**Data Processing Pipeline**

1. Data Cleaning and Quality Assurance:

We began by inspecting for missing values, outliers, and redundant attributes. Missing entries were imputed using median values to preserve distributional integrity, while non-informative features, such as constant decelerations, were dropped. Features showing high correlation (|r| > 0.9) were pruned to minimize multicollinearity.

2. Transformation and Normalization:

Several CTG features (Fetal Movements, Light/Severe/Prolonged Decelerations, Variance) exhibited strong skewness. Therefore, we apply a log(1+x) transformation to reduce this skew and improve normality. Standard scaling was applied to ensure uniform feature contribution during model training, allowing for stabilized convergence in SVMs and neural networks.

3. Label Encoding and Stratification:

Categorical target labels were encoded as numerical classes (1: Normal, 2: Suspect, 3: Pathologic). Stratified splitting ensured consistent class proportions across training and testing sets, while balanced class weighting addressed the dominance of Normal cases.

### **Feature Insights and Reasoning**

Exploratory data analysis revealed several clinically meaningful relationships that guided both feature selection and model design. Variability metrics such as ASTV, MSTV, ALTV, and MLTV were the strongest indicators of fetal well-being. Normal fetuses typically exhibited moderate variability in heart rate, while pathological cases showed patterns that were either too flat (indicating low responsiveness) or too erratic (indicating instability). Similarly, higher detections of Accelerations (AC) and Fetal Movements (FM) were closely associated with healthy outcomes, which reflects active and well-oxygenated fetuses, whereas their absence often signaled potential distress. The mean and variance of Fetal Heart Rate (FHR) show greater variability observed in at-risk fetuses. Although features such as Severe (DS) and Prolonged Decelerations (DP) did not appear often, their presence typically occurred in high-risk scenarios. In conclusion, fetal health is best captured through variability and responsiveness measures, where we emphasize these features in our modeling approach.

**Model Design and Rationale**

Our modelling strategy was to balance interpretability, performance, and clinical relevance. We began with baseline models, Logistic Regression and Decision Trees, to establish explainable benchmarks and gain an intuitive understanding of feature importance. From this, we then implemented ensemble models such as Random Forest and Gradient Boosting, which captured complex, nonlinear interactions between physiological variables and demonstrated robustness to noise and feature overlap. To further enhance predictive performance, we use Neural Networks equipped with Batch Normalization, Dropout, and L2 regularization to mitigate overfitting and improve generalization. We tried to fine-tune hyperparameters such as network depth, learning rate, and dropout ratio, ensuring a strong balance between accuracy and stability. Throughout all experiments, we used Balanced Accuracy and Macro-F1 Score to evaluate the model accuracy and precision, and ensure that each fetal health class, Normal, Suspect, and Pathologic, was represented fairly despite the dataset’s inherent imbalance.