

Preterm Birth Prediction

using Explainable Machine Learning: The Mutaba'ah Study

Wasif Khan^{1,2}, Nadirah Ghenimi³, Romona D Govender³, 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, Al Ain, UAE

4 Zayed Centre for Health Sciences, United Arab Emirates University, Al Ain, UAE

5 Institute of Public Health, College of Medicine and Health Sciences, United Arab Emirates University, Al Ain, UAE

Background

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 in a pregnant woman can be useful for preventing catastrophic outcomes for her and her baby. With advances in machine learning (ML)-based algorithms, several PTB prediction models have been proposed. However, existing ML-based models for PTB prediction work in a black-box manner and lack proper interpretations for clinicians, which limits their use in clinical settings.

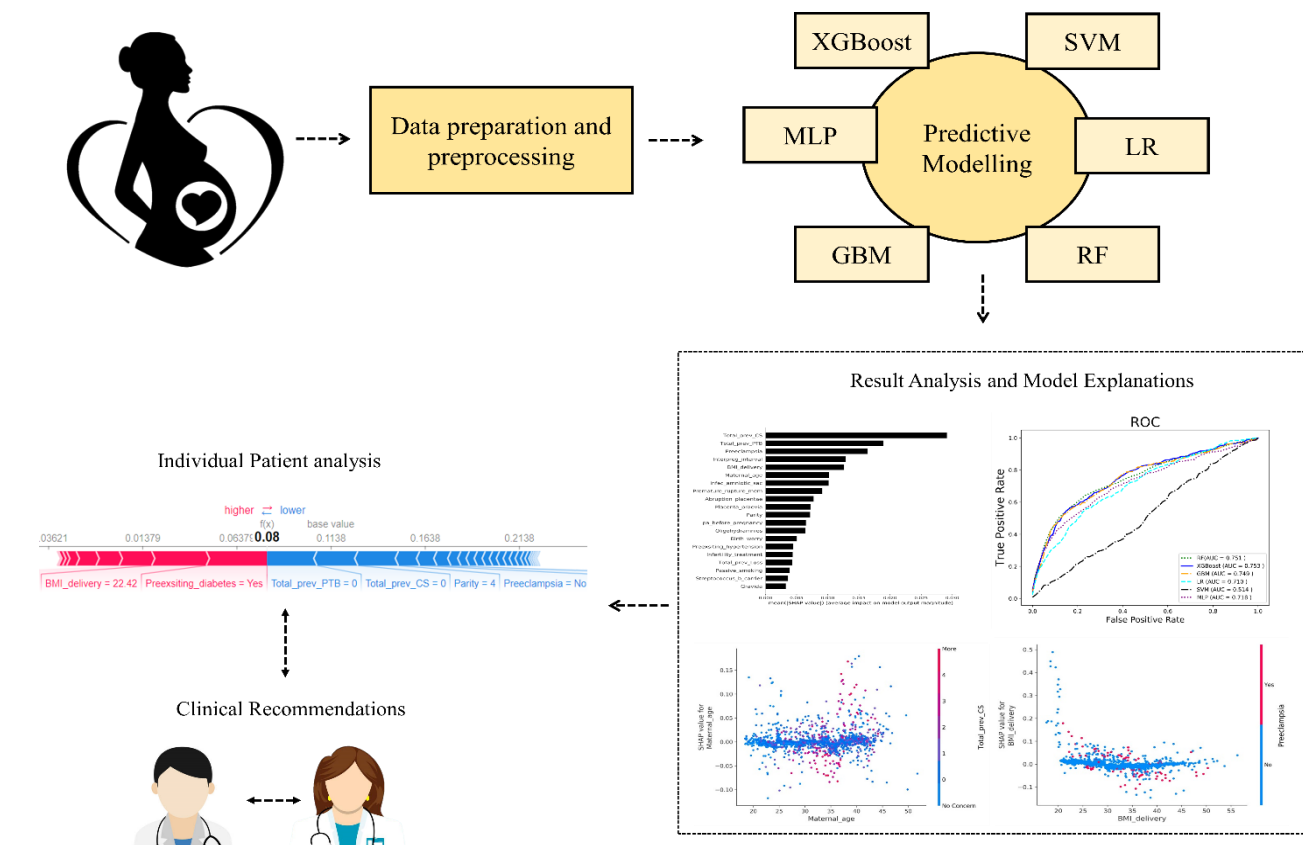
Our study aims to identify risk factors and predict PTB in a population of Emirati pregnant women using ML models and to explain and interpret this prediction via SHAP and LIME algorithms applicable to the clinical domain.

Method

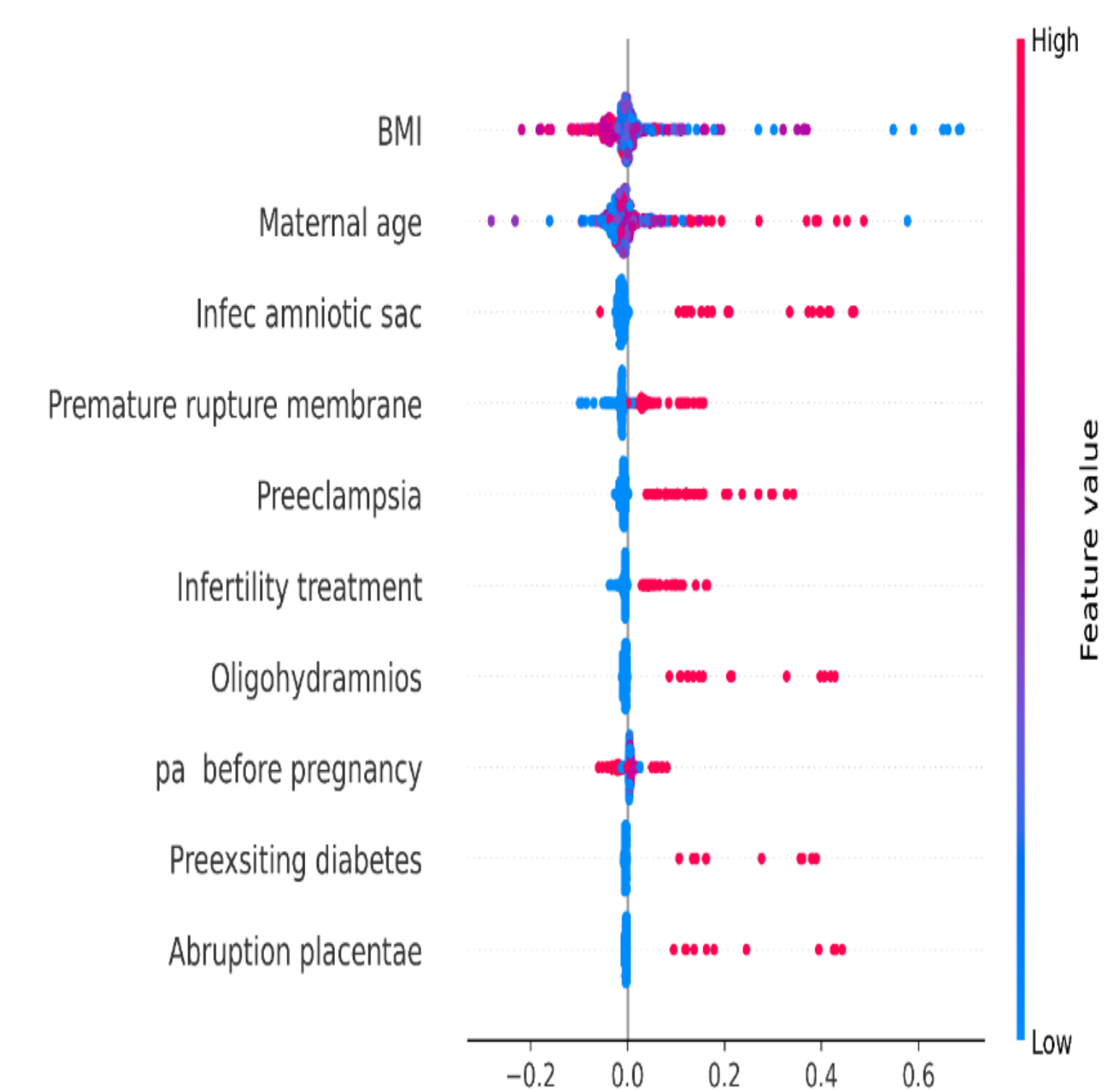
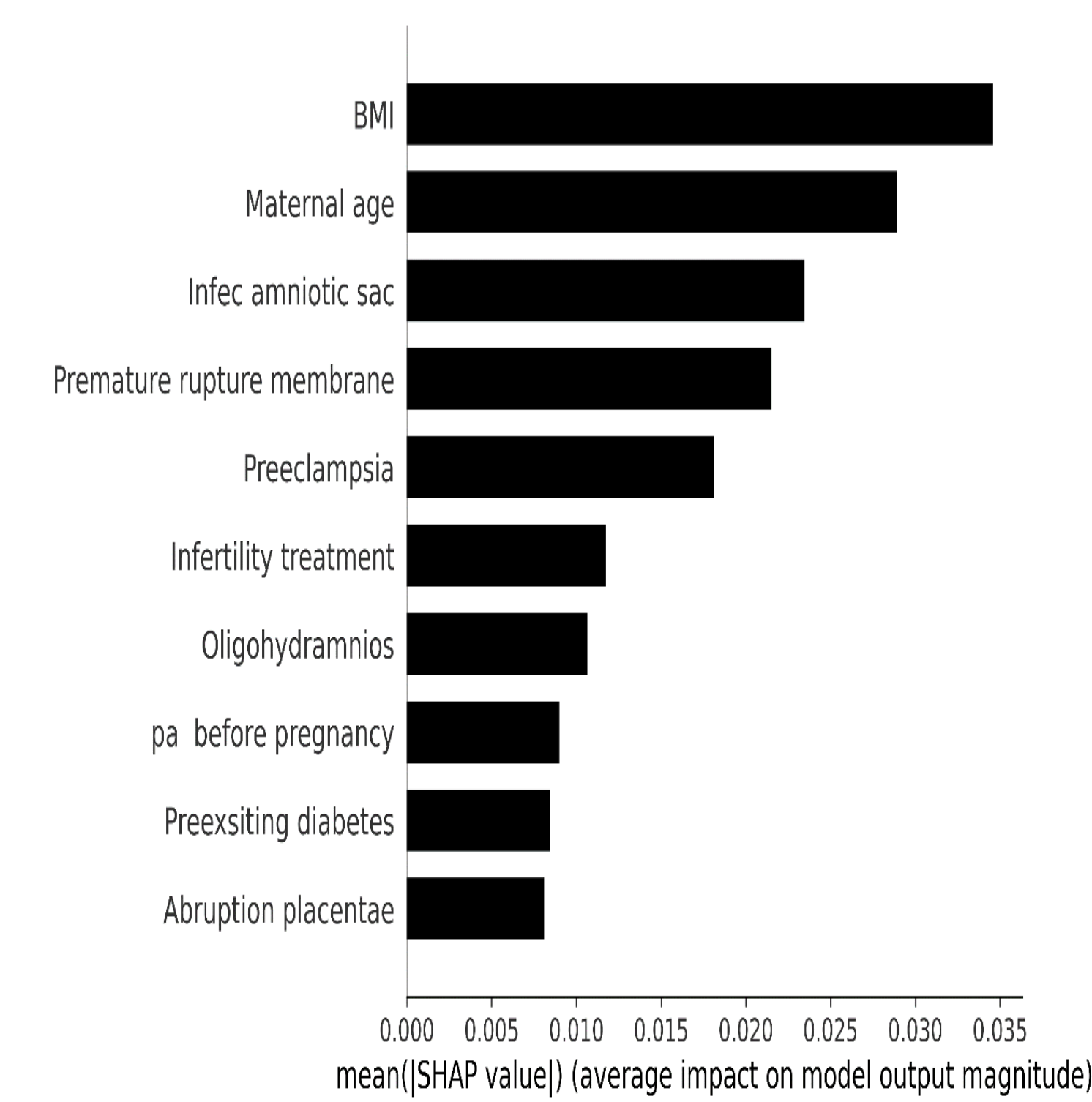
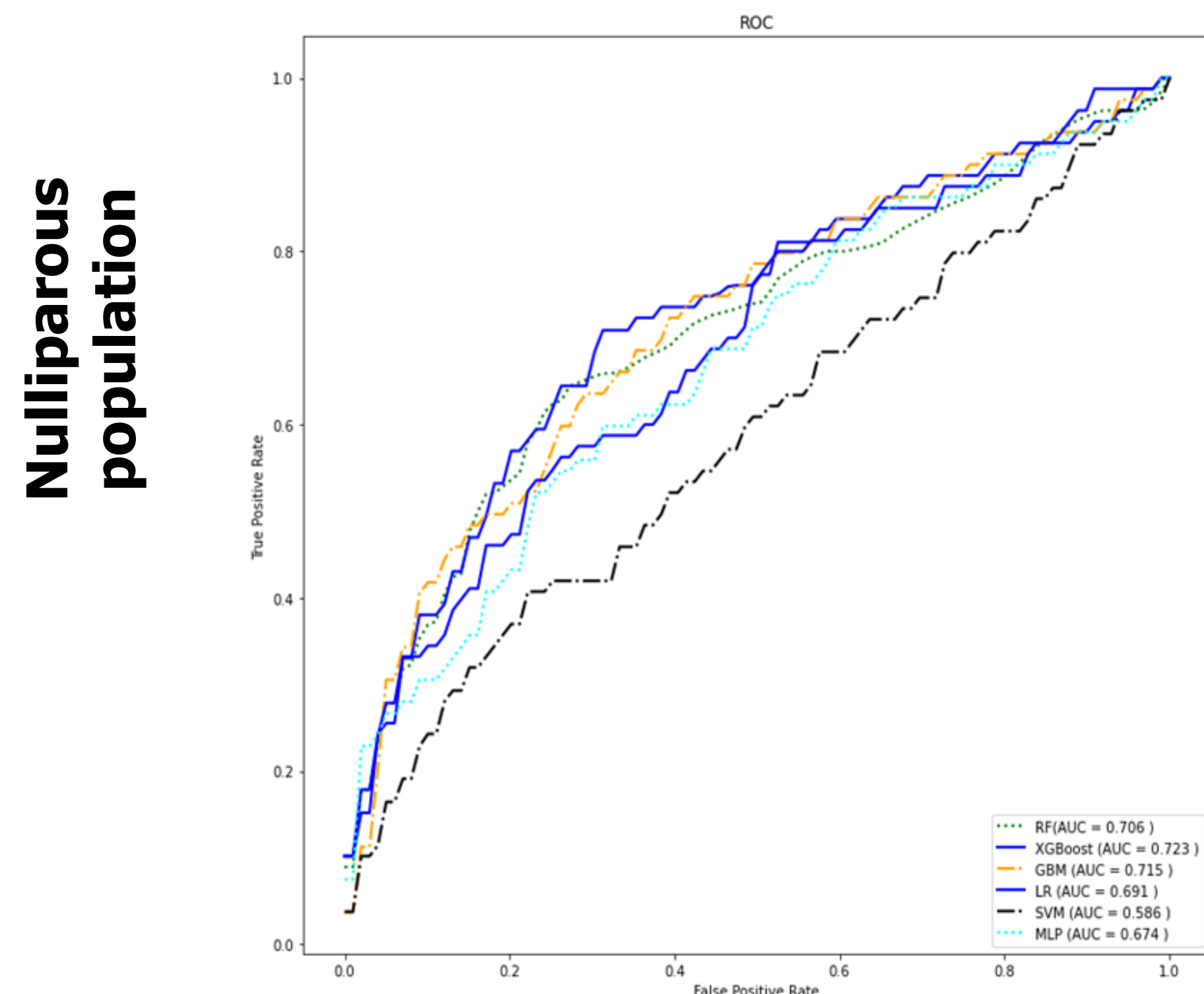
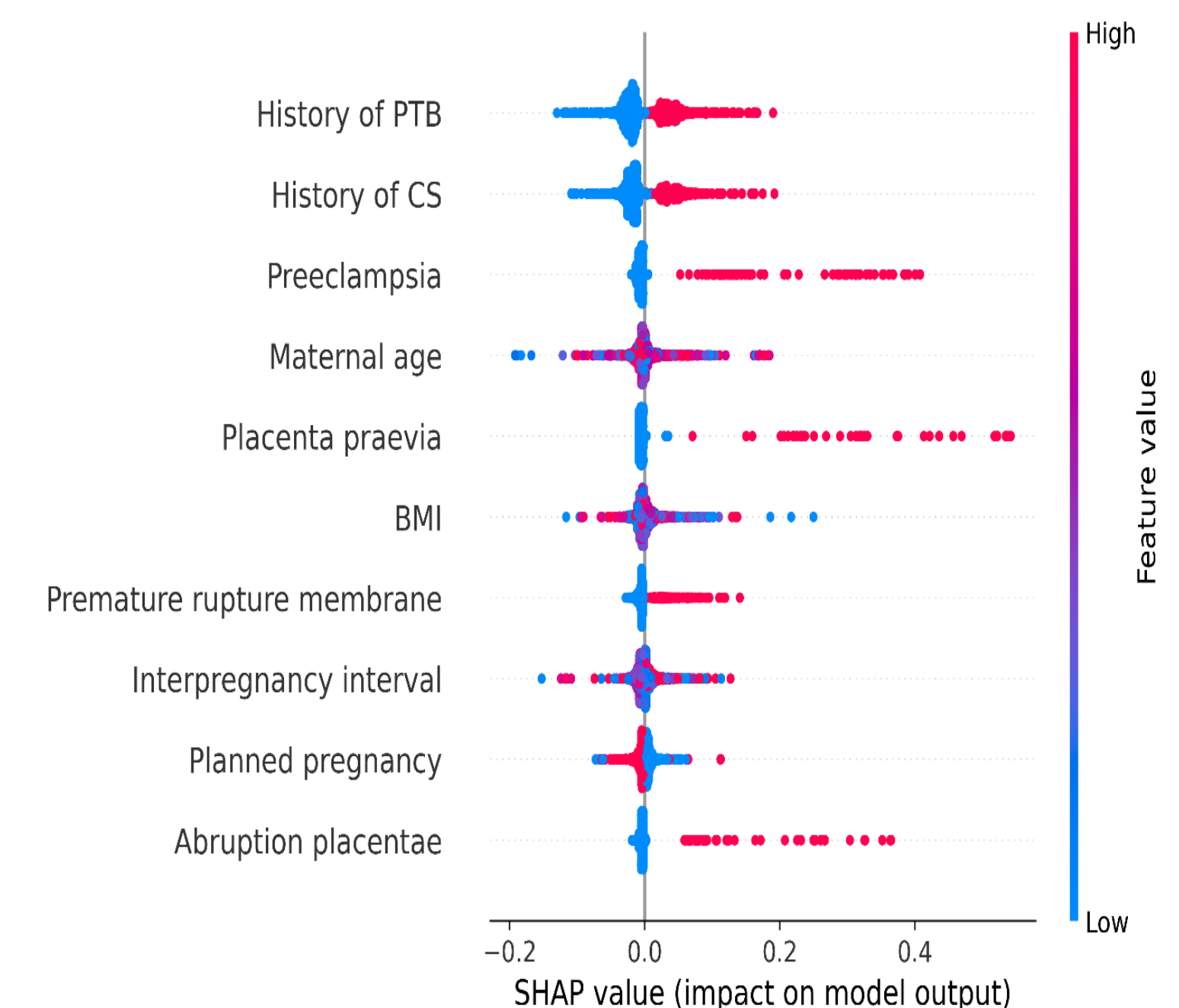
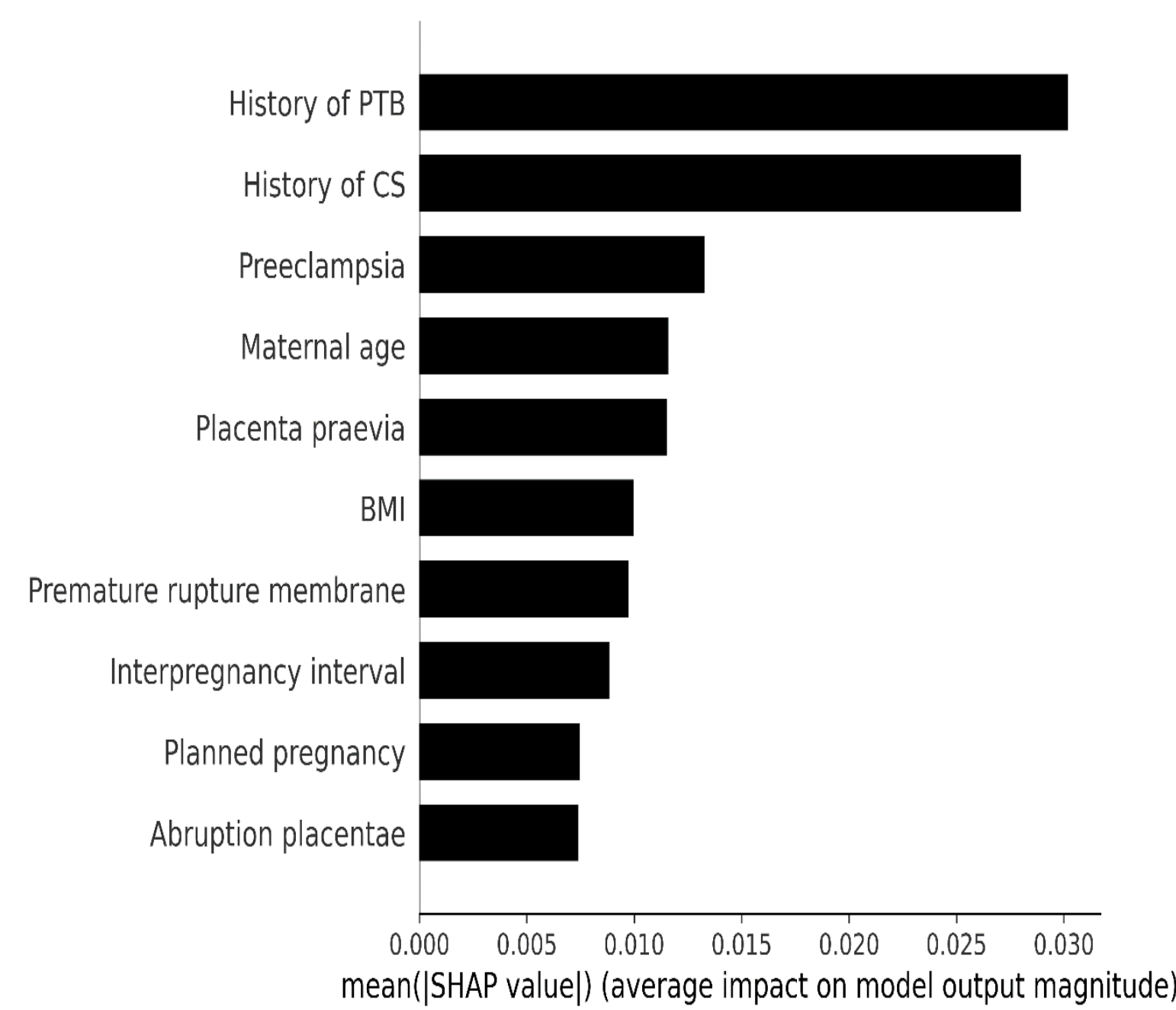
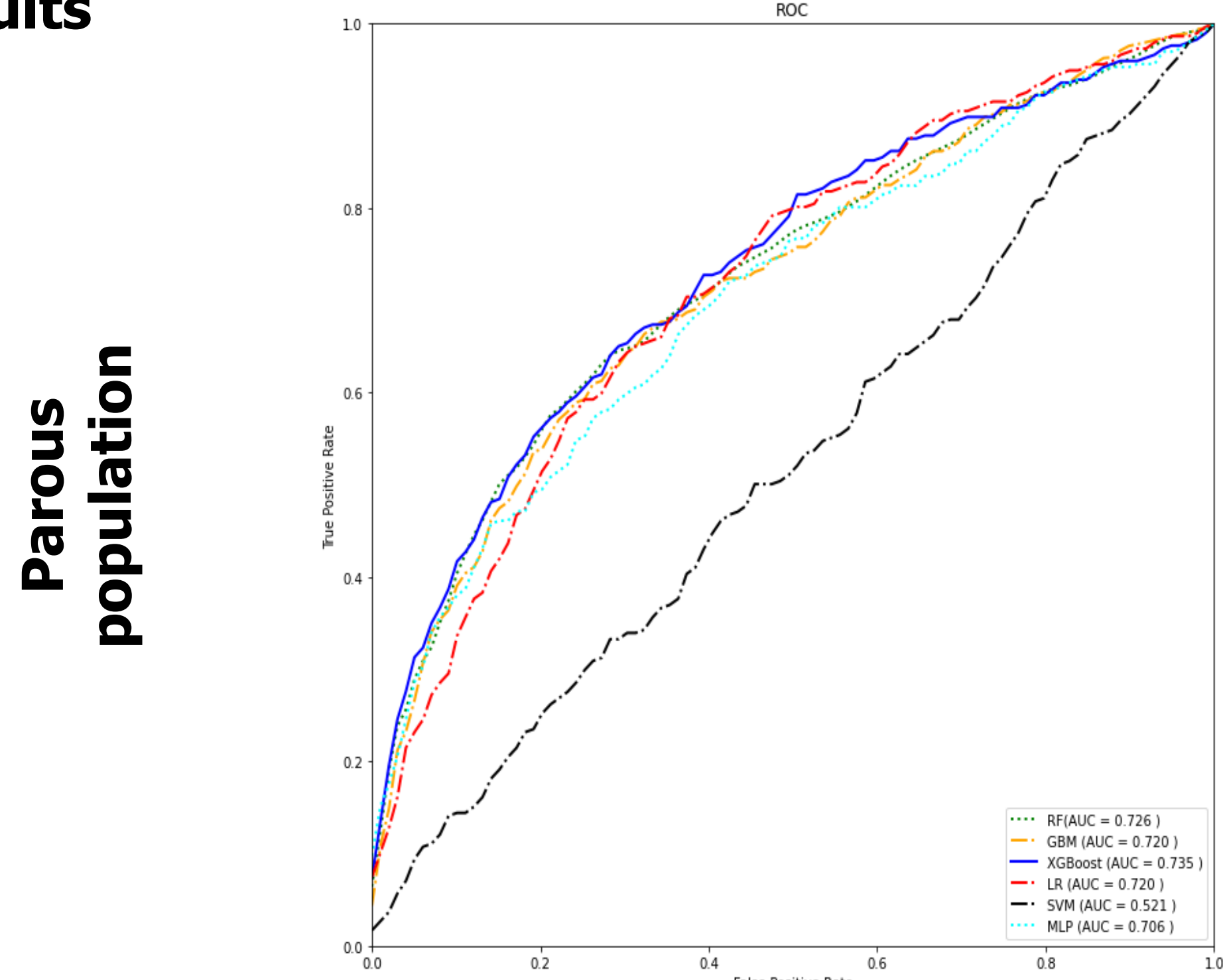
A dataset of 3,509 pregnant women from the Emirati population in (UAE) was used. Parous population (n=2708) and Nulliparous population (n=801) were studied.

Based on medical and AI literature, 35 risk factors associated with PTB were selected.

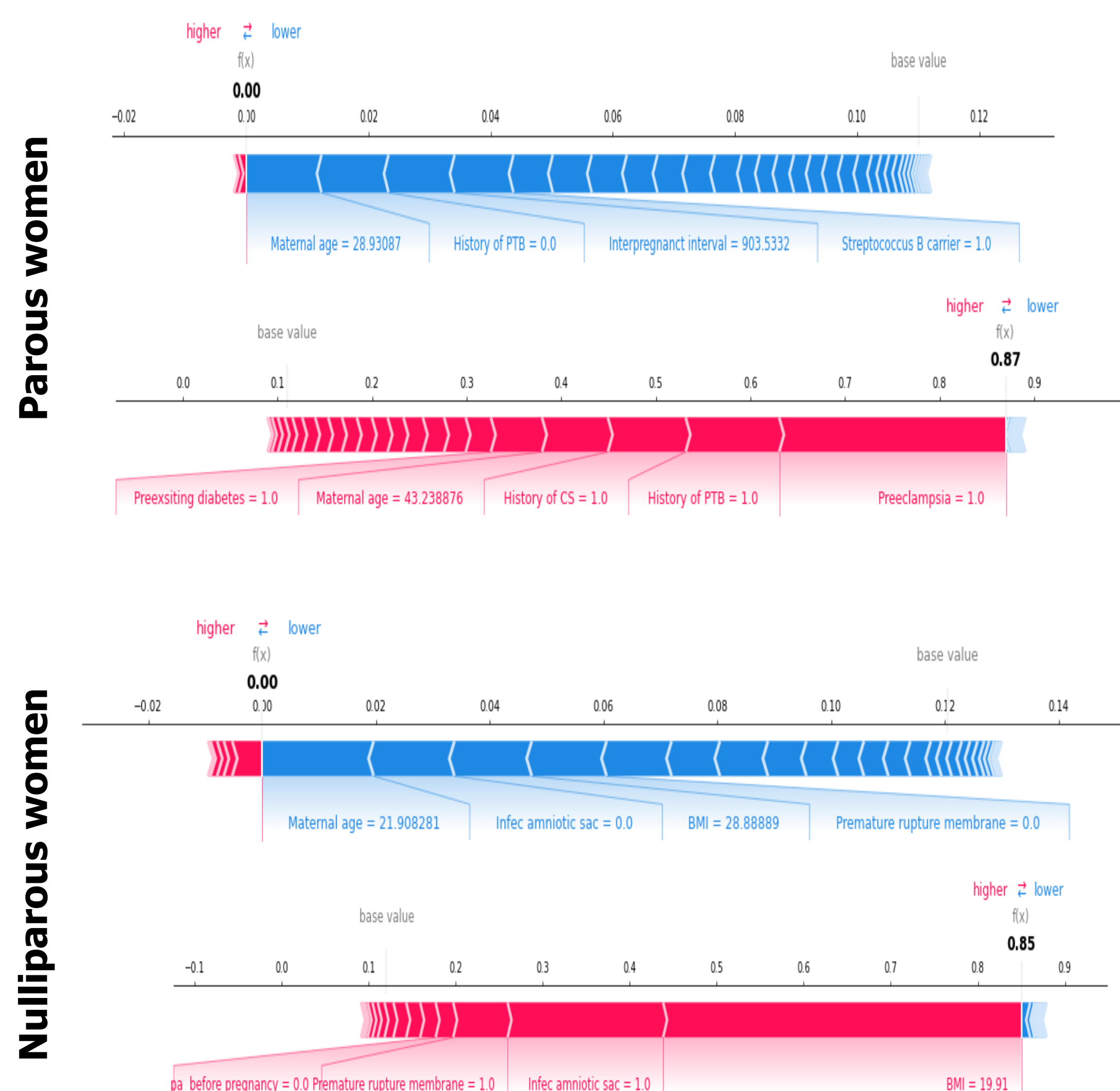
To evaluate the performance of the PTB prediction method, six ML algorithms were tested. The most important risk factors were identified using the SHAP feature attribution framework. An individual patient analysis was performed using SHAP and Local Interpretable Model-agnostic Explanations (LIME) to determine the risk factors associated with each patient.



Results



Individual risk analysis



Conclusions

- ❑ The overall prevalence of preterm birth in this study was 11.23% (11% parous population and 12.1% in nulliparous population).
- ❑ The XGBoost model achieved the best performance (AUC of 0.735 in the parous population, AUC of 0.723 in the nulliparous population).
- ❑ The main risk factors associated with preterm birth in this parous population are previous preterm birth, 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 (during pregnancy) are the most relevant risk factors.
- ❑ The trained Machine Learning prediction model can potentially be used as a screening tool for preterm birth prediction in our population

References

1. Prediction and Prevention of Spontaneous Preterm Birth: ACOG Practice Bulletin, Number 234. *Obstet Gynecol.* 2021;138(2):e65-e90. doi:10.1097/AOG.0000000000004479
2. Belaghi RA, Beyene J, McDonald SD. Prediction of preterm birth in nulliparous women using logistic regression and machine learning. *PLOS ONE.* 2021;16(6):e0252025. doi:10.1371/JOURNAL.PONE.0252025A.
3. Al Haddad et al., "Mutaba'ah—Mother and Child Health Study: protocol for a prospective cohort study investigating the maternal and early life determinants of infant, child, adolescent and maternal health in the United Arab Emirates," *BMJ Open*, vol. 9, no. 8, p. e030937, 2019.

Acknowledgments

- The authors would like to acknowledge and thank all the participants of The Mutaba'ah Study, and Tawam Hospital
- This work was supported by a grant from United Arab Emirates University (31R239)