Springboard–DSC Program
Capstone Project 2 Proposal

Can fetal cardiac abnormalities be accurately predicted based on data obtained from cardiotocography?

By Michael Bobal
November 2022

The Problem
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, totalling 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). An accurate determination is a critical step for parents, obstetricians, surgeons, and hospitals in terms of preparedness for medical intervention (C-section to save the life of the fetus, possible cardiac surgery in the first year of life after birth, etc.)

The Data
Using CTG data obtained from the research paper, Ayres de Campos et al. [1], and the UCI Machine Learning Laboratory [2], we aim to build several models that can be trained to accurately predict the Fetal CTG Class from an input characterizing a CTG.

Envisioned Approach
The plan is to use several supervised machine learning algorithms to predict the classification of our main target, based on the features of the CTG reading:

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

While the dependent variable will be our categorical target:
NSP - Fetal CTG Class (N=normal; S=suspect; P=pathologic)

The available dataset also provides an alternative target, known as the FHR Pattern Class Code, which is a number between 1 and 10. We will explore the use of this target as well.

Multiple classification algorithms will be used and they will be compared with respect to appropriate performance metrics. We will also explore so-called ensemble methods [3, 4].

Finally, we will explore some interpretability approaches to characterize how the variation of independent variables affect the target/s [5]

Deliverables
As required, the deliverables for this project will be: all Jupyter notebooks that will be developed (one per project phase), a written final report, and a presentation slide deck. All of these deliverables will be available from my GitHub repository,

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

[2] Ayres de Campos et al. (2000). UCI Machine Learning Repository . Irvine, CA: University of California, School of Information and Computer Science.
https://archive.ics.uci.edu/ml/datasets/cardiotocography

[3] https://machinelearningmastery.com/tour-of-ensemble-learning-algorithms/

[4] https://sebastianraschka.com/pdf/lecture-notes/stat479fs18/07_ensembles_slides.pdf

[5] https://christophm.github.io/interpretable-ml-book/