Fetal Health Classification

Jeff Nicolas Olarte
A20421822
Avinash Sankar
A20525471

Abstract

The goal of this project is to examine data from cardiotocograms (CTGs) and to construct models that can identify the state of the fetus's health. Using these models can help healthcare professionals quickly make the decisions and therefore increase the likelihood of the fetus's survival.

Introduction

The United Nations as part of their Sustainable Development Goals are aiming to "end preventable deaths of newborns and children under 5 years of age"[1]. Most of these deaths occur in impoverished areas. According to a 2022 report tracking stillbirth statistics, there are over 5,000 babies that are stillborn after 28 weeks of pregnancy every day. This number increases to 1.9 million deaths per year[2]. What this report emphasizes is how preventable these deaths are. One of the causes of these preventable deaths that the report identifies is the lack of quality data for stillbirths. The aim of our project is to construct models from health readings of fetuses. These models are to be employed by healthcare professionals in order to quickly classify the health of a fetus and therefore take earlier action to prevent perinatal mortality.

Data Understanding

The data we are working with comes from a Kaggle dataset which was classified by professional obstetricians. The dataset consists of 2,126 fetal CTGs which are classified under 3 categories: Normal, Suspect, and Pathological. A CTG that is classified as 'Normal' describes a fetus that is in good health. A CTG that is classified as 'Suspect' describes a fetus who's health requires further investigation. A CTG that is classified as 'Pathological' describes a fetus that requires the administration of healthcare to prevent fetal mortality. The dataset consists of 21 features that we will utilize to construct our model. Refer to the kaggle dataset within the 'Sources' section for the description of each feature.

A cursory look at the dataset reveals that it is complete as referenced by figures 1, 2, and 3.

```
data.isna().sum()# here we are checking whether we have missing values

baseline value                                                      0
accelerations                                                       0
fetal_movement                                                      0
uterine_contractions                                                0
light_decelerations                                                 0
severe_decelerations                                                0
prolongued_decelerations                                            0
abnormal_short_term_variability                                     0
mean_value_of_short_term_variability                                0
percentage_of_time_with_abnormal_long_term_variability              0
mean_value_of_long_term_variability                                 0
histogram_width                                                     0
histogram_min                                                       0
histogram_max                                                       0
histogram_number_of_peaks                                           0
histogram_number_of_zeroes                                          0
histogram_mode                                                      0
histogram_mean                                                      0
histogram_median                                                    0
histogram_variance                                                  0
histogram_tendency                                                  0
fetal_health                                                        0
dtype: int64
```

Figure 1

Figure 1 displays the number of CTGs that have a value missing in regards to a specific feature. As displayed, there are no missing values. Figure 2 also confirms this fact by showing each CTG has no missing values for every feature.

	baseline value	accelerations	fetal_movement	uterine_contractions	light_decelerations	severe_decelerations
0	False	False	False	False	False	False
1	False	False	False	False	False	False
2	False	False	False	False	False	False
3	False	False	False	False	False	False
4	False	False	False	False	False	False
...	...	...	...	...	...	...
2121	False	False	False	False	False	False
2122	False	False	False	False	False	False
2123	False	False	False	False	False	False
2124	False	False	False	False	False	False
2125	False	False	False	False	False	False

Figure 2

Figure 3 shows that for every column in the dataset, there are 2126 values total. This means that every CTG has no missing values for all columns and we can therefore conclude that this dataset is complete.
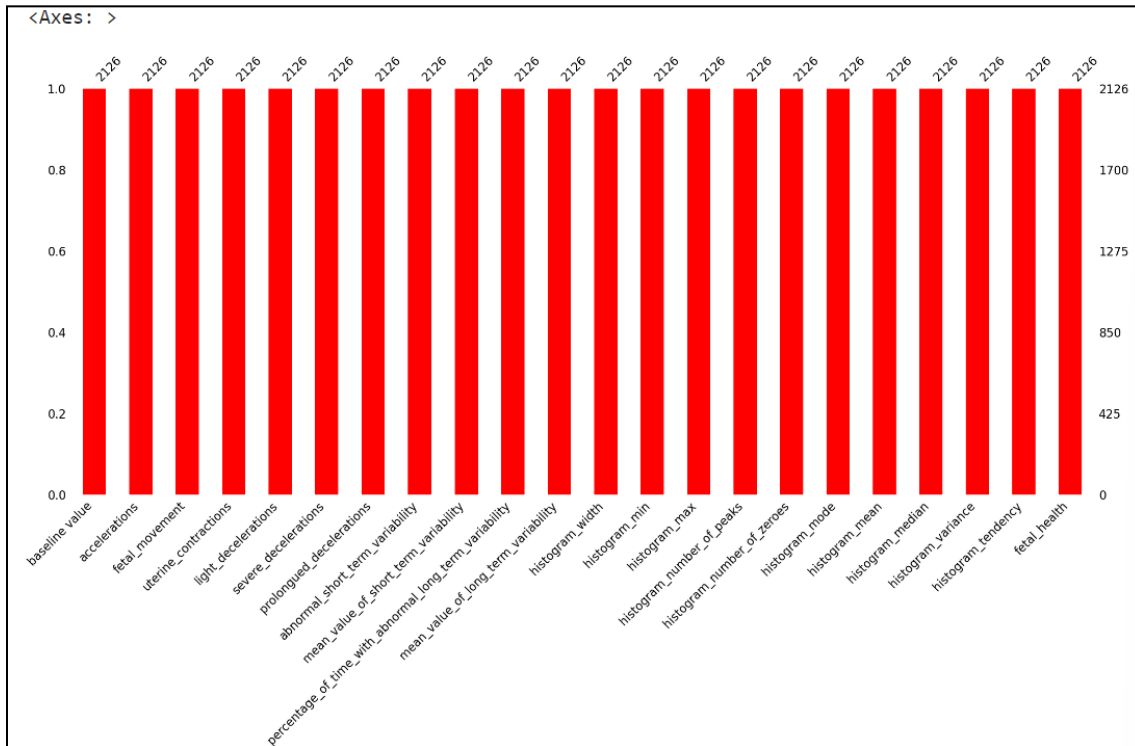
Figure 3

We must now examine what type of data populates the columns of the dataset. From Figure 4, we can see that all the features are floating point numbers. Keep in mind that the attribute we are trying to predict, fetal_health, which is represented by floating point numbers, is nominal in nature. That is to say that fetal_health is characterized by categories.

```
data.dtypes

baseline value                                                 float64
accelerations                                                  float64
fetal_movement                                                 float64
uterine_contractions                                           float64
light_decelerations                                            float64
severe_decelerations                                           float64
prolongued_decelerations                                       float64
abnormal_short_term_variability                                float64
mean_value_of_short_term_variability                           float64
percentage_of_time_with_abnormal_long_term_variability         float64
mean_value_of_long_term_variability                            float64
histogram_width                                                float64
histogram_min                                                  float64
histogram_max                                                  float64
histogram_number_of_peaks                                      float64
histogram_number_of_zeroes                                     float64
histogram_mode                                                 float64
histogram_mean                                                 float64
histogram_median                                               float64
histogram_variance                                             float64
histogram_tendency                                             float64
fetal_health                                                   float64
dtype: object
```

Figure 4

After we evaluated the dataset for completeness, we turned our attention to the statistics for each attribute in the dataset. Figure 5 shows the statistics of some of the attributes and gives us a better understanding as to how the value of these attributes are distributed.

```
data.describe()
```

	baseline value	accelerations	fetal_movement	uterine_contractions	light_decelerations
count	2126.000000	2126.000000	2126.000000	2126.000000	2126.000000
mean	133.303857	0.003178	0.009481	0.004366	0.001889
std	9.840844	0.003866	0.046666	0.002946	0.002960
min	106.000000	0.000000	0.000000	0.000000	0.000000
25%	126.000000	0.000000	0.000000	0.002000	0.000000
50%	133.000000	0.002000	0.000000	0.004000	0.000000
75%	140.000000	0.006000	0.003000	0.007000	0.003000
max	160.000000	0.019000	0.481000	0.015000	0.015000

Figure 5

Data Visualization

These graphs below represent the distribution of values for each feature. The horizontal axis represents the values of the attribute. The vertical axis represents how many CTGs have that specific value.
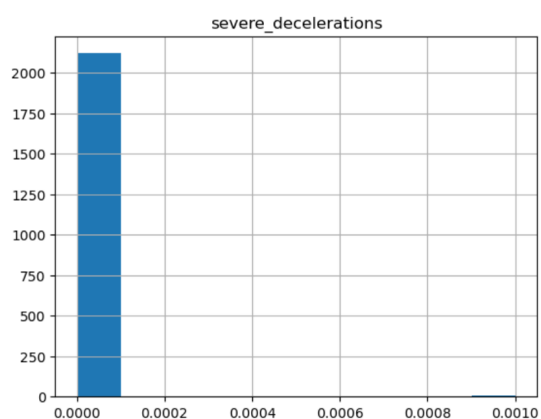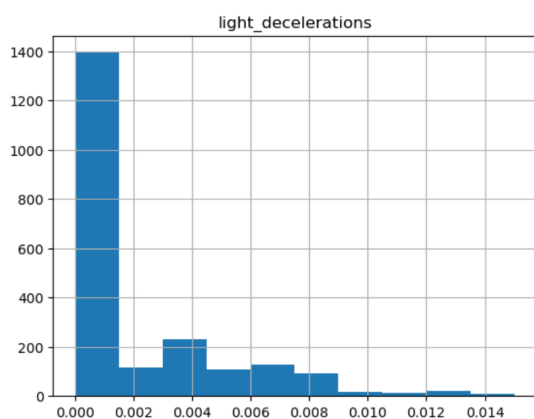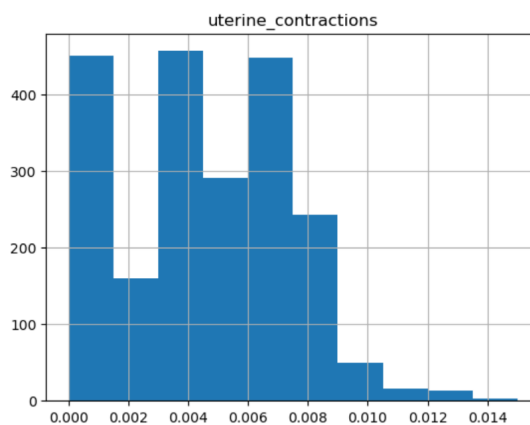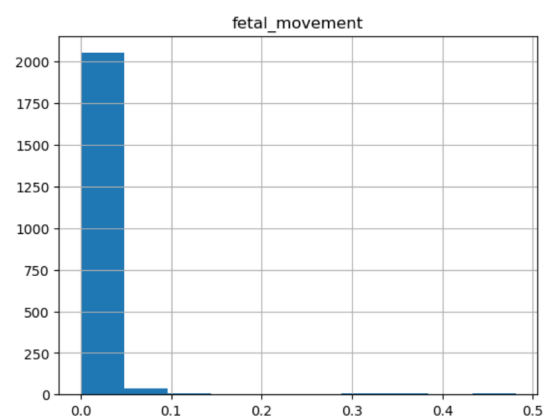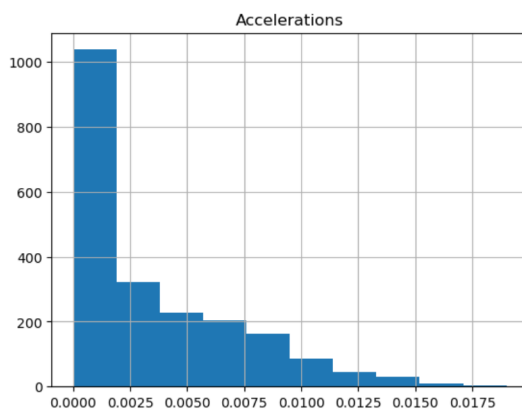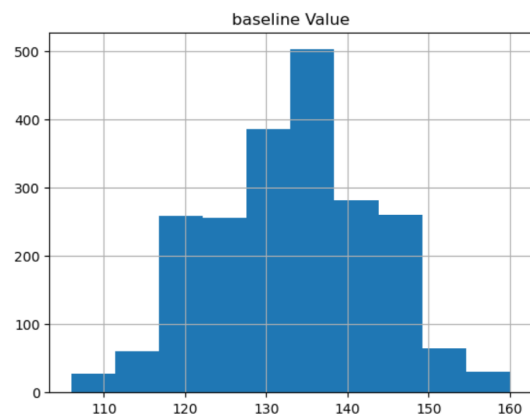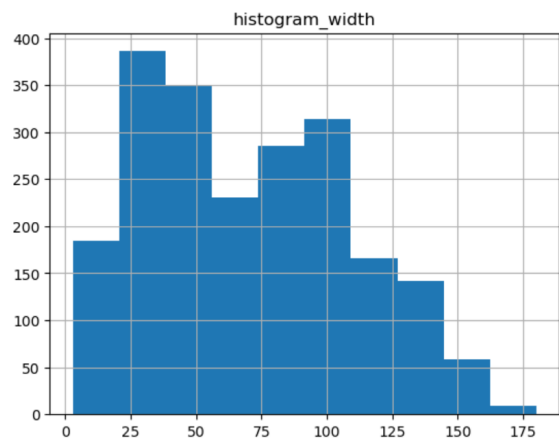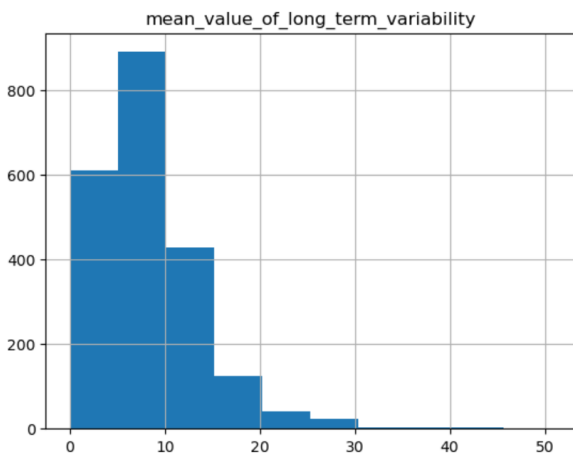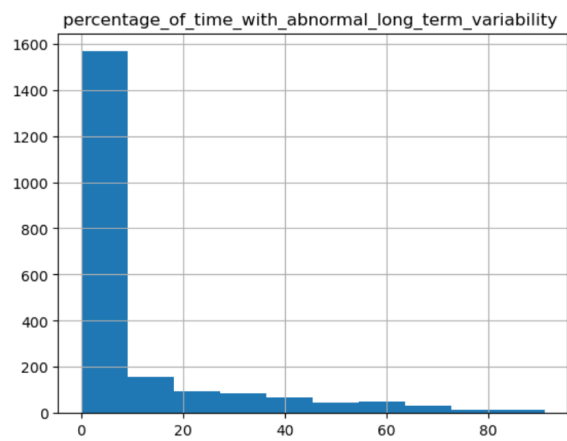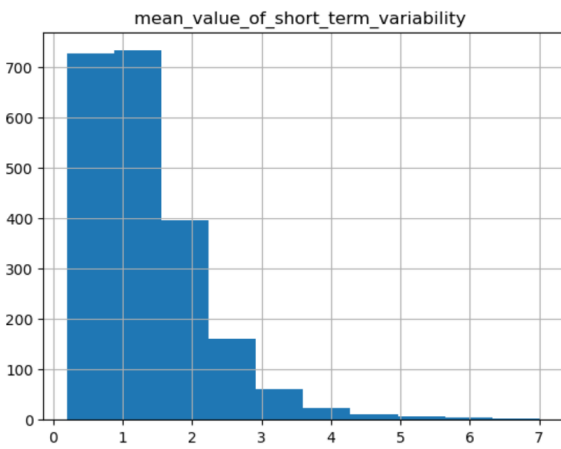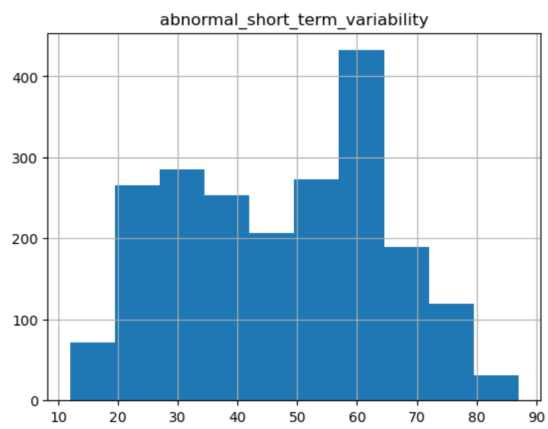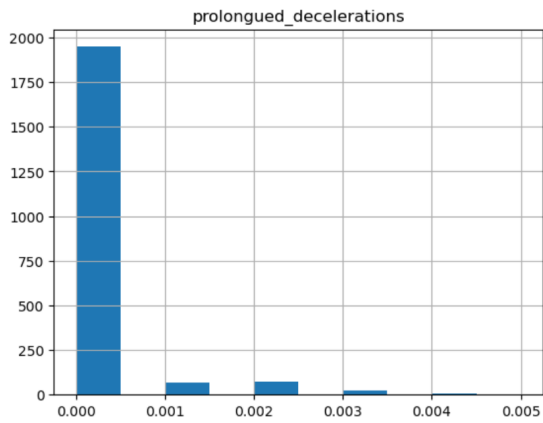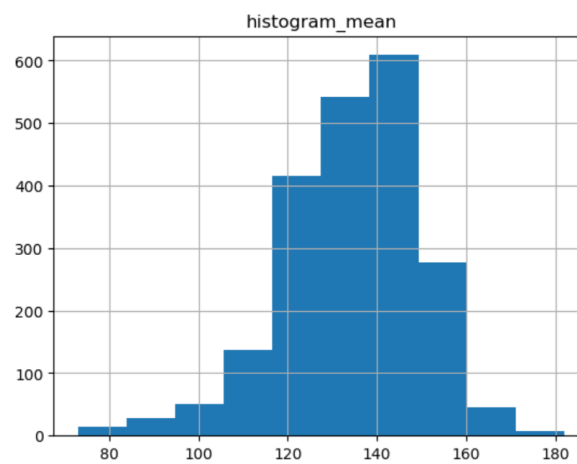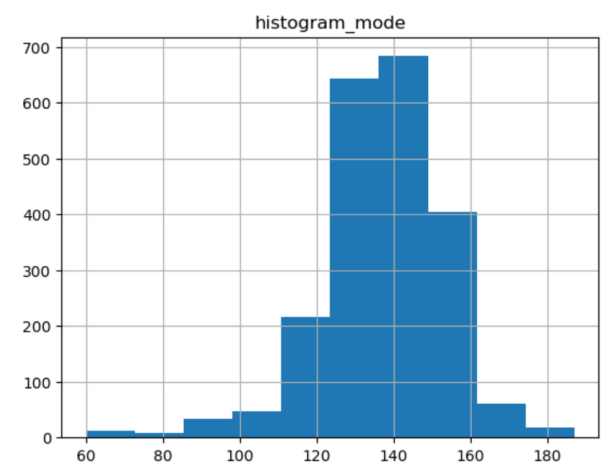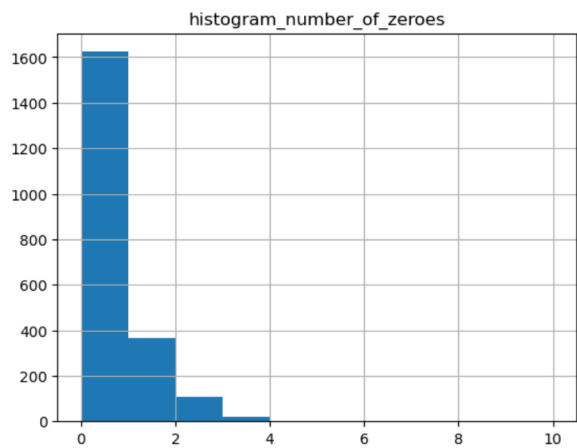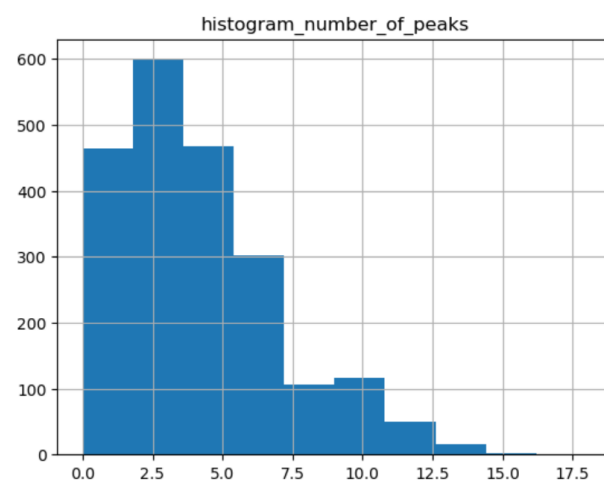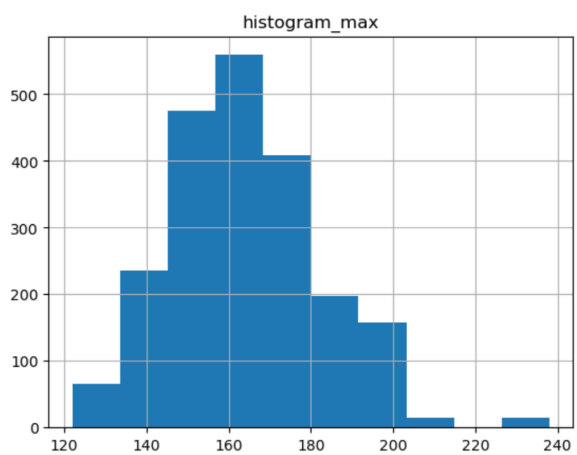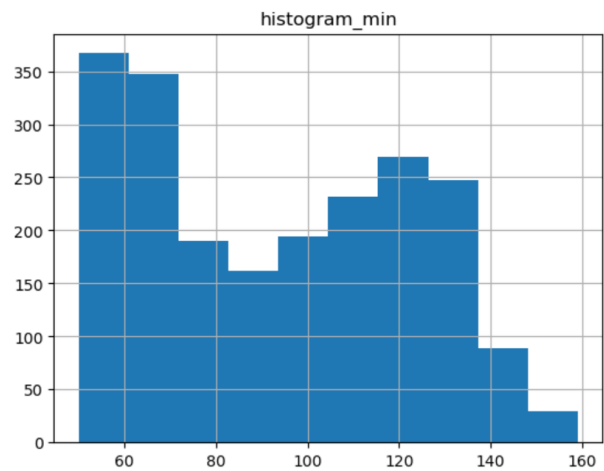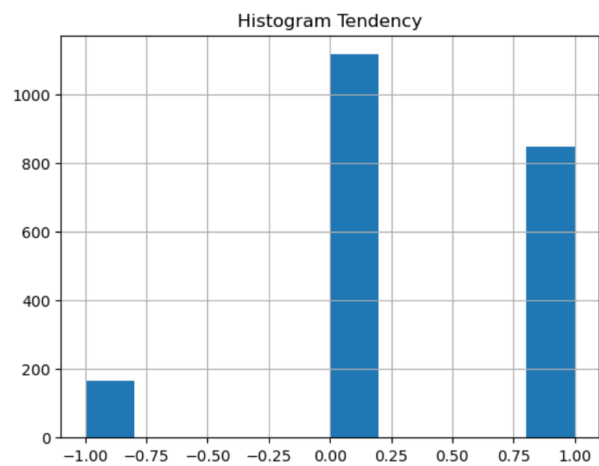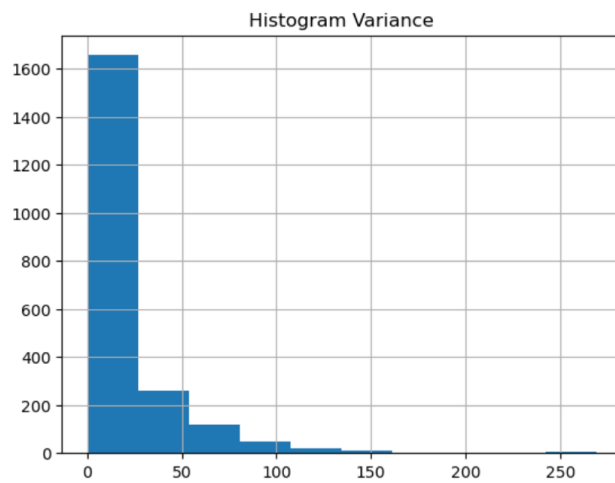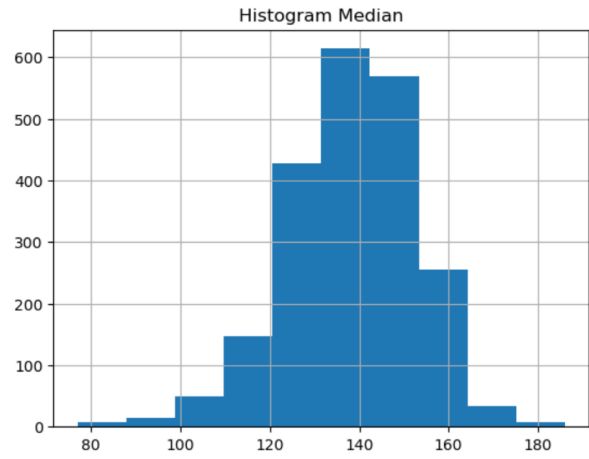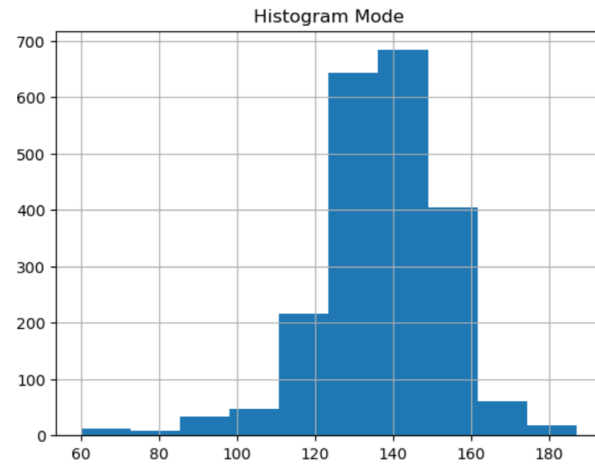
Another thing to note is that we visualized each feature and showed how which values for each feature belonged to a certain classification of 'fetal_health'

Feature Selection

After visualizing these distributions, we continued using boxplots to better visualize these distributions (Figure 6).
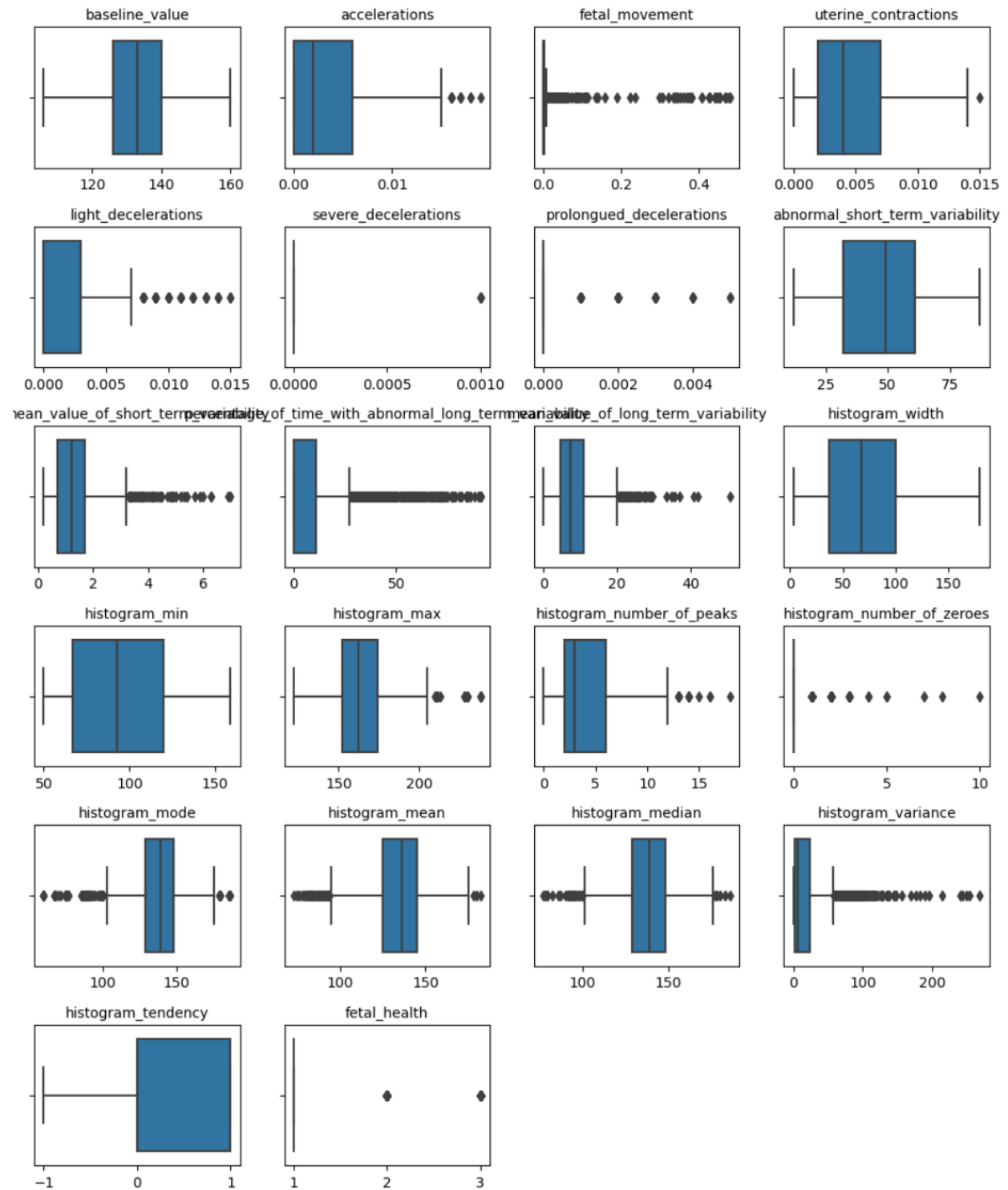
Figure 6

We observed several attributes having outliers. We however chose to keep outliers as that data was extracted by professionals as previously mentioned. Some factors we can attribute to the outliers are the conditions of the mother and any distress a fetus undergoes.

Correlations between dataset attributes were then visualized. Figure 7 describes the results.

Figure 7

We then listed which features were the most positively and negatively correlated with each other in Figure 8.

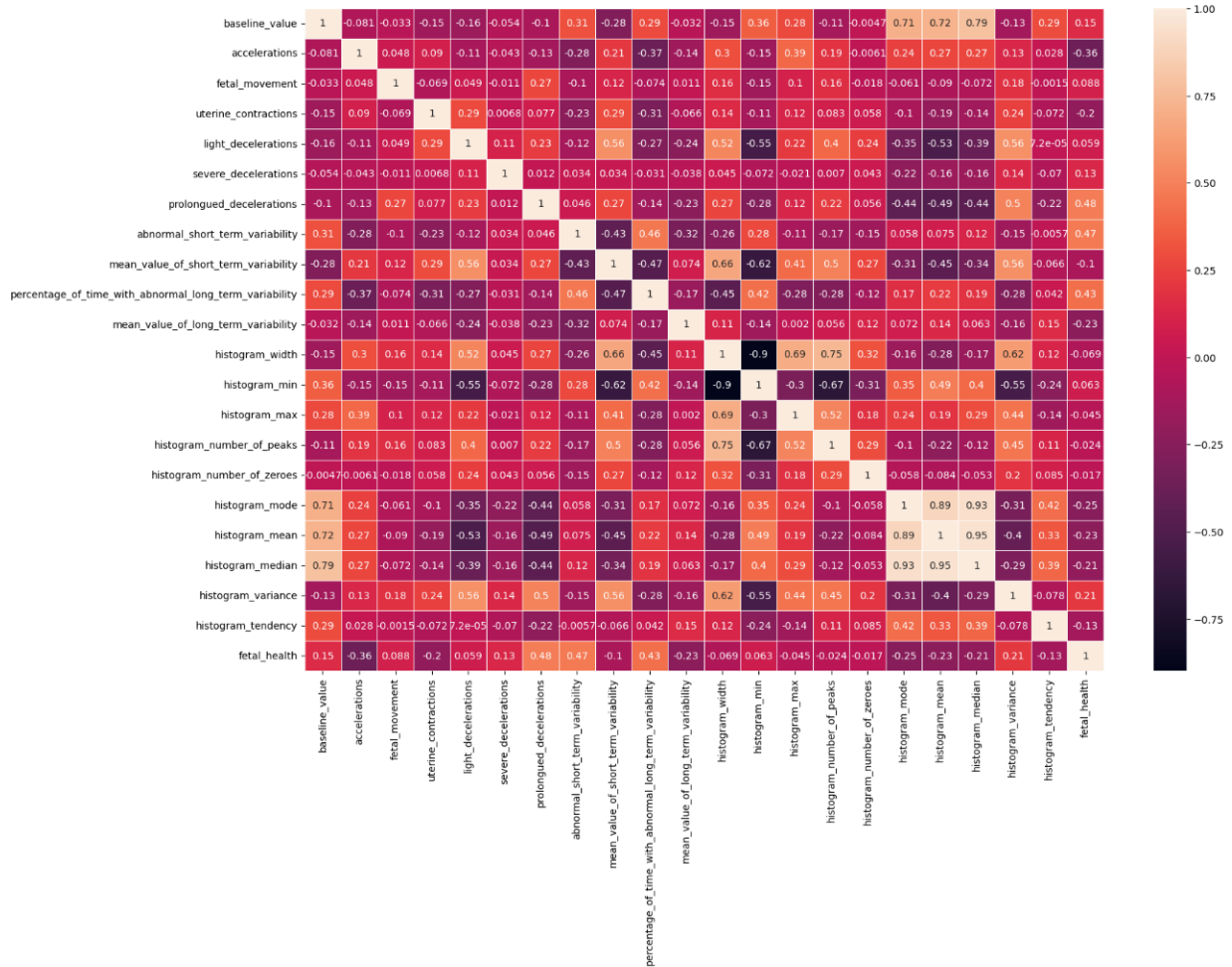|	level_0	level_1	corr_value
37	histogram_mean	histogram_median	0.948251
33	histogram_mode	histogram_median	0.933399
36	histogram_mean	histogram_mode	0.893412
2	baseline_value	histogram_median	0.789246
18	histogram_width	histogram_number_of_peaks	0.747071
34	histogram_mean	baseline_value	0.723121
31	histogram_mode	baseline_value	0.708993
25	histogram_max	histogram_width	0.690769
10	mean_value_of_short_term_variability	histogram_width	0.660847
19	histogram_width	histogram_variance	0.615884
7	light_decelerations	histogram_variance	0.564289
3	light_decelerations	mean_value_of_short_term_variability	0.562170
13	mean_value_of_short_term_variability	histogram_variance	0.555852
4	light_decelerations	histogram_width	0.520467
30	histogram_number_of_peaks	histogram_max	0.517652
8	prolongued_decelerations	histogram_variance	0.503301
27	histogram_number_of_peaks	mean_value_of_short_term_variability	0.501430
6	light_decelerations	histogram_mean	-0.527354
45	histogram_variance	histogram_min	-0.545091
20	histogram_min	light_decelerations	-0.553534
21	histogram_min	mean_value_of_short_term_variability	-0.622569
29	histogram_number_of_peaks	histogram_min	-0.670287
22	histogram_min	histogram_width	-0.898519

Figure 8

From this table we can determine that 'histogram_mean', 'histogram_mode', and 'histogram_median' are highly positively correlated with each other. We can also determine that 'histogram_min' and 'histogram_width' are highly negatively correlated with each other.

We now look at how the predictors correlate with the target attribute 'fetal_health' in Figure 9.

```
fetal_health                                               1.000000
prolongued_decelerations                                   0.484859
abnormal_short_term_variability                            0.471191
percentage_of_time_with_abnormal_long_term_variability     0.426146
histogram_variance                                         0.206630
baseline_value                                             0.148151
severe_decelerations                                       0.131934
fetal_movement                                             0.088010
histogram_min                                              0.063175
light_decelerations                                        0.058870
histogram_number_of_zeroes                                -0.016682
histogram_number_of_peaks                                 -0.023666
histogram_max                                             -0.045265
histogram_width                                           -0.068789
mean_value_of_short_term_variability                      -0.103382
histogram_tendency                                        -0.131976
uterine_contractions                                      -0.204894
histogram_median                                          -0.205033
mean_value_of_long_term_variability                       -0.226797
histogram_mean                                            -0.226985
histogram_mode                                            -0.250412
accelerations                                             -0.364066
Name: fetal_health, dtype: float64
```

Figure 9

From the table we can see that 'prolongued_decelerations', 'abnormal_short_term_variability', and 'percentage_of_time_with_abnormal_long_term_variability' are most positively correlated with 'fetal_health' with correlation values of 0.48, 0.47, 0.43 respectively. We considered predictors with a correlation value below 0.1 to be not correlated with 'fetal_health'.

Feature Removal

We then continued on to examine which predictors had the least distinct values and determined that 'severe_decelerations' and 'prolongued_decelerations' had the least distinct values. The reason as to why these 2 predictors are singled out as having the least distinct values is because we established a rule as to what predictors to exclude from the model. We determined that any predictor that has 10 or less distinct values should not be considered in the creation of our models. 'severe_decelerations' and 'prolongued_decelerations' failed to pass this criteria to be included in the model.

Dataset Preparation

After dropping the predictors previously mentioned in the previous section, we had the remaining predictors displayed in this table (Figure 10). The remaining predictors were then scaled to improve the model. Figure 11 shows how each predictor has been rescaled:
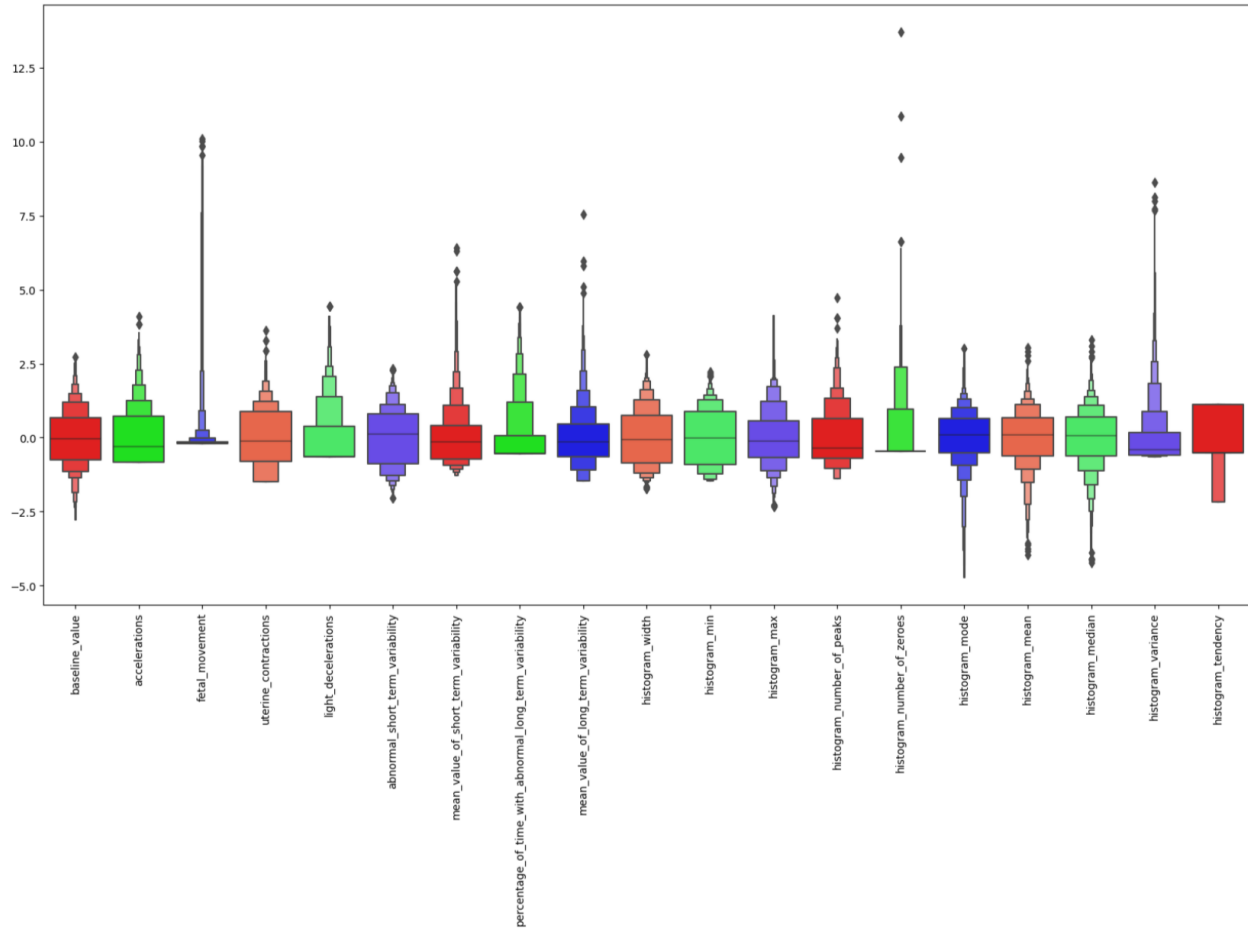

Figure 11

After dropping the above mentioned predictors from the dataset, we split the dataset into 2 datasets. Dataset 'X' holds the predictors while dataset 'Y' holds the target.

	count	mean	std	min	25%	50%	75%	max
baseline_value	2126.0	1.069490e-15	1.000235	-2.775197	-0.742373	-0.030884	0.680604	2.713428
accelerations	2126.0	-4.010589e-17	1.000235	-0.822388	-0.822388	-0.304881	0.730133	4.093929
fetal_movement	2126.0	-1.336863e-17	1.000235	-0.203210	-0.203210	-0.203210	-0.138908	10.106540
uterine_contractions	2126.0	-1.336863e-16	1.000235	-1.482465	-0.803434	-0.124404	0.894142	3.610264
light_decelerations	2126.0	-5.347452e-17	1.000235	-0.638438	-0.638438	-0.638438	0.375243	4.429965
abnormal_short_term_variability	2126.0	-7.352747e-17	1.000235	-2.035639	-0.872088	0.116930	0.815060	2.327675
mean_value_of_short_term_variability	2126.0	6.684315e-17	1.000235	-1.282833	-0.716603	-0.150373	0.415857	6.417893
entage_of_time_with_abnormal_long_term_variability	2126.0	-5.347452e-17	1.000235	-0.535361	-0.535361	-0.535361	0.062707	4.412293
mean_value_of_long_term_variability	2126.0	2.406354e-16	1.000235	-1.455081	-0.637583	-0.139975	0.464263	7.555172
histogram_width	2126.0	-3.007942e-17	1.000235	-1.731757	-0.858765	-0.075640	0.758838	2.812936
histogram_min	2126.0	-4.679021e-17	1.000235	-1.474609	-0.899376	-0.019608	0.893996	2.213648
histogram_max	2126.0	-1.203177e-16	1.000235	-2.342558	-0.670314	-0.112899	0.555999	4.123453
histogram_number_of_peaks	2126.0	-1.671079e-16	1.000235	-1.379664	-0.701397	-0.362263	0.655137	4.724738
histogram_number_of_zeroes	2126.0	2.757280e-17	1.000235	-0.458444	-0.458444	-0.458444	-0.458444	13.708003
histogram_mode	2126.0	1.069490e-16	1.000235	-4.729191	-0.516077	0.094519	0.644055	3.025381
histogram_mean	2126.0	-6.684315e-16	1.000235	-3.951945	-0.616458	0.089126	0.666422	3.039749
histogram_median	2126.0	2.673726e-16	1.000235	-4.223849	-0.628514	0.062897	0.685166	3.312527
histogram_variance	2126.0	-5.347452e-17	1.000235	-0.649208	-0.580173	-0.407586	0.179212	8.635997
histogram_tendency	2126.0	-1.069490e-16	1.000235	-2.162031	-0.524526	-0.524526	1.112980	1.112980

Figure 10

The datasets X and Y were then split even further into training and testing sets with the training set consisting of 70% of all the CTGs. With the training and testing set in place, several models were trained.

Random Forest

After the model had completed training, we examined its performance. The model had an accuracy of 0.94, precision of 0.94, recall of 0.94, and an F1 score of 0.94 as well. The confusion matrix for this model also showed promising results (Figure 11)
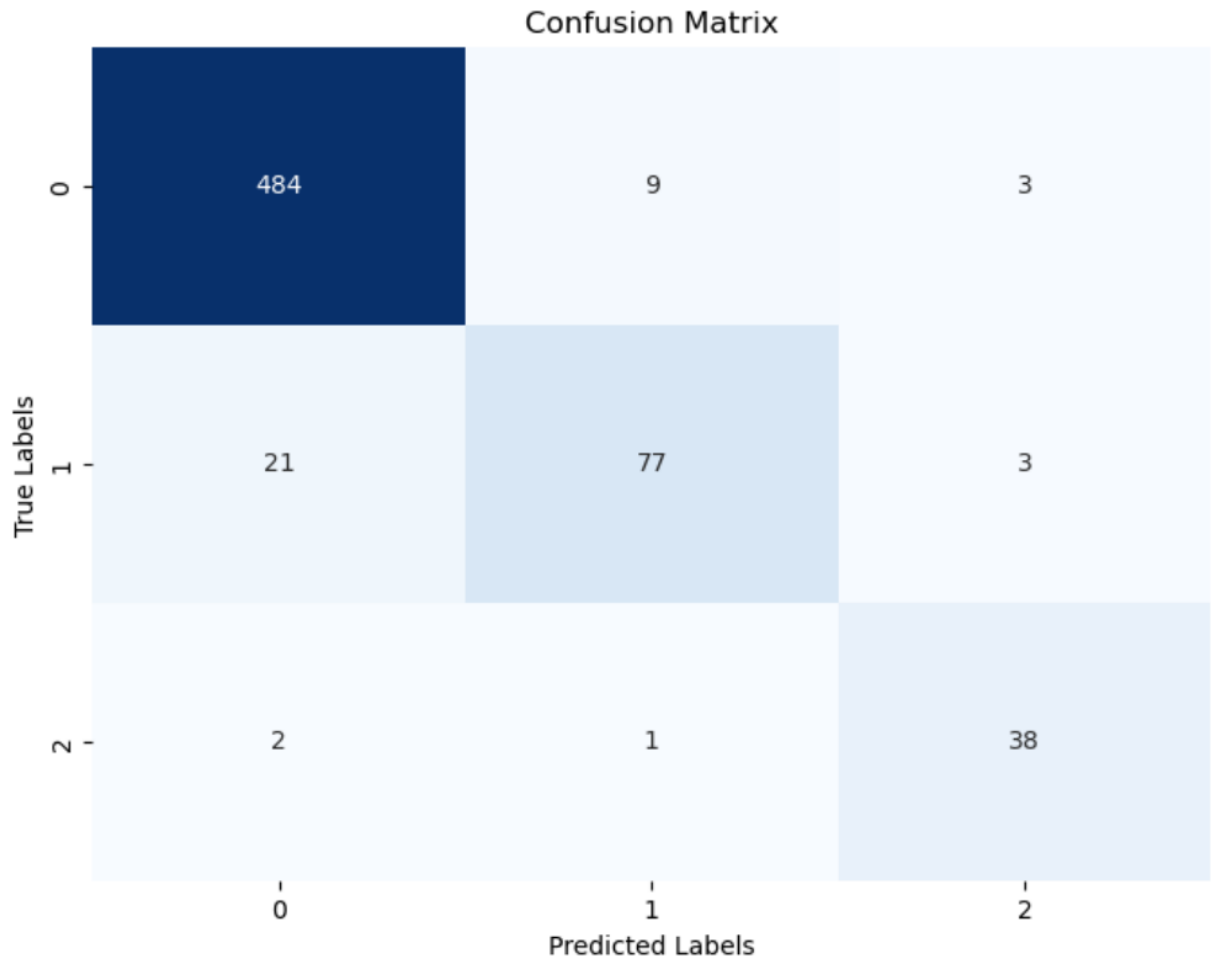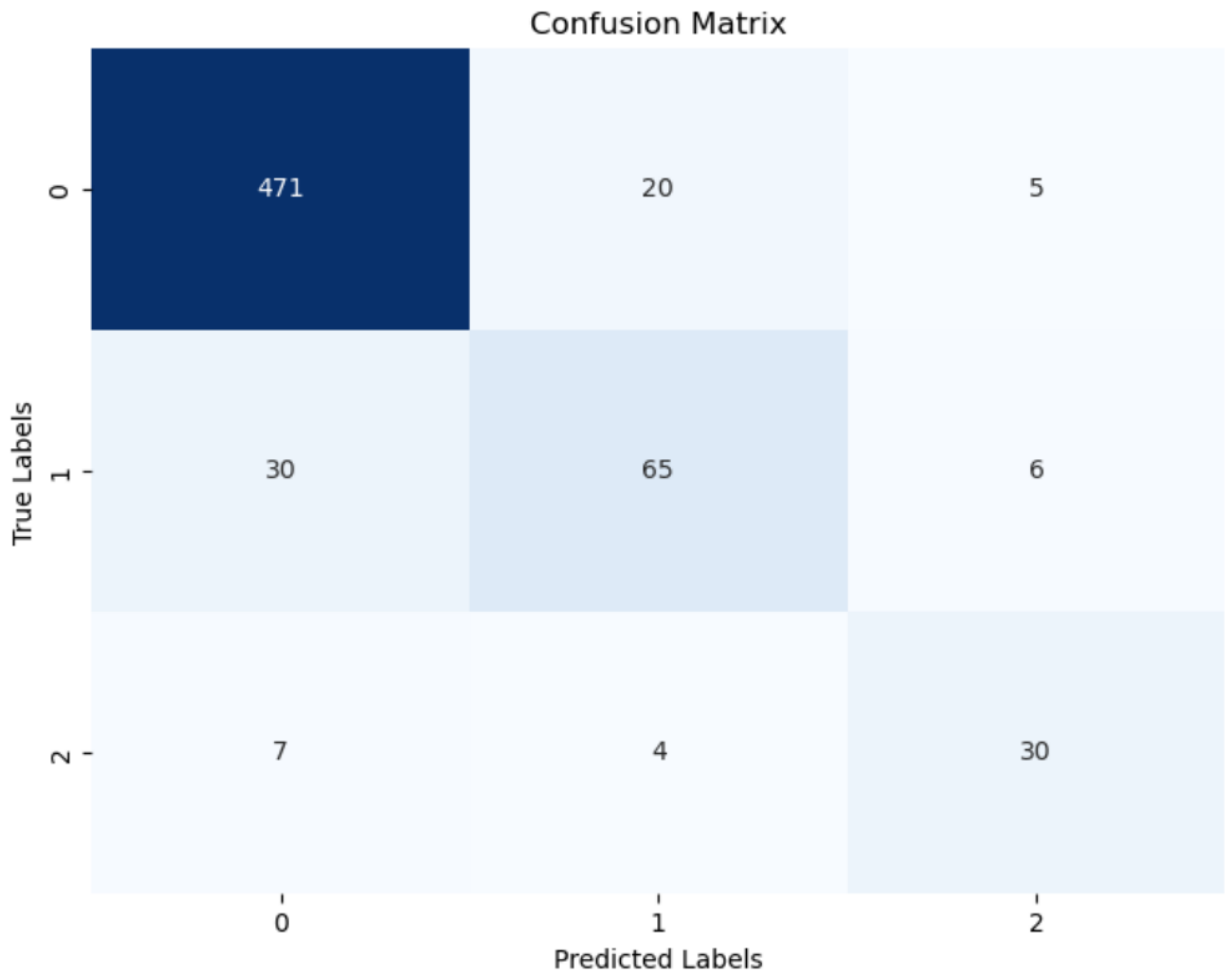
Confusion Matrix

|	0	1	2
0	484	9	3
1	21	77	3
2	2	1	38

True Labels / Predicted Labels

Figure 11

K-Nearest Neighbors

From trying multiple values for k, we determined that k=3 had good accuracy when compared with other values of k. Here is the confusion matrix for k=3.

Confusion Matrix

	0	1	2
0	471	20	5
1	30	65	6
2	7	4	30

True Labels (rows) / Predicted Labels (columns)

Logistic Regression

Our logistic regression model was reported to have an Accuracy of 0.876, Precision of 0.879, Recall of 0.876, and an F1 score of 0.876.
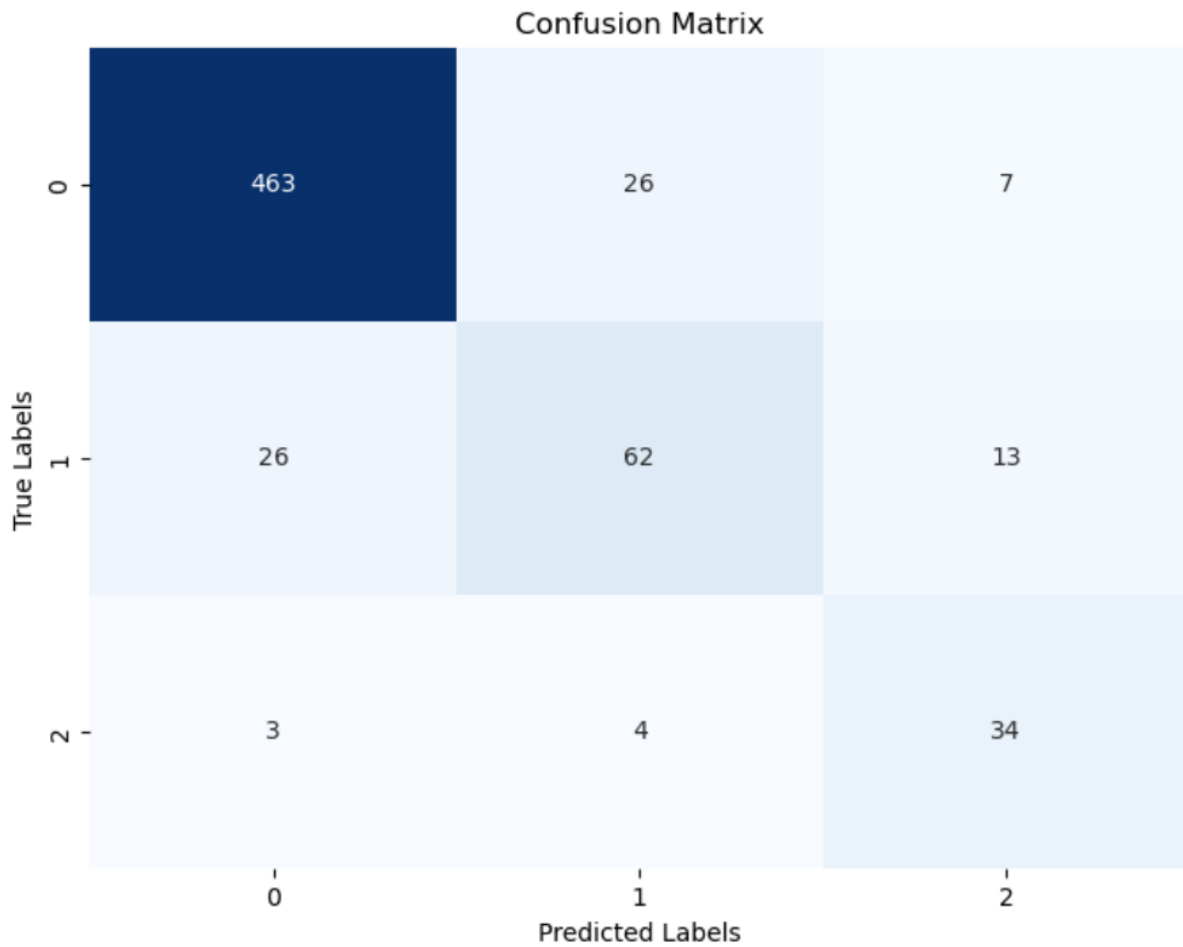

Confusion Matrix

Gradient Boosting Classifier

This model had an Accuracy of 0.95, Precision of 0.95, Recall of 0.95, and an F1-Score of 0.95. Below is the confusion matrix for the model's performance.

Confusion Matrix

True Labels / Predicted Labels	0	1	2
0	480	11	5
1	14	87	0
2	2	0	39
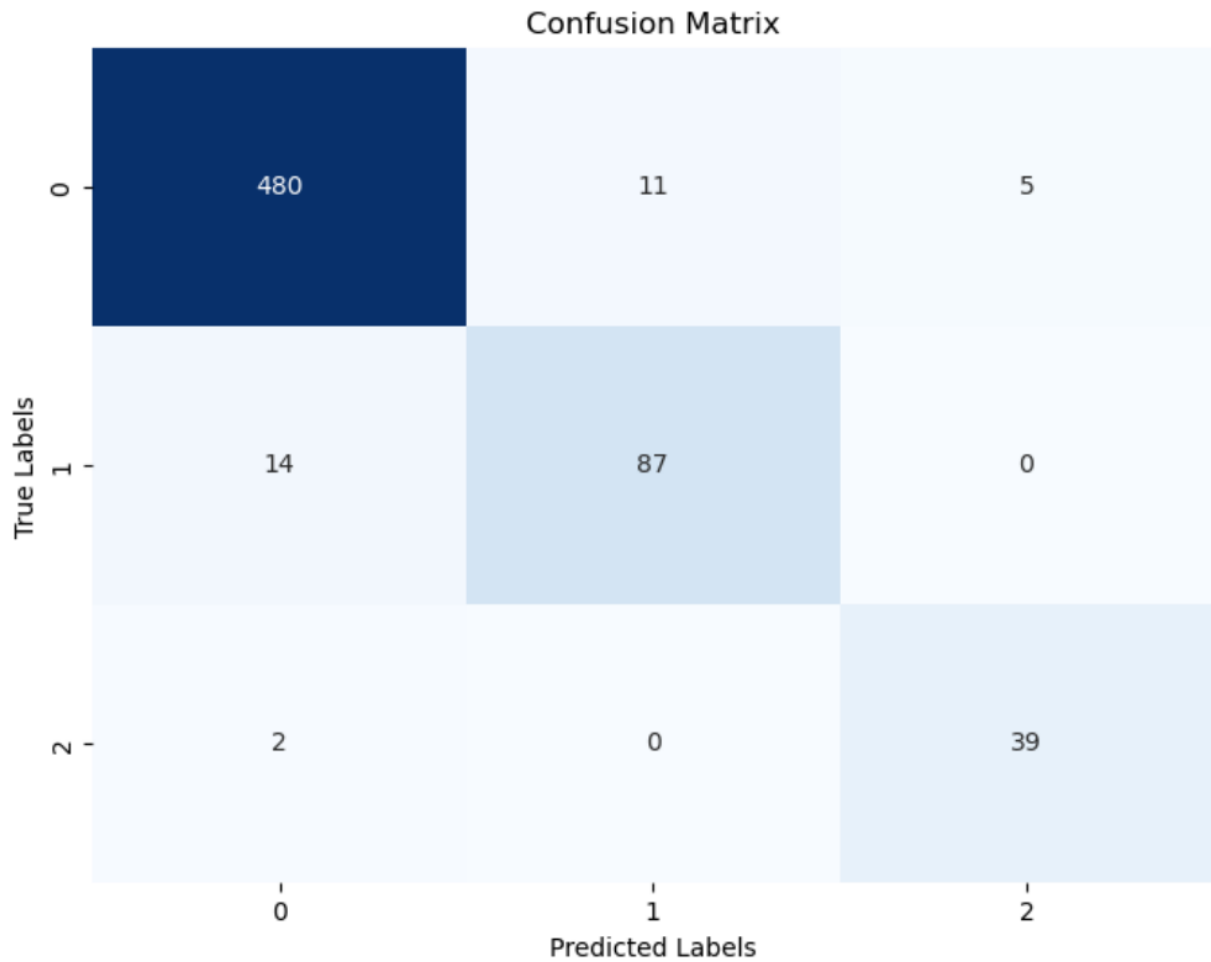
Decision Tree Classifier

This model had an Accuracy of 0.923, Precision of 0.923, Recall of 0.923, and an F1-Score of 0.923. The confusion matrix for the model's accuracy is below.
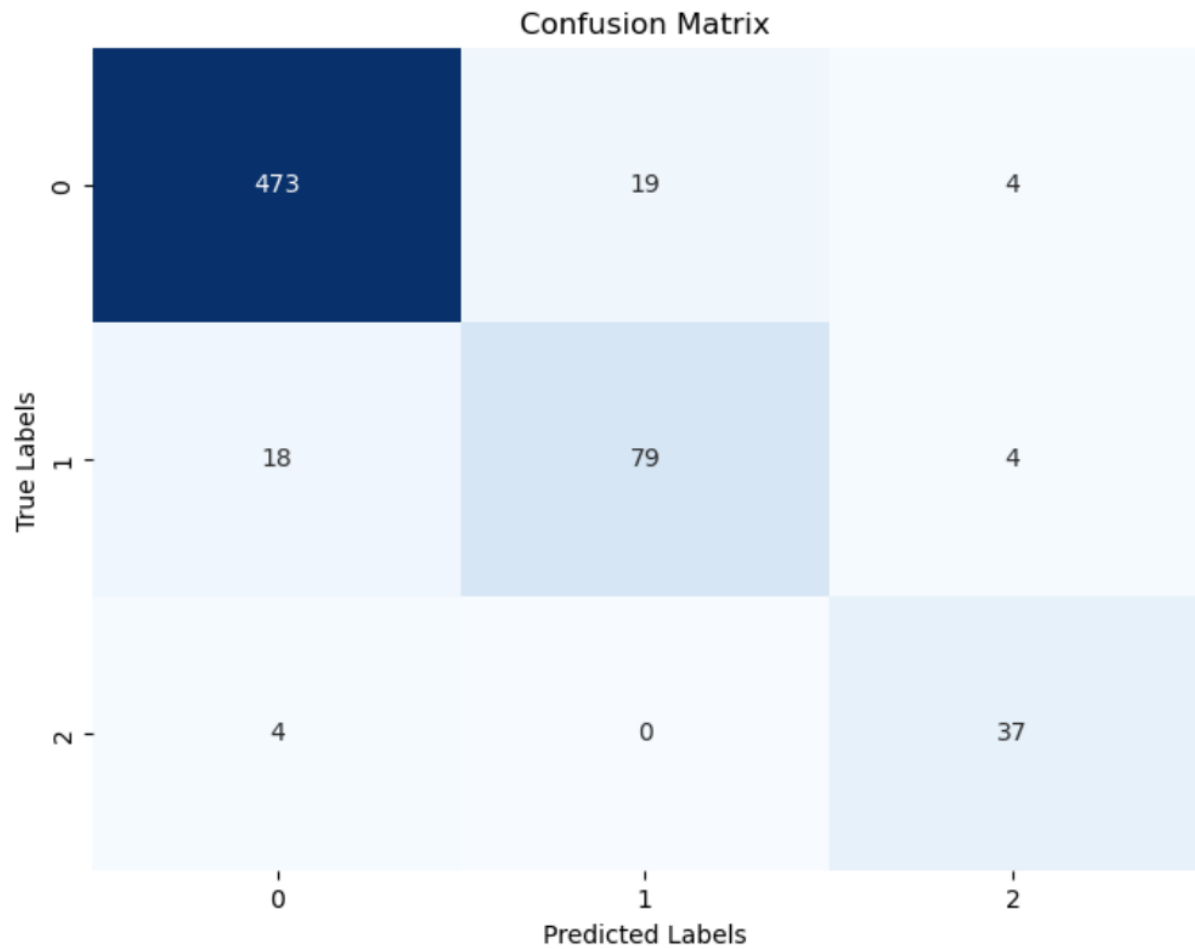
Conclusion

Based on the models trained and tested, we can conclude that Gradient Boosting Classifier had the best performance.

Sources

https://www.unwomen.org/en/news/in-focus/women-and-the-sdgs/sdg-3-good-health-well-being#:~:text=By%202030%2C%20end%20preventable%20deaths,25%20per%201%2C000%20live%20births. [1]

https://childmortality.org/wp-content/uploads/2023/03/UN-IGME-Stillbirth-Report-2022.pdf [2]

https://www.kaggle.com/datasets/andrewmvd/fetal-health-classification/ [3]

Link to github repository for project: https://github.com/JNTO/CS584-Class-Project