Midterm Project

Fetal Health Classification

By:
Vidusha Wilpita
&
Ikenna Odinye

Project Outline

- Problem Statement and Importance
- Dataset description
- EDA
- Model selection
- Model evaluation
- Findings/Conclusion

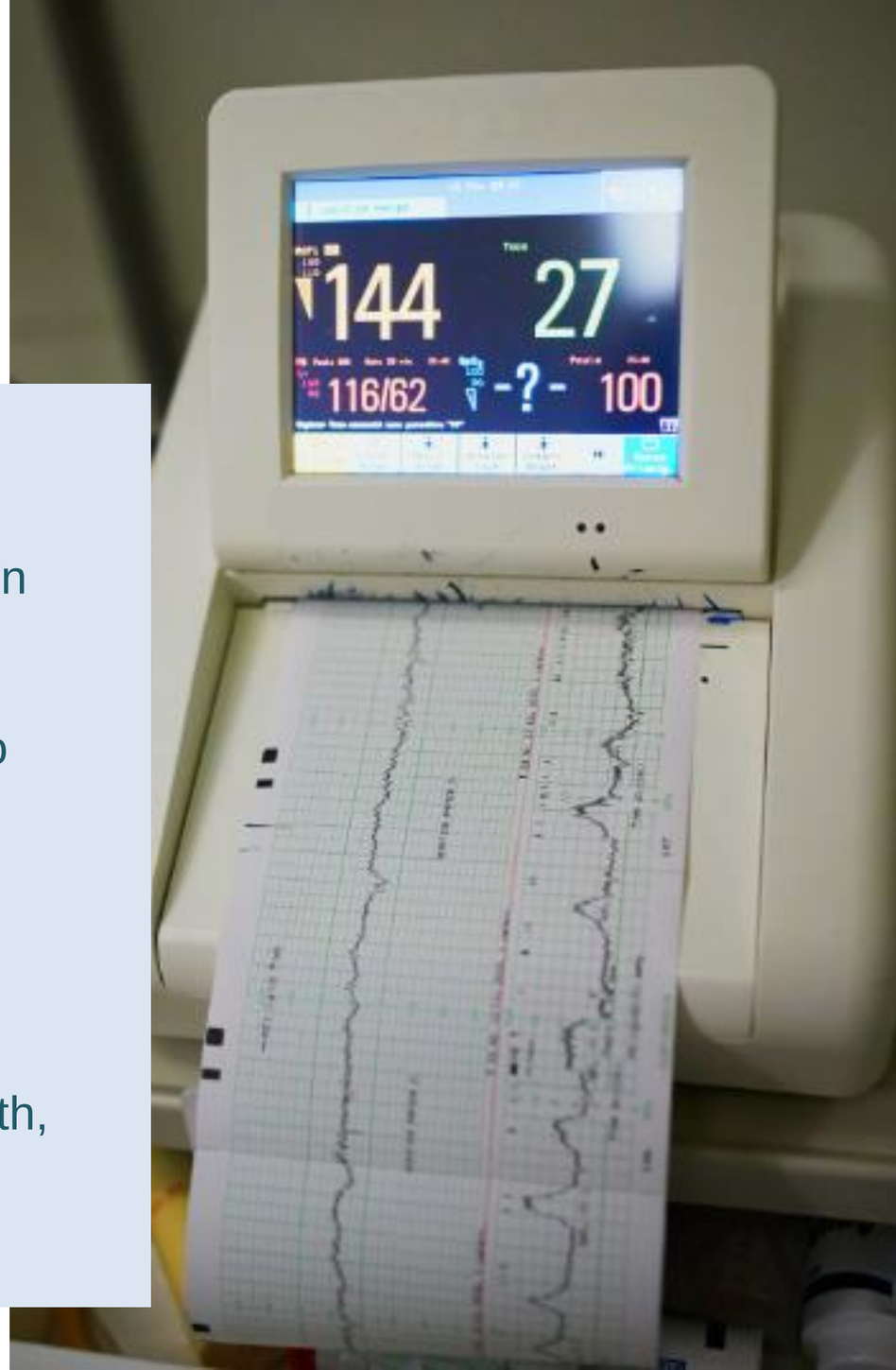
Problem Statement and project importance

Problem:

- Under-five mortality is greatly influenced by the death of children in the first month of their life. Perinatal mortality is largely caused by intrapartum complications. However, Cardiotocograms (CTGs), which are simple and affordable, can serve as a monitoring tool to assess fetal health and identify high-risk women during labor.

Why is it important?

- To help achieve the MDG 4 goal of globally reducing under-five mortality by two-thirds. Additionally, to reduce maternal mortality, which accounts for 295,000 deaths during pregnancy and childbirth, with 94% occurring in low-resource settings, many of which are preventable.



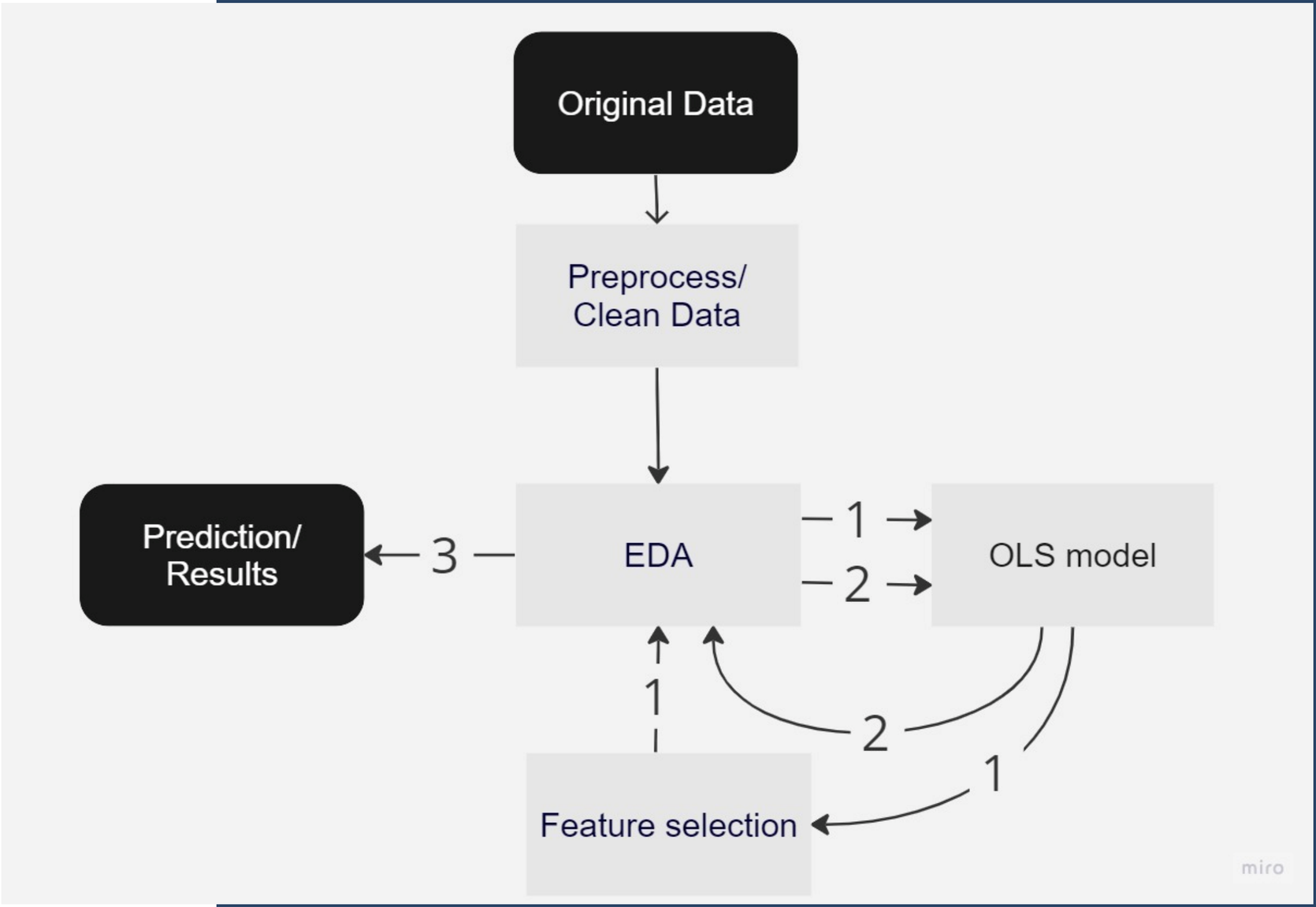
Dataset Description

- Our dataset consists of 2126 CTG exam records of pregnant women in their third trimester. It includes 22 features, with Fetal Health classified into three classes (Normal as 1, Suspect as 2, and Pathological as 3) by expert obstetricians.

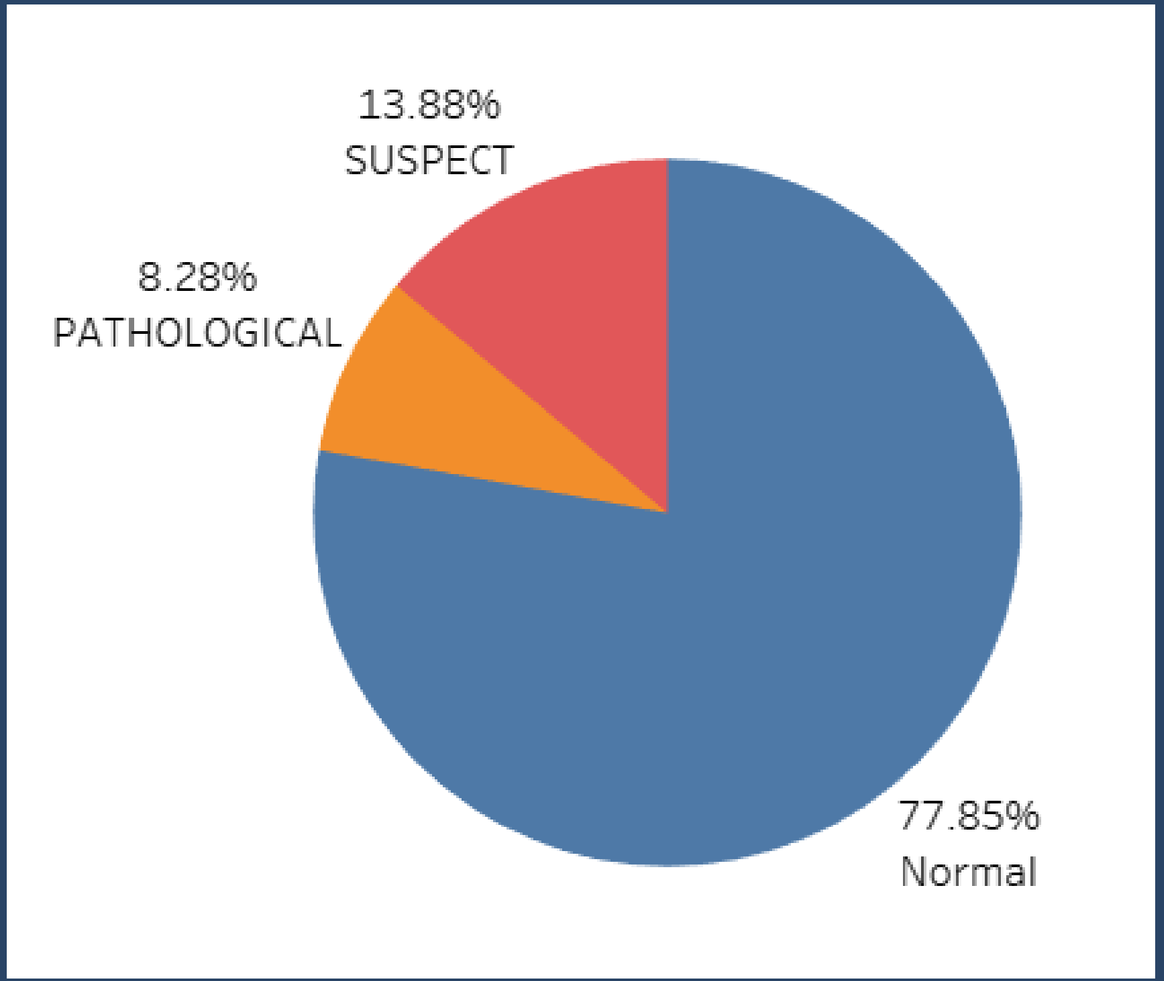
Project Objectives

1. To identify the most significant features for detecting high-risk fetal health conditions.
2. To analyze the precision of CTG model algorithms in correctly classifying instances.
3. Based on the model evaluation result, propose the global scale-up of CTG usage, with a focus on regions burdened by high newborn and maternal mortality.

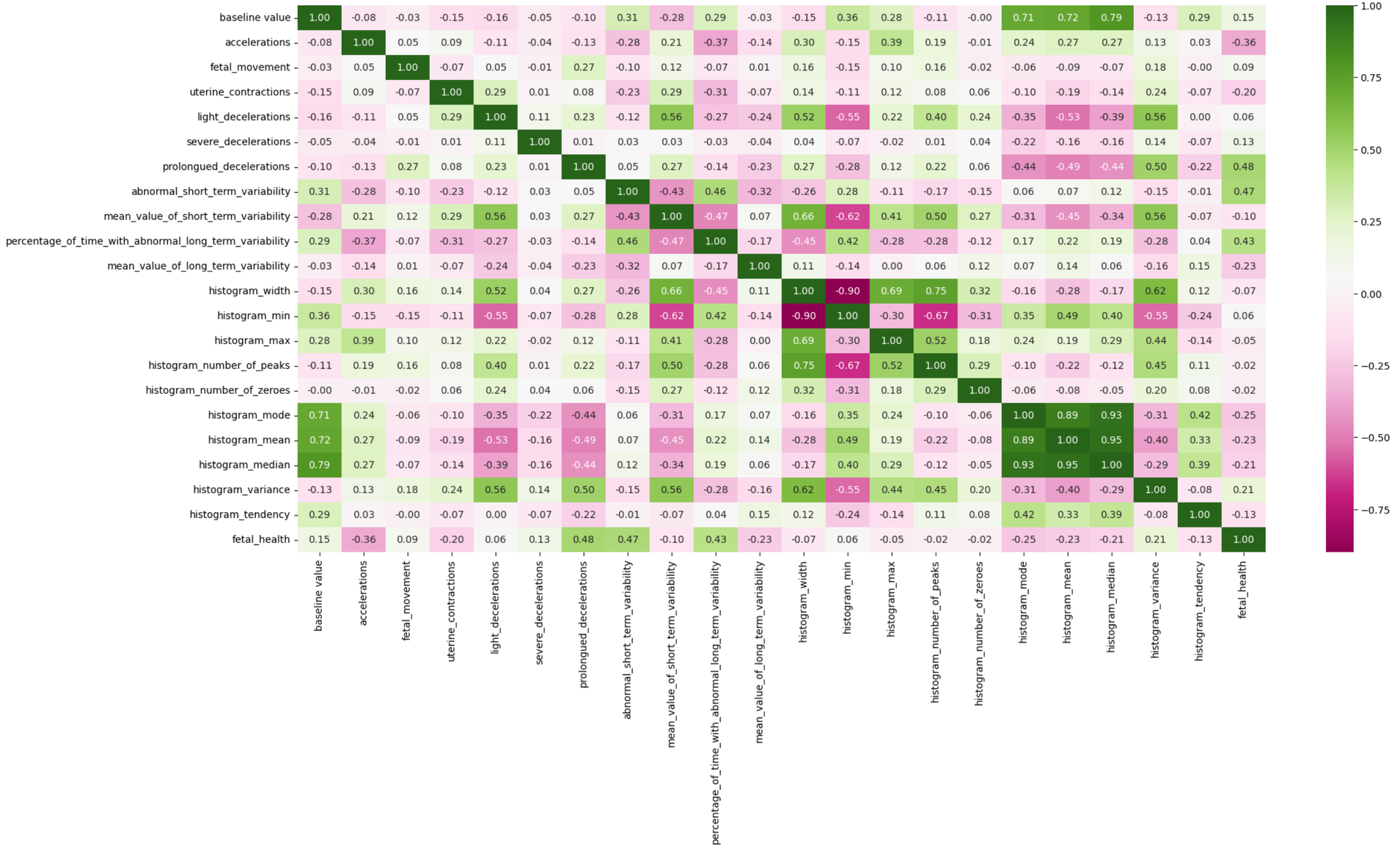
Approach


Original Data
Preprocess/
Clean Data
Prediction/
Results
3
EDA
1
2
OLS model
1
2
1
Feature selection
miro

Fetal Health Category



Heatmap of Correlation



OLS Regression Results

OLS Regression Results

Dep. Variable:	fetal_health	R-squared:	0.607
Model:	OLS	Adj. R-squared:	0.604
Method:	Least Squares	F-statistic:	162.8
Date:	Tue, 11 Jul 2023	Prob (F-statistic):	0.00
Time:	10:07:41	Log-Likelihood:	-986.93
No. Observations:	2126	AIC:	2016.
Df Residuals:	2105	BIC:	2135.
Df Model:	20		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.2620	0.135	1.937	0.053	-0.003	0.527
baseline value	0.0155	0.002	6.995	0.000	0.011	0.020
accelerations	0.1378	3.690	0.037	0.970	-7.099	7.375
fetal_movement	0.0842	0.192	0.439	0.661	-0.292	0.460
uterine_contractions	-22.8929	3.226	-7.096	0.000	-29.220	-16.566
light_decelerations	-3.4471	5.267	-0.654	0.513	-13.776	6.882
severe_decelerations	761.1237	156.006	4.879	0.000	455.182	1067.065
prolongued_decelerations	362.2050	23.525	15.397	0.000	316.071	408.339
abnormal_short_term_variability	0.0086	0.001	12.716	0.000	0.007	0.010
mean_value_of_short_term_variability	0.0017	0.016	0.102	0.919	-0.030	0.034
percentage_of_time_with_abnormal_long_term_variability	0.0117	0.001	18.742	0.000	0.010	0.013
mean_value_of_long_term_variability	0.0058	0.002	2.697	0.007	0.002	0.010
histogram_width	0.0002	0.000	0.619	0.536	-0.000	0.001
histogram_min	0.0034	0.001	5.653	0.000	0.002	0.005
histogram_max	0.0036	0.001	5.106	0.000	0.002	0.005
histogram_number_of_peaks	-0.0029	0.004	-0.669	0.504	-0.012	0.006
histogram_number_of_zeroes	0.0124	0.013	0.958	0.338	-0.013	0.038
histogram_mode	-0.0037	0.002	-2.405	0.016	-0.007	-0.001
histogram_mean	-0.0035	0.002	-1.456	0.146	-0.008	0.001
histogram_median	-0.0112	0.003	-3.724	0.000	-0.017	-0.005
histogram_variance	0.0028	0.000	5.988	0.000	0.002	0.004
histogram_tendency	0.0825	0.023	3.537	0.000	0.037	0.128

Omnibus:	237.892	Durbin-Watson:	0.888
Prob(Omnibus):	0.000	Jarque-Bera (JB):	514.911
Skew:	0.679	Prob(JB):	1.54e-112
Kurtosis:	4.992	Cond. No.	4.65e+15

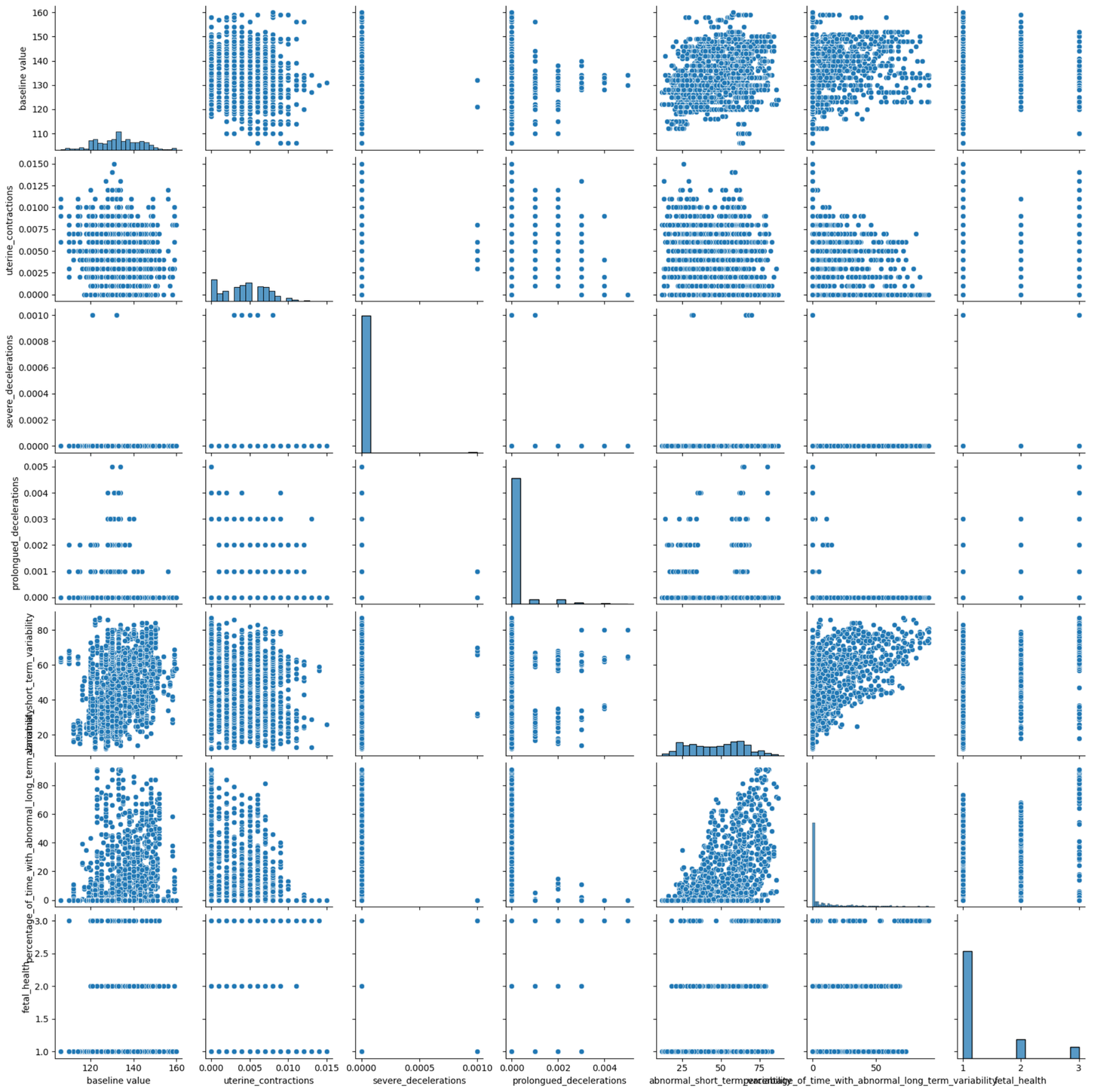
Selected Features

Our EDA revealed most significant features that affect fetal health as:

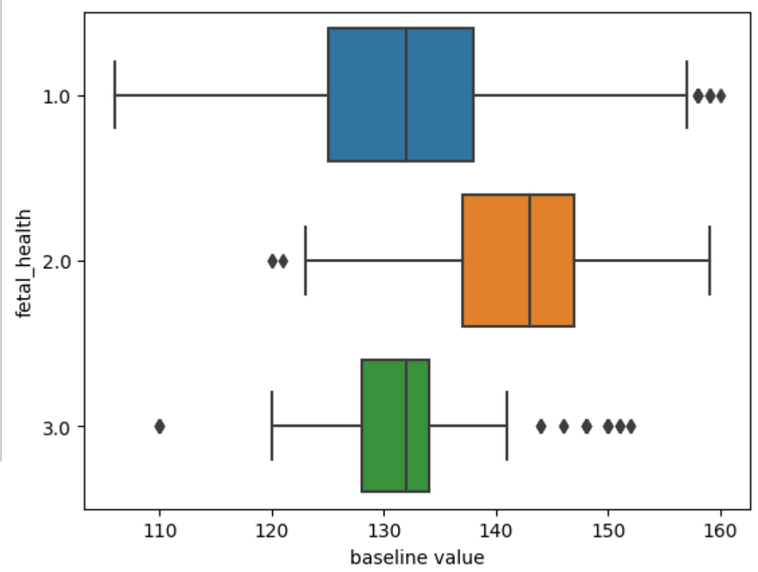
- ✓ **Baseline value** - FHR baseline (beats per minute)
- ✓ **Uterine contractions** - Number of uterine contractions per second
- ✓ **Severe Decelerations** - Number of severe decelerations per second
- ✓ **Prolongued Decelerations** - Number of prolonged decelerations per second
- ✓ **Abnormal short-term variability** - Percentage of time with abnormal short-term variability

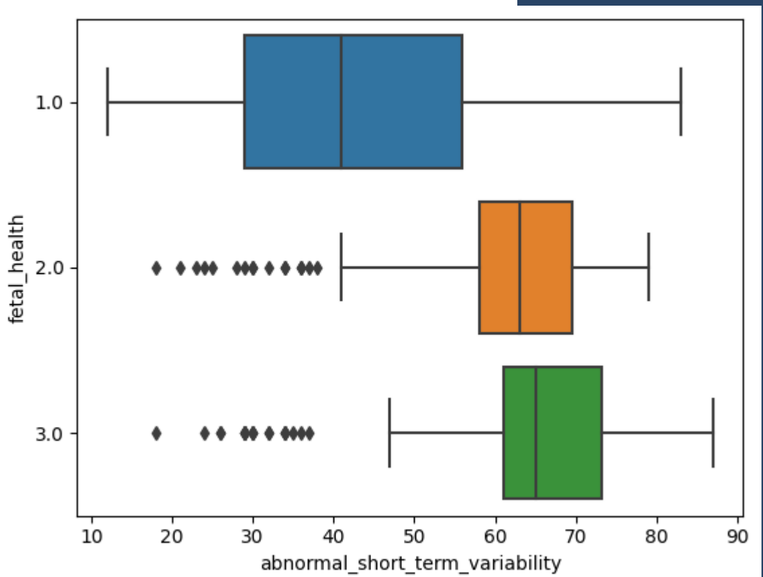
Percentage of time with abnormal long-term variability (identified using the p-value of OLS Regression model)

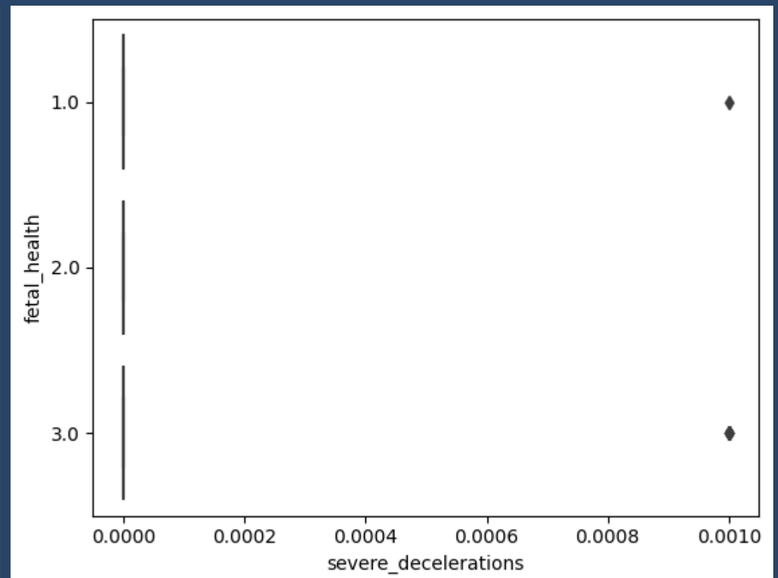
Pairplot for significant features

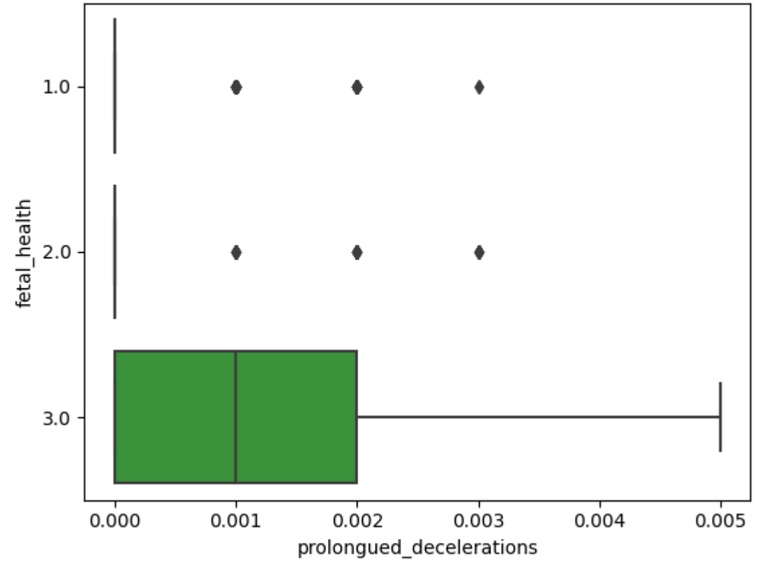


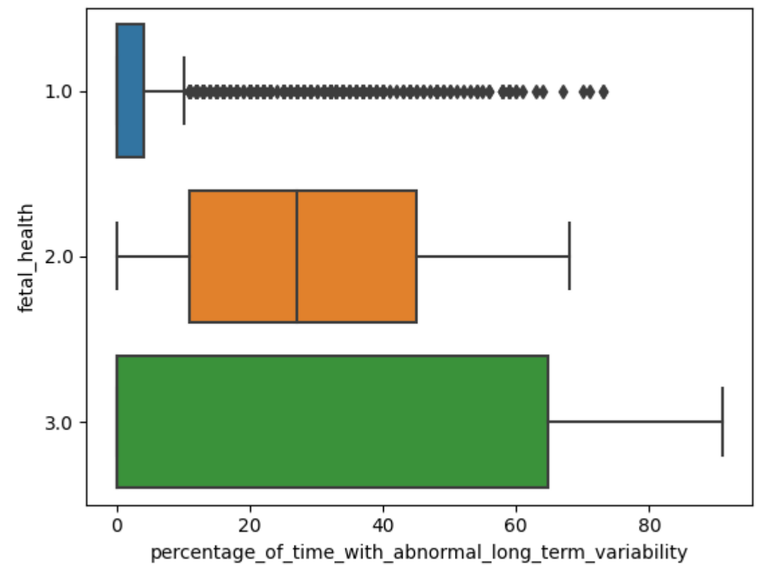
Boxplots of each selected Feature

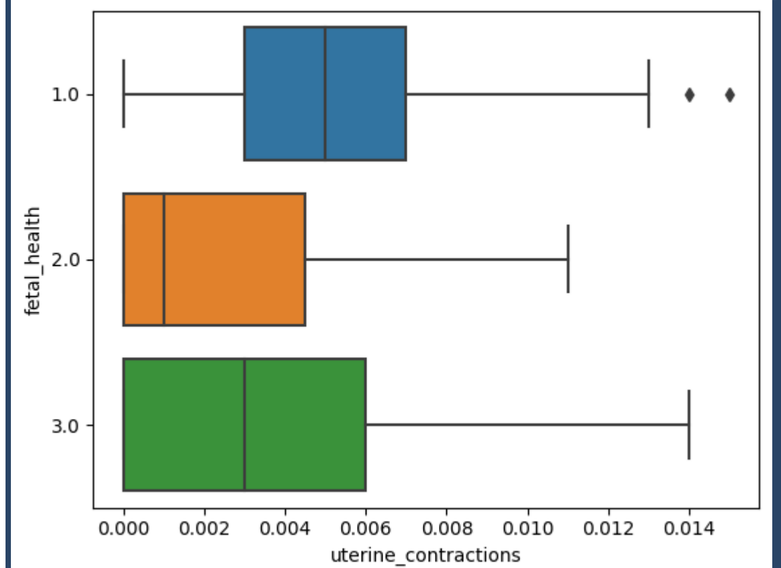












T-tests Results

Feature	Category 1	Category 2	Result	> / < 0.05
Baseline value	normal	Suspect + pathological	4.02e-32	<
	suspect	pathological	1.24e-30	<
Prolonged Decelerations	normal	suspect	0.020	<
Abnormal short-term variability	normal	suspect	8.19e-85	<
	normal	pathological	2.42e-67	<
	suspect	pathological	0.031	<
Percentage of time with abnormal long-term variability	normal	suspect	1.90e-151	<
	normal	pathological	8.57e-47	<
	suspect	pathological	0.013	<
Severe Decelerations	normal	suspect	0.673	>
	normal	pathological	5.83e-12	<
	suspect	pathological	0.0014	<
Uterine Contractions	normal	suspect	2.196e-41	<
	normal	pathological	1.10e-5	<
	suspect	pathological	2.25e-6	<

Model Evaluation Result

- 70% accuracy rate

Conclusion

- Model represents proficient ability to classify various classes

Challenges

- Choosing a suitable project topic
- Limited time

Future Goals

- Deeper analysis
- Model evaluations

Thank you!