Introduction and challenges

Cardiotocography (CTG) is vital for detecting fetal distress, but subtle patterns are often missed in practice [1]. We developed a machine learning classifier using the UCI CTG dataset (n=2,126) to distinguish Normal, Suspect, and Pathologic states. The main challenges were class imbalance (8% Pathologic) and ensuring clinical interpretability.

Data Cleaning and Preprocessing: We removed missing values, duplicates, corrected data types, and excluded physiologically impossible values while retaining outliers to reflect emergencies. A stratified 80/20 split preserved class proportions.

Feature Engineering:
Clinically meaningful features were added, such as ASTV/ALTV ratio, deceleration severity score, AC per UC, and total abnormal variability. This enhanced predictive power while remaining interpretable.

Class Imbalanced Handling: We applied sample weighting (class_weight='balanced') (inverse class frequency), making rare Pathologic cases ~10× more influential during training.

Models and Results:
Four models were tested: Logistic Regression, Decision Tree, Random Forest, and XGBoost. Optimized for **balanced accuracy**, **macro F1**, and **Pathologic recall (>85%)**, the best model was **XGBoost** with:

- Balanced Accuracy: **94.7%**
- Macro F1-Score: **93.1%**
- Pathologic Recall: **97.1%** (34/35 identified, none misclassified as Normal)

Five-fold stratified cross-validation (random_state=42) confirmed robustness with mean

- **Balanced Accuracy 91.5% (±1.9%)**
- **Macro F1 88.9% (±1.9%)**
- **Pathologic Recall 92.1% (±1.4%)**, all folds exceeding the 85% safety threshold.

To ensure interpretability, we performed **SHAP analysis** on the final XGBoost model. Both SHAP and feature-importance rankings highlighted clinically meaningful predictors (ASTV, ALTV, Median heart rate).SHAP explainability showed predictions consistent with clinical guidelines: **ASTV >70** strongly indicated Pathologic risk, amplified when **ALTV >40**, reflecting compromised autonomic function. The only misclassified Pathologic case (ASTV=77, ALTV=4) was conservatively labelled *Suspect*, not Normal, illustrating the model's cautious handling of atypical presentations.
Engineered features such as **total_abnormal_var** ranked among the top predictors, validating the value of domain knowledge. This supports both **medical safety** and **trust in predictions**.

Conclusion

With **97.1% Pathologic recall**, low variance across folds, and interpretable SHAP insights, XGBoost delivers clinically safe, transparent, and efficient (6.9ms latency, 0.53MB model) fetal distress detection

References

[1] ACOG (2009). Intrapartum Fetal Heart Rate Monitoring. Practice Bulletin No. 106.
[2] Campos, D. & Bernardes, J. (2000). Cardiotocography [Dataset]. UCI Machine Learning Repository. https://doi.org/10.24432/C51S4N
[3] He H, Garcia EA (2009). Learning from Imbalanced Data. IEEE Trans Knowl Data Eng, 21(9):1263-1284.