Hypoxia-Identifying Explainable AI for Risk Tracking in Fetal Cardiotocography

Jasmine Dhaliwal | Lauren DiPaola | Abigail Pomeroy | Swetha Saravanan
Faculty Advisor: Professor Lu Wang, Ph. D | Graduate Student Advisor: Ruiheng Yu, B. E.
University of Houston | Cullen College of Engineering | Department of Biomedical Engineering

BACKGROUND

- **Intrauterine** hypoxia is a leading cause of stillbirths worldwide.
- **Cardiotocography** (CTG) is used to diagnose fetal hypoxia as **Normal, Suspect, or Pathologic** (NSP) based on fetal heart rate and uterine contractions.
- CTG is used worldwide during labor but is **subjectively interpreted**, often inconsistently. Clinicians could benefit from an **AI tool which provides a second opinion** or standard.

With an approach that emphasizes both performance and explainability, **AI could pave the way for standardization of CTG analysis, ideally translating into better clinical outcomes in labor and delivery.**

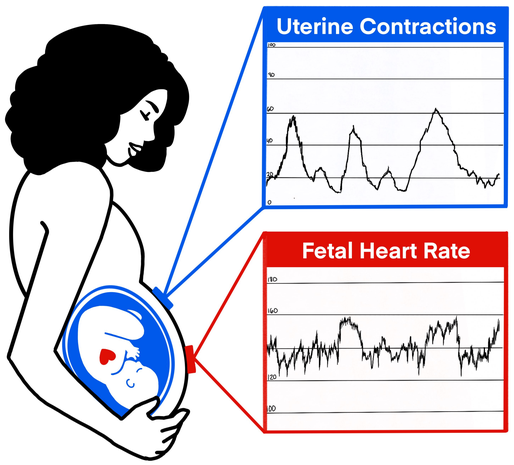

FIGURE 1: Sample CTG Waveforms

OBJECTIVES

- **Develop a machine learning (ML) model** that can **predict fetal NSP condition from CTG data** with high performance and interpretability.
- Meet or exceed the **F1 score of 0.9561** achieved in prior research.
- **Identify key influences in model prediction** using model explainability tools, such as SHAP and LIME.

RESULTS

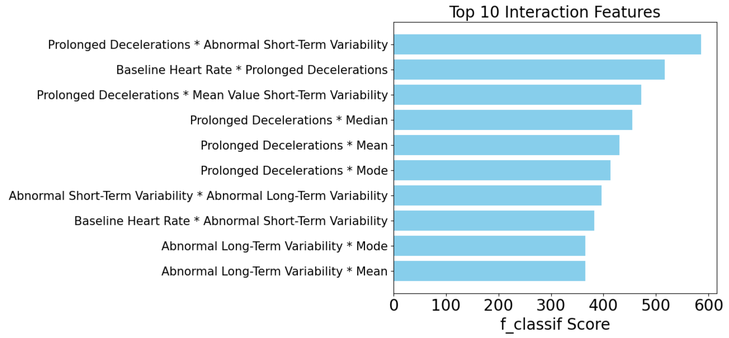

FIGURE 2: Feature interactions derived from the dataset ranked by statistical association to NSP classes, independent of model behavior.

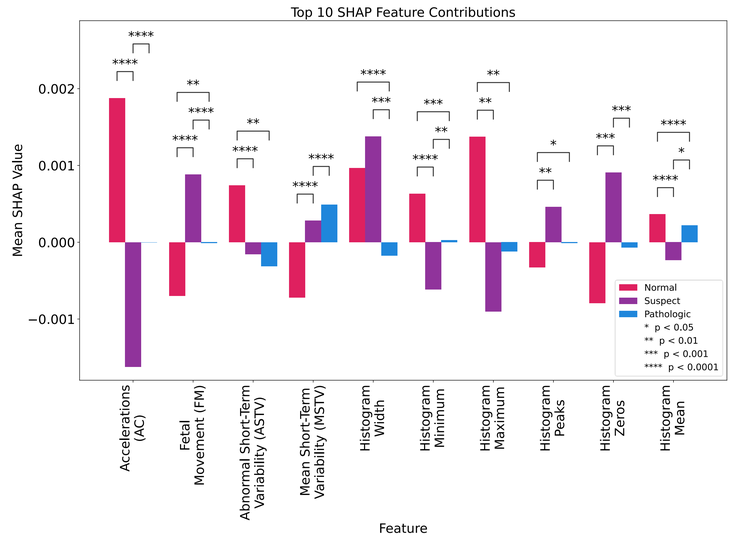

FIGURE 3: Average SHAP feature contribution to HEART-fC predictions across NSP classes.

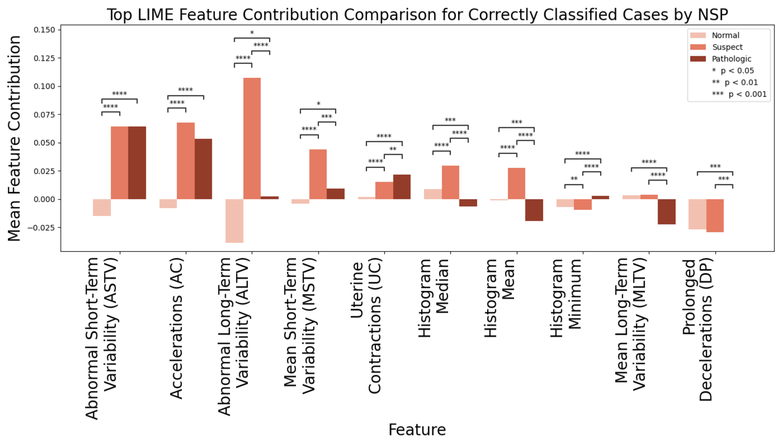

FIGURE 4: Average LIME feature contribution to HEART-fC predictions across NSP classes.

METHODS

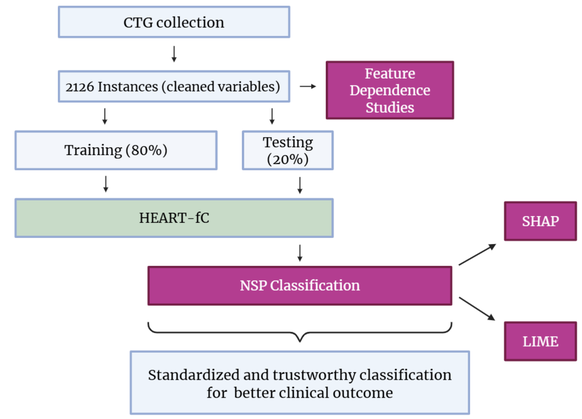

FIGURE 5: General System Workflow

CONCLUSION

- HEART-fC **correctly identified 95.80% of Normal, 90.62% of Suspect, and 96.55% of Pathologic instances.** Overall weighted F1 score was 0.9526.
- All **SHAP** class comparisons of feature importance for the **top 10 variables were statistically significant** except AC for N – P, ASTV for S – P, MSTV for N – P, Width for N – S, Max for S – P, Peaks for S – P, & Zeros for N – P. **(Fig. 3)**
- All **LIME** class comparisons (N – S, N – P, P – S) of feature contributions for the **top 10 variables were statistically significant** except AC for P – S, MLTV for N – S, & DP for N – S. **(Fig. 4)**
- **HEART-fC** demonstrates the development and application of **explainable, trustworthy AI in fetal medicine.**
- **Future Directions**: CLASS variable, Software as a Medical Device (SaaMD), XAI implementation into other healthcare fields.

ACKNOWLEDGEMENTS

We would like to thank Dr. Lu Wang, Ruiheng Yu, and the UH DHAI lab for their guidance in the development of our model. We would also like to thank Dr. Yuncheng Du and Dr. Jerome Schultz for their advice along the way and their consistent support of our endeavor. We would also like to thank the entire Biomedical Engineering department of the University of Houston, for their role in our development as researchers.