Fetal Health Classification using Cardiotocography

Kimber Ashman

March 27, 2022

Abstract

The goal of this report is to determine a machine learning method to classify fetal health into three health categories.

keywords: *fetal health, cardiotocography*

---

1. Introduction

Cardiotocography, or the monitoring of fetal heart rate, movement, and uterine contractions, is a non-invasive ultrasound method used by obstetricians to access fetal health[1]. In this study, we evaluated different machine-learning methods to examine Cardiotocogram (CTG) data categorized as normal, suspect, and pathological by three obstetricians. In the future, such a classification algorithm could be used to support clinical decision making.

2. Methods

This CTG dataset contains 2126 records of features extracted from Cardiotocogram exams. Each record was classified by three experts as:[2]

- Normal

- Suspect

- Pathological

The features include:

- **baseline value:** FHR baseline (beats per minute)

- **accelerations:** Number of accelerations per second. Defined as the number of increases in FHR over the baseline lasting 15–120 s and reaching a peak of at least 15 bpm in 60 min.

- **fetal movement:** Number of fetal movements per second

- **uterine contractions:** Number of uterine contractions per second. Defined as the number of periods lasting a max of 254 s where on upward slope of 17 s was detected reaching a peak lasting more than 90s.

- **light decelerations:** Number of light decelerations per second. Defined as the number of decreases in FHR under the baseline lasting 15-120s.

- **severe decelerations:** Number of severe decelerations per second. Defined as number of decreases in FHR under the baseline lasting 120-300s.

- **prolonged decelerations:** Number of prolonged decelerations per second. Defined as the number of decelerations lasting more than 300s in 60 min.

- **abnormal short term variability:** Percentage of time with abnormal short term variability. Defined as the percentage of subsequent FHR signals differing 1 bpm.

- **mean value of short term variability:** Mean value of short term variability. Defined as the mean difference between adjacent FHR signals at 4 Hz on the fetal monitor, after removal of adjacent signals that differ 15 bmp.

- **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. Defined as the mean difference between max and min FHR in a 1 min sliding window, in segments free of accelerations or deceleration.

- **histogram width:** Width of FHR histogram

- **histogram min:** Minimum (low frequency) of FHR histogram

- **histogram max:** Maximum (high frequency) of FHR histogram

- **histogram number of peaks:** Number of histogram peaks

- **histogram number of zeroes:** Number of histogram zeros

- **histogram mode:** Histogram mode

- **histogram mean:** Histogram mean

- **histogram median:** Histogram median

- **histogram variance:** Histogram variance

- **histogram tendency:** Histogram tendency

There is no missing data among the features; however, the data is unbalanced with the number of normal measurements far exceeding suspect or pathological (Figure 1).

Based on the correlation matrix (Figure 2), it is clear that "accelerations","prolonged decelerations", "abnormal short term variability", "percentage of time with abnormal long term variability" and "mean value of long term variability" are the features with higher correlation with our target feature: "fetal health." Because the features exist at different scales, we rescaled them before building the models. The data was split according to a 0.33 test to train ratio. We compared 4 supervised learning models:

1. K-nearest neighbor (KNN).

2. Naive bayes

3. Decision tree

4. Random forest

For the KNN model, we evaluated the error curves and found a k value of 11 gives the best accuracy.

We selected a Random Forest model because it had the best accuracy (Table **??**).

We used scikit's GridSearchCV to find the best parameters for the Random Forest Model. The maximum depth was estimated to be 6 and the number of estimators was 100.
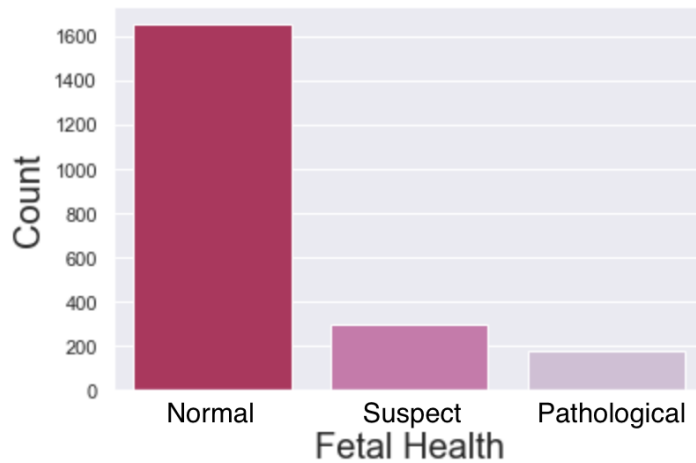
Figure 1: Class distribution

Model	Accuracy
KNN	0.863
Naive Bayes	0.833
Decision Tree	0.886
Random Forest	0.904

Table 1: Accuracy metrics

3. Results and discussion

The final accuracy was predicted to be 92%. Because the dataset is unbalanced, using accuracy as a performance metric may be misleading. As a result, we also analyzed the confusion matrix of the Random Forest model to understand the distribution of correct and incorrect predictions (Figure 3). Of the normal class, 94% were true positives. In the case of infant morbidity and mortality, it is probably more important to correctly identify suspected and pathological cases. The model does not perform as well across these cases, most likely due to the class imbalance. In the future, more data could be collected to improve the class balance.

4. Conclusions

We built a Random Forest model to classify fetal health according to CTG measurements.

References

[1] Diogo Ayres de Campos, João Bernardes, Antonio Garrido, Joaquim Marques de sá, and Luis Pereira-leite. Sisporto 2.0: A program for automated analysis of cardiotocograms. *Journal of Maternal-Fetal Medicine*, 9(5):311–318, 2000.

[2] João Monteiro-Santos, Teresa Henriques, Inês Nunes, Célia Amorim-Costa, João Bernardes, and Cristina Costa-Santos. Complexity of cardiotocographic signals as a predictor of labor. *Entropy*, 22(1), 2020.
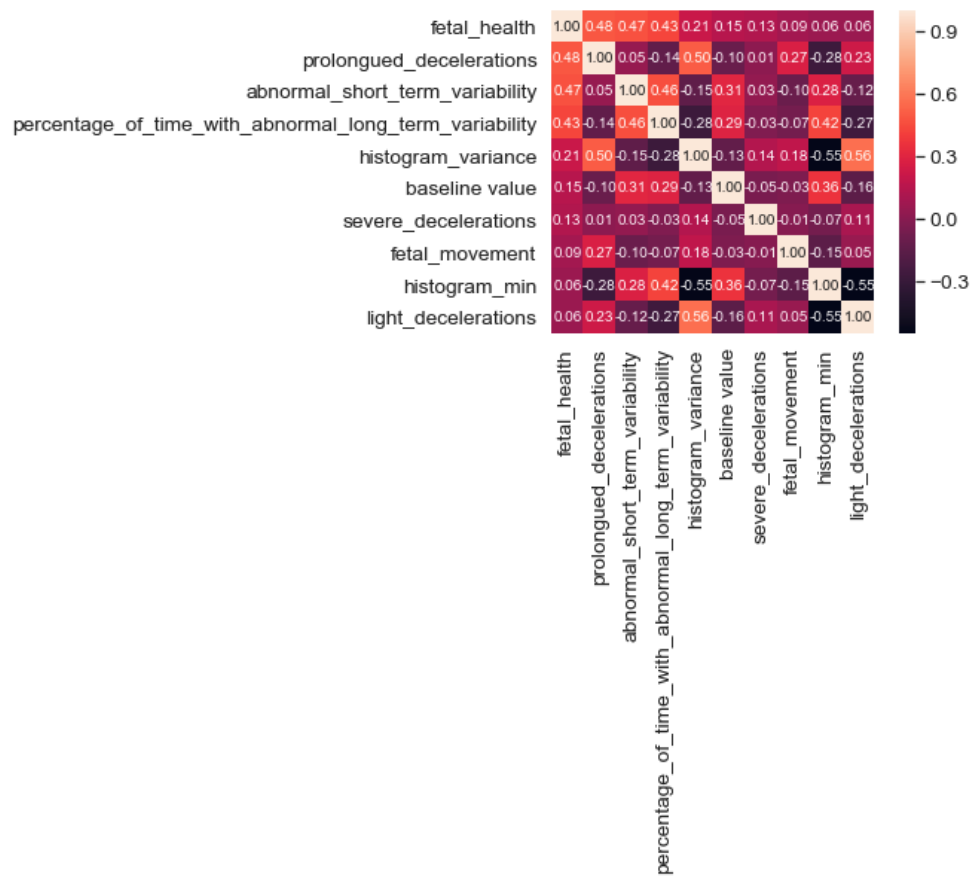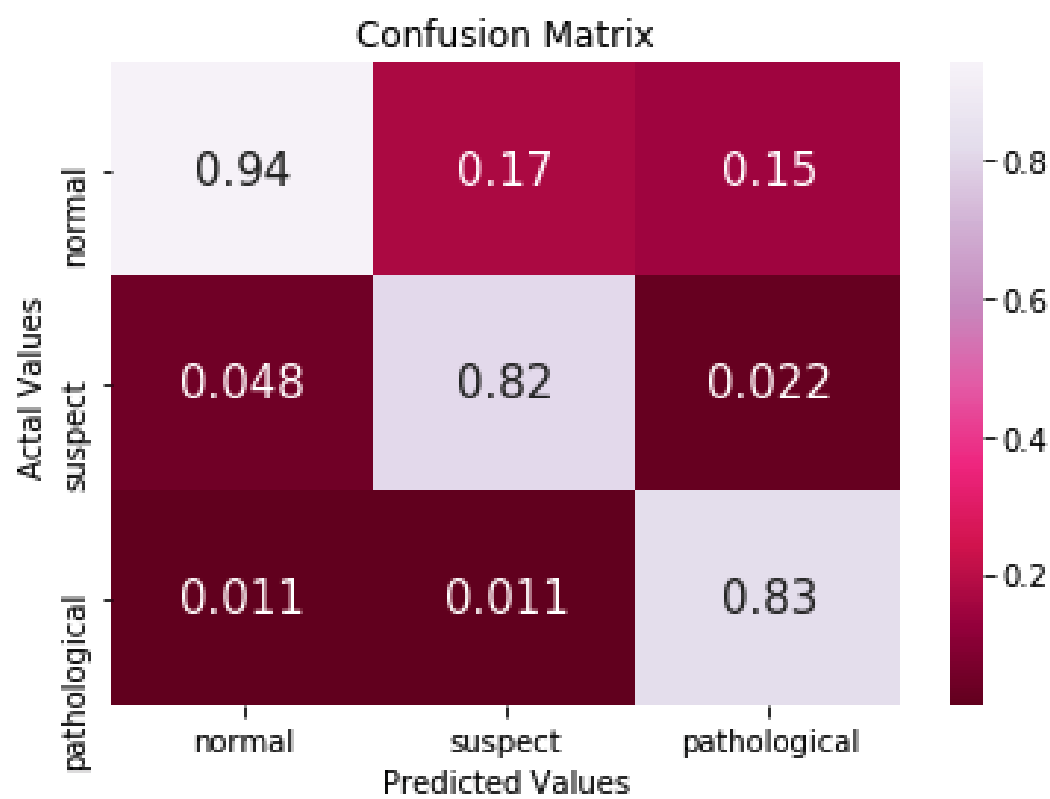
Figure 2: Correlation matrix

Figure 3: Confusion matrix