# Foetal Health using CTG data

Analysis, Classification and Prediction of foetal health through supervised learning models.

*HSLU Luzern*

*Master in Applied Information and Data Science*

Sasa Ljubisavljevic
Jonas Renfer
Daniele Buson

24th March 2023

## *Introduction*

Reduction of child mortality is reflected in several of the United Nations' Sustainable Development Goals and is a key indicator of human progress.
In the last three decades, the world made remarkable progress in child survival. In the 90's, 1 in 11 children died before the age of five. Nowadays, the ratio is 1 in 26 children. (1)
The UN expects to lower this percentage to 25 deaths in 1000 births by 2030.
Another aspect, alongside this, that is crucial in the Sustainable Development Goals is maternal mortality. About 287 000 women died during and following pregnancy and childbirth in 2020. Almost 95% of all maternal deaths occurred in low and lower middle-income countries in 2020, and most could have been prevented. (2)
In this context, Cardiotocography (CTGs) is a simple and cost accessible option to assess foetal 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 foetal heart rate (FHR), foetal movements, uterine contractions and more.

## *Question*

1) "*Can we create a model to classify and predict foetal health, based on CTG results?"*
2) *"In case the answer to the first question was affirmative, can the obtained model be used as a tool to assist doctors and obstetricians to prevent child and maternal mortality?*"

These are the questions this project seeks to answer.
For what concerns the first question, we believe that it is possible to create a model. However, we do not take it for granted. On the contrary, we would like to show how to do that through the analysis of data and machine learning algorithms. We are going to discuss this topic more deeply in the "Data & Methods" section.
The second question is the real challenge of this project. Once the model is done, we would like to analyse it in order to understand its real potential. Again, more information can be found in the "Data & Methods" section.

## *Project Structure*

1. Introduction
2. Dataset Overview
    - 2.1. Table
    - 2.2. Correlation Plot
3. EDA
    - 3.1. Heart Rate
    - 3.2. Accelerations and Decelerations
    - 3.3. Foetal and Uterine Movements
    - 3.4. Short Term Variability (STV) vs Long Term Variability
    - 3.5. Heart Beat Histogram
    - 3.6. Foetal Health Condition
4. Machine Learning Models
    - 4.1. Creation of Test Data and Train Data
        - 4.1.1. SVM Model
        - 4.1.2. Random Forest Classifier Model
        - 4.1.3. Multinomial Logistic Regression Model
5. Conclusion
    - 5.1. Answer to the project's question
    - 5.2. Possible implementations for the future
    - 5.3. (IF possible) Shiny App, to present the result interactively

### ***Data & Methods***

The dataset (5)  consists of 2126 records of features extracted from a CGT exam. Three expert obstetricians categorised them into three groups (1 = Normal, 2 = Suspect, 3 = Pathological).
The dataset columns are defined as follow:

- ***baseline.value*** : baseline foetal heart rate
- ***accelerations*** : number of accelerations per second
- ***fetal_movement*** : number of foetal movements per second
- ***uterine_contractions*** : number of uterine contractions per second
- ***light_decelerations*** : number of LDs per second (3)
- ***severe_decelerations*** : number of SDs per second (3)
- ***prolonged_decelarations***: number of PDs per second (3)
- ***abnormal_short_term_variability*** : percentage of time with abnormal short term variability (4)
- ***mean_value_of_short_term_variability*** : mean value of short term variability (4)
- ***percentage_of_time_with_abnormal_long_term_variability*** : percentage of time with abnormal long term variability (4)
- ***mean_of_long_term_variability*** : mean value of long term variability (4)
- ***histogram_width*** : width of the histogram made using all values from a record
- ***histogram_min*** : histogram minimum value
- ***histogram_max*** : histogram maximum value
- ***histogram_number_of_peaks*** : number of peaks in the exam histogram
- ***histogram_number_of_zeros*** : number of zeroes in the exam histogram
- ***histogram_mode*** : histogram mode
- ***histogram_mean*** : histogram mean
- ***histogram_median*** : histogram median
- ***histogram_variance*** : histogram variance
- ***histogram_tendency*** : histogram trend
- ***fetal_health*** : foetal health, 1 = normal, 2 = suspect, 3 = pathological

For what concerns the methods, we have decided to develop our project in R, showing our result through Rmarkdown. As you can imagine, the methods we are going to use depend on the purpose that we want to achieve.
In the "Overview" section we would like to let the reader know general aspects of our dataset. We have, therefore, decided to show the table and a paired scatterplot (using the function ggpairs). In the EDA we decided to group different columns under the same subject. The graphical analysis is conducted using box plots or histograms (depending on graphical or statistical considerations) from the library ggplot2.
Through indexing we are going to separate the dataset into two subsets, one for testing and one for training. This operation could be iterated to have more strong evidence of the result.
We have opted to create three different models, using supervised learning algorithms such as SVM, Random Forest Classifier and Multinomial Logistic Regression.
In conclusion, we are going to evaluate these models looking at their precision or recall (or even a combination of the two, using the F-score). This point will be further investigated to arrive at the best solution for our problem.
Time permitting, we would love to create a Shiny App to show what we achieved in a more dynamic and entertaining way.

### ***Sources***


(1) https://data.unicef.org/topic/child-survival/under-five-mortality/
(2) https://www.who.int/news-room/fact-sheets/detail/maternal-mortality
(3) https://www.ncbi.nlm.nih.gov/books/NBK557393/
(4) https://pubmed.ncbi.nlm.nih.gov/11874624/
(5) https://www.kaggle.com/datasets/andrewmvd/fetal-health-classification