

Abstract

Problem

Understanding fetal neurological development is crucial for early detection of potential health issues, yet current methods for assessing autonomic nervous system maturation are often invasive, limited in scope, or require high level professionals to interpret complex diagnostic data, making widespread early screening challenging in many clinical settings. This study investigated whether heart rate variability (HRV) features from cardiotocogram (CTG) data could be used to predict fetal health status and serve as non-invasive indicators of ANS maturation.

Approach

From the fetal health dataset (Ayres de Campos et al., 2000) containing 2,126 CTG records, all features were used except for histogram features and light decelerations. The histogram features were removed because other variables already provided equivalent heart rate data, and light decelerations was removed due to high correlation with another feature, which could cause collinearity. In the second part of the study, only HRV-related features were analyzed. Machine learning models were trained on these features, with hyperparameters optimized using GridSearchCV and performance evaluated through K-Fold cross-validation.

Results

The best performing model for predicting fetal health using CTG data was a gradient boosting model optimized with GridSearchCV, achieving 95% training accuracy and 90% test accuracy. Additionally, abnormal short-term variability emerged as the most important predictor of fetal health. All models showed no statistically significant performance differences (t -test $p > 0.05$), reinforcing the robustness of the findings.

Conclusion

HRV features from CTG data can accurately predict fetal health and provide clinically relevant insights into ANS maturation. These findings support the integration of ML-driven HRV analysis into prenatal monitoring systems, enabling earlier detection of neurological immaturity or dysfunction and advancing non-invasive assessment in obstetric care.

Background

This project uses the public Kaggle Fetal Health Classification dataset, which contains 2,126 records of features extracted from cardiotocography (CTG) exams labeled by three expert obstetricians into three classes: Normal, Suspect, and Pathological. The ability to automate CTG interpretation through machine learning could supplement, and in some settings speed up or improve, expert review, making fetal health assessments more accessible and consistent. This is especially important given the United Nations' goal of ending preventable deaths of newborns and children under five by 2030, with CTGs providing a cost-effective, non-invasive means of monitoring fetal well-being.

My first research question asks: Can we accurately predict fetal health status using cardiotocographic features? My second question, inspired by my interest in neuroscience, asks: Can variability measures in fetal heart rate act as non-invasive indicators of autonomic nervous system (ANS) maturation? Prior work by Schneider et al. (2009) supports this investigation, showing that heart rate variability patterns reflect stages of ANS development before birth, with different patterns emerging before and after 32 weeks of gestation. This suggests that CTG-based variability metrics may serve as non-invasive biomarkers for fetal neurological development, providing both clinical and developmental insight.

Results

Research Question #1: Can we accurately predict fetal health status using cardiotocographic features?

Figure #1 shows the imbalance between the fetal health classes within the dataset. This can be expected because fetuses are usually healthy and normal; however, this can also create inaccurate scores for our model. Fortunately, gradient boosting deals with this sort of imbalance because it iteratively focuses on misclassified examples and can be configured with class weights or sampling techniques to give more attention to minority classes, which is why it performed so well.

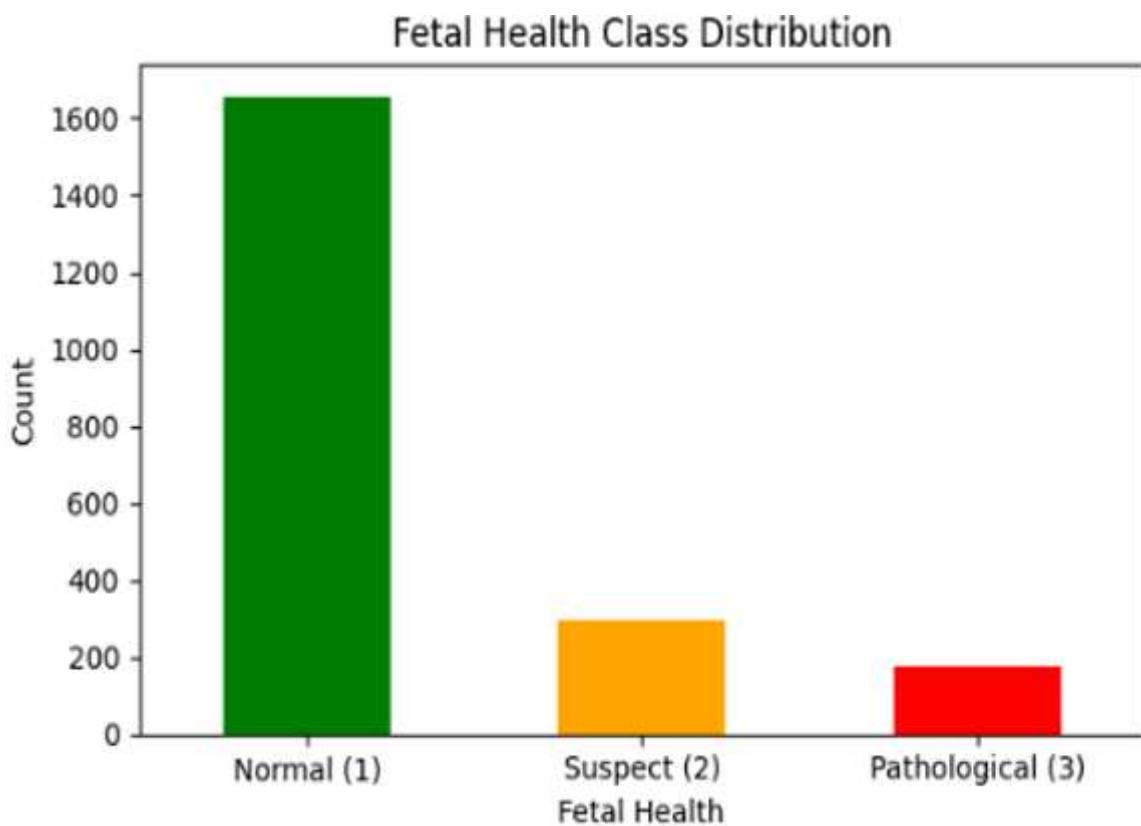


Figure #2 is a correlation matrix that shows the correlation between every feature. The redder the square, the more positively correlated it is to that feature, while the bluer the square, the more negatively correlated it is.

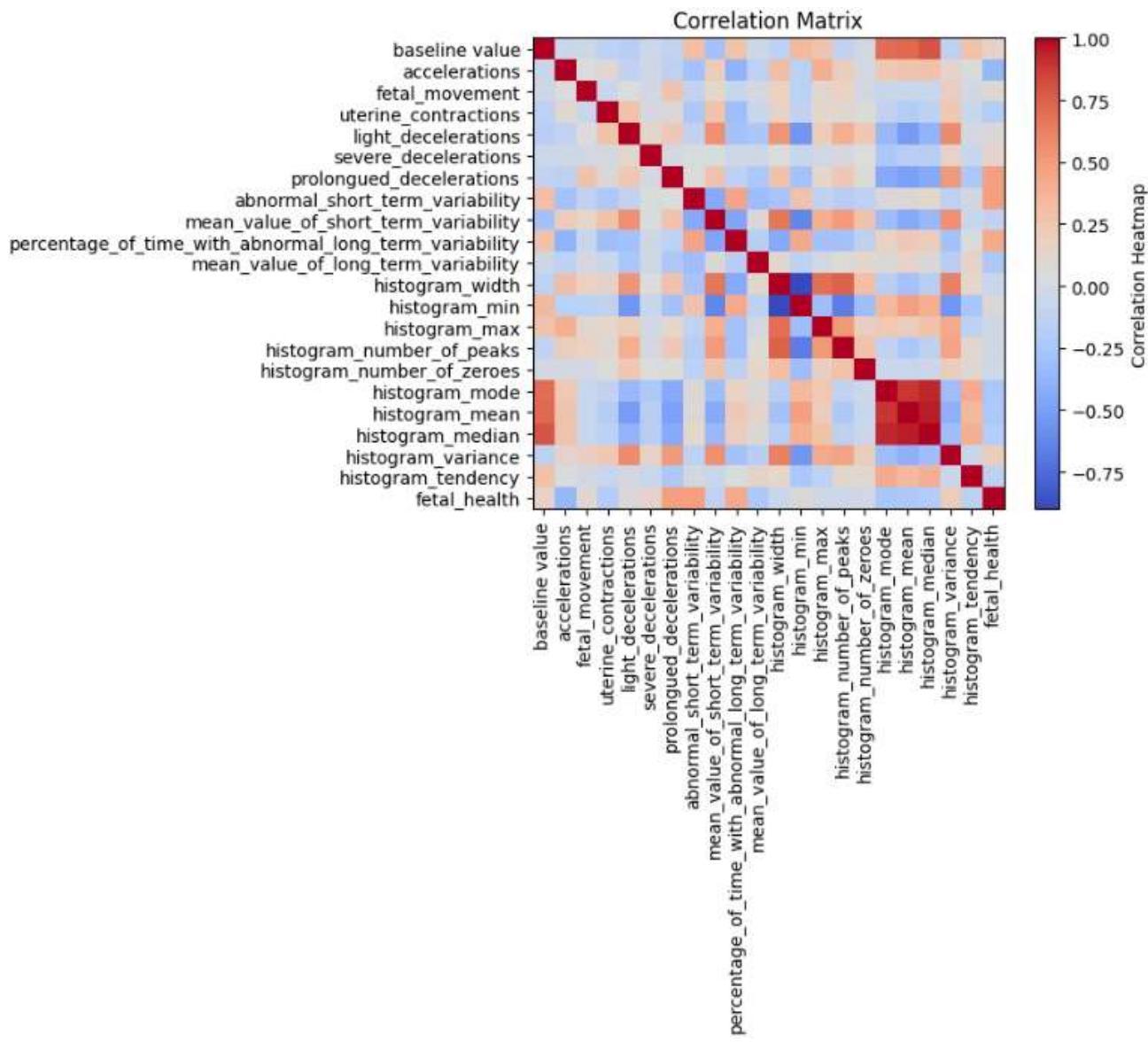


Figure #3 shows the gradient boosting GridSearchCV model's predictions for each class of fetal health. Its most frequent misclassification is predicting suspect cases as normal.

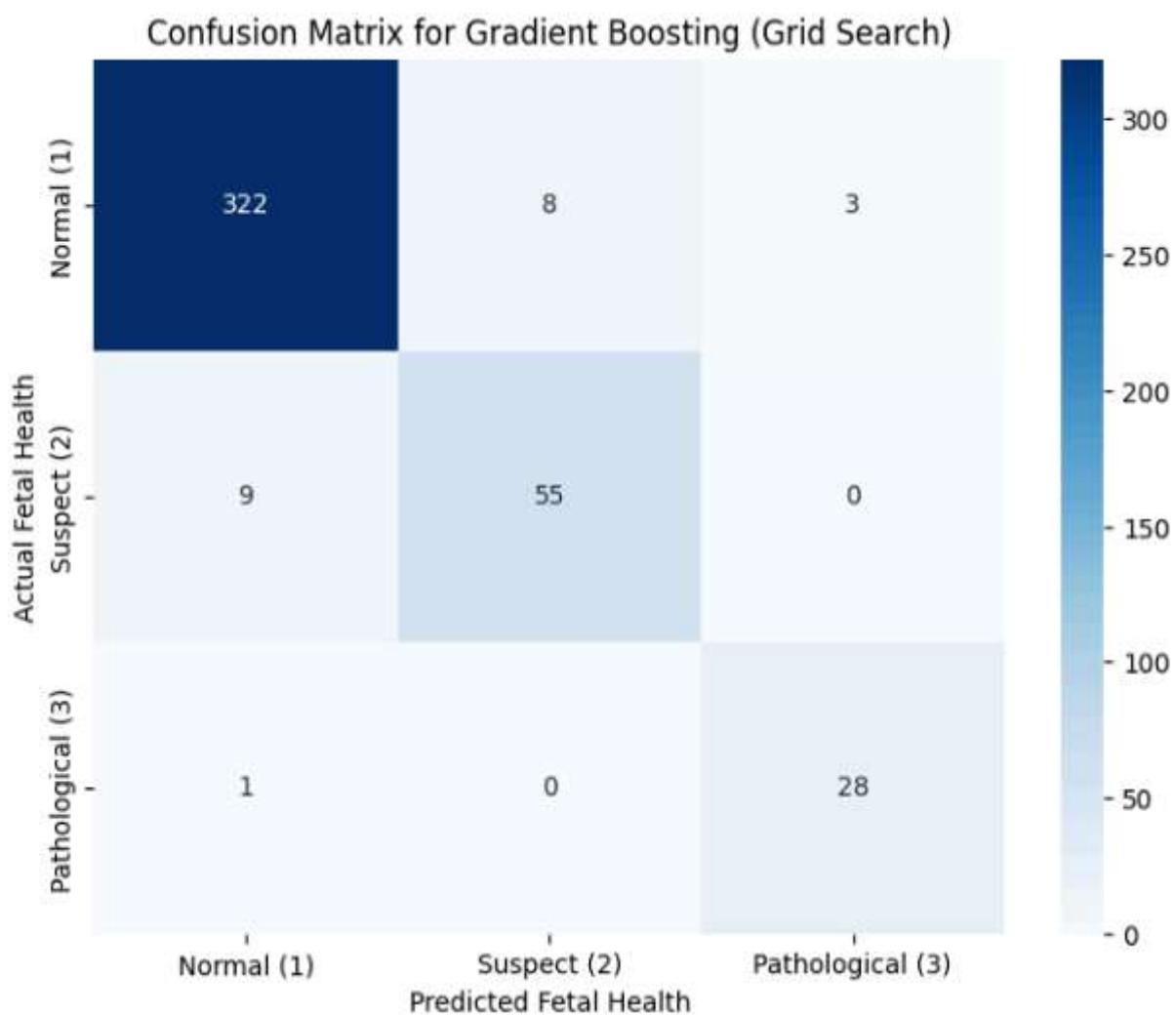


Figure #4 shows the classification report scores for each model tested on all of the features to predict fetal health. The models are sorted in order by accuracy, with Gradient Boosting (Grid Search) achieving the highest accuracy at 95.07%, while KNN performed the poorest with an accuracy of 89.96%.

	Accuracy	Precision (weighted)	Recall (weighted)	F1-score (weighted)
KNN (k=4)	0.899061	0.899935	0.899061	0.898674
Decision Tree	0.922535	0.924926	0.922535	0.923522
Random Forest	0.946009	0.944710	0.946009	0.944523
Gradient Boosting	0.948357	0.947965	0.948357	0.948015
Gradient Boosting (Grid Search)	0.950704	0.950790	0.950704	0.950667

Research Question #2: Can variability measures in fetal heart rate be non-invasive indicators of autonomic nervous system (ANS) maturation?

Figure #5 shows the distribution of abnormal short-term variability across the three fetal health classes, revealing that normal fetuses (class 1.0) tend to have lower short-term variability values clustered around 20-40, while suspect and pathological cases show broader distributions extending to higher values. This

visualization is particularly valuable because short-term variability is the most correlated feature to fetal health, making it a key indicator for distinguishing between healthy and at-risk pregnancies.

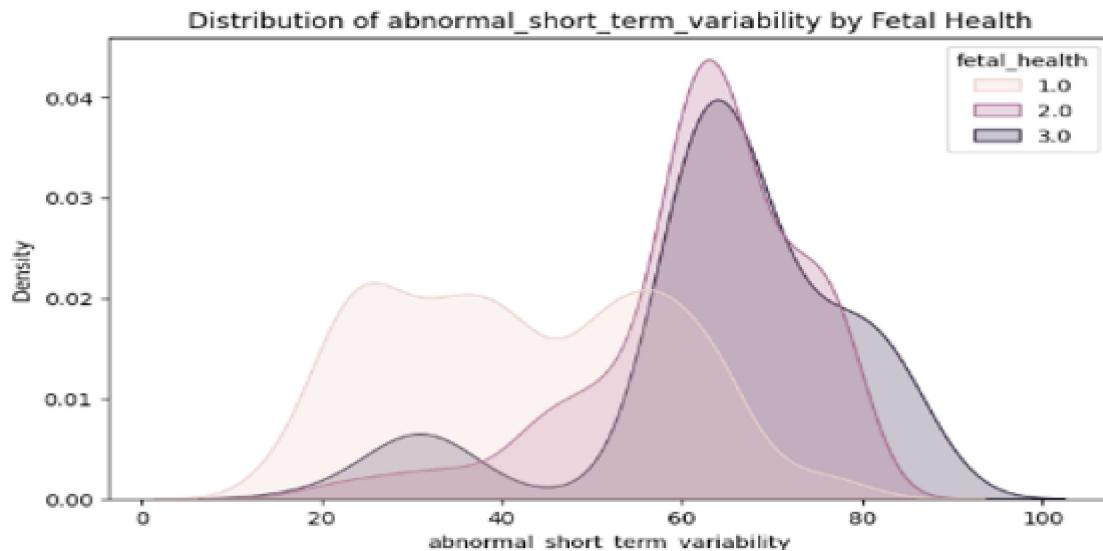


Figure #6 shows the feature importance of variability measures in predicting fetal health. The chart reveals that abnormal short-term variability is by far the most important feature with an accuracy of approximately 40%.

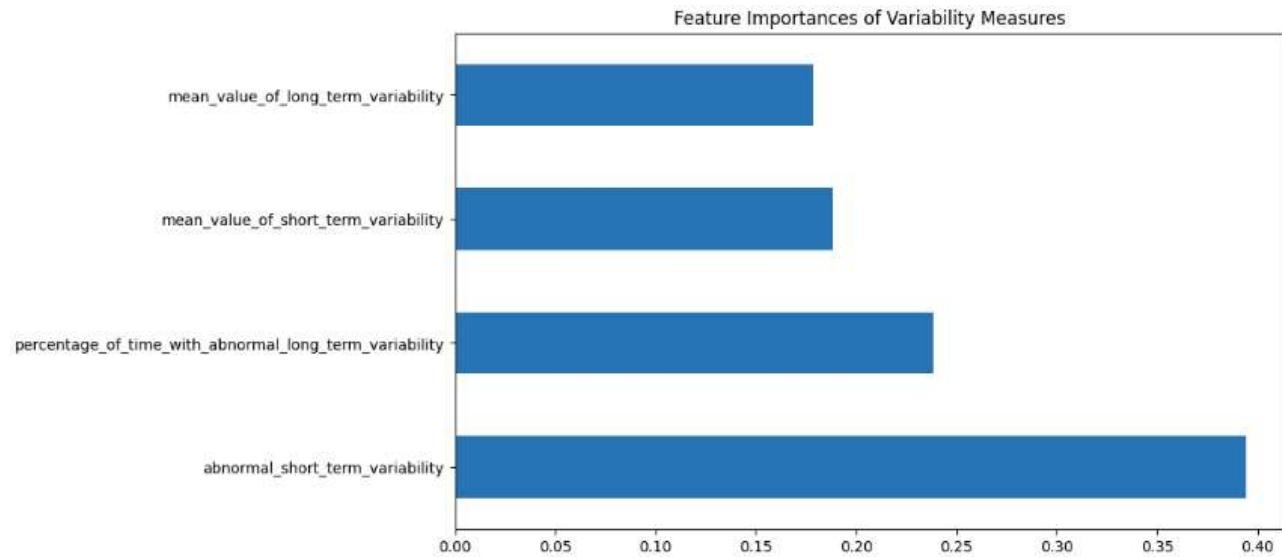


Figure #7 is the classification report scores for the gradient boosting GridSearchCV model using only the variability features. The mean accuracy is 93% which is very close to the 95% from the model trained on all the features, suggesting that variability features are the key predictors of fetal health.

Figure #7 XGR Model Classification Report on Variability Features

Gradient Boosting Classification Report (Variability Features Only):				
	precision	recall	f1-score	support
1.0	0.94	0.97	0.95	333
2.0	0.87	0.75	0.81	64
3.0	0.92	0.79	0.85	29
accuracy			0.93	426
macro avg	0.91	0.84	0.87	426
weighted avg	0.93	0.93	0.93	426

Discussion

Methods The project was implemented in Python using Google Colab. The dataset used required no cleaning. Histogram-based features were removed because other features already captured heart rate information, which helped improve model performance.

Exploratory data analysis was conducted to identify the top five correlated features. The feature "light decelerations" was removed due to its high correlation (56%) with the mean value of short-term variability to reduce multicollinearity and prevent redundant information that could lead to model overfitting.

Five machine learning models were developed and evaluated: K-Nearest Neighbors, Decision Tree, Random Forest, Gradient Boosting, and Gradient Boosting with GridSearch Cross-Validation. To compare model performances, a T-test was performed and showed no statistically significant difference between the models.

All experiments used a random seed of 42 to ensure reproducibility. The best-performing model was further evaluated using stratified K-fold cross-validation with 5 folds, achieving an average accuracy of 90% on unseen data.

For Research Question 1, we addressed it by training and evaluating machine learning models on CTG features, achieving high predictive accuracy and confirming that these measurements can reliably classify fetal health status.

For Research Question 2, we analyzed heart rate variability metrics from the dataset, demonstrating their association with autonomic nervous system maturation and showing that abnormal short-term variability detected by our models may signal fetal ANS dysfunction.

Impact: This work can enable earlier detection of fetal neurological issues, support the development of automated prenatal screening tools, and advance personalized care by integrating machine learning with clinical understanding of fetal development. By reducing reliance on specialized manual interpretation, such models can make fetal health assessment more accessible in low-resource settings and help standardize evaluations across different healthcare providers.

Next Steps: The findings should be validated on larger datasets, implemented within clinical monitoring systems, and expanded to include additional heart rate variability (HRV) parameters to further enhance prediction accuracy.

Code and Data Availability

- The code and the presentation for the project can be found here on Github.
<https://github.com/ShaneBollinger/Fetal-ANS-Health-Prediction-Project>
- The Fetal Health Classification dataset can be found here.
<https://www.kaggle.com/datasets/andrewmvd/fetal-health-classification>.

Acknowledgements

- I would like to acknowledge **Noah Royce** for being my **TA** and for helping me during the development of my project along with providing constructive feedback that greatly improved my project and presentation.
- I would like to acknowledge **Kyla Gabriel**, my **Mentor**, for regularly checking in on my progress and offering guidance throughout the program.
- I would like to acknowledge The Coding School (<https://the-cs.org/>) for giving me the opportunity to be part of this amazing **program** and for teaching me valuable data science skills that will benefit me throughout my career.
- I would like to acknowledge my fellow peers, **Ella Odinec, Kalista Yeba, and Valentina Shahinian**, who were part of the **Reproductive Medicine Research Team** for their collaboration on this project.
- I would like to acknowledge the **Summer Youth Employment Program** for **funding** this project and my efforts in the program.
- I would like to acknowledge the **reference paper** cited below which informed and supported my research. ("Fetal Risk Using Cardiotocographic Data." International Journal of Applied and Basic Medical Research, vol. 9, no. 4, 2019, p. 226, https://doi.org/10.4103/ijabmr.ijabmr_370_18. Accessed 11 Dec. 2021.)
- I would like to acknowledge the **dataset** from Ayres de Campos et al. (2000), SisPorto 2.0: A Program for Automated Analysis of Cardiotocograms, J Matern Fetal Med 5:311–318, which was essential for the development of this project.