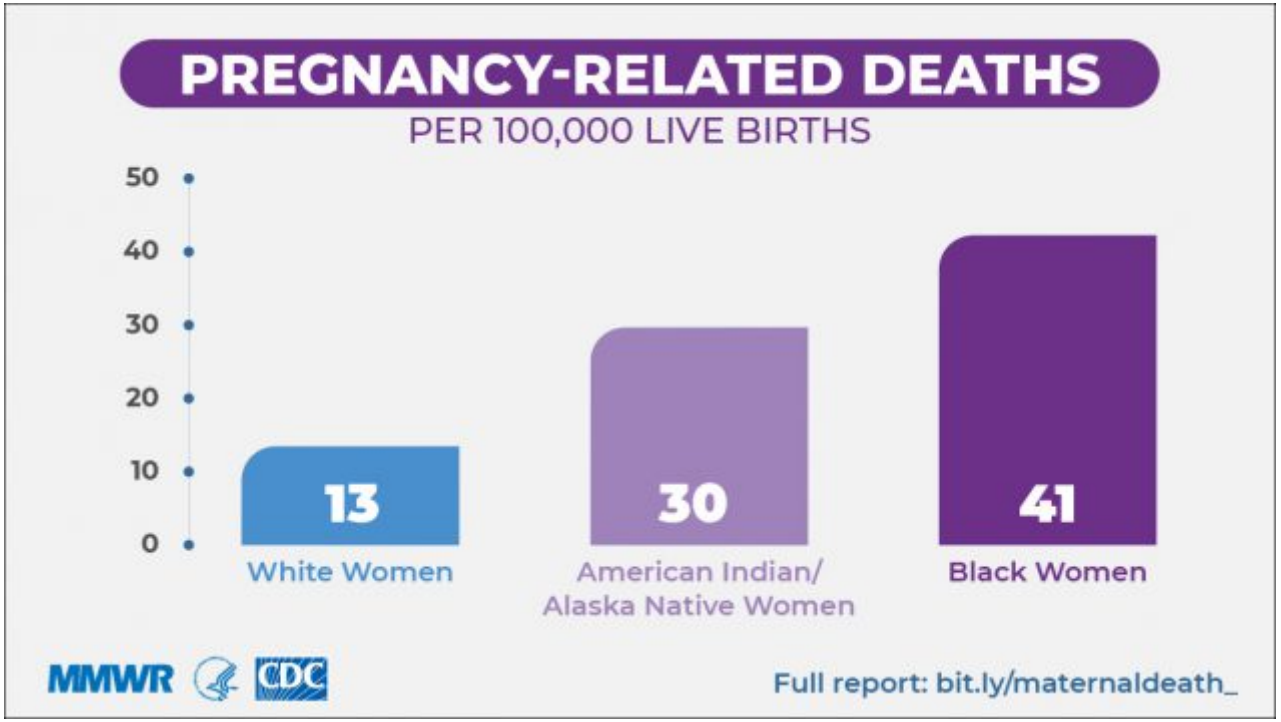
Problem : Maternal Morbidity on the Rise According to the University of Minnesota, "Maternal deaths in the first months of the COVID-19 pandemic increased 33%—and even higher in Black and Hispanic women..." CITATION: https://www.cidrap.umn.edu/news-perspective/2022/06/maternal-deaths-climbed-33-during-covid-19.

Maternal Morbidity is already high and lacking in equity.



Given the most recent overturning of Roe vs. Wade in the United States, it's crucial that medical research is performed with the goal of decreasing maternal morbidity. It's especially important that health care solutions for birthing parents be accessible and low cost.

Business Case : Maternal Morbidity Research NYU Langone in NYC is doing research on Machine Learning and it's applications in lethal fetal diagnoses. Early detection in these cases provide the best possible medical outcomes for birthing parents and provide the most choices on how the birthing parents would like to handle the diagnosis.

Proposed Solution : An Early Detection AI for Pathological Diagnoses Priorotizing recall as the risk of a potentially lethal pregnancy going undetected would result in 1 and/or two deaths, and the risk of a false positive would result in further medical diagnostics.

NOTE: Pathological diagnosis in this data does not necessarily mean a lethal prenatal diagnosis, but that is the worst possible case that this data can represent.

My stakeholder: consider the following three stakeholders--

Stacey has a pre-existing medical condition putter her at high risk during her pregnancy. Where she lives, there are restrictions set in place, preventing her from getting an abortion after the first trimester.

Alex has religious beleifs and knows she will want to carry her pregnancy to term and meet her child face to face if at all possible, even if that child has a fetal prenatal diagnosis.

NYU Langone is looking to provide the best possible outcomes to their patients, no matter how they decide to proceed with their respective diagnoses. Clinicians time needs to be prioritized wisely. Early interventions always provide less invasive and less risky outcomes. NYU Langone is looking to save resources and better allocate those resources by implementing early detection tools for folks like Stacey and Alex.

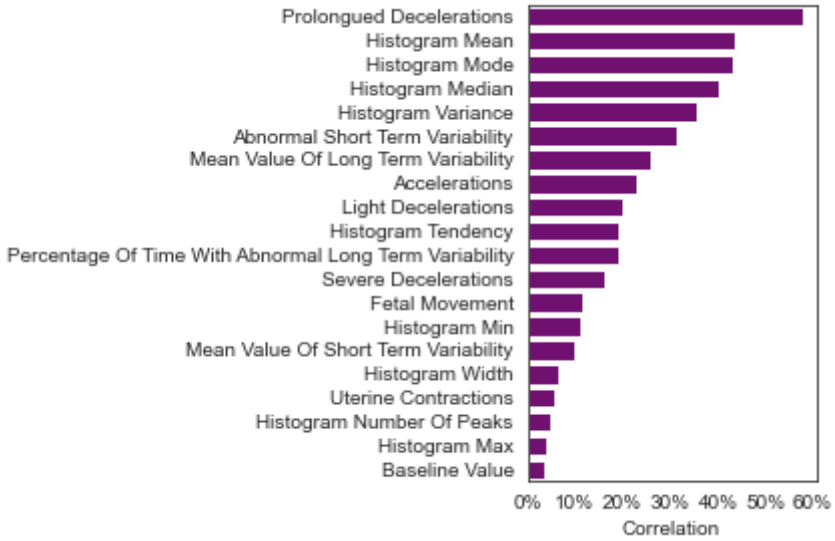
Data This dataset contains 2126 records of features extracted from Cardiotocogram exams, which were then classified by three expert obstetricians into 3 classes:

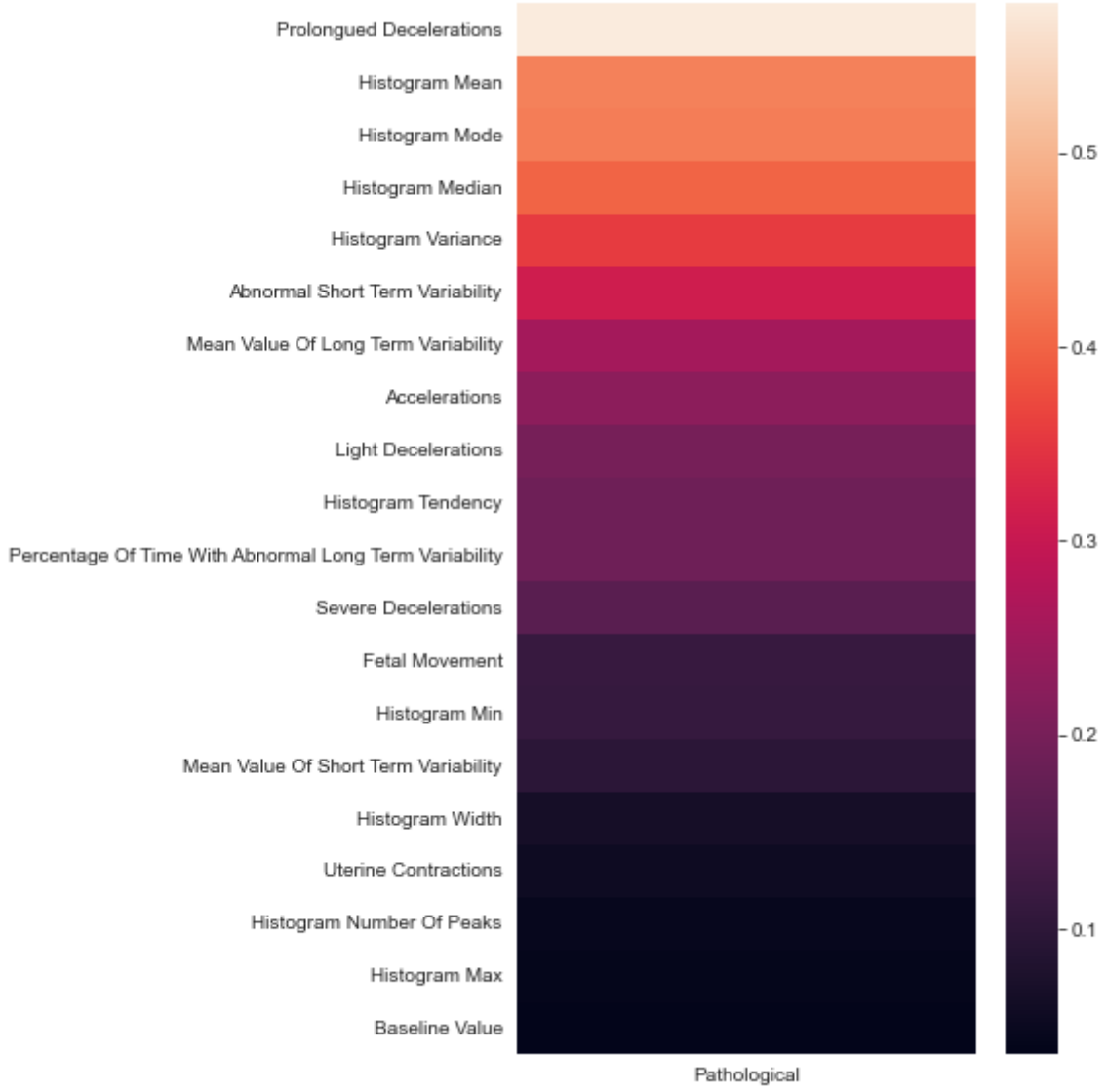
Normal

Suspect

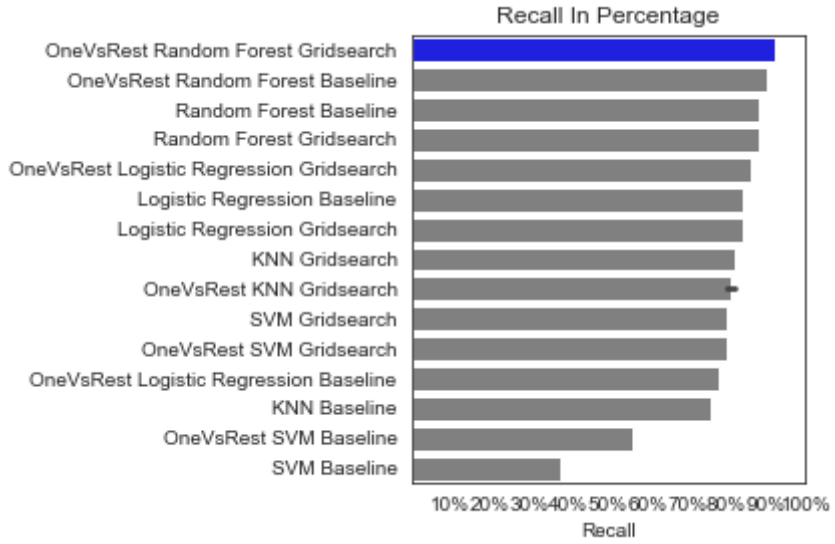
Pathological

Exploring the correlations between the Target and the Pathological Class:


Prolongued Decelerations
Histogram Mean
Histogram Mode
Histogram Median
Histogram Variance
Abnormal Short Term Variability
Mean Value Of Long Term Variability
Accelerations
Light Decelerations
Histogram Tendency
Percentage Of Time With Abnormal Long Term Variability
Severe Decelerations
Fetal Movement
Histogram Min
Mean Value Of Short Term Variability
Histogram Width
Uterine Contractions
Histogram Number Of Peaks
Histogram Max
Baseline Value
0% 10% 20% 30% 40% 50% 60%
Correlation


Prolongued Decelerations
Histogram Mean
Histogram Mode
Histogram Median
Histogram Variance
Abnormal Short Term Variability
Mean Value Of Long Term Variability
Accelerations
Light Decelerations
Histogram Tendency
Percentage Of Time With Abnormal Long Term Variability
Severe Decelerations
Fetal Movement
Histogram Min
Mean Value Of Short Term Variability
Histogram Width
Uterine Contractions
Histogram Number Of Peaks
Histogram Max
Baseline Value
0.5
0.4
0.3
0.2
0.1
Pathological

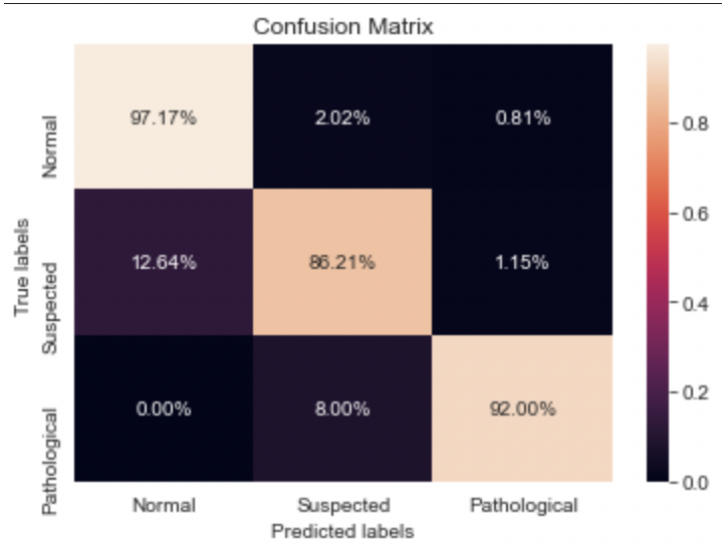
METHODS:



I tested iterations of Logistic Regression, KNN, Support Vector Machines, and Random Forest, with and without gridsearch crossvalidations, both as is and inside a One Vs All wrapper. I made recall of the Pathological class my target metric.

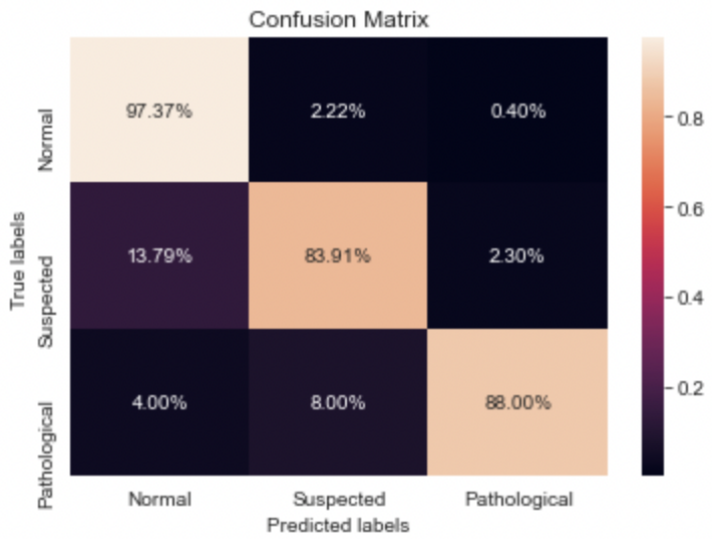
RESULTS:

I have two final models to propose:



The above model had a 90% precision of the Pathological class with a 92% Recall in the Pathological class, using all features.

By using SHAP to reduce dimensionality I was able to produce another model for roll out.



The above model had a 91% precision of the Pathological class with a 88% Recall in the Pathological class, using a subset of features.

I'm proposing to do A/B testing on the two models to see which has the most utility.

**FUTURE WORK: **

1. Retune Models
2. Reorganize Data Collection Methods
3. Iterative test Dimensionailty Reduction
4. Secondary test for Suspected Diagnoses

Recommendations:

Birthing centers, hospitals and women's health centers should be on the look out for some early indicators:

1. A low histogram median.
2. A low short term variability.

3. A high prolonged decelerations.

These signs all point to fast tracking that patient for further diagnostic testing.

These are the simplest-to-interpret, most expressive features in my research targeting Pathological Diagnoses.

For data collection:

1. We can reduce the amount of data we are collecting for the purposes of this model, as proven by the model with reduced dimensions.

2. We can be more specific in our data collection towards lethal prenatel diagnoses specifically.

SOURCES I'll be using Cardiotocogram information from CITATION: https://www.kaggle.com/datasets/andrewmvd/fetal-health-classification.

Further Reading: https://onlinelibrary.wiley.com/doi/10.1002/1520-6661(200009/10)9:5%3C311::AID-MFM12%3E3.0.CO;2-9

This data was published in 2000, by Marques de SÃ¡, J.P., Biomedical Engineering Institute, Porto, Portugal. Bernardes, J., Faculty of Medicine, University of Porto, Portugal. Ayres de Campos, D., Faculty of Medicine, University of Porto, Portugal. This is an appropriate data source for this problem/solution because it has a mix of Cardiotocogram information that range from normal to suspected diagnosis to pathological diagnosis.

CITATION: https://www.kaggle.com/datasets/andrewmvd/fetal-health-classification

CITATION : Dua, D. and Graff, C. (2019). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

This is a paper by the original collectors of the data about their Machine Learning results.

REPO NAV:

--.ipynb_checkpoints

--images

--plots

--.DS_Store

--Abnormal Short Term Variability.png

--Accelerations.png

--Baseline Value.png

--Fetal Movement.png

--Final_model_test_grid.png

--Histogram Max.png

--Histogram Mean.png

--Histogram Median.png

--Histogram Min.png

--Histogram Mode.png

--Histogram Number Of Peaks.png

--Histogram Number Of Zeroes.png

--Histogram Tendency.png

--Histogram Variance.png

--Histogram Width.png

--Light Decelerations.png

--Machines Ranked by Recall in Pathological Class.png

--Mean Value Of Long Term Variability.png

--Mean Value Of Short Term Variability.png

--OVR_RF_test_gridsearch_cm.png

--Percentage Of Time With Abnormal Long Term Variability.png

--Prolongued Decelerations.png

--SHAP Importances.png

--Severe Decelerations.png

--Target Correlation Barplot.png

--Target Correlation.png

--Uterine Contractions.png

--cdc_mm_by_race.jpeg

-- PDFs

--.DS_Store

-- Predicting Pathological Prenatal Diagnoses.pdf

--final_notebook - Jupyter Notebook.pdf

--github.pdf

--.DS_Store

--State-restrictions-on-abortion-map-June-2022-thumbnail.webp

--dan-meyers-f1WMJR8pLqo-unsplash.jpg

--sharon-mccutcheon--iaHr12PVQg-unsplash.jpg

--stephen-andrews-kHAqo7qXoJw-unsplash.jpg

--.DS_Store

--Predicting Pathological Prenatal Diagnoses.pdf

--.gitattributes

--README.md

--fetal_health.csv

--final_notebook.ipynb