 0.91 with a random forest on a real Iranian cohort; Rueangket et al. [16] found neural networks maximised recall at the cost of specificity; Du et al. [7] built an individualised risk tool with gradient boosting. Jacob et al. [11] identified prior EP, prior genital surgery, endometriosis, and psychiatric disorders as salient risk factors in a German cohort of $n = 100{,}197$; the anonymised extract released via Dryad [10] underlies the present study.

In this work we develop and evaluate an interpretable machine-learning framework for early EP risk stratification from routine clinical variables. Our contributions are:

- **Real-data evaluation with strict separation.** All metrics are computed on a test set that is held out before any model fitting; results are reported from

a single results file so that tables, figures, and text are mutually consistent.

- **Structure-preserving synthetic augmentation.** A class-conditional Gaussian copula generator preserves both the marginal prevalence and the pairwise co-occurrence structure of the clinical variables when generating synthetic training samples.

- **Feature analysis.** We study pairwise interaction terms, composite counts, and a learned composite risk score, together with feature selection.

- **Comprehensive evaluation.** Discrimination (AUC, PR-AUC), threshold metrics, statistical testing (DeLong, McNemar, Friedman), calibration (Brier score, slope, intercept), decision-curve analysis, and SHAP interpretability.

2. Related Work

2.1. Ectopic Pregnancy: Epidemiology and Risk Factors

EP occurs in 1−2% of all pregnancies and accounts for 9% of pregnancy-related deaths in the United States [8]. Established risk factors include prior EP, previous pelvic or tubal surgery, pelvic inflammatory disease (PID), endometriosis, irregular menstruation, advanced maternal age, and smoking [3, 9, 11, 18]. Menstrual symptoms – dysmenorrhoea, mid-cycle pain, and abnormal bleeding – have been highlighted in multiple studies as early clinical discriminators [11]. Huang et al. [9] demonstrated a strong nationwide association between PID and EP incidence, while Chong et al. [18] quantified the endometriosis risk (OR = 2.6) through meta-analysis.

2.2. Machine Learning for EP Prediction

Aghayari et al. [1] conducted a comprehensive ML study of EP prediction, evaluating Logistic Regression, Decision Trees, Random Forests, and XGBoost on a real clinical dataset of 2,060 patients from Iran. Random Forest achieved the best performance (AUC 0.9065, accuracy 87.13%). Their study is significant in introducing interpretable ML to EP triage but is limited by the absence of interaction-based features and a learned composite risk score. Rueangket et al. [16] compared stepwise logistic regression against Naive Bayes, k-NN, and neural network classifiers for EP diagnosis using emergency presentation data, finding that neural networks provided superior recall but at the cost of reduced specificity. Du et al. [7] proposed an individualized risk prediction tool for EP within the first 10 weeks of gestation using gradient boosting and lasso-based variable selection, achieving competitive AUC.

2.3. Ensemble Methods in Clinical Prediction

The theoretical justification for ensemble learning rests on the variance–bias decomposition and on the observation that combining classifiers with uncorrelated errors reduces total generalisation error [6, 19]:

$$\text{Generalisation Error} = \text{Bias}^2 + \text{Variance} + \text{Irreducible Noise}.$$

Ensemble methods have demonstrated consistent improvements over single models across oncology [2], cardiac risk stratification, and ICU outcome prediction. Soft-voting ensembles preserve calibrated probability estimates, which is essential for downstream clinical risk thresholding [17].

2.4. Interpretability in Clinical Machine Learning

SHapley Additive exPlanations (SHAP) [12] provide a game-theoretic decomposition of individual predictions, assigning each feature a contribution value that satisfies efficiency, symmetry, and null player axioms. SHAP values are now routinely used to audit clinical models for face validity and to support regulatory submissions [2]. Logistic Regression coefficients, when features are standardised, offer a complementary global interpretability mechanism that is directly actionable by physicians [5].

3. Methodology

3.1. Dataset and preprocessing

We use the Dryad extract of Jacob et al. [10]: $n = 2060$ records with one continuous variable (age) and 12 binary history and symptom variables (prior ectopic pregnancy; prior genital surgery; psychiatric disease; vulvitis; endometriosis; cervical erosion/ectropion; non-inflammatory vaginal disorders; absent/scanty/rare menstruation; irregular menstruation; vaginal bleeding; mid-cycle pain; dysmenorrhoea). The cohort is a matched case–control sample (1030 EP, 1030 controls). There were 0 missing cells, so no imputation was required. 1488 records share an identical feature vector; because the predictors are sparse binaries, these are distinct patients with the same recorded profile and were retained. Continuous and binary features were standardised using statistics fit on the training partition only.

3.2. Train/test separation

Data were split by stratified sampling into 1442 training and 618 test records. The test set is held out before any fitting and is used once, for final reporting. All transformers, the feature selector, every classifier, and the synthetic generator are fit on training data only. Model selection and calibration use 5-fold stratified cross-validation within the training partition, with identical folds across models.

3.3. Synthetic Training Augmentation

Synthetic data were generated solely to increase the diversity of the training set and were never used for validation or testing. The independent test set was separated before model development, and all reported results were obtained exclusively from previously unseen real patient data.

A class-conditional Gaussian copula model was fitted separately for the EP and non-EP classes using the training partition. Binary variables were generated by thresholding latent Gaussian variables according to their observed prevalence, while age was sampled from the fitted Gaussian distribution and truncated to the clinically plausible range (16–45 years). This approach preserves both feature prevalence and the dependency structure among predictors.

The quality of the synthetic data was assessed by comparing feature prevalence, age distribution, and pairwise correlations with the original training data. Finally, four training strategies – no augmentation, class weighting, SMOTE, and Gaussian copula augmentation – were compared using identical train/test partitions, with performance evaluated only on the untouched real-world test set.

3.4. Feature engineering and selection

Feature engineering was performed to investigate whether clinically motivated feature transformations could improve predictive performance. Four pairwise interaction terms (age × irregular menstruation, prior genital surgery × vaginal bleeding, prior EP × mid-cycle pain, and psychiatric disease × endometriosis) were selected *a priori* based on previously reported clinical associations rather than exhaustive combinatorial search. This strategy reduces the risk of overfitting while preserving clinical interpretability.

In addition, two composite count variables and a logistic-regression-derived risk score were evaluated. The contribution of each engineered feature set was assessed using 5-fold cross-validated ROC-AUC on the training partition (Table 1). Feature-selection methods, including ANOVA-F, mutual information, and recursive feature elimination, were also investigated. Because none of the engineered features consistently improved cross-validated performance, the final models were trained using the original 13 routine clinical variables.

3.5. Models and evaluation

We evaluate logistic regression (ℓ_2), random forest, gradient boosting, SVM (RBF), a small multilayer perceptron (32–16 units, early stopping), and an unweighted soft-voting ensemble of the five. On the test set we report accuracy, recall (sensitivity), precision (PPV), specificity, NPV, F1, balanced accuracy, MCC, ROC-AUC, and PR-AUC, with percentile bootstrap 95% confidence intervals for AUC. Model comparison uses the DeLong test (paired AUCs), McNemar's test (paired predictions), and a Friedman test across cross-validation folds. Calibration is assessed by the reliability curve, Brier score, calibration slope, and calibration intercept; clinical

utility by decision-curve analysis. Interpretability uses multivariable odds ratios and SHAP.

4. Results

4.1. Feature analysis

Table 1 reports 5-fold cross-validated AUC for incremental feature sets on the training partition. Figure 1 visually summarizes the feature engineering and training strategy experiments. The left panel confirms that neither interaction features, composite variables, nor the learned risk score improved cross-validated ROC-AUC compared with the original 13 routine clinical variables. The right panel compares alternative training strategies evaluated on the independent real-world test set. Class weighting and SMOTE produced almost identical performance to the baseline models, whereas Gaussian-copula augmentation resulted in only small model-dependent changes. These findings indicate that the predictive limitation originates primarily from the available clinical variables rather than from insufficient sample size.

Table 1. Cross-validated AUC for incremental feature sets (5-fold, training partition, logistic regression).

Feature set	**CV AUC (mean ± SD)**
Original 13 features	0.5902 ± 0.0271
+ Interaction terms (f_1–f_4)	0.5875 ± 0.0262
+ Composite counts (f_5–f_6)	0.5875 ± 0.0263
+ Learned risk score (f_7)	0.5876 ± 0.0263

4.2. Training strategies

Table 2 reports held-out test AUC under four training strategies. Class weighting and SMOTE produce test AUC close to the unaugmented models, and copula augmentation produces small, model-dependent differences.

Table 2. Held-out test AUC by training strategy.

Model	**None**	**Class weight**	**SMOTE**	**Copula aug.**
LogisticRegression	0.578	0.578	0.578	0.587
RandomForest	0.597	0.597	0.597	0.595
GradientBoosting	0.600	0.600	0.600	0.590
SVM_RBF	0.605	0.605	0.605	0.605
MLP	0.589	0.589	0.589	0.616

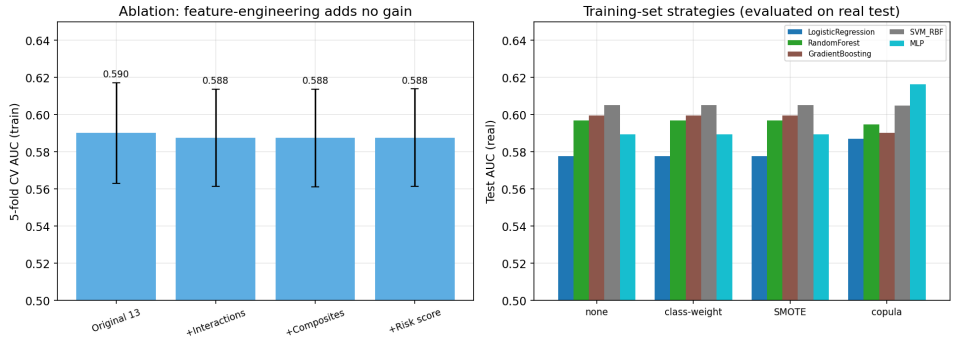


Figure 1. Feature engineering ablation (left) and training strategy comparison (right).

4.3. Validation of Synthetic Data Quality

We assessed whether the class-conditional Gaussian-copula generator reproduces the statistical properties of the real training data. The generator was fit on the training partition only, and a synthetic replicate of the same size per class was drawn. Agreement was quantified along three axes: binary feature prevalence, pairwise correlation structure, and the age distribution.

Feature prevalence. Table 3 compares the prevalence of each binary feature between the real and synthetic training data, separately for controls and EP cases. The synthetic marginals closely track the real ones, with a maximum absolute prevalence error of 2.5 percentage points across all features and classes (Fig. 2).

Table 3. Binary feature prevalence (%) in the real vs. synthetic training data, by class.

	Control		EP	
Feature	**Real**	**Synthetic**	**Real**	**Synthetic**
Prior ectopic pregnancy	0.4	0.1	10.5	10.4
Prior genital surgery	1.2	1.7	2.9	3.6
Psychiatric disease	2.8	2.4	6.4	7.1
Vulvitis	3.9	3.9	4.4	3.7
Endometriosis	0.6	0.4	1.1	2.2
Cervical erosion/ectropion	3.3	3.7	5.7	6.1
Noninflammatory vaginal disorder	9.3	7.2	14.0	16.5
Absent/scanty menstruation	5.5	3.7	6.2	5.8
Irregular menstruation	6.9	5.7	9.7	11.1
Vaginal bleeding	1.7	1.7	3.5	4.2
Mid-cycle pain	0.1	0.0	0.6	0.7
Dysmenorrhoea	3.3	4.6	5.0	6.0

Correlation structure. Because interaction-aware modelling requires that feature co-occurrence be preserved, we compared the pairwise correlation matrices of the real and synthetic data within each class (Table 4, Fig. 3). The mean absolute off-diagonal correlation error is 0.055 for controls and 0.066 for EP cases, confirming that the copula reproduces the observed association structure rather than treating the features as independent.

Table 4. Agreement between real and synthetic pairwise correlation matrices (off-diagonal elements).

Class	Frobenius norm	Mean abs. error	Max abs. error
control	1.379	0.055	0.315
EP	0.971	0.066	0.233

Age distribution. The continuous age variable was compared using a two-sample Kolmogorov–Smirnov (KS) test and by matching the first two moments (Table 5, Fig. 4). Synthetic and real age means agree to within 0.4 years in both classes, and the KS statistic is small in both cases ($D \leq 0.076$), indicating close distributional agreement.

Table 5. Age distribution: real vs. synthetic (mean $\pm$ SD) and two-sample Kolmogorov–Smirnov test.

Class	Real (mean $\pm$ SD)	Synthetic (mean $\pm$ SD)	KS D	KS p
control	31.30 ± 5.72	30.93 ± 5.95	0.076	0.030
EP	31.84 ± 6.85	31.91 ± 6.47	0.071	0.054

4.4. Predictive performance

Table 6 reports test-set performance. The soft-voting ensemble attains the highest AUC (0.609; 95% CI $[0.563, 0.653]$). To illustrate the prediction behaviour of the best-performing ensemble model, Figure 5 presents the confusion matrix obtained using the default probability threshold of 0.5. The model correctly identified 244 non-EP cases and 118 EP cases, while misclassifying 65 controls as EP and 191 EP cases as non-EP. The relatively high number of false negatives reflects the limited discriminative information contained in routine clinical variables alone.

4.5. Statistical comparison

Pairwise DeLong tests against the ensemble (LogisticRegression $p = 0.087$, RandomForest $p = 0.154$, GradientBoosting $p = 0.345$, SVM_RBF $p = 0.764$, MLP $p = 0.320$) and a Friedman test across cross-validation folds ($\chi^2 = 5.00$, $p = 0.416$;

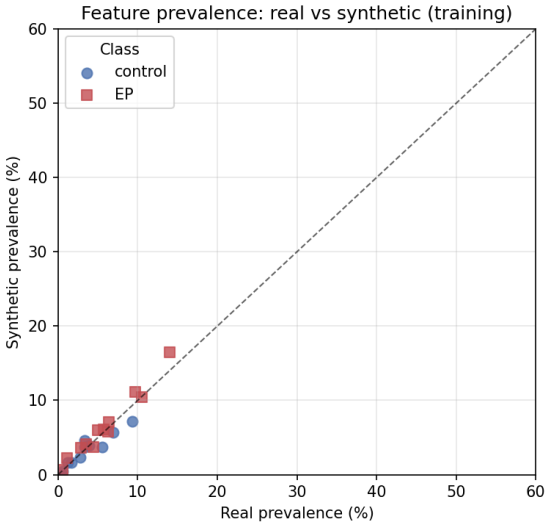


Figure 2. Real vs. synthetic feature prevalence (training partition); the dashed line is the identity.

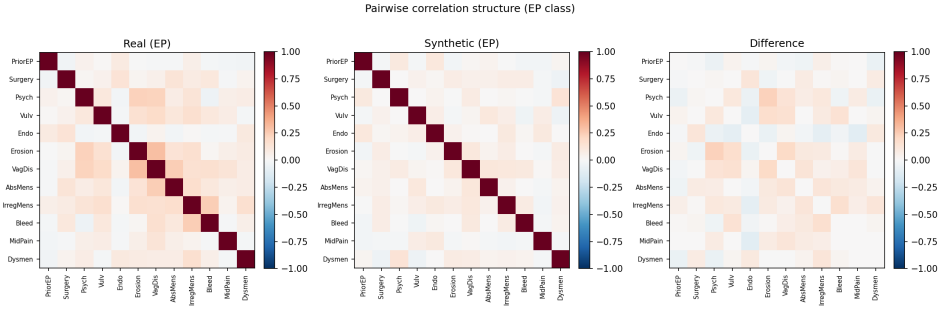


Figure 3. Pairwise correlation structure for the EP class: real (left), synthetic (centre), and their difference (right).

Table 6. Test-set performance ($n = 618$). Best AUC in bold.

Model	ROC_AUC	PR_AUC	Acc.	Recall	Spec.	Prec.	F1	Bal.Acc	MCC	NPV
LogisticRegression	0.578	0.615	0.561	0.298	0.825	0.630	0.404	0.561	0.145	0.540
RandomForest	0.597	0.633	0.578	0.356	0.799	0.640	0.457	0.578	0.173	0.554
GradientBoosting	0.600	0.629	0.573	0.440	0.706	0.599	0.507	0.573	0.151	0.558
SVM_RBF	0.605	0.619	0.573	0.440	0.706	0.599	0.507	0.573	0.151	0.558
MLP	0.589	0.623	0.594	0.366	0.822	0.673	0.474	0.594	0.211	0.564
Ensemble	0.609	0.640	0.586	0.382	0.790	0.645	0.480	0.586	0.188	0.561

McNemar's test likewise) indicate no statistically significant differences among the models.

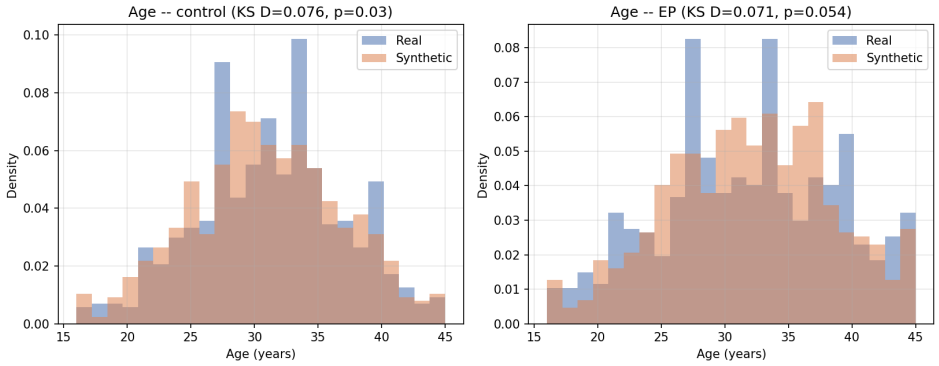


Figure 4. Age distributions, real vs. synthetic, for controls (left) and EP cases (right), with two-sample KS statistics.

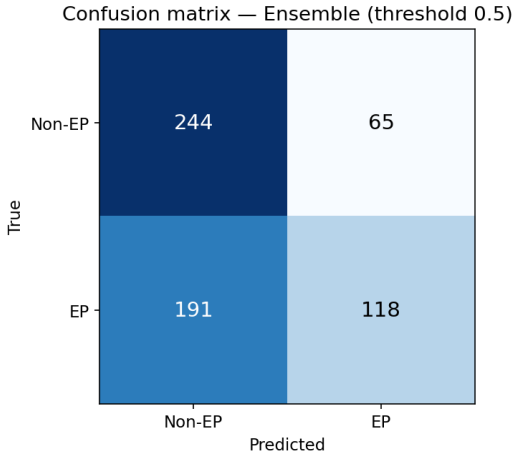


Figure 5. Confusion matrix of the soft-voting ensemble evaluated on the independent test set using a probability threshold of 0.5.

4.6. Calibration and clinical utility

The ensemble is well calibrated (Brier score = 0.237, calibration slope = 1.088, intercept = −0.009; Fig. 7a). Decision-curve analysis (Fig. 7b) shows a positive net benefit over the "treat all" and "treat none" strategies across clinically plausible threshold probabilities.

4.7. Interpretability

Multivariable odds ratios (Table 8) and SHAP (Fig. 8) agree that prior ectopic pregnancy is the dominant predictor (adjusted OR 27.8, 95% CI [8.7, 88.8]), followed by psychiatric disease, prior genital surgery, and vaginal bleeding. Age was

Table 7. 5-fold stratified cross-validation (training partition, identical folds).

Model	AUC (mean ± SD)	Accuracy (mean ± SD)
LogisticRegression	0.590 ± 0.027	0.563 ± 0.026
RandomForest	0.621 ± 0.031	0.583 ± 0.034
GradientBoosting	0.601 ± 0.038	0.576 ± 0.025
SVM_RBF	0.598 ± 0.025	0.589 ± 0.017
MLP	0.599 ± 0.029	0.575 ± 0.020
Ensemble	0.608 ± 0.029	0.582 ± 0.026

not significantly associated with EP ($p = 0.19$), consistent with the source epidemiology [11].

Table 8. Adjusted odds ratios (multivariable logistic regression; $^*p < 0.05$). Age is per year.

Feature	OR	95% CI	p
ectopic_past	27.79	[8.70, 88.78]	0.000*
Mid-cycle pain	3.16	[0.34, 29.57]	0.313
Genital surgery in the past	2.25	[1.00, 5.05]	0.050
vaginal bleeding	1.94	[0.94, 4.00]	0.074
psychiatric_disease	1.88	[1.07, 3.32]	0.029*
Dysmenorrhea	1.38	[0.79, 2.40]	0.258
Erosion and ectropion of cervix uteri	1.36	[0.78, 2.39]	0.277
Noninflammatory disorders of vagina, unspecified	1.34	[0.93, 1.94]	0.116
Endometriosis	1.14	[0.30, 4.36]	0.847
irregular menstruation	1.10	[0.72, 1.66]	0.665
age	1.01	[0.99, 1.03]	0.190
Absent, scanty, and rare menstruation	0.92	[0.57, 1.48]	0.742
Vulvitis	0.91	[0.52, 1.59]	0.732

5. Discussion

The proposed framework provides a reproducible and interpretable evaluation of early ectopic pregnancy (EP) prediction using routinely available clinical variables. Under strict train/test separation, the best-performing ensemble achieved a test ROC-AUC of 0.609, indicating that routine demographic, medical history, and symptom variables alone provide only modest discrimination for early EP.

The ablation study showed that interaction terms and the learned composite risk score produced no meaningful improvement in predictive performance, sug-

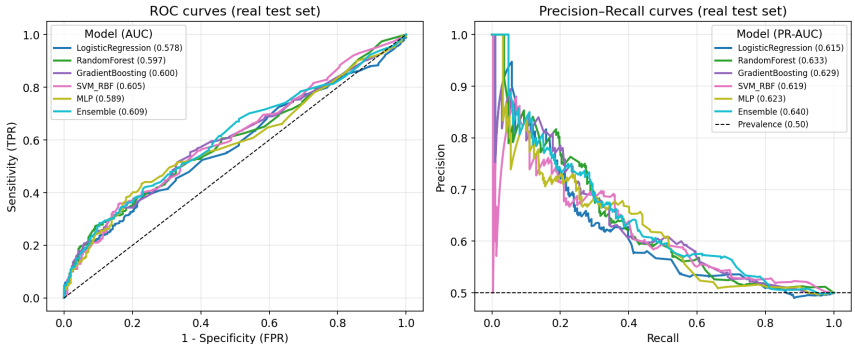
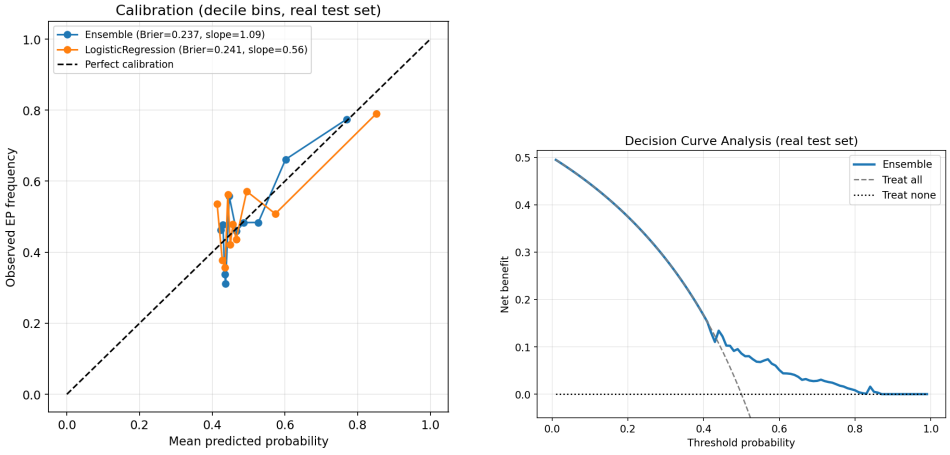


Figure 6. ROC and precision–recall curves on the test set.



(a) Calibration. **(b)** Decision curve analysis.

Figure 7. Calibration and clinical utility (best model, test set).

gesting that the original clinical variables already capture most of the available predictive information. Similarly, Gaussian-copula augmentation preserved the statistical characteristics of the training data but resulted in only small performance differences, indicating that limited predictor information rather than sample size is the principal constraint.

The identified predictors are consistent with previous clinical evidence. Prior ectopic pregnancy remained the strongest risk factor in both multivariable logistic regression and SHAP analysis, while previous genital surgery and psychiatric disease also contributed to prediction. Compared with studies reporting substantially higher AUC values, the present work relies exclusively on routine clinical variables and intentionally excludes laboratory biomarkers and ultrasound findings, providing a realistic estimate of the predictive capability of basic clinical information.

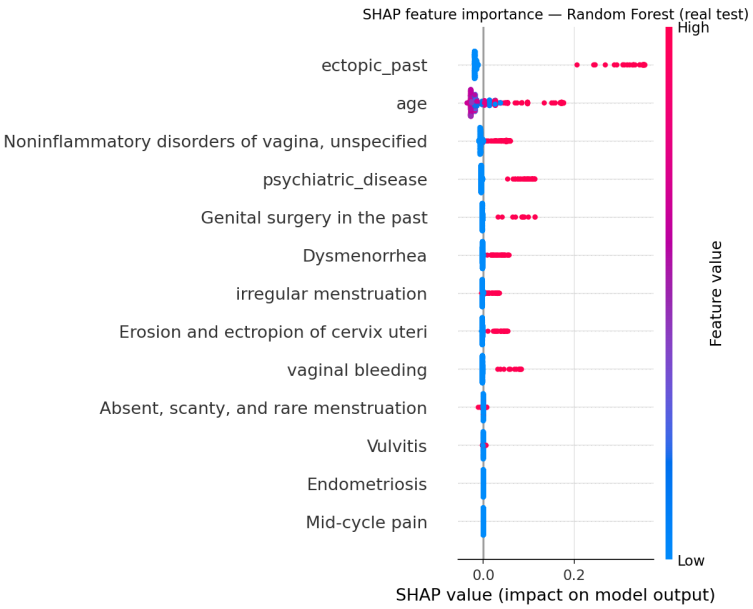


Figure 8. SHAP feature importance (random forest, test set). Prior ectopic pregnancy is the dominant contributor.

5.1. Limitations

The study has some limitations. First, the matched case–control cohort does not reflect the true clinical prevalence of EP, requiring recalibration before deployment in routine practice. Second, the available predictors exclude important diagnostic information such as serum β-hCG measurements and ultrasound findings, which likely explains the modest predictive performance. Finally, although the Gaussian-copula generator preserved the statistical characteristics of the training data, synthetic samples were used exclusively for training augmentation, while all reported results were obtained from an independent real-world test set.

6. Conclusion

We presented a reproducible and interpretable machine-learning framework for early EP risk stratification using 13 routine clinical variables. The results demonstrate that synthetic data are most appropriately used for training augmentation, whereas reliable performance assessment should remain based exclusively on independent real-world patient data. Under strict train/test separation, the best-performing ensemble achieved a test ROC-AUC of 0.609 (95% CI: 0.563–0.653), while multivariable logistic regression and SHAP consistently identified prior ectopic pregnancy as the strongest predictor. Although predictive performance from

routine clinical variables alone was modest, the proposed framework provides a transparent benchmark for early EP prediction and supports future integration of laboratory biomarkers, imaging data, and prospective external validation.

References

[1] A. Aghayari, A. Sorayaie Azar, M. Taheri-Anganeh, S. Sadeghpour, Y. Mohammadpour, J. Bagherzadeh Mohasefi, H. Ghasemnejad-Berenji: *Prediction of ectopic pregnancy using interpretable machine learning algorithms*, BMC Pregnancy and Childbirth 25 (2025), p. 1119, doi: 10.1186/s12884-025-08179-7.

[2] R. O. Alabi, M. Elmusrati, I. Leivo, A. Almangush, A. A. Mäkitie: *Machine learning explainability in nasopharyngeal cancer survival using LIME and SHAP*, Scientific Reports 13 (2023), p. 8984, doi: 10.1038/s41598-023-35795-0.

[3] A.-M. N. Andersen, J. Wohlfahrt, P. Christens, J. Olsen, M. Melbye: *Maternal age and fetal loss: population based register linkage study*, British Medical Journal 320.7251 (2000), pp. 1708–1712, doi: 10.1136/bmj.320.7251.1708.

[4] J. Bouyer, J. Coste, H. Fernandez, J.-L. Pouly, N. Job-Spira: *Sites of ectopic pregnancy: a 10 year population-based study of 1800 cases*, Human Reproduction 17.12 (2002), pp. 3224–3230, doi: 10.1093/humrep/17.12.3224.

[5] N. R. Cook: *Statistical evaluation of prognostic versus diagnostic models: beyond the ROC curve*, Clinical Chemistry 54.1 (2008), pp. 17–23, doi: 10.1373/clinchem.2007.096529.

[6] T. G. Dietterich: *Ensemble methods in machine learning*, in: Multiple Classifier Systems, ed. by J. Kittler, F. Roli, vol. 1857, Lecture Notes in Computer Science, Springer, 2000, pp. 1–15, doi: 10.1007/3-540-45014-9_1.

[7] T. Du, H. Chen, R. Fu, Q. Chen, Y. Wang, B. W. Mol, R. Li, J. Qiao: *An individualized risk prediction tool for ectopic pregnancy within the first 10 weeks of gestation based on machine learning algorithms*, Frontiers in Medicine 12 (2025), p. 1726606, doi: 10.3389/fmed.2025.1726606.

[8] E. Hendriks, R. Rosenberg, L. Prine: *Ectopic pregnancy: diagnosis and management*, American Family Physician 101.10 (2020), pp. 599–606, url: https://www.aafp.org/pubs/afp/issues/2020/0515/p599.html.

[9] C.-C. Huang, C.-C. Huang, S.-Y. Lin, C.-Y. Chang, W.-C. Lin, C.-H. Chung, W.-C. Chien, C.-H. Lin: *Association of pelvic inflammatory disease (PID) with ectopic pregnancy and preterm labor in Taiwan: a nationwide population-based retrospective cohort study*, PLoS ONE 14.8 (2019), e0219351, doi: 10.1371/journal.pone.0219351.

[10] L. Jacob, M. Kalder, K. Kostev: *Data from: Risk factors for ectopic pregnancy in Germany: a retrospective study of 100,197 patients*, 2017, doi: 10.5061/dryad.pv805, url: https://datadryad.org/dataset/doi:10.5061/dryad.pv805.

[11] L. Jacob, M. Kalder, K. Kostev: *Risk factors for ectopic pregnancy in Germany: a retrospective study of 100,197 patients*, GMS German Medical Science 15 (2017), Doc19, doi: 10.3205/000260.

[12] S. M. Lundberg, S.-I. Lee: *A unified approach to interpreting model predictions*, in: Advances in Neural Information Processing Systems, ed. by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett, vol. 30, Curran Associates, Inc., 2017, pp. 4765–4774, url: https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf.

[13] K. Mullany, M. Minneci, R. Monjazeb, O. Coiado: *Overview of ectopic pregnancy diagnosis, management, and innovation*, Women's Health 19 (2023), p. 17455057231160348, doi: 10.1177/17455057231160348.

[14] H. Murray, H. Baakdah, T. Bardell, T. Tulandi: *Diagnosis and treatment of ectopic pregnancy*, Canadian Medical Association Journal 173.8 (2005), pp. 905–912, doi: 10.1503/cmaj.050222.

[15] S. A. Razavi, H. Pakniat, A. Emami, M. Mirabedini, M. Mahdavi, M. Mirzadeh: *Ectopic pregnancy and its associated risk factors: a case-control study in Qazvin*, International Journal of Fertility and Sterility 16.3 (2022), pp. 202–206, doi: 10.22074/ijfs.2021.537225.1167.

[16] P. Rueangket, K. Rittiluechai, A. Prayote: *Predictive analytical model for ectopic pregnancy diagnosis: statistics vs. machine learning*, Frontiers in Medicine 9 (2022), p. 976829, doi: 10.3389/fmed.2022.976829.

[17] B. Van Calster, L. Wynants, J. F. M. Verbeek, J. Y. Verbakel, E. Christodoulou, A. J. Vickers, M. J. Roobol, E. W. Steyerberg: *Reporting and interpreting decision curve analysis: a guide for investigators*, European Urology 74.6 (2018), pp. 796–804, doi: 10.1016/j.eururo.2018.08.038.

[18] P. J. Yong, S. Matwani, C. Brace, A. Quaiattini, M. A. Bedaiwy, A. Albert, C. Allaire: *Endometriosis and ectopic pregnancy: a meta-analysis*, Journal of Minimally Invasive Gynecology 27.2 (2020), 352–361.e2, doi: 10.1016/j.jmig.2019.09.778.

[19] Z.-H. Zhou: *Ensemble learning*, in: Encyclopedia of Biometrics, ed. by S. Z. Li, A. K. Jain, Springer, 2015, pp. 411–416, doi: 10.1007/978-0-387-73003-5_293.