



BAN 5600

Advanced Big Data Computing and Programming

Fetal Health Classification

Students: Maryam Ahmadi
Ritu Shrestha

Instructor: Professor Hamid Ahady

**Graduate School of Management
Clark University**

March 24th, 2023

Table Of Contents

1. Problem Importance	3
2. Problem Definition	3
3. Literature Review	4
4. Dataset Description	5
3.1 Dataset Features	5
5. Methodology	6
6. Descriptive Study	6
7. Implementations and Analysis	13
8. Results	14
9. Limitations	15
10. Recommendation	15
11. Software	15
6. References	16

Fetal Health Classification

Project Proposal for BAN 5600- Advanced Big Data Computing and Programming

Team members: Maryam Ahmadi- Ritu Shrestha

Authors: Maryam Ahmadi- Ritu Shrestha

GitHub link: <https://github.com/writu6/FetalHealthClassification>

Dataset: <https://www.kaggle.com/datasets/andrewmvd/fetal-health-classification>

Abstract

1. Problem Importance

The United Nations' Sustainable Development Goals include objectives to reduce child mortality, which is viewed as a significant measure of human advancement. The UN has set a target for countries to end preventable deaths of newborns and children under 5 years by 2030, with a goal to lower under-5 mortality to 25 per 1,000 live births. Maternal mortality, which is responsible for 295,000 deaths during and after pregnancy and childbirth, is also a concern, especially in low-resource areas where 94% of these deaths occur. According to a report published by the Centers for Disease Control and Prevention's National Center for Health Statistics, the total fetal mortality rate in the U.S. was 5.68 deaths for every 1,000 live births and fetal deaths in 2021 measured among fetuses at 20 weeks of gestation or more.

To address these issues, Cardiotocography (CTGs) provide a simple and affordable means of evaluating fetal health and can aid healthcare professionals in preventing maternal and child mortality. CTG equipment utilizes ultrasound pulses to assess fetal heart rate (FHR), fetal movements, uterine contractions, and other factors.

2. Problem Definition

The main purpose of CTG is to monitor the well being of the fetus especially in the third trimester and to detect early signs of fetal distress. However, it is to be noted that the output of CTG is a graph which requires domain experts to understand and analyze. Hence, we propose to develop a machine learning based on features extracted by CTG outputs to have an alternative approach to the traditional method of reading CTG output graphs.

Research Questions:

- Existing models could be efficient as expected?
- Can prediction methods like classification techniques be efficient in fetal health prediction?

- What is the performance of each classification model? Which one is more accurate?
- What is the suitable manner for evaluating the accuracy rate resulting from the prediction method?

In this project, we aim to use machine learning algorithms to accurately classify fetal health into three categories: Normal, Suspect, and Pathological, using the features extracted from CTG exams. We used both black box and white box machine learning algorithms, including dense models, Random Forest, XGBoost, and others, to classify fetal health. We also used feature importance analysis to detect the most important factors. Dense models offer high accuracy but are not interpretable, while Random Forest and XGBoost models offer interpretability and allow healthcare professionals and mothers to understand the factors that contribute to fetal health. We conclude that developing an accurate classification model can help healthcare professionals identify potential issues early and take appropriate measures to prevent child and maternal mortality.

3. Literature Review

Cardiotocography (CTG) is a commonly used method for fetal monitoring, which has been found to be useful in reducing neonatal mortality rates (Lakshmi & Suhasini, 2020). CTG involves the measurement of fetal heart rate (FHR) and uterine contractions to assess fetal well-being. CTG output is displayed in the form of a graph that requires interpretation by a healthcare professional.

Recently, machine learning (ML) techniques have been applied to CTG data to develop prediction models for fetal health. Several studies have shown that ML algorithms can be used to analyze CTG data for predicting fetal distress (Goya et al. 2017; Xu et al., 2018). Goya et al. (2017) developed a deep neural network (DNN) model to predict fetal distress based on CTG features. The DNN model achieved an accuracy of 93.2%, which was higher than other ML models such as support vector machine (SVM) and random forest (RF).

Other studies have also shown promising results in using ML techniques for fetal health prediction. For example, Zhang et al. (2019) developed a prediction model for fetal distress based on CTG and other clinical features using a SVM algorithm. The model achieved an accuracy of 95.1%, which was higher than the accuracy of CTG alone.

In summary, ML techniques have shown promising results in predicting fetal distress based on CTG data. Several studies have developed ML models that have achieved high accuracy rates for fetal distress prediction. However, further research is needed to validate these models in different clinical settings and to compare their performance with that of traditional CTG interpretation by healthcare professionals. The evaluation of ML models for fetal health prediction can be done using different metrics such as sensitivity, specificity, and AUC-ROC.

4. Dataset Description

The dataset used in the research is the “Fetal Health Classification” from the Kaggle repository.

Data Set Characteristics:	Multivariate	Number of Instances:	2126
Attribute Characteristics:	Categorical, Integer, Double	Number of Attributes:	22
Associated Tasks:	Classification	Missing Values?	No

The classification goal is to find machine learning algorithms that are able to classify fetal health in three categories. Namely **Normal**, **Suspect** and **Pathological**.

3.1 Dataset Features

The dataset contains the following features:

baseline_value: Baseline Fetal Heart Rate (FHR)

accelerations: Number of accelerations per second

fetal_movement: Number of fetal movements per second

uterine_contractions: Number of uterine contractions per second

light_decelerations: Number of LDs per second

severe_decelerations: Number of SDs per second

prolongued_decelerations: Number of PDs per second

abnormal_short_term_variability: Percentage of time with abnormal short term variability

mean_value_of_short_term_variability: Mean value of short term variability

percentage_of_time_with_abnormal_long_term_variability: Percentage of time with abnormal long term variability

mean_value_of_long_term_variability: Mean value of long term variability record

histogram_width: Width of the histogram made using all values from a record

histogram_min: HHistogram minimum value

histogram_max: Histogram maximum value

histogram_number_of_peaks: Number of peaks in the exam histogram

histogram_number_of_zeroes: Number of zeroes in the exam histogram

histogram_mode: Hist mode

histogram_mean: Hist mean

histogram_median: Hist Median

histogram_variance: Hist variance

histogram_tendency: Histogram trend

fetal_health: Fetal health: 1 - Normal 2 - Suspect 3 - Pathological

5. Methodology

In this project, we aim to use machine learning algorithms to accurately classify fetal health into three categories: Normal, Suspect, and Pathological, using the features extracted from CTG exams. We will use both black box and white box machine learning algorithms, including dense models, Random Forest, XGBoost, and others, to classify fetal health. We also use feature importance analysis to detect the most important factors. Dense models offer high accuracy but are not interpretable, while Random Forest and XGBoost models offer interpretability and allow healthcare professionals and mothers to understand the factors that contribute to fetal health.

6. Descriptive Study

The dataset was then used in databricks for some visual analysis.

1. Count Observation

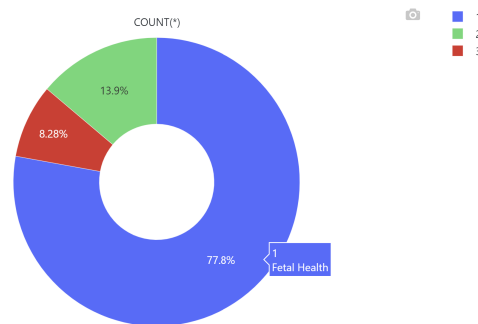


Fig: Count of Each Class

Here, we can see that the fetal health class 1 has the highest count among the 2126 observations, which is good as it denotes the fetuses are healthy and normal. However, this will create certain biases while developing a model. As a result we will need to resample the dataset.

2. Correlation Matrix

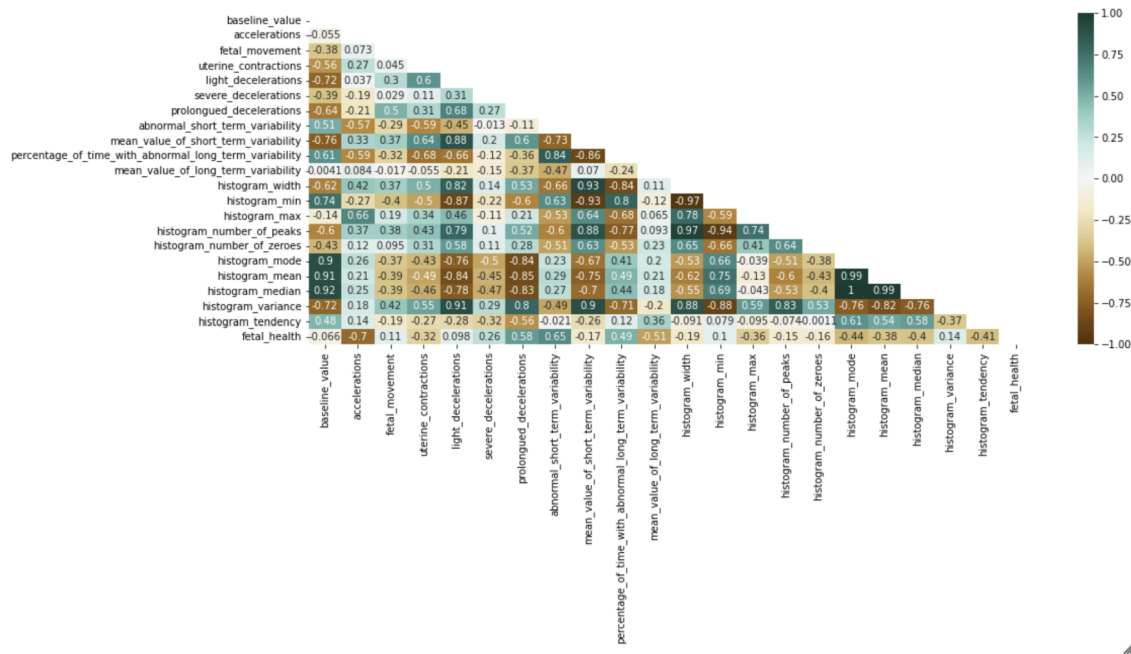
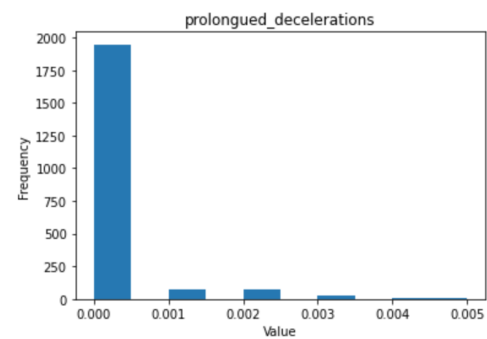
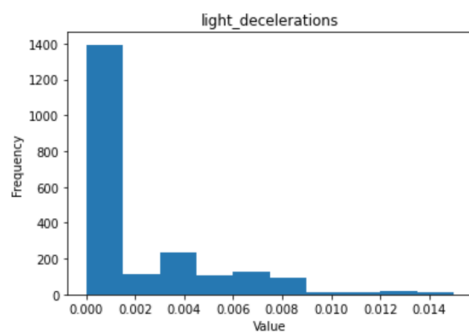
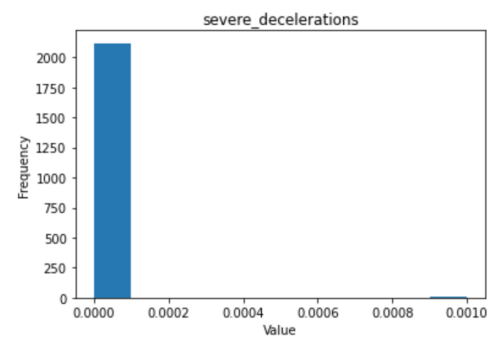
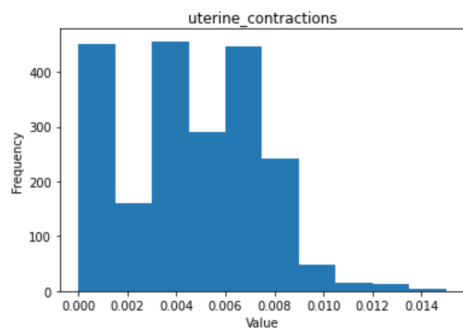
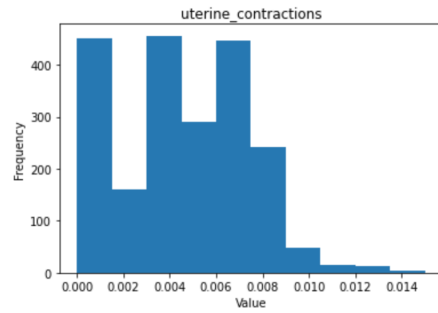
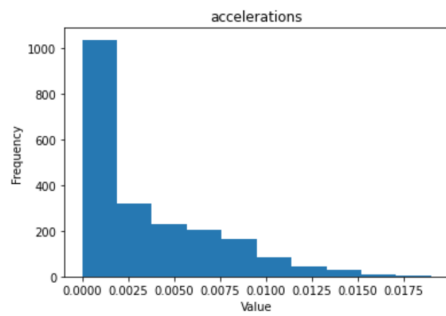
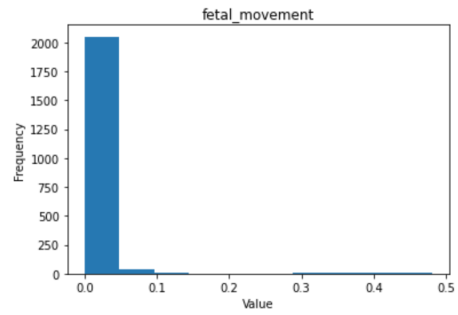
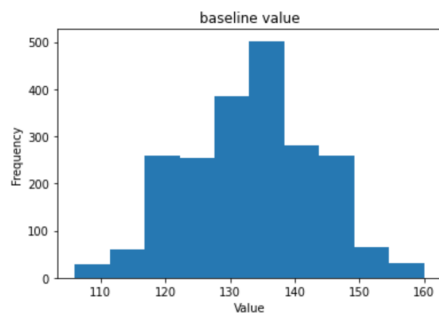


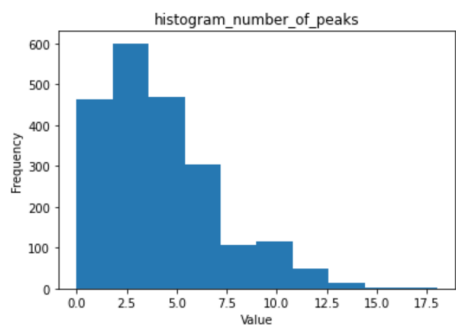
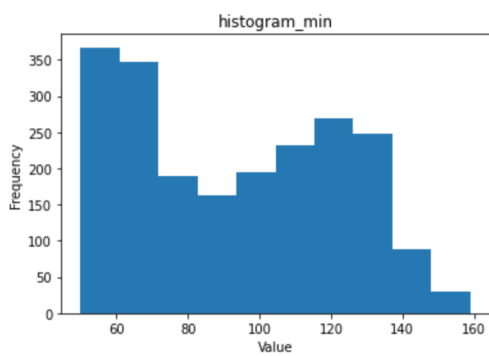
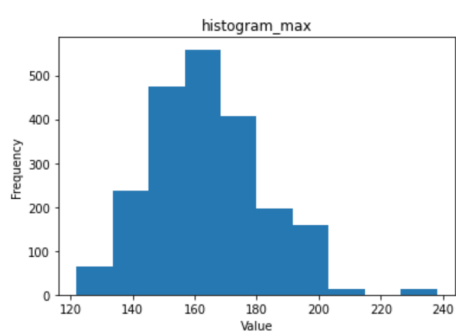
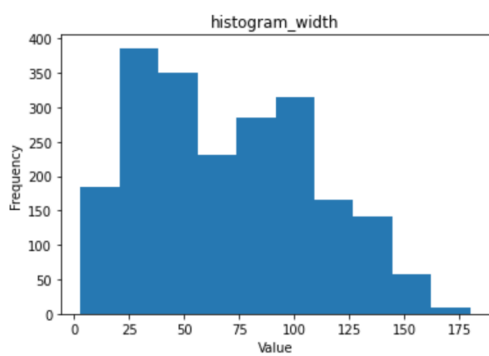
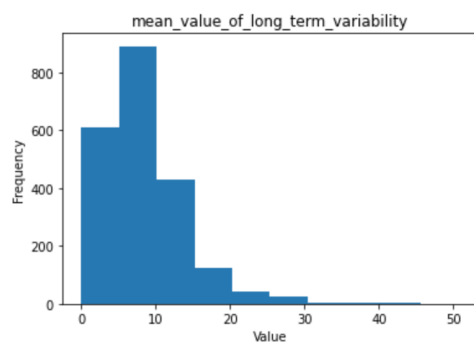
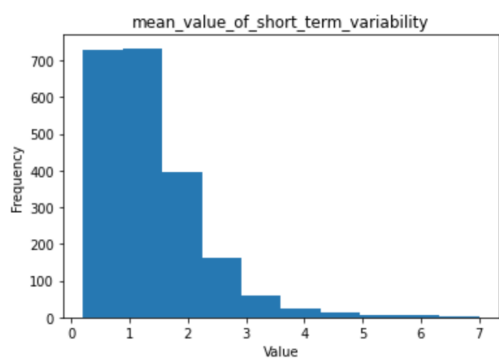
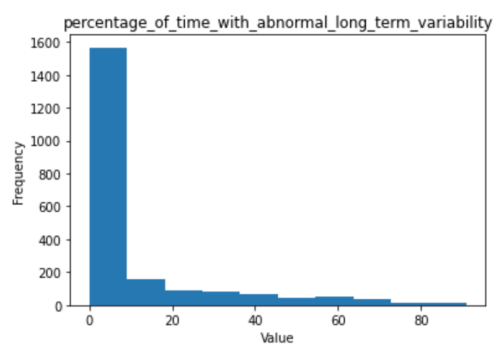
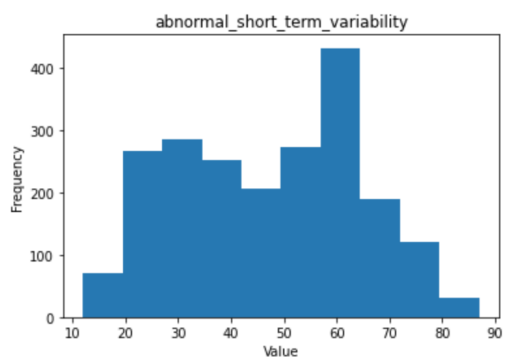
Fig: Correlation Matrix

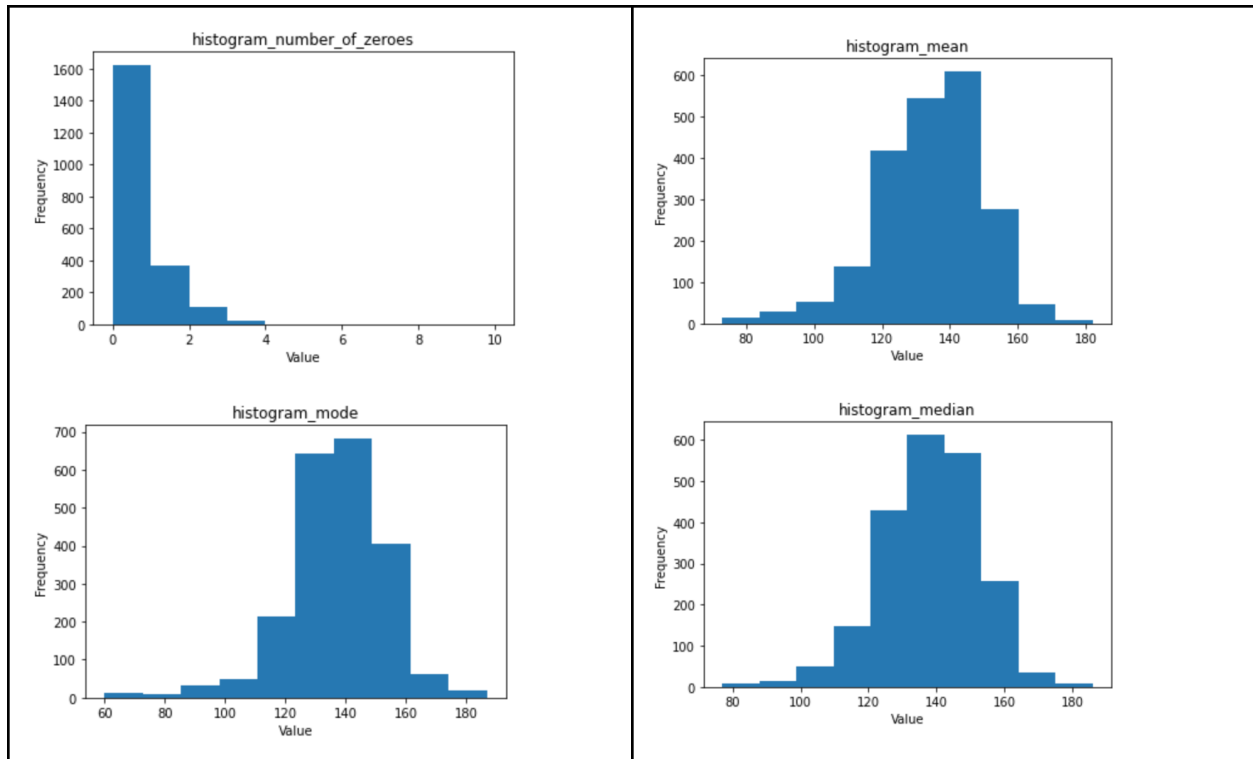
From the correlation matrix, we can see that the fetal health is heavily related to abnormal short term variability which specifies the percentage of time with abnormal short term variability.

3. Histogram

Histograms are graphical representations of the distribution of numerical data. They show the frequency or count of data values that fall within specified intervals or "bins" of the data range. Histograms can help to visualize the shape of the data distribution, including the center, spread, and skewness of the data.

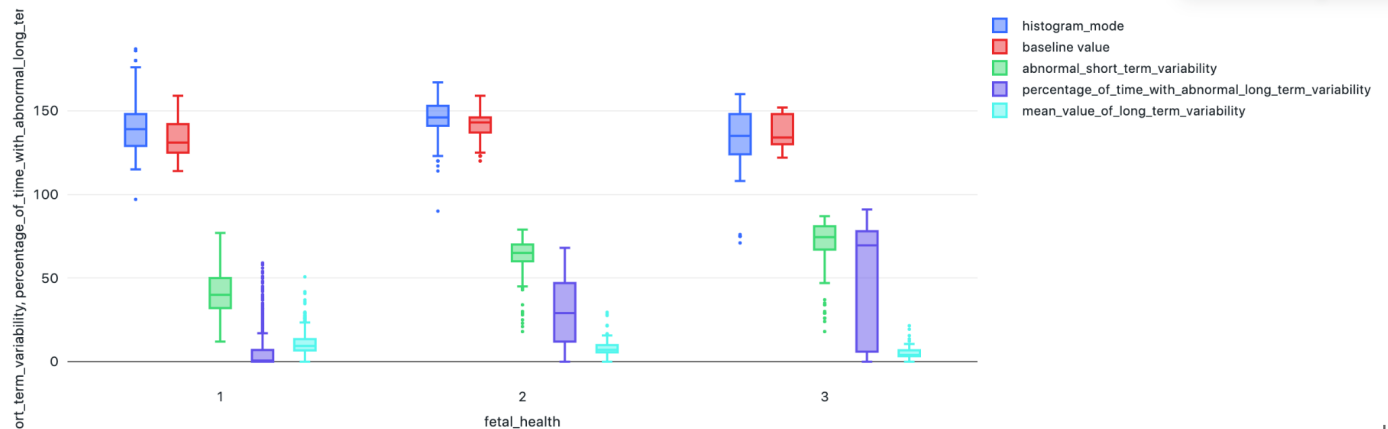






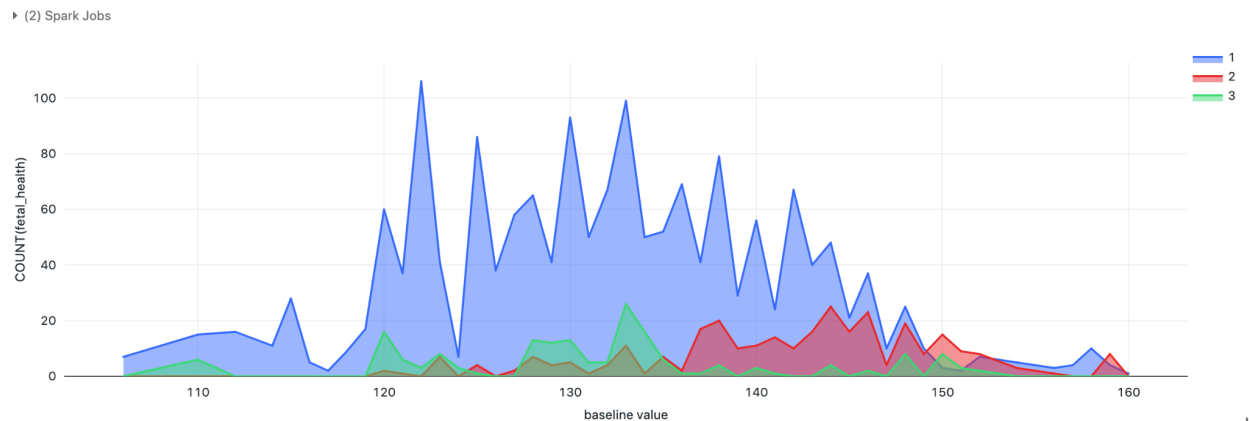
The above histogram shows the data distribution of each variable. While most of the variables follow a normal distribution, there are some skewness in some of them. For example, accelerations, light_deceleration, mean value of short term variability, mean value of long term variability, number of peaks, and number of zeroes are skewed to the right. Also it seems that there are outliers in some variables such as severe decelerations, and fetal movements. Which means that further consideration is required to detect and remove the outliers. Therefore, we will check the variables with box-plot in the next section.

4. Box-Plot for outlier detection

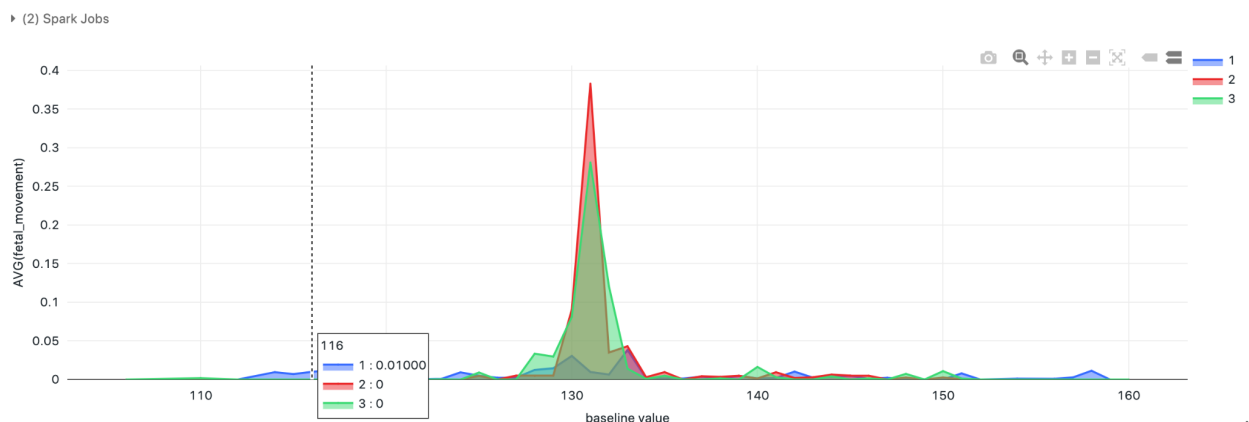


In box-plot, we can see there are outliers for each variables which is shown by the classes that is fetal health. It is important to consider outliers based on their class type since the dataset is unbalanced and the nature of the study tries to find the abnormal cases which are unhealthy fetals. Therefore, although some outliers are observable, we will not remove all of them. We will remove some outliers in the way that do not negatively affect the main goal of the project.

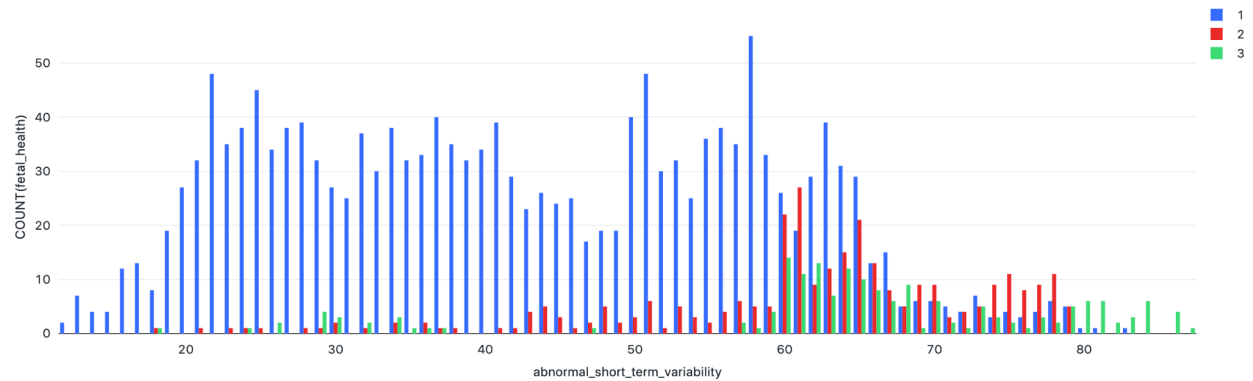
5. Visualization



The chart shows distribution of fetal health based on baseline value. Each color represents a class. From the chart, we understand that for healthy fetals which are shown with blue color, baseline value includes a wide range, while for the suspicious fetals, it is shifted to higher amount of baseline value with a concentration on 140 to 150.



This chart shows average fetal movement vs. baseline value for each class of fetal health. While fetal movement for healthy fetals is less than 0.05, it has higher values for suspicious and pathology class with a concentration on baseline value around 130. It says that unhealthy fetals abnormally show higher movements.



This bar chart show the number of fetal heath cases based on abnormal short term variability. While the value for healthy fetals has a range of 0 to 65, unhealthy cases has a shift to the right with higher values between 40 to 90.

6. Scatter Plot

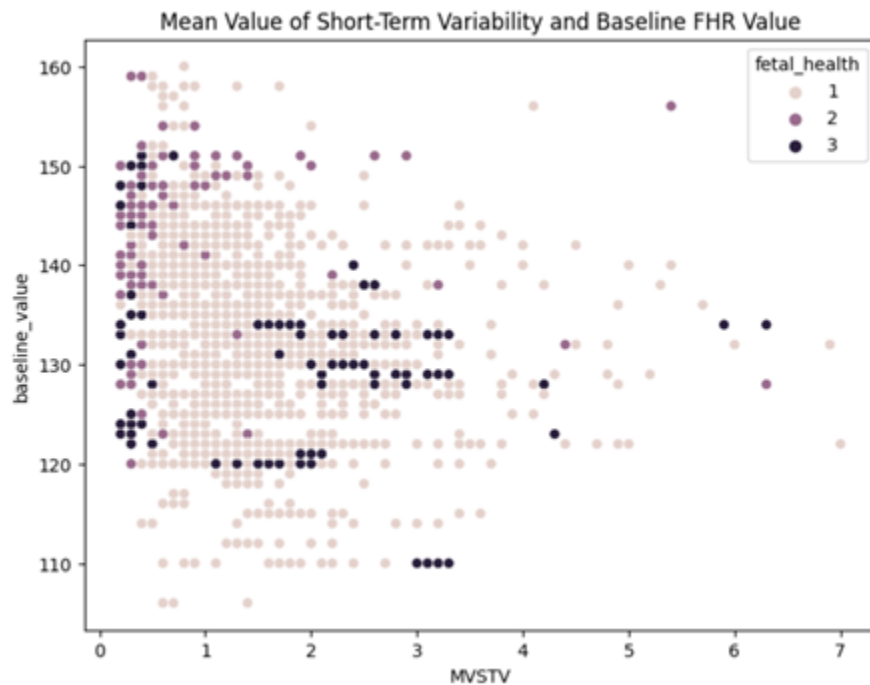


Fig: Scatter Plot of Mean Value of Short-Term Variability and Baseline Fetal Heart Rate Value

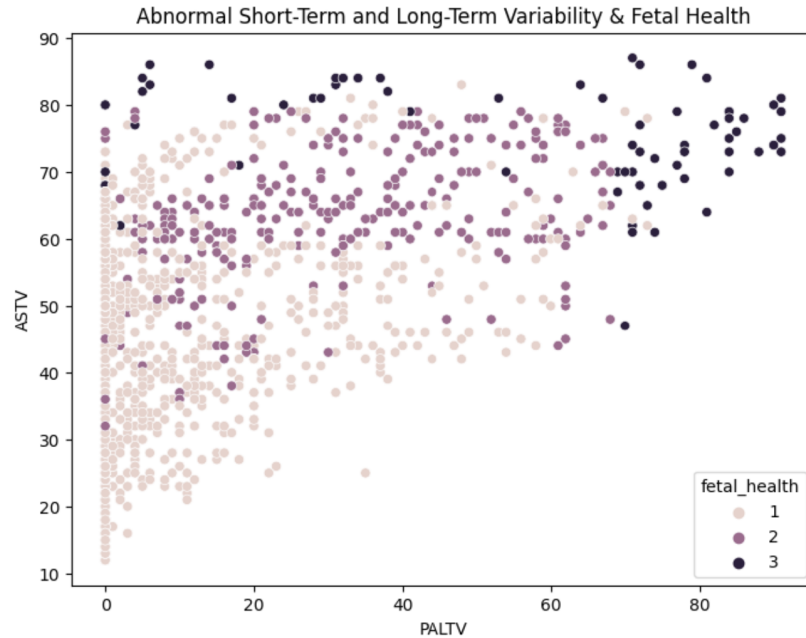


Fig: Scatter Plot of Percentage of Abnormal Short-Term and Percentage of Abnormal Long-Term of Fetal Health Status

7. Implementations and Analysis

We started our project by checking for missing values in our dataset and found that there were no missing values. Next, we analyzed the class distribution of our target variable "fetal_health" and found that it was imbalanced. Class 1 had 1655 observations, class 2 had 295 observations, and class 3 had 176 observations. To address this issue, we performed under-sampling and oversampling of the data to balance the classes for each model we trained.

We trained three models for our project: Decision Tree, Pruned Decision Tree, and Random Forest. We implemented these models using two different platforms: Databricks Pyspark environment and Amazon EC2.

In the Databricks Pyspark environment, we first implemented the Decision Tree model. We performed under-sampling of the data to balance the classes, created a feature vector, split the data into training and testing sets, trained the model, made predictions on the test data, and calculated the accuracy. Next, we implemented the Pruned Decision Tree model. We performed under-sampling of the data to balance the classes, created a feature vector, split the data into training and testing sets, trained the model with a maximum depth of 7 and minimum info gain of 0.1, made predictions on the test data, and calculated the accuracy. Here is the result:

Method/Measurement	Accuracy	Recall	F1	Precision
Under Sampling	0.8013	0.78	0.75	0.75
Over Sampling	0.8193	0.82	0.77	0.74
All Observations included	0.9047	0.9047	0.8998	0.9012
Pruned Tree	0.9064	0.9064	0.9034	0.9018

Lastly, we implemented the Random Forest model. We performed over-sampling of the data to balance the classes, created a feature vector, split the data into training and testing sets, trained the model with 10 trees, made predictions on the test data, and calculated the accuracy.

Method/Measurement	Accuracy	Recall	F1	Precision
Under Sampling	0.8078	0.81	0.76	0.74
Over Sampling	0.8489	0.85	0.79	0.74
All Observations included	0.9080	0.9080	0.9008	0.9027

In the Amazon EC2 environment, we followed a similar process to implement the models. After checking variables correlation, we removed two variables, histogram mode and histogram median, which were highly correlated with histogram mean, to see if it improves models performance or not. We performed under-sampling and oversampling of the data to balance the classes for each model. We trained the models with the same parameters as in the Databricks Pyspark environment and obtained better accuracies due to removing highly correlated variables.

8. Results

From the results, we can see that the pruned decision tree was the best-performing model with an accuracy of 90%. However, as much as accuracy is important, we have given the recall score as the primary basis for our selection of the model. Misclassification is a huge issue for our project as our project is trying to classify the health status of the fetus. Any wrong misclassification will have a severe impact on the future of the fetus and their mother. Hence, we have chosen recall score as our primary metrics for the selection of the model. As we can see in the Implementation and Analysis section, the recall score for the class 3 (Pathological) is 100% in the Pruned Decision Tree, meaning there was no misclassification of this status class at all, which is great. The other two classes, Class 1 (Normal) and Class 2(Suspect) are 96% and 94% respectively. Therefore, we conclude that Pruned Decision Tree is the best performing model for the fetal health classification.

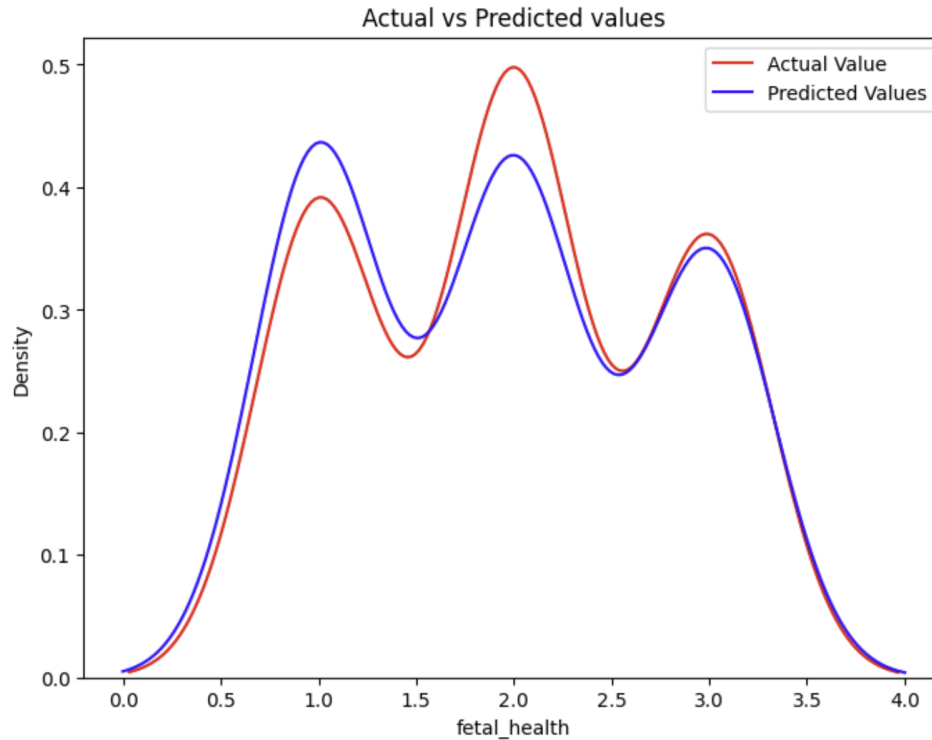


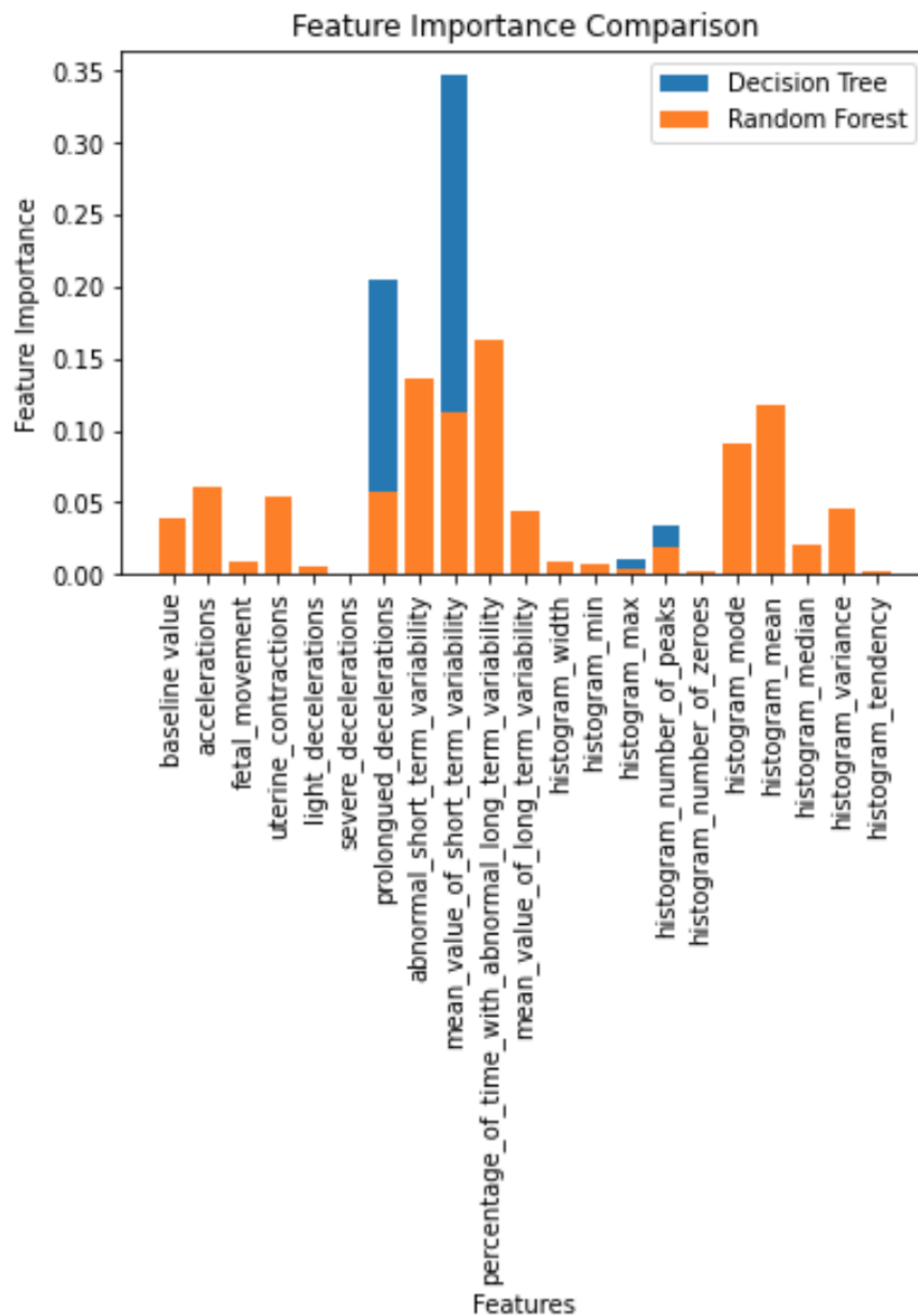
Fig: Actual vs Predicted values for Pruned Decision Tree

Another result that we concluded from our analysis was that Percentage of Abnormal Short-Term Variability is the key factor for determining the health status of the fetus, followed closely by histogram mean and prolonged decelerations. Therefore, we conclude that these three features are crucial to the fetal health and thus, more attention to these should be given.

Feature importance:

The feature importance values represent the relative importance of each feature in the model's prediction. The Decision Tree Feature Importance shows that the most important feature is the mean value of short-term variability (0.347), followed by prolonged decelerations (0.205), and abnormal short-term variability (0.134). The Random Forest Feature Importance shows a different ranking, where the most important feature is the mean value of short-term variability (0.113), followed by abnormal short-term variability (0.137) and histogram mean (0.118).

Both models show that the mean value of short-term variability and abnormal short-term variability are important features in predicting fetal health. The Decision Tree model places more emphasis on prolonged decelerations, while the Random Forest model emphasizes histogram mean and mode. It is important to note that some features, such as acceleration and severe decelerations, have little or no importance in either model.



9. Limitations

We had one main limitation for our project and that was the data were relatively low. We understand that the medical data is very private and is not released for public use. As a result, we had very little data to work with. This could result in overfitting the data, which is why the accuracy is so high. It is also important to note that the fetal health data are relatively new in comparison to other medical data such as radiology reports.

Black box models such as deep learning might perform better for our dataset but as we need interpretability and feature importance, we prefer decision tree and random forest models.

10. Recommendation

As we see that the feature Percentage of abnormal short-term variability is of high importance for our models, we would like to recommend the health care system to give extreme focus on this feature. This finding could be used in conjunction with advanced medical research, such as being done by the National Library of Medicine [5]. More study can be done on the features highlighted by the feature importance so as to be able to use these as the perfect guidelines for further improvement.

11. Software

In this project, we will use various software and platforms to explore the dataset and build classification models. To this aim, Pyspark, Jupyter notebook, Databricks, and AWS will be implemented for data preprocessing, visualization, and fitting machine-learning models on the dataset. For better collaboration between the team members, Github has been used for the collaboration.

Few List of Libraries: pyspark.sql, matplotlib, pyspark.ml.

Github link:<https://github.com/writu6/FetalHealthClassification>

6. References

- 1.Oded Koren, Michal Koren, Or Peretz, A procedure for anomaly detection and analysis, Engineering Applications of Artificial Intelligence, 10.1016/j.engappai.2022.105503, 117, (105503), (2023)
<https://geekymedics.com/how-to-read-a-ctg/>
<https://www.usnews.com/news/health-news/articles/2023-01-12/fetal-mortality-rate-holds-steady-despite-covid-pandemic>
- 2.Alfirevic Z, Devane D, Gyte GM, Cuthbert A. Continuous cardiotocography (CTG) as a form of electronic fetal monitoring (EFM) for fetal assessment during labour. Cochrane Database Syst Rev. 2017 Feb 3;2(2):CD006066. doi: 10.1002/14651858.CD006066.pub3. PMID: 28157275; PMCID: PMC6464257.
- 3.Guo R., Hou W., Dong Y., Yu Z., Stites J., and Weiner C. P. 2010. Brain injury caused by chronic fetal hypoxemia is mediated by inflammatory cascade activation. Reprod. Sci. 17:540–548. doi:10.1177/1933719110364061
- 4.Zhao Z, Zhang Y, Deng Y. A Comprehensive Feature Analysis of the Fetal Heart Rate Signal for the Intelligent Assessment of Fetal State. J Clin Med. 2018 Aug 20;7(8):223. doi: 10.3390/jcm7080223. PMID: 30127256; PMCID: PMC6111566.

5.<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4794822/>