Fetal Health Classification

Introduction

In the past 30 years, the worldwide child mortality rate has decreased by 60%. While this is great progress, still in 2019 over 5 million children (about twice the population of Mississippi) died before reaching the age of 5-years old and half of these deaths were among newborns. To combat child mortality, the United Nations has set the Sustainable Development Goal (SDG) to reduce newborn mortality to 12 deaths per 1,000 live births by the end of 2030[1]**.** For this, low-cost detection systems that can identify fetuses at high risk of death are required.

Fortunately, Cardiotocogram (CTG) can be used to collect vital information about fetal health during pregnancy, such as uterine contractions, fetal baseline heart rate (BHR), acceleration, deceleration, and variability of BHR at a low cost. CTGs as shown in figure 1, graph the fetal heart rate and uterine contraction as well as showing the relationship between the two. A normal BHR is 110-160 bpm and any heart rate outside of this range could suggest an abnormality. Abrupt increases/decreases in BHR are also abnormal and are described as accelerations/decelerations. Uterine contractions are assessed by the duration and intensity of the contraction with 2 in 10 minutes being normal[2].

Objective

Our objective is to be able to predict if fetus health is at high risk solely based on data collected by the CTG and determine which variables are most important in making this prediction.

Exploratory Data Analysis (EDA)

The dataset (source:Kaggle) contains 22 variables associated with 2126 CTG reports (Table 1). To begin the EDA, the outputs for the fetal health variable, "suspect" and "pathological" were combined into one output of "abnormal". This was done in order to transform the problem into a binary classification problem, with the model predicting whether fetal health is "normal" or "suspect." Following that, histograms and density plots of all variables were created to help visualize their distribution and the effect they will have on the analysis (Figure 2). The density plots revealed that the variables *acceleration*, *uterine contractions*, *severe deceleration*, *prolonged deceleration,* and *abnormal short-term variability* could be good indicators for model prediction due to the high variance of values dependent on fetal health. Variables like *prolonged & severe decelerations* and *accelerations* show a clear demarcation for the normal and suspect cases. Next, we explored the correlation amongst predictor variables and observed a very high linear relationship between *histogram mean, median, and mode* (Figure 3 – heatmap of correlation). We used the variance inflation factor (VIF) to address the remaining variables' non-obvious correlations (Figure 4). *Histogram max, histogram width, and histogram min* all returned *infinity* VIF, indicating that these variables are perfectly multicollinear. This is consistent with our previous findings in EDA.

Data Preprocessing

- **Missing values identification**: Fortunately, our dataset was mostly clean and contained no missing values.
- **Standardization**: We standardized predictor variables to bring them within a range of 0 to 1. Finally, dummy variables were created for the target variable, Fetal Health, to facilitate analysis. It was encoded to 0, and 1 (Normal = 0, Suspect = 1) for further analysis.

Solution

1. Evaluation matrix

In addition to the overall accuracy of the classifier, we focus on reducing False Negatives and hence give preference to recall value as well. Since our goal is to flag as many real suspicious test results as possible, we are willing to sacrifice some accuracy for increased recall. Therefore, after fitting the model, we conducted a **cost analysis** *assuming that the cost of 'false negative' is three times that of 'false positive'* and identified a threshold (typically <0.5) that minimizes the overall cost.

2. Modelling

Logistic Regression

We divided the data into training and testing sets, using a 70:30 split ratio. We fit a Logistic Regression model to all variables to ascertain their significance (Figure 5). Not surprisingly, the model is unable to resolve the significance of *histogram max, histogram min*, and *histogram width*, which corresponds to our findings in the VIF test above. Besides, the variables that came out to be statistically significant showed significant variation in the Normal/Suspect class during EDA (Figure 2).

Then we added two interaction variables: **Uterine contract light decel** and **Uterine contract accel** because these variables are important indicators in medical theory. Accelerations occurring alongside uterine contractions indicate the presence of a healthy fetus. Light deceleration associated with uterine contraction is called early deceleration, which is due to increased fetal intracranial pressure causing increased vagal tone. As a result, this deceleration is considered physiological rather than pathological.

Restricting Logistic Regression to the significant variables and the two additional interaction terms (Figure 6), we built a model with a lower BIC value owing to lower complexity. Moreover, the low P values of the interaction terms corroborated our hypothesis of paired effect. This model had an AUC of 0.965 (Figure 7). By conducting the cost analysis mentioned previously, we determined the optimal threshold for the training set should be 0.284 (Figure 8). Model accuracy on test data was 90% (Figure 9), significantly higher than the baseline classifier's accuracy of 77%. However, class 1 recall was 87%, indicating room for growth.

As per our modified Logistic Regression model, the following are the most important predictors of Fetal Health (Figure 10):

- **Abnormal short-term variability, prolonged decelerations**: both contribute to an abnormal CTG result. Abnormal short-term variability, such as less than 5bmp for fetal heartbeat rate raises suspicion for fetal tachycardia, congenital heart abnormalities, and other factors, necessitating additional investigations and possible intervention. Prolonged decelerations do not categorically predict poor fetal health but evoke obstetricians to recommend a caesarean delivery.
- **Histogram Mean**: Based on initial EDA, we observed clear differentiation for both classes in terms of mean value and density and expected it to be an important predictor. However, it turned out to be one of the less important factors in this model.

Decision Tree

We also tried Decision Tree model and followed a similar process of cost analysis. The two interaction variables were omitted because it take care of interactions automatically. We tuned our model on the training dataset using Grid Search and cross-validation to find optimal hyperparameters (Figure 11). The AUC was 0.995 (Figure 12) and cost-cutting threshold was 0.200 (Figure 13). Accuracy in the test set (93%) outperformed both the baseline classifier (77%) and logistic regression (Figure 14). Additionally, the recall rate for the positive class was increased to 91%.

The most important predictors of Fetal Health, as determined by our Decision Trees model, are as follows (Figure 15):

- **Abnormal short-term variability** is consistently the most important feature followed by the mean value of short-term variability, which makes sense given that increased short term variability increases the likelihood of fetal unhealth.
- **Histogram Mean**, which we expected to be a significant feature, is third in row.
- Two out of the top five variables are consistent with Logistic Regression.

Random Forest

Similar approach as decision trees. The AUC was 0.999 (Figure 16) and cost-cutting threshold was 0.345 (Figure 17). Accuracy on the test set (94%) significantly outperformed the baseline classifier (77%) and logistic regression and was slightly higher than decision tree (Figure 18). The **recall rate** has been increased to **92%,** the highest of the three models.

The most important predictors in Random Forest model are (Figure 19):

- The most significant feature remains **abnormal short-term variability**.

- Top five characteristics are mostly consistent with the top features of decision tree model. However, as expected from EDA (Figure 2), the variable *mean value of long-term variability* is pushed out of the top five range.
- Histogram attributes are consistently ranked lower, which is consistent with EDA since they are highly correlated with the histogram mean.

Random Forest was chosen as the final model post comparing the performance metrics for all the models (Figure 20)

Insights

The data analysis results, in conjunction with the prediction models, are meaningful in both the academic and medical fields.

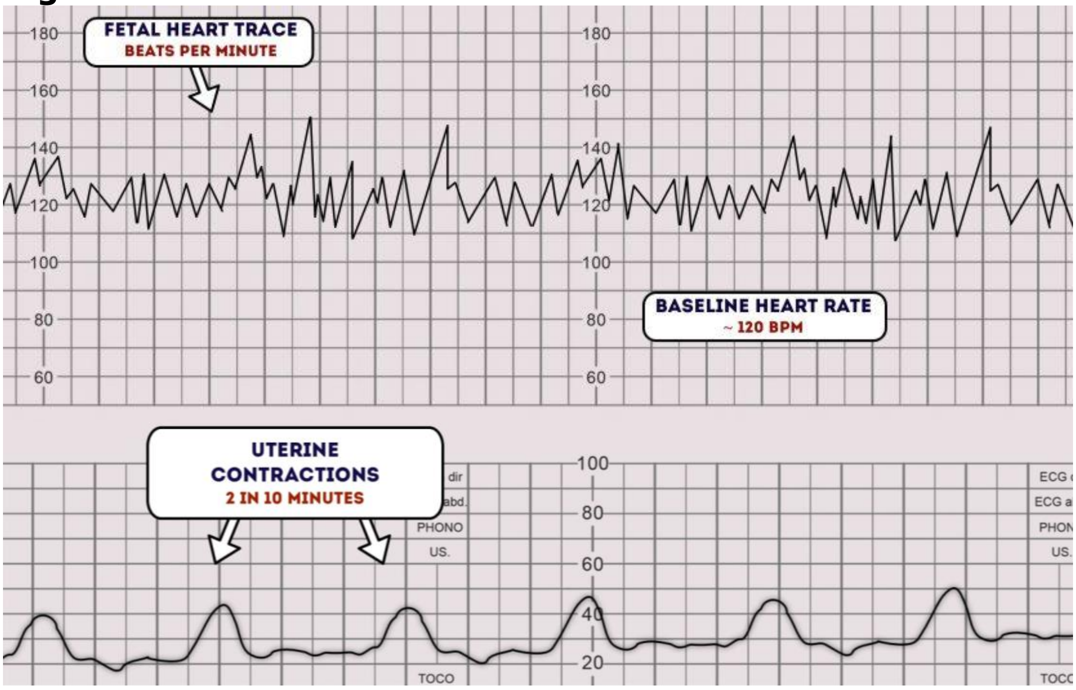
- **Academically**, data mining on medical cases is a relatively new area of research with scant coverage in the literature. Our research examined the association between CTG test characteristics and fetal health. Analogies can be drawn with other types of medical imaging that share similar procedure such as the analysis of Covid-19 infections on a CT image using deep learning models. Additionally, the performance of the various models used in this project can be used as a benchmark for future research, and the models can be enhanced to become more powerful.

- For **medical purposes**, the outcome can be used to develop effective tools for the gynecology industry. To begin, our model can be used by the CTG instrument to perform automatic prediction, screening, and advance warning of abnormal conditions. The machine monitors more patients 24 hours a day, seven days a week at a lower cost than if the physician reads the chart manually, instantly flagging promising cases for further investigation. Additionally, it can help reduce diagnostic delay, which results in missed opportunities for intervention. Second, hospitals can use a continuous flow of new data to train models in order to uncover previously unknown relationships and patterns between variables and infant health. Medical experiments can be used to further validate the findings, which may result in new medical discoveries.

To summarize, we analyzed the CTG data by developing three classification models. The outcome confirmed the hypothesis regarding the most influential factors affecting fetal health and established an order of importance. The project's outcome has resulted in a relatively robust model that significantly improves the prediction of suspect cases in CTG and contributes to the reduction of infant mortality. With the potential to save the lives of infants and pregnant women, we believe this project is enormously significant.

Appendix

Figure 1. CTG: Baseline heart rate and Uterine Contractions



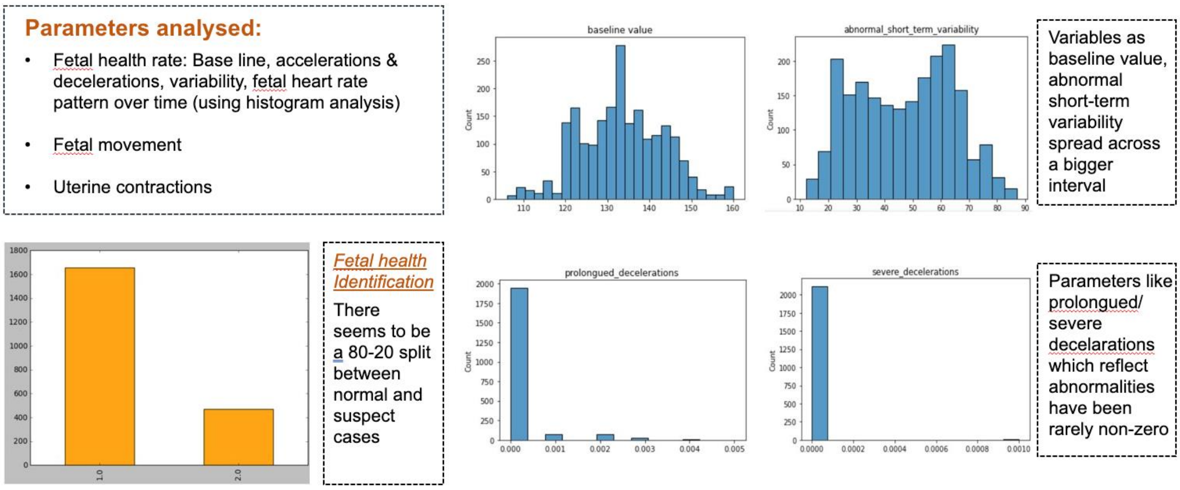
Reference:

[1] Dr Lewis Potter·Data Interpretation·Last updated:May 11, 2021. (2021, May 11). *How to Read a CTG: CTG Interpretation*. Geeky Medics. https://geekymedics.com/how-to-read-a-ctg/.

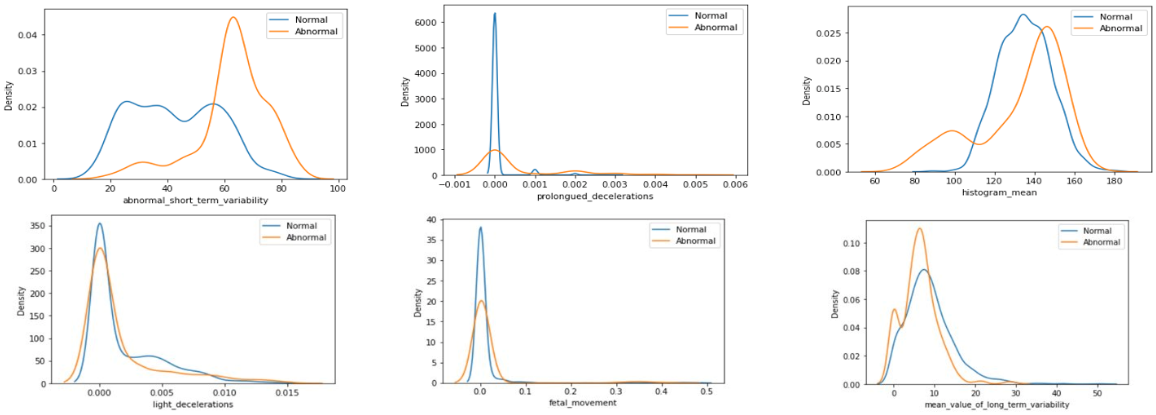
[2] "Child Survival and the SDGs." *UNICEF DATA*, 20 July 2021, data.unicef.org/topic/child-survival/child-survival-sdgs/.

Figure 2. Histograms and Density Plots of Variables

Histograms: univariate analysis



Density Plots: bivariate analysis wrt fetal health identification

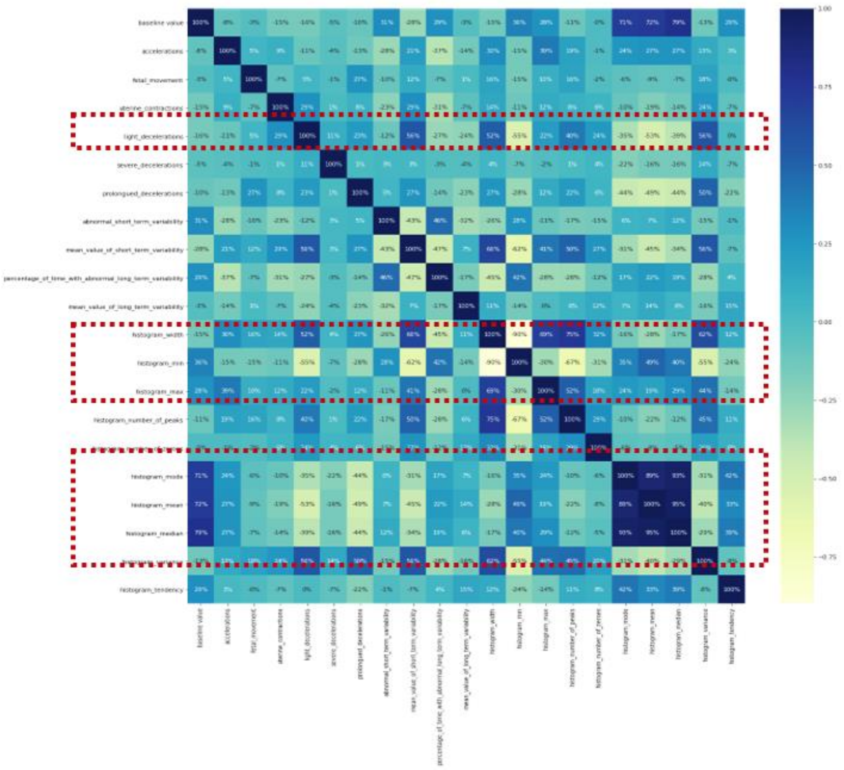


Key Takeaways:

- Variables like abnormal short-term variability, prolonged decelerations, and histogram mean show a clear demarcation for the normal and suspect cases
- Other parameters as light decelerations, fetal movement, mean value of long term variability do not show a clear demarcation for normal and suspect class

Figure 3. Correlation Analysis - Heatmap of Correlation

Histogram mean, median, and mode are strongly correlated



Highly correlated variables:

- Light decelerations with mean short term variability, histogram width and variance
- Histogram width, variance & max values
- Histogram mean, mode and median along with baseline value

Figure 4. VIF Results Table

Histogram width, max, and min showed inf VIF

	columns	VIF
11	histogram_width	inf
13	histogram_max	inf
12	histogram_min	inf
17	histogram_mean	1150.248307
0	baseline value	923.943113
16	histogram_mode	470.576981
7	abnormal_short_term_variability	15.796705
8	mean_value_of_short_term_variability	9.425245
14	histogram_number_of_peaks	6.862846
10	mean_value_of_long_term_variability	6.244875

Key Takeaways from VIF:

- Histogram width, max, min, mean, mode and baseline value seem to have very high VIF
- Other variables associated to abnormal variability (short & long) also have VIF > 4

Figure 5. Logistic Regression Model of All Variables

Significance of *histogram max, min*, and *width* are Nan

```
Optimization terminated successfully.
         Current function value: 0.180238
         Iterations 10
                                        Results: Logit
=====================================================================================================
Model:                          Logit                         Pseudo R-squared:              0.651
Dependent Variable:             fetal_health_enc              AIC:                           578.3897
Date:                           2021-08-06 19:41              BIC:                           689.7987
No. Observations:               1488                          Log-Likelihood:                -268.19
Df Model:                       20                            LL-Null:                       -768.09
Df Residuals:                   1467                          LLR p-value:                   4.3134e-199
Converged:                      1.0000                        Scale:                         1.0000
No. Iterations:                 10.0000
-----------------------------------------------------------------------------------------------------
                                                        Coef.  Std.Err.    z     P>|z|   [0.025   0.975]
-----------------------------------------------------------------------------------------------------
const                                                  -9.3797      nan     nan     nan      nan      nan
baseline value                                         -0.3051   1.8371 -0.1661 0.8681  -3.9057   3.2956
accelerations                                         -14.7370   2.2919 -6.4300 0.0000 -19.2291 -10.2449
fetal_movement                                          2.0825   1.3721  1.5178 0.1291  -0.6067   4.7716
uterine_contractions                                   -3.7931   0.6621 -5.7287 0.0000  -5.0908  -2.4953
light_decelerations                                     1.2855   1.0701  1.2013 0.2296  -0.8119   3.3828
severe_decelerations                                    1.5035   1.4801  1.0158 0.3097  -1.3974   4.4044
prolongued_decelerations                               14.7829   1.9580  7.5499 0.0000  10.9452  18.6206
abnormal_short_term_variability                         7.4968   0.9063  8.2719 0.0000   5.7205   9.2731
mean_value_of_short_term_variability                   -2.1629   1.8631 -1.1609 0.2457  -5.8145   1.4887
percentage_of_time_with_abnormal_long_term_variability  2.7545   0.6354  4.3348 0.0000   1.5091   4.0000
mean_value_of_long_term_variability                     0.8740   2.0127  0.4342 0.6641  -3.0708   4.8187
histogram_width                                        -1.0257      nan     nan     nan      nan      nan
histogram_min                                           1.5033      nan     nan     nan      nan      nan
histogram_max                                           5.4269      nan     nan     nan      nan      nan
histogram_number_of_peaks                               1.4225   1.0629  1.3383 0.1808  -0.6607   3.5058
histogram_number_of_zeroes                             -0.7010   1.6789 -0.4175 0.6763  -3.9916   2.5896
histogram_mode                                         -4.6303   3.2784 -1.4124 0.1578 -11.0558   1.7952
histogram_mean                                         12.8332   3.5156  3.6504 0.0003   5.9428  19.7237
histogram_median                                       -5.2282   4.5168 -1.1575 0.2471 -14.0811   3.6246
histogram_variance                                     10.0290   2.1469  4.6715 0.0000   5.8213  14.2368
histogram_tendency                                      0.3592   0.5574  0.6444 0.5193  -0.7333   1.4516
=====================================================================================================
```

Figure 6. Logistic Regression Model of significant variables and two interaction terms

- BIC value has decreased (less complex model)
- The two interaction terms are significant

```
Optimization terminated successfully.
         Current function value: 0.191581
         Iterations 9
                                              Results: Logit
=============================================================================================================
Model:                            Logit                          Pseudo R-squared:             0.629
Dependent Variable:               fetal_health_enc               AIC:                          592.1463
Date:                             2021-08-06 19:41               BIC:                          650.5034
No. Observations:                 1488                           Log-Likelihood:               -285.07
Df Model:                         10                             LL-Null:                      -768.09
Df Residuals:                     1477                           LLR p-value:                  3.8565e-201
Converged:                        1.0000                         Scale:                        1.0000
No. Iterations:                   9.0000
-------------------------------------------------------------------------------------------------------------
                                                          Coef.  Std.Err.    z     P>|z|   [0.025   0.975]
-------------------------------------------------------------------------------------------------------------
uterine_contractions                                     -4.7572   0.7094 -6.7059 0.0000  -6.1477  -3.3668
accelerations                                           -16.0174   2.6950 -5.9433 0.0000 -21.2996 -10.7352
prolongued_decelerations                                 14.1863   1.5336  9.2503 0.0000  11.1805  17.1921
abnormal_short_term_variability                           6.9646   0.7330  9.5016 0.0000   5.5280   8.4012
percentage_of_time_with_abnormal_long_term_variability    2.6724   0.5584  4.7861 0.0000   1.5780   3.7668
histogram_mean                                            7.6190   1.1106  6.8600 0.0000   5.4422   9.7958
histogram_variance                                        9.9725   1.7617  5.6608 0.0000   6.5197  13.4254
uterine_contract_accel                                   13.6833   5.9694  2.2922 0.0219   1.9834  25.3832
light_decelerations                                      -1.3291   1.1662 -1.1397 0.2544  -3.6148   0.9566
uterine_contract_light_decel                              5.1252   2.4141  2.1231 0.0337   0.3938   9.8567
const                                                    -9.2179   0.9718 -9.4855 0.0000 -11.1226  -7.3133
=============================================================================================================
```

Figure 7. Logistic Regression – ROC Curve (AUC 0.965)

0.9659684163520481

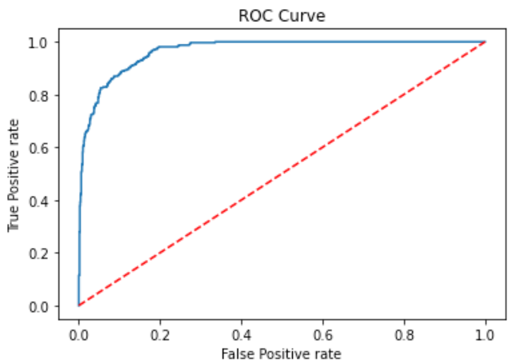
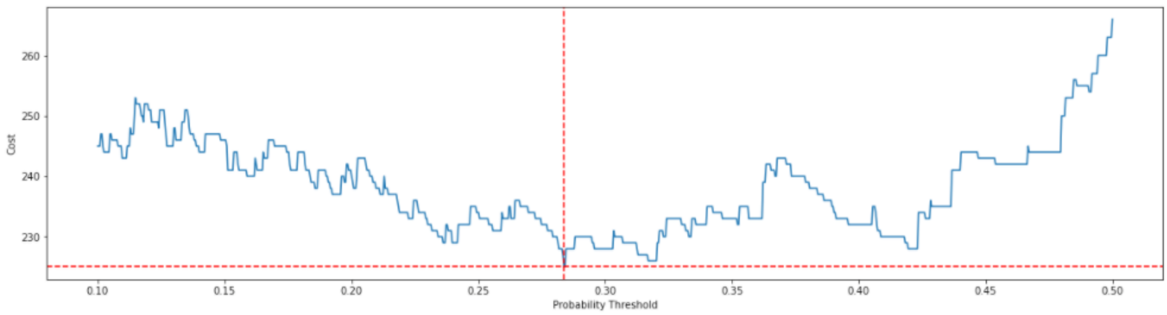


Figure 8. Logistic Regression – Cost Analysis (Threshold 0.284)

Identify the threshold probability cutoff and performance on training set



0.28378378378378377

Figure 9. Logistic Regression - Result (Accuracy 90%)

Accuracy in test data (90%) is much better than baseline classifier (77%)

	precision	recall	f1-score	support
0.0	0.96	0.91	0.94	1173
1.0	0.73	0.87	0.79	315
accuracy			0.90	1488
macro avg	0.85	0.89	0.87	1488
weighted avg	0.91	0.90	0.91	1488

Figure 10. Logistic Regression - Feature Importance

Abnormal short-term variability is the most important feature

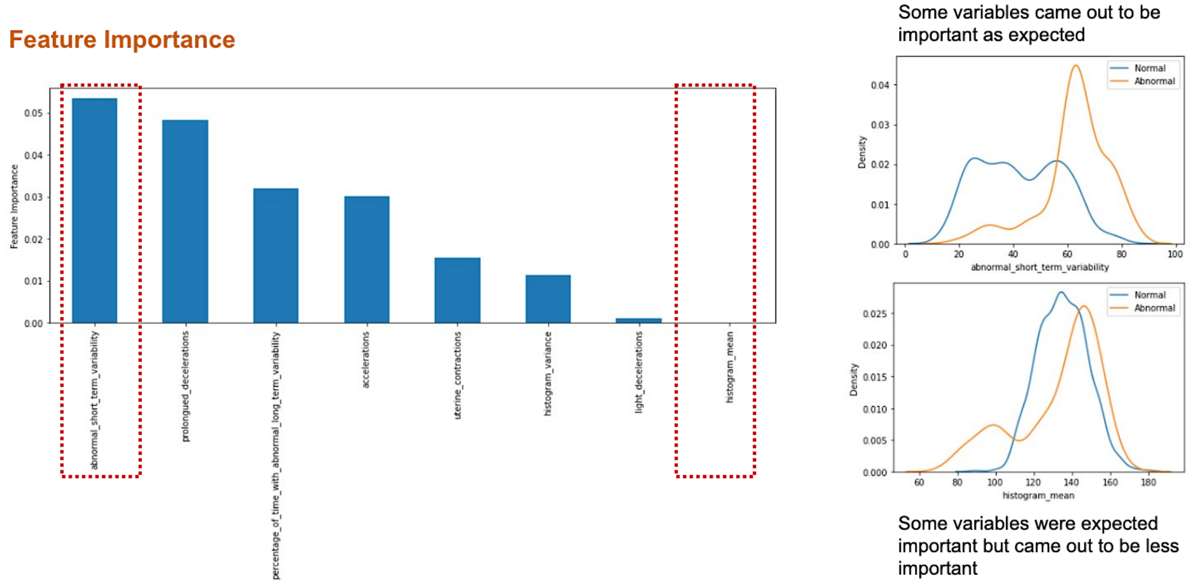


Figure 11. Decision Trees - Model Illustration

Optimal model parameters: max depth=7, min samples split=10

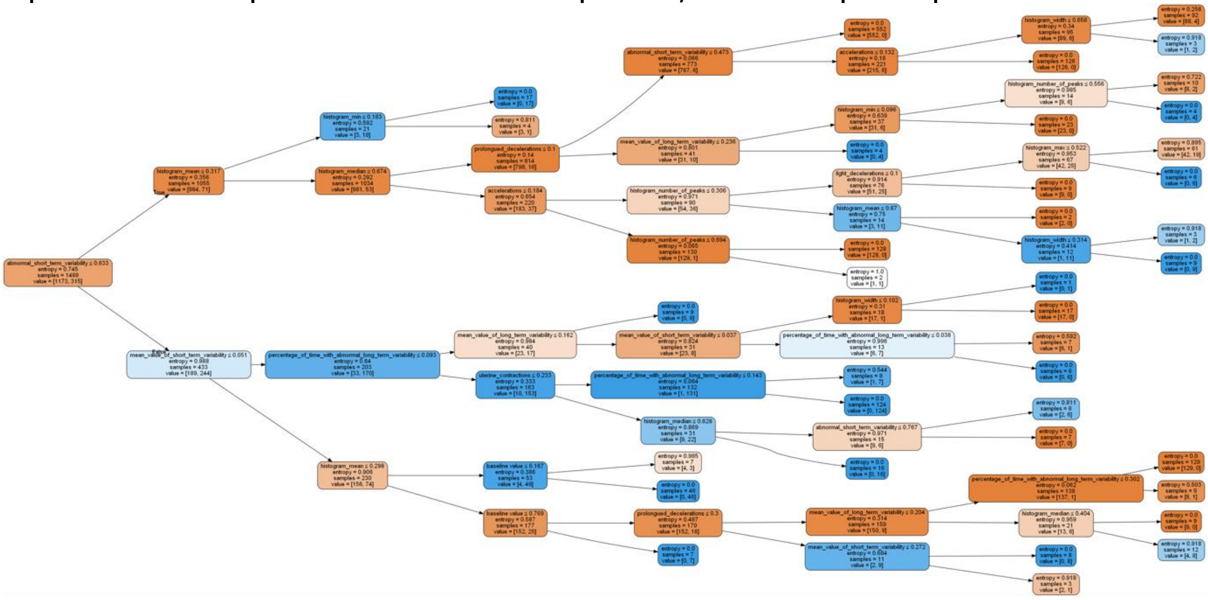


Figure 12. Decision Trees - ROC Curve (AUC 0.995)

0.9954004790321926

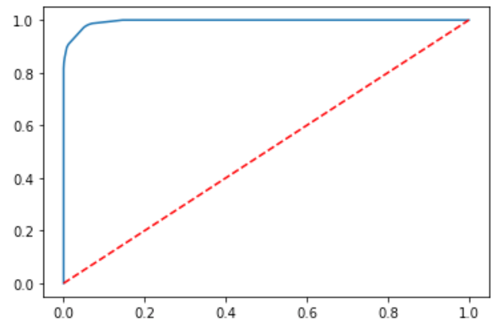
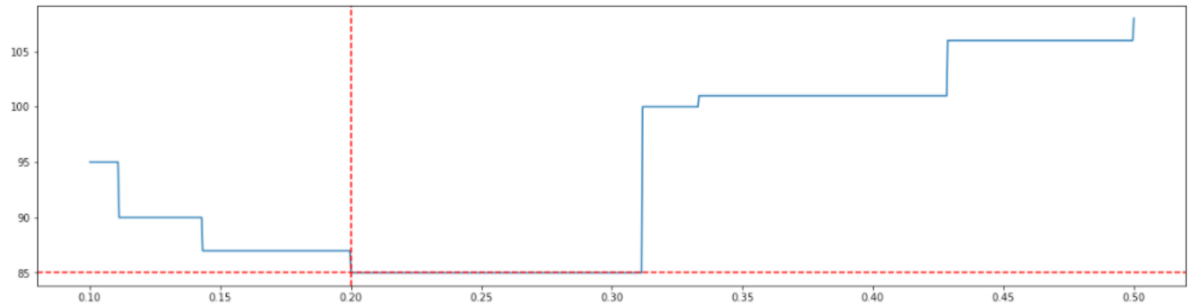


Figure 13. Decision Trees - Cost Analysis (Threshold 0.200)

Identify the threshold probability cutoff and performance on training set



0.2001001001001001

Figure 14. Decision Trees - Result (Accuracy 93%)

Accuracy in test data (93%) is better than logistic regression (90%)

	precision	recall	f1-score	support
0.0	0.97	0.93	0.95	482
1.0	0.81	0.91	0.86	156
accuracy			0.93	638
macro avg	0.89	0.92	0.90	638
weighted avg	0.93	0.93	0.93	638

Figure 15. Decision Trees - Feature Importance

Abnormal short-term variability is the most important feature

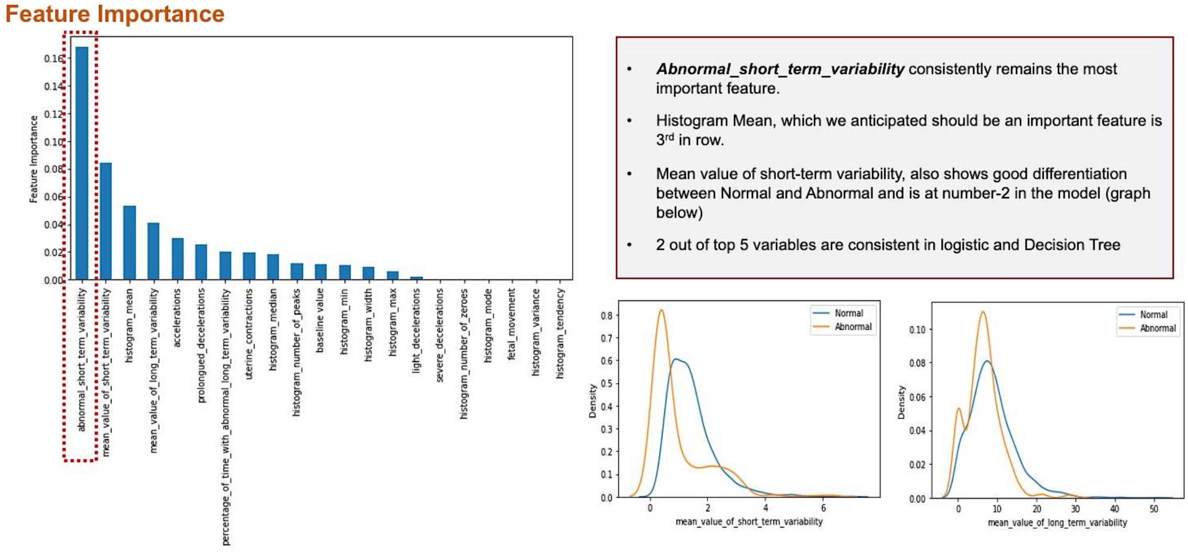


Figure 16. Random Forest - ROC Curve (AUC 0.999)

0.9999986468017159

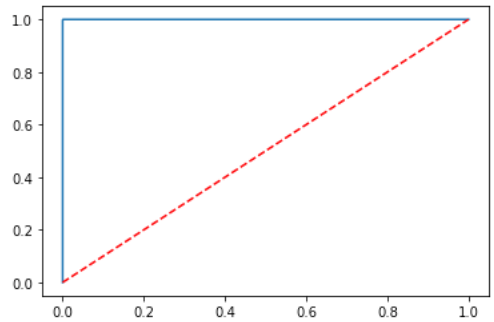
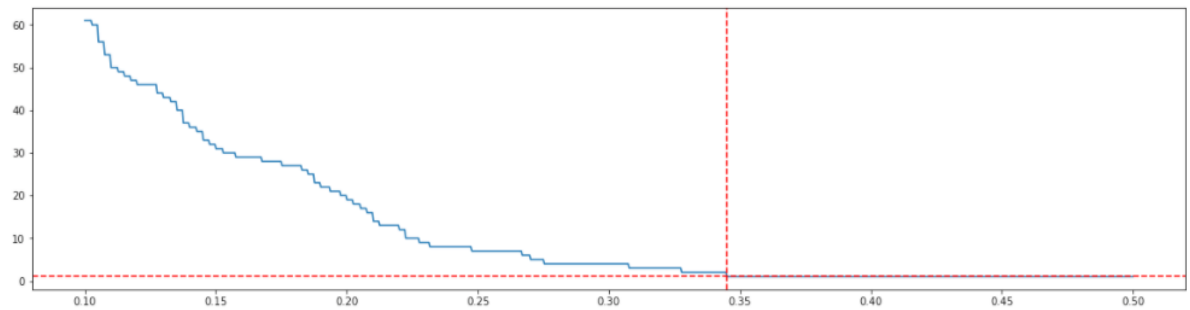


Figure 17. Random Forest - Cost Analysis (Threshold 0.345)

Identify the threshold probability cutoff and performance on training set



0.3450450450450451

Figure 18. Random Forest - Result (Accuracy 94%)

Accuracy in test data (94%) is even better than decision trees (93%)

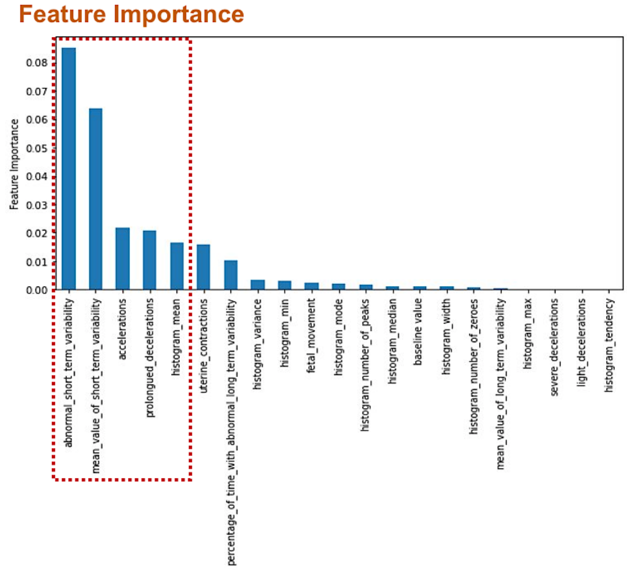
```
              precision    recall  f1-score   support

         0.0       0.97      0.95      0.96       482
         1.0       0.86      0.92      0.89       156

    accuracy                           0.94       638
   macro avg       0.92      0.93      0.93       638
weighted avg       0.95      0.94      0.94       638
```

Figure 19. Random Forest - Feature Importance

Abnormal short-term variability is the most important feature



- ***Abnormal_short_term_variability*** consistently remains the most important feature.
- If we compare the top 5 features, they are roughly consistent with decision trees. However, as expected variable mean_value_of_long_term_variability is thrown out of the top 5 range.
- Other histogram attributes are consistently ranked lower which is consistent with EDA as they were correlated with histogram_mean.

Figure 20. Model Results Comparison

Random Forest was chosen as the final model post comparison

Model Selection

Parameter	Logistic	Decision Tree	Random Forest
Overall Accuracy	89%	93%	94%
AUC	0.965	0.99	0.99
For Class 1 in hold-out dataset i.e., Abnormal Fetus			
Precision	73%	81%	86%
Recall	87%	91%	92%
F1 Score	79%	86%	89%

Important Features

- Abnormal Short-Term Variability
- Mean Value of Short-Term Variability
- Acceleration
- Prolonged Deceleration
- Histogram Mean

Table 1. Fetal Health Data Columns
Each row represents a CTG report extracted from cardiotocograms and classified by expert obstetricians, as described in:

Baseline Value: Fetal baseline heart rate (BHR)
Accelerations: Number of accelerations per second
Fetal Movement: Number of fetal movements per second
Uterine Contractions: Number of uterine contractions per second
Light Decelerations: Number of light decelerations/second
Severe Decelerations: Number of severe decelerations/second
Prolongued Decelerations: Number of prolounged decelerations/second
Abnormal Short Term Variability: % of time with abnormal short-term variability
Mean Value of Short Term Variability: Mean of short term variability
Abnormal Long Term Variability: % of time with abnormal long-term variability
Mean Value of Long Term Variability: Mean of long term variability
Histogram Width: Range of BHR
Histogram Min: Minimum BHR
Histogram Max: Max BHR
Histogram Number of Peaks: Number of peaks in BHR
Histogram Number of Zeros: Number of zeros in BHR
Histogram Mode: Mode of BHR
Histogram Mean: Mean of BHR
Histogram Median: Median of BHR
Histogram Variance: Variance in BHR
Histogram Tendency: Tendency of BHR
Fetal Health: 1 = Normal, 2 = Suspect

Reference:
Ayres de Campos et al. (2000) SisPorto 2.0 A Program for Automated Analysis of Cardiotocograms. J Matern Fetal Med 5:311-318