

Interaction-Aware Clinical Feature Engineering with Ensemble Learning for Early Ectopic Pregnancy Prediction

Dhafer G. Honi^{abc*}, Ahmed Abed Mohammed^{de}, Laszlo Szathmary^f

^aDoctoral School of Informatics, University of Debrecen, Debrecen, Hungary
dhafer.honi@inf.unideb.hu

^bDepartment of Computer Science, College of Education for Pure Sciences, University of Basrah, Basrah, Iraq

^cDepartment of Business Management, Al-imam University College, Balad, Iraq

^dCollege of Computer Science and Information Technology, University of Al-Qadisiyah, Al Diwanayah, Iraq a.alsherbe@qu.edu.iq

^eCollege of Technical Engineering, Islamic University, Najaf, Iraq

^fDepartment of IT, University of Debrecen, Debrecen, Hungary
szathmary.laszlo@inf.unideb.hu

Abstract

This paper proposes an interaction-aware machine learning architecture for early ectopic pregnancy risk prediction. Engineered composite scores and explicit interaction terms address the limitation of modeling risk factors independently. Six models were evaluated, with Logistic Regression achieving the highest ROC AUC (0.978) and a soft-voting ensemble attaining the highest accuracy (0.915), outperforming recent benchmarks. SHAP analysis confirms clinical validity by identifying patient age and engineered composite risk scores as dominant predictors. The results demonstrate that engineered interaction features improve both predictive performance and model transparency.

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

1. Introduction

Ectopic pregnancy is a leading cause of first-trimester maternal mortality, with early diagnosis challenged by non-specific symptoms and inconclusive initial assessments [2, 4, 5]. Machine learning offers promising tools for early risk stratification, with recent models such as Random Forest and CatBoost reporting AUCs up to 0.930 [1, 3]. However, these approaches often model predictors separately, neglecting clinically important interactions.

This paper proposes an interaction-aware feature engineering method that integrates composite scores and explicit interaction terms into an interpretable framework. SHAP ensures clinically viable decisions. The work contributes interaction-aware features, strong predictive performance, and interpretability analyses that verify clinical intuition.

2. Methods

2.1. Dataset and Feature Engineering

The dataset includes demographics, gynecological symptoms, menstrual abnormalities, and obstetric/surgical risk factors. Categorical data are one-hot encoded, age-standardized, and split into 70% training and 30% test. Two novel feature types capture complex risk patterns: a Composite Score, which is an enhanced risk score from weighted integration of key indicators (history of ectopic pregnancy, genital surgery, dysmenorrhea) derived from univariate logistic regression coefficients, and Interaction Terms, which model critical clinical relationships, including *age_irregular_interaction* and *multiple_risk_interaction*.

2.2. Models and Evaluation

Six classification models are evaluated: Logistic Regression, Random Forest, Gradient Boosting, SVM-RBF, soft-voting Ensemble, and Deep Neural Network. ROC AUC is the primary metric, with accuracy, F1-score, and Average Precision also reported. Hyperparameter tuning uses 5-fold cross-validation.

3. Results

3.1. Predictive Performance

All models performed well, with Logistic Regression achieving the highest ROC AUC (0.978) and the soft-voting ensemble attaining the highest accuracy (0.915). The proposed Logistic Regression model outperformed recent benchmarks (Table 1).

3.2. Interpretability and SHAP Analysis

SHAP techniques provide robust explanations, evaluating feature importance for global model behavior and individual cases. SHAP analysis for the Logistic Regression model identifies patient age as the most influential predictor, followed by the engineered *enhanced_risk_score*. Engineered interaction features show moderate importance, indicating ectopic pregnancy risk is influenced by interactions between age, menstrual history, and co-occurring risk factors.

Table 1. Comparison with recent literature benchmarks.

Study (Year)	Best Model	AUC	Acc	F1	AP
Aghayari et al. (2025)	Random Forest	0.9065	0.843	0.821	0.901
Du et al. (2025)	CatBoost	0.930	0.872	0.865	0.915
Proposed Study		0.978	0.902	0.901	0.914

4. Discussion and Conclusion

This interaction-sensitive model for ectopic pregnancy risk prediction achieved strong performance (ROC AUC 0.978). Engineered interaction features improved accuracy, with Logistic Regression outperforming more complex models. SHAP explanations confirmed clinical validity, identifying age as the most influential predictor with engineered features providing meaningful contributions. Using readily available clinical data enhances interprofessional applicability. This transparency and the results underscore the importance of combining strategic feature engineering with explainable methods to improve clinical decision support.

References

- [1] A. AGHAYARI, A. AZAR, et AL.: *Prediction of ectopic pregnancy using interpretable machine learning algorithms*, BMC Pregnancy Childbirth 25 (2025), p. 1119, DOI: [10.1186/s12884-025-08179-7](https://doi.org/10.1186/s12884-025-08179-7).
- [2] G. CONDOUS, E. KIRK, et AL.: *Prediction of ectopic pregnancy in women with a pregnancy of unknown location*, Ultrasound Obstet Gynecol 29.6 (2007), pp. 680–687, DOI: [10.1002/uog.4015](https://doi.org/10.1002/uog.4015).
- [3] X. DU, R. LI, et AL.: *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](https://doi.org/10.3389/fmed.2025.1726606).
- [4] E. HENDRIKS, R. ROSENBERG, et AL.: *Ectopic pregnancy: diagnosis and management*, Am Fam Physician 101.10 (2020), pp. 599–606, URL: <https://www.aafp.org/pubs/afp/issues/2020/0515/p599.html>.
- [5] H. MURRAY, H. BAAKDAH, et AL.: *Diagnosis and treatment of ectopic pregnancy*, CMAJ 173.8 (2005), pp. 905–912, DOI: [10.1503/cmaj.050222](https://doi.org/10.1503/cmaj.050222).