Preterm Birth Prediction using Explainable Machine Learning: The Mutaba'ah Study

Wasif Khan[1,2], Nadirah Ghenimi[3], Romona D Govender[3], Nazar Zaki[1,2], Amir Ahmad[1,4], Mohammad M. Masud[1,2,4,], Luai A. Ahmed[4,5]

[1]*College of Information Technology, United Arab Emirates University, UAE*

[2]*Big Data Analytics Center, United Arab Emirates University, UAE*

[3] *Department of Family Medicine, College of Medicine and Health Sciences, United Arab Emirates University, UAE*

[4]*Zayed Centre for Health Sciences, United Arab Emirates University, UAE*

[5]*Institute of Public Health, College of Medicine and Health Sciences, United Arab Emirates University, UAE*

ABSTRACT

Preterm birth (PTB) is a complex pregnancy challenge that can lead to high perinatal mortality and long-term morbidity for infants. Early prediction of PTB would be useful for clinicians. With advances in machine learning algorithm (ML), several PTB prediction models have been proposed. However, existing ML-based models lack proper interpretations for clinicians.

Purpose: The aim is to identify risk factors and predict PTB in a population of pregnant woman using ML models and explain the predictions via the SHapley Additive exPlanations (SHAP) algorithm.

Study Design: A dataset of 3,509 pregnant women from the Emirati population in the United Arab Emirates (UAE) was analyzed. Based on medical and Artificial Intelligence (AI) literature, 35 risk factors associated with PTB were selected. To evaluate the performance of the PTB prediction method, six ML algorithms were tested. The most important risk factors were identified using the SHAP feature attribution framework. Additionally, an individual patient analysis was performed using SHAP and Local Interpretable Model-agnostic Explanations (LIME).

Results: The overall incidence proportion of PTB was 11.23% (11% parous women and 12.1% in nulliparous women). The XGBoost model achieved the best performance (AUC of 0.735 in the parous women, AUC of 0.723 in the nulliparous women) compared to other ML models tested. The main risk factors associated with PTB in the parous women were previous PTB, previous caesarian section, pre-eclampsia during pregnancy and maternal age. In nulliparous women, body mass index (BMI) at delivery, maternal age, and infection of amniotic sac were the most relevant risk factors.

Conclusion: The trained ML prediction model can potentially be used as a screening tool for PTB prediction in this population. The SHAP analysis can assist clinicians in understanding the impact of each risk factor for each patient and providing appropriate care to reduce the morbidity and mortality related to PTB.

Keywords: PTB prediction, Machine Learning model, risk assessment, SHAP, LIME algorithms.