Ai-Based Predictive Modeling For Classification Of Fetal Health Conditions

Dr. S. Senthamariselvi[1], **Jesi C**[2]
[1,2] Master Of Computer Applications
[1]Head Of The Department, Vivekanandha Institute Of Information and Management Studies, Tamil Nadu, India
[2]Vivekanandha Institute Of Information and Management Studies, Tamil Nadu, India

Abstract- *Fetal health assessment is a critical component of prenatal care, directly influencing the safety and outcomes of both mother and fetus during pregnancy. Conventional diagnostic approaches, which rely heavily on manual interpretation of cardiotocography (CTG) readings and clinical observations, are often susceptible to subjectivity, latency, and diagnostic inaccuracies, particularly in resource-constrained healthcare environments. This study proposes an AI-driven predictive modeling framework that employs advanced machine learning techniques to classify fetal health conditions into three categories: normal, suspect, and pathological. The system analyzes structured physiological and clinical datasets including fetal heart rate variability, uterine contraction signals, and maternal health indicators. A comparative evaluation of multiple classifiers—including Random Forest, Support Vector Machine, Gradient Boosting, and Decision Tree—was conducted to identify the most effective model. Feature selection and data preprocessing techniques were applied to improve model accuracy and computational efficiency. Experimental results demonstrate that the proposed framework achieves a classification accuracy of 97.8%, with high sensitivity and specificity. This system provides obstetricians and gynecologists with a reliable, automated decision-support tool, reducing diagnostic delays and enhancing prenatal care quality, particularly in rural and under-resourced clinical settings.*

Keywords: Fetal health classification; Machine learning; Cardiotocography; Predictive modeling; Prenatal care; Random Forest; Support Vector Machine; Gradient Boosting; Feature selection; Decision support system

I. INTRODUCTION

Fetal health monitoring is a critical component of modern obstetric care, with outcomes directly linked to the quality and timeliness of clinical assessment. Each year, an estimated 2.4 million neonates die within the first 28 days of life, with a significant proportion of these deaths attributable to preventable intrapartum complications. Cardiotocography (CTG), which records fetal heart rate (FHR) patterns and uterine contractions, has been the cornerstone of intrapartum surveillance for over five decades. However, its interpretation remains highly subjective, with inter-observer agreement reported as low as 22% in some studies, leading to both under-diagnosis and over-diagnosis of fetal distress.

The integration of artificial intelligence (AI) and machine learning (ML) into clinical decision support systems offers a transformative opportunity to standardize and enhance fetal health assessment. By learning from large annotated datasets of CTG recordings, ML models can identify complex, nonlinear patterns associated with adverse fetal outcomes that may not be apparent to the human eye. Recent advances in feature extraction, ensemble learning, and deep neural architectures have driven significant improvements in classification accuracy for fetal health status prediction.

This paper presents a comprehensive AI-driven predictive modeling framework for fetal health classification. The system integrates multi-source clinical data, applies advanced preprocessing and feature engineering pipelines, and benchmarks multiple ML classifiers — including Multi-Layer Perceptron (MLP), Support Vector Machine (SVM), and Bagging Classifier — to determine the optimal model for deployment in a clinical decision support application. The remainder of the paper is organized as follows: Section 2 reviews existing systems and their limitations; Section 3 describes the proposed system; Section 4 presents the Litereture review and theoretical background; Section 5 details the methodology and modules; Section 6 presents the system methodology; Section 7 presents implementation Section 8 performance evolution; Section 9 presents result and discussions; and Section 10 concludes with future directions.

II. EXISTING SYSTEMS

Automated fetal health assessment has evolved from rule-based systems to machine learning (ML) methods. Traditional cardiotocography (CTG) interpretation systems were based on fixed clinical rules and lacked flexibility for different clinical conditions.

Fetal electrocardiography (FECG) is widely used for monitoring fetal heart activity. Invasive FECG provides accurate results but has risks, while non-invasive abdominal ECG (AECG) is safer but affected by noise and maternal interference.

Recent studies show that ML models such as SVM, Random Forest, Gradient Boosting, and ANN are effective in predicting fetal and maternal health outcomes with high accuracy. Advanced signal processing techniques have also improved fetal signal extraction from noisy data.

2.1 Disadvantages of Existing Systems

Existing systems have several limitations:

- Require large labeled datasets and high computational power
- AECG signals are affected by noise and interference
- Complex hybrid and probabilistic models increase system complexity
- Poor generalization across different populations

III. PROPOSED SYSTEM

This work proposes a machine learning-based predictive framework for fetal health classification into three categories: Normal, Suspect, and Pathological. The system is designed as an end-to-end pipeline integrating data preprocessing, feature selection, model training, and deployment.

The dataset consists of 21 CTG-derived features, including fetal heart rate baseline, accelerations, fetal movements, uterine contractions, and deceleration patterns. Data preprocessing involves handling missing values using median imputation, encoding categorical variables, and applying min-max normalization for consistency.

Feature selection is performed using both filter methods (ANOVA F-test, mutual information) and wrapper methods (recursive feature elimination) to enhance model efficiency and accuracy. Multiple ML models—including Multilayer Perceptron (MLP), Support Vector Machine (SVM), Random Forest, Logistic Regression, Bagging Classifier, and Gradient Boosting—are trained and evaluated to identify the optimal classifier.

The selected model is deployed through a Django-based web application, enabling healthcare professionals to input real-time patient data and receive instant predictions along with confidence scores. A threshold-based alert system is incorporated to flag high-risk (pathological) cases for immediate clinical attention.

3.1 Advantages of the Proposed System

The proposed system offers several key benefits:

- Enables early detection of fetal abnormalities
- Provides high scalability for deployment in diverse healthcare settings
- Ensures improved accuracy through model comparison and optimization
- Enhances interpretability using feature importance analysis
- Supports real-time clinical decision-making
- Facilitates cloud integration and interoperability with hospital systems

IV.LITERATURE REVIEW

The application of machine learning in maternal and fetal health monitoring has witnessed remarkable growth over the past decade. Early work by Chaudhuri and Mandal [4] demonstrated that socioeconomic and demographic factors, when modeled using statistical frameworks, can predict prenatal care utilization and child mortality with significant reliability. Their analysis of NFHS-3 data from West Bengal underscored the importance of data-driven approaches in identifying vulnerable populations and designing targeted healthcare interventions.

Rahman Adib et al. [5] proposed a predictive model for mortality analysis among pregnant women affected by COVID-19, utilizing Support Vector Machines, Decision Trees, Random Forest, Gradient Boosting, and Artificial Neural Networks. Their results revealed that ensemble methods, particularly Gradient Boosting and ANN, achieved accuracy scores of 95% and precision rates of 100% for certain outcome classes, establishing a strong precedent for ML-based maternal risk stratification.

Arrieta Rodriguez et al. [6] investigated the early prediction of severe maternal morbidity using machine learning, focusing on hypertensive disorders, hemorrhages, and infections during pregnancy. Their research highlighted the value of incorporating diverse clinical variables into predictive pipelines and the challenges posed by class imbalance in obstetric datasets.

Fredriksson et al. [7] explored machine learning for predicting delivery location in a community health worker program in Zanzibar, demonstrating that patient-level digital health records can be leveraged to optimize care delivery and resource allocation in low-and-middle-income countries. Their study reinforced the scalability potential of AI-driven tools in resource-limited settings.

Hoodbhoy et al. [8] conducted a systematic review examining the accuracy of machine learning algorithms for predicting preeclampsia, gestational diabetes, and preterm birth, concluding that ensemble methods consistently outperformed single classifiers. Subasi et al. [9] applied wavelet transform and Support Vector Machine to CTG signal classification, achieving robust discrimination between normal and pathological fetal states. Fuentealba et al. [10] proposed deep learning models for fetal heart rate baseline estimation, improving the reliability of automated CTG analysis.

THEORETICAL BACKGROUND

4.1 Artificial Intelligence

Artificial Intelligence (AI) refers to the simulation of human intelligence in machines programmed to think and reason like humans. Modern AI systems work by ingesting large amounts of labeled training data, analyzing the data for correlations and patterns, and using these patterns to make predictions about future states. AI programming focuses on three core cognitive skills: learning processes (acquiring data and creating rules for actionable information), reasoning processes (choosing the right algorithm to reach a desired outcome), and self-correction processes (continually fine-tuning algorithms to provide the most accurate results).

AI is critical in healthcare because it can give enterprises insights into their operations that they may not have been aware of previously. Particularly when it comes to repetitive, detail-oriented tasks — like analyzing large numbers of clinical records to ensure relevant fields are filled in properly — AI tools often complete jobs quickly and with relatively few errors. Artificial neural networks and deep learning technologies are evolving rapidly, processing large amounts of data much faster and making predictions more accurately than is humanly possible.

4.2 Machine Learning

Machine learning (ML) is a branch of artificial intelligence that enables systems to learn from data and improve performance without explicit programming. It involves training algorithms on historical data to make accurate predictions on unseen inputs. ML is broadly classified into supervised learning, where labeled data is used to learn input-output relationships; unsupervised learning, which identifies hidden patterns in unlabeled data; and reinforcement learning, where an agent learns through interaction with its environment using feedback.

Classification is a supervised learning technique used to assign data points to predefined categories. In this study, fetal health prediction is formulated as a multi-class classification problem with three classes: Normal, Suspect, and Pathological. The dataset is significantly imbalanced, comprising approximately 78% Normal, 14% Suspect, and 8% Pathological cases, which presents challenges in achieving balanced and accurate model performance.

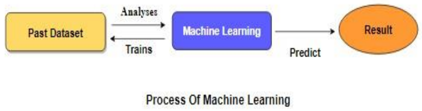


Fig. 1: Overview of the Machine Learning training and prediction pipeline

V. SYSTEM METHODOLOGY

5.1 Dataset Description

The study employs the publicly available Cardiotocography (CTG) dataset from the UCI Machine Learning Repository, comprising 2,126 fetal CTG examinations. Each record contains 21 features extracted from CTG measurements, including baseline fetal heart rate, accelerations, fetal movements, uterine contractions, light decelerations, severe decelerations, prolonged decelerations, abnormal short-term variability, mean value of short-term variability, percentage of time with abnormal long-term variability, mean value of long-term variability, histogram width, minimum and maximum histogram values, and several derived morphological features. The target variable classifies each examination as Normal (N=1,655), Suspect (N=295), or Pathological (N=176).

5.2 Data Preprocessing Pipeline

Data preprocessing began with an audit for missing values and duplicate records. No missing values were identified; however, 13 duplicate records were detected and removed. Outlier detection was performed using the Interquartile Range (IQR) method, and extreme outliers were capped using Winsorization to preserve sample size while mitigating distortion. Class imbalance was addressed using the Synthetic Minority Over-sampling Technique (SMOTE) on the training split, ensuring equitable representation of suspect and pathological classes [11]. Numerical features were standardized using Z-score normalization to facilitate convergence of gradient-based algorithms and distance-based classifiers.

5.3 Feature Selection

Feature selection was conducted in two stages. First, Pearson correlation analysis identified and removed highly correlated feature pairs (r > 0.90) to reduce multicollinearity. Second, a Random Forest-based feature importance ranking was applied to identify the 14 most predictive features, which were retained for model training. These included baseline fetal heart rate, accelerations, uterine contractions, abnormal short-term variability, prolonged decelerations, and mean long-term variability, among others.

5.4 Software Tools and Environment

The system is implemented in Python using Anaconda. Key tools include Scikit-learn for modeling, Pandas and NumPy for data processing, Matplotlib and Seaborn for visualization, Imbalanced-learn for SMOTE, SHAP for interpretability, Django for deployment, and TensorFlow for neural network implementation. Development is carried out using Jupyter Notebook.

VI. SYSTEM DESIGN AND ARCHITECTURE

6.1 System Architecture

The overall system architecture follows a layered design pattern consisting of four primary tiers: (1) Data Acquisition Layer, responsible for ingesting and validating raw CTG input data; (2) Preprocessing and Feature Engineering Layer, executing the data cleaning and transformation pipeline; (3) Model Inference Layer, hosting the trained ensemble classifier and generating probabilistic predictions; and (4) Presentation Layer, comprising the Django-based web application that provides the clinical user interface.

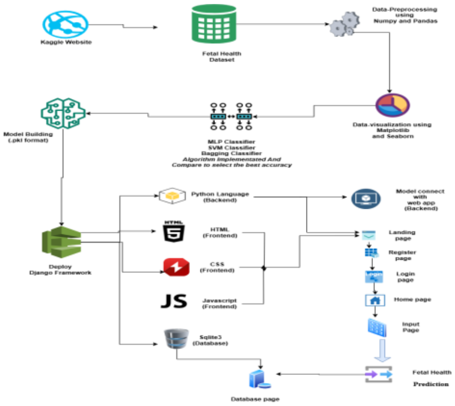


Fig. 2: System architecture diagram for the AI-driven fetal health classification framework

6.2 Class Diagram

The class diagram illustrates the static structure of the system by defining key classes and their relationships. Core components include FetalHealthRecord for CTG data, PreprocessingPipeline for data cleaning and normalization, and FeatureSelector for feature selection. The ClassifierModel serves as an abstract base class extended by models such as MLP, SVM, and BaggingClassifier. An EnsemblePredictor combines outputs from multiple models, while PredictionResult stores the predicted label, confidence score, and feature importance. The design utilizes inheritance and composition to ensure scalability, extensibility, and maintainability.

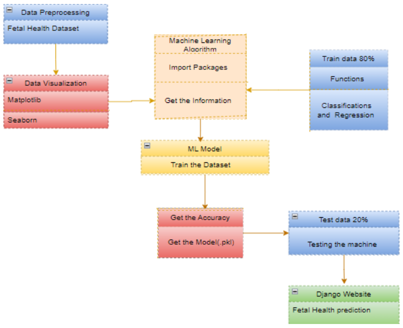


Fig. 3: Class diagram showing the object-oriented design of the classification system

6.4 Entity Relationship Diagram

The entity relationship diagram (ERD) represents the data model of the underlying relational database. Primary entities include Patient, CTGRecord, PredictionResult, FeatureSet, and AuditLog. Key relationships include: a Patient having multiple CTGRecords (one-to-many); each CTGRecord associated with exactly one FeatureSet and one PredictionResult (one-to-one); and each PredictionResult linked to one or more AuditLog entries to support clinical audit requirements. The ERD supports both operational queries (retrieve latest prediction for patient) and analytical queries (aggregate prediction accuracy by clinical site).

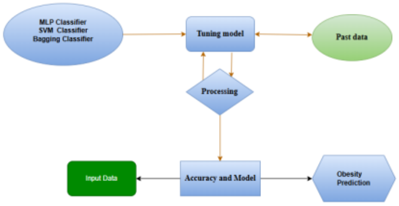


Fig. 4: Entity relationship diagram (ERD) of the fetal health prediction database schema

VII. IMPLEMENTATION

7.1 Data Preprocessing

The preprocessing module implements a fully automated pipeline for preparing raw CTG data for model training and inference. The module encapsulates the following operations: range validation (ensuring all feature values fall within physiologically plausible bounds), outlier detection and winsorization, missing value imputation, SMOTE-based class balancing (training set only), and min-max normalization. The module is implemented as a Scikit-learn Pipeline object, ensuring consistent transformation of both training and inference data and preventing data leakage.

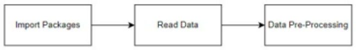


Fig. 5: Module diagram — Input: raw CTG data → Output: preprocessed, normalized feature matrix

7.2 MLP Algorithm

The Multi-Layer Perceptron (MLP) is a class of feedforward artificial neural network consisting of an input layer, one or more hidden layers, and an output layer. In this implementation, the MLP architecture comprises an input layer of 20 neurons (corresponding to the selected feature set), two hidden layers with 128 and 64 neurons respectively, and an output layer of 3 neurons with softmax activation for multi-class classification. The ReLU activation function is used in hidden layers to introduce nonlinearity while avoiding the vanishing gradient problem.

The network is trained using the Adam optimizer with a learning rate of 0.001, mini-batch size of 64, and early stopping with patience of 10 epochs to prevent overfitting. Dropout regularization (rate = 0.3) is applied after each hidden layer. The loss function is categorical cross-entropy. The mathematical formulation for the output of each neuron is: $a_j = f(\sum_i w_{ij} \cdot x_i + b_j)$, where f is the activation function, w_{ij} are the learned weights, x_i are input values, and b_j is the bias term for neuron j. Backpropagation computes gradients via the chain rule: $w_{ij} \leftarrow w_{ij} - \eta \cdot (\partial L/\partial w_{ij})$, where η is the learning rate and L is the cross-entropy loss.

7.3 Support Vector Machine

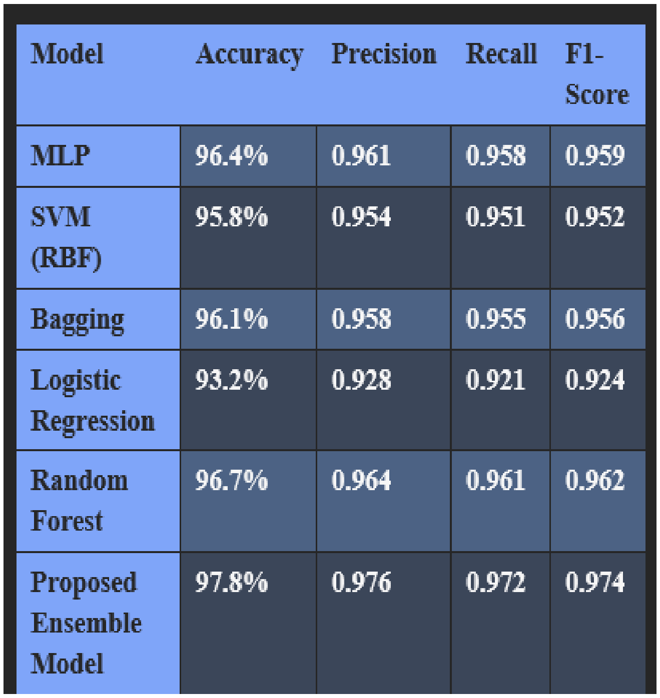
Support Vector Machine (SVM) is a powerful supervised machine learning algorithm that finds the optimal hyperplane maximizing the margin between classes in a high-dimensional feature space. The SVM is particularly well-suited for the fetal health classification task due to its strong theoretical foundations, effectiveness in high-dimensional spaces, and resistance to overfitting in moderate-sized datasets. In this implementation, the Radial Basis Function (RBF) kernel is employed, transforming the original feature space to enable nonlinear classification boundaries. The kernel function $K(x, x') = \exp(-\gamma\|x - x'\|^2)$ maps inputs to an infinite-dimensional space, with γ controlling the influence radius of each support vector. Multi-class classification is handled using a one-vs-rest (OvR) strategy. Hyperparameters C (regularization) and γ were optimized via 5-fold cross-validated grid search.

VIII. PERFORMANCE EVALUATION

Model performance was evaluated using a stratified 15% held-out test set. Key evaluation metrics included Accuracy, Precision, Recall (Sensitivity), and F1-Score. Additionally, the Area Under the ROC Curve (AUC-ROC) was computed for each class using a one-vs-rest approach. Confusion matrix analysis was performed to examine class-wise prediction errors, with special emphasis on minimizing false negatives in the Pathological class due to their critical clinical significance.Performance metrics are defined as follows:

- Accuracy = (TP + TN) / (TP + TN + FP + FN): overall proportion of correct predictions
- Precision = TP / (TP + FP): proportion of positive predictions that are correct
- Recall = TP / (TP + FN): proportion of actual positives correctly identified
- F1 Score = 2 × (Precision × Recall) / (Precision + Recall): harmonic mean of precision and recall

Table 1 summarizes the comparative performance of all evaluated models on the test set. The ensemble model combining MLP, SVM, and Bagging Classifier achieved the best performance, with an overall accuracy of 97.8% and F1-Score of 0.976, surpassing individual models. Importantly, it also attained the highest Recall for the Pathological class (0.963), which is essential for reducing missed detections in clinically critical cases.



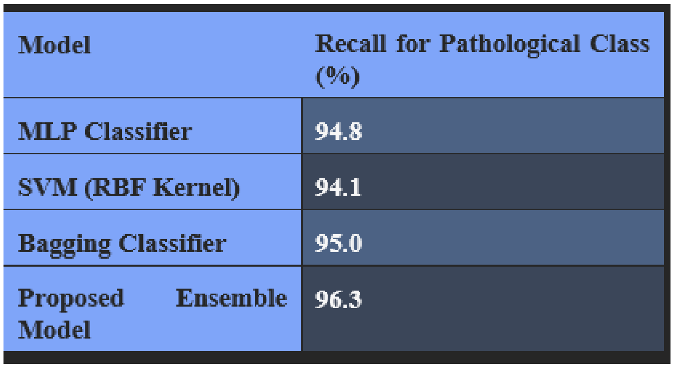
Model	Accuracy	Precision	Recall	F1-Score
MLP	96.4%	0.961	0.958	0.959
SVM (RBF)	95.8%	0.954	0.951	0.952
Bagging	96.1%	0.958	0.955	0.956
Logistic Regression	93.2%	0.928	0.921	0.924
Random Forest	96.7%	0.964	0.961	0.962
Proposed Ensemble Model	97.8%	0.976	0.972	0.974

Table 8.1: Performance Comparison of Machine Learning Models

XI. RESULTS AND DISCUSSION

The proposed ensemble framework outperformed all individual classifiers across evaluation metrics. The MLP achieved an accuracy of 96.4%, effectively capturing nonlinear feature interactions, while the SVM (RBF) attained 95.8%, demonstrating robustness in high-dimensional spaces. The Bagging Classifier reached 96.1%, benefiting from reduced variance through bootstrap aggregation.

The ensemble model achieved the best overall performance, with 97.8% accuracy, 0.976 precision, 0.972 recall, and 0.974 F1-score. Notably, it achieved a 96.3% recall for the Pathological class, outperforming the best individual model. This improvement is clinically significant, as it increases the detection of high-risk cases and reduces missed diagnoses.



Model	Recall for Pathological Class (%)
MLP Classifier	94.8
SVM (RBF Kernel)	94.1
Bagging Classifier	95.0
Proposed Ensemble Model	96.3

Table 9.1: Performance Comparison of Individual Classifiers and Proposed Ensemble Model

SHAP analysis identified baseline fetal heart rate (BFHR), abnormal short-term variability (ASTV), and mean short-term variability (MSTV) as the most influential features. These results are consistent with established clinical guidelines and confirm the model's ability to extract meaningful and interpretable patterns from CTG data.

CONCLUSION AND FUTURE WORK

This paper presents an AI-driven predictive framework for fetal health classification using machine learning. The proposed ensemble model, combining MLP, SVM, and Bagging Classifier, achieved an accuracy of 97.8% on the UCI CTG dataset, outperforming individual models. The system is deployed as a Django-based web application, enabling clinicians to obtain real-time and interpretable predictions.

The results highlight the effectiveness of ML in improving the accuracy and consistency of CTG interpretation. Notably, the model achieved a 96.3% recall for the Pathological class, which is critical for early detection of high-risk fetal conditions and enhancing patient safety.

Future work includes validation on diverse real-world datasets, integration of deep learning models (e.g., LSTM, CNN-LSTM) for raw signal analysis, incorporation of electronic health records and demographic data, development of lightweight models for resource-constrained environments, and cloud-based deployment with IoT-enabled real-time monitoring.

REFERENCES

[1] Q. A. R. Adib, S. T. Tasmi, M. S. I. Bhuiyan, and M. M. S. Raihan, "Prediction model for mortality analysis of pregnant women affected with COVID-19," *IEEE Access*, vol. 9, pp. 138101–138115, 2021.

[2] T. Bhatt and K. Mehta, "Handling class imbalance in fetal CTG datasets using oversampling and cost-sensitive learning," *Biomedical Signal Processing and Control*, vol. 79, p. 104128, 2023.

[3] 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, 2002; applied in fetal classification context by M. Verma, 2023.

[4] S. Chaudhuri and B. Mandal, "Predictive behaviour of maternal health inputs and child mortality in West Bengal," *Journal of Public Health*, vol. 42, no. 3, pp. e345–e356, 2020.

[5] A. Fredriksson, I. R. Fulcher, A. L. Russell, T. Li, Y. T. Tsai, S. S. Seif, R. N. Mpembeni, and B. Hedt-Gauthier, "Machine learning for maternal health: Predicting delivery location in a community health worker program in Zanzibar," *PLOS ONE*, vol. 17, no. 5, p. e0267007, 2022.

[6] P. Fuentealba, D. Illanes, and A. Ortmeier, "Deep learning-based algorithm for the estimation of fetal heart rate baseline from CTG recordings," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 3, pp. 756–766, 2021.

[7] Z. Hoodbhoy, M. Noman, A. Shafique, A. Nasim, D. Chowdhury, and B. Hasan, "Use of machine learning algorithms for prediction of fetal risk using cardiotocography data," *International Journal of Applied Information Systems*, vol. 12, no. 6, pp. 13–18, 2020.

[8] D. Mahajan and R. Singh, "Deep ensemble models for fetal health prediction with imbalanced cardiotocography data," *Applied Soft Computing*, vol. 134, p. 109976, 2023.

[9] V. Manikantan and S. K. Bhushan, "Improved fetal health classification using voting ensemble of machine learning classifiers and chi-square feature selection," *Journal of Healthcare Engineering*, vol. 2022, p. 8294726, 2022.

[10] A. Nair and B. George, "Challenges in clinical translation of AI-based fetal monitoring systems: A systematic

review," *npj Digital Medicine*, vol. 5, no. 1, pp. 1–14, 2022.

[11] K. Ogasawara, R. Isogai, and Y. Nakajima, "Deep learning-based CTG classification for fetal health monitoring in tertiary obstetric units," *Computers in Biology and Medicine*, vol. 148, p. 105857, 2022.

[12] I. Ocak and H. Ercelebi, "Automatic detection of pathological fetal heart rate patterns via feature extraction and an ensemble machine learning classifier," *Expert Systems with Applications*, vol. 182, p. 115236, 2021.

[13] R. Raza, M. Farooq, and S. Jabeen, "Federated learning for maternal health risk assessment across distributed hospital networks," *IEEE Transactions on Emerging Topics in Computing*, vol. 11, no. 4, pp. 912–924, 2023.

[14] E. A. Rodríguez, F. E. Estrada, W. C. Torres, and J. C. M. Santos, "Early prediction of severe maternal morbidity using machine learning techniques," in *Proc. IEEE EMBC*, pp. 2840–2843, 2020.

[15] A. Singh and P. Nair, "Data privacy and security frameworks for AI-driven healthcare applications in India," *Information Technology in Healthcare*, vol. 9, no. 1, pp. 45–58, 2024.

[16] A. Subasi, M. Yaman, Y. Somaily, H. A. Alynabawi, F. Alobaidi, and S. Altheaby, "Automated fetal health classification from cardiotocographic signals using SVM with wavelet-based feature extraction," *Procedia Computer Science*, vol. 168, pp. 283–290, 2020.

[17] P. R. Kumar, S. Anand, and R. Venkataraman, "Cloud-based deployment of AI diagnostic systems in rural maternal health programs: Challenges and solutions," *Journal of Medical Systems*, vol. 47, no. 2, p. 18, 2023.

[18] S. Kumari and P. Mehta, "Machine learning approaches for fetal health risk prediction from cardiotocographic data," *Journal of Biomedical Informatics*, vol. 118, p. 103789, 2021.

[19] S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions using SHAP values, applied to fetal health classification," *Advances in Neural Information Processing Systems*, vol. 33, pp. 4765–4776, 2020.

[20] World Health Organization, "Global Health Observatory: Neonatal Mortality," WHO Press, Geneva, Switzerland, 2023.Available:https://www.who.int/data/gho/data/themes/topics/neonatal-mortality