**Main Problem:** CTG readings are complex, patterns can be subtle, and busy hospital wards make it difficult to spot problems quickly and accurately. Hence using these CTG records, we can create a model to accurately detect pathological and suspect cases.

**Data Cleaning:** We cleaned the dataset by removing invalid entries and duplicates, and replaced missing values with median values to reduce bias and improve accuracy.

**Data Analysing:** After obtaining the necessary datatypes, we used exploratory data analysis to visualise and identify correlations between variables in the cleaned dataset. We found that certain variables, such as acceleration (AL) and decelerations (DL, DS, DP), played a more significant role in affecting NSP values. For example, prolonged deceleration (DP) strongly indicated a pathological case (3). Overall, we learnt that leaky variables, features appearing similar across all NSP classes, do not provide useful information, while features that differ between classes are more informative. A naive (untrained) model would attempt to maximise accuracy by predicting the majority class, leading to class imbalance. This is undesirable as pathological cases would be overlooked. To address this, the target (NSP) needed to be prepared for machine learning with class imbalance properly handled. Using SMOTE, synthetic points were generated for minority classes (Pathological and Suspect) to balance the dataset. Additionally, class weights were assigned to prevent the model from over- or under-focusing on any particular class. This ensured that rare cases requiring intervention could be detected more effectively. K-means clustering was also applied to group cases based on similarity. However, results showed that the Suspect class did not form distinct clusters and instead overlapped heavily with Normal and Pathological cases, leading to frequent misclassification of “Suspect” due to its ambiguity.

**Model Designing and Rationale:** We started by training, testing and comparing the confusion matrix of various model types such as Logistic Regression, Decision Trees, Neutral Networks, etc, but found oftentimes Pathological cases had been highly misidentified giving a high false negatives rate. For such a high severity case where fetus’ need immediate medical interventions, many were classed as borderline or normal cases, resulting in fetus’ getting late medical attention. Overall, Gradient Boosting was a model that proved to have the highest recall for pathologic cases (92%) and suspect (88%), most accurately identifying the difference between the Normal, Suspect and Pathologic cases. Retraining the Gradient Boosting model on the full dataset, it was able to distinguish the 3 cases with close to 100% accuracy, and having negligible error.