International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(2) 2026 pp. 2163-2179
https://doi.org/10.70917/ijcisim-2026-5547
Article

Comparative Performance Assessment of HyCIRRF and Traditional Machine Learning Models for Preterm Pregnancy Risk Prediction

Pritika Goel[1,*], Sachin Ahuja[2]

[1,2]Department of Computer Science and Engineering, Chandigarh University, NH-05 Chandigarh-Ludhiana Highway, Mohali, Punjab 140413, India
[1]goelpritika320@gmail.com; 2sachinahuja11@gmail.com
* Corresponding author: goelpritika320@gmail.com

Abstract: Nowadays, high-risk maternal pregnancy is a serious healthcare limitation by the combined influence of gestational, demographic, clinical, and physiological factors. Precision and timely prediction of maternal healthcare risks is crucial for reducing neonatal and maternal complications. Moreover, existing prediction method often exhibit limited capability in integrating heterogeneous maternal fetal data and providing personalize risk assessment. This research study proposes a hybrid crow search and independent component analysis optimized regression random forest (HyCIRFF) model for maternal health risk prediction and compares its performance with tradition machine learning (ML) methods, as support vector machine (SVM), XGBoost, logistic regression (LR), decision tree (DT) with particle swarm optimization (PSO), BiLTCN, and decision tree (DT). The proposed outline was calculated using the publicly available Kaggle Maternal Health Risk Dataset, which comprises clinical metrics like age, heart rate, temperature, BP, and blood glucose, etc., collected through Internet-of-Things (IoT) based on maternal monitoring systems. The database was divided into 183 training and 203 testing samples. The proposed HyCIRRF framework integrates Fast Independent Component Analysis (FastICA) for feature transformation, CatBoost for feature ranking, Crow Search Optimization (CSO) for optimal feature selection, and a Regression Random Forest Classifier with a soft-voting ensemble plan for classifying low-medium, and high-risk pregnancies. Simulation outcomes demonstrate that the proposed model attain superior prediction performance, attaining 93.0% precision, 94.0% recall, 93.0 % F1-Score, 93.0% accuracy, and a 7.0% error rate. In comparison, the optimized SVM-RBF with PSO model attained 86.0% accuracy, 83.3 % precision, 81.7% recall, 83.3% F1-score, and 14.0% error rate, while other baseline techniques that merging feature transformation, meta-heuristic optimization, and ensemble learning (EL) significantly improves maternal health risk prediction while maintaining clinical interpretability and supporting early decision-making in real-time healthcare applications.

Keywords: HyCIRRF, Maternal healthcare system, Preterm birth, Regression Random Forest, Traditional Machine Learning Models

1. Introduction

Delivery that occurs before 37 weeks of gestation is referred to as preterm birth (PTB). It is a major contributor to developmental issues, neonatal sickness, and mortality. According to gestational age, PTB is generally classified as very preterm (28 to <32 weeks), extremely preterm (<28 weeks) and moderate-to-late preterm (32 to 37 weeks) [1]. In 2020, nearly 13.4 million infants in all over world were delivered prematurely, representing more than one in ten live births. The burden is serious because complications associated with prematurity caused approximately 900,000 child deaths in 2019. Infants who survive may also experience long-term problems, including hearing, learning, and visual impairments [2]. Prematurity is biggest reason of mortality among children below 5 years of age. The chances of survival are widely different in different environments. Nearly one in two babies born in low-income countries at or before 32 weeks are at risk of dying due to lack of cost-effective and accessible care, including assistance with breastfeeding, infection control, and respiratory management. While in high income countries most premature infants

survive, in many middle-income countries, unequal access to medical technology can result in a significant disability burden for surviving preterm infants.

The accurate and timely prediction of PTB is therefore an important aspect of this research. The study compares well-known ML methods to understand their applicability in estimating the risk of PTB using clinical variables [3]. Because they can process a variety of health information, including electronic health records (EHRs) and measurements of transvaginal ultrasound, and because they can be more accurate than conventional assessment methods, ML-based prediction systems are being progressively adopted to decision-support tools [4]. EHR-based models have also demonstrated the potential of scalable prediction across independent patient cohorts and across various healthcare settings. The most frequently used prediction algorithms for PTB are SVM, LR, RF and similar classifiers [5].

Most of the previous research on PTB and maternal-risk prediction has focused on classical ML methods as SVM [4], LR [6], DT [6], RF [7] and boosting-based classifiers [8]. These methods are applicable to structured clinical data, but may not be effective if the maternal variables have nonlinear interactions and overlapping risk patterns. A single model may not provide a sufficiently balanced clinical decision-support outcome, as some baseline models have a high recall while others have better precision or accuracy. This limitation inspires the proposed hybrid scheme of independent feature extraction, optimized feature selection and ensemble-based risk classification.

Comparative research introduces combines CSO, Regression Random Forest (RRF) and Fast Independent Component Analysis (Fast-ICA) for PTB pregnancy-risk prediction. In contrast to using only simple feature selection, Fast-ICA is used as a feature-transformation step to reduce dimensionality and extract independent components from the original clinical variables. This transformation can enhance the information quality provided to the classifier and can aid in making more accurate predictions. CSO is then used to find an optimized reduced feature subset and then the complementary classifiers' output is combined using ensemble learning to get more stable predictions. By combining these, HyCIRRF aims to deliver more robust PTB risk classification than individual conventional ML methods.

In this study, the proposed HyCIRRF model is compared with LR, DT, SVM, RF, XGBoost, CatBoost, LightGBM and GBDT under the same experimental protocol. The proposed model produced 93% accuracy, 93% precision, 94% recall, 93% F1-score and a 7% error rate, showing a stronger balance between correct classification and high-risk case detection than the selected traditional models. The comparison supports the use of Fast-ICA based feature transformation, CatBoost feature ranking, CSO-based feature optimization and RRF-LR soft voting for more reliable PTB and pregnancy-risk prediction.

The further research article is written as follows. Section 2 shows the systematic and comparative review of literature on preterm-birth pregnancy-risk prediction and traditional machine-learning models. Section 3 defines the materials and methods, with the dataset, proposed framework, mathematical formulation, and baseline methods. Section 4 presents the results and comparative analysis. Section 5 shows the findings. Section 6 conclude the research with future scope.

2. Literature Review

PTB is one of the most difficult pregnancy outcomes to predict, as it is affected by many interacting maternal, clinical, biochemical, behavioral, and socioeconomic factors. Traditional risk assessment techniques are primarily based on known clinical risk factors and physician clinical judgment, but these methods may not adequately account for the nonlinear relationships between pregnancy-related variables that are heterogeneous. In current years, Deep Learning (DL) and ML framework have thus been the focus of many studies for the prediction of PTB, especially with the use of antenatal surveillance data, hospital records, preschool survey data, Electronic Health Records (EHRs), and routinely available maternal health indicators.

Yu et al. [9] built ML models for singleton pregnancies to predict PTB and showed that ensemble-based models can effectively predict PTB based on the routine antenatal variables. Similarly, Ding et al. [10] used the survey data

on large-scale from Shenzhen, China, and demonstrated that advanced classifiers, such as XGBoost, could not only improve the prediction of the risk of PTB but also provide interpretable analysis of the key predictors. Zhang et al. [11] compared the traditional ML models with DL models, and found that DL models have a better ability to capture complex patterns in pregnancy data.

Several recent studies have also expanded the prediction of PTB by adding clinical information and high-risk pregnancy data. Kloska et al. [12] discussed the different ML techniques for predicting PT and highlighted the importance of simplicity and interpretability of the model when working with a small number of samples. Qian et al. [13] built models for prediction of PTB using EHR and showed that the efficiency of the prediction can be increased through the use of longitudinal clinical information. Pang et al. [14] used interpretable ML models using systemic inflammatory markers to target women with gestational diabetes mellitus, where clinical characteristics of subgroups could complement the assessment of risk for PTB.

Besides PTB-specific research, maternal health risk prediction studies offer valuable methodological support for pregnancy risk modelling. Khadidos et al. [15] provided an ensemble ML framework for health risk prediction of maternal health, while Desuky et al. [16] optimized the parameters and addressed data imbalance in the prediction process. Al Mashrafi et al. [17] compared various ML models that could be used for maternal risk classification, and Malde et al. [18] demonstrated that sparse vital-sign data can be used to predict maternal health risk in low resource areas. Tzimourta et al. [19] also pointed out the role of Artificial Intelligence (AI) in supporting midwives to detect maternal risk through routine physiological measurements.

Umoren et al. [20] created a PTB risk framework for efficient birth results in a clinical setting in Nigeria using DTCR and achieved an accuracy of 89.2% on 120 clinical data samples, compared to an SVM comparator. The study highlighted the utility of simple decision rules using routinely collected parameters like blood pressure, blood glucose level, and weight, trimester and pregnancy month in early risk assessment in rural healthcare settings. Zhang et al. [21] built an EHR-based PTB prediction model with 5,411 subjects and compared LR, DTree, Naive Bayes, SVM and AdaBoost models. Their AdaBoost model was the most accurate (0.954) and had the highest AUC (0.93) and the most important predictors were preterm premature rupture of membranes, placenta previa, multiple pregnancies, and Streptococcus agalactiae colonization and prenatal hemorrhage.

Overall, recent research indicates that there is a clear shift from statistical risk assessment to prediction models using ML and DL for the classification of PTB and risks of maternal health. Ensemble models derived on trees such as XGBoost, RF, CatBoost, or LightGBM have been shown to be useful in modeling nonlinear relationships and heterogeneous clinical data, while DL models like transformers and LSTM have been found to be promising for extracting complex temporal patterns from large-scale hospital and EHR datasets. Interpretability methods have also become increasingly important, particularly SHAP, as clinical prediction models must explain the role of maternal, obstetric, biochemical and inflammatory predictors before they can be trusted in real-world healthcare settings.

But there are still some disadvantages. Many PTB prediction studies have been created on a single center cohort, retrospective data, sample size is small, or in a particular high-risk group, like Gestational Diabetes Mellitus (GDM) patients. The pregnancy risk model developed by Umoren et al. [17] demonstrates that an interpretable decision-tree model can be used to assist rural maternal risk classification, but its limited number of input variables, the small samples (120) in the clinic-based data, and the absence of external validation limit its generalizability to broader pregnancy risk screening. Likewise, Zhang et al. [18] reported high AdaBoost accuracy with EHR data, but also experienced EHR-related retrospective leakage, class imbalance, decreased accuracy for real preterm births in external validation, and poor generalizability due to validation using similar hospital EHRs. In addition, the performance measures vary from study to study, and are not easily comparable. Although some models may have reasonable Area under the Curve (AUC) or accuracy, they can also have a low sensitivity or low positive predictive value, which means that some high-risk cases may be missed or that some unnecessary clinical alerts may be triggered. In addition, wide-range maternal health risk models are not necessarily designed to predict PTB, but rather to categorize pregnancy risk as low, medium, or high, and thus may not be as useful for PTB-specific prevention strategies. There is a need for an

interpretable, clinically deployable, robust, clinically useful, PTB risk prediction framework that utilizes routinely collected maternal, clinical, obstetric, EHR-based, and vital-sign data, and that is able to handle class imbalance, missing values, external validation, multicenter generalizability, prediction for a specific pregnancy stage, and clinical utility. This can help to determine whether a pregnancy is at risk early and prompt actions to be taken to prevent poor maternal and neonatal outcomes.

3. Materials and Methods

The flow of the research work based PTB risk prediction is represented in the flowchart defined in Figure 1.

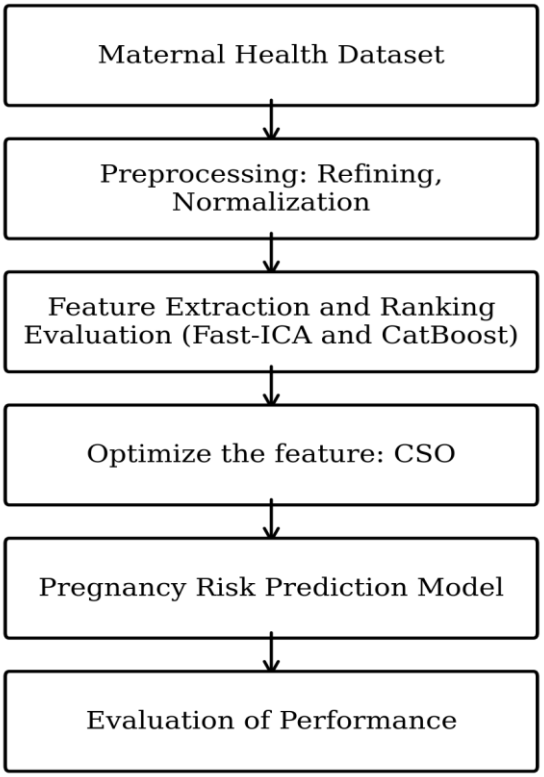

Figure 1. Flow of the Proposed Work

The dataset used of this research analysis comprises of secondary database attained from Kaggle with Maternal health Risk Database. This Maternal health risk database shows the patient based information such as blood pressure, age, ID, etc., After the database collection, it developed the preprocessing step to enhance the quality of data, and removing the unwanted information of the actual database. This is the main step to remove missing values and refine the database. Then, it developed the Fast-ICA method used for feature extraction to extract the reliable feature in the minimum time consumption. When feature extraction is done then it developed the feature ranking approach using CatBoost, it introduces the feature optimization method using Crow Search Optimization (CSO) method. These methods help to improve the feature quality and refine the feature set. Then, this optimized implemented the fitness function with involve the ensemble learning model using regression RF to enhanced the performance and PTB risk prediction in early stage. Further, it evaluate the evalustion metrics, as recall, f1-score, accuracy, precision, etc. Then, it validates and compares the performance metrics with existing implemented methods, such as LR, RF, SVM, etc.

3.1. Steps of the Proposed Work

The proposed steps are discussed in detailed below:

3.1.1. Dataset Collection

The experimental assessment is based on the Maternal Risk assessment Dataset of Kaggle data source and related public maternal-health benchmark repositories [22]. The aligned proposed protocol has 1,014 maternal records, six regular clinical predictors and a single categorical outcome variable. The outcome variable, RiskLevel, is coded as a three level outcome: mid risk, low risk, and high risk. There are no missing data in the provided protocol, which can be used for controlled comparative evaluation without imputation bias. The detail dataset defined in Table 1.

Table 1. Dataset Variables and Clinical Relevance

Feature	Description	Clinical relevance
Age	Maternal age in years	Very young or advanced maternal age may increase pregnancy-related risk.
SystolicBP	Systolic blood pressure	Elevated systolic pressure may indicate hypertensive complications.
DiastolicBP	Diastolic blood pressure	Diastolic elevation may indicate maternal cardiovascular stress.
BS	Blood sugar level	Abnormal glucose level may indicate gestational diabetes or metabolic risk.
BodyTemp	Temperature of body	Increased temperature may indicate infection or inflammatory response.
HeartRate	Maternal heart rate	Abnormal heart rate may reflect physiological stress.
RiskLevel	Low, mid, or high risk label	Three-class maternal-risk classification output.

3.1.2. Data Preprocessing

This process is used to refine the dataset, identify the missing values in the PTB dataset. Duplicate data were identified in the database; none were detected that shows the completeness in the database. By standardizing all integral variables to a normal scale, Z-score normalization enhances convergence and eliminates the predominance of high-magnitude features. To ensure that the actual class separation was handled within the subsets, an 80:20 database was divided into training and testing sets.

3.1.3. Feature Extraction and Importance

Fast-ICA was applied to transform interrelated clinical variables into statistically independent components. Unlike classical dimensionality reduction methods that mainly maximize variance, Fast-ICA focuses on non-Gaussianity, which enables separation of independent information sources within the maternal health data.

After feature extraction, the contribution of each clinical variable/component to PTB risk prediction was estimated using CatBoost. This method is suitable for structured healthcare datasets because it is robust and can identify relationships among variables while supporting feature-importance analysis.

3.1.4. Feature Selection

This process is established the best features that is the least redundant and increases classification performance accuracy rate. This selection process is used for CSO approach. It is a meta-heuristic approach which is inspired by the food-hiding and retrieval behavior of crows.

3.1.5. Hybrid Ensemble Learning (EL) Model

It is combines tree-based method that is probabilistic classifiers capable of handling both linear and non-linear model. This method is used to classify the prediction of pregnancy risk and improve the performance accuracy. The proposed framework has detail description in the section 3.2.

3.2. HyCIRRF Model

The research model is a sequential in nature and converts Maternal Health raw data into clinical important risk prediction. The workflow includes preprocessing, multi-stage feature engineering through extraction and ranking, optimized feature selection, and ensemble learning classification for pregnancy-risk prediction.

Initial raw records are cleaned, normalized, and separated into training and testing subsets. Fast-ICA then extracts features by reducing redundancy and identifying statistically independent latent components. CatBoost ranks the clinical features according to their predictive contribution, which also improves interpretability. The ranked features are further filtered using CSO, which searches for the most reliable subset by maximizing predictive accuracy.

The Hybrid EL classifier model, which is made up of RRF and LR, is then given an optimal feature set. The Soft Voting Architecture is a probabilistic structure that uses the results of both approaches to determine the ultimate classification of pregnancy risk in three stages: Low, High, and Mid. This multi-stage architecture increases the prediction accuracy rate and provides little overfitting. The HyCIRRF method's framework is depicted in Figure 2.

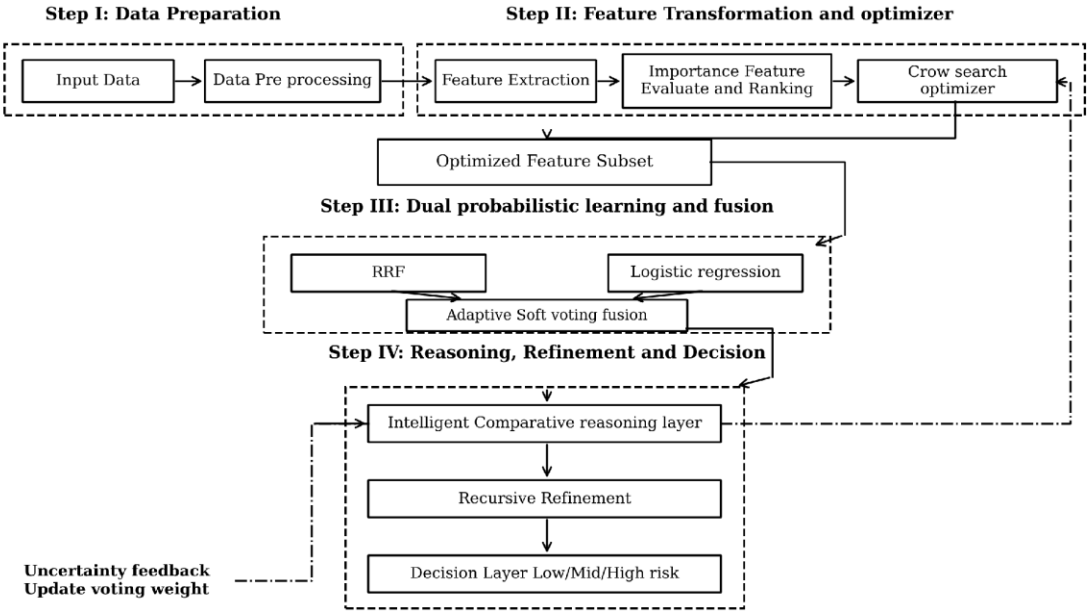

Figure 2. Workflow of the HyCIRRF Framework

3.2.1. Mathematical Formulation

This section formalizes the HyCIRRF pipeline. Eq (1) represents the supervised multi-class maternal-risk prediction dataset, where each clinical observation is associated with one of three risk classes.

$$D = \{(x_i, y_i)\}_{i=1}^{n},\ x_i \in \mathbb{R}^d,\ y_i \in C = \{\text{low, mid, high}\} \quad (1)$$

Where x_i denotes the i^{th} clinical feature vector, y_i denotes its corresponding risk class, d denotes number of input observations, and n is the complete count of observations.

After preprocessing and standardization, Fast Independent Component Analysis transforms each standardized observation x'_i into an independent latent component space. The transformation is expressed as in Equation (2):

$z_i = W\, x'_i, \quad Z = X'W^T$ (2)

Where W is the ICA unmixing matrix, x'_i is the normalized clinical feature vector, and Z represents the transformed independent component matrix.

The CatBoost-based feature/component importance score is calculated in Equation (3):

$$\phi_j = \frac{\Delta L_j}{\sum_{q=1}^{k} \Delta L_q}, \sum_{j=1}^{k} \quad \phi_j = 1 \quad (3)$$

Where ΔL_j denotes the prediction-loss change associated with the j^{th} component and k is the total indivisuals of transformed components.

In the CSO module in Equation (4), each crow represents a candidate feature subset using a binary mask:

$b_i = [b_i1, \ldots, b_ik], \; b_{ij} \in \{0,1\}, \; S_i = \{j \mid b_{ij} = 1\}$ (4)

Where $b_{ij} = 1$indicates that the j^{th} feature/component is selected, and S_i denotes the selected subset.

The subset optimization objective is defined in Equation (5):

$$J(S) = \alpha F1_{macro}(S) + \beta AUC_{macro}(S) - \gamma \frac{|S|}{k} - \lambda C(S) \quad (5)$$

Where $F1_{macro}$ and AUC_{macro} reward balanced classification and discrimination, $|\frac{|S|}{k}$ penalizes unnecessary feature expansion, $C(S)$ denotes computational cost, and α, β, γ, λ are weighting coefficients.

The optimized feature subset S^* is then passed to the RRF and LR learners. Their class probabilities are fused using adaptive soft voting in Equation (6):

$$P_f(x_i) = w_1 P_{RRF}(x_i, S^*) + w_2 P_{LR}(x_i, S^*), \; w_1 + w_2 = 1 \quad (6)$$

Where P_{RRF} and P_{LR} are the class-probability outputs of the RRF and LR learners, respectively.

The final predicted risk class is selected as shown in Equation (7):

$$\hat{y}_i = arg P_f\,(c|x_i) \quad (7)$$

Clinical performance is evaluated using performance metric in Equation (8), (9), (10), (11), and (12). These metrics are showed as:

$$Accu = \frac{TP+TN}{TP+FN+FP+TN} \quad (8)$$

$$Pre = \frac{TP}{FP+TP} \quad (9)$$

$$Recall = \frac{TP}{TP+FN} \quad (10)$$

$$F1 = \frac{2*PR}{R+P} \quad (11)$$

$$Error\ Rate = 1 - Accuracy \quad (12)$$

3.3. Baseline Methods

The following subsections describe the baseline methods used for comparative evaluation. LR provides a linear probabilistic baseline, DT offers rule-based classification, SVM evaluates margin-based decision boundaries, and RF/XGBoost-based models capture nonlinear feature interactions in tabular clinical data. These frameworks were selected due to widely used in PTB and maternal-risk classification and provide a fair reference against the proposed HyCIRRF framework. All methods were evaluated using performance metrics on the same training and testing split.

3.3.1. Logistic Regression (LR)

This models a binary dependent variable, described as 0/1 or Yes/No, from one or more independent variables. The equation (13) and (14) are defined as:

$$y/(1 - y) = \exp(A + B_1x_1 + B_2x_2 + \cdots + B_nx_n + \varepsilon) \quad (13)$$

$$\log(y/(1 - y)) = A + B_1x_1 + B_2x_2 + \cdots + B_nx_n + \varepsilon \quad (14)$$

In this formulation, x1, x2, ... xn denote the predictor variables, A is the intercept, B1, B2, ... Bn are regression coefficients, and ε shows the residual error. The model explains how a change in the dependent variable corresponds to a unit change in the independent variables. In the PTB dataset, the birth outcome is encoded in binary form [18].

3.3.2. Decision Tree (DT)

It is normally used supervised ML methods which separate the data into two parts in the DT [6]. This approach of dividing starts with a binary data split and endures until it is possible to make no further divisions. This ML method considered both regression and classification concepts. The decision outcome show a tree based chart which is called as DT. In the DT decision nodes, leaves, and branches are mainly three parts.

3.3.3. XGBoost Method

It is an enhanced gradient-boosting algorithm [7] [24]. It builds an ensemble of sequential decision trees, In which each new tree attempts to correct errors made by earlier trees by minimizing a differentiable loss function, including log loss in binary classification. The main XGBoost aim is given as:

$$Fn\ (\theta) = \sum_{I=1}^{N} \quad L(Y1, YI') + \sum_{K=1}^{t} \quad \Omega\ (FnK) \quad (15)$$

In eq (15) shows the $L(Y1, YI')$ loss function. The $\Omega\ (FnK)$ presents the regularization term, and t is the no. of trees. XgBoost is called for its speed and efficiency in managing structured information. In the blending architecture, XgBoost gives maximum accuracy and considers complex designs.

3.3.4. Support Vector Machine (SVM)

It is a robust supervised ML method used extensively for regression and classification tasks. It works in n-dimension space, verifying the optimized hyperplane which can discern data-points into two categories or classes as optimally as possible and so makes it especially reliable for prediction healthcare applications [4][25].

3.4. Comparative Methods

To ensure a rigorous comparison, the proposed model is evaluated against seven machine-learning and optimization-based techniques. These methods cover linear classification, tree-based learning, kernel-based classification, gradient boosting, swarm-optimized feature/model selection, deep temporal convolutional learning, and hybrid optimization-based classification shown in Table 2.

Table 2. Comparative Analysis of Traditional Machine-Learning Methods

Ref	Methods	Motivation	Strength	Limitation

[6]	LR	Used as a simple baseline model for classification problems.	Easy to implement, fast training, interpretable coefficients, works better for linearly separable data.	Performs poorly on complex non-linear relationships; sensitive to multicollinearity and feature scaling.
	DT	Used to model decision rules and capture non-linear patterns.	Readable tree structure; supports numeric and categorical variables; requires minimal preprocessing.	Prone to overfitting; unstable when data changes slightly; may have lower accuracy than ensemble methods.
[4]	SVM Linear	Used when the data is expected to be linearly separable or high-dimensional.	Effective in high-dimensional spaces; memory efficient; good generalization with proper regularization.	Not suitable for strongly non-linear data; sensitive to feature scaling; can be slow on large datasets.
	SVM Polynomial	Used to capture non-linear relationships using polynomial kernel transformation.	Can model complex decision boundaries; useful when interactions between features are important.	Computationally expensive; sensitive to kernel degree and hyperparameters; risk of overfitting.
	RF	Used to improve prediction accuracy by combining multiple decision trees.	Reduces overfitting compared with one tree; handles nonlinear structure; tolerant of noise and outliers.	Less transparent than a single tree; higher computation; parameter tuning may be needed.
	XGBoost	Used for high-performance gradient boosting with regularization.	High accuracy; handles missing values; includes regularization; efficient and widely used in structured/tabular data.	Requires careful hyperparameter tuning; less interpretable; can overfit if not properly controlled.
	CatBoost	Used for gradient boosting, especially with categorical features.	Handles categorical variables effectively; reduces preprocessing effort; strong performance on tabular data.	Training can be slower on large datasets; model interpretation is more difficult; parameter tuning may still be needed.
[8]	LightGBM	Used for fast and scalable gradient boosting on large datasets.	Very fast training; memory efficient; performs well on large-scale tabular data; supports categorical features.	Can overfit on small datasets; sensitive to parameter settings; less interpretable.
	GBDT	Used to build an additive ensemble of decision trees to improve prediction accuracy.	Strong predictive performance; captures non-linear relationships; flexible for classification and regression.	Slower than simpler models; sensitive to hyperparameters; can overfit without proper regularization.

4. Results

4.1 Evaluation Criteria and Parameter Configuration

To increase the level of reproducibility, the evaluation should be done with an 80:20 training-test split, normalizing the data by z-scores, fitting the Fast-ICA, ranking the features and selecting the CSO features, and tuning the model should be done only in the training partition. Report accuracy, macro-averaged precision, macro-averaged

recall, macro F1-score, error rate, class-wise sensitivity, and macro-AUC and 95% confidence intervals (if class-probability outputs are available) on the untreated test partition. The exact values of the hyperparameters that led to the already reported 93% result are not provided with the manuscript, so they cannot be checked retroactively. Therefore, Table 3 shows the fixed Parameter Configuration for Reproducible Reruns and External Validation.

Table 3. Fixed Parameter Configuration for Reproducible Reruns and External Validation

Component/model	Fixed configuration or tuning rule
Common evaluation protocol	Stratified 80:20 split; random seed = 42; stratified five-fold cross-validation within the training partition only; Z-score parameters fitted on training data; primary tuning metric = macro F1-score.
Fast-ICA	n_components = 6; algorithm = 'parallel'; tol = 1e-4; whiten = 'unit-variance'; max_iter = 1000; random_state = 42.
CatBoost feature ranking/baseline	iterations = 300; learning_rate = 0.05; depth = 6; loss_function = MultiClass; random_seed = 42; verbose = False.
Crow Search Optimization (CSO)	Population = 20 crows; max iterations = 50; flight length = 2.0; binary sigmoid transfer; seed = 42; probability of awareness = 0.10; early stopping when the convergence-rate criterion is satisfied for five consecutive iterations.
RRF/RF classifier	n_estimators = 300; criterion = 'gini'; max_features = 'sqrt'; min_samples_split = 2; min_samples_leaf = 1; class_weight = 'balanced'; random_state = 42.
Logistic Regression (LR)	C = 1.0; solver = 'lbfgs'; penalty = 'l2'; max_iter = 1000; multinomial probability output.
Adaptive soft voting	RRF:LR weights selected by inner cross-validation from {0.7:0.3, 0.6:0.4, 0.5:0.5, 0.4:0.6, 0.3:0.7}; final class selected by maximum fused probability.
Decision Tree (DT) baseline	criterion = 'gini'; min_samples_leaf in {1, 2, 4}; max_depth in {None, 3, 5, 10}; random_state = 42.
SVM baselines	C in {0.1, 1, 10}; gamma = 'scale'; probability = True; linear kernel and polynomial kernel with degree = 3.
XGBoost baseline	n_estimators in {100, 300}; max_depth in {3, 6}; learning_rate in {0.05, 0.10}; colsample_bytree = 0.80; subsample = 0.80; objective = multi:softprob; eval_metric = mlogloss; random_state = 42.
LightGBM and GBDT baselines	LightGBM: n_estimators in {100, 300}, num_leaves in {15, 31}, learning_rate in {0.05, 0.10}. GBDT: n_estimators in {100, 200}, max_depth in {2, 3}, learning_rate in {0.05, 0.10}. Random seed = 42.

4.2 CSO Objective-Function and Convergence Assessment

In the case of the fixed rerun protocol, the maximization objective of Equation (5) is assessed using the equivalent minimization loss of Equation (16). The first two terms are penalties for weak class-balanced prediction and discrimination, the third term is a penalty for unnecessarily large feature subsets, and the fourth term is a penalty for the normalized computational cost. All objective terms should be computed by stratified cross-validation in the training partition, with the test set completely excluded from the optimization process.

$$F(S) = 0.50[1 - F1\_macro(S)] + 0.30[1 - AUC\_macro(S)] + 0.15(|S|/k) + 0.05[C(S)/C\_max] \quad (16)$$

Convergence is assessed from the relative change in the best objective value between two successive iterations, as defined in Equation (17), where epsilon = 1e-12 prevents division by zero.

$$CR\_t = |F\_best(t-1) - F\_best(t)| \,/\, max(|F\_best(t-1)|, epsilon) \quad (17)$$

The CSO run is converged if CR_t < 1e-4 for 5 consecutive iterations or the run is stopped at 50 iterations. The following should be included in a complete convergence report: Median and interquartile range of the best objective value over at least 30 independent seeds; iteration at which 95% of the total objective improvement is achieved; Mean runtime. Since the final classification metrics are the only ones that are preserved in the current manuscript materials, a numerical convergence curve or a measured convergence-rate value cannot be reconstructed without rerunning the experiment, and no convergence value is invented in this revision.

4.3 External and Real-Time Validation Scope

The Kaggle and UCI versions of the maternal health data used in this study seem to be repository mirrors of the same underlying data set and should not be viewed as separate data sets without record level verification and deduplication; otherwise, duplicated observations can lead to over-estimation of performance. Recently a truly independent IoT-enabled maternal-health dataset with 6,058 quality-checked records from nine healthcare facilities and expert validated high, mid and low-risk labels has been made public [23]. It has the same age, temperature, systolic and diastolic blood pressure, heart rate, and glucose-related measurements as the present study, but direct pooling still needs to be unit harmonized, feature mapped, labeled, and patient independent checked. No external or real-time performance value is claimed here since the raw external cohort was not provided with the supplied study materials and the HyCIRRF pipeline was not re-run on the same. A valid follow-up should use the full development cohort to train the pipeline and the harmonized external cohort to test the pipeline, without any changes, or a prospective IoT stream should be collected with patient-level grouping and timestamps to ensure that measurements from the same patient are not in both the training- testing sets.

It confirms the evaluation and validation of the performance comparison between proposed HyCIRRF, and existing models such as SVM, LR, DT, XGBoost, etc. These methods were trained and calculated under identified conditions to remove difficult variables. Normally, the database was reliably divided into 813 training, and 203 testing data values for different traditional ML models. This research work used for Python simulation tool with graphical representation. The calculation parameters were used to calculate method efficiency in a manner which involve the mathematical accuracy, precision, recall, etc., and clinical implication. The research model implemented was calculated with five performance metrics like precision, accuracy, f1-score, recall, and error rate. The outcomes are given Table 4, and visualization is delivered in Figure 3.

Table 4. Evaluation of the Proposed HyCIRRF Model

Parameters	**Values (%)**
Accuracy (acc)	93
Precision (pre)	93
Recall (rec)	94
F1-score	93
Error	7

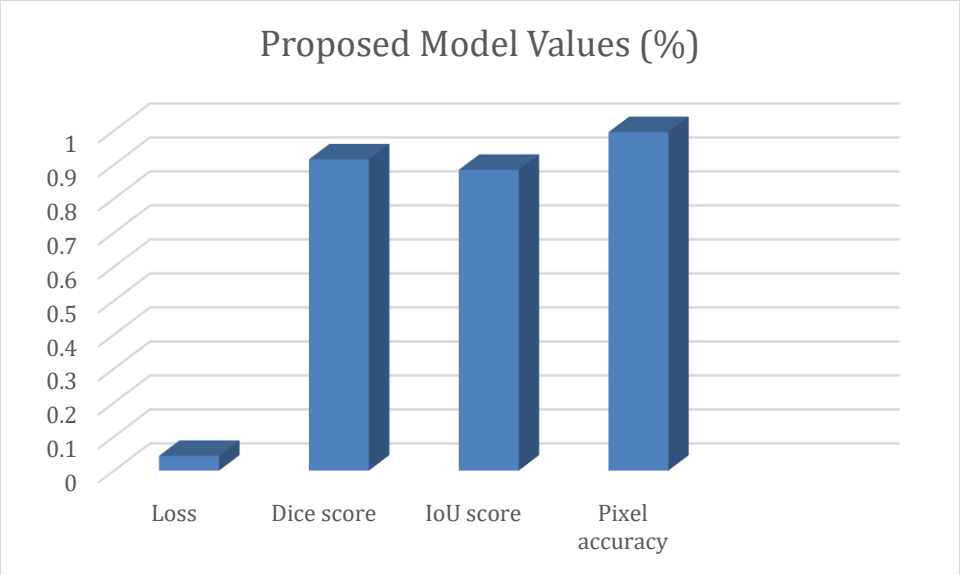

Figure 3. Evaluation Performance of the HyCIRRF Model

The results of the suggested HyCIRRF model is shown in figure using five important classification metrics namely precision, accuracy, recall, F1-score, and error rate. The method demonstrated high accuracy, F1 score and precision, of 93.0% and a recall score of 94.0%, demonstrating its strong predictive power and balanced classification performance. Additionally, the low error rate of 7.0% indicates the effectiveness and reliability of the suggested model in determining high-risk pregnancy cases with minimal misclassification.

4.4. Comparative Result Analysis

As shown in Table 5, the suggested HyCIRRF model achieved the strongest overall efficiency among the compared methods, with 0.94 recall, 0.93 accuracy, 0.93 F1-score and 0.93 precision. The best existing accuracy in the comparative table was obtained by XgBoost [8] with 0.83 accuracy and 0.84 F1-score, while SVM Linear [4] also achieved a competitive F1-score of 0.84 with 0.82 accuracy. Compared with XgBoost [8], HyCIRRF improved accuracy by 0.10 and F1-score by 0.09, which demonstrates the benefit of combining Fast-ICA feature extraction, CatBoost ranking, CSO feature selection and RRF-LR soft voting. Although LR [6] produced the highest recall of 1.00, its precision was only 0.65 and its F1-score was 0.79, indicating a higher possibility of false-positive risk classification. Similarly, DT [6], SVM Polynomial [4], RF [4], RF [7], LightGBM [8] and GBDT [8] showed lower accuracy and F1-score values than the proposed model. Overall, the results confirm that HyCIRRF provides a more balanced and reliable prediction performance for maternal/PTB risk classification than the traditional and boosting-based baseline models considered in this study.

Table 5. Quantitative Performance Analysis of the Proposed and Existing Models

Ref	Models	Accuracy	Recall	Precision	F1-score
[6]	LR	0.78	1.00	0.65	0.79
	DT	0.67	0.91	0.56	0.69
[4]	SVM Linear	0.82	0.86	0.83	0.84
	SVM Polynomial	0.66	0.71	0.69	0.70
	RF	0.74	0.86	0.73	0.79

	LR	0.80	0.82	0.82	0.82
	XGBoost	0.76	0.82	0.77	0.79
	Catboost	0.76	0.82	0.77	0.79
	DT	0.74	0.71	0.80	0.75
[7]	RF	0.79	0.68	0.72	0.70
[8]	XgBoost	0.83	0.84	0.84	0.84
	LR	0.79	0.80	0.79	0.80
	LightGBM	0.77	0.79	0.79	0.78
	GBDT	0.77	0.79	0.77	0.78
HyCIRRF Model (Proposed)		0.93	0.94	0.93	0.93

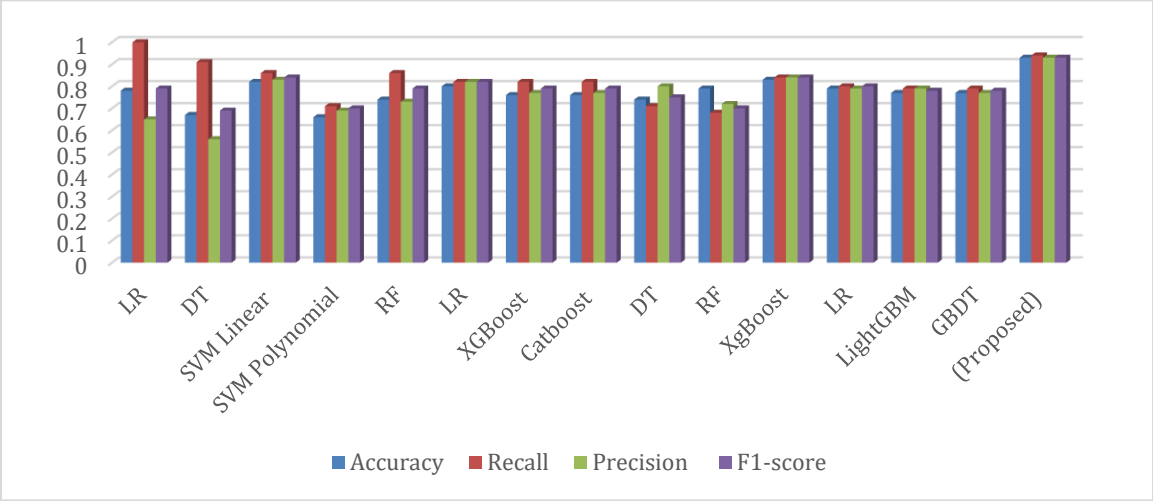

Figure 4. Efficiency Comparison of the Proposed HyCIRRF Framework with Existing Methods

Figure 4 visualize that the proposed HyCIRRF model achieved strong and balanced performance for maternal risk prediction. It obtained 93% accuracy, 93% precision, 94% recall, 93% F1-score, and only 7% error rate. The high recall is especially important in maternal healthcare because correctly identifying high-risk pregnancy cases can support early clinical intervention and reduce the chances of adverse outcomes.

5. Discussion

The proposed model achieved 93% accuracy, 94% recall, 93% precision, and 93% F1-score, which are better than all the benchmark methods in all the evaluated metrics. In addition, the model had a 7% error rate, which indicates the reliability of the model for predicting maternal health risks. The proposed HyCIRRF improved accuracy by 15 percentage points (19.2% relative improvement) compared to Logistic Regression (LR) model reported in [6] with 78% accuracy. The F1 score improved from 0.79 to 0.93, which is a 17.7% improvement. The most interesting result is that LR in [6] had a perfect recall (1.00), but a low precision (0.65), which means that many false positives were generated. HyCIRRF, on the other hand, had a more balanced efficiency with a 94% recall and 93% precision.

The results of the proposed HyCIRRF model were compared with the Decision Tree (DT) model in [6] and the accuracy was improved by 26 percentage points (38.8% relative gain) and F1-score was improved by 34.8% with the proposed model.

The Linear SVM was the most successful model among the models reported in [4] with an accuracy of 82% and an F1 score of 0.84. The accuracy of HyCIRRF was 11 percentage points higher than these values (13.4% relative improvement), and the F1-score was 10.7% higher. The improvements were even more significant when compared to the Polynomial SVM, which saw a 27% increase in accuracy (40.9%) and a 32.9% increase in F1-score.

In ensemble-based approaches in [4], the proposed model also outperformed other models.

HyCIRRF outperforms Random Forest (74% accuracy, 0.79 F1-score) with 19 percentage points (25.7%) higher accuracy and 17.7% higher F1-score. Likewise, the proposed model outperformed XGBoost and CatBoost (both 76% accuracy and 0.79 F1-score) by 22.4% and 17.7% in terms of accuracy and F1-score, respectively.

In [7] the accuracy of the RF model was 79% and the F1-score was 0.70. It was outperformed by HyCIRRF in accuracy (17.7% relative gain) and F1-score (32.9%) which shows that the classification consistency was significantly better.

The best alternative model was the XGBoost classifier in [8] with 83% accuracy, 84% precision, 84% recall, and 84% F1-score. Despite this strong baseline, HyCIRRF achieved improvements of 10 percentage points in accuracy (12.0% relative improvement), 9 percentage points in precision (10.7%), 10 percentage points in recall (11.9%), and 9 percentage points in F1-score (10.7%). The proposed model outperformed LightGBM and GBDT by about 16-21% on the important evaluation parameters.

Overall, the results show that the proposed HyCIRRF framework efficient than the classical ML techniques and advanced ensemble learners. It has a balanced precision and recall, which helps reduce both false-+ve and false--ve predictions, making it particularly well suited for maternal healthcare applications where accurate prediction of high-risk pregnancies is paramount for timely medical intervention and increased outcomes of maternal healthcare.

5.1. Limitations

There are a number of limitations to this study. First, the reported 93% accuracy is from internal evaluation of one relatively small, cross-sectional public dataset, so the estimate may be sensitive to the specific split of the data and should not be interpreted as clinical readiness or generalizability to the broader population. There was no prospective, temporal, multicentre, real-time or independent external validation performed in the reported experiment. Furthermore, the three level RiskLevel target is not a directly adjudicated preterm-birth outcome, and therefore should be interpreted as maternal-health risk stratification, unless the PTB-specific outcome labels are available.

Secondly, the six input variables exclude the following clinically relevant risk pathways: gestational age, obstetric history, parity, previous preterm birth, cervical length, laboratory markers, medication, lifestyle, socioeconomic context, and longitudinal changes, which could reduce the sensitivity of the model to clinically relevant pathways. Fast-ICA can enhance the separability of correlated information, but also changes the original variables into latent components, which may not be directly interpretable in clinical settings. Other features, such as CatBoost ranking, CSO search, RRF, and soft voting also add to the complexity of the model and multiple layers where overfitting or data leakage can happen if not fitted strictly within the training folds.

Third, the actual versions of the software, the historical hyperparameter log, random seeds, selected feature subset, the hyperparameter iteration trace of the CSO, runtime, calibration, confidence intervals, class-wise confusion matrix, and repeated-run variability were not included in the manuscript materials. Therefore, it is difficult to fully evaluate the advantage reported at this time because it is not yet reproducible or statistically stable. The comparative table also includes results from various studies and data settings, which will facilitate the context of comparison, but will not be considered a head-to-head comparison of the same cohort and experimental protocol.

Lastly, there was no assessment of fairness across age, geography, socioeconomic status, and clinical subgroups, of clinical utility, alert burden, or of decision-curve benefit. Prior to deployment, the model must be validated, calibrated, tested at independent sites, split at the patient level, have an error analysis that is transparent, and be reviewed by clinicians for false negative high-risk cases.

6. Conclusions and Future Scope

This study proposed the HyCIRRF model for maternal risk prediction using a hybrid framework that combines Fast-ICA feature extraction, CatBoost feature ranking, CSO-based feature selection, and Regression Random Forest–Logistic Regression soft voting classification. The proposed model achieved 93% F1-score, 93% accuracy, 94% recall, 93% precision, and only 7% error rate, showing strong and balanced classification performance. Compared with current models like LR, DT, SVM, RF, XGBoost, LightGBM, and GBDT, HyCIRRF provided better overall performance, especially in maintaining both high recall and high precision. These results confirm that the proposed model can effectively support early maternal health risk classification and may assist healthcare professionals in identifying low, mid, and high-risk pregnancy cases.

In future work, the proposed model can be improved by validating it on larger and more diverse maternal health datasets collected from different hospitals and clinical environments. Cross-validation and external validation should be performed to confirm the generalizability and robustness of the model. Further research can also include additional clinical, demographic, lifestyle, and laboratory features to improve prediction reliability. Moreover, explainable AI approach can be integrated to make the model decisions more transparent for healthcare professionals. Finally, the HyCIRRF framework can be implemented in a real-time clinical decision-support system for early pregnancy risk screening and timely medical intervention.

Nomenclature

Acc	Accuracy
AI	Artificial Intelligence
AUC	Area Under the Curve
BP	Blood Pressure
BS	Blood Sugar
CatBoost	Categorical Boosting
CSO	Crow Search Optimization
DL	Deep Learning
DT	Decision Tree
DTCR	Decision Tree Classification and Regression
EHR	Electronic Health Records
EL	Ensemble Learning
Fast-ICA	Fast Independent Component Analysis
GBDT	Gradient Boosting Decision Tree
GDM	Gestational Diabetes Mellitus
HyCIRRF	Hybrid Crow Search and Independent Component Analysis Optimized Regression Random Forest
ICA	Independent Component Analysis
IoT	Internet of Things
LightGBM	Light Gradient Boosting Machine
LR	Logistic Regression
LSTM	Long Short-Term Memory

ML	Machine Learning
pre	Precision
PSO	Particle Swarm Optimization
PTB	Preterm Birth
RBF	Radial Basis Function
rec	Recall
RF	Random Forest
RRF	Regression Random Forest
SHAP	SHapley Additive exPlanations
SVM	Support Vector Machine
XGBoost	Extreme Gradient Boosting

Acknowledgments

The authors acknowledge the UCI Machine Learning Repository and Kaggle for access to the Maternal Health Risk dataset used in this study.

Data Availability Statement

The data used in this study are publicly available from the UCI Machine Learning Repository as the Maternal Health Risk dataset, cited in Reference [22].

Author Biographies

Pritika Goel is affiliated with the Department of Computer Science and Engineering, Chandigarh University, Mohali, India. Her work in this article concerns machine-learning-based maternal-health risk prediction.

Sachin Ahuja is affiliated with the Department of Computer Science and Engineering, Chandigarh University, Mohali, India. His work in this article concerns machine-learning-based clinical decision support.

REFERENCES

1. World Health Organization, “Preterm birth,” May 10, 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/preterm-birth
2. E. Bradley, H. Blencowe, A. B. Moller, Y. B. Okwaraji, F. Sadler, A. Gruending, et al., “Born too soon: Global epidemiology of preterm birth and drivers for change,” Reproductive Health, vol. 22, Suppl. 2, Art. no. 105, 2025.
3. C. Huang, X. Long, M. van der Ven, M. Kaptein, S. G. Oei, and E. van den Heuvel, “Predicting preterm birth using electronic medical records from multiple prenatal visits,” BMC Pregnancy and Childbirth, vol. 24, Art. no. 843, 2024.
4. A. Kloska, A. Harmoza, S. M. Kloska, T. Marciniak, and I. Sadowska-Krawczenko, “Predicting preterm birth using machine learning methods,” Scientific Reports, vol. 15, no. 1, Art. no. 5683, 2025.
5. B. F. Narice, M. Labib, M. Wang, V. Byrne, J. Shepherd, Z. Q. Lang, et al., “Developing a logistic regression model to predict spontaneous preterm birth from maternal socio-demographic and obstetric history at initial pregnancy registration,” BMC Pregnancy and Childbirth, vol. 24, Art. no. 688, 2024, doi: 10.1186/s12884-024-06892-3.
6. R. K. Saroj and M. Anand, “Environmental factors prediction in preterm birth using comparison between logistic regression and decision tree methods: An exploratory analysis,” Social Sciences & Humanities Open, vol. 4, no. 1, Art. no. 100216, 2021, doi: 10.1016/j.ssaho.2021.100216.
7. A. Hassan, S. Nawaz, S. Tahira, and A. Ahmed, “Preterm birth prediction using an explainable machine learning approach,” in Artificial Intelligence and Applications, Mar. 2025.
8. X. Teng, M. Liu, Z. Wang, and X. Dong, “Machine learning prediction of preterm birth in women under 35 using routine biomarkers in a retrospective cohort study,” Scientific Reports, vol. 15, no. 1, Art. no. 10213, 2025.
9. Q. Y. Yu, Y. Lin, Y. R. Zhou, X. J. Yang, and J. Hemelaar, “Predicting risk of preterm birth in singleton pregnancies using machine learning algorithms,” Frontiers in Big Data, vol. 7, Art. no. 1291196, 2024, doi: 10.3389/fdata.2024.1291196.
10. L. Ding, X. Yin, G. Wen, D. Sun, D. Xian, Y. Zhao, et al., “Prediction of preterm birth using machine learning: A comprehensive analysis based on large-scale preschool children survey data in Shenzhen of China,” BMC Pregnancy and Childbirth, vol. 24, Art. no. 810, 2024, doi: 10.1186/s12884-024-06980-4.

11. F. Zhang, L. Tong, C. Shi, R. Zuo, L. Wang, and Y. Wang, “Deep learning in predicting preterm birth: A comparative study of machine learning algorithms,” Maternal-Fetal Medicine, vol. 6, no. 3, pp. 141-146, 2024, doi: 10.1097/FM9.0000000000000236.
12. A. Kloska, A. Harmoza, S. M. Kloska, T. Marciniak, and I. Sadowska-Krawczenko, “Predicting preterm birth using machine learning methods,” Scientific Reports, vol. 15, Art. no. 5683, 2025, doi: 10.1038/s41598-025-89905-1.
13. L. Qian, H. Jia, Z. Chang, Y. Hu, C. Chen, X. Li, et al., “Predicting the risk of preterm birth with machine learning and electronic health records in China,” BMC Medical Informatics and Decision Making, vol. 25, Art. no. 415, 2025, doi: 10.1186/s12911-025-03254-7.
14. Q. Pang, L. Peng, J. Wu, Y. Wang, R. Zhang, Z. Liu, et al., “Comparison of interpretable machine learning models using systemic inflammation index to predict preterm birth in gestational diabetes mellitus,” International Journal of Women’s Health, vol. 18, pp. 1-16, 2026, doi: 10.2147/IJWH.S541610.
15. A. O. Khadidos, F. Saleem, S. Selvarajan, Z. Ullah, and A. O. Khadidos, “Ensemble machine learning framework for predicting maternal health risk during pregnancy,” Scientific Reports, vol. 14, Art. no. 21483, 2024, doi: 10.1038/s41598-024-71934-x.
16. A. S. Desuky, S. Hussain, M. A. Cifci, L. M. El Bakrawy, O. Mzoughi, and N. Kraiem, “Parameter optimization based Mud Ring Algorithm for improving the maternal health risk prediction,” IEEE Access, vol. 12, pp. 167245-167261, 2024, doi: 10.1109/ACCESS.2024.3495518.
17. S. S. Al Mashrafi, L. Tafakori, and M. Abdollahian, “Predicting maternal risk level using machine learning models,” BMC Pregnancy and Childbirth, vol. 24, Art. no. 873, 2024, doi: 10.1186/s12884-024-07030-9.
18. A. Malde, B. Payne, M. A. Brown, and T. Montgomery-Csoban, “A machine learning approach for predicting maternal health risks in lower-middle-income countries using sparse data and vital signs,” Future Internet, vol. 17, no. 5, Art. no. 190, 2025, doi: 10.3390/fi17050190.
19. K. D. Tzimourta, V. C. Pezoulas, T. P. Exarchos, and D. I. Fotiadis, “Maternal health risk detection: Advancing midwifery with artificial intelligence,” Healthcare, vol. 13, no. 7, Art. no. 833, 2025, doi: 10.3390/healthcare13070833.
20. I. Umoren, F. Chigozirim, A. Silas, and B. Ekong, “Modeling and prediction of pregnancy risk for efficient birth outcomes using Decision Tree Classification and Regression model,” ResearchGate preprint, Apr. 2022.
21. Y. Zhang, S. Du, T. Hu, S. Xu, H. Lu, C. Xu, J. Li, and X. Zhu, “Establishment of a model for predicting preterm birth based on the machine learning algorithm,” BMC Pregnancy and Childbirth, vol. 23, Art. no. 779, 2023, doi: 10.1186/s12884-023-06058-7.
22. University of California Irvine Machine Learning Repository, “Maternal Health Risk,” 2023. [Online]. Available: https://archive.ics.uci.edu/dataset/863/maternal+health+risk
23. M. M. Hossain, N. M. Nayan, and M. A. Kashem, “A comprehensive maternal health risk prediction dataset from IoT-enabled medical cyber-physical systems in developing countries: supporting machine learning and deep learning applications for clinical decision support,” BMC Medical Informatics and Decision Making, vol. 26, Art. no. 79, 2026, doi: 10.1186/s12911-026-03343-1.
24. L. Bercovich, “Risk factors for preterm labor,” M.S. thesis, Dept. Gynecol. Obstet., School of Medicine, Univ. of Zagreb, Zagreb, Croatia, 2026.
25. C. Brickmann *et al*., “PROMISE—Impact of maternal and neonatal risk factors on the respiratory outcome of extremely preterm infants following PPROM in the second trimester of pregnancy,” *Front. Pediatr.*, vol. 14, Art. no. 1776970, 2026.