**Fetal Distress Classification Using Cardiotocography (CTG) Data**

**1. Introduction**

The objective of this project is to develop a machine learning model that can assist clinicians in detecting fetal distress using real-world cardiotocography (CTG) recordings. The model aims to classify fetal conditions into Normal, Suspect, and Pathologic categories. CTG signals are widely used in obstetrics to monitor fetal well-being, with key features such as the baseline fetal heart rate (LB), accelerations (AC), decelerations (DS, DP, DR, DL), and variability indices (ASTV, MSTV, ALTV) offering important diagnostic insights.

**2. Data Preprocessing Pipeline**

The CTG dataset comprises 2,126 samples and 21 features. The preprocessing workflow begins with a data integrity check to ensure there are no missing or anomalous values. Physiologically implausible readings, such as extremely low or high fetal heart rates, are removed to eliminate noise.

All numerical features are standardized using z-score normalization to ensure comparability between different variables. The data is then divided into training and testing subsets, typically using an 80/20 stratified split, to preserve class balance and enable reliable model validation.

Feature engineering focuses on ensuring that each variable aligns with its clinical meaning, such as confirming that accelerations reflect short-term increases in fetal heart rate and variability measures capture the degree of fluctuation between beats. Additional derived features, such as total deceleration counts, may also be included to capture the combined effects of stress indicators.

**3. Model Design and Rationale**

The problem is formulated as a three-class classification task. Several models are considered, including Logistic Regression, Random Forest, and Gradient Boosting classifiers. Logistic Regression provides interpretable results and helps understand how clinical variables contribute to fetal distress, while Random Forests capture complex nonlinear relationships and offer feature importance measures. Gradient Boosting (such as XGBoost) is often chosen for its superior accuracy on structured datasets like CTG.

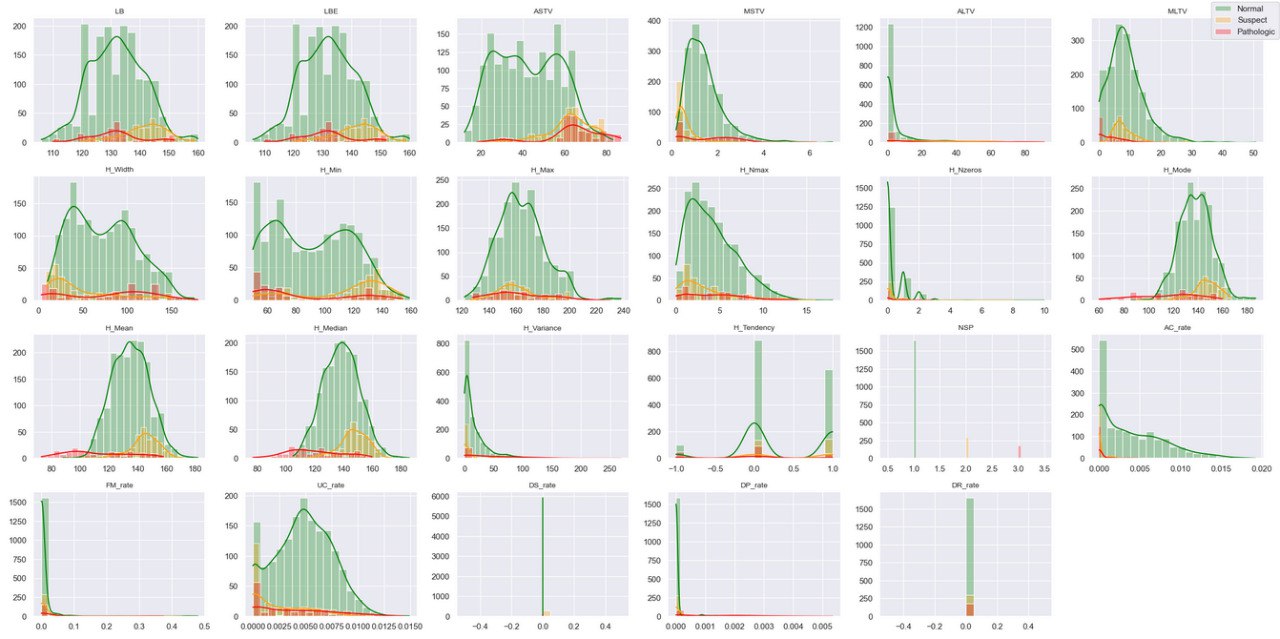
To further improve robustness and generalization, a stacked ensemble model was implemented. This approach combines the predictive strengths of Random Forest, XGBoost, and an MLP neural network, allowing the meta-learner to learn optimal weighting among their outputs.

The ensemble effectively balances linear interpretability and nonlinear flexibility, improving macro F1-score and reducing variance introduced by any single classifier. In medical diagnosis contexts such as fetal health classification, this ensures that borderline or ambiguous cases are more reliably identified, reducing the risk of false negatives in the “Suspect” and “Pathologic” categories.

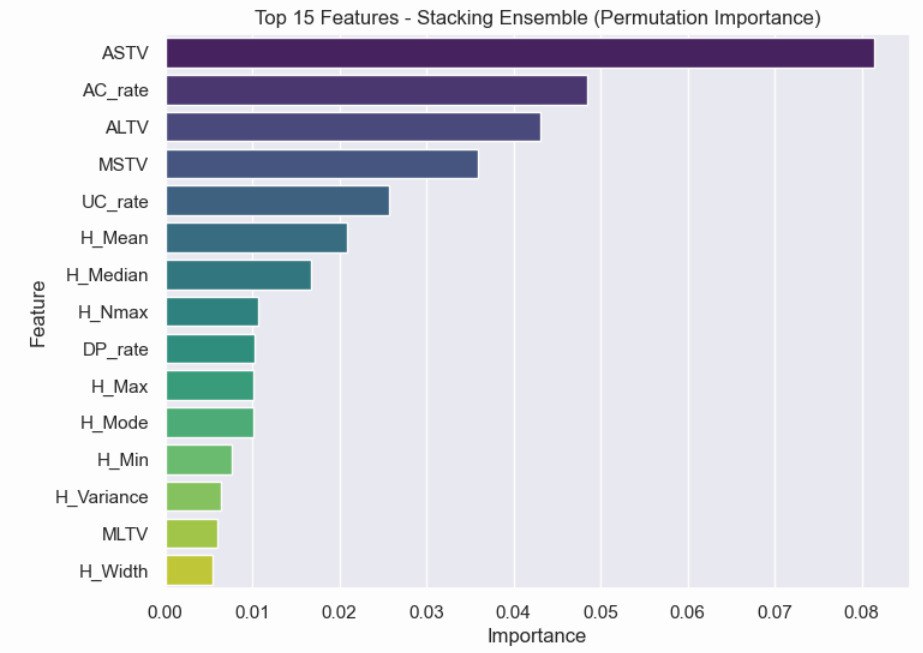
The final model is tuned using cross-validation to optimise parameters such as tree depth and learning rate. To address the inherent class imbalance, class weighting or stratified sampling techniques are applied. The emphasis is on building a model that not only achieves strong predictive performance but also maintains interpretability to support clinical decision-making.

**4. Evaluation and Interpretation**

Model performance is assessed using accuracy, precision, recall, and F1-score, with a particular focus on the model’s ability to correctly identify Pathologic cases. Confusion matrices and ROC curves are used to visualize classification performance. Feature importance and SHAP (Shapley Additive exPlanations) values are employed to interpret the model and communicate results in a clinician-friendly manner.



The feature distribution plots reveal strong class imbalance, with most samples belonging to the Normal category. Several features, including ASTV, MSTV, and ALTV, display partial separation between classes, while others show significant overlap, suggesting complex multivariate relationships. Many features are right-skewed and non-Gaussian, reinforcing the use of ensemble and tree-based classifiers that handle non-linear feature interactions without distributional assumptions. Overall, these visual insights support the motivation for applying SMOTE to address imbalance and for adopting robust ensemble models for classification.”



The final system highlights medically relevant patterns such as reduced variability and frequent severe decelerations as key predictors of fetal distress. This ensures that the solution not only performs well statistically but also aligns with clinical intuition.

**5. Conclusion**

The resulting model demonstrates that a well-prepared CTG dataset, combined with robust machine learning techniques, can effectively assist in fetal monitoring. Through careful preprocessing, model selection, and interpretability analysis, the project bridges technical modeling with clinical relevance, supporting safer and more informed obstetric decisions.

**Bibliography**

1. Ayres-de-Campos, D., et al. UCI Cardiotocography Dataset Documentation. UCI Machine Learning Repository.
2. American College of Obstetricians and Gynecologists (ACOG). Practice Bulletin: Intrapartum Fetal Heart Rate Monitoring.
3. Magen, N., et al. (2020). Machine Learning Models for Fetal Distress Classification Using CTG Signals. IEEE Access.
4. Georgoulas, G., et al. (2006). Classification of Fetal Heart Rate Signals Using Support Vector Machines. Computers in Biology and Medicine.
5. Chudáček, V., et al. (2014). Automatic Evaluation of Intrapartum Fetal Heart Rate Recordings: A Comprehensive Review. Physiological Measurement.