BIOS/EPID 511: Healthcare Data Science
Fall 2024

Final Project Report on 'Fetal Health Classification'

Ajay Sreekumar

GitHub Repository: https://github.com/BIOS-511/Fetal-Health-Classification

Literature Review

Importance of Public Health Question

The project addresses the critical issue of fetal health monitoring using cardiotocogram (CTG) data to predict fetal health outcomes. This topic is significant for several reasons:

1. **Maternal and Fetal Mortality Reduction**: Accurate and timely analysis of CTG data aligns with global health goals, such as the United Nations Sustainable Development Goals for maternal and child health. Preventable fetal distress is a leading cause of complications during labor.
2. **Advancements in Accessible Health Monitoring**: CTG is a cost-effective, widely accessible method for fetal monitoring, and applying machine learning (ML) techniques can enhance diagnostic accuracy and reliability.
3. **Integration of Historical Data with AI**: The ability to use over 50 years of CTG data in advanced ML models exemplifies how historical medical data can improve current health outcomes.

Supporting Studies

- **Frasch et al. (2021)**
    - **Objective**: Developed a deep learning model to detect early fetal distress.
    - **Methodology**: Leveraged 50 years of electronic fetal monitoring (EFM) data. CTG tracings were digitized and analyzed using convolutional neural networks (CNN). These models were trained to process fetal heart rate (FHR) patterns and uterine contractions in real time.
    - **Findings**: The model achieved a 94% accuracy rate in identifying early signs of fetal distress. This accuracy is a major improvement over traditional diagnostic methods. The study also showed the ability to prevent injuries through precise detection of distress, bridging historical medical records with cutting-edge AI technology.
    - **Significance**: Demonstrates the feasibility of real-time, AI-driven fetal health diagnostics that can directly impact clinical practice.
- **Georgieva et al. (2013)**
    - **Objective**: Investigated machine learning approaches for classifying CTG data.
    - **Dataset**: Used 634 CTG recordings manually labeled by nine obstetricians.

- **Methodology**: Tested several machine learning models, including decision trees, SVMs, and neural networks. The classification performance was evaluated using precision, recall, and F1 scores.
- **Findings**: The SVM model achieved the best performance due to its ability to handle complex feature relationships effectively. Key features included baseline fetal heart rate, variability, and acceleration metrics.
- **Significance**: Provided strong evidence for ML models' utility in real-time monitoring and diagnosis, emphasizing their practical application in clinical settings.

- **Reddy et al. (2023)**
  - **Objective**: Addressed the challenges of noisy signals and imbalanced datasets in fetal health classification.
  - **Dataset**: Consisted of 2,000+ CTG records collected from a multicenter study.
  - **Methodology**: Employed a convolutional neural network (CNN) to classify CTG patterns. Advanced preprocessing techniques were applied to handle data noise and class imbalances effectively.
  - **Findings**: CNN outperformed traditional machine learning algorithms, particularly in handling the noisy and imbalanced CTG data. It highlighted the importance of features like decelerations, which are critical for classifying fetal health.
  - **Significance**: Reinforced the role of deep learning approaches in fetal health diagnostics, demonstrating their ability to overcome limitations of traditional methods.
- **Spilka et al. (2022)**
  - **Objective**: Compared the performance of various machine learning models for classifying fetal health.
  - **Dataset**: Included 2,126 CTG records.
  - **Methodology**: Utilized Random Forest (RF), Decision Trees, and AdaBoost classifiers. These models evaluated features such as prolonged decelerations, baseline variability, and abnormal heart rate patterns.
  - **Findings**: RF and AdaBoost outperformed Decision Trees, achieving higher classification accuracy. The study identified severe and prolonged decelerations as critical indicators of abnormal fetal health states.
  - **Significance**: Highlighted the potential of ensemble learning approaches in identifying critical features for fetal health classification, emphasizing the value of feature engineering and model selection.

Statistical Question

The statistical question posed is:

How can automated analysis of CTG data improve the prediction of fetal health outcomes and contribute to reducing fetal and maternal mortality?

Data Description

1. How Was the Data Created?

The dataset was derived from cardiotocograms (CTGs), a widely used diagnostic tool during pregnancy. These CTGs were generated using ultrasound sensors to record three main physiological signals:

- **Fetal Heart Rate (FHR)**: Captures the baby's heartbeat and identifies variations crucial for assessing well-being.
- **Uterine Contractions**: Tracks the mother's contractions to evaluate their influence on fetal activity.
- **Accelerations**: Represents transient increases in the fetal heart rate, which can indicate fetal well-being.

The data was curated and labeled by expert obstetricians into three categories based on clinical observations:

1. **Normal**: Indicates no apparent signs of distress.
2. **Suspect**: Potential signs of fetal stress requiring closer monitoring.
3. **Pathological**: Clear evidence of distress requiring immediate intervention.

2. Response Variable and Variable of Interest

- **Response Variable**:

The primary variable of interest is **fetal_health**, which is a categorical variable indicating fetal condition as:

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

- **Definition**:

These categories were defined based on established clinical criteria such as baseline heart rate, variability, accelerations, decelerations, and uterine contractions as interpreted by obstetricians.

3. Confounding Variables

Confounding variables were identified as factors potentially influencing the relationship between independent features and the fetal_health variable:

- **Uterine Contractions**:
    - **Impact**: Directly affect fetal heart rate variability and the occurrence of decelerations.
- **Maternal Health Factors** (e.g., gestational diabetes, hypertension, maternal age):
    - **Impact**: Could influence baseline FHR and variability metrics.
    - **Challenge**: Maternal health data was not included in the dataset, limiting adjustments for these confounders.
- **Measurement Noise**:
    - **Impact**: Variability caused by ultrasound sensor inaccuracies or environmental factors.

These were considered confounders because they could obscure or distort the true relationship between FHR metrics and fetal health classification. Addressing these involved analyzing interactions between features to minimize noise.

4. Issues Related to the Dataset

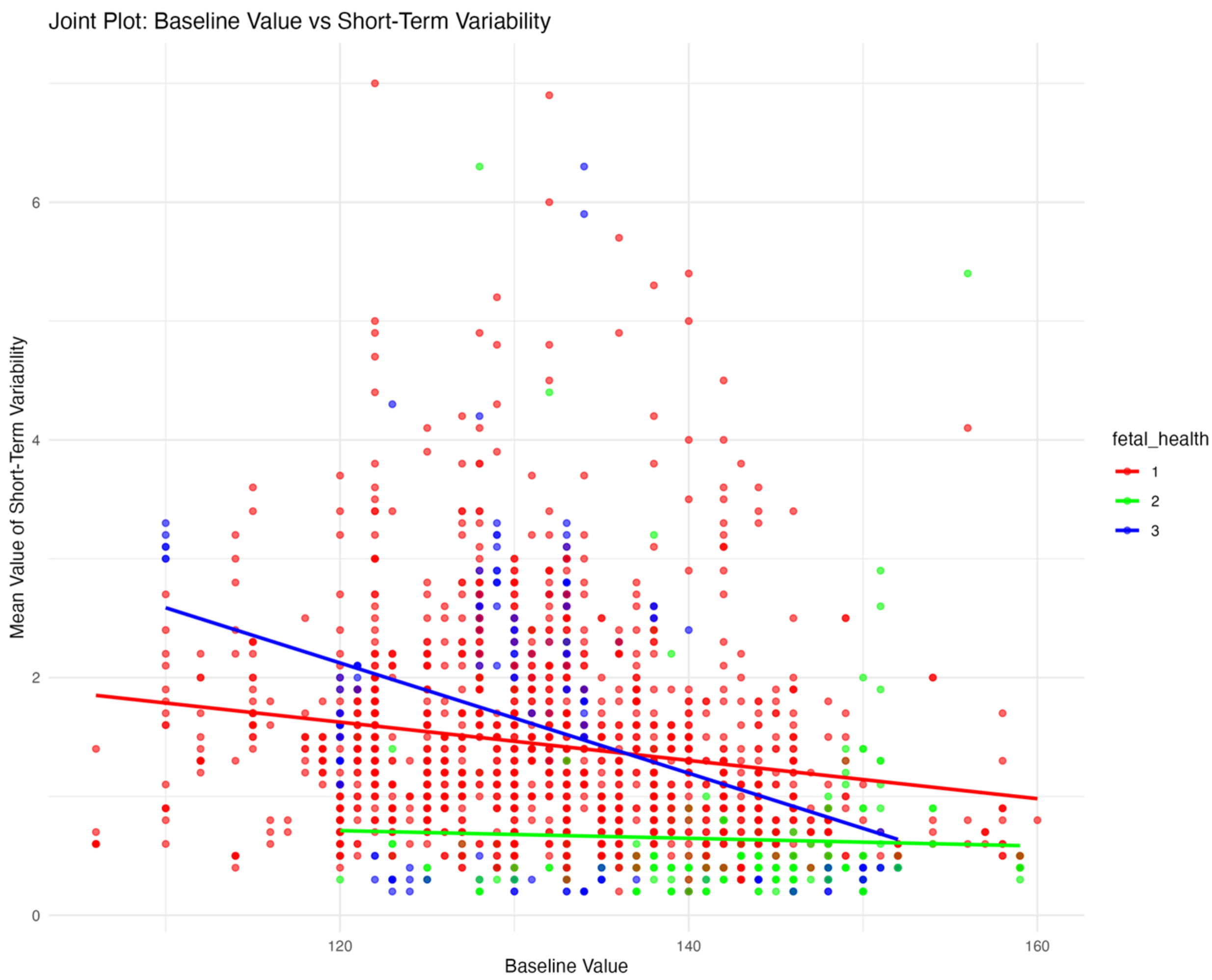
- **Class Imbalance**:
    - The dataset had a majority of samples labeled as Normal, with fewer in Suspect and Pathological categories.
    - **Solution**: Synthetic Minority Oversampling Technique (SMOTE) was applied to balance class distributions by generating synthetic samples for minority classes.
- **Missing Data**:
    - The dataset had minimal missingness, as CTG recordings are typically complete due to their clinical context.
    - Missing values were handled through imputation where necessary or excluded from analyses to ensure robust results.

Data Analysis

1. Exploratory Data Analysis (EDA)

The exploratory analysis revealed key insights into the dataset's structure and relationships between variables:

- **Scatter Plot Analysis**:



Visualized the relationship between Baseline FHR and Short-Term Variability across fetal health classes to assess how these relationships differ among fetal health categories: **Normal, Suspect, and Pathological**.

Visual Elements of the Plot:

1. **Scatter Plot:** Each point represents an observation, color-coded by fetal health status.
2. **Trend Lines:** Linear model fits for each health category, highlighting the trend without the distraction of confidence intervals.

The data points are color-coded based on fetal health classes:

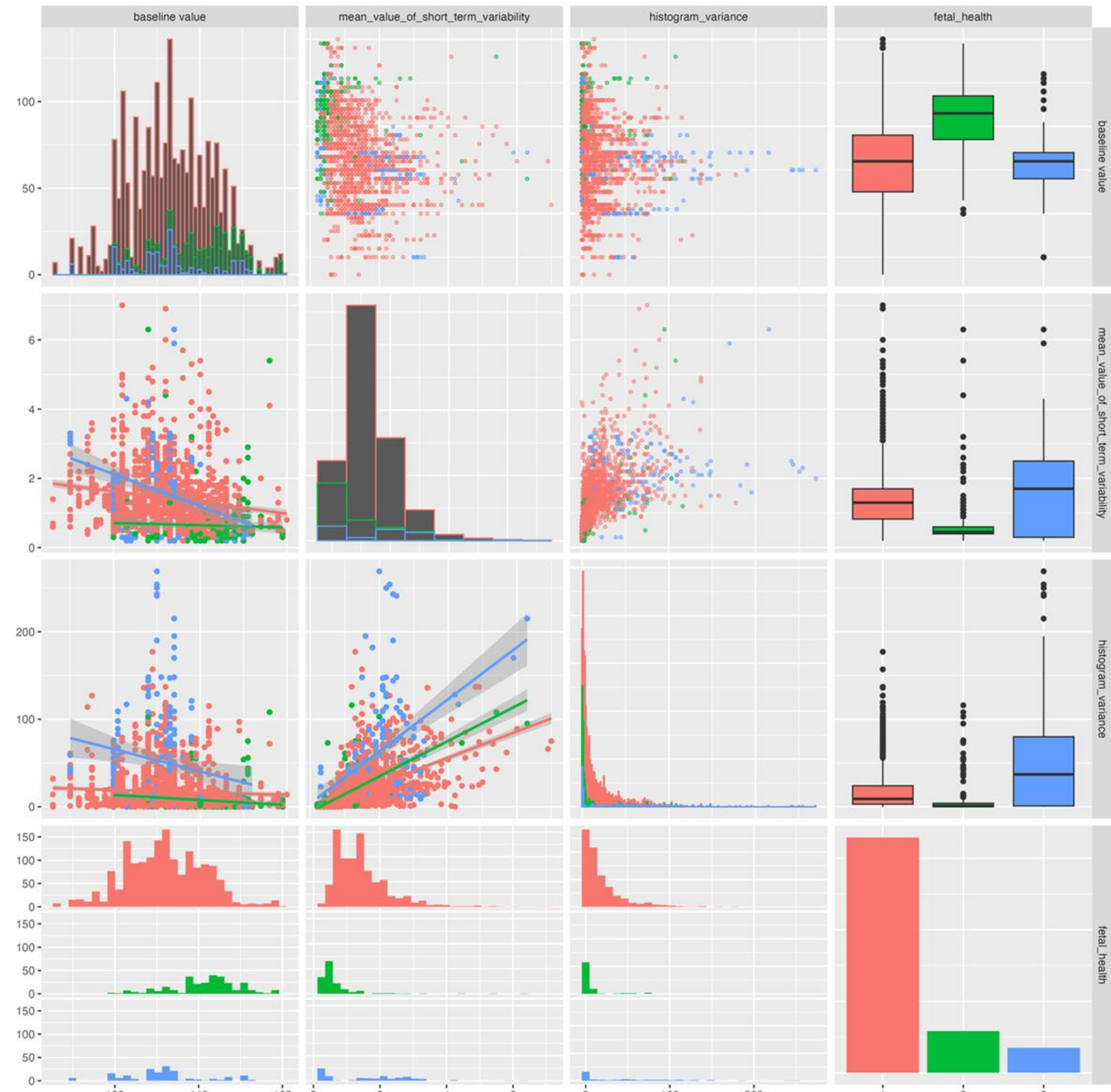
- Red (Class 1 - Normal)
- Green (Class 2 - Suspect),
- Blue (Class 3 - Pathological).

Key Insights:

- A negative correlation is evident for Class 1 (Normal) and Class 3 (Pathological), indicating that as Baseline FHR increases, Short-Term Variability decreases.
- Class 3 (Pathological) shows significantly lower baseline values compared to Class 1 and Class 2.

This highlights how Baseline FHR and Short-Term Variability are critical features for differentiating fetal health categories.

- **Pairwise Plot Analysis**:



This plot visualizes pairwise relationships between key features like Baseline Value, Short-Term Variability, Histogram Variance, and Fetal Health Class

Key Observations:

1. **Baseline Value Distribution:**

- Class 1 (Normal) has a wider and higher distribution of Baseline FHR.
- Class 3 (Pathological) is clustered toward lower baseline values.

2. **Box Plots:**

- Variance in **Histogram Variance** and **Short-Term Variability** is significantly higher for Class 3, suggesting instability in FHR for pathological cases.

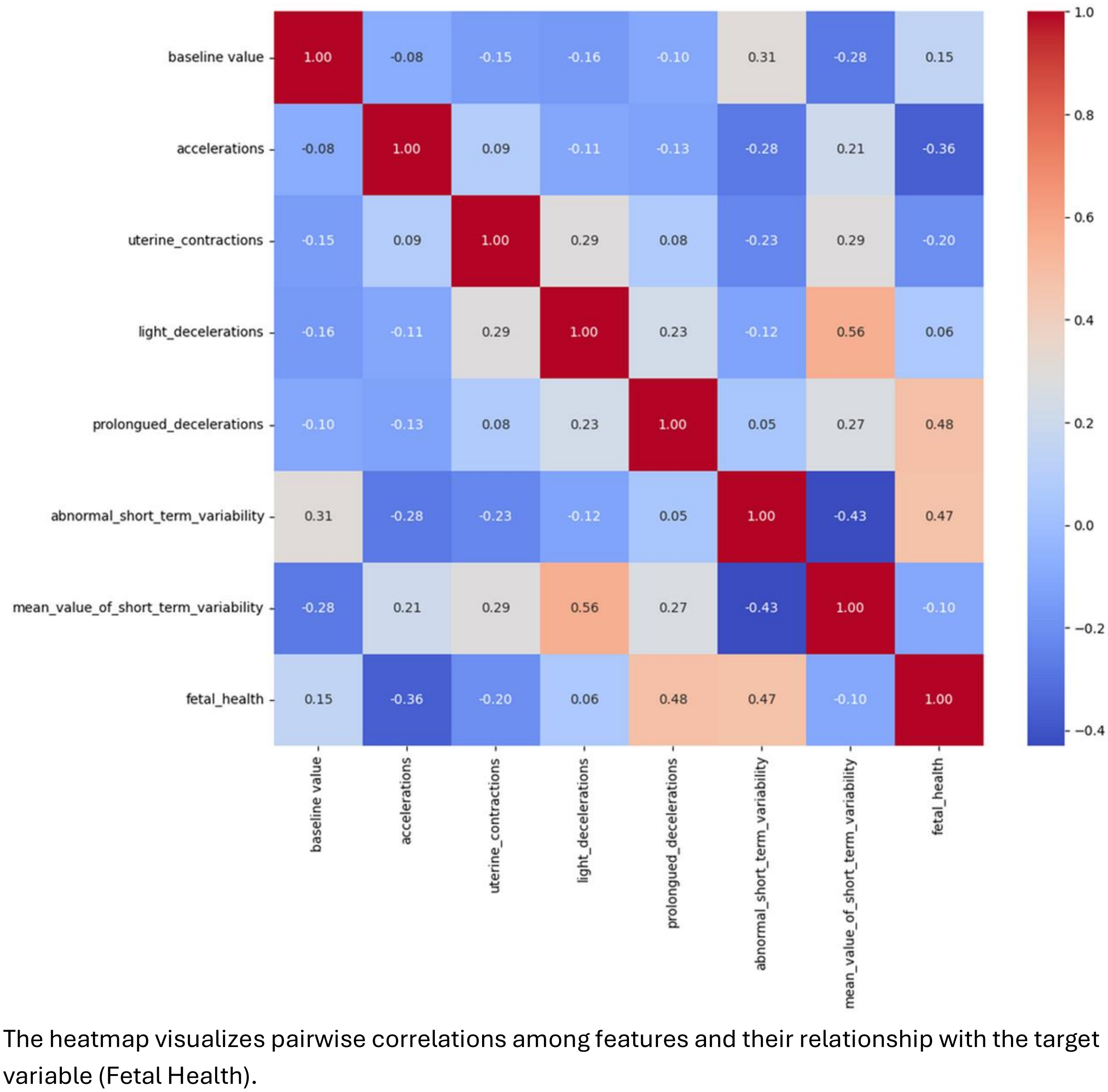
3. **Histograms:**

- The **imbalance** in class distribution is clearly visualized, with Class 1 dominating the dataset.

4. **Correlation Insights:**

- Features like Baseline Value and Histogram Variance show non-linear relationships across classes, further reinforcing the **need** for **non-linear classifiers** (e.g., Gradient Boosting and XGBoost).

- **Heatmap Analysis**:



The heatmap visualizes pairwise correlations among features and their relationship with the target variable (Fetal Health).

Key Observations:

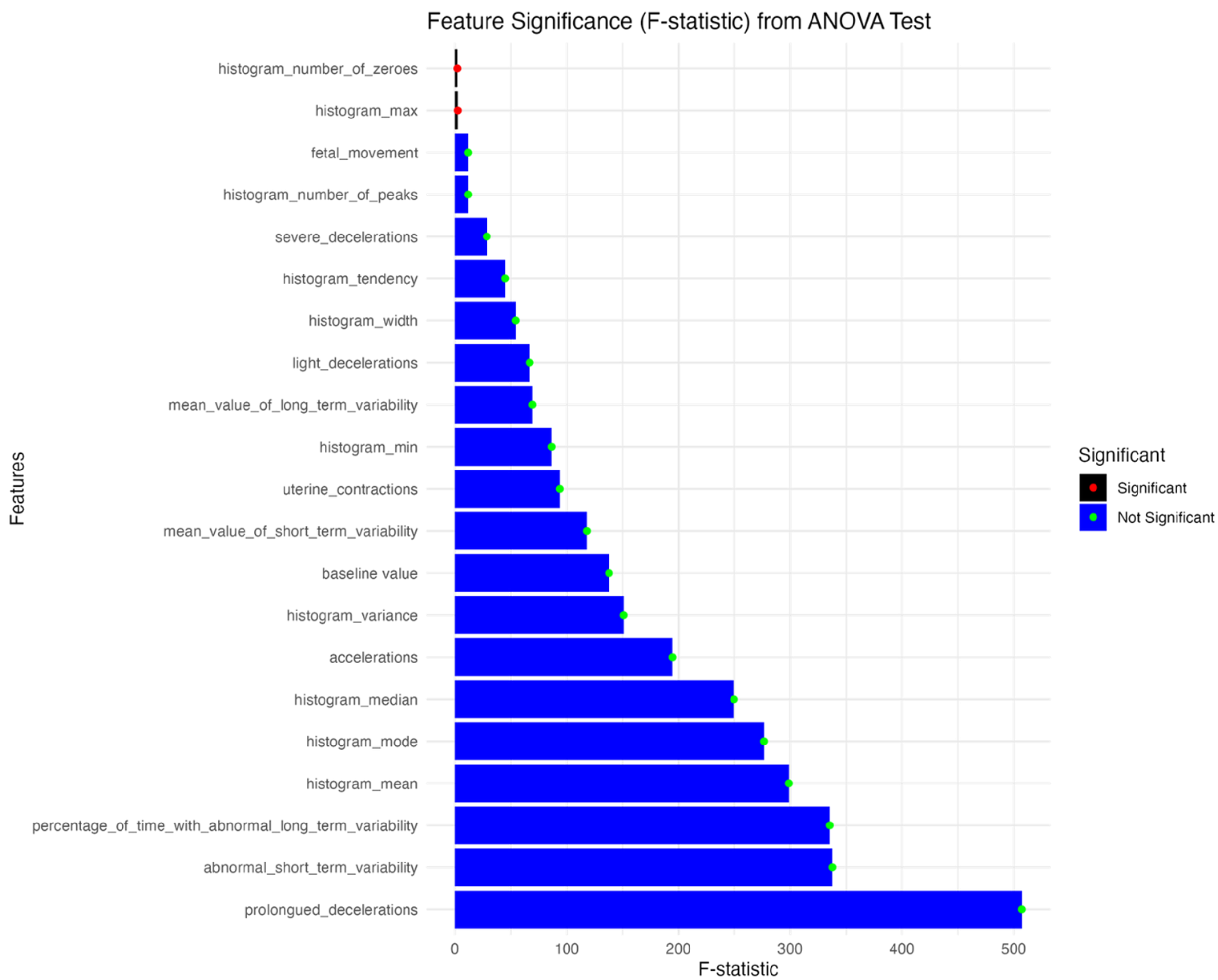
1. **Strong Correlations:**

- **Baseline Value** shows moderate positive correlation with Fetal Health (0.15).

- **Prolonged Decelerations** and **Abnormal Short-Term Variability** have strong correlations with Fetal Health, indicating their critical role in classification.

2. **Feature Interdependence:** Mean Value of Short-Term Variability is positively correlated with Light Decelerations (0.56), suggesting shared variability.

- **ANOVA Testing Analysis**:



Examining the impact of various physiological and monitoring features on fetal health using ANOVA (Analysis of Variance) F-statistics.

Methodology:

- Conducted ANOVA tests for each feature against fetal health categories.
- ANOVA and non-linear model applications addressed the variance in key features like Histogram variances and Short-term variability.
- Assessed the statistical significance and strength of associations through F-statistics.
- Features are grouped as **Significant** (red markers) or **Not Significant** (blue markers) based on their contribution to fetal health classification.

Key Findings:

- **Highly Significant Features:** Prolonged Decelerations, Abnormal Short-Term Variability, and Percentage of Time with Abnormal Long-Term Variability have the highest F-statistics, making them the most significant predictors of fetal health.
- **Features with Lower Impact:** Histogram number of zeroes, histogram max, and fetal movement.
- **Insights:** Focus on the top-ranked features can lead to better model performance and interpretability.

2. Analytical Methods

To answer the public health question, the following models and techniques were applied:

- **Logistic Regression**:
  - Used as a baseline model to interpret linear relationships between features and the target variable.
  - **Justification**: It provides simplicity and interpretability, making it useful for feature contribution analysis.

Logistic Regression and Variable Correlation:

Logistic regression works best when the relationship between the predictor variables (features) and the log-odds of the target (outcome) is **linear**. If some variables are non-linearly correlated with the target, the model will struggle to capture that relationship. Non-linearly correlated variables can reduce performance because the model cannot effectively fit non-linear patterns. However, it doesn't mean these variables are inherently "bad"; they just require transformations to make them interpretable by the model.

What Happens If You Use Only Linearly Correlated Variables?

If you filter out non-linearly correlated variables and keep only linearly correlated ones, the logistic regression model may perform better **if the linear variables contain most of the predictive information**. However, removing non-linearly correlated variables could result in **loss of important information**, especially if those variables hold predictive power that could be captured using transformations.

Alternative Approaches

If feature engineering to transform non-linear variables is complex or ineffective, consider using more flexible models that can handle both linear and non-linear relationships, like:

- **Random Forest Classifier**:
  - Constructed multiple decision trees to classify fetal health.
  - **Justification**: Effective at handling imbalanced data and providing insights into feature importance.
- **Gradient Boosting and XGBoost**:
  - Applied boosting techniques to sequentially minimize prediction errors.
  - **Justification**: Ideal for capturing complex, non-linear relationships in the data. XGBoost was particularly optimized for handling large datasets and avoiding overfitting.

3. Analytical Results:

The performance of each model was evaluated both before and after applying SMOTE for class imbalance:

Accuracy Assessment

Before SMOTE

	Accuracy	Precision	Recall	F-1 Score
Logistic Regression	0.875	0.866	0.875	0.866
Random Forest	0.927	0.924	0.927	0.923
Gradient Boosting	0.936	0.935	0.936	0.935
XGBoost	0.929	0.927	0.929	0.927

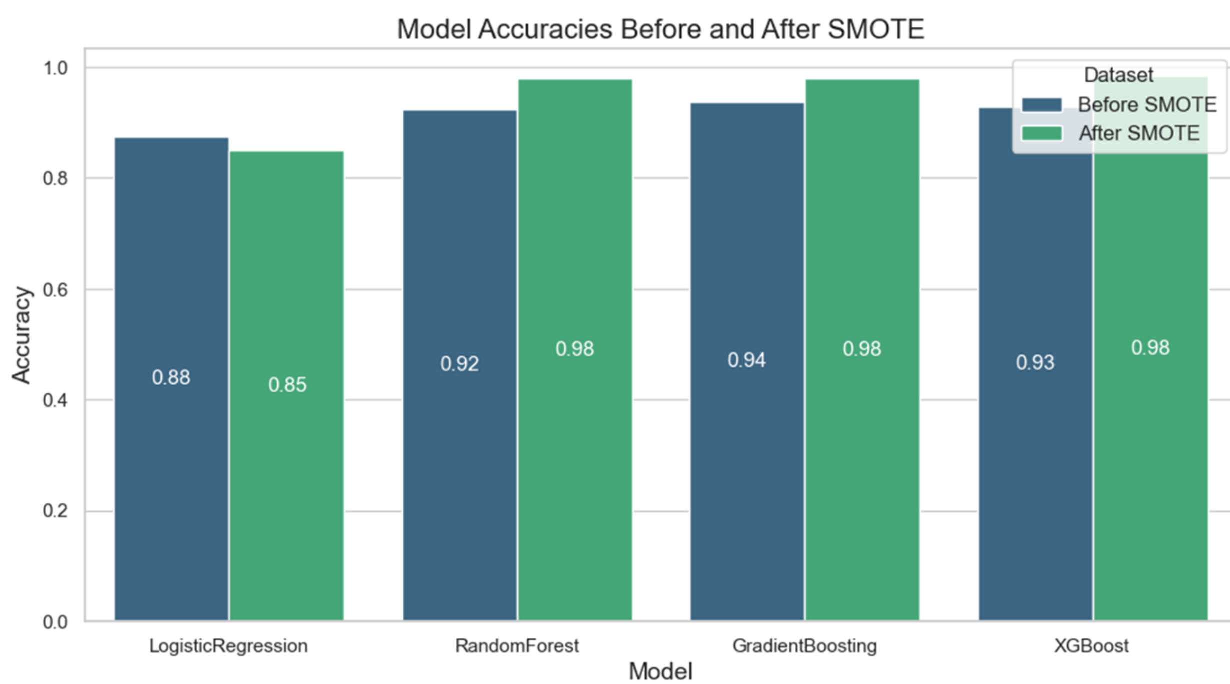
Accuracy Assessment

After SMOTE

	Accuracy	Precision	Recall	F-1 Score
Logistic Regression	0.849	0.851	0.849	0.849
Random Forest	0.978	0.979	0.978	0.978
Gradient Boosting	0.98	0.981	0.98	0.981
XGBoost	0.983	0.984	0.983	0.984

Impact of SMOTE: Balancing the dataset improved precision and recall, particularly for minority classes (Suspect and Pathological).

Performance Highlights:



1. Superior Model Performance:

- **Gradient Boosting, XGBoost** and **Random Forest** models achieved the highest accuracy (**98%**), showcasing their ability to handle complex, non-linear relationships within the dataset.
- These models effectively captured **non-linear patterns** in the dataset, addressing the complexity of fetal health classification.

2. Logistic Regression:
   - Served as a **baseline model** with moderate performance (85% accuracy), providing a linear perspective on feature contributions.

General Observations:

1. **Overfitting Mitigation:**

- To address potential overfitting due to the small dataset size, a 5-fold cross-validation was used, ensuring robust evaluation across all models.

2. **Dataset-Specific Challenges:**

- High accuracy values may reflect the smaller dataset's limited variability. However, SMOTE balanced the dataset effectively, allowing minority class patterns to emerge.

3. **Public Health Significance:**

- For public health data, prioritizing high accuracy reduces critical false negatives, enabling safer decision-making for fetal health interventions.
- Results were consistent with published studies, such as Frasch et al. (2021) and Spilka et al. (2022), which emphasized similar significant predictors like decelerations and variability metrics.

Quantitative Performance Summary:

Trade-offs:

- XGBoost offered marginally better precision, while Gradient Boosting demonstrated a slight edge in recall and F1-Score, emphasizing its suitability for imbalanced classes.

Conclusions

Summary of Findings

This project successfully demonstrated the potential of machine learning techniques for analyzing cardiotocogram (CTG) data to predict fetal health outcomes. Through a combination of exploratory data analysis and advanced modeling techniques, several key insights were obtained:

1. **Significant Predictors**: Prolonged Decelerations, Abnormal Short-Term Variability, and Percentage of Time with Abnormal Long-Term Variability were identified as the most critical features for fetal health classification.
2. **Model Performance**: Gradient Boosting and XGBoost emerged as the best-performing models, achieving an accuracy of 98% after addressing class imbalance with SMOTE. These models effectively captured non-linear relationships in the data.
3. **Addressing Data Challenges**:
    a. Class imbalance was successfully mitigated using synthetic oversampling techniques.
    b. The dataset's small size was accounted for through cross-validation, minimizing overfitting risks.

Consistency With Published Studies

The results align with existing research, such as Frasch et al. (2021), which also highlighted the significance of variability metrics and decelerations in predicting fetal distress. This consistency validates the robustness and clinical relevance of the models used in this project.

Public Health Impact

The findings underscore the importance of leveraging automated analysis for fetal health diagnostics, particularly in resource-limited settings where CTG is a cost-effective and accessible tool. By enabling real-time, AI-driven predictions, this approach has the potential to reduce preventable neonatal and maternal complications, contributing to global efforts in improving maternal and child health outcomes.

Limitations and Future Directions

While the project achieved high accuracy, several limitations were noted:

- **Limited Dataset Size**: A larger and more diverse dataset could enhance the model's generalizability.
- **Missing Maternal Health Data**: Incorporating maternal health factors like gestational diabetes and hypertension could improve predictive performance.
- **Exploration of Advanced Techniques**: Future research could explore ensemble models or transfer learning to further optimize predictions.

Conclusion

This project illustrates the transformative potential of AI in maternal-fetal healthcare. By bridging advanced analytics with clinical practice, it paves the way for scalable, cost-effective solutions to reduce maternal and neonatal mortality rates worldwide.