## Datathon Academic Report: Early Detection of Fetal Distress

Objective

The model was trained using Cariotocography (CTG) data to classify each CTG trace into the respective NSP (Normal, Suspect, or Pathologic) classes so as to detect potential fetal distress. In order to prevent class imbalance and biased decision-making, the performance of the model has been evaluated using the Balanced Accuracy and Macro F1 scores.

Data Processing (Cleaning, Transformation, Preparation)

The CTG data was initially converted from .xls file format to .xlsx to for compatibility with pandas. Afterwards, the initial headers of .xlsx file were dropped and the first rows are reassigned to be dataframe’s headers. Columns with duplicated names were also renamed by appending “\_2” to ensure columns could be differentiated. Additionally, columns with missing values (NaN) were dropped as data completeness is crucial to ensure that the medical datasets remain accurate and reliable (Liu, 2023). The datatypes were then converted into float64 to ensure consistency for easier parsing.

The preprocessed dataset was then used to plot a correlation map, revealing that the data distribution is relatively symmetric. The correlation map allows us to infer that the significant patterns that allow us to perform feature selection by identifying the features with highest correlation with NSP. The feature analysis revealed that the ASTV was one of the strongest predictors of pathological cases, followed by MSTV and LB.

Model Design and Rationale

The initial model makes use of a Random Forest (RF) Classifier to ensure that the different correlation relationships can be represented by aggregating multiple decision trees (Sun, 2024). A key feature of the Random Forest model is that overfitting data is reduced by introducing randomness in feature selection so as to then identify the best features to use as a split point.

1. The preprocessed data is scaled before being divided by Stratified Shuffle Split so as to preserve the original proportions across varying training and test sets and ensure fair representation of the cases from differing NSP classes.
2. In order to find the optimal parameters, GridSearch was then used to search over hyperparameter combinations. Cross validation is then used to further adjust and improve the parameters. After the configurations were validated, the best parameters were selected by the best mean Balanced Accuracy scores so as to prioritise class fairness.

The RF model was then refined using SMOTE to ensure that the minority classes Suspect and Pathological classes were adequately represented (Matharaarachchi, 2024). After the SMOTE resampling was used, there was improved recall for minority classes which then improved the overall Macro F1 score.

The subsequent model using XGBoost makes use of gradient boosting, where new decision trees attempt to correct errors from previous trees so as to improve accuracy across minority-class patterns such as Suspect and Pathological cases (Matharaarachchi, 2024). By reducing the bias towards the majority Normal class, the model was able to form more refined decision boundaries, accounting for the increased Balanced Accuracy scores. The XGBoost model was then further adjusted by using SMOTE which then produced higher Balanced Accuracy scores.

Evaluation

After comparing the models, the RF+SMOTE model yields the smoothest ROC curves as well as the highest AUC values across all the classes. As the RF+SMOTE model delivers the best Balanced Accuracy and macro F1 scores, we can conclude that the model consistently demonstrates a stronger and more reliable ability to detect all fetal health states. While the XGBoost’s model offers gradient-based refinement, the RF model combined with the SMOTE’s class-balance correction collectively provided a fairer and more recall-sensitive model. These improvements are particularly crucial in medical classification tasks, where under-detection of abnormal cases can have significant clinical consequences. As our model focuses on variability and decelerations, our model successfully adheres to medical guidelines. By combining high test accuracy, minimal overfitting, and inherent robustness, the RF+SMOTE model is the most dependable and deployable solution for the critical task of fetal health classification.

References

Liu, M. (2023, August). *Handling missing values in healthcare data: A systematic review of deep learning-based imputation techniques*. ScienceDirect. <https://www.sciencedirect.com/science/article/abs/pii/S093336572300101X>

Matharaarachchi, S. (2024, December). *Enhancing smote for imbalanced data with abnormal minority instances*. ScienceDirect. <https://www.sciencedirect.com/science/article/pii/S2666827024000732>

Sun, Z. (2024, March 1). *An improved random forest based on the classification accuracy and correlation*

*measurement of decision trees*. ScienceDirect. <https://www.sciencedirect.com/science/article/abs/pii/S0957417423020511>