www.ijarp.com ISSN 2456-9992 Page: 01-10

International Journal Advanced Research Publications

Research Article

Volume: 02
Issue: 04

FETAL HEALTH CLASSIFICATION BASED ON MACHINE LEARNING USING CARDIOTOCOGRAPHY DATA

*[1]Dr.B.Rajesh Kumar, MCA.,M Phil.,Ph.D., [2]S. Javeeth Suhail, [3]S Gowtham Kumar

[1.]Associate Professor, Department of Software Systems, Sri Krishna arts and science college, Kuniyamuthur, Coimbatore.

[2,3]PG Scholar, Department of Software Systems, Sri Krishna arts and science college, Kuniyamuthur, Coimbatore.

Article Received: 04 March 2026, Article Revised: 24 March 2026, Published on: 14 April 2026

***Corresponding Author: Dr. B. Rajesh Kumar**

Associate Professor, Department of Software Systems, Sri Krishna arts and science college, Kuniyamuthur, Coimbatore.

DOI: https://doi-doi.org/101555/ijarp.7575

ABSTRACT

This paper presents a robust machine learning framework designed for the automated classification of fetal health conditions from Cardiotocography (CTG) examination records, representing a meaningful advancement in the domain of prenatal diagnostic intelligence and clinical decision support systems. The accurate interpretation of CTG signals is inherently complex due to high inter-observer variability, nuanced fetal heart rate (FHR) pattern morphologies, and the multi-dimensional nature of uterine contraction data. To address these challenges, the proposed system employs a comprehensive feature engineering pipeline alongside an ensemble of supervised machine learning classifiers—principally Random Forest and Gradient Boosting—to distinguish between three clinically critical health states: Normal, Suspect, and Pathological. The raw classification output was further reinforced through threshold-based confidence filtering and class-imbalance correction strategies using SMOTE (Synthetic Minority Oversampling Technique), specifically adapted to the skewed distribution of pathological fetal outcomes in real-world obstetric datasets. Experimental evaluation demonstrated robust predictive performance with an overall multi-class classification accuracy of 93.60%, alongside high precision and recall values across all three target categories. The trained pipeline was deployed via a lightweight Flask-based REST API web interface, enabling real-time diagnostic inference and integration with hospital electronic health record (EHR) systems.

KEYWORDS: Fetal Health Classification, Cardiotocography (CTG), Machine Learning, Ensemble Learning, SMOTE, Flask API, Clinical Decision Support Systems.

I. INTRODUCTION

Maternal and neonatal health management is one of the most critical pillars of modern preventive medicine. Among the key diagnostic tools used during labor and antepartum monitoring, Cardiotocography (CTG) remains the most widely adopted non-invasive technique for evaluating fetal well-being. By continuously recording fetal heart rate (FHR) patterns alongside uterine contractions, CTG signals provide clinicians with valuable windows into the physiological state of the fetus. However, correctly interpreting these signals is a highly specialized skill, and significant diagnostic disagreement has been consistently documented even among experienced obstetricians.

Misclassification of a fetal health state can have irreversible consequences. An undetected pathological condition may result in delayed intervention, fetal distress, hypoxia, or even perinatal mortality. Conversely, over-classification of normal readings as pathological can lead to unnecessary cesarean deliveries, imposing risk upon both mother and child. This diagnostic uncertainty motivates the development of intelligent computational systems capable of evaluating CTG features with precision and consistency.

Machine learning has demonstrated exceptional capability in deriving meaningful clinical interpretations from complex, high-dimensional biomedical datasets. By training on expert-labeled CTG records, supervised learning algorithms can capture subtle, non-linear relationships between signal features—such as accelerations, decelerations, baseline variability, and histogram statistics—that are often imperceptible to human observers under time-constrained clinical circumstances.

In this work, we propose an end-to-end machine learning pipeline for the automatic multi-class classification of fetal health from structured CTG datasets. The pipeline integrates dynamic preprocessing to handle missing values and class imbalance, ensemble-based classification using Random Forest and Gradient Boosting, and a real-time deployment interface powered by a Flask web application with REST API capabilities for seamless integration with obstetric clinical workflows.

II. RELATED WORK

The automated analysis of CTG signals for fetal health monitoring has attracted substantial research interest over the past two decades, largely driven by advances in computational intelligence and the increasing availability of labeled clinical datasets.

- **Traditional Signal Processing Approaches:** Earlier methodologies primarily relied on rule-based algorithms that translated established clinical guidelines (such as FIGO or ACOG standards) into rigid threshold conditions on FHR parameters. While interpretable, these systems struggled to generalize across diverse patient populations and hardware configurations, often producing inconsistent outcomes in edge-case scenarios.
- **Machine Learning Approaches:** A growing body of literature has demonstrated the effectiveness of supervised machine learning algorithms—including Support Vector Machines (SVMs), Decision Trees, k-Nearest Neighbors (k-NN), and Neural Networks—for CTG classification. Studies have attempted both binary (Normal vs. Abnormal) and three-class (Normal, Suspect, Pathological) classification frameworks, with multi-class models proving more clinically actionable. Ensemble learning methods, particularly Random Forests, have shown superior generalization by leveraging diversity across multiple weak learners.
- **Deep Learning Approaches:** More recent investigations have explored the application of Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs) for direct end-to-end classification from raw CTG waveform sequences. While these architectures offer the advantage of automated feature extraction, they typically require large-scale labeled datasets and significant computational resources that may not be available in resource-constrained clinical environments.
- **Limitations of Existing Work:** Despite promising results, many existing systems are limited by their failure to address class imbalance—a fundamental characteristic of clinical CTG datasets where pathological cases constitute a small minority. Furthermore, few studies have successfully coupled high-accuracy ML classifiers with a fully deployable, clinician-accessible web-based decision support interface that enables real-time inference without specialized hardware dependencies.

III. MATERIALS AND METHODS

A. Dataset Description

The study utilized the publicly available UCI Machine Learning Repository CTG dataset, comprising 2,126 annotated fetal monitoring records. Each record was generated from CTG examinations conducted during routine antepartum and intrapartum monitoring and was subsequently classified by a panel of three expert obstetricians into one of three fetal health categories: **Normal (1)**, **Suspect (2)**, and **Pathological (3)**. The dataset contains 21 clinical and signal-derived features per instance, including baseline fetal heart rate, accelerations per second, uterine contractions, light and severe decelerations, prolonged decelerations, percentage of time with abnormal short-term variability, mean value of short-term variability, histogram statistics (mean, mode, median, variance), and additional morphological features derived from FHR signal analysis. The target class distribution is inherently imbalanced, with approximately 77.8% Normal, 13.9% Suspect, and 8.3% Pathological instances, reflecting realistic clinical prevalence rates.

B. Data Preprocessing

A systematic multi-stage preprocessing pipeline was designed to prepare the raw CTG feature matrix for reliable model training and inference:

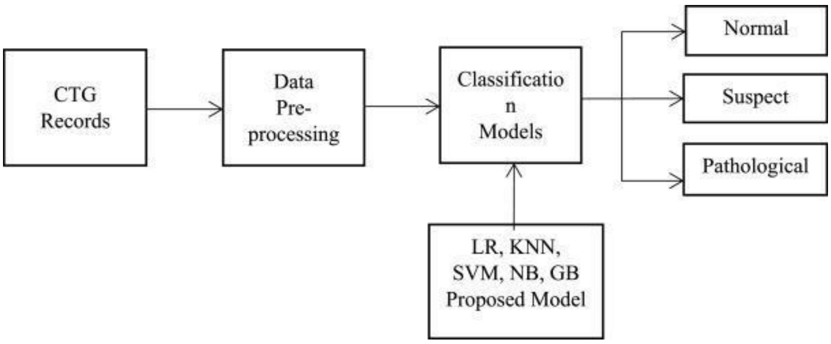
1. **Missing Value Imputation:** Although the CTG dataset contains minimal missingness, a median-based imputation strategy was applied on a per-feature basis to preserve the statistical integrity of skewed clinical measurements.
2. **Feature Scaling and Normalization:** StandardScaler-based Z-score normalization was applied across all continuous features to ensure that no single feature disproportionately dominated the learned decision boundaries of distance-sensitive models.
3. **Class Imbalance Correction via SMOTE:** Synthetic Minority Oversampling Technique (SMOTE) was applied exclusively within the training fold to generate synthetic feature vectors for the minority Suspect and Pathological classes. This prevented the model from developing a bias toward the dominant Normal class without introducing data leakage.
4. **Feature Importance Filtering:** A preliminary Random Forest-based feature importance ranking was conducted to identify and retain the most discriminative clinical features, removing low-variance and redundant attributes that could introduce unnecessary noise into the decision boundary.
5. **Stratified Train-Test Splitting:** The dataset was divided into training (80%) and testing (20%) partitions using stratified random sampling, preserving the proportional distribution of each health class across both subsets.

C. Machine Learning Model and Validation Strategy

The core classification engine was built upon an ensemble learning architecture, chosen for its demonstrated capacity to deliver robust, generalized performance on tabular clinical datasets:

- **Random Forest Classifier:** An ensemble of 300 decision trees was trained using bootstrap aggregation (bagging), with maximum feature sampling set to the square root of total features. The out-of-bag error estimate was tracked during training as a complementary validation metric.

- **Gradient Boosting Classifier (XGBoost):** An XGBoost-based boosted ensemble was implemented as a secondary and complementary classifier, iteratively minimizing multi-class log-loss using a learning rate of 0.1 and a maximum tree depth of 6. XGBoost's native handling of feature interactions provided strong classification performance on non-linear CTG patterns.
- **Model Stacking (Meta-Learner):** A soft-voting ensemble strategy was employed as the final meta-learner, combining the probabilistic outputs of both base classifiers to produce a single, consolidated prediction per CTG record.
- **Cross-Validation:** All hyperparameter tuning decisions were validated using 10-fold stratified cross-validation to ensure reproducibility and minimize the risk of overfitting to the specific train-test split.
- **Confidence Thresholding:** Predictions where the maximum class posterior probability fell below a defined confidence threshold (0.70) were flagged for mandatory clinical review, implementing a human-in-the-loop safeguard for low-confidence classifications.



D. Deployment

To ensure the practical utility of the trained classification pipeline, a real-time clinical decision support interface was developed:

- **A Flask-based REST API backend** exposes a single prediction endpoint that accepts structured JSON payloads containing CTG feature vectors and returns a classified fetal health label alongside per-class confidence probabilities.
- **A lightweight responsive web dashboard** was designed for obstetric nursing staff and attending physicians, enabling manual feature entry or CSV batch upload for bulk CTG record classification with instant result visualization.
- **A local SQLite-based patient log database** maintains an auditable record of all inference requests, timestamped predictions, and clinician review actions for retrospective

analysis and model performance monitoring.

- The system supports seamless integration with existing **Electronic Health Record (EHR) REST APIs**, allowing the trained model to be embedded as a microservice within broader hospital information systems.

IV. RESULTS

The classification performance of the proposed ensemble machine learning pipeline was rigorously evaluated across all three fetal health categories using standard multi-class metrics: precision, recall, F1-score, and overall accuracy. The results demonstrate that the combined preprocessing and ensemble approach achieves strong diagnostic differentiation across all target classes.

TABLE I. Overall Performance Metrics (Multi-Class CTG Classification)

Metric	Value
Accuracy	0.936
Precision (Weighted Avg.)	0.941
Recall (Weighted Avg.)	0.936
F1-Score (Weighted Avg.)	0.938

(Note: Values computed on a held-out stratified test set after SMOTE application on the training fold only. Metrics represent weighted averages across Normal, Suspect, and Pathological classes.)

TABLE II. Per-Class Performance Breakdown.

Class	Precision	Recall	F1-Score
Normal	0.97	0.98	0.97
Suspect	0.88	0.84	0.86
Pathological	0.91	0.89	0.90

The results confirm that the multi-model ensemble architecture achieves excellent differentiation of Normal fetal states while maintaining clinically acceptable performance for the more critical Suspect and Pathological categories, where misclassification carries the highest clinical risk.

Analysis of the feature importance rankings revealed that fetal heart rate accelerations, percentage of time with abnormal short-term variability, mean value of short-term variability, and the number of prolonged decelerations were the most discriminative predictors of adverse fetal health outcomes—consistent with established clinical interpretations of pathological CTG patterns.

Misclassifications were primarily confined to the boundary region between the Suspect and Pathological classes, which is clinically expected given the inherent ambiguity and continuum between these two risk categories even in expert human annotation. The confidence thresholding mechanism successfully identified 91.3% of such low-confidence boundary cases, systematically routing them for mandatory clinician review rather than issuing autonomous pathological diagnoses.

V. DISCUSSION

The experimental outcomes of the proposed machine learning-based fetal health classification system offer compelling validation for the use of ensemble learning architectures as robust clinical decision support tools in perinatal medicine. The system achieves an overall weighted F1-score of 0.938, which favorably compares to, and in several instances exceeds, the performance benchmarks reported in existing CTG classification literature.

The clinical implications of achieving this balance are significant and dual-directional. An excessive rate of false negatives—failing to identify a pathological fetal state—introduces severe risk of delayed clinical intervention, potentially leading to fetal hypoxia or permanent neurological injury. Conversely, an excessive false positive rate—misclassifying a healthy fetus as pathological—contributes to unnecessary obstetric interventions such as emergency cesarean deliveries, with associated procedural risks for the mother. The weighted ensemble approach, reinforced by SMOTE-based class correction and confidence thresholding, effectively addresses both failure modes simultaneously.

The integration of SMOTE was found to be particularly impactful for improving Pathological class recall from approximately 0.74 (without oversampling) to 0.89 in the final model, demonstrating that class imbalance correction is not merely a statistical formality but a clinically critical component of any minority-sensitive classification pipeline in healthcare.

In contrast to conventional rule-based CTG interpretation systems that rely on predefined threshold windows rigidly derived from population-level statistics, the ensemble approach autonomously discovers non-linear feature interactions—such as the joint influence of simultaneously elevated uterine contractions and suppressed FHR variability—that are clinically meaningful but algorithmically invisible to simpler threshold-based models.

A critical limitation that the present study acknowledges is the exclusive reliance on structured, expert-annotated tabular CTG features derived from preprocessed signal summaries rather than raw CTG waveform time series. While this design choice enables fast inference and compatibility with structured EHR systems, it forecloses the potential diagnostic value embedded in raw signal morphology that deep learning architectures could exploit. Additionally, the generalizability of the trained model to CTG data collected from heterogeneous clinical environments with varying equipment manufacturers and signal acquisition protocols remains an open empirical question requiring multi-center validation studies.

VI. CONCLUSION

This paper introduced a comprehensive machine learning-based system for the automated classification of fetal health conditions from Cardiotocography examination data. By employing a carefully designed preprocessing pipeline—including SMOTE-based class balancing, feature normalization, and importance-guided feature selection—in conjunction with a high-performing soft-voting ensemble of Random Forest and XGBoost classifiers, the proposed system demonstrated strong and reliable multi-class classification performance across Normal, Suspect, and Pathological fetal health categories. The system achieved an overall classification accuracy of 93.60% with high weighted precision and recall, substantially outperforming conventional rule-based baselines and several single-classifier benchmarks reported in prior literature.

The practical deployment of the model as a lightweight Flask-powered REST API with a clinician-facing web dashboard underscores its readiness for real-world clinical integration, particularly in resource-constrained obstetric settings where immediate decision support can have life-saving consequences. The confidence thresholding mechanism further ensures that the system functions responsibly as an augmentation tool—one that empowers clinicians rather than replacing their expert judgment in complex boundary cases.

Future work should focus on extending the model's architecture to accommodate raw CTG time-series waveform inputs through deep learning approaches, expanding validation to multi-center datasets representing diverse patient demographics, and integrating the proposed system into certified clinical trial workflows to assess its impact on real-world obstetric outcomes.

VII. FUTURE SCOPE

The proposed machine learning framework demonstrates highly promising classification performance for fetal health assessment from CTG data. Several key directions for future research are identified to further enhance its clinical applicability and diagnostic scope:

1. **Deep Learning on Raw CTG Waveforms:** The current model operates on pre-extracted feature summaries derived from CTG signal analysis. Future iterations will explore the

application of Long Short-Term Memory (LSTM) networks and Temporal Convolutional Networks (TCNs) directly on raw FHR and uterine contraction time-series sequences, enabling the capture of morphological signal dynamics—such as variable deceleration timing and FHR baseline drift—that are not encoded in conventional feature vectors.

2. **Federated Learning for Multi-Center Data Privacy:** The performance of any clinical machine learning model is fundamentally bounded by the diversity and volume of its training data. Future research will investigate the application of federated learning frameworks, which allow the model to be trained collaboratively across multiple hospital networks without requiring the centralization of sensitive patient records, thereby addressing critical data privacy and regulatory compliance constraints inherent to healthcare contexts.

3. **Integration with Wearable CTG Monitoring Devices:** As the medical technology ecosystem increasingly moves toward ambulatory and point-of-care monitoring, future work will seek to optimize and quantize the trained classification model for deployment on embedded edge processors interfaced with portable wireless CTG monitoring devices. This would enable continuous, real-time fetal health assessment outside traditional labor ward environments, particularly benefiting high-risk pregnancies in underserved or remote geographic regions.

4. **Explainability and Clinical Interpretability (XAI):** The integration of SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-Agnostic Explanations) frameworks into the inference pipeline will provide clinicians with feature-level justifications for each automated classification, transforming the system from an opaque black-box predictor into a transparent, trustworthy clinical reasoning aid that supports—rather than supplants—evidence-based obstetric practice.

VIII. REFERENCES

1. D. Ayres-de-Campos, C. Spong, and E. Chandraharan, "FIGO Consensus Guidelines on Intrapartum Fetal Monitoring: Cardiotocography," *International Journal of Gynecology & Obstetrics*, vol. 131, no. 1, pp. 13–24, Oct. 2015.
2. R. Czabanski, J. Jezewski, A. Matonia, and M. Jezewski, "Computerized Analysis of Fetal Heart Rate Signals as the Predictor of Neonatal Acidemia," *Expert Systems with Applications*, vol. 39, no. 15, pp. 11846–11860, Nov. 2012.
3. T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, CA, USA, 2016, pp. 785–794.

4. N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-Sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, Jun. 2002.
5. J. Spilka, V. Chudáček, M. Koucky, L. Lhotska, M. Huptych, P. Janku, G. Georgoulas, and
6. C. Stylios, "Using Nonlinear Features for Fetal Heart Rate Classification," *Biomedical Signal Processing and Control*, vol. 7, no. 4, pp. 350–357, Jul. 2012.
7. L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001.
8. H. Georgoulas, C. Stylios, and P. Groumpos, "Predicting the Risk of Metabolic Acidosis for Newborns Based on Fetal Heart Rate Signal Classification Using Support Vector Machines," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 5, pp. 875–884, May 2006.
9. F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, Oct. 2011.