Final Project

Andrea Hannah, Tara Linney
MMC
School of Applied Computational Sciences
handrea22@email.mmc.edu; tlinney@email.mmc.edu

Abstract**—**Fetal health of a baby in utero has an impact on child and maternal mortality. Child mortality is a concern in countries all over the world. By analyzing and classifying fetal health, countries can end preventable deaths of newborns and babies under the age of 5 years old. This issue of fetal health is something that affects both the fetus and the maternal health of its carrier. Thus, by solving this problem, we could save not just one life, but two. By analyzing the Cardiotocograms (CTGs) data to evaluate the data attributes of baseline value or fetal heart rate (FHR), fetal movements, and uterine contractions, we build machine learning models to be able to accurately distinguish between healthy and unhealthy fetuses in utero. In identifying unhealthy fetuses, providers and parents will be able to figure out how to help the mother and the baby. Using 2126 observations, the supervised machine learning algorithms of the Decision Tree Classifier and the Random Forest Classifier, which works best with multiple predictor variables, were applied to the data. Since one predictor variable, prolonged decelerations, is chosen for evaluation based on correlation diagrams, the Decision Tree Classifier is the better performing model for the dataset in this instance. The addition of new predictor variables pertaining to demographics and lifestyle at the least will help obstetric and gynecological clinicians best predict fetal health for patients in a clinical setting.

Keywords—Fetal Health; Maternal Health; Cardiotocograms; Prolonged Decelerations; Big Data; Obstetrics; Gynecology;

I. INTRODUCTION

The causes of fetal health issues can catch an expecting mother off guard. With child mortality being such a heart wrenching topic for many expectant mothers, one must understand how health metrics can be used to determine the health of a fetus while in utero. While in utero, there are a number of interventions that can take place to aid in the health of a fetus. In addition to the health of a fetus, we must also look at the health of the mother, as fetal health has an impact on maternal health, and vice versa. In order to understand this relationship, a data analyst must examine cardiotocogram (CTGs) in order to predict fetal health while the fetus is still in utero.

In this study, we focus on a dataset of cardiotocogram data that contains 2126 measurements that were extracted from cardiotocograms and classified by obstetricians. Furthermore, we look at the baseline fetal heart rate, fetal movement and accelerations to predict conclusions on fetal health.

II. DATASET AND FEATURES

We used a dataset found on Kaggle, whose author was Ayres de Campos et al. (2000) SisPorto 2.0 A Program for Automated Analysis of Cardiotocograms. There were a total of 2126 measurements in this dataset, which were extracted from cardiotocograms and classified by expert obstetricians. This dataset includes the following features:

- baseline value
- accelerations
- fetal movement
- uterine contractions
- light decelerations
- severe decelerations
- prolonged decelerations
- abnormal short term variability
- mean value of short term variability
- percentage of time with abnormal long term variability

To begin to understand the dataset, we created a Heatmap to explore the correlations between fetal health and the aforementioned variables. The heatmap shown below allowed which variables had the highest correlation with fetal health.

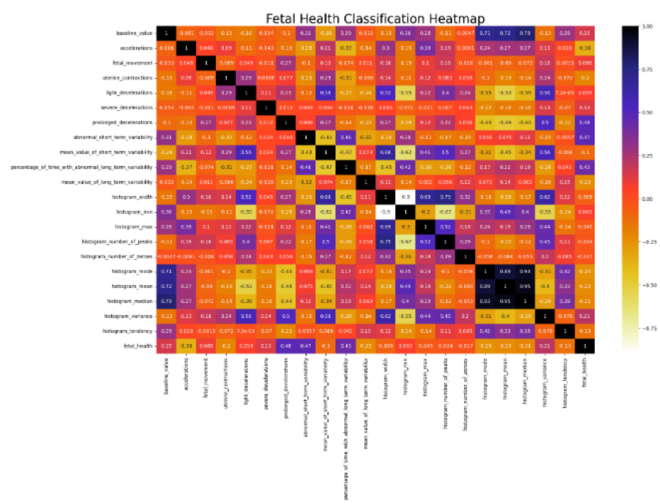

Figure 1

To continue our exploration, we selected the following variables with the respective correlations:

- prolonged decelerations: 0.48
- abnormal short term variability: 0.47

- percentage of time with abnormal long-term variability: 0.42
- baseline value: 0.14
- severe decelerations: 0.13

We then classified the fetal health variable into three categories:

1 = Normal
2 = At Risk
3 = At Risk

Through exploratory data analysis, we were able to see a clear link between prolongued decelerations and fetal health. As the graph below illustrates, as prolonged decelerations increase, the fetal health trends higher up in the "At Risk" category.

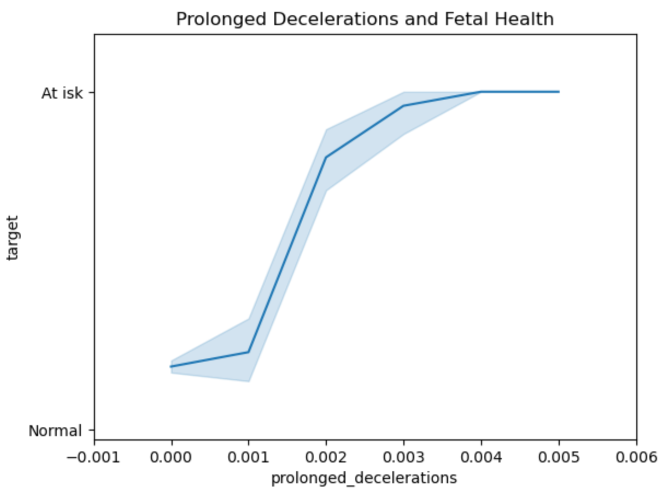

Figure 2

We then examined the abnormal short-term variability and its correlation to fetal health, and saw that there was a somewhat strong correlation between the two. Thus, we chose to include this variable in our final models.

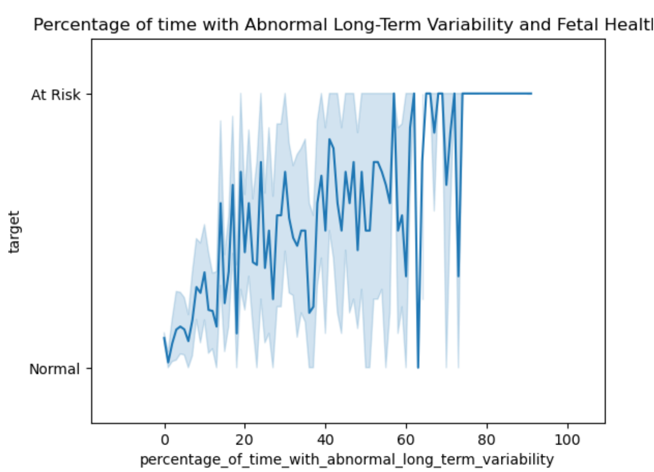

Figure 3

In examining the correlation between percentage of time with abnormal long-term variability and fetal health, we noticed an inconsistent pattern in its correlation. Thus, we decided not to use 'percentage of time with abnormal long-term variability' in our final models.

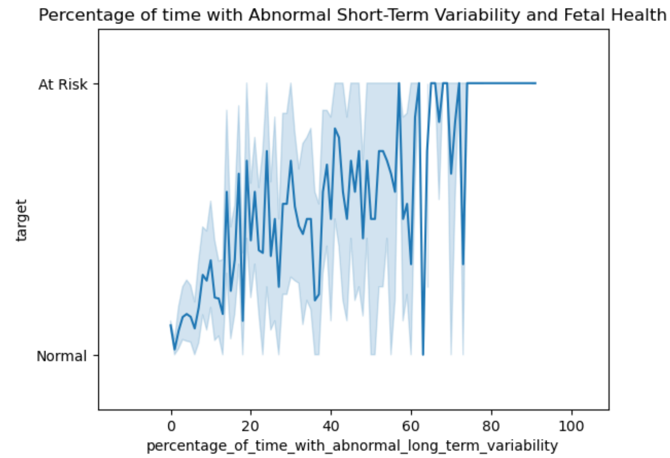

Figure 4

When looking at the baseline value, also known as the fetal heart rate, and its correlation to fetal health, we noticed that there was an inconsistent pattern in its correlation, with several peaks and valleys. Thus, we decided not to use 'baseline value' in our final models.

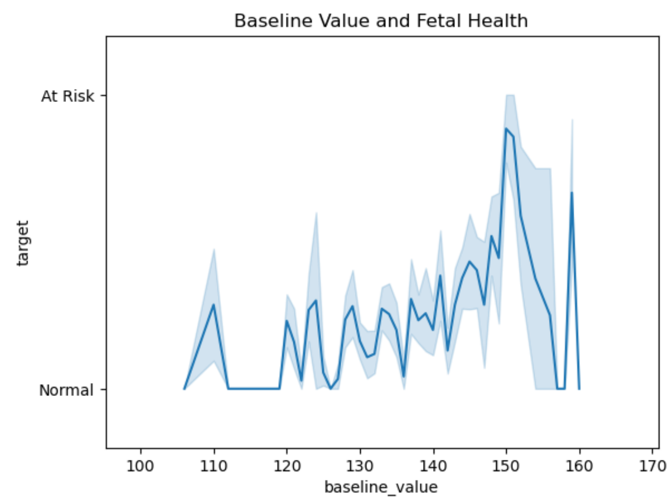

Figure 5

Finally, we examined the correlation between severe decelerations and fetal health. In this examination, we noticed a very clear correlation between severe decelerations and fetal health. Thus, we chose to use this variable in our final model.

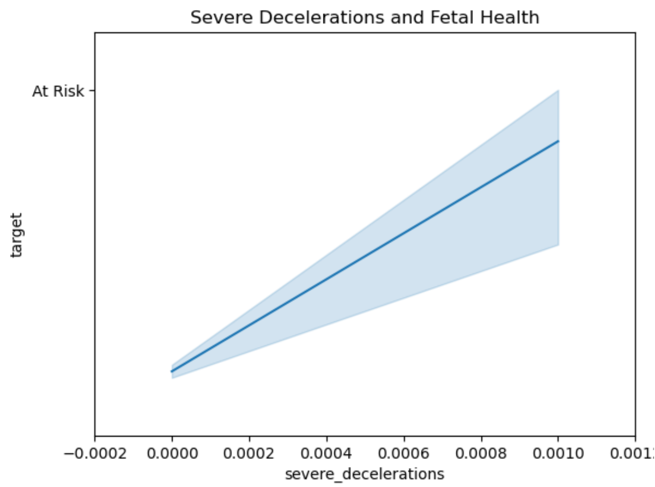

Figure 6

III. METHODS

We used the Decision Trees Classifier and Random Forest Classifier model. Decision Trees Classifier is a supervised machine learning algorithm, which uses labeled data sets to train algorithms that classify data or predict outcomes accurately. Decision trees use a tree-like flowchart model of decisions and the corresponding possible consequences. A decision tree consists of three types of nodes:

1. Root Node - represents the feature all other nodes split from

2. Decision Nodes - represent a test on a feature or attribute for a decision to be made on. As the tree depth increases, the loss entropy should decrease and the information gain should increase until we end up with a pure leaf or end node

3. Leaf/End Nodes - represents the outcome with reduced uncertainty

It is advantageous to use decision trees because they are easily interpretable and understandable after brief explanation. However, a disadvantage is that model calculation can become complex if many values are uncertain. Since our data has no missing values, there is little to no uncertainty to impact decision trees as a selected model. Accuracy of the decision tree model is increased when the depth increases.

Random Forest Classifier is a supervised machine learning technique consisting of many decision trees. Random Forest uses ensemble learning to combine many weak classifiers to provide solutions for complex problems. Random Forest is a bagging method that uses a subset of the original dataset to make predictions, which is an advantage to help to limit overfitting, and creates multiple decision trees with a different set of observations. This classifier involves bootstrapping, which is row and feature sampling with a replacement before training the model. A disadvantage in using this model is when using one feature, Low Bias and High Variance increases with the depth of the decision trees.

Machine Learning Models Comparison:

Decision Trees Classifier

- A single decision tree has faster computation
- Uses rules to predict from input that is a dataset with features
- May experience overfitting if maximum depth is reached.

Random Forest Classifier

- Computation is slower
- Randomly selects observations, builds decision trees and averages
- May experience overfitting if not enough features selected; otherwise, the bagging method that yields output based on majority vote/ranking fixes overfitting

IV. EXPERIMENTS & RESULTS

Performance for each model showed good results (weighted accuracy score):

0.87 for Random Forest | 0.86 for Decision Trees

Random Forest Classifier	Evaluation Parameter		
	Precision	Recall	F1-score
0	0.70	0.67	0.68
1	0.91	0.92	0.92
Weighted Avg	0.87	0.87	0.87
		Accuracy:	0.87

Figure 7

Note that the Fetal Health Classification is binary, where Normal Fetal Health = 0, and At Risk Fetal Health = 1. The Random Forest model is better at predicting At Risk fetal health than at risk fetal health. As seen in figure 7, the Random Forest model predicted Fetal Health with 87% accuracy and the same for precision, recall, and f-1 score for weighted average. However, the Decision Trees model predicted Fetal Health with 86% accuracy and recall, but 85% for precision and F-1 score for weighted average. Therefore, the decision trees model underwent hyperparameter tuning by applying a grid search cross validation in an attempt to improve the model.

The optimal hyperparameters for a decision tree using this dataset included setting criterion for measuring information gain with the machine learning algorithm to 'gini', as opposed to changing this default to entropy. The gini index measures if a node will not be classified correctly if chosen at random, whereas entropy is the measure of impurity or uncertainty of a node. Other optimal hyperparameter changes included setting the maximum depth level of the tree to 5. The splitting nodes default remained as 'best'.

After hyperparameter tuning, Decision Trees modeling performance did not increase for any metrics. Overall, the Random Forest classification model was the best performing model in detecting at-risk fetal health and was 1% more accurate than the Decision Trees as seen in Figure 8. The Decision Tree model is better at predicting normal fetal health than at-risk fetal health.

Decision Tree Classifier & Tuned DT	Evaluation Parameter		
	Precision	Recall	F1-score
0	0.89	0.93	0.91
1	0.70	0.59	0.64
Weighted Avg	0.85	0.86	0.85
		Accuracy:	0.86

Figure 8

A confusion matrix was also used on all models, Figures 9, 10. On the both runs, for Decision Trees, pre and post tuning, there were 313 true positives outcomes, and 23 false positive outcomes for predictions. There were 37 false negatives and 53 true negatives

Decision Tree Classifier & Tuned DT	Actual Values	
	Actual Positive	Actual Negative
Predicted Positive	313 (TP)	23 (FP)
Predicted Negative	37 (FN)	53 (TN)

Figure 9

For Random Forest, there were 310 true positive outcomes, and 26 false positive outcomes for positive predictions. There were 30 false negatives and 60 true negatives. The Random Forest model performed better than Decision Trees in this metric.

Random Forest Classifier	Actual Values	
	Actual Positive	Actual Negative
Predicted Positive	310 (TP)	26 (FP)
Predicted Negative	30 (FN)	60 (TN)

Figure 10

V. CONCLUSION AND FUTURE WORK

We find and agree that the application of machine learning models like Decision Trees and Random Forest for the detection of at-risk fetal health is an effective method that with additional data collection and demographic features can easily be used in clinical settings. Model performance can improve from moderate to excellent with additional patient data. Aside from this, we are aware that considerable advancements have been made in medical care, however the rate of maternal mortality and morbidity and preterm birth have been rising in the United states over the past few years. This prompts new research and tools to alleviate this issue. We know that maternal and infant mortality rates in the US are much higher than their counterparts of similarly populated and wealthy countries. And of course the most at risk populations are those of people of color where they experience an increased risk for poor maternal and infant health outcomes in comparison to their American peers of European descent.

It is imperative to close the gap on this disparity and collect data that includes demographic attributes. For we all know that race and ethnicity has correlating trends with family income, which correlates with lifestyle and health access and outcomes.

The below missing features could improve research and help with eliminating disparities in obstetrics and gynecological clinical practice.

- Demographics
- Stage of pregnancy during which measurements were taken
- Number of fetuses in utero

- Gender of fetus
- Lifestyle of mother
- Results of non-stress test
- Previous pregnancies or stillbirths of mother

In addition to the above, increasing the number of observances, i.e. patients, in our dataset will help improve model performance in predicting at-risk fetal health, which only had a moderate performance as opposed to good performances by all models for normal fetal health predictions. More observances trained and tested for our models should drive the number or ratio of false positives and false negatives down, as well as improve precision and recall when we have more at-risk fetal health data collected and observed. Lastly, we can tune the random forest classifier to improve its performance, which we opted out of for this research since the random forest model performed better on its first model run and was not outperformed altogether after the decision tree model was tuned for optimal hyperparameters.

REFERENCES

[1] Ayres-de-Campos, D., Spong, C.Y., Chandraharan, E. and (2015), FIGO consensus guidelines on intrapartum fetal monitoring: Cardiotocography. International Journal of Gynecology & Obstetrics, 131: 13-24. https://doi.org/10.1016/j.ijgo.2015.06.020

[2] Baccianella, S., A. Esuli, & F. Sebastiani, "SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining", Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC), Vol. 10, 2010.

[3]