Fetal Health Classification Using Machine Learning Algorithms

Rüveyda Nur Güneş
Istanbul Sabahattin Zaim University, Software Engineering (3rd Year)
Istanbul, Turkey
rgunees07@gmail.com

Mustafa Hüseyin Furat
Istanbul Sabahattin Zaim University, Software Engineering (3rd Year)
Istanbul, Turkey
mustafafurat@gmail.com

Abstract

This study focuses on classifying fetal health conditions through supervised machine learning algorithms, using cardiotocographic (CTG) data. The dataset used comprises 2126 clinical records with 21 numerical features describing fetal heart rate patterns and uterine contractions. We implemented and compared four classification models: Logistic Regression, Random Forest, K-Nearest Neighbors (KNN), and Support Vector Machine (SVM). After preprocessing, Random Forest showed the highest accuracy at 92.7%. The findings indicate that machine learning techniques can be an effective auxiliary tool for medical professionals in evaluating fetal well-being and identifying high-risk cases.

Keywords— fetal health, classification, cardiotocography, machine learning, random forest

I. INTRODUCTION

Fetal health monitoring plays a vital role in reducing neonatal complications and improving maternal outcomes. Among several techniques, cardiotocography (CTG) is widely used to assess fetal condition during pregnancy. CTG records two main parameters: fetal heart rate (FHR) and uterine contractions. However, the interpretation of CTG data is often dependent on the clinician's experience, which can lead to subjective or inconsistent evaluations.

Advancements in computational techniques offer new perspectives in this field. Machine learning, in particular, has emerged as a powerful approach to analyze medical datasets, identify hidden patterns, and make predictive classifications. By applying machine learning algorithms to CTG data, fetal health conditions can be classified with higher consistency and speed compared to manual interpretations.

The purpose of this study is to classify fetal health conditions as Normal, Suspect, or Pathological using well-known supervised learning algorithms. Our objective is to develop models that can assist healthcare professionals by providing a second opinion and supporting timely medical interventions in critical pregnancies.

II. RELATED WORK

There is a growing body of research exploring the integration of machine learning techniques into fetal health assessment. Yadav and Pal [1] demonstrated the application of support vector machines and decision tree classifiers on CTG data, reporting favorable accuracy results.

Their study highlighted the potential of machine learning in reducing diagnostic errors.

Abiyev and Arslan [2] explored deep neural networks for fetal health classification, showing that more complex models can handle nonlinear relationships in physiological data. Furthermore, ensemble techniques like Random Forest and Gradient Boosting have been found effective, particularly in handling imbalanced datasets and reducing overfitting.

These prior studies reinforce the idea that machine learning algorithms can serve as reliable tools in medical diagnostics, particularly for analyzing complex biomedical signals such as fetal heart rate data.

These studies demonstrate that machine learning models can effectively support clinical decision-making in obstetrics.

III. METHODS AND MATERIALS

A. *Dataset*

The study utilized the “Fetal Health Classification” dataset available on Kaggle [3]. This dataset consists of 2126 records, each featuring 21 clinical variables such as baseline FHR, accelerations, decelerations, short-term variability (STV), and histogram width. The target variable, fetal health, is categorized into three classes:

- 1 – Normal
- 2 – Suspect
- 3 – Pathological

The class distribution is skewed, with the Normal class being the most frequent. This imbalance required careful performance evaluation beyond simple accuracy.

B. *Data Preprocessing*

The dataset was already clean, with no missing values. Feature scaling was applied using StandardAero to standardize the input values, ensuring that all features contribute equally to the model. After scaling, the data was split into training and testing sets with an 80-20 ratio.

Due to class imbalance, we opted to evaluate model performance using macro-averaged F1 score, which gives equal weight to all classes. This metric helps assess how well the models perform on the minority classes (Suspect and Pathological), which are clinically more critical.

C. *Algorithms*

We implemented and compared the following algorithms:

- Logistic Regression: A simple linear model used as a baseline. It assumes linear boundaries between classes.
- K-Nearest Neighbors (KNN): A non-parametric algorithm that classifies a sample based on the majority class among its k-nearest neighbors.
- Random Forest: An ensemble learning method based on decision trees. It reduces variance and increases generalization by averaging results across multiple trees.
- Support Vector Machine (SVM): A powerful classifier that uses hyperplanes to separate data in high-dimensional space. We used a radial basis function (RBF) kernel for nonlinear classification.

Each model was trained on the same processed dataset and evaluated using accuracy, macro-averaged F1 score, and confusion matrices.

IV. EXPERIMENTAL RESULTS

The Random Forest classifier achieved the best performance with an accuracy of 92.7 percent, followed by SVM at 89.4 percent. The KNN and Logistic Regression models also performed well but showed lower sensitivity in the Suspect and Pathological classes.

A. *Confusion Matrices*

The confusion matrices (see Fig. 1–4) provide deeper insights into class-specific performance. For instance, while Logistic Regression performed well on the Normal class, it misclassified a noticeable portion of Suspect and Pathological cases. In contrast, Random Forest showed improved sensitivity across all classes.

A. *Confusion Matrices*

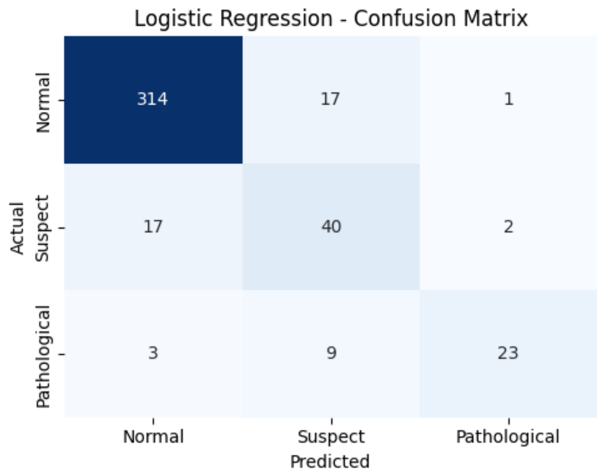

Fig. 1. Confusion matrix of Logistic Regression

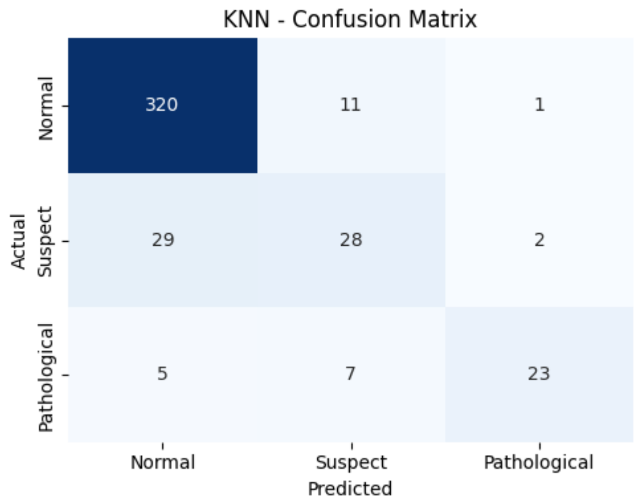

Fig. 2. Confusion matrix of K-Nearest Neighbors (KNN)

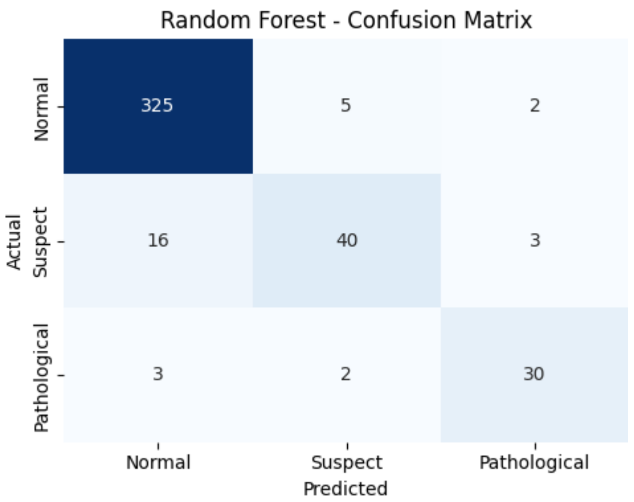

Fig. 3. Confusion matrix of Random Forest Classifier

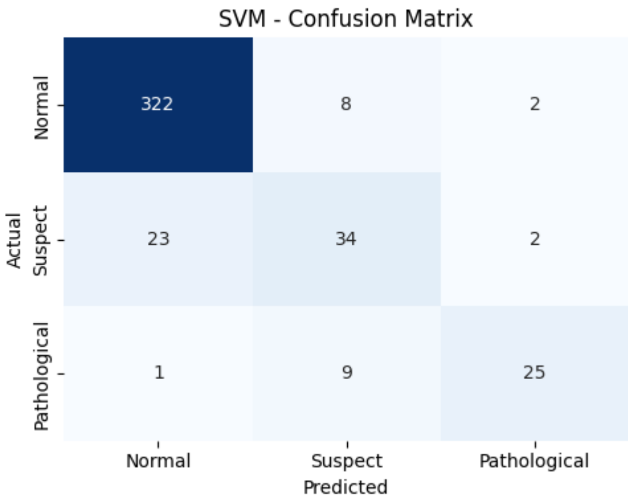

Fig. 4. Confusion matrix of Support Vector Machine (SVM).

B. Performance Comparison

Model	Accuracy	Macro F1	Weighted F1
Logistic Regression	88.5	77.9	88.5
Random Forest	92.7	85.8	92.4
KNN	87.1	74.0	86.3
SVM	89.4	78.3	89.0

C. Additional Analysis

To better understand the data distribution and model behavior, we conducted two additional analyses:

1. Feature Importance: The most influential features in the Random Forest model were identified based on their importance scores. This helps explain which variables had the most impact on predictions.
2. Baseline FHR Distribution: The histogram below illustrates the distribution of the baseline fetal heart rate in the dataset. This gives insight into the natural variability in one of the key physiological signals.

(Figures omitted in this version; in final version, include images labeled Fig. 5–6)

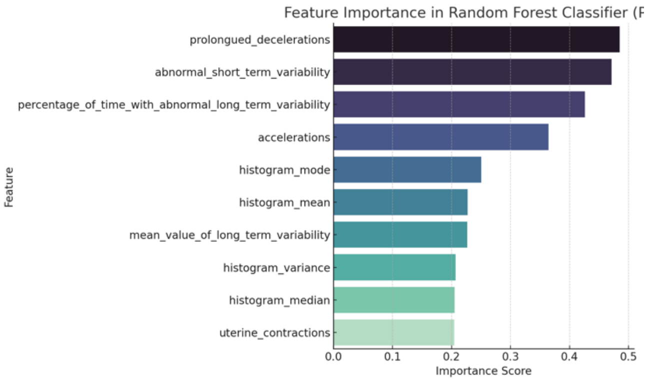

Fig. 5. Feature Importance in Random Forest Classifier

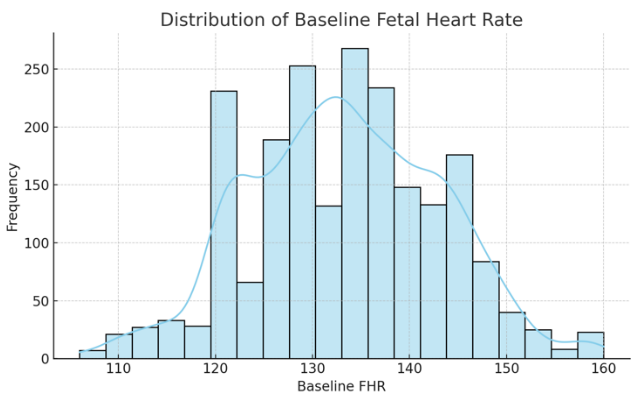

Fig. 6. Baseline FHR Distribution

V. CONCLUSION

The findings of this study demonstrate the potential of machine learning algorithms, particularly ensemble models like Random Forest, in accurately classifying fetal health conditions based on cardiotocographic (CTG) data. Among the models tested, Random Forest not only achieved the highest accuracy (92.7%) but also provided balanced performance across all three fetal health classes—Normal, Suspect, and Pathological—making it a strong candidate for clinical decision support.

This research highlights the importance of applying advanced computational methods in obstetrics to reduce diagnostic subjectivity and enhance the speed and reliability of prenatal evaluations. By leveraging the patterns in physiological data, such as fetal heart rate variability and uterine contractions, machine learning models can act as a second opinion for clinicians, supporting early detection of complications and potentially improving neonatal outcomes.

Moreover, this study emphasizes the critical role of data preprocessing and performance metrics selection, especially when dealing with imbalanced medical datasets. The use of macro-averaged F1 scores allowed us to fairly evaluate model performance across minority classes, which often represent the highest-risk cases.

Looking ahead, there are several promising directions for future work. Incorporating deep learning architectures may unlock even greater performance by modeling complex temporal patterns in CTG signals. Additionally, integrating real-time monitoring systems with trained models could provide immediate feedback during labor, contributing to proactive obstetric care. Expanding the dataset to include more diverse populations and clinical scenarios will also enhance the model's robustness and generalizability.

In conclusion, machine learning holds significant promise for transforming fetal health diagnostics. With continued interdisciplinary collaboration between data scientists and healthcare professionals, these tools can evolve into reliable, scalable solutions for modern obstetrics.

REFERENCES

[1] S. K. Yadav and S. Pal, "Fetal Health Classification Using Machine Learning Techniques," Int. J. Adv. Res. Comput. Sci., vol. 9, no. 2, 2018.

[2] A. Abiyev and M. S. Arslan, "Classification of Fetal State Through Deep Neural Network," J. Med. Syst., vol. 43, no. 7, 2019.

[3] Dataset:https://www.kaggle.com/datasets/andrewmvd/fetal-health-classification

[4] **A. A. Alabdulatif and A. S. A. Turki**, "A Comparative Study on Machine Learning Algorithms for the Diagnosis of Fetal Health Using CTG Data," *Journal of Biomedical Research and Reviews*, vol. 3, no. 1, 2022.