



Fetal Health Classification

KRITIKA GARG
UIN:662632356

BSTT 426: Healthcare analytics
Final Project Report

PROBLEM STATEMENT

Fetal and maternal mortality rate are key indicators of the progress of any society. This project aims to reduce the fetal and maternal mortality rate by examining fetal health using results from **Cardiotocogram (CTGs)** exams. CTGs evaluate fetal health based on various parameters like uterine contractions, fetal movement, decelerations etc.

A **multiclass classification model** that can predict fetal health based on these features beforehand will give will give the doctor **enough time to take necessary actions** and prevent death of newborns and their mothers.

Relevant prior work:

A retrospective population-based study was performed by Daly et al. on 524 women to study the diagnostic capacity of cardiotocography (CTG) as the primary investigator of the clinical scenario. It was concluded that CTG is reliable screening technique to indicate fetal wellbeing. Also, in the cases where initial CTG was abnormal, abnormal outcomes were more common.

Based on these results, it can be believed that an algorithm that predicts based on the results of CTG exams can help reduce fetal mortality rate.

Dataset Description:

The dataset extracts from Cardiotocograms (CTGs) exams performed to assess fetal health. It contains **2126 records of 21 features** which are finally classified into **three classes (fetal health)** by the obstetricians:

- Normal (Class 1)
- Suspect (Class 2)
- Pathological (Class 3)

The dataset contains the following features:

1. Baseline value
2. Accelerations
3. Fetal movement
4. Uterine contractions
5. Light decelerations
6. Severe decelerations
7. Prolonged decelerations
8. Abnormal short-term Variability
9. Mean value of short-term variability
10. Percentage of time with abnormal long-term variability
11. Mean value of long-term variability
12. Histogram width
13. Histogram min
14. Histogram max
15. Histogram number of peaks
16. Histogram number of zeroes
17. Histogram mode
18. Histogram mean
19. Histogram median
20. Histogram variance
21. Histogram tendency

This dataset was retrieved from Kaggle and the results were from a study namely “A Program for Automated Analysis of Cardiotocograms” performed by Ayres de Campos et al. in 2000.

Methodology:

The following steps were performed during the project:

1. **Data pre-processing and standardization:** Features were standardized using `sklearn.preprocessing.StandardScaler` such that they have zero mean and unit standard deviation.
2. **Exploratory Data Analysis (EDA):** The distribution and correlation among features were identified. Statistical description for the numerical variables was calculated. Data visualization was also performed.
3. **Predictive Model Building:** Predictive model was created using two classification algorithms: KNN and Random Forest. Among the two, the best was selected based on their ability to classify fetal health with lowest prediction error.
4. **Parameter tuning:** `GridSearchCV` was used to tune the parameters that will give the best possible score for each model.
5. **Evaluation of the classification results:** The performance of each model was evaluated through confusion matrix and classification report. Parameters such as accuracy, precision, recall were looked into. Finally, the best model was used to identify important features that influence fetal health.

Analysis of results:

There were no missing values in the dataset. Hence, after the standardization of the features, EDA was performed. The following table showed the statistical description of the features (*Fig. 1*). The results show that the feature “severe deceleration” has very low variance as compared to the other features. It can be inferred that this will not be useful for the future analysis and was thus dropped.

	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

Fig 1: Statistical description of numerical features

Furthermore, correlation of the variables was studied among themselves as well with the target variable to get an understanding of the most probable important features. From the results, it was identified that “Prolonged decelerations”, “Abnormal short-term Variability”, and “Percentage of time with abnormal long-term variability” have highest correlation with the target variable i.e., fetal health (*Fig. 2*). A heatmap was used to study the correlation among the variables. Most of the variables had moderate correlation except “histogram mean” and “histogram median” (*Fig. 3*). However, these variables were not removed because they showed high variability (*Fig. 1*).

Finally, the visualization of the distribution of the target variable was made to reveal whether the data is balanced or unbalanced. The results showed that Class 1 was more common than the others and hence it was concluded that the dataset is not balanced (*Fig 4*). As a result, it was decided that along with the accuracy of the prediction model, precision and recall will be considered for model evaluation.

	fetal_health
fetal_health	1
prolongued_decelerations	0.484859
abnormal_short_term_variability	0.471191
percentage_of_time_with_abnormal_long_term_variability	0.426146
histogram_variance	0.20663
baseline_value	0.148151
fetal_movement	0.08801
histogram_min	0.0631749
light_decelerations	0.0588705
histogram_number_of_zeroes	-0.0166818
histogram_number_of_peaks	-0.023666
histogram_max	-0.0452654
histogram_width	-0.0687888
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

Fig 2: Correlation of attributes with the target variable

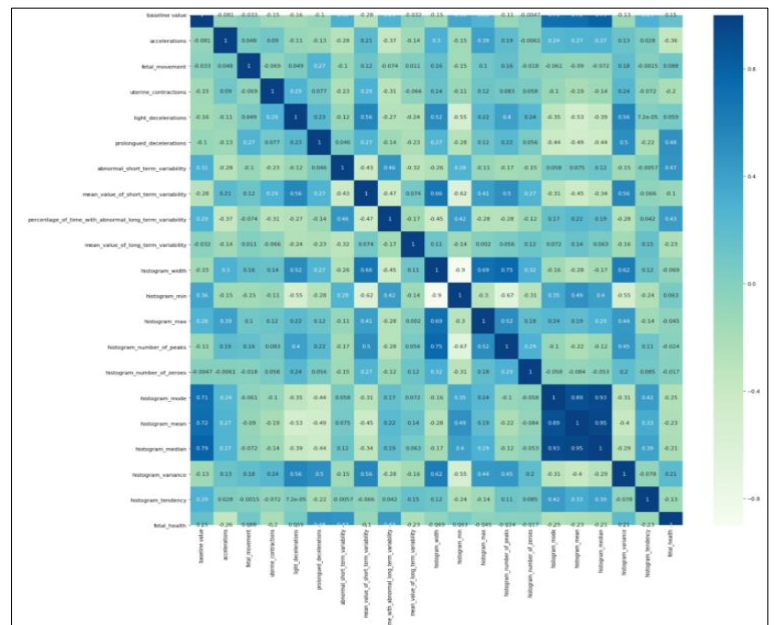


Fig 3: Heatmap to study the correlation of attributes among themselves

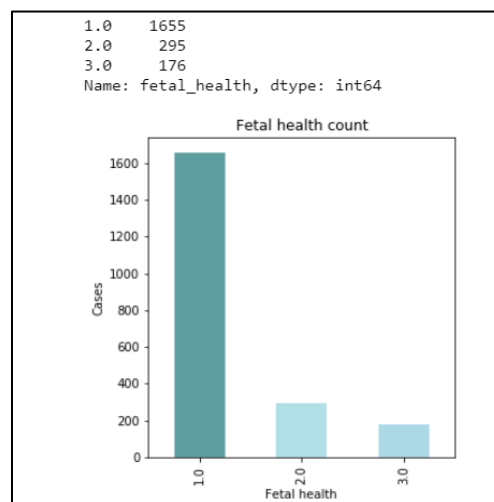


Fig 4: Distribution of the fetal health classes (1,2,3)

For the process of model building, the dataset was split into testing and training data. 70% of the data was used for training the model and the remaining 30 % was used for testing and evaluating the performance of the model.

a) KNN model building:

For KNN model, the parameter “n_neighbours” was tuned to optimize its performance. This was done using two techniques: Accuracy curve and GridSearchCV. As both the methods provided different optimal parameters and model accuracy for the model, it can be inferred that one is better and rigorous than the other and thus it must be the preferred method for future parameter tuning.

In this case, optimization using accuracy curve gave an accuracy of 89.8% with n_neighbours = 7 (Fig 5). This gives good accuracy on training as well as testing data. Thus, it avoids overfitting and underfitting of the data. However, the accuracy was similar to that with default parameters. (89.4%). On the other hand, GridSearchCV was able to secure a score of 91.3%. This method chose n_neighbours=1. However, it also defined other optimal parameters to improve the model performance (Fig. 6).

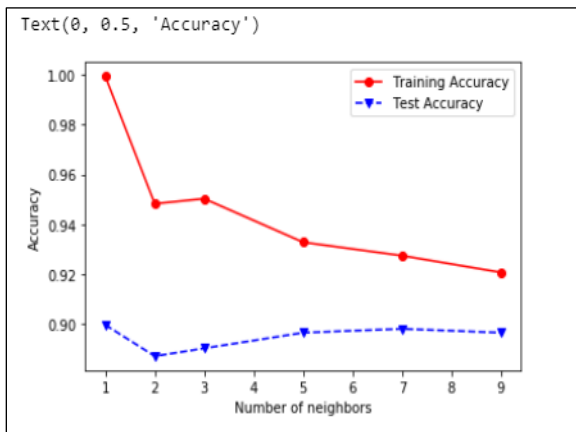


Fig 5: Accuracy curve to optimize KNN model

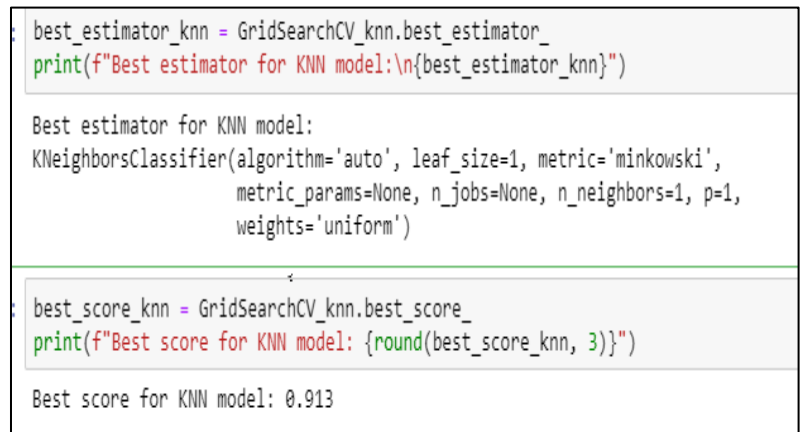


Fig 6: GridSearchCV to optimize KNN model by identifying the optimal parameters

b) Random Forest model building:

As GridSearchCV performed better to optimize the KNN model, it was also used for the Random forest parameter tuning. It was successful in bringing up the accuracy of the baseline model (with default parameters) from 91.7% to 94% (Fig. 7 and 8).

```
# Baseline model of RF with default parameters:

random_forest = RandomForestClassifier()
random_forest_mod = random_forest.fit(X_train, Y_train)
print(f"Baseline Random Forest: {round(random_forest_mod.score(X_test, Y_test), 3)}")

pred_random_forest = random_forest_mod.predict(X_test)

Baseline Random Forest: 0.917
```

Fig 7: Random forest model using default parameters

```
best_params_RF = GridSearchCV_RF.best_params_
print(f"Best parameter values for RF model:\n{best_params_RF}")

Best parameter values for RF model:
{'criterion': 'gini', 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 300}

best_score_RF = GridSearchCV_RF.best_score_
print(f"Best score for RF model: {round(best_score_RF, 3)}")

Best score for RF model: 0.94
```

Fig 7: Random forest model using GridSearchCV-selected parameters

However, as the dataset not unbalanced, precision and recall were calculated using the confusion matrix in order to get the true picture. The results showed that the average score for the KNN model was 0.89 as compared to 0.93 for the random forest (Fig. 8 and 9). Hence, it can be concluded that the random forest performed better to predict fetal health based on the CTGs features.

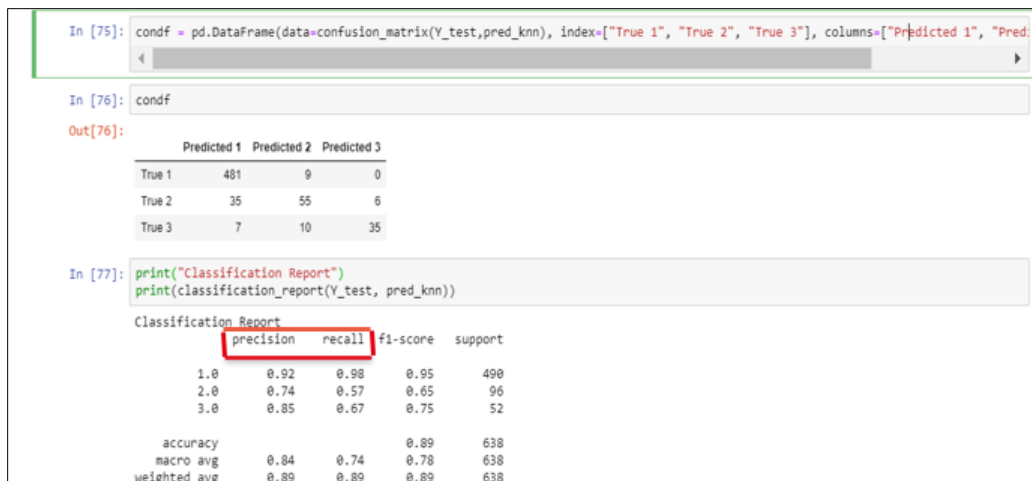


Fig 8: KNN model: Confusion matrix and classification

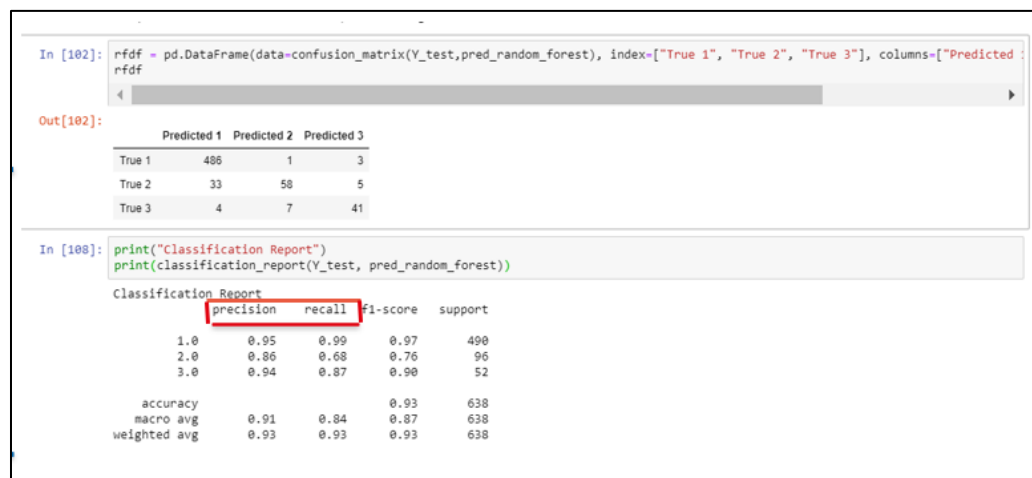


Fig 9: Random forest: Confusion matrix and classification

Conclusions:

Based on the analysis results, following conclusions can be made:

1. Random forest performed better and was more efficient than KNN classification algorithm to classify fetal health (*Fig 10*). This is mainly because random forest is an ensemble method and has multiple trees with each tree working on a subset of the data. This makes the algorithm robust, stable and unbiased.



Fig 10: Summary of scores from KNN and Random forest classification algorithm

2. Along with being accurate, Random Forest was also much faster than KNN model. The computation time of Random forest (1.1 mins) is also much better than KNN classification model (3.3 min). The difference is also large: 3 times (*Fig.11 and 12*).

```
# Fit model with train data
GridSearchCV_knn.fit(X_train, Y_train);
```

Fitting 3 folds for each of 1160 candidates, totalling 3480 fits

```
[Parallel(n_jobs=-1)]: Using backend LokyBackend with 8 concurrent workers.
[Parallel(n_jobs=-1)]: Done 34 tasks | elapsed: 6.4s
[Parallel(n_jobs=-1)]: Done 184 tasks | elapsed: 22.9s
[Parallel(n_jobs=-1)]: Done 434 tasks | elapsed: 44.5s
[Parallel(n_jobs=-1)]: Done 784 tasks | elapsed: 1.1min
[Parallel(n_jobs=-1)]: Done 1234 tasks | elapsed: 1.6min
[Parallel(n_jobs=-1)]: Done 1784 tasks | elapsed: 2.1min
[Parallel(n_jobs=-1)]: Done 2434 tasks | elapsed: 2.5min
[Parallel(n_jobs=-1)]: Done 3184 tasks | elapsed: 3.1min
[Parallel(n_jobs=-1)]: Done 3480 out of 3480 | elapsed: 3.3min finished
```

Fig 11: KNN model: Computation time to fit the model

```
# Fit model with train data
GridSearchCV_RF.fit(X_train, Y_train);
```

Fitting 3 folds for each of 36 candidates, totalling 108 fits

```
[Parallel(n_jobs=2)]: Using backend LokyBackend with 2 concurrent workers.
[Parallel(n_jobs=2)]: Done 46 tasks | elapsed: 28.3s
[Parallel(n_jobs=2)]: Done 108 out of 108 | elapsed: 1.1min finished
```

Fig 12: Random Forest: Computation time to fit the model

- As RF performed much better than KNN model, it was used to identify the important features that impact fetal health. The top 3 features based on their importance score were: “Mean value of short-term Variability”, “Abnormal short-term Variability”, and “Histogram mean” (Fig. 13 and 14).
- Although the correlation of “Histogram mean” was very high, the high variance in the variable made it an important feature to predict fetal health. Hence, it can also be concluded that all the factors must be considered before dropping a variable.

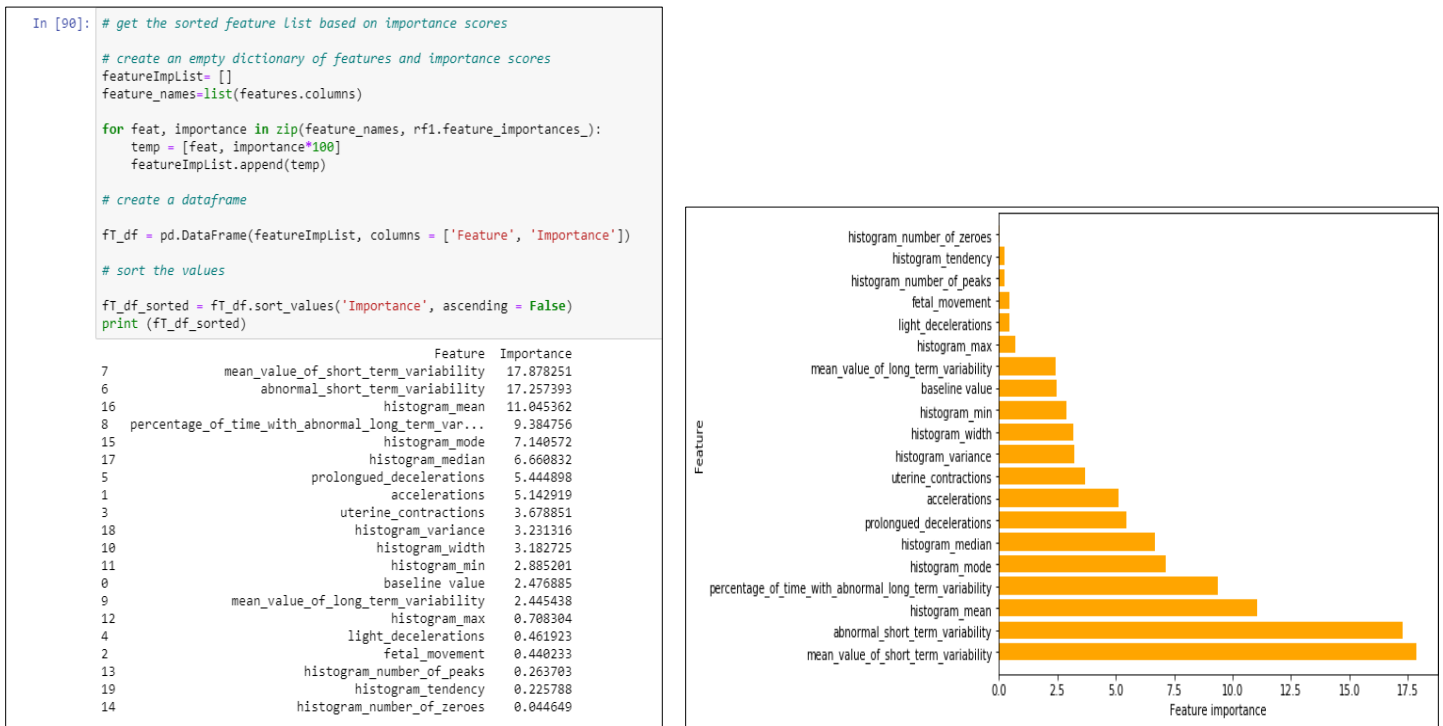


Fig 13 and 14: Selecting important features that effect fetal health using RF

Discussion and future directions:

A multiclass RF model can assist doctors to prevent deaths of newborns and mothers by predicting health using results from simple and cost-accessible CTGs. It can offer the following advantages:

- Simple and fast
- Can be made widely available
- Can be applied in low-resource settings where mortality rate is high.

Similar models can also be developed for other exams to improve accuracy of prediction. This will also allow the doctors to give personalized treatments to the patients as the faulty parameters specific to each individual can also be identified using the model.

However, CTGs can only work as primary screening indicator and thus after getting a suspect or pathological condition, further investigation is required to confirm the diagnosis.

References:

1. Shee, Hassan & Cheruiyot, Kipruto & Kimani, Stephen. (2014). Application of k-Nearest Neighbour Classification in Medical Data Mining. 4.
2. Fawagreh, K., Gaber, M.M. Resource-efficient fast prediction in healthcare data analytics: A pruned Random Forest regression approach. *Computing* 102, 1187–1198 (2020). <https://doi.org/10.1007/s00607-019-00785-6>
3. Daly N, Brennan D, Foley M, O'Herlihy C. Cardiotocography as a predictor of fetal outcome in women presenting with reduced fetal movement. *Eur J Obstet Gynecol Reprod Biol.* 2011 Nov;159(1):57-61. doi: 10.1016/j.ejogrb.2011.07.002. Epub 2011 Sep 6. PMID: 21899943.
4. Ayres de Campos et al. (2000) SisPorto 2.0 A Program for Automated Analysis of Cardiotocograms. *J Matern Fetal Med* 5:311-318
5. Data retrieved from <https://www.kaggle.com/andrewmvd/fetal-health-classification>