Springboard – DCS

Capstone Project 2

Predicting Fetal Cardiac Health Outcomes

Using Cardiotocogram Data

By Michael Bobal

January 2023

1) Introduction

According to the Centers for Disease Control (CDC), in the United States, congenital heart defects (CHDs) are the most common types of birth defects. CHDs affect nearly 1% of children born, totaling nearly 40,000 cases per year. Of those 1% of children born with CHDs, about 25% have a critical CHD. Infants with critical CHDs will need surgery within their first year of life in order to prevent unnecessary death. Children in the fetal stage of development can also suffer from in-utero hypoxia - and even death - if certain fetal cardiac issues are not detected in time.

Cardiotocography (CTG) is a non-invasive medical test used to assess fetal heart rhythm, as well as uterine contractions in the mother. Obstetricians use information acquired from CTG to assess fetal health, make determinations about the necessity of preventative C-sections, and inform parents about possible health issues which may require surgical intervention for the child.

Currently, highly-trained physicians are required for the reading of CTG data, where they determine if the fetus’ CTG reading class (a.k.a. Fetal CTG Class) is normal, suspect, or pathologic (N, S, or P, respectively). Having access to a fast, high-fidelity classification system based on this CTG data would be an ideal scenario.

An accurate diagnostic determination is a critical step for our stakeholders: parents, obstetricians, surgeons, and hospitals. These parties must be vigilant in case the need arises for medical intervention (C-section to save the life of the fetus, possible cardiac surgery in the first year of life after birth, etc.). Therefore, the development of such a system was the goal of this project.

The result of our efforts have concluded that the data supplied to the data science team is insufficient. The raw data was poorly configured, confusingly explained, and contained repeat and time-overlapping entries of the same patients, which made leakage-free modeling impossible.

The notebooks detailing the process which led to our conclusions can be found at the following link:

https://github.com/mikebobal/springboard/tree/main/Capstone2/Fetal%20health%20idea

(2) Approach

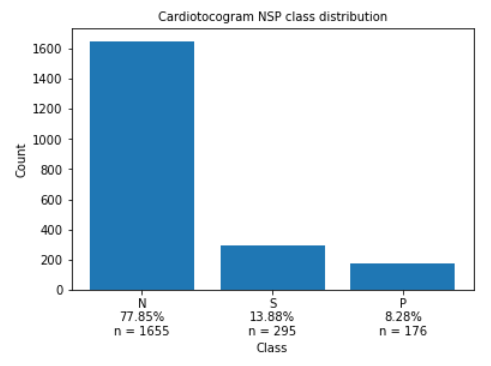
(2.1) Data Acquisition and Wrangling

The data came to us looking fairly clean. We found it in the University of California Irvine Machine Learning Repository. There were no missing values, but there were extra values due to some aggregate value rows. There were also many unneeded columns (evidenced by their exclusion from the “variables” list provided by the data documentation), which we removed. This cleaned dataset was saved for the next step.

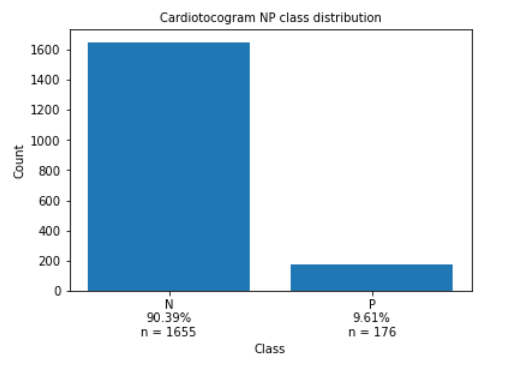
(2.2) Storytelling and Inferential Statistics

The data files presented us with 2126 CTG instances. Given the general reliability of the data source and the fact that the originating research was peer reviewed and published, we understood this to mean that the study used 2126 unique patients to compile their data. At the start of this project, we did not have access to the full research paper, and the researchers who conducted the study would not respond to our attempted communications. The study at hand was conducted in Portugal over 20 years ago, so we decided not to wait for a response that would never come.

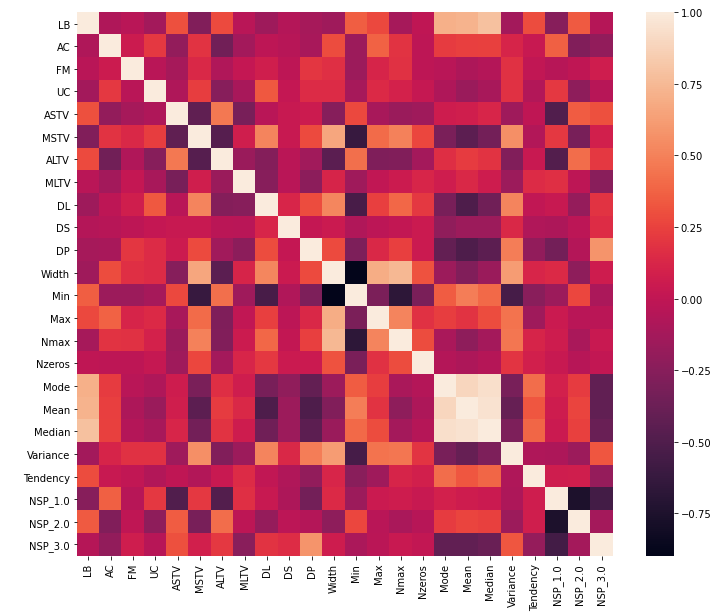
We discovered from the outset that the data was highly imbalanced on the diagnostic class. The researchers had medical experts classify each CTG instance as Normal, Suspect, or Pathological.



After considering the goal of our project, the decision was made to eliminate the Suspect class. In the health field, it is important to know if a patient has a serious health issue. Giving the diagnosis of “suspect” serves no cogent purpose. At this juncture, the plan was to use our best model at the end of the project to reclassify each Suspect class instance as either Normal or Pathological.



Afterward, we visualized how the features correlated with each other and with our target.



There were a few highly correlated variables seen. Particularly, Mean, Median, and Mode (all aggregate measures of central tendency).

(2.3) Baseline Modeling

Discuss the model/s you chose to build a baseline,

including a summary of the obtained results.

Leave the comparative analysis for section (3).

(2.4) Extended Modeling

Discuss what motivated the implementation of

other models, hopefully more advanced, including

a summary of the obtained results. Leave the

comparative analysis for section (3).

(3) Findings

Present a summary of all your results in

a table like the following:

PM1 PM2 PM3 ...

------------------------------------------------

Model 1 X X X

------------------------------------------------

Model 2 X X X

------------------------------------------------

Model 3 X X X

...

Where 'PM' stands for 'Performance Metric'.

Present a discussion of these results with two

goals in mind:

\* To highlight the relative merits (or lack thereof)

of each model. Including, if appropriate, the

identification of the "best model".

\* To connect this analysis with the original

business problem. In other words, to discuss

the implications of these results in the context

of the original business problem.

(4) Conclusions and Future Work

Your conclusions tie together the original

business problem, the data science problems

you decided to model the former with, and the

results you obtained.

Future work can be presented as a bulleted

list of approaches you would like to have tried

given more time (and resources--when applicable).

(5) Recommendations for the Clients

From the CPGD, it is required that you

offer "up to 3 concrete recommendations on how

your client[s] can use your findings."

This is the way to "close the loop" between

the original business problem, the data

science problems you decided to pursue, and

the results you obtained.

In "real-life" projects, this is where us data

scientists communicate the results to our clients

in a way that they suggest (or prescribe, in some

cases) concrete business interventions.

I would recommend that you present your recommendations

as a bulleted list.

(6) Consulted Resources

This is just a list of articles, packages,

books, and any other resource you used in

this project.

