Journal of Engineering Technology and Applied Physics

Early Detection of Fetal Distress using CTG and Machine Learning to Improve Maternal and Child Health

Md Farman Ali[1, *], Kamrul Hasan[2], Md Ikramul Haque[2], Md Serajun Nabi[3], Md Tasin Tazwar[4], Mohammad Hossain[1] and Jahid Hassan Noor[1]

[1]*School of Computing and Informatics, Albukhary International University, Kedah, Malaysia.*
[2]*Computer Science and Engineering, Dhaka University of Engineering & Technology, Gazipur, Bangladesh.*
[3]*Faculty of AI and Engineering, Multimedia University, Cyberjaya, Malaysia.*
[4]*Department of Electrical Engineering and Computer Science, Florida Atlantic University, Florida, United States of America.*
[*]*Corresponding author*: farman.ali@student.aiu.edu.my, *ORCiD*: 0009-0007-6731-322X
https://doi.org/10.33093/jetap.2026.8.1.9
Manuscript Received:2 July 2025, Revised: 1 September 2025, Accepted: 25 November 2025, Published: 15 March 2026

***Abstract*—This study addresses the critical challenge of fetal distress identification, with either maternal or neonatal death coming first in this case. Manual CTG interpretation, being subjective, leads to disadvantages, and a high degree of inter-observer variability may result in misdiagnosis or late clinical intervention with detrimental consequences. Thus, our research fills the gap wherein there has been very little comprehensive and rigorous comparative analysis of a large number of machine learning models. We systematically studied a wide and diverse range of classifiers in our comparison, including tree-based ensembles such as XGBoost, Random Forest, CatBoost, and LightGBM, but also Gradient Boosting, Decision Tree, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Logistic Regression, AdaBoost, Naive Bayes, and Convolutional Neural Networks (CNN). These models were trained and tested on a very large and open-source CTG dataset to validate their predictive power. Our findings reveal that the XGBoost model demonstrated superior performance with an impressive accuracy of 99.04%, while CatBoost, LightGBM, and Random Forest, which had intense predictive powers well above traditional diagnostic means. The above-mentioned accuracy-driven models have found their aptitude in capturing highly complicated nonlinear patterns occurring in CTG data and therefore hold a promising prospect to be developed and applied toward a large-scale and automated diagnostic aid. The successful implementation of these novel techniques would have huge potential for improving the quality of prenatal care and clinical decision-making in resource-poor areas, where expert supervision may be widely scarce.**

Keywords*—Cardiotocography, Fetal health, Classification, Machine learning models, Maternal health.

I. INTRODUCTION

A. Background

The realm of fetal health monitoring has benefited from ML techniques for better accuracy and early detection in recent developments. Almadi et al. experimented with different ML classifiers on cardiotocography (CTG) data to monitor fetal health as Normal, Suspect, and Pathological, achieving the highest accuracy of 92 percent with the Random Forest, thereby showing the potential of data-driven decisions in obstetric care [1]. Mondal et al. presented an ensemble learning-based architecture encompassing stacking, bagging, and voting mechanisms after hyperparameter tuning of models, which led to an impressive classification accuracy of 99.78 percent, solving the problem of data imbalance in an advanced manner with sampling methodologies such as SMOTE and SMOTE-ENN [2]. Similarly, Puri and Reddy developed ML-based predictive models based on CTG data and showed that ensemble methods could outperform individual classifiers, with an accuracy of almost 99 percent, stressing the significance of automation in limiting diagnostic errors in clinical environments [3]. Adding to these efforts, Olayemi and Olasehinde centered their attention on feature selection and performance evaluation of different ML models on CTG datasets

https://doi.org/10.33093/jetap.2026.8.1.9

Published by MMU PRESS. URL: https://journals.mmupress.com/index.php/jetap/index

and have claimed that the Categorical Boosting model with selected features gave the highest MCC score of 0.6321, thereby reinforcing the need for careful feature engineering and robust evaluation metrics for predicting fetal health in resource-constrained scenarios [4].

Child and maternal mortality continue to present immense challenges to global health, especially in low-resource settings where about 94% of pregnancy-related deaths occur, mostly preventable if timely medical monitoring is ensured. Foremost among these strategies for the implementation of Sustainable Development Goals by the UN, upon which these deaths are slated to be reduced by 2030, is the accurate assessment of fetal distress. CTG monitors fetal heartbeat and movements and uterine contractions in a continuous manner and hence is non-invasive and very cheap, providing a whole lot of time for health workers to intervene and solve complications while they are still early. The study tries to support machine learning models on CTG data in order to create new methods for early detection of fetal distress, thus expediting clinical decision-making, with the main aim of lowering preventable maternal and neonatal mortality.

B. Problem Statement

Fetal hypoxia during labor can bring forth severe neonatal complications, including cerebral palsy, stillbirth, and delayed development. Though CTG is a brilliant monitoring device, interpreting it is subjective and highly a poor positive predictive value. Existing ML research over CTG data intends to enhance diagnostic accuracy but still suffers from all kinds of methodological problems, such as synergetics with scarce datasets, an arbitrarily defined notion of hypoxia, limited external validation, and non-interpretable models [1, 5].

C. Objectives

1. To develop and evaluate machine learning models for the classification of fetal health using CTG data.

2. To evaluate and compare the performance of different classifiers to identify the most effective model for the early detection of fetal distress.

3. To propose scalable, interpretable solutions suitable for use in low-resource settings.

D. Research Questions

1. How can Machine Learning models be developed and effectively utilized to classify fetal health using CTG data?

2. Which Machine Learning classifiers demonstrate the highest accuracy and reliability for the early detection of fetal distress?

3. What scalable and interpretable machine learning solutions can be proposed for fetal health monitoring in low-resource settings?

E. Scope

This study is limited to supervised ML models applied to a publicly available CTG dataset. The study focuses on classification performance, dataset imbalance, and practical implications for deployment in clinical settings.

II. LLITERATURE REVIEW

Huang *et al.* [6] explore the interpretability of fetal status assessment using cardiotocography (CTG) by identifying key CTG features and their causal relationships with fetal health through structural equation modeling. The findings validate and supplement existing fetal monitoring guidelines, enhancing evidence-based prenatal care.

Chiou *et al.* describe [7] the limitations of traditional cardiotocography (CTG) interpretation due to subjectivity and variability, which have led to unnecessary clinical interventions with limited neonatal benefit. Prior studies explored machine learning for CTG analysis, but often reduced complex signal data to rule-based features, ignoring important temporal and contextual cues critical for accurate fetal distress prediction.

Sheakh *et al.* stated [8] that Various machine learning algorithms, particularly Random Forest and neural networks, have demonstrated high accuracy in classifying fetal health from CTG data across multiple studies. These approaches effectively identify risk factors for maternal and child mortality by analyzing heart rate and uterine contraction patterns from publicly available datasets like UCI and Kaggle.

Francis *et al.* said [5] that the traditional CTG interpretation is limited by observer variability and poor predictive value, leading to misdiagnoses. Machine learning shows strong potential for improving fetal hypoxia detection, but clinical implementation requires diverse datasets, standard outcome benchmarks, and model interpretability.

Valderrama *et al.* emphasize [9] that despite widespread use, conventional fetal monitoring methods like CTG and ultrasound offer limited improvements in perinatal outcomes in low-resource settings. Emerging low-cost technologies and mobile health (mHealth) solutions are gaining attention for their potential to improve access and effectiveness of fetal cardiac monitoring in low- and middle-income countries.

Katebijahromi *et al.* [10] underscore the urgent need for accessible fetal cardiac monitoring in low-resource settings, highlighting that conventional technologies are often too costly or complex. It emphasizes the potential of low-cost 1D Doppler ultrasound combined with AI-driven models to provide effective, interpretable, and scalable monitoring solutions for underserved populations.

Puri Reddy *et al.* highlight [11] that, with conventional interpretation, interferences due to human errors, by their very nature, introduce variability; hence, the use of machine learning models such as Random Forest, SVM, and logistic regression improves diagnostic accuracy. High predictive performance has been reported with these models, while the ensemble and hybrid techniques add robustness to the classification of fetal health.

Zbelo *et al.* emphasize [12] that insufficient monitoring of fetal heart rate in low-resource settings is a major cause of intrapartum stillbirth. Moreover, it has been found that the participation of laboring mothers in fetal monitoring using handheld Doppler devices could promote the timely detection of fetal distress and thus reduce neonatal mortality.

Rajeev Madiraju and Utkarsh Upadhyay [13] mention that while traditional CTG interpretation uses subjective analysis, inconsistent by nature, machine learning methods like logistic regression, KNN, and gradient boosting show better results in decision-making related to determining fetal health status. They thus decrease the variability caused by the observers and assist in better predictive diagnostics, especially in low-resource settings.

According to Salini *et al.* [14], manual cardiotocography analysis is labor-intensive and prone to misinterpretation, with a consequent need to create machine learning models (Random Forest, SVM, and logistic regression) to improve accuracy and decision-making in fetal health classification. Previous studies prove that ML performs better than traditional approaches, particularly when considering early detection and optimal use of resources.

Hoodbhoy *et al.* [15] present the argument against traditional CTG being subject to subjectivity and interobserver differences, making it hard to timely identify fetal risks. Machine learning approaches like XGBoost and Random Forest have shown that they can classify fetal states more accurately, allowing for the early detection of complications in settings with fewer resources.

Frasch *et al.* [16] emphasize the lingering doubts about electronic fetal monitoring's effectiveness and varying interpretations, often associating them with unnecessary interventions and untoward consequences. Recent studies significantly indicated applying deep learning models to scan CTG images, clocking a 93-plus percent accuracy in early-stage fetal distress detection and thus the provision of timely obstetric care for prevention.

On the other hand, Nivedithaa *et al.* [17] emphasize that manual CTG interpretation is often subjective and prone to error, thereby delaying interventions for treatment. Recent studies show that machine learning models are being used to improve the classification accuracy of fetal health, especially ensemble techniques like Random Forest, XGBoost, and CatBoost, and also contribute toward better clinical decision-making using methods like SHAP and LIME.

Bansal *et al.* show [18] that CTG is one of the major methods of fetal monitoring, yet manual interpretations are error-prone and subjective. Recent research has shown that machine learning models, especially XGBoost, Random Forest, and deep neural networks, invariably increase the accuracy of classifying fetal health and, accordingly, allow for timely and objective prenatal care.

According to Kori *et al.* [19], classical CTG analysis is subjective and, therefore, prone to human error; hence, attempts were made to develop newer machine learning algorithms like Random Forest, SVM, and ensemble techniques to help in the classification of fetal health conditions. Recent studies indicate that blending and stacking algorithms yield superior predictions compared to individual models, thereby enhancing clinical reliability. Relying on CTG data, Almadi *et al.* [1] propose a machine learning classification of fetal health into Normal, Suspect, and Pathological. The results of the study show that Random Forest and Logistic Regression models trained on balanced datasets improve prediction accuracy, thereby facilitating early intervention and lessening maternal-fetal risks.

Rishard *et al.* [20] have undertaken the development, conduct, and evaluation of an online CTG training course for healthcare providers in Sri Lanka with promising results in improving documentation and interpretation skills. While the study marks the training course as feasible and effective even in low-resource settings, motivation and infrastructural issues hampered completion rates.

Mondal *et al.* [2] demonstrated a robust ensemble learning framework to classify fetal health status using CTG data into normal, suspect, and pathological classes. Through advanced sampling, tuning hyperparameters, and ensemble modeling procedures like stacking and soft voting, they reached an accuracy level of up to 99.78 percent, further stressing the potential of ensemble AI in prenatal risk prediction.

Jayalakshmi and Rajakumar *et al.* [21] proposed the Fetal Health Classifier Algorithm (FHCA), by means of CTG data and machine learning methods such as SVM, Random Forest, and MLP to discern fetal states with 95 percent accuracy. The study profoundly emphasizes the potential of FHCA to enhance prenatal care and curb maternal-child mortality in low-resource settings.

Chiou *et al.* [8] developed deep learning models using CTG data to predict fetal hypoxia and found that models trained on objective pH values performed better than those trained with subjective Apgar scores. With stable performance under temporal distribution shifts, their method further strengthens the case of AI-enabled CTG analysis for continuous and intermittent monitoring scenarios.

Puri and Reddy *et al.* [3] offered a framework for prediction using ML methods like Logistic Regression, SVM, Random Forest, and Naive Bayes on CTG data to classify fetal health conditions. It was found through this study that Logistic Regression was the best-performing algorithm in terms of accuracy (with 99.5 percent), thereby making a strong argument for the use of machine learning in early diagnosis and ensuring better maternal-fetal outcomes.

Mennickent *et al.* [22] have reviewed machine learning techniques for the prediction and management of pregnancy diseases such as gestational diabetes, preeclampsia, and preterm birth. The study emphasizes ML's importance in early diagnosis, biomarker discovery, and risk estimation while listing obstacles such as data quality, standardization, and translation into practice.

Lee *et al.* [23] proposed the 1D and 2D ResNet CNN models for fetal status classification in real-time using CTG data on mobile and server environments. The models had achieved high acceptance with an accuracy of 98.7%, having ensured that prenatal monitoring becomes a remote capability with clinching reliability and performance caliber.

An ensemble-based ML framework was proposed using CTG data and Multi-SVM, XGBoost, Random Forest, and Neural Network classifiers by Alkurdi and Abdulazeez [24]. The study seemed to register good classification accuracies, with XGBoost clearly dominating the rest, thus proving that the model effectively supports prenatal diagnostics and reduces human subjectivity during fetal health assessment.

In low-resource settings, Olayemi and Olasehinde [4] applied six machine learning models - LR, DT, RF, GB, XGB and CB on CTG data with feature selection and cross-validation to predict fetal health. Categorical Boosting was observed to be the best model with an MCC score of 0.6321, thus validating the importance of ML in early risk detection within emerging healthcare systems.

Das *et al.* [25] proposed stage-specific fetal health classifiers for CTG data and soft-computing methods involving SVM, Random Forest, and MLP. Their models provide up to 98 percent accuracy and excellent sensitivity toward suspicious and pathological cases, proven to work well in both stages of labor, with great agreement with clinical annotations.

Goel *et al.* [26] evaluated multiple ML algorithms—Random Forest, Gradient Boosting, Decision Trees, and k-NN—on CTG data for classifying fetal health states. The Random Forest model achieved the highest accuracy (98.49 percent), demonstrating strong potential for automated prenatal diagnostics and decision support systems.

Sirisha *et al.* [27] proposed a Dynamic Multi-Layer Perceptron (DMLP) model for fetal health classification using CTG data, achieving 97 percent accuracy. Through advanced preprocessing (e.g., SMOTE, scaling) and comparative analysis with traditional classifiers, the study demonstrated DMLP's superior performance and potential in real-time, explainable prenatal diagnostics.

Bansal *et al.* [18] evaluated eight ML models on CTG data for fetal health classification, with XGBoost achieving the highest accuracy at 99 percent. The study confirms the effectiveness of ensemble methods in real-time prenatal diagnostics and emphasizes the need for robust systems that support early intervention and improved maternal-fetal outcomes.

Harish and Reeja [28] conducted an extensive review of 25 studies applying machine learning and deep learning techniques to CTG and ultrasound data for fetal health monitoring. The study highlights the strengths and limitations of models like ANN, SVM, Q-Net, and 3D U-Net, emphasizing their diagnostic potential and challenges in preprocessing and real-world implementation.

III. METHODOLOGY

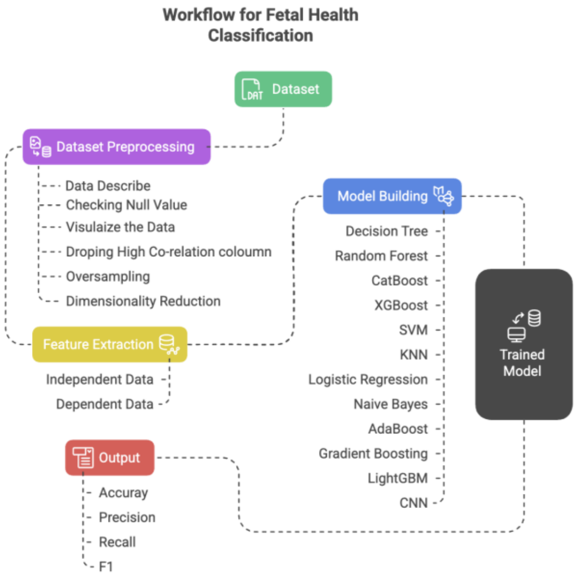

Fig. 1. Workflow for Fetal Health classification.

Figure 1 shows that the study employs a quantitative model approach, utilizing CTG readings as numerical data and incorporating several mathematical models for computation. This study begins by preprocessing the raw dataset to ensure its suitability for applying various machine learning algorithms. Key features are extracted to optimize model performance. The given Fig. 1 describes the workflow of this proposed solution. To predict fetal health status, multiple algorithms are employed, including Logistic Regression [29], K-Nearest Neighbors (KNN) [30], Convolutional Neural Networks (CNN) [31], AdaBoost [32], XGBoost [33], Naive Bayes[34], Gradient Boosting[35], LightGBM [36], Decision Tree [37], Random Forest [38], CatBoost [39], and Support Vector Machine(SVM) [40]. Each model is trained and tested with proper performance parameters, for example, metrics of accuracy, precision, and recall, so as to ensure dependable predictions. The proposed workflow is aimed at providing accurate and timely fetal health assessments so that healthcare professionals can act upon such information to improve pregnancy outcomes. As future work, the system will integrate the inclusion of more data to better fine-tune the models and ultimately improve system implementation.

A. Data Acquisition

The study used the "Fetal Health Prediction" dataset by Saumya Gupta from Kaggle [41]. The dataset is made use of CTG recordings, which record data concerning the fetal heart rate and uterine contractions required to observe fetal health. It entails parameters from tests performed on unborn babies, such as baseline data on heart rates and test results concerning accelerations and decelerations. The publicly accessible and anonymized data set is suitable for fetal health classification using machine learning and ensures ethical compliance.

B. Data Cleaning

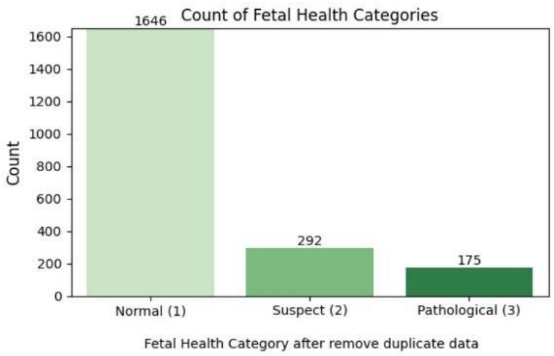

Fig. 2. Categorical view after removing duplicates.

Figure 2 shows that the dataset was carefully checked for duplicates and missing values, which were removed or imputed to ensure clean, complete input features. This preprocessing step helped improve the accuracy and reliability of the machine learning models.

C. Data Preprocessing

The data came from the Fetal Health Stroke dataset, which is utilized in numerous research endeavours. The only way to run the model is with restricted data. And use data from a trustworthy source to obtain a forecast. Because of this, the Fetal dataset was used to construct the dataset. To predict the foetus's health, that collection employs several techniques.

The starting value, fetal movement, uterine contractions, braking, accelerations, etc. A subset of the data is chosen from the earliest training data using the filtering approach in machine learning and data visualization applications.

D. Data Engineering

Figure 3 shows that a crucial step in improving the models' predictive capabilities was feature engineering. From the cardiotocographic dataset, pertinent features were chosen, such as baseline fetal heart rate, variability, accelerations, and decelerations—all of which are crucial markers of fetal health. The Synthetic Minority Over-Sampling Technique (SMOTE), which generates synthetic samples for underrepresented classes to produce a more impartial and balanced dataset, was used to solve the problem of class imbalance. By lowering bias toward the majority class and making sure that the input features were both instructive and indicative of the underlying data distribution, this method enhanced the model's capacity to correctly categorize all fetal health statuses.

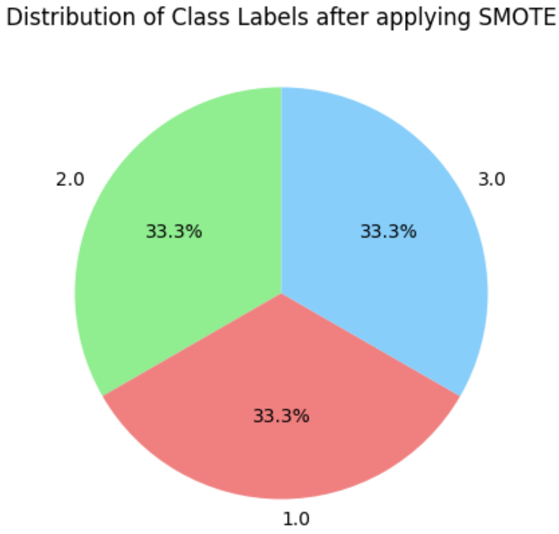

Fig. 3. Balance the categorical data using SMOTE.

E. Model Architecture

The categorical data is split into two datasets, one for training the models and the other for testing model performance.

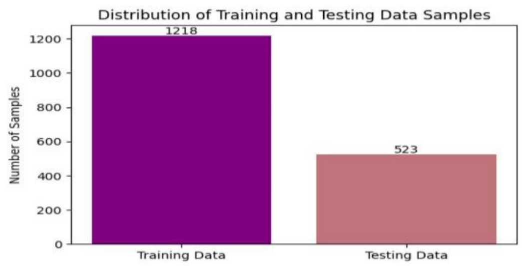

Fig. 4. Dataset divided into train and test data.

This study utilizes multiple supervised machine learning models, Random Forest, SVM, Naïve Bayes, and Logistic Regression, analysing key features from cardiotocography.

The data is used to accurately predict fetal health. These algorithms work together to handle data variability and enhance early diagnosis through robust classification. Some algorithms perform very well on this dataset, as shown below.

(1) Decision Tree: Decision trees employ supervised learning and partition data into branches depending on feature values. It results in a model that resembles a decision tree to perform predictions for classification or regression.

(2) Random Forest: Random forest builds many decision trees using an ensemble learning approach and then aggregates their results to increase the accuracy of predictions and reduce overfitting; hence,

it is a popular choice for regression as well as classification problems.

(3) CatBoost: CatBoost is a gradient boosting algorithm that efficiently and automatically deals with categorical features. It avoids overfitting by means of symmetric trees and ordered boosting, thus giving great precision with little pre-processing.

(4) XGBoost: The powerful gradient boosting technique XGBoost can be used for classification and regression tasks with the highest accuracy and efficiency. It sequentially builds an ensemble of decision trees that use gradient descent to minimize error and maximize performance.

(5) LightGBM: Tree-based learning techniques are used by LightGBM, a quick and effective gradient boosting framework, to increase prediction accuracy. It supports GPU and parallel learning and is built for great performance with big datasets.

(6) SVM: Support Vector Machines (SVMs) are supervised max-margin models with associated learning algorithms that analyze data for classification and regression analysis.

(7) KNN: k-Nearest Neighbors is a supervised machine learning algorithm used for both classification and regression tasks.

(8) Logistic Regression: It is a supervised machine learning algorithm used for classification problems. Unlike linear regression, which predicts continuous values, it predicts the probability that an input belongs to a specific class.

(9) Naïve Bayes: It is part of a family of generative learning algorithms, meaning that it seeks to model the distribution of inputs of a given class or category.

(10) Gradient Boosting: This is an ensemble machine learning technique that constructs a strong predictive model by combining multiple weak predictive models, typically decision trees, in a sequential and additive manner.

(11) AdaBoost: The Adaptive Boosting algorithm is an ensemble learning method primarily used for classification tasks, though it can also be adapted for regression.

(12) CNN: Convolutional Neural Networks (CNNs) are deep learning models that are particularly useful for image and signal analysis since they automatically extract and learn hierarchical features from data. They are effective at classification tasks because they employ convolutional layers to identify patterns like edges and textures.

Among all the models above, XGBoost performs best, followed respectively by Random Forest, CatBoost, and LightGBM.

IV. RESULTS AND DISCUSSION

A. Environmental Setting

This study is set in low-resource healthcare situations where advanced medical equipment is rare. Cardiotocography (CTG) is used in the study as a low-cost, non-invasive means of monitoring foetal health on a continuous basis. CTG data is critical for early detection of foetal distress, especially in resource-limited situations where prompt interventions can greatly reduce maternal and newborn mortality. By applying machine learning approaches to publicly accessible data, the study hopes to enhance early detection and provide scalable solutions, coinciding with global initiatives to improve maternal health outcomes in disadvantaged regions

B. Performance Measurement Metric

To get the best results, the current study used seven different machine learning methods. The effectiveness of the trained model has been evaluated in this study using a variety of metrics for performance measurement, which are described below:

(1) Accuracy: Classification accuracy measures how well a model correctly identifies cases. It is calculated as:

$$Accuracy = (TP + TN)/(TP + TN + FP + FN) \tag{1}$$

(2) Precision: Precision measures the proportion of true positive predictions out of all positive predictions. A model with higher precision is well-designed. It is calculated as:

$$Precision = TP/(TP + FP) \tag{2}$$

(3) Recall: Recall measures how well a model identifies actual positives. It is calculated as follows:

$$Recall = TP/(TP + FN) \tag{3}$$

, where TP (True Positive): Correctly classified as normal, suspicious, or pathological.

FP (False Positive): Mislabelled as pathological or suspicious when they were normal.

FN (False Negative): Misdiagnosed as normal when they were pathological or suspicious.

C. Performance Analysis

The successful implementation of the advanced techniques holds significant promise for enhancing the quality of prenatal care and improving clinical decision-making, particularly in resource-limited settings.

Figure 5 shows the overall accuracy of several models on testing data. The XGBoost model achieved the highest accuracy of 99% after completing all data preprocessing steps, including cleaning and removing outliers, duplicate data, and NULL values. Random Forest, CatBoost, and LightGBM achieved 98% accuracy. Where Gradient Boosting, Decision Tree achieved 97% and KNN, SVM, Logistic Regression, CNN achieved 96%, 95% and 92% respectively, while Naive Bayes performed the worst score 87%.

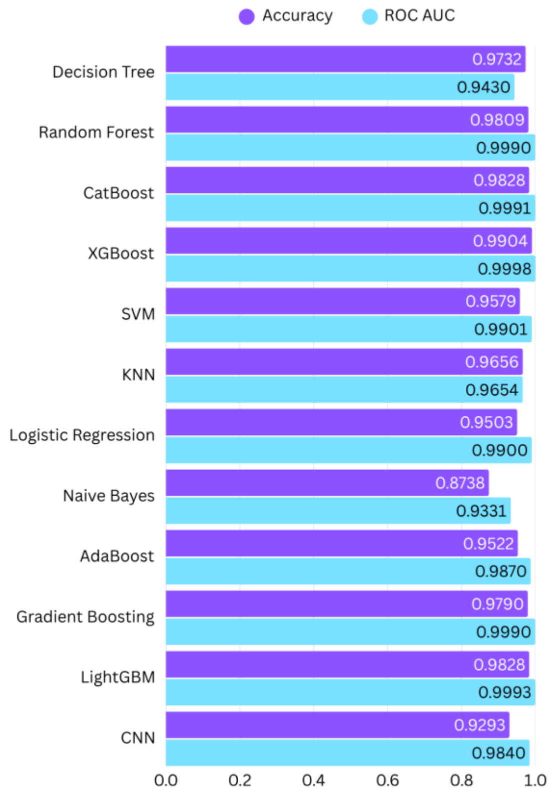

Fig. 5. Overall accuracy of several models on testing data.

Table I. Model accuracy comparison.

Model	*Accuracy*	*ROC AUC*
Decision Tree	0.9732	0.9433
Random Forest	0.9809	0.9990
CatBoost	0.9828	0.9991
XGBoost	0.9904	0.9998
SVM	0.9579	0.9901
KNN	0.9656	0.9654
Logistic Regression	0.9503	0.9900
Naive Bayes	0.8738	0.9331
AdaBoost	0.9522	0.9870
Gradient Boosting	0.9790	0.9990
LightGBM	0.9828	0.9993
CNN	0.9293	0.9840

Table I shows the comparison of performances among different models on the test datasets.

D. Performance Comparison: Previous Model vs Proposed Model.

In all the models, XGBoost scores the highest, achieving precision and recall values of nearly 1.0 across the classes, with 0.99 (Normal), 0.96 (Suspect), and 0.97 (Pathological), thereby being the most reliable model for accurately classifying the fetal state. The Decision Tree and SVM models show improvement in their detection rates as well, especially on the Pathological state. Overall, while Naive Bayes has improvements for the Normal and Pathological states, the Suspect state still has low precision and recall scores.

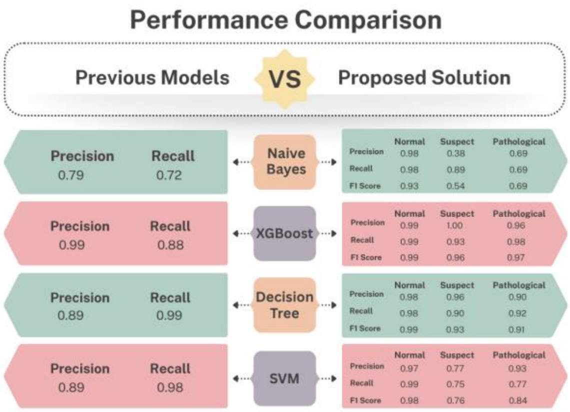

Fig. 6. Performance comparison: Previous model vs proposed solution.

Figure 6 represents the performance of prior and proposed versions of four machine learning models—Naive Bayes, XGBoost, Decision Tree, and Support Vector Machine (SVM)—for the early detection of fetal distress from Cardiotocography (CTG). The proposed models show a statistically significant improvement for the evaluation metrics of Precision, Recall, and F1 Score, measured individually per class (Normal, Suspect, Pathological).

Overall, this comparison demonstrates that the proposed machine learning models, particularly the XGBoost, provide advancements in monitoring fetal distress, which is an important way to promote improved fetal mortality outcomes for mothers and babies living in precarious socio-economic status or low-resourced healthcare settings.

E. Limitations

While this study demonstrates promising results in fetal health classification using machine learning on CTG data, several limitations must be acknowledged. First, the dataset used is limited in size and diversity, which may affect the generalizability of the models to broader clinical populations. Additionally, the data was collected from a single source and lacked real-time physiological variability or annotations from multiple clinical experts, potentially introducing bias. The study also assumes ideal data quality and does not account for noise or artifacts commonly encountered in real-world settings. Ultimately, although several models were evaluated, model interpretability remains a challenge—particularly for deep learning approaches like CNNs—which can hinder adoption in clinical environments where transparency is critical.

F. Future Work

Future research will focus on addressing current limitations by incorporating larger, multi-source datasets that reflect a more diverse demographic and clinical environment. Real-time CTG signal processing and noise reduction techniques will also be integrated to improve model robustness. Moreover, attention will be given to enhancing model explainability through techniques such as SHAP or LIME, enabling clinicians to understand and trust the decisions made by machine learning algorithms. Another avenue involves developing lightweight, mobile-compatible frameworks that support deployment in low-resource healthcare settings, aligning to reduce maternal and neonatal mortality in underserved regions.

V. CONCLUSION

This study presents a comprehensive evaluation of various machine learning models applied to cardiotocography (CTG) data for early detection of fetal distress. The XGBoost model demonstrated the highest predictive performance, achieving 99 percent accuracy, followed closely by CatBoost, LightGBM, and Random Forest. These results highlight the potential of machine learning in transforming prenatal care, especially in low-resource settings where early intervention can significantly reduce preventable child and maternal mortality. Despite existing challenges, the integration of scalable and interpretable ML-based tools into clinical workflows holds promise for enhancing diagnostic accuracy and supporting healthcare.

ACKNOWLEDGEMENT

We gratefully acknowledge the support and guidance of our mentors and peers throughout this study. We also extend our appreciation to the providers of the publicly available CTG dataset used in this research.

FUNDING STATEMENT

The authors received no funding from any party for the research and publication of this article.

AUTHOR CONTRIBUTIONS

Md Farman Ali: Conceptualized the study, designed the overall research framework, supervised the work, and provided critical guidance and revisions for all sections.
Kamrul Hasan: Led the restructuring of the manuscript for logical flow and coherence, conducted an extensive review and unification of correlated writings.
Md Ikramul Haque: Conducted formatting, proofreading, and the preliminary review of the manuscript to ensure compliance with journal requirements.
Jahid Hassan Noor: Performed grammatical corrections, ensured citation consistency, and revised the paper according to the feedback of the reviewers.

CONFLICT OF INTERESTS

No conflict of interest was disclosed.

ETHICS STATEMENTS

Our publication ethics follow The Committee of Publication Ethics (COPE) guidelines. https://publicationethics.org/

REFERENCES

[1] M. Almadi, F. Alotaibi, R. Almudawah, A. Ali, Y. Nasser and N. Nasser, "Data-Driven Machine Learning Models for Enhanced Fetal Health Classification and Monitoring," in *2025 8th Int. Conf. Data Sci. and Machine Learn. Appl.*, pp. 189–192, 2025.

[2] S. Mondal, R. Maity, A. Nag and S. Ghosh, "Fetal Health Risk Prediction using Ensemble-Based Machine Learning Approaches," *Knowled. and Inform. Syst.*, vol. 67, pp. 7227–7261, 2025.

[3] C. Puri and K. T. V. Reddy, "Smart Healthcare: Early Diagnosis of Fetal Health Conditions using Machine Learning and Cardiotocographic Data," in *2024 2nd DMIHER Int. Conf. Artif. Intellig. Healthcare, Educ. and Indust.*, pp. 1–5, 2024.

[4] O. C. Olayemi and O. O. Olasehinde, "Machine Learning Prediction of Fetal Health Status from Cardiotocography Examination in Developing Healthcare Contexts," *J. Comput. Sci. Res.*, vol. 6, pp. 43–53, 2024.

[5] M. T. Rayhana, A. S. Arefina and S. A. Chowdhury, "Automatic Detection of Fetal Health Status from Cardiotocography Data using Machine Learning Algorithms," *J. Bangladesh Acad. Sci.*, vol. 45, no. 2, pp. 155–167, 2022.

[6] F. Francis, S. Luz, H. Wu, S. J. Stock and R. Townsend, "Machine Learning on Cardiotocography Data to Classify Fetal Outcomes: A Scoping Review," *Comput. Biol. and Medic.*, vol. 172, pp. 108220, 2024.

[7] L. Huang *et al.*, "Investigating the Interpretability of Fetal Status Assessment using Antepartum Cardiotocographic Records," *BMC Medic. Informat. and Dec. Making*, vol. 21, pp. 1–11, 2021.

[8] N. Chiou *et al.*, "Development and Evaluation of Deep Learning Models for Cardiotocography Interpretation," *NPJ Women's Health*, vol. 3, no. 1, pp. 21, 2025.

[9] M. A. Sheakh *et al.*, "Child and Maternal Mortality Risk Factor Analysis Using Machine Learning Approaches," in *2023 11th Int. Symp. Digit. Forens. and Secur.*, Chattanooga, TN, USA, pp. 1–6, 2023.

[10] C. E. Valderrama, N. Ketabi, F. Marzbanrad, P. Rohloff and G. D. Clifford, "A Review of Fetal Cardiac Monitoring, with A Focus on Low-and Middle-Income Countries," *Physiolog. Measure.*, vol. 41, no. 11, pp. 11TR01, 2020.

[11] N. Katebijahromi, Detection of Adverse Events in Pregnancy Using A Low-Cost 1D Doppler Ultrasound Signal, *PhD Dissertation*, Emory University, 2021.

[12] M. G. Zbelo, Fetal Heart Rate Self-Monitoring by Mothers in Labor, *PhD Dissertation*, Walden University, 2024.

[13] R. Madiraju, U. Upadhyay, C. Muralidharan and R. Bharati, "Fetal Health Analysis Based on CTG," in *Proc. 11th Int. Conf. Commun. Sign. Process*, Melmaruvathur, India, pp. 1706–1711, 2025.

[14] Y. Salini, S. N. Mohanty, J. V. N. Ramesh, M. Yang and M. M. V. Chalapathi, "Cardiotocography Data Analysis for Fetal Health Classification using Machine Learning Models," *IEEE Access*, vol. 12, pp. 26005–26022, 2024.

[15] Z. Hoodbhoy *et al.*, "Use of Machine Learning Algorithms for Prediction of Fetal Risk Using Cardiotocographic Data," *Int. J. Appl. and Basic Medic. Res.*, vol. 9, no. 4, pp. 226–230, 2019.

[16] M. G. Frasch, S. B. Strong, D. Nilosek, J. Leaverton and B. S. Schifrin, "Detection of Preventable Fetal Distress During Labor from Scanned Cardiotocogram Tracings using Deep Learning," *Front. in Pediatr.*, vol. 9, pp. 736834, 2021.

[17] G. Niveditha, B. U. Maheswari and R. P. de Prado, "Elevating Maternal Healthcare: Synergy of Cardiotocography, Machine Learning Models and Interpretive Analysis," *Proc. Comput. Sci.*, vol. 258, pp. 2787–2797, 2025.

[18] R. Bansal, I. B. Sonowal and R. K. Bhukya, "Predicting Fetal Health using Cardiotocography and Machine Learning: A Study on Prenatal Care," in *2025 Int. Conf. Innov. Comput. and Eng.*, Greater Noida, India, pp. 1–6, 2025.

[19] A. Kori, "Enhancing Fetal Health Prediction Using Machine Learning and Ensemble Methods," in *2025 Int. Conf. Pervas. Computat. Technol.*, Greater Noida, India, pp. 116–120, 2025.

[20] M. Rishard *et al.*, "Development and Evaluation of An Online Cardiotocography Course Tailored to LMIC Settings: A

Feasibility Study Conducted in A Tertiary Care Hospital in Sri Lanka," *BMC Res. Notes*, vol. 18, no. 1, pp. 163, 2025.

[21] P. R. Jayalakshmi and P. S. Rajakumar, "Enhancing Fetal Health Classification for Child and Maternal Mortality Prevention Using FHCA," *Afr. J. Bio. Sci.*, vol. 6, no. 6, pp. 7017-7028, 2024.

[22] D. Mennickent *et al.*, "Machine Learning Applied in Maternal and Fetal Health: A Narrative Review Focused on Pregnancy Diseases and Complications," *Front. Endocrinol.*, vol. 14, pp. 1130139, 2023.

[23] K. S. Lee *et al.*, "Real-Time Classification of Fetal Status Based on Deep Learning and Cardiotocography Data," *J. Medic. Syst.*, vol. 47, no. 1, pp. 82, 2023.

[24] A. Alkurdi and A. M. Abdulazeez, "Comprehensive Classification of Fetal Health Using Cardiotocogram Data Based on Machine Learning," *The Indonesian J. Comput. Scit*, vol. 13, no. 1, pp. 277-300, 2024.

[25] S. Das, H. Mukherjee, K. Roy and C. K. Saha, "Fetal Health Classification from Cardiotocograph for Both Stages of Labor—A Soft-Computing-Based Approach," *Diagnostics*, vol. 13, no. 5, pp. 858, 2023.

[26] A. Goel, I. B. Sonowal, V. K. Saxena and R. K. Bhukya, "Machine Learning Models for Fetal Health Classification using Cardiotocography: Towards Improved Prenatal Care and Outcomes," in *2024 5th Int. Conf. Emer. Technol.*, Belgaum, India, pp. 1–6, 2024.

[27] U. Sirisha *et al.*, "Dynamic Multi-Layer Perceptron for Fetal Health Classification using Cardiotocography Data," *Comput., Mater. & Contin.*, vol. 80, no. 2, pp. 2301-2330, 2024.

[28] S. Harish and S. R. Reeja, "Comprehensive Review of Fetal Health Monitoring using DL and ML Techniques with Ultrasound and Cardiotocography Data," in *2024 Int. Conf. Emer. Syst. and Intellig. Comput.*, Bhubaneswar, India, pp. 550–555, 2024.

[29] E. Bisong, *Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners*, Apress, pp. 243–250, 2019.

[30] Y. Liao and V. R. Vemuri, "Use of K-Nearest Neighbor Classifier for Intrusion Detection," *Comput. & Secur.*, vol. 21, no. 5, pp. 439–448, 2002.

[31] L. Xie and A. Yuille, "Genetic CNN," in *Proc. the IEEE Int. Conf. Comput. Vision*, Venice, Italy, pp. 1379–1388, 2017.

[32] R. E. Schapire, "Explaining AdaBoost," in *Empirical Inference*, pp. 37–52, Springer Nature, 2013.

[33] S. Ramraj, N. Uzir, R. Sunil and S. Banerjee, "Experimenting XGBoost Algorithm for Prediction and Classification of Different Datasets," *Int. J. Contr. Theory and Appl.*, vol. 9, no. 40, pp. 651–662, 2016.

[34] D. Berrar, "Bayes' Theorem and Naive Bayes Classifier," in *Encyclopedia of Bioinformatics and Computational Biology*, S. Ranganathan, M. Gribskov, K. Nakai, and C. Schönbach, Edn. Oxford, Elsevier, pp. 403–412, 2018.

[35] C. Bentejac, A. Csörgő and G. Martínez-Muñoz, "A Comparative Analysis of Gradient Boosting Algorithms," *Artif. Intellig. Rev.*, vol. 54, pp. 1937–1967, 2021.

[36] G. Ke *et al.*, "LightGBM: A Highly Efficient Gradient Boosting Decision Tree," *Neural Inform. Process. Syst.*, vol. 30, 2017.

[37] B. Charbuty and A. Abdulazeez, "Classification Based on Decision Tree Algorithm for Machine Learning," *J. Appl. Sci. and Technol. Trends*, vol. 2, no. 01, pp. 20–28, 2021.

[38] S. J. Rigatti, "Random Forest," *J. Insuran. Medic.*, vol. 47, no. 1, pp. 31–39, 2017.

[39] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush and A. Gulin, "CatBoost: Unbiased Boosting with Categorical Features," *Neural Inform. Process. Syst.*, vol. 31, 2018.

[40] R. Xiao, J. Wang and F. Zhang, "An Approach to Incremental SVM Learning Algorithm," in *Proc. the 12th IEEE Int. Conf. Tools with Artif. Intellig.*, DC, United States, pp. 268–273, 2000.

[41] S. Gupta, "Fetal Health Prediction," Kaggle. [Available online on 22 Nov. 2024] https://www.kaggle.com/code/saumyagupta2025/fetal-health-prediction-eda-acc-98/notebook.