

casanave / **fetal_health** Public

phase 3 project, first commit

 0 stars  0 forks

 Star

 Unwatch

 Code  Issues  Pull requests  Actions  Projects  Wiki  Security  Insights  Settings

 main 

 casanave Merge branch 'main' of https://github.com/casanave/fetal_health ... 3 minutes ago  21

[View code](#)

README.md 

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

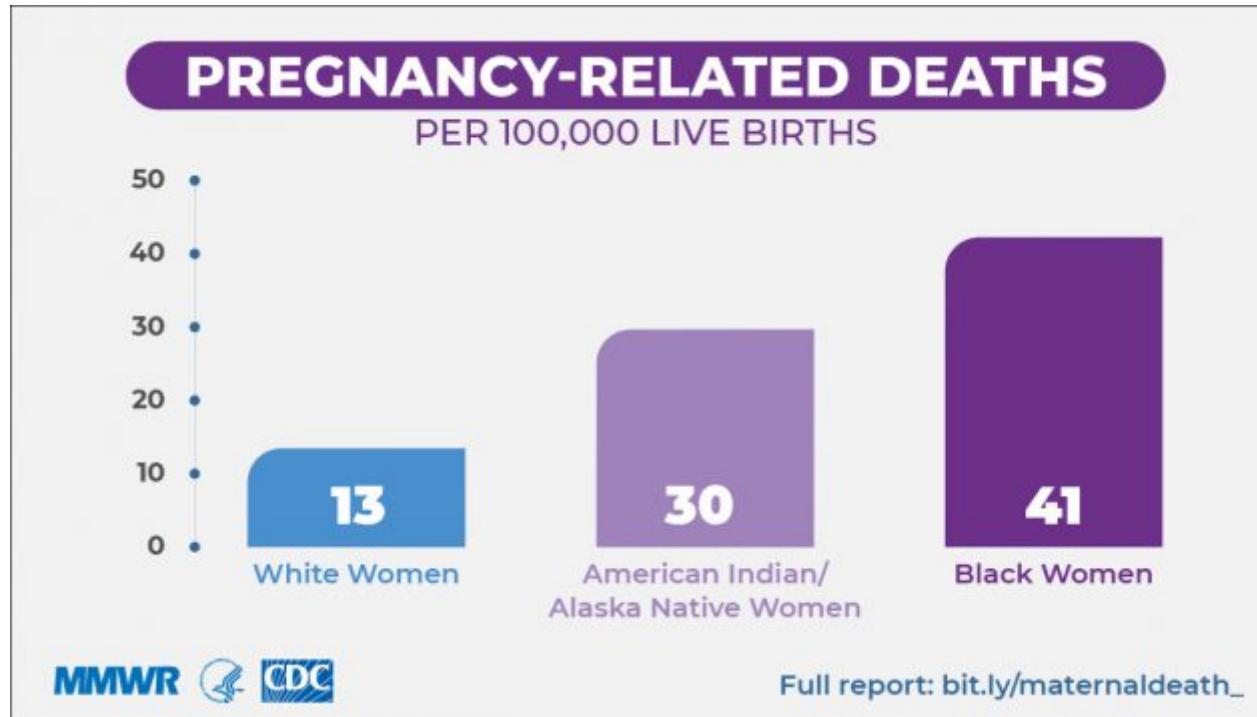
--.DS_Store
--State-restrictions-on-abortion-map-June-2022-thumbnail.webp
--dan-meyers-f1WMJR8pLqo-unsplash.jpg
--sharon-mccutcheon--iaHr12PVQg-unsplash.jpg
--stephen-andrews-kHAqo7qXoJw-unsplash.jpg

--.DS_Store
--Predicting Pathological Prenatal Diagnoses.pdf
--.gitattributes
--README.md
--fetal_health.csv

--final_notebook.ipynb

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.

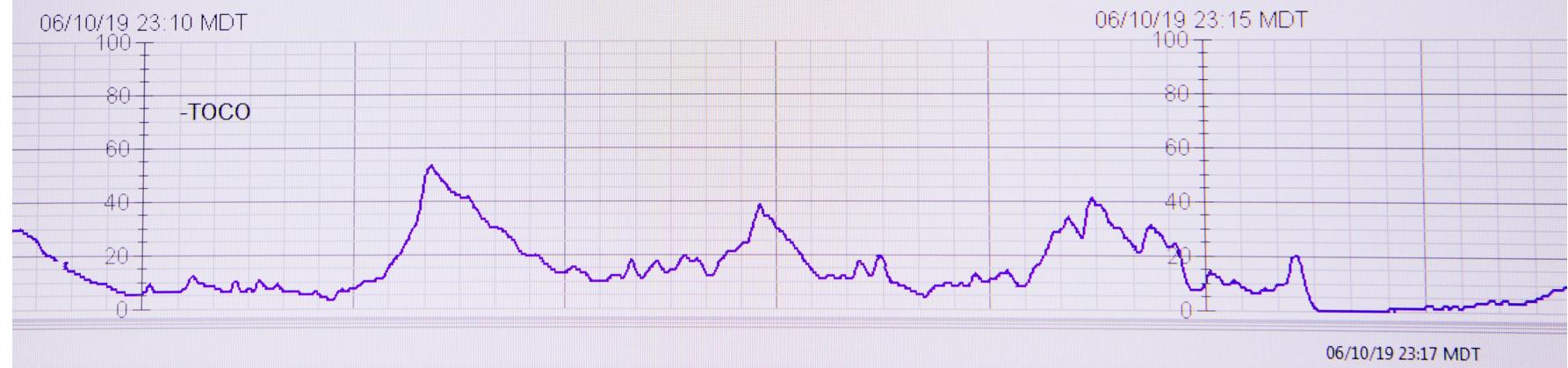
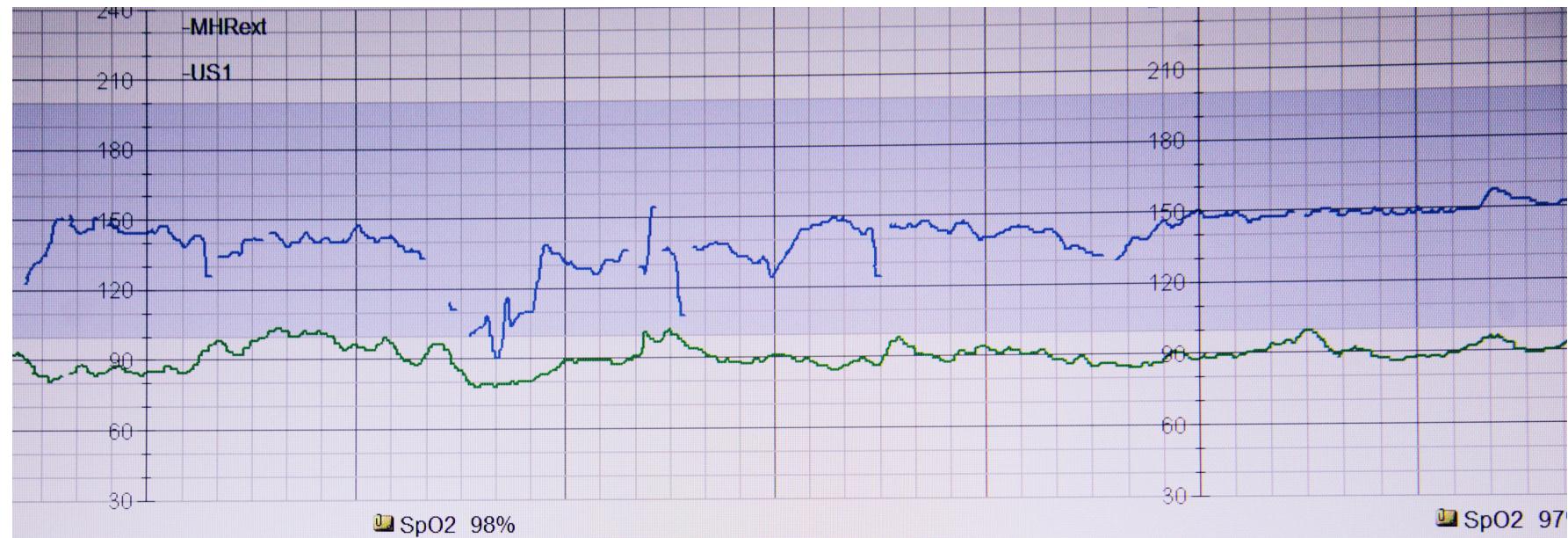


Duke University recently published a paper estimating that a total ban on abortion would increase Maternal morbidity rates dramatically.

"...I find that in the first year of such a ban, estimated pregnancy-related deaths would increase from 675 to 724 (49 additional deaths, representing a 7% increase), and in subsequent years to 815 (140 additional deaths, for a 21% increase). Non-Hispanic Black people would experience the greatest increase in deaths (a 33% increase in subsequent years). Estimated pregnancy-related deaths would increase for all races and ethnicities examined."

CITATION <https://read.dukeupress.edu/demography/article/58/6/2019/265968/The-Pregnancy-Related-Mortality-Impact-of-a-Total>

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 I've been hired by NYU Langone to do design a model. I've decided that this model will focus on having a high recall as it's primary metric. This is because 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. This model is designed to flag pathological diagnoses as early as possible, for a doctor to confirm the diagnosis. This model is NOT designed to replace the diagnostic capabilities of a doctor, but to save time and resources for birthing parents and hospitals.

I'll be working with Cardiotocogram information specifically, and provide NYU Langone with my final result as proof of concept for further research.

Data Understanding I'll be using Cardiotocogram information from CITATION:

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

The original data is here: <https://archive.ics.uci.edu/ml/datasets/cardiotocography>

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.

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. Therefore I'm going to be focusing on the "Pathological" class within this data, especially, and bear in mind that the more generalized and broad use cases for this model will be early detection of all pathological diagnoses.

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. She wants to know as soon as possible if there is reason to suspect she has a fatal prenatal diagnosis, so that she can avoid carrying to term. Early diagnosis of fetal prenatal conditions will give her the time she needs, and prevent mental and physical trauma.

Alex has religious beliefs 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. She wants to know as soon as possible if there is reason to suspect she has a fatal prenatal diagnosis, so that she can make funeral arrangements. Early diagnosis of fetal prenatal conditions will give her the time she needs and prevent mental and physical trauma.

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.

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.

From Kaggle on the Size of the Data and Classes within:

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

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

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

1.0.5 Attribute Information: LB - FHR baseline (beats per minute)

AC - # of accelerations per second

FM - # of fetal movements per second

UC - # of uterine contractions per second

DL - # of light decelerations per second

DS - # of severe decelerations per second

DP - # of prolonged decelerations per second

ASTV - percentage of time with abnormal short term variability

MSTV - mean value of short term variability

ALTV - percentage of time with abnormal long term variability

MLTV - mean value of long term variability

Width - width of FHR histogram

Min - minimum of FHR histogram

Max - Maximum of FHR histogram

Nmax - # of histogram peaks

Nzeros - # of histogram zeros

Mode - histogram mode

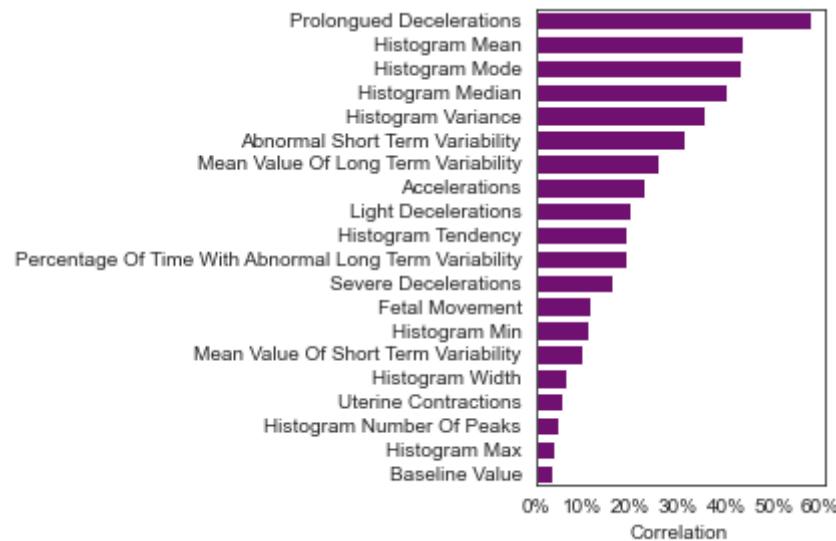
Mean - histogram mean

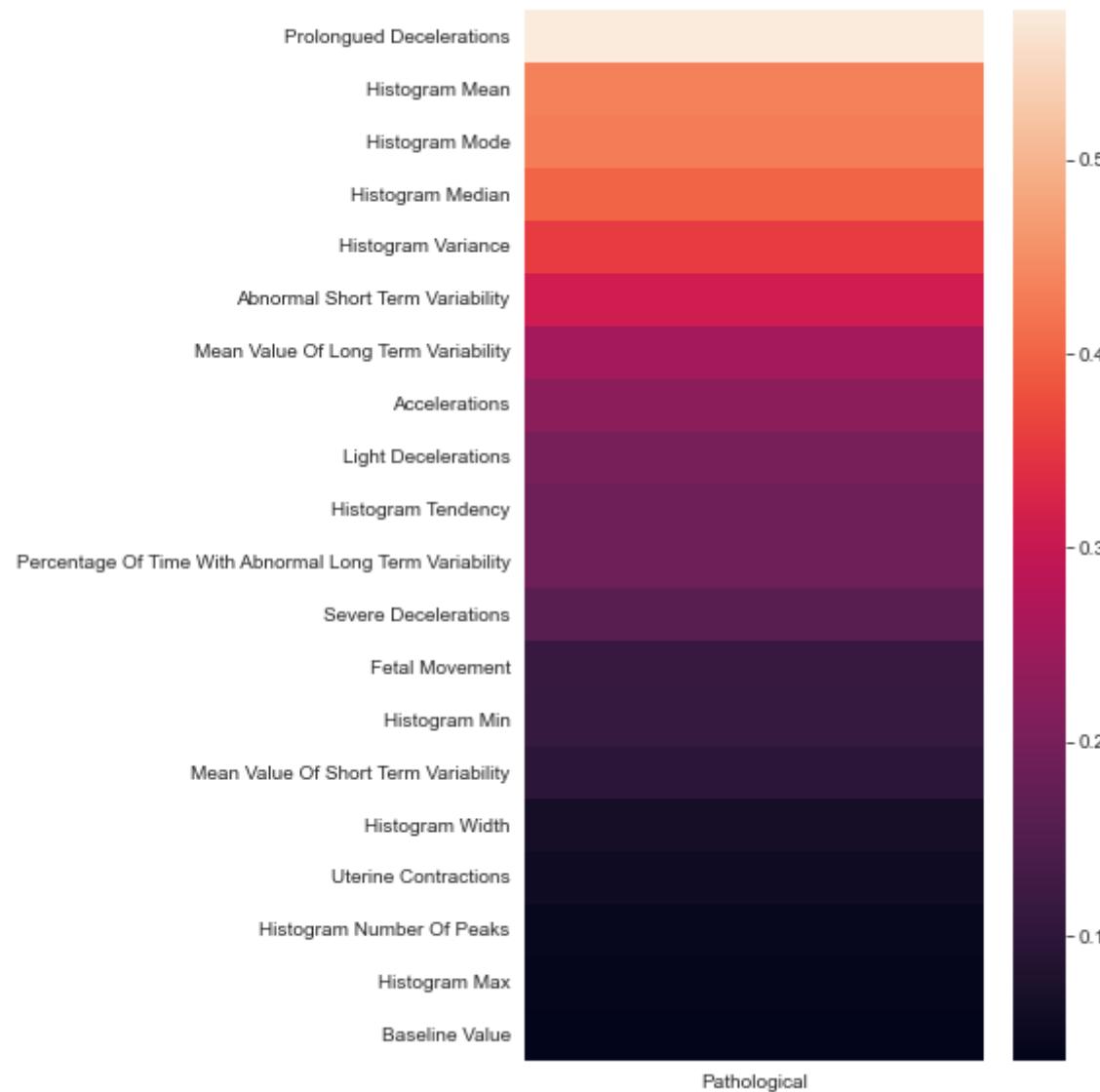
Median - histogram median

Variance - histogram variance

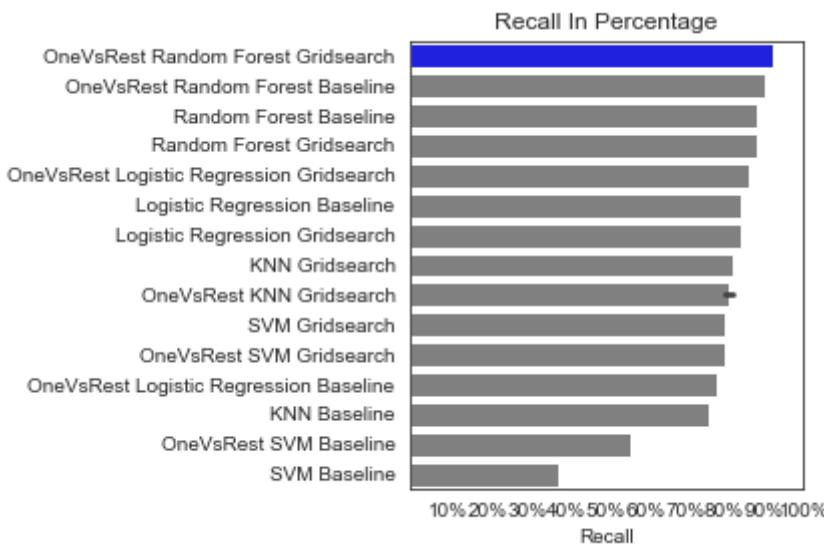
Tendency - histogram tendency

Exploring the correlations between the Target and the Pathological Class:





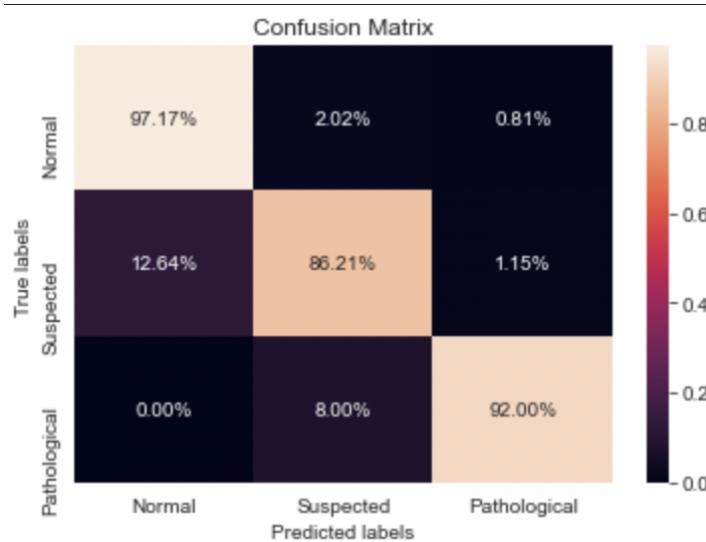
METHODS:



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.

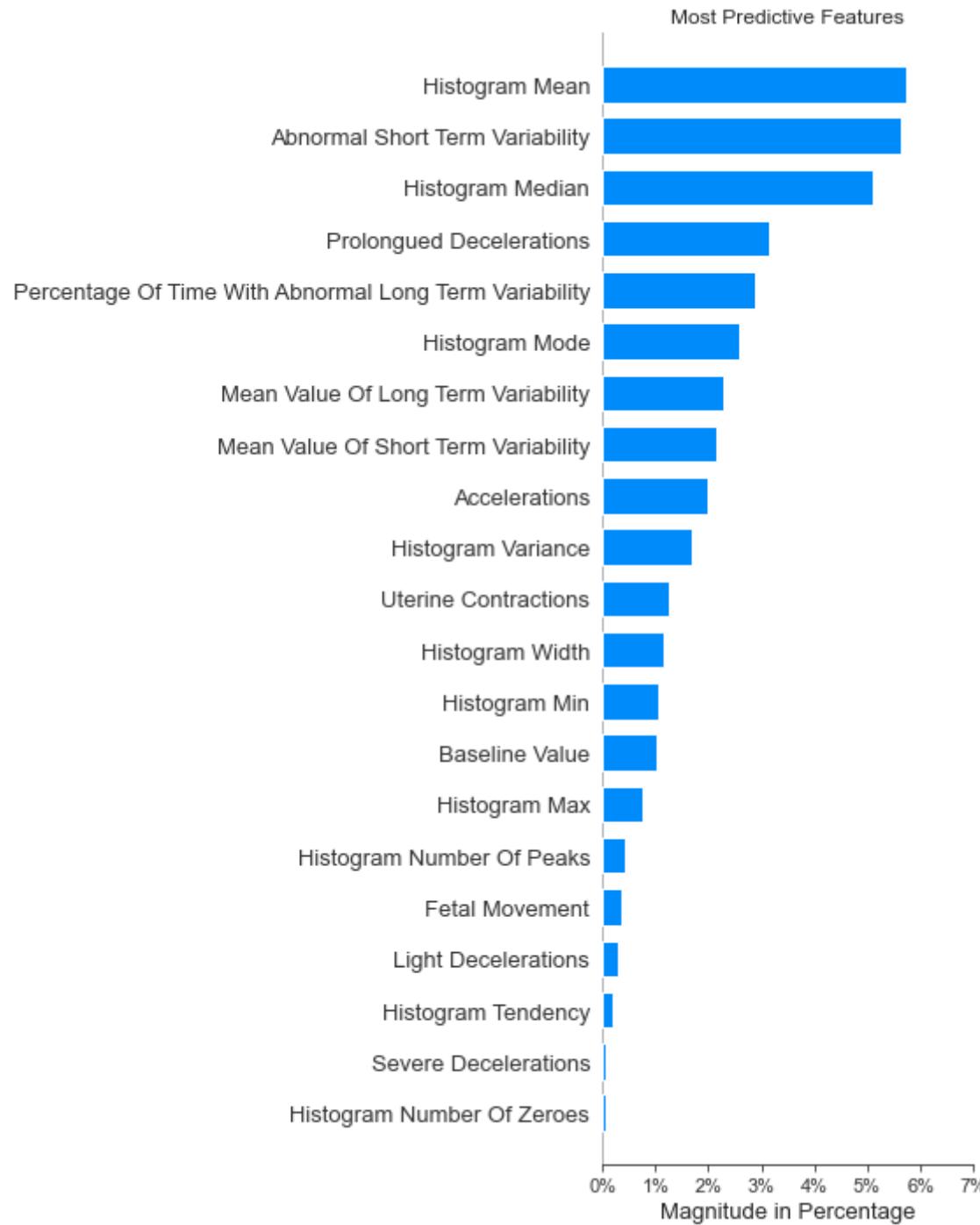
RESULTS:

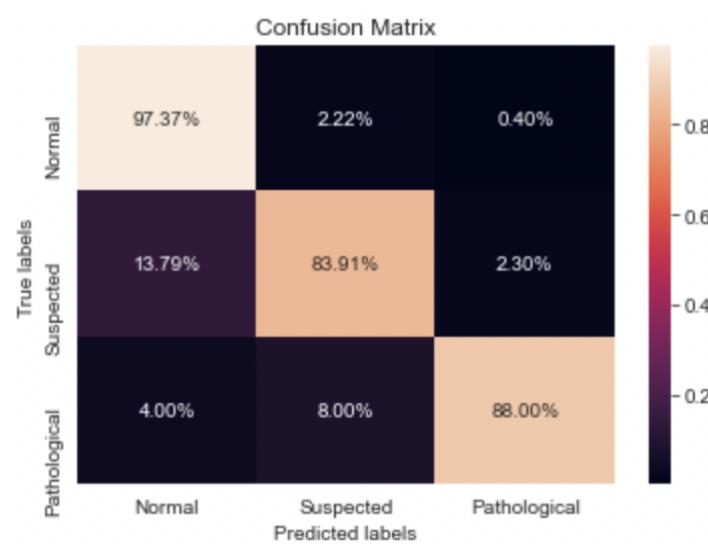
I have two final models to propose:



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. I'd like to revisit these models, especially my final model to retune them and see if I can't improve performance. Some of my models may have improved recall for the target class, or be more well rounded than my final model if they were less overfit.
2. I'd like to speak with the data collection team and see if we could negotiate a new class or indicator in the data to mark lethal prenatal diagnoses. Only a fraction of the pathological diagnoses our current data represents are cases of lethal prenatal diagnoses.
3. I'd like to iteratively test how I can reduce the dimensionality without sacrificing performance. Reducing the amount of tests needed to run for this model to produce reliable results will help the model's utility in the real world.

4. Most importantly: I want to make a secondary model. Right now, about 16% of Suspected cases were misclassified as Normal by my final model. I would like to make another model more sensitive to the "Suspected" class and have someone's metrics run through both models as a safety for that 16% of Suspected cases.

4.2 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 prenatal diagnoses specifically.

4.3 Lastly: This final model should be implemented by birthing centers, hospitals and women's health clinics as an early screening detection tool towards reducing Maternal Mortality rates, giving birthing parents a quicker diagnosis, and the best possible outcome in accordance to their own wishes.

Releases

No releases published

[Create a new release](#)

Packages

No packages published

[Publish your first package](#)

Languages

- Jupyter Notebook 100.0%