

# NORTHEASTERN UNIVERSITY

**Student** – Sajal Gangrade

**Email** – gangrade.s@northeastern.edu

Module 1 – Understanding Baby Health

**Class Number** - ALY 6020

**Class Name** – Predictive Analytics

**CRN Number** – 80249

# Professor – Justin Grosz

# Date – May 25, 2025

**2**

# Understanding Baby Health

# Introduction

Fetal health monitoring plays a critical role in ensuring the safety of both the baby and the mother during pregnancy and delivery. Clinicians often rely on cardiotocography (CTG) data to assess the baby's condition, using signals like fetal heart rate and uterine contractions. However, interpreting these signals in real time can be challenging — especially when early warning signs are subtle or inconsistent.

The goal of this project was to build a predictive model that helps distinguish between normal and potentially risky fetal conditions using historical CTG data. To achieve this, we applied a supervised machine learning algorithm — K-Nearest Neighbors (KNN) — which classifies new observations based on the similarity to known, labeled cases. By analyzing patterns across a range of physiological indicators, this model can serve as a supporting tool for clinical decision-making.

We focused on identifying which physiological signals are most predictive of abnormal outcomes, building a model that not only performs well but also makes intuitive sense to practitioners. The broader goal is to explore whether such a model could be realistically used in a hospital setting to help triage and prioritize care during labor and delivery.

# Dataset Overview

The dataset used in this project contains cardiotocography records of 2,126 fetal monitoring sessions. Each row represents a single session, recorded during pregnancy, and includes several numerical variables derived from fetal heart rate and uterine activity signals. These variables are used by clinicians to assess fetal well-being in real time.

After a preliminary review, we focused on the following core features:

* **Baseline value** – the average fetal heart rate, typically measured in beats per minute.
* **Accelerations** – increases in fetal heart rate, usually associated with fetal movement and healthy oxygenation.
* **Fetal movement** – spontaneous movements detected by the CTG machine.
* **Uterine contractions** – pressure changes indicating contractions of the uterus.
* **Light, severe, and prolonged decelerations** – drops in fetal heart rate, categorized by duration and severity.
* **Abnormal short-term variability** – fluctuations in heart rate variability, often linked to fetal distress.

**3**

The target variable in the original dataset was a multiclass variable labeled fetal\_health, with values:

* 1 for normal fetal condition
* 2 for suspect
* 3 for pathological

Since the goal was to create a binary classification model to flag any form of abnormal condition, we redefined the target variable:

* 1 = normal
* 0 = not normal (includes both suspect and pathological)

This simplification reflects a real-world triage system — where the priority is to distinguish normal from potentially risky cases as early as possible.

# Data Preprocessing

Before training the model, we carefully prepared the dataset to ensure that the inputs were both accurate and interpretable. In a healthcare setting, where clinical data is often noisy and incomplete, this step was essential. One of our primary focuses during preprocessing was how to handle missing values — and we treated this task with caution, feature by feature.

### **Handling Missing Values: Field-by-Field Justification**

The dataset contained missing values in five features. Rather than applying a one-size-fits-all imputation method, we examined the **distribution** and **clinical context** of each variable and then decided the most appropriate way to handle its missing data.

#### baseline\_value

* **Missing values**: 32
* **Distribution**: Nearly symmetric with low skew (skewness ≈ 0.03)
* **Decision**: Imputed using the **mean**
* **Why**: Since the distribution was roughly normal, using the mean helped preserve the central tendency. There were no extreme outliers that could distort it.
* **Support**: Visualized via histogram and boxplot — values ranged realistically from ~106 to 160 bpm, consistent with expected heart rate ranges during pregnancy.

#### A graph of a normal distribution AI-generated content may be incorrect.

#### 4

#### fetal\_movement

* **Missing values**: 21
* **Distribution**: Highly right-skewed (skewness ≈ 7.79)
* **Decision**: Imputed using the **median**
* **Why**: A few extremely high values were present, but most readings were close to 0. Using the mean would have pulled imputed values unrealistically high. The median was better suited to reflect the typical (and clinically normal) absence of movement.
* **Support**: Confirmed by boxplot, which showed a long tail and tight cluster near 0.

#### A graph of a person's body AI-generated content may be incorrect.

#### light\_decelerations

* **Missing values**: 14
* **Distribution**: Moderately skewed (skewness ≈ 1.71)
* **Decision**: Imputed using the **median**
* **Why**: Although not as extreme as fetal movement, the distribution was still right skewed. Median preserved the central tendency without being influenced by outliers.
* **Support**: Boxplot revealed some extreme values, but most entries were close to 0, consistent with clinical norms.

#### A graph of a bar and a box AI-generated content may be incorrect.

#### 5

#### severe\_decelerations

* **Missing values**: 52
* **Distribution**: Extremely sparse, most values were 0 (skewness ≈ 17.14)
* **Decision**: Imputed with **0**
* **Why**: These events are rare and usually indicate acute fetal distress. In most sessions, the value was zero, so a missing value likely meant that the event simply didn’t occur or wasn’t recorded — not that it should be estimated.
* **Support**: Both histogram and boxplot showed a dense cluster at 0, with almost no non-zero entries.

#### A graph of a graph of a number of deceleration AI-generated content may be incorrect.

#### prolongued\_decelerations

* **Missing values**: 23
* **Distribution**: Highly right skewed and sparse (skewness ≈ 4.30)
* **Decision**: Imputed with **0**
* **Why**: As with severe\_decelerations, these are rare but important clinical markers. Missing values were interpreted as the likely absence of an event.
* **Support**: Histogram showed that nearly all values were zero, and outliers were few and distinct.

A graph of a box and a box

AI-generated content may be incorrect.

**6**

**Summary of Imputation Methods**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Feature | Missing Values | Distribution Skewness | Imputation Method | Reasoning |
| Baseline\_value | 32 | 0.03 (symmetric) | Mean | Normal-like distribution |
| Fetal\_movement | 21 | 7.79 (highly skewed) | Median | Skewed with many near-zero values |
| Light\_decelerations | 14 | 1.71 (moderately skewed) | Median | Avoid influence of mild outliers |
| Severe\_decelerations | 52 | 17.14 (sparse) | 0 | Rare event, zeros dominate |
| Prolongued\_deceleration | 23 | 4.30 (sparse) | 0 | Rare event, zeros dominate |

# Feature Selection and Exploratory Data Analysis

Once the data was cleaned and preprocessed, the next step was to identify which features were most relevant to predicting fetal health. Rather than relying solely on statistical methods, we combined clinical intuition with visual analysis to make sure the selected variables would be both predictive and explainable to medical professionals.

**Initial Screening of Features**

The dataset included multiple physiological indicators derived from cardiotocography. While all of them had some potential relevance, using too many features in a KNN model can increase computational load and reduce interpretability — especially when some features may be noisy or redundant.

So, we decided to narrow down our model input to three carefully selected features. To do this, we plotted boxplots of each variable against the target (normal vs. not normal), allowing us to visually inspect whether the feature showed meaningful separation between the two classes.

A group of blue squares with black text

AI-generated content may be incorrect.

**7**

### **Selected Features: What We Chose and Why**

#### ****1. Accelerations****

* **What it measures**: Spikes in fetal heart rate, often associated with movement or healthy reactivity.
* **What we saw**: Clear separation in the boxplot — normal cases showed higher acceleration values, while abnormal cases clustered close to zero.
* **Why we chose it**: Accelerations are widely recognized as a sign of fetal well-being. The model can use low acceleration to flag potential risk.

**A graph with a blue square and a white background

AI-generated content may be incorrect.**

#### ****2. Prolongued Decelerations****

* **What it measures**: Extended drops in heart rate, which may signal fetal distress or oxygen deprivation.
* **What we saw**: Most values were 0, but when non-zero values occurred, they appeared almost exclusively in the not-normal class.
* **Why we chose it**: Although rare, these events are highly predictive when they happen. They help the model catch edge cases that other features might miss.

**A graph of a diagram

AI-generated content may be incorrect.**

**8**

#### ****3. Abnormal Short-Term Variability****

* **What it measures**: The irregularity in beat-to-beat heart rate intervals.
* **What we saw**: Clear distinction — normal cases had higher variability scores, while not-normal cases showed reduced variability.
* **Why we chose it**: Clinically, reduced heart rate variability is a red flag and often precedes other signs of distress.

**A diagram of a graph

AI-generated content may be incorrect.**

### **Why We Didn’t Choose Other Features**

* **Fetal Movement**: Too many zero values and high skew made it less reliable as a predictor.
* **Uterine Contractions**: Did not show strong separation between classes; more useful in a supporting role than as a primary indicator.
* **Light and Severe Decelerations**: Too infrequent or overlapping across both classes.

After this visual and clinical screening, we selected a **balanced mix:**

* One feature that increases with health (accelerations)
* One that sharply flags distress (prolongued\_decelerations)
* One that captures subtle rhythm changes (abnormal\_short\_term\_variability)

This mix gives our KNN model a strong foundation for distinguishing between normal and not-normal fetal conditions — even when the signs are subtle or intermittent.

**9**

# Model Building and Evaluation

With our features selected and dataset cleaned, we built a classification model using the **K-Nearest Neighbors (KNN)** algorithm. KNN is a simple but powerful method that predicts a class based on the “closeness” of new data points to previously labeled ones. Because KNN is distance-based, it relies heavily on properly scaled and relevant features — which we ensured during preprocessing.

We trained and tested three different models, each using a different value for **K** (number of neighbors):

* **K = 5**: Highly localized predictions — sensitive to noise, may overfit
* **K = 15**: Balanced model — captures both local patterns and general trends
* **K = 30**: More generalized — smoother predictions, but might underfit

Each model was evaluated on a **20% test set,** using the same data split for a fair comparison.

### **Evaluation Metrics Used**

We evaluated each model using:

* **Confusion matrix** – to show how many normal/not-normal cases were correctly or incorrectly classified
* **Accuracy** – overall correctness
* **Precision, Recall, F1-Score** – to understand performance for each class, especially important in healthcare where false negatives can be dangerous

**Model 1: KNN with K = 5**

A screenshot of a computer

AI-generated content may be incorrect.

**Interpretation**

This model was slightly biased toward identifying abnormal cases — which can be a good thing in clinical settings. It was very strong at catching risky cases (92% recall for not normal), but slightly weaker at identifying normal cases.

**10**

**Model 2: KNN with K = 15**

**A screenshot of a computer

AI-generated content may be incorrect.**

**Interpretation**

This was the most **balanced** model. It maintained high accuracy while achieving almost equal performance across both classes. It avoided overfitting and underfitting and generalizes well. We selected **K = 15 as the final model** for its stability and clinical practicality.

**Model 3: KNN with K = 30**

**A screenshot of a computer

AI-generated content may be incorrect.**

**Interpretation**

K = 30 slightly underfit the data. It smoothed out noise but sacrificed some responsiveness. While still strong, it offered no major advantage over K = 15 and slightly reduced model sensitivity.

**11**

**Comparison Summary**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **K Value** | **Accuracy** | **Normal F1** | **Not Normal F1** | **Best for….** |
| **5** | **88.73%** | **75%** | **93%** | **Catching risky cases aggressively** |
| **15** | **88.97%** | **75%** | **93%** | **Balanced, stable, generalizable** |
| **30** | **88.50%** | **74%** | **93%** | **Smoother but less responsive** |

# Real – World Interpretation and Clinical Use

**How well does the model classify Fetal Health?**

Based on our evaluation results, the KNN model (with **K = 15**) achieved nearly **89% accuracy**, with balanced precision and recall for both “normal” and “not normal” cases. These results are especially promising in a clinical context, where the cost of missing a risky case can be high.

Notably:

* The model correctly identified **93% of not-normal cases**, which means it rarely misses potentially dangerous situations.
* It also maintained a respectable **75% F1-score for normal cases**, ensuring it doesn't overwhelm clinicians with unnecessary alerts.

This balance is important: over-warning can lead to alarm fatigue; under-warning can risk fetal harm. Our model performed strongly on both fronts.

### **Are the Selected Features Clinically Meaningful?**

Each of the three features used in the final model is grounded in clinical research and daily medical practice:

1. **Accelerations**  
   Typically reflect fetal responsiveness. Lack of accelerations is one of the earliest signs of fetal distress. The model uses this as a key signal of normality.
2. **Prolongued Decelerations**  
   When present, they are serious and usually indicate oxygen deprivation. The model accurately flags these rare events, giving weight to high-risk conditions.
3. **Abnormal Short-Term Variability**  
   Variability in heart rate is a crucial indicator of neurological and physiological function. Reduced variability, when detected, significantly increases the risk profile — and our model correctly uses this to distinguish cases.

This model could be used as a real-time decision-support tool in hospitals. It can help clinicians quickly flag high-risk fetal patterns for further review, especially in high-volume or understaffed settings. While it wouldn’t replace medical judgment, it offers an interpretable, data-driven second opinion to support early intervention.

**12**

# Practical Strengths and Limitations of Using KNN in Healthcare

**Strengths**

1. **Simplicity**  
   KNN is intuitive — it classifies a new case by comparing it to similar past cases. This makes it easy to explain and implement, especially in clinical settings where transparency matters.
2. **Interpretability**  
   The model doesn’t rely on complex equations or hidden weights. Instead, predictions are based on observable data from known patients. This gives healthcare providers confidence in what the model is “thinking.”
3. **No Training Time**  
   KNN is known as a "lazy learner," meaning it doesn’t require time-consuming training. This allows it to be updated or re-run with new data easily — a useful trait in evolving hospital environments.

### **Limitations**

1. **Sensitive to Feature Scale**  
   Since KNN relies on distance calculations, variables with large scales can overpower smaller ones. We addressed this by standardizing all features — but this step is critical and must be maintained in any future use.
2. **Affected by Irrelevant Features**  
   If unimportant variables are included, they can distort distance calculations and reduce accuracy. That’s why we carefully selected only three clinically relevant features based on visual and domain analysis.
3. **Computational Cost with Large Datasets**  
   As the number of records grows, KNN can become slower because it compares a new case to every previous one. For small to medium datasets like ours, it works well — but for deployment on large hospital databases, optimization would be needed.

# Conclusion

This project demonstrated how a simple, interpretable model like K-Nearest Neighbors (KNN) can be used to assist in fetal health classification using cardiotocography data. After carefully cleaning the dataset and selecting three clinically relevant features — accelerations, prolonged decelerations, and abnormal short-term variability — we built and evaluated multiple KNN models to find the best balance between accuracy and reliability.

The final model, using **K = 15**, achieved nearly **89% accuracy**, with strong recall for identifying abnormal fetal conditions. These results suggest that such a model could be used in real-world hospital settings as a decision-support tool, helping prioritize care and flag early signs of fetal distress.

By combining clinical intuition with machine learning, this approach reinforces the idea that interpretable models — when built responsibly — can play a meaningful role in supporting patient care.

**13**

# References

* Ayres-de-Campos, D., Spong, C. Y., & Chandraharan, E. (2015). **FIGO consensus guidelines on intrapartum fetal monitoring: Cardiotocography.** International Journal of Gynecology & Obstetrics, 131(1), 13–24. <https://doi.org/10.1016/j.ijgo.2015.06.020>
* Hasan, M. K., & Sultana, M. (2020). **Fetal Health Classification from Cardiotocography Data Using Machine Learning Techniques.** Indonesian Journal of Electrical Engineering and Computer Science, 19(1), 382–389. <https://doi.org/10.11591/ijeecs.v19.i1.pp382-389>
* Han, J., Kamber, M., & Pei, J. (2011). **Data Mining: Concepts and Techniques (3rd ed.).** Morgan Kaufmann.
* Altman, D. G., & Bland, J. M. (1994). **Diagnostic tests 1: Sensitivity and specificity.** BMJ, 308(6943), 1552. <https://doi.org/10.1136/bmj.308.6943.1552>
* Scikit-learn Developers. (n.d.). **K-Nearest Neighbors — scikit-learn 1.4.1 documentation.** <https://scikit-learn.org/stable/modules/neighbors.html>
* Singh, A., Thakur, N., & Sharma, A. (2016). **A review of supervised machine learning algorithms.** 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), 1310–1315. IEEE.
* Chudáček, V., Spilka, J., Koucký, M., Lhotská, L., & Huptych, M. (2011). **Use of artificial intelligence in cardiotocography classification.** Acta Polytechnica, 51(3), 12–17.
* Patel, N., & Desai, A. (2020). **Comparative analysis of classification techniques on CTG dataset.** International Journal of Advanced Trends in Computer Science and Engineering, 9(2), 1488–1493. <https://doi.org/10.30534/ijatcse/2020/12922020>