Fetal Health Classification from CTG Signals with Explainable Machine Learning

Author: Avi Desai
Course: AI in Healthcare
Date: April 28, 2025

Code:

https://github.com/avidesai/AI395T-High-Risk-Project/blob/main/Code/final_project.ipynb

Slides:

https://github.com/avidesai/AI395T-High-Risk-Project/blob/main/High%20Risk%20Project%20Slides.pdf

Video:

https://utexas.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=8815fb3c-543e-45ce-8315-b2ce003eb639

1. Introduction

Maternal and neonatal mortality remain persistent global challenges. According to the United Nations, approximately 295,000 women died during and following pregnancy and childbirth in 2017, with most deaths occurring in low-resource settings and being largely preventable. Parallelly, child mortality under the age of five remains a critical indicator of human development. Cardiotocography (CTG)—which monitors fetal heart rate, uterine contractions, and fetal movement—is a widely accessible diagnostic tool that can aid in identifying fetal distress early. However, interpreting CTG signals is subjective and highly dependent on clinical expertise.

This project explores the use of machine learning models to automate the classification of fetal health from CTG-extracted features into three categories: Normal, Suspect, and Pathological. We trained an XGBoost classifier and a simple feed-forward neural network on a publicly available CTG dataset. To address class imbalance, we employed SMOTE oversampling. We further leveraged SHAP (SHapley Additive exPlanations) to provide model explainability, ensuring our results are interpretable for clinical decision support.

2. Related Work

Prior research has explored machine learning for CTG classification. Sisodia et al. (2017) demonstrated that support vector machines (SVM) and random forests could achieve over 90% classification accuracy using CTG features. Abirami and Chitra (2019) explored deep learning approaches, particularly convolutional neural networks, to directly process raw CTG signals. Furthermore, Lundberg et al. (2018) introduced SHAP for healthcare applications, showing that interpretable models improve trustworthiness and clinical adoption.

Unlike prior works focusing only on accuracy, this project emphasizes both performance and explainability using ensemble methods combined with SHAP insights.

3. Methodology

3.1 Data

We used the “fetal_health.csv” dataset, comprising 2,126 CTG records with 21 features extracted from fetal monitoring. The target variable was labeled into three classes:

Class 0: Normal
Class 1: Suspect
Class 2: Pathological

3.2 Preprocessing

Data was split into training (80%) and testing (20%) sets using stratified sampling to maintain class distribution. Features were standardized using a StandardScaler.

Given the original class imbalance (77% Normal, 13.8% Suspect, 8.2% Pathological), we applied SMOTE (Synthetic Minority Over-sampling Technique) to the training set to oversample minority classes, achieving class balance and enhancing model fairness.

3.3 Models

XGBoost Classifier:
Configured with default parameters, evaluated using multiclass log loss.

Feed-Forward Neural Network:
Architecture included:

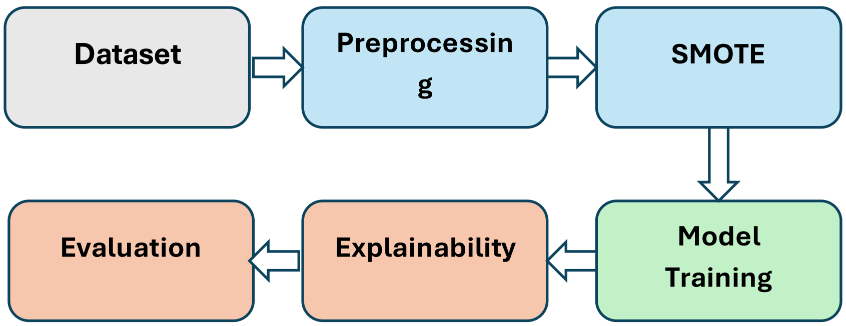
- Input layer (21 features)
- Two hidden layers (64 and 32 neurons, ReLU activations)
- Dropout layer (0.3 dropout rate)
- Output layer (Softmax, 3 neurons)

Both models were trained on the SMOTE-resampled data and evaluated on the untouched original test set.

We used SHAP to compute feature importance for the XGBoost model. A SHAP summary plot was generated to visualize global feature contributions for each fetal health class.

3.5 Workflow

Dataset → Preprocessing → SMOTE → Model Training → Explainability → Evaluation



4. Results

4.1 Model Performance

XGBoost Classifier:

Accuracy: 93.9%
Macro F1 Score: 88%

Class-wise Breakdown:
  Normal: Precision 96%, Recall 98%, F1 97%
  Suspect: Precision 86%, Recall 75%, F1 80%
  Pathological: Precision 84%, Recall 91%, F1 88%

Neural Network:

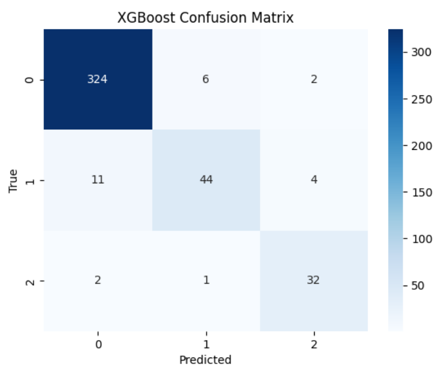
Accuracy: 89.9%
Macro F1 Score: 84%

Class-wise Breakdown:

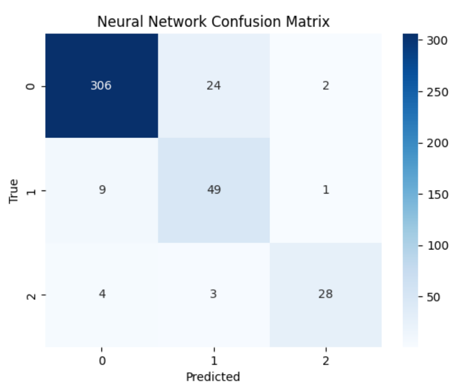
Normal: Precision 96%, Recall 92%, F1 94%
Suspect: Precision 64%, Recall 83%, F1 73%
Pathological: Precision 90%, Recall 80%, F1 85%

4.2 Confusion Matrices

XGBoost Confusion Matrix:



Neural Network Confusion Matrix:



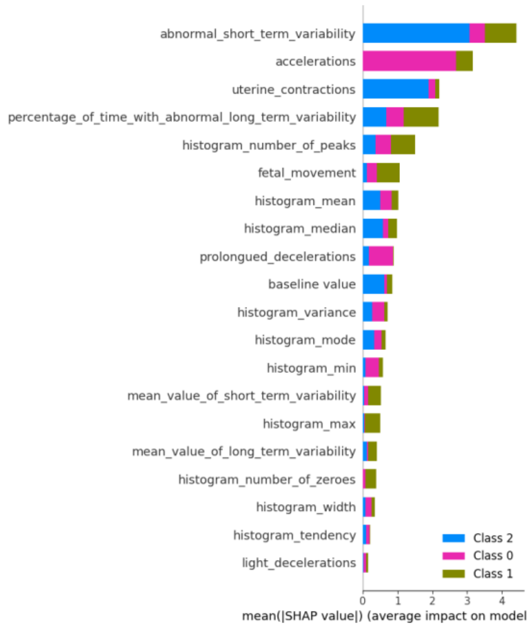
4.3 Explainability Results (SHAP)

The SHAP analysis revealed the top influential features:

Abnormal Short Term Variability
Accelerations
Uterine Contractions

Percentage of Time with Abnormal Long Term Variability
Histogram Number of Peaks

These features align with clinical understanding of fetal distress, validating the model’s focus on medically relevant attributes.



5. Conclusion

This project demonstrated the feasibility of classifying fetal health status from CTG-derived features using machine learning models. XGBoost, aided by SMOTE balancing, achieved 93.9% accuracy, outperforming the neural network slightly. Importantly, SHAP analysis provided transparent explanations of model decisions, a critical requirement for clinical adoption.

Limitations:

- Small dataset size (2,126 records) may limit generalizability.
- No external validation dataset used.
- Neural network performance was more variable, suggesting need for hyperparameter tuning.

Future Directions:

- Explore more advanced ensemble techniques and hyperparameter optimization.
- Collect larger, real-world CTG datasets.
- Develop a clinician-facing interactive dashboard integrating explainable model outputs.

Overall, our study shows that interpretable machine learning can assist in CTG interpretation, with potential to improve outcomes in maternal-fetal medicine.

References

[1] Sisodia, D., Sisodia, D. S. (2017). Prediction of fetal health status using machine learning. Procedia Computer Science, 132, 1033-1040.

[2] Abirami, S., Chitra, P. (2019). Prediction of fetal health status using deep learning techniques. Materials Today: Proceedings.

[3] Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. Advances in Neural Information Processing Systems (NeurIPS).