

Is Interpretable Machine Learning for Preeclampsia Prediction Better Than the Standard Approach?



Tomas Bosschieter¹, Harsha Nori², Ian Painter³, Rich Caruana², Vivienne Souter³

¹ Stanford University, Stanford, CA ² Microsoft Research, Seattle, WA ³ Foundation for Health Care Quality, Seattle, WA



INTRODUCTION

- Accurate assessment of preeclampsia (PE) risk early in pregnancy has the potential for preventive interventions.

OBJECTIVES

- To develop and externally validate an interpretable machine learning (ML) model for prediction of preterm PE (PE with birth at < 37 weeks' gestation).
- To compare the performance of the interpretable ML model with the rules-based approach recommended by the US Preventive Services Task Force (USPSTF).

METHODS

- Retrospective study of **158,406 singleton and twin births** (20-43 weeks' gestation) at **20 US hospitals** (Jan 2016 – Dec 2021).
- 111,694 births at 12 hospitals were used as training data, and 46,712 births at 8 different hospitals were used for external validation.
- An Explainable Boosting Machine (EBM) was trained to predict preterm PE using **22 features** available at the first prenatal visit, **10 of which are also used in the USPSTF approach**.
- The EBM was also trained on these **10 features** for a direct comparison against the USPSTF approach.

RESULTS

- Preterm PE occurred in 2.3% (Table 1).
- BMI, nulliparity, and maternal age were the strongest predictors of the preterm PE in the EBM model (Figure 1).
- Compared to the USPSTF, the EBM demonstrated higher sensitivity for preterm preeclampsia: 74.4% versus 67.7% at a 37.7% screen positive rate (SPR).
- Using all 22 features, the EBM had a 76.5% sensitivity at a 37.7% SPR. With both feature sets, the EBM made statistically significantly more correct predictions compared to the USPSTF approach (McNemar's test: $p < 0.001$).

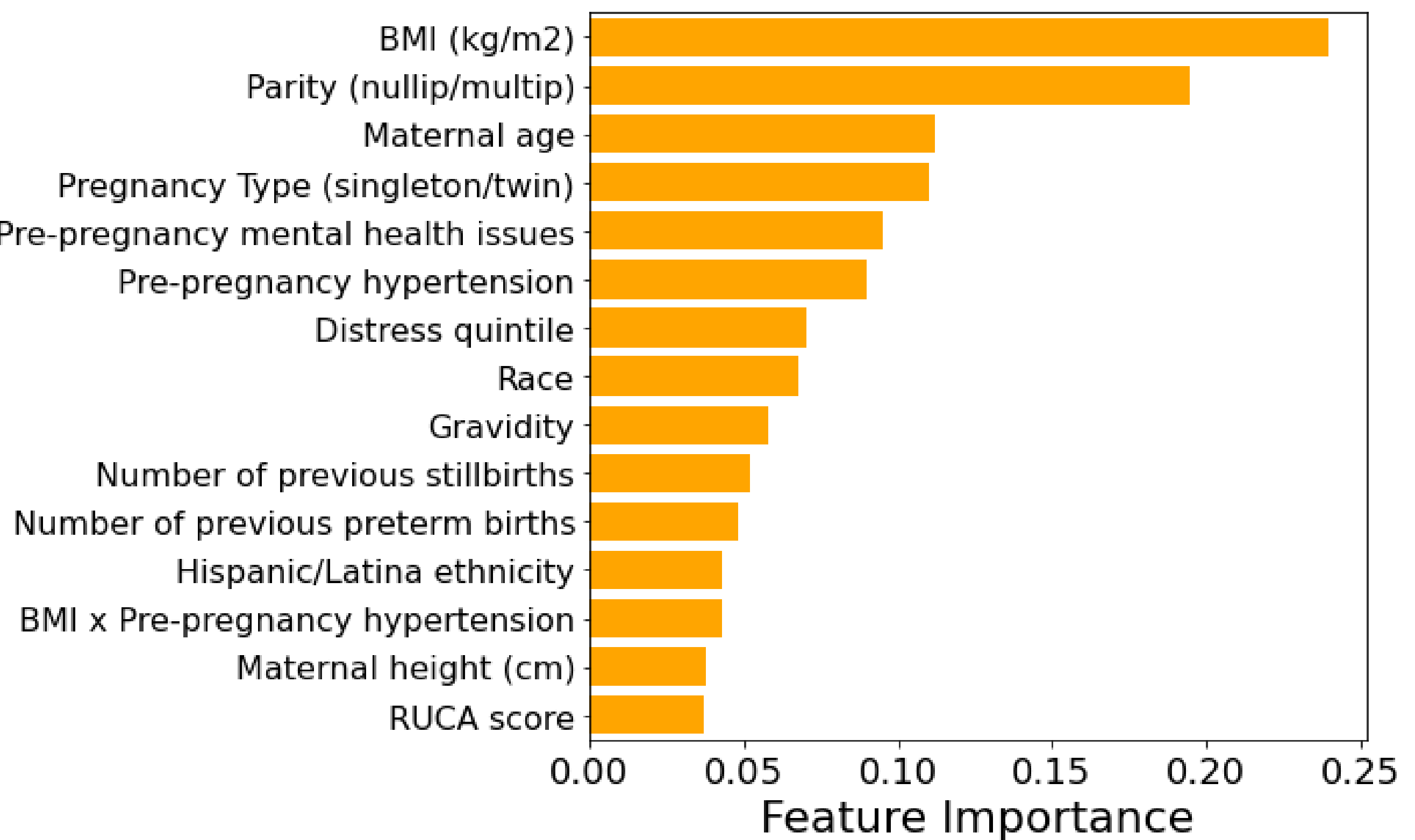
Table 1 Characteristics of study population.

	Preterm PE N = 3634 (2.3%)	No preterm PE N = 154,772 (97.7%)
Median initial BMI (kg/m ²)	28.9	25.7
Median age (years)	32.0	30.0
Median height (cm)	163.0	163.0
Nulliparous	48.8%	40.0%
Self-reported race	55.7% White 11.6% Asian 8.3% Black 2.5% Native Hawaiian, Pacific islander	54.7% White 16.6% Asian 5.5% Black 1.7% Native Hawaiian, Pacific islander
	2.1% American Indian, Alaska Native	1.4% American Indian, Alaska Native
	12.1% Other or Mixed Races	13.4% Other or Mixed Races
	7.8% Missing	6.6% Missing
Self-reported Hispanic or Latino ethnicity	14.4%	16.2%

Table 2 Both the 10 and the 22 variable EBMs had a higher sensitivity for preterm PE than the USPSTF approach at the same screen positive of 37.7%.

	Screen Positive Rate n (%)	Sensitivity (95% confidence interval)
USPSTF High risk	17620 (37.7%)	67.7% (65.5-69.9%)
EBM using 10 variables	17620 (37.7%)	74.3% (72.0-76.7%)
EBM using 22 variables	17620 (37.7%)	76.5% (74.7-78.3%)

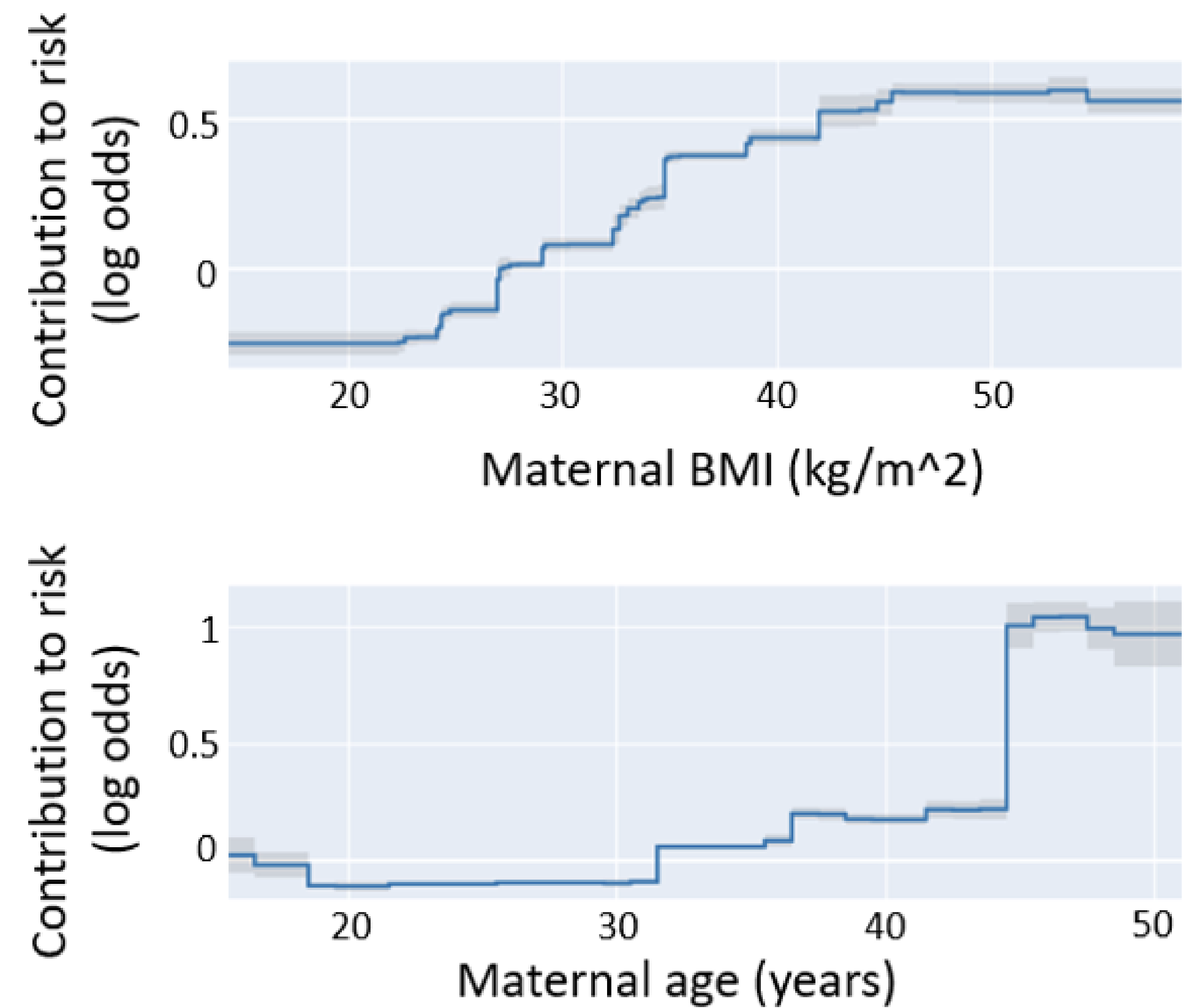
Figure 1 Most important predictors of preterm preeclampsia. Each feature's average contribution to risk (Shapley value) is visualized.



* The *Distress quintile* is from Economic Innovation Group's Distressed Communities Index (<https://eig.org/distressed-communities/>).
* The *RUCA* score refers to Rural-Urban Commuting Area Codes (<https://depts.washington.edu/uwruca/ruca-uses.php>).

Our interpretable ML outperformed the USPSTF rules-based approach, even when using the same 10 clinical variables

Figure 2 Contribution of maternal age and BMI to preterm preeclampsia risk. The higher the BMI and age, the higher the risk.



CONCLUSION

- Our interpretable machine learning model (EBM) outperformed the USPSTF rules-based approach, even when using the same 10 clinical variables. Additional variables led to improved discrimination.
- The EBM also generalized well in the external validation cohort.
- The strongest predictors of preterm PE risk were BMI, nulliparity and maternal age.
- Clinical information available at the first prenatal visit can be used to accurately evaluate risk of preterm supporting the potential to enhance risk prediction and management strategies for preterm preeclampsia.

This work was supported by Azure sponsorship credits granted by Microsoft's AI for Good Lab.