**Introduction**

Fetal distress during labor can lead to severe complications if not detected early. Manual interpretations of fetal heart rate (FHR) and uterine contractions (UC) patterns is prone to errors hence we designed a solution that automatically identifies signs of fetal distress by analysing cardiotocography (CTG) datasets in order to detect subtle indicators of fetal compromise that may otherwise go unnoticed.

**Data Cleaning**

The raw cardiotocography (CTG) dataset initially contained extraneous metadata rows, duplicated columns and potential label leakage from medical indicators such as CLASS and SUSP. To ensure integrity analytically, all the empty records and unnamed columns were systematically removed. Missing values were identified and addressed using median imputation, preserving the dataset's central tendencies and removing any bias. Any potential outliers in physiological parameters, such as FHR and accelerations, were assessed using boxplots and z-scores in order to ensure that no extreme values would warp the model performance. After verification, the target variable representing fetal state (Normal, Suspect, Pathologic) was numerically encoded to facilitate supervised learning and ease data processing

**Relationship Analysis**

Multiple analyses were used to understand the interrelationships between fetal features and maternal distress levels. A boxplot matrix was generated to visualise the distribution of physiological indicators across all three fetal states, Normal, Suspect and Pathological. A heat map quantified strong positive relationships between ASTV and ALTV (r ≈ 0.46) and moderate correlations of both with LB suggesting that increased baseline rate often coincides with greater variability. These relationships align with clinical understanding that increased baseline heart rate is often accompanied by reduced variability, a potential sign of fetal hypoxia. The dataset was then stratified into training and testing subsets to ensure balanced model representation. A Principal Component Analysis (PCA) was used to show a 2D understanding of the fetal state. Normal cases concentrated to the center while Pathologic cases were scattered elsewhere, reflecting deviation. These findings were further confirmed using k-means clustering with 63-94% of clusters aligning with our findings.

**Model Analysis**

Using multiple algorithms to test such as K-nearest neighbours and random forest, Gradient Boosting was the most accurate with highest balanced accuracy and F! Macro at 0.91 and 0.90 respectively. The table below explains some advantages.

| **Medical Concern** | **Gradient Boosting Advantage** |
| --- | --- |
| Subtle patterns that may be overlooked | Captures interactions among LB, ASTV, DS, etc. |
| Rare distress cases | Focuses on hard-to-classify samples |
| Real-world noise | Regularized and robust |
| Need for trust | Feature interpretability via SHAP |
| Early detection | Learns nuanced differences between “Suspect” and “Pathologic |

In conclusion, the model effectively detects subtle fetal distress patterns, enhancing real-time decision support during labor and contributing to improved perinatal outcomes.