International Journal of Statistical Sciences *ISSN 1683-5603*
Vol. 25(2), November, 2025, pp 59-80
DOI: https://doi.org/10.3329/ijss.v25i2.85735

Comparative Evaluation of Machine Learning Models for Maternal Health Risk Prediction Using IoT-Based Data

Md. Mahfuz Uddin and Md. Rezaul Karim*

Department of Statistics, University of Rajshahi, Rajshahi-6205, Bangladesh
Emails: mahfuz.ru.stat.58@gmail.com; mrkarim@ru.ac.bd
*Correspondence should be addressed to Md. Rezaul Karim
(Email: mrkarim@ru.ac.bd)

[Received July 21, 2025; Accepted 20 October, 2025]

Abstract

Maternal health risk prediction is crucial for reducing maternal mortality, a key concern of the United Nations' Sustainable Development Goals (SDGs). The main objective of the study is to evaluate the predictive performance of various machine learning (ML) models, identify the most significant physiological risk factors, and provide a data-driven approach for improving maternal healthcare interventions. This study utilizes ML models to classify maternal risk levels using an Internet of Things (IoT)-based dataset obtained from the UCI ML repository, which was collected from rural areas of Bangladesh. The dataset comprises 1014 observations with six physiological independent variables: Age, Systolic Blood Pressure, Diastolic Blood Pressure, Blood Sugar, Body Temperature, and Heart Rate. The categorical dependent variable represents the Risk Level. The Boruta algorithm and the Regularized Random Forest (RRF) method are applied for feature selection to enhance model efficiency. Various ML algorithms, including Multinomial Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), XGBoost, Naïve Bayes, Ranger, and LogitBoost, are evaluated using 5-fold cross-validation. Performance metrics, including accuracy, sensitivity, specificity, precision, F1-score, false discovery rate (FDR), and area under the curve (AUC), are considered to compare the effectiveness of the models. Results indicate that among all ML models, Random Forest emerges as the top-performing model, achieving the highest accuracy (89.1%), macro-AUC (0.970), and weighted-AUC (0.968), and balanced performance across all classes. Blood Sugar is identified as the most critical predictor, followed by Systolic Blood Pressure and Age. Heart Rate and Body Temperature contribute minimally. The findings highlight the potential of ML techniques in enhancing the early detection of maternal health risks, thereby enabling timely interventions to improve healthcare outcomes.

Keywords: Maternal health, Machine learning, IoT-based monitoring, Risk prediction, Feature selection, Ensemble methods.

AMS Classification: 62P10, 68T09.

1. Introduction

The maternal mortality ratio is the number of women who die during pregnancy for every 100,000 live births [1]. Maternal mortality continues to be a major worldwide health challenge, particularly in developing nations like Bangladesh, where rural areas lack adequate access to high-quality maternal healthcare. According to the World Health Organization (WHO), over 300,000 women lost their lives in 2017 due to pregnancy-related causes, which is 808 women per day [2]. Two-thirds (200000) of maternal fatalities worldwide occur in sub-Saharan Africa, while just 19% (57000) occur in South Asia, according to a regional analysis of maternal deaths [3]. Most pregnant women who die from known and preventable pregnancy and delivery complications live in low- and middle-income countries, including developing countries, where the risk factors for maternal mortality are poorly understood. Maternal health problems can occur during pregnancy, childbirth, and the postpartum period [4]. Pregnancy-related health problems are more common in women and can sometimes lead to miscarriage and death. All phases of pregnancy must be pleasurable to ensure the health of the woman and their unborn children. A woman's chances of remaining healthy in the future are increased if she maintains her health during her pregnancy and after giving birth. These women give birth to children who have a positive impact on their childhood, adolescence, and adulthood [5].

Recent advancements in Internet of Things (IoT)-based healthcare systems have enabled the real-time monitoring of maternal health indicators, allowing for timely risk assessment and improved clinical decision-making [6]. Various physiological factors, including age, blood pressure, blood sugar levels, body temperature, and heart rate, have been identified as significant predictors of maternal health risks [7]. However, traditional clinical assessment methods often fail to provide early warning signals, leading to delayed interventions and increased mortality rates. Techniques for machine learning have become increasingly effective tools for the prediction of healthcare, offering higher accuracy, automation, and the ability to handle complex data relationships [8]. Study [9] aimed to employ ML algorithms to assess pregnancy risk levels based on key risk factors and found that the logistic model tree (LMT) outperforms other models, demonstrating its strong predictive ability. Study in [10] introduced decision tree-based ML models, including Bagging, Decision Tree, Random Forest, and AdaBoost, to support pregnancy risk prediction within IoT-based healthcare systems. With 50 trees and an accuracy of 82%, the Random Forest model performed best on the Maternal Health Risk dataset, which was taken from Kaggle. It had a high recall for both high-risk (88%) and low-risk (85%) classes. However, the recall for medium-risk was 73%. The study [11] showed that the decision tree algorithm produced better performance among six algorithms for the data set. However, few studies have examined the performance of different machine learning algorithms in predicting maternal mortality risk using IoT-based data in rural Bangladesh.

The IoT-based research [12] integrated a medical device for maternal healthcare from Bangladesh perspective. In addition to developing the technology, they collected patient data in its raw form and uploaded it to a cloud system. It was processed in the cloud, and graphs representing the resulting data were displayed. The danger levels were deduced from these graphs. The main objective of that research was to minimize the expense of maternity healthcare in developing nations such as Bangladesh. As a result, maternal healthcare facilities were improved, resulting in a reduction in maternal and infant mortality. Several of the system's modules were introduced, including sensor-based data collection, cloud-based data accumulation and communication link establishment, cloud-based raw data storage, and open-source software for data analysis. The purpose of the article [13] was to provide healthcare practitioners with accurate tools for early risk assessment and action by investigating the critical field of maternal health risk prediction. The dataset under investigation comprised 1102 meticulously collected instances, sourced from the nearest hospital, and included 12 key characteristics. XGBoost outperformed the other nine algorithms used, employing ML techniques. The primary goal of that study was to simplify the accurate and thorough classification of maternal health risk factors, enabling prompt and targeted interventions. The research [14] introduced an ANN-based approach for the prediction of maternal health risks applying clinical and IoT-based health records. A deep learning model, DT-BiLTCN, was developed to predict maternal health risks using 1,218 IoT-based health records. It integrated decision trees, BiLSTM, and TCN. Class imbalance is addressed using the Synthetic Minority Oversampling Technique (SMOTE). The model achieved the highest performance via SVM. Exploratory data analysis highlighted that systolic and diastolic blood pressure, heart rate, and age are the most significant indicators of maternal health risks during pregnancy.

Although several prior studies have deployed ML techniques to predict maternal health outcomes, there exists a significant disparity in the application of advanced feature selection techniques, namely Boruta algorithm, Regularized Random Forest (RRF), LASSO, Information Gain, Relief, etc., for systematically identifying the most influential predictors of maternal health risk [9-14]. Moreover, existing literature has seldom incorporated ensemble and boosting algorithms, such as XGBoost, Ranger, and LogitBoost, which are particularly effective in capturing complex nonlinear patterns within clinical and sociodemographic data. This review of literature highlights a substantial methodological shortcoming, particularly in leveraging cutting-edge machine learning frameworks to enhance predictive accuracy and gain deeper insights into maternal risk profiles. To address these limitations, this study employs various statistical tools and techniques as an extension of previous studies, including descriptive analysis, correlation analysis, normality test, non-parametric test, advanced feature selection techniques (Boruta algorithm and RRF), and the examination of feature effects and contributions to the models. It also applies nine ML algorithms including Multinomial Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, K-Nearest Neighbor, Extreme Gradient Boosting (XGboost), Naïve Bayes, Ranger, and LogitBoost to predict maternal health risk levels based on real-world data collected from hospitals, community clinics, and maternal healthcare centers in rural Bangladesh. Our research aims to (i) evaluate the predictive performance of nine ML models, (ii) identify the most significant physiological risk factors, and (iii) provide a data-driven approach for improving maternal healthcare interventions. The findings of this study can contribute to enhancing early risk detection, supporting healthcare decision-making, and ultimately reducing maternal mortality rates. The systematic workflow diagram of the study process is presented in Figure 1.

The remainder of the study is structured as follows: Section 2 defines the materials and methods, including the data source, variables, statistical analysis, feature selection methods, and ML algorithms. Section 3 presents data analysis results and their interpretations. Section 4 outlines the discussion. Section 5 summarizes the study's conclusions and outlines future extensions.

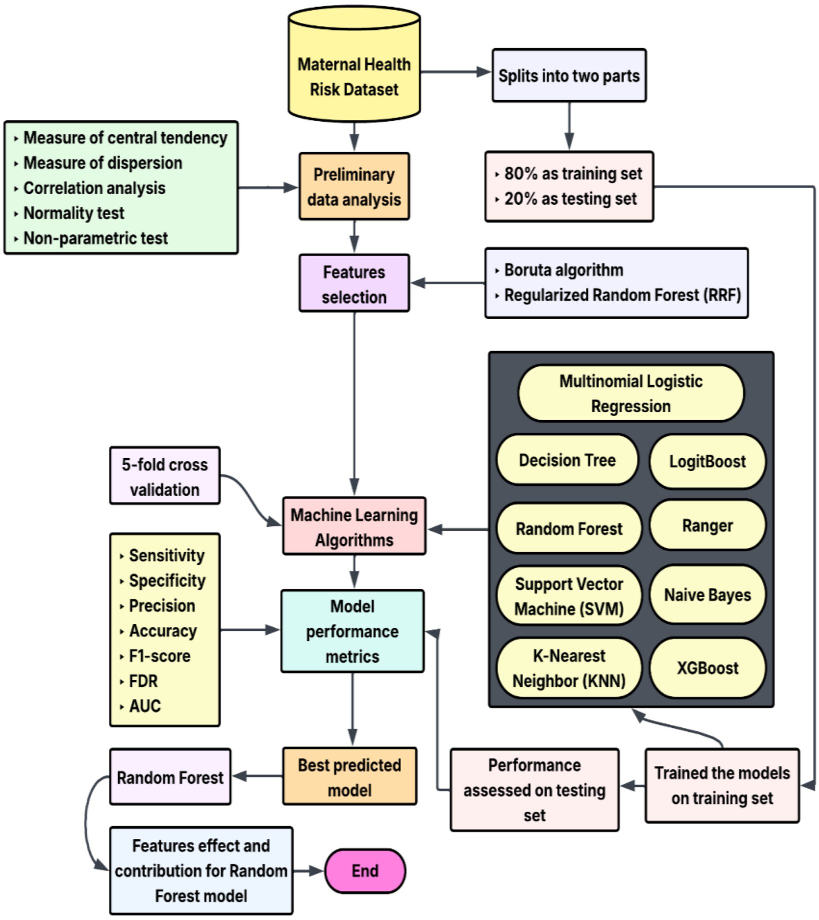

Figure 1: Systematic diagram of the proposed study

2. Materials and Methods

2.1 Data source and study variables

The study's dataset, which included 1014 observations and seven variables, was sourced from the UCI Machine Learning Repository, a publicly accessible dataset repository [15]. An IoT-based risk monitoring system was used to gather data from a variety of hospitals, community clinics, and maternity healthcare institutions located in rural areas of Bangladesh. The quantitative independent variables are Age, Systolic Blood Pressure (Systolic BP), Diastolic Blood Pressure (Diastolic BP), Blood Sugar (BS), Body Temperature (Body Temp), and Heart Rate. The categorical dependent variable is Risk Level (high risk, low risk, mid risk) with 272 patients classified as high-risk, 406 as low-risk, and 336 as mid-risk. Maternal mortality is one of the primary concerns of the UN's SDGs, and all of the independent variables are significant risk factors for this issue.

2.2 Statistical analysis

Various graphical representations (such as density plots and histograms) and measurements of central tendency, dispersion, skewness, and kurtosis are used to examine the properties of the variables. Pearson's correlation coefficient is used to quantify the association between each independent variable, and the R corrplot tool is used to display the correlation matrix as a heatmap. For every continuous variable, non-parametric and normality tests are also conducted.

2.3 Feature selection

In order to increase model accuracy by eliminating features that are not relevant, feature selection is essential in regression and classification. This study applies the following feature selection methods.

2.3.1 Boruta algorithm

The Boruta algorithm, a well-known technique for feature selection, is based on Random Forest[16]. Decision trees are constructed using random copies of the original variables and shadow features. The decrease in model performance resulting from the shadow features is used to evaluate feature significance. Z-scores, which aid in identifying key predictors, are computed by dividing the mean accuracy loss by its standard deviation. In the Boruta approach, the Z-score is considered the most important metric. Consequently, the importance of original attributes is ascertained by using the set of shadow attribute importance as a guide. The importance of original traits and shadow qualities is then contrasted [17]. An in-depth explanation can be found in [18].

2.3.2 Regularized random forest (RRF) method

The RRF method is a sophisticated feature selection technique that improves on conventional random forests by encouraging sparse models through the use of regularization [19]. Assume that $\mathrm{gain}(X_j)$ is the evaluation metric for a particular feature X_j. Define F as the collection of features that were previously applied to splits in a tree model. The following is the definition of the modified measure:

$$\mathrm{gain}_R(X_j) = \begin{cases} \lambda.\mathrm{gain}(X_j) & X_j \notin F \\ \mathrm{gain}(X_j) & X_j \in F \end{cases},$$

The regularization parameter $\lambda \in [0, 1]$ is used in this context. A smaller value of λ means that traits outside of F are penalized more severely. The tree regularization framework is the process of identifying the splitting feature at each tree node using $\mathrm{gain}_R(.)$. More information is available in [19,20].

2.4 Machine learning techniques

Nine ML techniques, discussed below, have been utilized for the prediction of maternal risk levels. To ensure accuracy and reliability, the proper performance criteria are used to train and assess the models.

2.4.1 Multinomial Logistic Regression

Multinomial logistic regression, an extension of binary logistic regression, is employed when the dependent variable is categorical and has more than two nominal categories. In the cases where the response variable has three distinct categories, the model estimates the relationship between the explanatory variables and the probabilities of each category of the response variable by constructing multiple logit (log-odds) functions. Specifically, for a three-category outcome, the model includes two logit equations, comparing each of the non-reference categories to a chosen reference category. This method expresses the dependent variable in terms of the logarithm of the likelihood of an event occurring based on the independent variables' values [21,22].

2.4.2 Decision Tree

The decision tree is a supervised learning technique used for tasks involving regression and classification. Its structure is similar to that of a flowchart, and by methodically dividing the data, it facilitates prediction and decision-making. The internal nodes that make up the tree stand in for feature tests, branches that indicate the results of these tests, and terminal leaf nodes that determine the final prediction in the structure of a number or a class label. The first split is started, and the root node encapsulates the entire dataset. Following that, a decision rule is applied by each internal node, and depending on the result, branches are sent to the nodes that follow. The process continues until it reaches the leaf nodes, at which point it stops dividing and yields the final prediction [23,24].

2.4.3 Random Forest

Random Forest, a supervised machine learning model, can be used for problems that involve both regression and classification. It creates multiple decision trees and aggregates their output to increase forecast accuracy and stability. By using replacement random sampling, bootstrap sampling is used to construct several subsets of the original dataset. At each split, a random subset of attributes is selected to minimize the correlation between trees, rather than utilizing all available features. A predetermined splitting criterion is then used to construct each decision tree [25,26].

2.4.4 Support Vector Machine (SVM)

SVM is a potent supervised machine learning method that is primarily used for classification tasks, while it is also capable of handling regression issues. The fundamental idea behind it is to determine the best hyperplane for categorizing the data with the largest margin, that is, the most separation between the hyperplane and the support vectors, or nearest data points from each class. SVM can handle both linear and nonlinear classification tasks through the use of kernel functions such as polynomial or radial basis function (RBF) kernels, which convert the data into higher dimensions where it is simpler to separate [27,28].

2.4.5 K-Nearest Neighbor (KNN)

K-Nearest Neighbor is a simple, instance-based, supervised, non-parametric machine learning method that may be applied to both regression and classification problems. It analyzes the k nearest neighbors in the feature space to identify the class or value of a particular data point. To determine the closest neighbors, the algorithm uses a distance measure, such as the Euclidean distance, and assumes that similar data points are located nearby. KNN performs effectively in a range of real-world circumstances because it is very adaptable and does not assume anything about the data distribution beforehand. However, the distance metric, the choice of k, and the existence of noisy or irrelevant characteristics can all affect how well it performs [29].

2.4.6 Extreme Gradient Boosting (XGBoost)

XGBoost is a widely used and powerful ML algorithm designed to minimize loss functions for both regression and classification tasks. It sequentially constructs an ensemble of decision trees, with each new tree seeking to fix the mistakes of the ones that came before it. It incorporates several advanced features, including regularization (L_1 and L_2), parallel processing, and sophisticated optimization techniques to increase performance and minimize the risk of overfitting. Additionally, it supports handling of missing values and tree pruning, making it robust and adaptable to diverse datasets. Due to its high predictive accuracy and scalability, it has become a preferred selection in numerous data science contests and real-world applications [30, 31].

2.4.7 Naive Bayes

Naive Bayes, based on the Bayes Theorem, is a straightforward but effective probabilistic machine learning technique that is typically employed for classification problems. To simplify calculation, it makes the "naive" assumption that the features are conditionally independent given the class name, which frequently works amazingly well in practice. The class with the highest probability is chosen by the algorithm, which determines the posterior probability of each class based on the likelihood of the input features. For high-dimensional data like text categorization and spam detection, it is exceptionally effective, scalable to big datasets, and efficient. Despite its simplicity, it frequently produces accuracy that is competitive with more intricate models, particularly when the independence assumption is roughly accurate [32,33].

2.4.8 Ranger

Ranger is a high-performance application of the Random Forest algorithm intended for efficient handling of large and high-dimensional datasets. It creates an ensemble of decision trees utilizing bootstrap aggregation and random feature selection, effectively capturing complex patterns in the data. Ranger is well-regarded for its speed, accuracy, and ability to manage large-scale classification and regression tasks efficiently. Because of its strong performance, it is a dependable option for a range of machine learning applications where predictive ability and computational economy are critical [34,35].

2.4.9 LogitBoost

LogitBoost is a classification algorithm that incrementally builds an additive logistic regression model using a boosting framework applied in a stage-wise manner. It works exceptionally well with the logistic loss function for binary classification. LogitBoost is a model for Additive Logistic Regression. LogitBoost's core idea is to apply boosting when building a logit model. This learning method is referred to as "weak" or "base" since it frequently uses different training cases. Consequently, a new weak prediction rule is produced by the basic learning process, leading to several rounds. Then, using the boosting process, numerous weak prediction rules must be combined to generate a single strong rule, which is typically much more accurate than a weak rule [36,37].

2.5 Model evaluation metrics

The model evaluation metrics, along with their computation formulae, are briefly described in Table 1. In this table, TP = True Positive, FN = False Negative, FP = False Positive, TN = True Negative) [38,39].

Table 1: Model evaluation metrics

Metrics	Formula	Description
Sensitivity	$\frac{TP}{TP + FN}$	Calculates the proportion of real positives that were accurately detected.
Specificity	$\frac{TN}{FP + TN}$	Calculates the proportion of real negatives that were accurately identified.
Precision	$\frac{TP}{TP + FP}$	Calculates the proportion of anticipated positives that are true.
Accuracy	$\frac{TP + TN}{TP + FP + TN + FN}$	Calculates the model's overall accuracy.
F1 score	$\frac{2TP}{2TP + FP + FN}$	Balancing both the harmonic mean of precision and sensitivity.
FDR	$\frac{FP}{TP + FP}$	Calculates the proportion of projected positive results that are false positives.

3. Results

3.1 Preliminary data analysis

Tables 2 and 3 provide the summary statistics for all quantitative variables in the dataset. Here, we observe that the age of the participants ranges from 10 to 70 years, with a mean age of 29.87 years (SD = 13.47). The distribution of age is positively skewed and platykurtic, indicating that a greater proportion of individuals are younger. Systolic Blood Pressure (Systolic BP) varies from 70 to 160 mmHg, with a mean of 113.20 mmHg (SD = 18.40) and a coefficient of variation (CV = 16.26%), suggesting moderate variability. The distribution is slightly negatively skewed and platykurtic. Diastolic Blood Pressure (Diastolic BP) ranges from 49 to 100 mmHg, with a mean of 76.46 mmHg (SD = 13.89), and the skewness value indicates a nearly symmetrical distribution. Blood Sugar (BS) levels range from 6.00 to 19.00 mmol/L, with a mean of 8.73 mmol/L (SD = 3.29), and skewness and kurtosis suggest a positively skewed and platykurtic distribution, indicating the presence of high blood sugar levels in some individuals. Body Temperature (Body Temp) is relatively stable, with a small range (98.00–103.00°F) and a mean of 98.67°F (SD = 1.37) and a low CV (1.39%), suggesting minimal variability. However, the positive skewness indicates that some individuals exhibit elevated body temperatures. Heart Rate varies widely from 7 to 90 bpm, with a mean of 74.30 bpm (SD = 8.09). The negative skewness suggests that most individuals have higher heart rates, while the high kurtosis indicates the presence of extreme values. [$\text{Skewness} = 0, > 0, < 0$, are said to be symmetric, positively skewed, and negatively skewed, respectively] and [$\text{Kurtosis} = 3, > 3, < 3$ are said to be mesokurtic, leptokurtic, and platykurtic, respectively.]

Table 2: Summary statistics of all quantitative variables

Statistics	Age (Years)	Systolic BP (mmHg)	Diastolic BP (mmHg)	BS (mmol/L)	Body Temp(F)	Heart Rate(bpm)
Minimum	10.00	70.00	49.00	6.00	98.00	7.00
Maximum	70.00	160.00	100.00	19.00	103.00	90.00
Range	60.00	90.00	51.00	13.00	5.00	83.00
1st quartile	19.00	100.00	65.00	6.90	98.00	70.00
Median	26.00	120.00	80.00	7.50	98.00	76.00
Mean	29.87	113.20	76.46	8.73	98.67	74.30
3rd quartile	39.00	120.00	90.00	8.00	98.00	80.00
SD	13.47	18.40	13.89	3.29	1.37	8.09
CV	45.11	16.26	18.16	37.74	1.39	10.89
Skewness	0.78	-0.25	-0.05	1.86	1.75	-1.04
Kurtosis	-0.40	-0.62	-0.95	2.28	1.43	8.33

Table 3: Summary statistics of all quantitative variables according to three categories

High risk (n = 272)						
Statistics	Age (Years)	Systolic BP (mmHg)	Diastolic BP (mmHg)	BS (mmol/L)	Body Temp(F)	Heart Rate(bpm)
Minimum	12.00	83.00	60.00	6.10	98.00	60.00
Maximum	65.00	160.00	100.00	19.00	103.00	90.00
Range	53.00	77.00	40.00	12.90	5.00	30.00
1st quartile	25.00	120.00	75.00	7.90	98.00	70.00
Median	35.00	130.00	90.00	11.00	98.00	77.00
Mean	36.22	124.20	85.07	12.12	98.90	76.74
3rd quartile	48.00	140.00	100.00	15.00	100.00	86.00
SD	13.03	20.23	14.11	4.17	1.56	8.69
CV	35.98	16.29	16.59	34.43	1.58	11.33
Skewness	0.01	-0.64	-0.57	0.22	1.30	-0.20
Kurtosis	-0.89	-0.63	-1.04	-1.35	-0.01	-0.89

Mid risk (n = 336)						
Minimum	10.00	70.00	50.00	6.00	98.00	60.00
Maximum	60.00	140.00	100.00	18.00	103.00	88.00
Range	50.00	70.00	50.00	12.00	5.00	28.00
1st quartile	19.00	100.00	65.00	6.80	98.00	70.00
Median	25.00	120.00	75.00	7.00	98.00	76.00
Mean	28.36	113.20	74.23	7.79	98.83	74.18
3rd quartile	32.00	120.00	80.00	7.80	100.00	78.00
SD	12.55	14.98	11.49	2.29	1.43	6.77
CV	44.26	13.24	15.48	29.32	1.45	9.12
Skewness	1.05	-0.78	0.05	2.87	1.27	-0.05
Kurtosis	0.27	-0.59	-0.89	7.99	-0.07	-0.34
Low risk (n = 406)						
Minimum	10.00	70.00	49.00	6.00	98.00	7.00
Maximum	70.00	129.00	100.00	11.00	103.00	88.00
Range	60.00	59.00	51.00	5.00	5.00	81.00
1st quartile	17.00	90.00	60.00	6.90	98.00	70.00
Median	22.00	120.00	75.00	7.50	98.00	70.00
Mean	26.87	105.90	72.53	7.22	98.37	72.77
3rd quartile	32.00	120.00	80.00	7.50	98.00	77.00
SD	13.12	15.89	13.05	0.65	1.11	8.29
CV	48.83	15.01	17.99	8.94	1.13	11.39
Skewness	1.27	-0.52	-0.17	1.54	2.89	-2.31
Kurtosis	0.75	-1.16	-1.05	9.78	6.97	17.81

Figure 2 illustrates that the distributions of Age, BS, and Body Temp are positively skewed, whereas Diastolic BP is nearly symmetric. We also observe that Systolic BP has a slightly negative skew, and Heart Rate exhibits a negatively skewed distribution. Among them, only the distribution of Heart Rate is leptokurtic, while the rest exhibit platykurtic behavior.

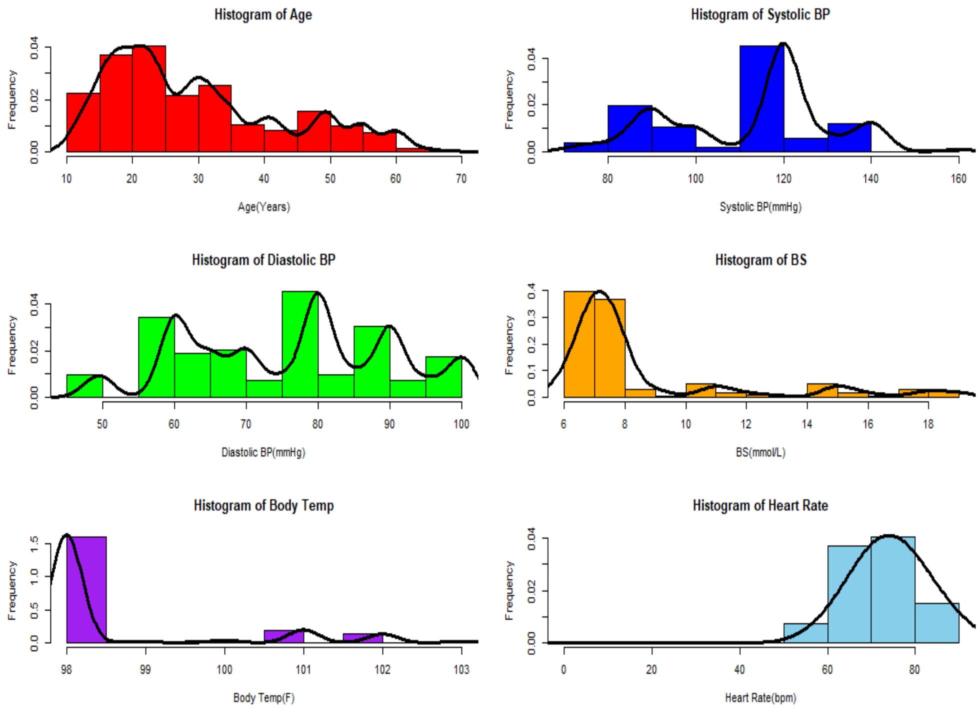

Figure 2: Histogram of each independent variable

The density plots in Figure 3 reveal distinct distributions of physiological variables across maternal health risk levels. High-risk cases are more frequent among younger individuals and those with elevated systolic and diastolic blood pressure, indicating hypertension as a key risk factor. Blood sugar levels are notably higher in high-risk individuals, suggesting a link to gestational diabetes. Heart rate variations are more pronounced in mid and high-risk groups, while body temperature remains relatively stable. These patterns highlight the critical role of hypertension, blood sugar, and age in maternal health risk, supporting the need for early monitoring and intervention strategies.

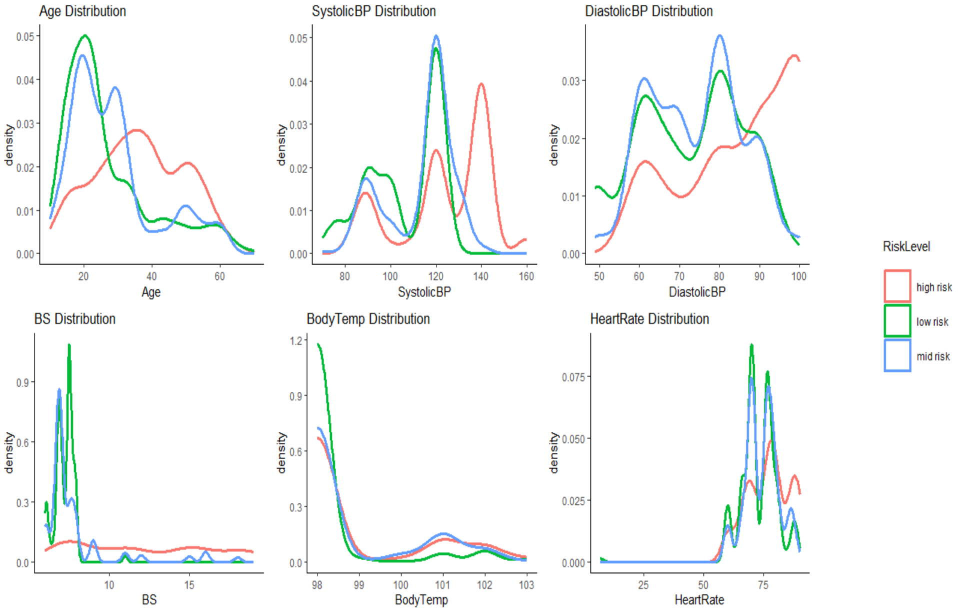

Figure 3: Density plot of each quantitative variable according to three categories of risk level

3.2 Correlation analysis

Figure 4 represents the correlation matrix, which reveals moderate positive correlations between Systolic BP and Diastolic BP (r = 0.79) and Age and BS (r = 0.47), indicating a potential relationship. Age also moderately correlates with Systolic BP (0.42) and Diastolic BP (0.40), suggesting increasing BP with age. However, Body Temp shows weak negative correlations with Age (-0.26), Systolic BP (-0.29), and Diastolic BP (-0.26), indicating minimal dependence. Heart Rate exhibits weak correlations with all variables, implying independence in this dataset.

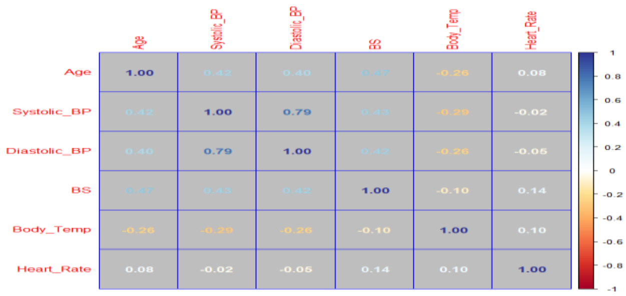

Figure 4: Correlation matrix among independent variables

3.3 Normality test

For every quantitative variable, the results of the normality tests using the Shapiro-Wilk and Kolmogorov-Smirnov (K-S) tests are shown in Table 4. Both tests indicate significant deviations from normality across all variables, as evidenced by D-values and W-values, with corresponding p-values < 0.001. The results suggest that the distributions of Age, Systolic BP, Diastolic BP, Blood Sugar (BS), Body Temperature, and Heart Rate significantly deviate from a normal distribution. According to these findings, non-parametric statistical techniques would be more suited for additional analysis.

Table 4: Normality test for each quantitative variable

Variables	One-sample Kolmogorov-Smirnov (K-S) test		Shapiro-Wilk test	
	D-value	P-value	W-value	P-value
Age	0.144	<0.001	0.916	<0.001
Systolic BP	0.278	<0.001	0.904	<0.001
Diastolic BP	0.145	<0.001	0.947	<0.001
BS	0.349	<0.001	0.674	<0.001
Body Temp	0.479	<0.001	0.528	<0.001
Heart Rate	0.149	<0.001	0.905	<0.001

3.4 Non-parametric test

The Kruskal-Wallis test is conducted in Table 5 to compare differences in continuous (independent) variables across the three Risk Level groups. The results indicate statistically significant differences for all variables (p-value < 0.001), suggesting that Age, Systolic BP, Diastolic BP, Blood Sugar (BS), Body Temperature, and Heart Rate vary significantly among the risk groups. These findings imply that these physiological measures are likely associated with different levels of maternal health risk.

Table 5: The Kruskal-Wallis test for comparing three groups of Risk Level.

Variables	H-statistic	P-value
Age	98.616	<0.001
Systolic BP	166.088	<0.001
Diastolic BP	133.758	<0.001
BS	303.744	<0.001
Body Temp	32.928	<0.001
Heart Rate	37.497	<0.001

3.5 Feature selection

The feature selection analysis in Figure 5 using the Boruta algorithm and Regularized Random Forest (RRF) highlights the most important predictors for maternal health risk classification. The Boruta algorithm confirms Blood Sugar (BS), Systolic Blood Pressure (Systolic BP), Body Temperature, Diastolic Blood Pressure (Diastolic BP), Age, and Heart Rate as significant features, with BS exhibiting the highest importance, followed by Systolic BP. Similarly, the RRF method identifies the same set of features, with BS showing the largest mean decrease in accuracy, indicating its strong predictive power. The consistency across both methods suggests that BS and Systolic BP are the most critical factors influencing maternal health risk, emphasizing the role of metabolic and cardiovascular indicators in risk assessment. Therefore, all six features contribute meaningfully to the prediction of maternal risk level.

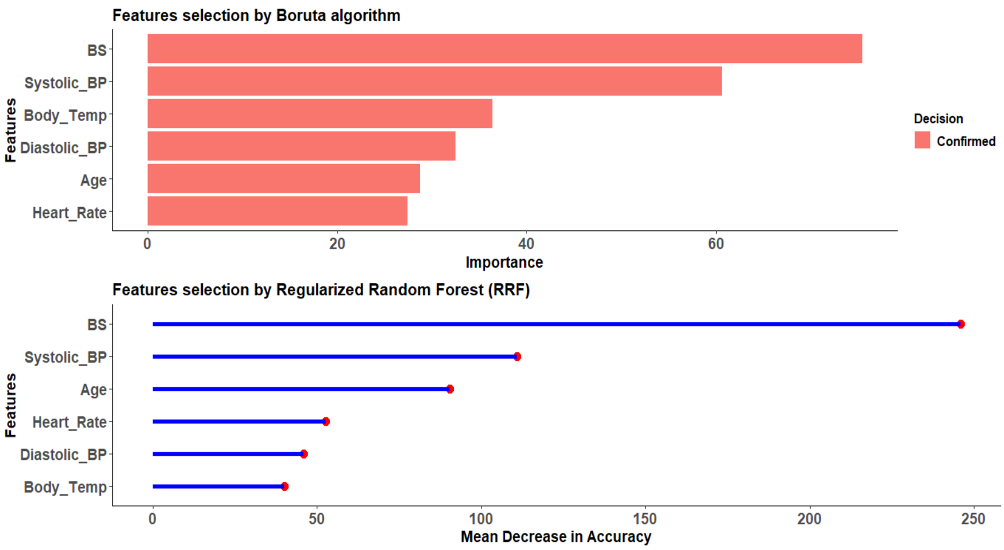

Figure 5: Feature selection by Boruta algorithm and RRF

3.6 Evaluation of model performance

First, 80% of the dataset is used for training, and 20% is used for testing. The training set is used to train the models, and the testing set is used to evaluate their performance. The classification performance of nine machine learning models is evaluated across three risk levels: high, low, and mid, using six key metrics: Sensitivity, Specificity, Precision, Accuracy, F1-score, and AUC, as shown in Table 6. Among all models, Random Forest and Ranger demonstrate the highest overall performance, achieving AUC values close to 1.000 across risk levels, indicating strong discriminative ability. XGBoost and LogitBoost also exhibit competitive results, particularly in high-risk classification, with AUC values exceeding 0.950. Conversely, Multinomial Logistic, KNN, and Naïve Bayes show weaker performance, especially in mid-risk classification, where their sensitivity values are notably lower. For high-risk classification, Random Forest and Ranger outperform other models, with high sensitivity (0.963 and 0.944, respectively) and specificity (0.973).

Decision Tree, SVM, and XGBoost maintain balanced sensitivity and specificity, ensuring robust predictions. In contrast, KNN and Multinomial Logistic struggle with lower sensitivity, leading to a higher probability of missing high-risk cases. In low-risk classification, Naïve Bayes exhibits the highest sensitivity (0.963), making it effective in identifying low-risk cases, but at the cost of lower specificity (0.678). Random Forest and Ranger maintain strong overall performance, while Multinomial Logistic and KNN have the weakest discrimination ability, reflected in their lower AUC values (0.655-0.726). Mid-risk classification poses the most significant challenge across models, but Random Forest, Ranger, and XGBoost perform best, with AUC values exceeding 0.92, indicating strong differentiation of mid-risk cases. Multinomial Logistic, Decision Tree, and Naïve Bayes models exhibited poor sensitivity (<0.45), struggling to classify mid-risk instances accurately. KNN and SVM also exhibit lower mid-risk performance, with sensitivity ranging from 0.43 to 0.54, resulting in higher misclassification rates. Overall, ensemble-based models (Random Forest, Ranger, and XGBoost) emerge as the most reliable classifiers, consistently delivering superior predictive performance across all risk levels.

Table 6: Class-wise performance metrics of nine machine learning models on testing set

Model names	Risk Level	Performance metrics					
		Sensitivity	Specificity	Precision	Accuray	F1 score	AUC
Multinomial Logistic Regression	high risk	0.611	0.959	0.846	0.866	0.709	0.889
	low risk	0.802	0.678	0.625	0.728	0.703	0.655
	mid risk	0.418	0.770	0.475	0.653	0.444	0.811
Decision Tree	high risk	0.870	0.885	0.734	0.881	0.797	0.937
	low risk	0.815	0.793	0.725	0.802	0.767	0.736
	mid risk	0.433	0.867	0.617	0.723	0.509	0.888
Random Forest	high risk	0.963	0.973	0.929	0.970	0.945	0.989
	low risk	0.901	0.926	0.890	0.916	0.896	0.958
	mid risk	0.821	0.933	0.859	0.896	0.839	0.981
SVM	high risk	0.870	0.946	0.855	0.926	0.862	0.955
	low risk	0.864	0.694	0.654	0.762	0.745	0.779
	mid risk	0.433	0.919	0.725	0.757	0.542	0.831
KNN	high risk	0.611	0.966	0.868	0.871	0.717	0.906
	low risk	0.654	0.653	0.558	0.653	0.602	0.726
	mid risk	0.537	0.756	0.522	0.683	0.529	0.783
XGBoost	high risk	0.852	0.932	0.821	0.911	0.836	0.965
	low risk	0.889	0.777	0.727	0.822	0.800	0.871
	mid risk	0.537	0.919	0.766	0.792	0.632	0.922
Naïve Bayes	high risk	0.759	0.946	0.837	0.896	0.796	0.918
	low risk	0.963	0.678	0.667	0.792	0.788	0.793
	mid risk	0.403	0.933	0.750	0.757	0.524	0.871
Ranger	high risk	0.944	0.973	0.927	0.965	0.936	0.985
	low risk	0.864	0.926	0.886	0.901	0.875	0.948
	mid risk	0.821	0.904	0.809	0.876	0.815	0.956
LogitBoost	high risk	0.822	0.949	0.860	0.914	0.841	0.961
	low risk	0.866	0.863	0.817	0.864	0.841	0.810
	mid risk	0.660	0.866	0.688	0.802	0.673	0.904

The ROC curves in Figure 6 demonstrate the classification performance of various ML models across various risk levels. Among them, Random Forest and Ranger exhibit the highest discriminative power, with AUC values exceeding 0.950 across all risk categories, making them the most reliable classifiers. XGBoost and LogitBoost also perform well, particularly in high-risk classification, although they slightly lag in mid- and low-risk cases. SVM and Decision Tree show moderate effectiveness, with Decision Tree struggling in low-risk classification. Naïve Bayes maintains competitive performance but falls short of the top models. In contrast, KNN and Multinomial Logistic perform the weakest, particularly in distinguishing low-risk cases. Overall, ensemble-based models, particularly Random Forest and Ranger, emerge as the most effective, while traditional methods, such as KNN and Multinomial Logistic Regression, exhibit limited predictive power, especially for mid- and low-risk groups.

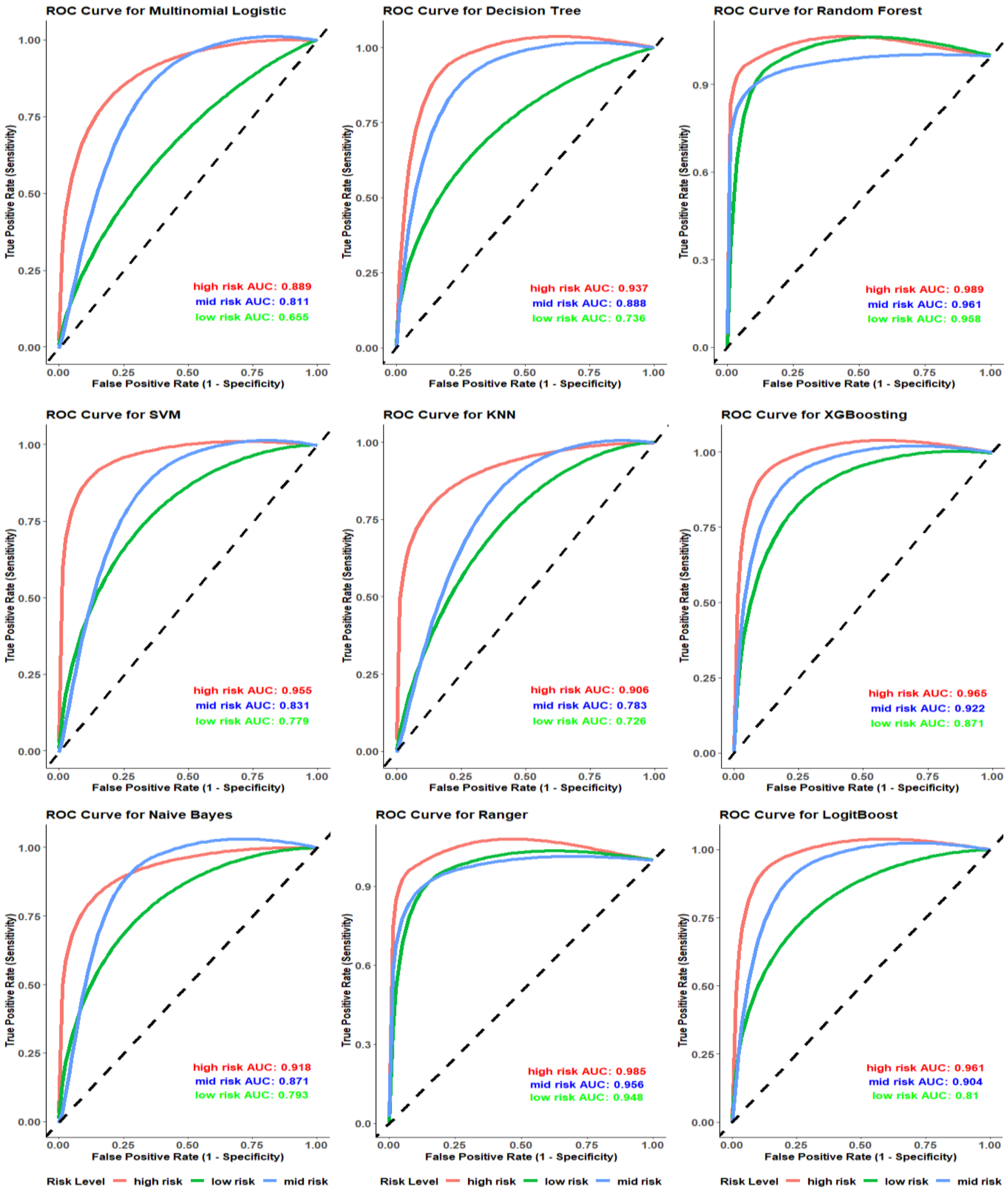

Figure 6: Class-wise ROC curves of nine machine learning models

In both macro-average and weighted-average evaluations in Tables 7 and 8, the Random Forest model consistently outperform other classifiers, achieving the highest sensitivity (0.895 macro, 0.891 weighted), specificity (0.944 macro, 0.941 weighted), precision (0.893 macro, 0.890 weighted), F1 score (0.894 macro, 0.890 weighted), and AUC (0.970 macro, 0.968 weighted), along with a high overall accuracy of 0.891. The Ranger model, a computationally efficient variant of Random Forest, performed strongly, with notable results (e.g., 0.877 macro sensitivity, 0.934 macro specificity, 0.874 macro precision, and 0.963 macro AUC), making it a viable alternative for high-dimensional data. XGBoost and LogitBoost also demonstrate competitive performance with macro AUCs of 0.922 and 0.893, respectively. In contrast, traditional models, such as Multinomial Logistic Regression and K-Nearest Neighbors, show relatively lower metrics, including macro F1 scores of 0.619 and 0.616, and macro AUCs of 0.779 and 0.799, respectively. These findings underscore the superior predictive capability and robustness of ensemble-based methods, particularly Random Forest and Ranger, in handling multiclass classification tasks with potential class imbalance. Here, the macro-average results highlight the balanced performance across all classes, whereas the weighted-average metrics reflect the models' robustness in the presence of class imbalance. Given the presence of imbalanced class distributions in the dataset, weighted-average metrics provide a more reliable assessment of model performance and are therefore emphasized in this study. Overall, ensemble-based models, particularly Random Forest and Ranger, prove to be the most effective classifiers for the dataset.

Table 7: Overall performance of nine machine learning models using macro-average method on testing set

Model names	Performance metrics using macro-average method						
	Sensitivity	Specificity	Precision	Overall accuracy	F1 score	FDR	AUC
Multinomial Logistic Regression	0.610	0.803	0.649	0.624	0.619	0.351	0.779
Decision Tree	0.706	0.848	0.692	0.703	0.691	0.307	0.851
Random Forest	0.895	0.944	0.893	0.891	0.894	0.107	0.970
SVM	0.722	0.853	0.745	0.723	0.716	0.255	0.857
KNN	0.605	0.792	0.649	0.604	0.616	0.351	0.799
XGBoost	0.759	0.876	0.772	0.762	0.756	0.228	0.922
Naïve Bayes	0.708	0.852	0.751	0.723	0.703	0.249	0.865
Ranger	0.877	0.934	0.874	0.871	0.875	0.126	0.963
LogitBoost	0.783	0.893	0.788	0.790	0.785	0.212	0.893

Table 8: Overall performance of nine machine learning models using weighted-average method on testing set

Model names	Performance metrics using weighted-average method						
	Sensitivity	Specificity	Precision	Overall accuracy	F1 score	FDR	AUC
Multinomial Logistic Regression	0.624	0.784	0.634	0.624	0.619	0.376	0.773
Decision Tree	0.703	0.842	0.692	0.703	0.689	0.297	0.849
Random Forest	0.891	0.941	0.890	0.891	0.890	0.109	0.968
SVM	0.722	0.836	0.731	0.723	0.709	0.277	0.848
KNN	0.609	0.771	0.633	0.604	0.613	0.396	0.790
XGBoost	0.762	0.865	0.765	0.762	0.754	0.238	0.919
Naïve Bayes	0.723	0.834	0.740	0.723	0.703	0.277	0.861
Ranger	0.856	0.916	0.860	0.871	0.854	0.144	0.948
LogitBoost	0.790	0.888	0.789	0.790	0.789	0.209	0.889

Figure 7 illustrates the overall comparative performance of different ML models based on key evaluation metrics using a weighted-average method. Random Forest and Ranger achieve the highest accuracy, sensitivity, and specificity, indicating their robustness in classification. LogitBoost and XGBoost also demonstrate strong predictive capabilities with high sensitivity and relatively low FDR. Decision Tree, SVM, and Naïve Bayes maintain moderate performance, balancing sensitivity and specificity. In contrast, Multinomial Logistic and KNN exhibit lower accuracy and F1-scores, accompanied by higher FDR, making them less reliable. These results highlight the superiority of ensemble-based models in predictive analytics for maternal health risk level.

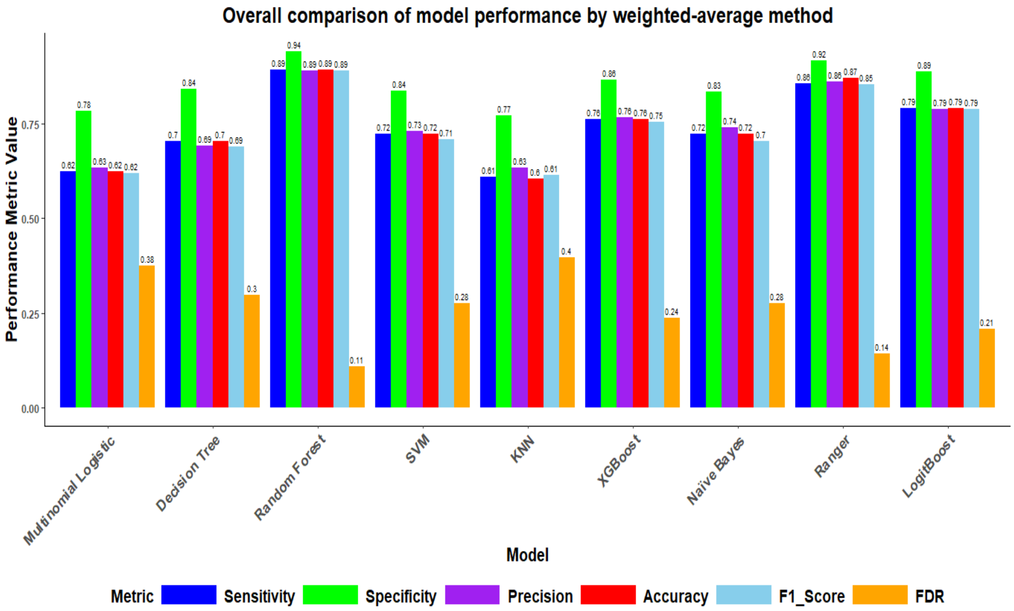

Figure 7: Overall comparison of models' performance by weighted-average method on testing set

3.7 Features effect and contribution to best-performing model

The Random Forest model, the top-performing model for classifying maternal health risks, is shown in Figure 8 along with the feature importance scores that were obtained from it. The graphic illustrates each physiological variable's proportional contribution to the model's predictions. With an importance score that is over 100, blood sugar (BS) stands out as the most significant element, highlighting its leading position in the decision-making process. Systolic Blood Pressure, Age, and Diastolic Blood Pressure also demonstrate meaningful contributions, although with notably lower importance scores. In contrast, Heart Rate and Body Temperature show minimal influence, indicating limited impact on the classification outcome. This feature ranking provides valuable insight into the variables most critical to maternal health risk prediction, with Blood Sugar identified as the primary determinant.

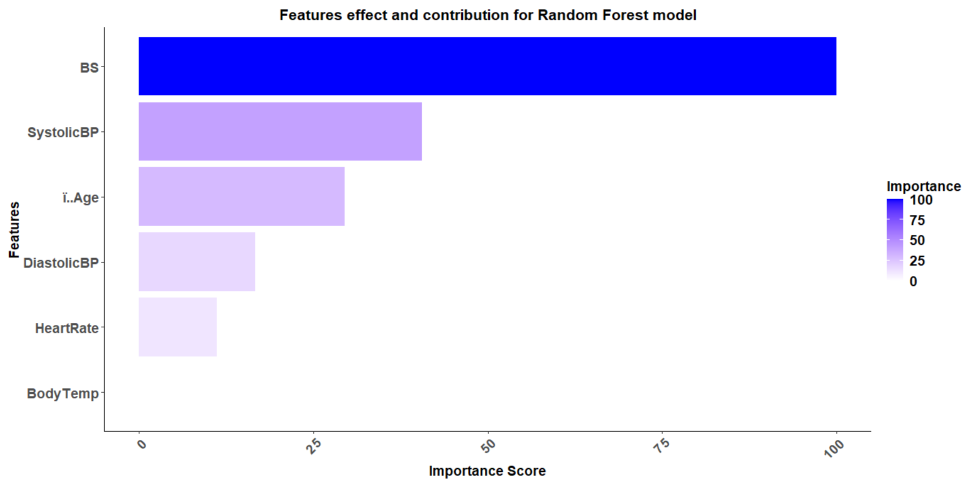

Figure 8: Feature effect and contribution for Random Forest model

4. Discussion

This study presents an IoT-based ML framework to classify maternal health risks using physiological data collected from rural areas in Bangladesh. The dataset, obtained from the UCI Machine Learning Repository, includes 1,014 observations and six independent physiological variables: Age, Systolic BP, Diastolic BP, Blood Sugar (BS), Body Temperature, and Heart Rate. Descriptive analysis reveals meaningful insights into the health profile of participants. The age range is from 10 to 70 years, with a mean of 29.87 years, indicating that most women are in their reproductive years. The positively skewed and platykurtic distribution reflects a younger population segment. Blood Sugar (mean = 8.73 mmol/L) exhibits a positively skewed distribution with high variability, indicating a significant number of women with elevated blood sugar levels, which highlights potential metabolic risks. Blood pressure variables exhibit moderate variability and are moderately correlated with age, consistent with the known physiological changes that occur during pregnancy. Normality tests confirm that none of the independent variables follow a normal distribution (p-value < 0.001), indicating the appropriateness of non-parametric techniques for preliminary analyses. The Kruskal-Wallis test reveals statistically significant differences across all risk levels (p-value < 0.001), underscoring the relevance of each physiological parameter in differentiating maternal health risks. The correlation matrix further shows strong relationships between systolic and diastolic BP (r = 0.79), and moderate correlations between age and blood sugar (r = 0.47), implying intertwined cardiovascular and metabolic effects in maternal health.

Feature selection through the Boruta algorithm and Regularized Random Forest consistently identifies Blood Sugar, Systolic BP, Diastolic BP, Age, Body Temperature, and Heart Rate as important predictors. Blood Sugar emerges as the most significant correlate, aligning with clinical literature that links hyperglycemia to complications during pregnancy. Machine learning models are trained on an 80:20 train-test split and evaluated using 5-fold cross-validation. Ensemble models, particularly Random Forest and Ranger, exhibit the highest classification performance across all risk levels (high, low, and mid). In the overall (macro and weighted average method) performance, again, Random Forest achieves outstanding metrics with weighted-AUC (0.968), overall accuracy (0.891), weighted-F1-score (0.890), and high sensitivity and specificity. These balanced scores reflect the model's robustness across both major and minor classes, minimizing bias toward the majority class. Ranger follows closely, offering a computationally efficient yet accurate alternative. These models exhibit robust discrimination, particularly for high-risk cases, crucial for timely intervention in maternal healthcare. On the contrary, traditional ML models, such as Multinomial Logistic Regression, KNN, and Naïve Bayes, show limited performance, especially for mid-risk classification. Their lower sensitivity values suggest a greater risk of misclassification, which could have severe implications in real-world healthcare applications. The macro average performance confirms consistent model performance across all risk categories, while the weighted average performance accounts for class imbalance, ensuring reliable predictions for underrepresented groups such as high-risk mothers.

The feature importance plot indicates that blood sugar (BS) is the most influential factor associated with maternal health risk. Elevated blood glucose levels during pregnancy may indicate gestational diabetes, a condition strongly linked to serious complications such as preeclampsia, preterm birth, and macrosomia. Moreover, poorly controlled blood sugar can increase susceptibility to infections and delayed wound healing during the postpartum period. Continuous glucose monitoring is therefore vital for mitigating adverse outcomes, particularly in high-risk pregnancies [9, 14]. Systolic blood pressure ranks as the second most significant predictor. It is a key indicator of hypertensive disorders in pregnancy, including gestational hypertension, preeclampsia, and eclampsia, all of which impair placental perfusion and heighten the risk of maternal and perinatal mortality [10,13]. Persistent elevation in systolic blood pressure may reflect cardiovascular stress and should be closely monitored during antenatal care. Maternal age is another important variable that plays a dual role in influencing health outcomes. Both adolescent mothers and those over 35 years of age face increased risks. Younger mothers are more prone to complications such as anemia, obstructed labor, and low birth weight, while older mothers are at greater risk for gestational diabetes, hypertension, and chromosomal abnormalities[11]. Identifying these age-related risks can support more tailored and effective prenatal interventions. Diastolic blood pressure, which reflects vascular resistance, also contributes significantly to maternal risk. Elevated diastolic values often indicate chronic hypertension, potentially leading to complications such as placental abruption, fetal growth restriction, or maternal stroke[10,14]. Its complementary role alongside systolic pressure reinforces the need for comprehensive blood pressure assessment during pregnancy. Heart rate serves as a moderate predictor, as abnormal maternal heart rate can be symptomatic of anemia, infections, or cardiovascular dysfunction, all of which may compromise maternal and fetal well-being. Extreme deviations in heart rate can disrupt oxygen delivery to the fetus and may indicate the need for urgent medical evaluation [13].

Lastly, although body temperature was found to be the least influential feature, it remains clinically relevant. Elevated body temperature is a common symptom of maternal infections such as urinary tract infections, malaria, or chorioamnionitis, which are known to increase the risk of preterm labor, sepsis, and maternal mortality if left untreated [12]. Even minor fluctuations in temperature should not be overlooked, particularly in resource-limited settings. To effectively prevent maternal health risks, both healthcare providers and government bodies must take coordinated action. Healthcare professionals should prioritize routine screening for blood sugar and blood pressure throughout pregnancy, as these are key indicators of conditions like gestational diabetes and hypertensive disorders. Special attention should be given to high-risk age groups, particularly adolescents and women over 35, through tailored antenatal care and close monitoring. Regular assessment of heart rate and body temperature is also essential for detecting underlying conditions, such as infections or cardiovascular stress. On a broader level, the government should strengthen maternal health services, especially in rural and underserved communities, by implementing nationwide screening programs and ensuring access to essential diagnostic tools. Investment in digital health solutions, including IoT-based monitoring systems, can enhance early detection and remote care capabilities.

Additionally, public awareness campaigns and targeted training for healthcare workers are crucial to ensuring the timely identification and management of maternal health risks. Collectively, these efforts can contribute significantly to minimizing maternal morbidity and mortality. Overall, the study highlights the feasibility and reliability of ML models, particularly ensemble-based techniques, in predicting maternal risk using physiological data. It highlights the potential for integrating IoT-enabled health monitoring systems in resource-constrained rural settings. Such models can support early risk detection, optimize resource allocation, and ultimately reduce maternal mortality rates, aligning with the Sustainable Development Goals (SDGs).

Limitations

The dataset included some implausible heart rate values (as low as 7 bpm), which are clinically unlikely since sustained rates below 30 bpm are typically life-threatening. These values were treated as anomalies, likely arising from data entry or measurement errors. This highlights a data quality limitation that has been considered when interpreting the results. Additionally, no hyperparameter tuning was performed, as most models demonstrated satisfactory performance using their default settings.

5. Conclusion and future research

This study demonstrates the effectiveness of ML models for the prediction of maternal health risk levels. Among the models, Random Forest and Ranger achieve the best performance, highlighting the strength of ensemble techniques, while simpler models like Multinomial Logistic Regression and KNN offer interpretability advantages that can aid clinical understanding. Leveraging an IoT-based maternal health dataset from the UCI Machine Learning Repository, this study integrates key physiological risk factors to enhance risk prediction. Feature selection using the Boruta algorithm and RRF method improves model performance and identifies critical determinants of maternal health risk. Beyond methodological contributions, the findings have significant practical implications for real-world deployment. The top-performing models, particularly Random Forest and Ranger, can be effectively integrated into IoT-enabled maternal health monitoring systems to facilitate timely risk assessment and support informed clinical decision-making. Simpler, interpretable models can also help healthcare providers understand individual risk factors, enabling targeted and informed interventions. Such applications have the potential to reduce maternal mortality, improve healthcare outcomes, and advance the United Nations' Sustainable Development Goals (SDGs). Future research could explore larger datasets and integrate additional clinical and demographic factors to further enhance predictive accuracy. It should also explore hybrid machine learning models, deep learning approaches, and the integration of AI-driven decision support systems with clinical expertise, which can help reduce maternal mortality and improve healthcare outcomes.

Acknowledgement: The author is thankful for the reviewer's professional guidance, which added depth to this research.

References

[1] Maternal mortality ratio (modeled estimate, per 100,000 live births) | Data. Accessed: Feb. 16, 2025. [Online]. Available: https://data.worldbank.org/indicator/SH.STA.MMRT?end=2017&start=2000&view=chart

[2] Maternal Mortality - Our World in Data. Accessed: Feb. 16, 2025. [Online]. Available: https://ourworldindata.org/maternal-mortality?source=post_page-----293a2e9374d0-------

[3] R. Mehboob et al., (2021). Maternal Mortality Ratio in Low Income Developing Countries, Glob. Women's Heal., Feb. 2021, doi: 10.5772/INTECHOPEN.95258.

[4] K. Finlayson, N. Crossland, M. Bonet, and S. Downe (2020). What matters to women in the postnatal period: A meta-synthesis of qualitative studies, PLoS One, vol. 15, no. 4, pp. 1–23, 2020, doi: 10.1371/journal.pone.0231415.

[5] J. L. Turban, D. King, J. Kobe, S. L. Reisner, and A. S. Keuroghlian (2022). Access to gender-affirming hormones during adolescence and mental health outcomes among transgender adults, PLoS One, vol. 17, no. 1 January, pp. 1–15, 2022, doi: 10.1371/journal.pone.0261039.

[6] K. R. Darshan and K. R. Anandakumar (2015). A comprehensive review on usage of Internet of Things (IoT) in healthcare system, 2015 Int. Conf. Emerg. Res. Electron. Comput. Sci. Technol. ICERECT 2015, pp. 132–136, Jun. 2016, doi: 10.1109/ERECT.2015.7499001.

[7] N. AlAmri and V. Smith (2022). The effect of formal fetal movement counting on maternal psychological outcomes: A systematic review and meta-analysis, Eur. J. Midwifery, vol. 6, no. March, pp. 1–10, 2022, doi: 10.18332/ejm/145789.

[8] H. K. Bharadwaj et al. (2021). A Review on the Role of Machine Learning in Enabling IoT Based Healthcare Applications, IEEE Access, vol. 9, pp. 38859–38890, 2021, doi: 10.1109/ACCESS.2021.3059858.

[9] M. Ahmed, M. A. Kashem, M. Rahman, and S. Khatun (2020). Review and Analysis of Risk Factor of Maternal Health in Remote Area Using the Internet of Things (IoT), Lect. Notes Electr. Eng., vol. 632, no. March, pp. 357–365, 2020, doi: 10.1007/978-981-15-2317-5_30.

[10] K. E. Setiawan, A. Kurniawan, and S. Y. Prasetyo (2024). Comparative analysis of machine learning decision tree-based models for predicting maternal health risks, Procedia Comput. Sci., vol. 245, no. C, pp. 57–64, 2024, doi: 10.1016/j.procs.2024.10.229.

[11] M. Ahmed and M. A. Kashem (2020). IoT Based Risk Level Prediction Model for Maternal Health Care in the Context of Bangladesh, 2020 2nd Int. Conf. Sustain. Technol. Ind. 4.0, STI 2020, no. February, 2020, doi: 10.1109/STI50764.2020.9350320.

[12] M. A. Kashem, M. Ahmed, and N. Mohammad (2023). Maternal HealthCare Using IoT-Based Integrated Medical Device: Bangladesh Perspective, J. Multidiscip. Acad. Pract. Stud., vol. 1, no. 4, pp. 377–391, 2023, doi: 10.35912/jomaps.v1i4.1793.

[13] H. A. Shifa, M. U. Mojumdar, M. M. Rahman, N. R. Chakraborty, and V. Gupta (2024). Machine Learning Models for Maternal Health Risk Prediction based on Clinical Data, Proc. 18th INDIAcom; 2024 11th Int. Conf. Comput. Sustain. Glob. Dev. INDIACom 2024, pp. 1312–1318, 2024, doi: 10.23919/INDIACOM61295.2024.10498822.

[14] A. Raza, H. U. R. Siddiqui, K. Munir, M. Almutairi, F. Rustam, and I. Ashraf (2022). Ensemble learning-based feature engineering to analyze maternal health during pregnancy

and health risk prediction, PLoS One, vol. 17, no. 11 November, 2022, doi: 10.1371/journal.pone.0276525.

[15] Maternal Health Risk - UCI Machine Learning Repository. Accessed: Feb. 16, 2025. [Online]. Available: https://archive.ics.uci.edu/dataset/863/maternal+health+risk

[16] Z. Jin, J. Shang, Q. Zhu, C. Ling, W. Xie, and B. Qiang (2020). RFRSF: Employee Turnover Prediction Based on Random Forests and Survival Analysis, Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 12343 LNCS, pp. 503–515, 2020, doi: 10.1007/978-3-030-62008-0_35.

[17] H. Ahmadpour, O. Bazrafshan, E. Rafiei-Sardooi, H. Zamani, and T. Panagopoulos (2021). Gully erosion susceptibility assessment in the kondoran watershed using machine learning algorithms and the boruta feature selection, Sustain., vol. 13, no. 18, 2021, doi: 10.3390/su131810110.

[18] M. B. Kursa and W. R. Rudnicki (2010). Feature selection with the boruta package, J. Stat. Softw., vol. 36, no. 11, pp. 1–13, 2010, doi: 10.18637/jss.v036.i11.

[19] E. Izquierdo-Verdiguier and R. Zurita-Milla (2019). An evaluation of Guided Regularized Random Forest for classification and regression tasks in remote sensing, Int. J. Appl. Earth Obs. Geoinf., vol. 88, no. June 2019, 2020, doi: 10.1016/j.jag.2020.102051.

[20] H. Deng and G. Runger (2012). Feature selection via regularized trees, Proc. Int. Jt. Conf. Neural Networks, 2012, doi: 10.1109/IJCNN.2012.6252640.

[21] A. Muslim, A. B. Mutiara, R. Refianti, C. M. Karyati, and G. Setiawan (2020). Comparison of accuracy between long short-term memory-deep learning and multinomial logistic regression-machine learning in sentiment analysis on twitter, Int. J. Adv. Comput. Sci. Appl., vol. 11, no. 2, pp. 747–754, 2020, doi: 10.14569/ijacsa.2020.0110294.

[22] K. Itano, K. Ueki, T. Iizuka, and T. Kuwatani (2020). Geochemical discrimination of monazite source rock based on machine learning techniques and multinomial logistic regression analysis, Geosci., vol. 10, no. 2, 2020, doi: 10.3390/geosciences10020063.

[23] H. H. Patel and P. Prajapati (2018). Study and Analysis of Decision Tree Based Classification Algorithms, Int. J. Comput. Sci. Eng., vol. 6, no. 10, pp. 74–78, 2018, doi: 10.26438/ijcse/v6i10.7478.

[24] B. Charbuty and A. Abdulazeez (2021). Classification Based on Decision Tree Algorithm for Machine Learning, J. Appl. Sci. Technol. Trends, vol. 2, no. 01, pp. 20–28, 2021, doi: 10.38094/jastt20165.

[25] Q. Ren, H. Cheng, and H. Han (2017). Research on machine learning framework based on random forest algorithm, AIP Conf. Proc., vol. 1820, no. February, 2017, doi: 10.1063/1.4977376.

[26] S. J. Rigatti (20170. “Random Forest,” J. Insur. Med., vol. 47, no. 1, pp. 31–39, Jan. 2017, doi: 10.17849/INSM-47-01-31-39.1.

[27] M. Tanveer, T. Rajani, R. Rastogi, Y. H. Shao, and M. A. Ganaie (2024). Comprehensive review on twin support vector machines, Ann. Oper. Res., vol. 339, no. 3, pp. 1223–1268, Aug. 2024, doi: 10.1007/S10479-022-04575-W/TABLES/8.

[28] R. Guido, M. C. Groccia, and D. Conforti (2023). A hyper-parameter tuning approach for cost-sensitive support vector machine classifiers, Soft Comput., vol. 27, no. 18, pp. 12863–12881, 2023, doi: 10.1007/s00500-022-06768-8.

[29] R. Wazirali (2020). An Improved Intrusion Detection System Based on KNN Hyperparameter Tuning and Cross-Validation, Arab. J. Sci. Eng., vol. 45, no. 12, pp. 10859–10873, 2020, doi: 10.1007/s13369-020-04907-7.

[30] R. Sibindi, R. W. Mwangi, and A. G. Waititu (2023). A boosting ensemble learning based

hybrid light gradient boosting machine and extreme gradient boosting model for predicting house prices, Eng. Reports, vol. 5, no. 4, pp. 1–19, 2023, doi: 10.1002/eng2.12599.

[31] C. Bentéjac, A. Csörgő, and G. Martínez-Muñoz (2021). A comparative analysis of gradient boosting algorithms, vol. 54, no. 3. Springer Netherlands, 2021. doi: 10.1007/s10462-020-09896-5.

[32] I. A. K. Aldio Nur Samsi, Yulian Findawati (2022). Analisa Sentimen Mahasiswa Terhadap Minat Berorganisasi Dengan Metode Naïve Bayes, Semin. Nas. Tek. Elektro, Sist. Informasi, dan Tek. Inform., pp. 171–176, 2022.

[33] N. Nurhasanah, D. E. Sumarly, J. Pratama, I. T. K. Heng, and E. Irwansyah (2022). Comparing SVM and Naïve Bayes Classifier for Fake News Detection, Eng. Math. Comput. Sci. J., vol. 4, no. 3, pp. 103–107, 2022, doi: 10.21512/emacsjournal.v4i3.8670.

[34] G. M. Rao, D. Ramesh, and A. Kumar (2020). RRF-BD: Ranger Random Forest Algorithm for Big Data Classification, Adv. Intell. Syst. Comput., vol. 990, pp. 15–25, 2020, doi: 10.1007/978-981-13-8676-3_2.

[35] Y. Shi et al., (2021). Hyperspectral band selection and modeling of soil organic matter content in a forest using the Ranger algorithm, PLoS One, vol. 16, no. 6 June, pp. 1–15, 2021, doi: 10.1371/journal.pone.0253385.

[36] H. Jain, A. Khunteta, and S. Srivastava (2020). Churn Prediction in Telecommunication using Logistic Regression and Logit Boost, Procedia Comput. Sci., vol. 167, no. 2019, pp. 101–112, 2020, doi: 10.1016/j.procs.2020.03.187.

[37] H. R. Pourghasemi, A. Gayen, S. Park, C. W. Lee, and S. Lee (2018). Assessment of landslide-prone areas and their zonation using logistic regression, LogitBoost, and naïvebayes machine-learning algorithms, Sustain., vol. 10, no. 10, 2018, doi: 10.3390/su10103697.

[38] R. Huang et al., (2022). Hossin, M. 1 and Sulaiman, M. N. 2 1, Med. Image Anal., vol. 80, no. 2, p. 102478, 2022.

[39] Confusion matrix - Wikipedia. Accessed: Feb. 17, 2025. [Online]. Available: https://en.wikipedia.org/wiki/Confusion_matrix.