We built a machine-learning model that classifies fetal well-being based on cardiotocography readings (dataset with around 2126 samples) into three classes: Normal (1), Suspect (2), and Pathologic (3) in order to help in early risk detection using non-invasive fetal monitoring data.
The dataset was highly imbalanced with a **much higher proportion of Normal to Suspect and Pathologic**, with pathologic represented the least. Thus, the performance was evaluated using Balanced Accuracy and Macro-F1 score, to ensure **equal weight** for each class regardless of frequency.

Methodology: After cleaning and keeping numeric CTG features, we performed a stratified 80-20 split to preserve class proportions. We explicitly framed the CTG features and considered how **clinicians prioritise/read a CTG** (fewer accelerations, more prolonged decels, and reduced long-term variability = higher risk). Similarly, our model considers **baseline > variability > accelerations and decelerations** to determine Normal/Suspect/Pathological risk. For instance, our model also noticed that **reduced variability + late deceleration had higher risk**.

We built an **Imbalanced-Learn Pipeline** comprising of StandardScaler -> SMOTE -> RandomForestClassifier(class_weight="balanced") to ensure no data leakage between sampling, scaling, and model fitting. We also ensured that the pipeline checked for missing values/duplicates and flagged out physiologically implausible outliers so that the model quality didn't suffer - allowing for excellent exception handling. At the same time, it was crucial to engineer "flat_trace_risk" when **ASTV/ALTV/Width/Variance are very low** (clinically **concerning** trace)

We also generated **correlation heatmaps and class-wise distributions/ boxplots** so that we could draw conclusions like "low short-term variability trends towards Pathologic group" as opposed to providing clinicians with raw coefficients. These visuals were also selected because they echo what clinicians already look out for in a CTG, namely - **spread, central tendency, co-movement of AC/FM with UC/decels, keeping the narrative intuitive**.

Class imbalance was handled via SMOTE oversampling and `class_weight='balanced'`.
Comparing between Logistic Regression, MLP, and Random Forest, **Random Forest** was selected as it can handle **mixed-scale features** and **non-linear relationships** in CTG data, while still being explainable through feature importance and SHAP values.

FEATURE	**INSIGHT GAINED FROM OUR MODEL**
LB (Baseline FHR) – Average fetal heart rate (bpm) where normal is 110 to 160	Extremes < 100 or > 170 often found in Pathological cases. Central tendency of FHR (Mean/Median/LB): Pathologic traces **skew lower overall** (mean ~112 bpm vs ~135 bpm in Normal), **consistent with more frequent/longer decels**.
AC (Accelerations) – short bursts of increased heart rate due to fetal movement	Low acceleration counts (<2 per min) often found in Suspect & Pathological Cases. AC>0 appears in **69.3% of Normal traces** but only **17–18% of Suspect/Pathologic.**
FM (Fetal Movements) – number of fetal movements per second of recording	Low FM positively correlates with high ASTV
UC (Uterine Contractions) – maternal uterine activity	Mild positive correlation with pathologic outcomes under stress conditions. **(Focused on this due to our rationale: False Positive > False Negative)**
ASTV (Abnormal Short-Term Variability) – % of time where SHORT-term variability is abnormally low	High ASTV commonly found in Pathological cases
ALTV (Abnormal Long-Term Variability) – % of time where LONG-term variability is abnormally low	Elevated ALTV indicated reduced autonomic response, signaling possible fetal hypoxia. mLTV<5 occurs in **70.7% Pathologic** vs **26.8% Normal**
DL, DS, DP – Diagnostic pattern summaries from obstetric rules	DL & DP positivity strongly aligned with Pathologic cases in exploratory data analysis. DP>0 jumps from **4.3% (Normal)** to **55.7% (Pathologic)**

Boxplots crafted to develop "feature exploration with clinical sense" revealed that babies with very low variability (high ASTV & ALTV) and lower AC & FM than normal were often Pathologic. High accelerations and moderate baseline FHR were strong indicators of Normal status. Correlation heatmap showed ASTV was positively correlated with ALTV (correlation coefficient r = around 0.72), both of which negatively correlated with AC and FM. Feature importance plots showed **ASTV, ALTV, DL, and AC as the strongest predictors**.

Results showed a **Balanced Accuracy** around 0.88, **Macro-F1** around 0.84, and a **confusion matrix** showing high recall for Normal class (0.95), with improved Pathological recall from 0.74 to 0.81 after SMOTE. **Per-class recall:** Normal **0.984**, Suspect **0.547**, Pathologic **0.879** (the rare but critical class is caught with high sensitivity).

Thus, our model correctly identified more than 80 % of Pathologic cases while keeping false alarms moderate. EDA and cross-fold consistency showed no overfitting. SHAP plots confirmed that higher ASTV and ALTV predict Pathologic cases. Hence, by addressing **class imbalance** with stratified sampling and possibly SMOTE, we successfully minimised the chance of a false negative (i.e., missing a distressed fetus).