Predicting Fetal Health from Cardiotocogram Results

Goal:

Create a predictive model from Cardiotocogram results to indicate to health care workers when clinical interventions are required

From the Kaggle description:

Reduction of child mortality is reflected in several of the United Nations' Sustainable Development Goals and is a key indicator of human progress.
The UN expects that by 2030, countries end preventable deaths of newborns and children under 5 years of age, with all countries aiming to reduce under-5 mortality to at least as low as 25 per 1,000 live births.

Parallel to notion of child mortality is of course maternal mortality, which accounts for 295 000 deaths during and following pregnancy and childbirth (as of 2017). The vast majority of these deaths (94%) occurred in low-resource settings, and most could have been prevented.

In light of what was mentioned above, Cardiotocograms (CTGs) are a simple and cost accessible option to assess fetal health, allowing healthcare professionals to take action in order to prevent child and maternal mortality. The equipment itself works by sending ultrasound pulses and reading its response, thus shedding light on fetal heart rate (FHR), fetal movements, uterine contractions and more.

Information source:

Project inspired by and data downloaded from Kaggle Dataset:
https://www.kaggle.com/andrewmvd/fetal-health-classification

Data from:

Ayres de Campos et al. (2000) SisPorto 2.0 A Program for Automated Analysis of Cardiotocograms. J Matern Fetal Med 5:311-318
https://onlinelibrary.wiley.com/doi/10.1002/1520-6661(200009/10)9:5%3C311::AID-MFM12%3E3.0.CO;2-9

Data Science Problem

Classification algorithms will be used to take the Cardiotocogram data and predict Fetal Health

Data consists of a .csv file with 22 columns, one of which is coded for fetal health as:

1 (Normal), 2 (Suspect) and 3 (Pathological)

F1 Score will be used as the primary metric for comparing models

Clean Data

Little data cleaning was necessary

Data had no missing or null values, no apparent typos or inappropriate entries

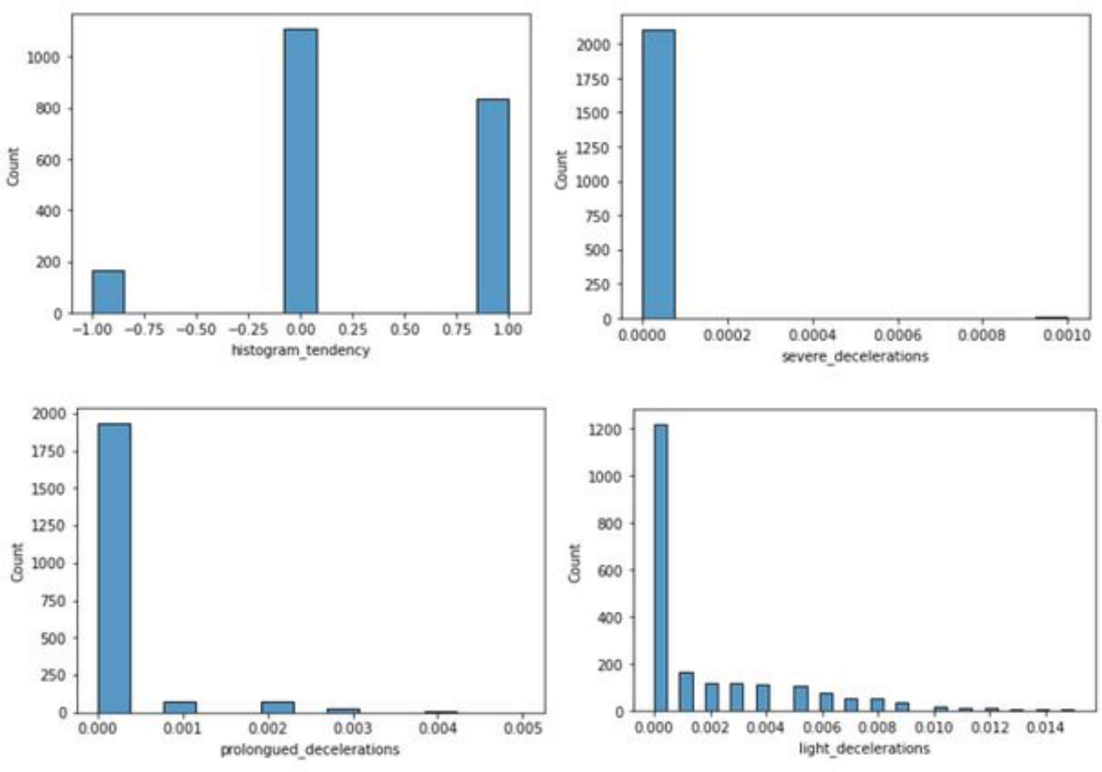
Thirteen rows deleted because they were duplicates of previous entries

Model Decision Points
Code categorical variables

“Light decelerations” - left as is

“Prolonged decelerations” and “histogram tendency” - one hot encoded

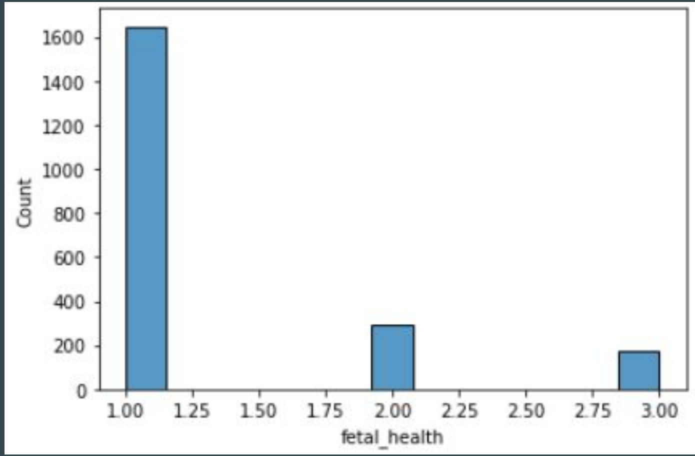
“Severe decelerations” - only 7 non-zero values. These coded as “1” due to correlation with clinical classification



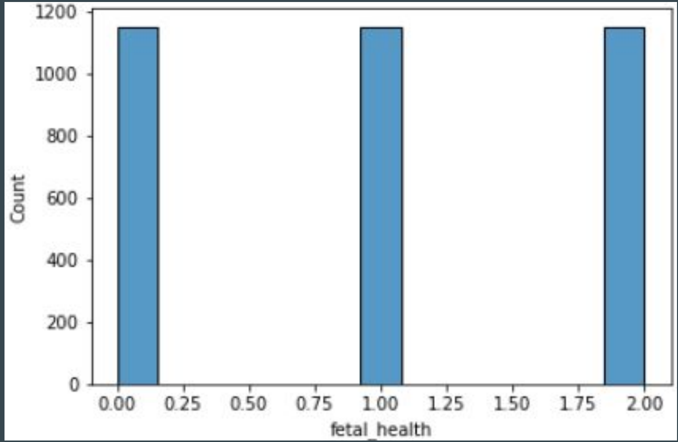
Bring Balance to the Force

Clinically relevant and borderline outputs significantly outnumbered by “normal” classification

Synthetic Minority Oversampling Technique (SMOTE) Balancing performed on Training Set

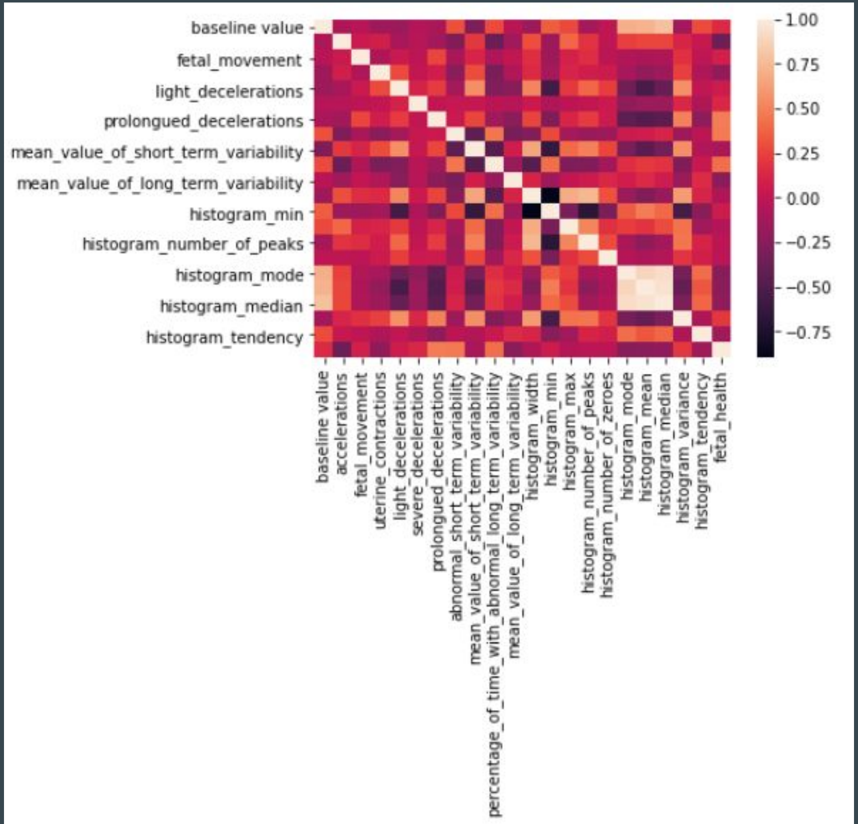


Unbalanced outputs



OverSampled Outputs

Study of factors



No significant auto correlation observed which would require removal of factors for classification.

No factors highly correlated with the output variable

Correlations between numerical factors

Model selection

Four different classification models were created

Models were evaluated based on the F1 scores

	KNN	Random Forest	Stock XGBoost	Tuned XGBoost
F1 - “normal”	.95	.96	.97	**.97**
F1 - “suspect”	.72	.78	.80	**.83**
F1 - “pathological”	.85	.94	.96	**.98**
weighted average F1	.91	.94	.95	**.95**

Best model is the Tuned XGBoost

Evaluation

The XGBoost Model provides a 0.95 F1 score, with 0.98 in the most important category of “pathological”.

This is adequate to provide important guidance to practitioners as to when interventions should be taken

Next Steps

Model is adequate for production

Can use this model to create product (app or web site) to assist practitioners

Further work

- Spend more time optimizing hyperparameters
  - Might be able to increase the accuracy with more work
  - Would need to be cautious of over-fitting
- Use more data sets
  - Different data sets could provide for broader application
  - Original data coded by only 3 experts