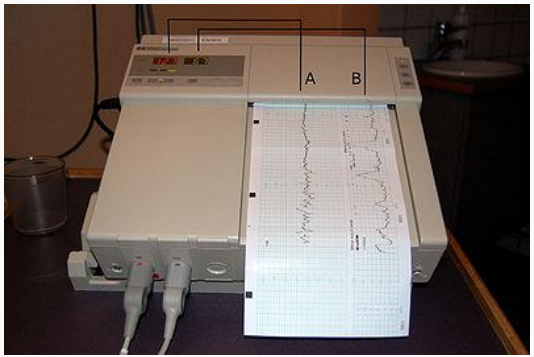
Fetal Health Classification

By: Aina Ahanmisi

Table of Contents

Introduction

- Every year, according to the Center for Disease Control and Prevention, about one million fetal deaths occur in the United States and almost three million in the world (DeSisto et al. 2021).
- The reduction of fetal death is an important goal because it has many beneficial effects for all.
- The objective of this project is to assist in the goal by properly identifying the status of a fetus early enough to reverse the causes leading to death through data exploration and classification.
- Target Audience



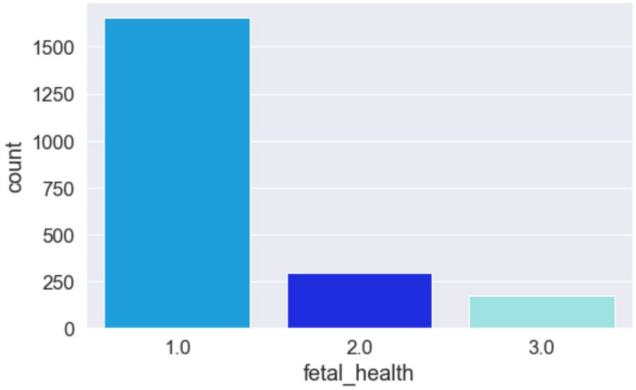
Data Description

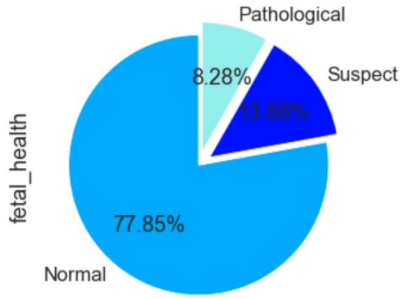
- Heart Rate
- Heart Acceleration
- Movement
- Uterine Contractions
- Decelerations
- Variability Calculations
- Histogram
- Fetal Health
  - Normal
  - Suspect
  - Pathological

	count	mean	std	min	25%	50%	75%	max
baseline value	2126.0	133.303857	9.840844	106.0	126.000	133.000	140.000	160.000
accelerations	2126.0	0.003178	0.003866	0.0	0.000	0.002	0.006	0.019
fetal_movement	2126.0	0.009481	0.046666	0.0	0.000	0.000	0.003	0.481
uterine_contractions	2126.0	0.004366	0.002946	0.0	0.002	0.004	0.007	0.015
light_decelerations	2126.0	0.001889	0.002960	0.0	0.000	0.000	0.003	0.015
severe_decelerations	2126.0	0.000003	0.000057	0.0	0.000	0.000	0.000	0.001
prolongued_decelerations	2126.0	0.000159	0.000590	0.0	0.000	0.000	0.000	0.005
abnormal_short_term_variability	2126.0	46.990122	17.192814	12.0	32.000	49.000	61.000	87.000
mean_value_of_short_term_variability	2126.0	1.332785	0.883241	0.2	0.700	1.200	1.700	7.000
percentage_of_time_with_abnormal_long_term_variability	2126.0	9.846660	18.396880	0.0	0.000	0.000	11.000	91.000
mean_value_of_long_term_variability	2126.0	8.187629	5.628247	0.0	4.600	7.400	10.800	50.700
histogram_width	2126.0	70.445908	38.955693	3.0	37.000	67.500	100.000	180.000
histogram_min	2126.0	93.579492	29.560212	50.0	67.000	93.000	120.000	159.000
histogram_max	2126.0	164.025400	17.944183	122.0	152.000	162.000	174.000	238.000
histogram_number_of_peaks	2126.0	4.068203	2.949386	0.0	2.000	3.000	6.000	18.000
histogram_number_of_zeroes	2126.0	0.323612	0.706059	0.0	0.000	0.000	0.000	10.000
histogram_mode	2126.0	137.452023	16.381289	60.0	129.000	139.000	148.000	187.000
histogram_mean	2126.0	134.610536	15.593596	73.0	125.000	136.000	145.000	182.000
histogram_median	2126.0	138.090310	14.466589	77.0	129.000	139.000	148.000	186.000
histogram_variance	2126.0	18.808090	28.977636	0.0	2.000	7.000	24.000	269.000
histogram_tendency	2126.0	0.320320	0.610829	-1.0	0.000	0.000	1.000	1.000
fetal_health	2126.0	1.304327	0.614377	1.0	1.000	1.000	1.000	3.000

Methodology: Data Preparation

- Cleansing
  - Missing Data
- Upsampling





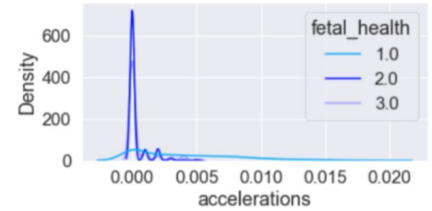
Methodology: Exploration

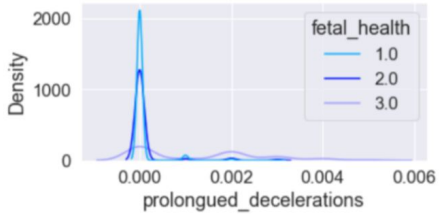
These features were picked by examining the heatmap and the relationship between predictor versus response.

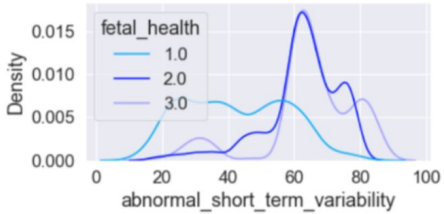
Table 4: Correlations Higher than 40%

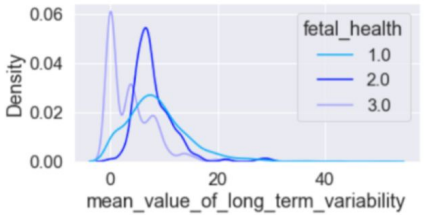
Feature	Correlation Percent
accelerations	-0.49
prolongued_decelerations	0.5
abnormal_short_term_variability	0.53
mean_value_of_long_term_variability	-0.4
histogram_mode	-0.44
histogram_mean	-0.44
histogram_median	-0.43

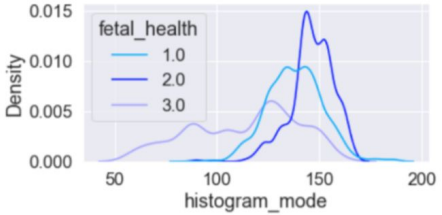
Distributions of Correlated Features versus Fetal Health

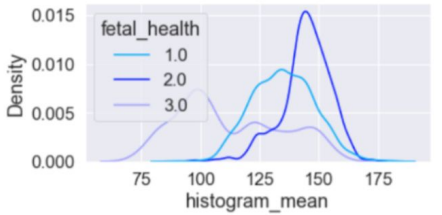


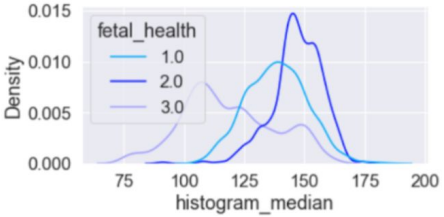




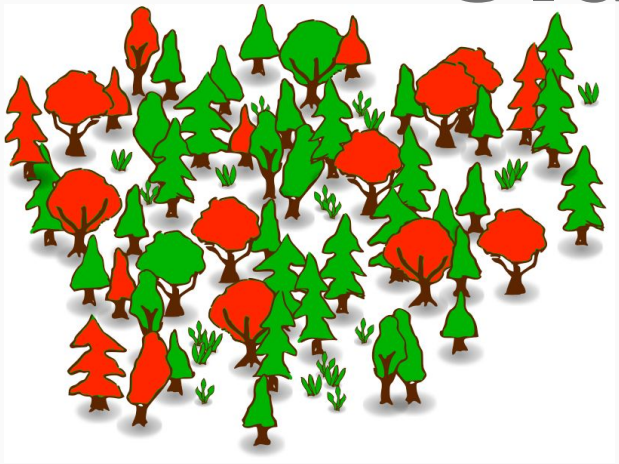




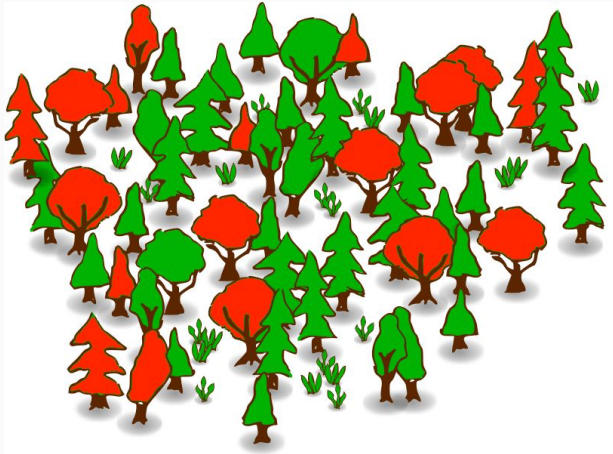




Classifications



- Linear Regression
- Random Forest
- Decision Tree



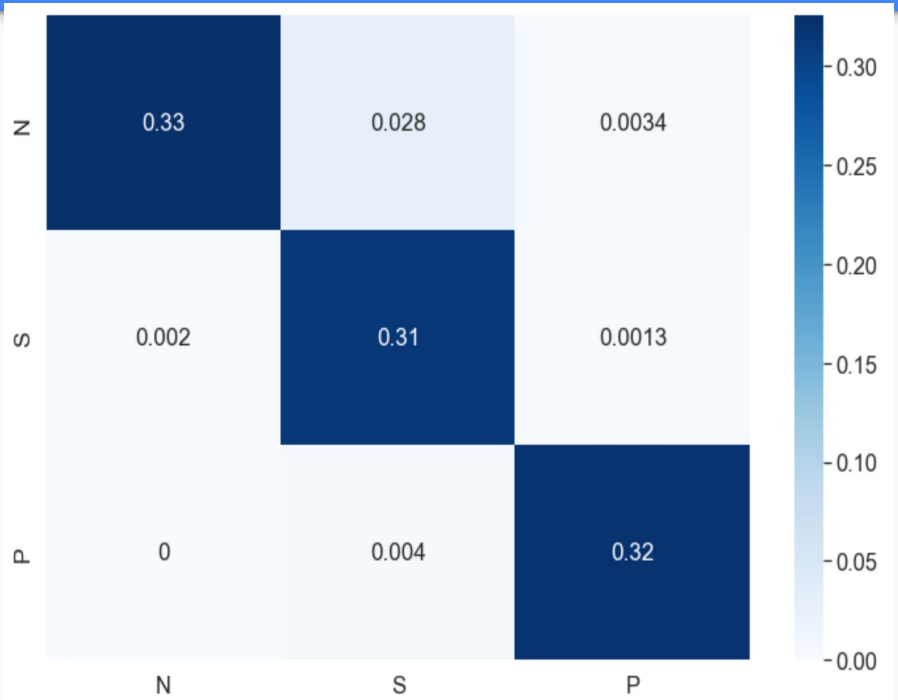
Methodology: Classification

- A total of three models were built and the highest accuracy was further optimized for a higher score.

Table 9: Accuracy Recap

Classifiers	**Accuracy**
Linear Regression	81%
Random Forest	85%
Decision Tree	72%
Optimal Random Forest	**96%**

Final Results



96% Accuracy

4% Error Rate

Based on the feature of importance, abnormal short term variability was the biggest influence to this model.

Recap or Recommendations

- The features that were parsed out based on the heatmap and predictor versus response relationship were all associated with the heart, so it would be good to mainly focus on the heart more than the other variables that were not used.
- Abnormal short term variability showed to have the most influence for such a high accuracy.
- Random forest was the best model out of all three, and increased to 96% accuracy after using optimal values.

References

Alam, Mahbubul (12 June 2020). *Dealing with missing data in data science projects*. Towards Data Science. https://towardsdatascience.com/dealing-with-missing-data-in-data-science-projects-e8ac7a4efdff

Ayres de Campos et al. (2000) SisPorto 2.0 *A Program for Automated Analysis of Cardiotocograms*. J Matern Fetal Med 5:311-318 (link). https://www.kaggle.com/andrewmvd/fetal-health-classification

Aznar, Pablo (2020). Decision Trees: Gini vs Entropy. QuantDare. https://quantdare.com/decision-trees-gini-vs-entropy/

Centers for Disease Control and Prevention. 2015. *The Challenge of Fetal Mortality*. https://www.cdc.gov/nchs/products/databriefs/db16.htm

Cravit, Rachel (2021). *What is a Decision Tree and How to Make One [Templates + Examples]*. Vengage. https://venngage.com/blog/what-is-a-decision-tree/

DeSisto, Carla; Martin, Joyce; Gregory, Elizabeth; Hoyert, Donna; Valenzuela, Claudia (2021). *Fetal Deaths*. Centers for Disease Control and Prevention. https://www.cdc.gov/nchs/data/nvss/webinar/fetal-death-webinar.pdf

References Continued…

Dildy, Gary (2006). *Management of prolonged decelerations*. OBG Management. Pg 30-33. https://cdn.mdedge.com/files/s3fs-public/Document/September-2017/1811OBGM_ArticlE2.pdf

Gross, J., & Groß, J. (2003). *Linear regression* (pp. 33-36 Vol. 175). Springer Science & Business Media. https://books.google.com/books?hl=en&lr=&id=IvAw_1MTASsC&oi=fnd&pg=PR1&dq=linear+regression&ots=GtgL6rawvv&sig=do4Oi2WG5gEDoPjyQE5avnPgvM0#v=onepage&q=linear%20regression&f=false

Elite Data Science (2019). *How to Handle Imbalanced Classes in Machine Learning*. Elite Data Science. https://elitedatascience.com/imbalanced-classes

Oshiro T.M., Perez P.S., Baranauskas J.A. (2012) *How Many Trees in a Random Forest?*. In: Perner P. (eds) Machine Learning and Data Mining in Pattern Recognition. MLDM 2012. Lecture Notes in Computer Science, vol 7376. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-31537-4_13

References Continued…

Myles, A. J., Feudale, R. N., Liu, Y., Woody, N. A., & Brown, S. D. (2004). An introduction to decision tree modeling. *Journal of Chemometrics: A Journal of the Chemometrics Society*, 18(6), 275-285. https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/abs/10.1002/cem.873

ProjectPro (2021). *How to deal with imbalance classes with upsampling in Python?* ProjectPro. https://www.projectpro.io/recipes/deal-with-imbalance-classes-with-upsampling-in-python

Sweha, Amir (1999). *Interpretation of the Electronic Fetal Heart Rate During Labor*. American Family Physician. https://www.aafp.org/afp/1999/0501/p2487.html

UpToData (2021). *NICHD definitions of FHR characteristics and patterns*. UpToData. https://www.uptodate.com/contents/image?imageKey=OBGYN%2F65859

Wilson, Debrea (2018). *Fetal Heart Monitoring: What's Normal, What's Not?* Healthline. https://www.healthline.com/health/pregnancy/abnormal-fetal-heart-tracings

Questions?