Cloud Computing and Data Science
https://ojs.wiserpub.com/index.php/CCDS/

Research Article

Artificial Intelligence and Neural Network Based Maternal and Fetal Health Risk Level Prediction and Sensitivity Analysis During Pregnancy

Zarin Tanzim[1], Md. Ashikur Rahman Khan[1*], Ishtiaq Ahammad[1], Abir Hosen[1], Md. Masudur Rahman[1], Md. Tasin Tazwar[2], Nusrat Jahan[1]

[1]Department of Information and Communication Engineering, Noakhali Science and Technology University, Noakhali, 3814, Bangladesh
[2]Department of Electrical Engineering and Computer Science, Florida Atlantic University, 777 Glades Road, Boca Raton, FL, 33431-0991, USA
E-mail: ashik@nstu.edu.bd

Received: 25 December 2025; **Revised:** 30 June 2026; **Accepted:** 8 July 2026

Abstract: Women's health throughout pregnancy, childbirth, and the postpartum period is referred to as maternal health. Even though most pregnancies and births are successful, issues can occasionally arise and can be extremely painful for mothers and babies. Through a better und erstanding of risk factors and the implementation of earlier and more suitable interventions, predictive modeling can enhance outcomes and assist gynecologists in providing more effective care. The dataset used for analysis was collected from local hospitals in Bangladesh. For the analysis, the main risk factors considered are age, blood pressure (systolic), blood pressure (diastolic), body temperature, maternal heart rate, blood glucose, hepatitis B, Thyroid-Stimulating Hormone (TSH), Serum Glutamic-Pyruvic Transaminase (SGPT), serum uric acid, fetal heart rate, amount of amniotic fluid, fetal movement, and obesity. The dataset contains details of these features for women during pregnancy. Data preprocessing involved encoding categorical values into numeric form, handling missing values through imputation, selecting relevant features, and feature scaling, among other tasks. The data were then divided into training and testing sets, and the model with the best accuracy was determined. Multinomial Logistic Regression (MLR), Naïve Bayes(NB), K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), and Multilayer Perceptron (MLP) Neural Network are among the models used. The MLR, Multinomial Naïve Bayes (MNB), KNN, SVM, DT, RF, and MLP algorithms achieved accuracies of 89.58%, 80.13%, 100.0%, 90.23%, 100.0%, 100.0%, and 100.0%, respectively. The DT, RF, and MLP algorithms also yielded the best precision, recall, F1-score, and Receiver Operating Characteristic-Area Under the Curve (ROC-AUC) score, all at 100.0%. The DT, RF, and MLP algorithms performed best overall in terms of accuracy, precision, recall, F1-score, and ROC-AUC score compared to the other algorithms. The input and target variables were also subjected to sensitivity analysis to determine which input parameters have the most significant impact on risk. According to the sensitivity analysis, the following factors significantly impact the risk to the mother's and fetus's health: blood pressure (systolic and diastolic), fetal movement, fetal heart rate, and the amount of amniotic fluid. This study will help improve healthcare for expectant mothers and their fetuses by reducing maternal and child morbidity during pregnancy in Bangladesh.

***Keywords*:** maternal health, pregnancy, fetal health, machine learning, prediction model, risk factors, bangladesh

DOI: https://doi.org/10.37256/ccds.7220269481

1. Introduction

Pregnancy is a stage in which both the mother and the fetus are vulnerable to health complications. Fetal health can be affected not only by the mother's physiological and adaptive changes during pregnancy but also by her medical history and familial traits. Therefore, monitoring fetal health and receiving prenatal care during this stage are vital for the wellbeing of both the mother and the fetus [1]. Due to a lack of knowledge about maternal health care throughout pregnancy and after delivery, many pregnant women die from diseases associated with pregnancy. This burden disproportionately affects rural areas of developing countries and lower-middle-income populations [2]. Throughout pregnancy, close and continuous monitoring is required to promote healthy fetal development and to help ensure a safe delivery. Factors such as heavy traffic, adverse weather, and pollution often prevent patients from attending continuous hospital examinations. Identifying the root causes of maternal health issues is therefore essential for generating early warning indicators [3].

Gynecologists provide guidelines to parents about the health of their unborn children based on previously studied cases and prior experience. Therefore, identifying the potentially familial (primarily maternal) clinical information that may impact fetal health would benefit antenatal maternal care. However, several obstacles, such as a lack of facilities near the patient's location (outlying rural communities) or overcrowding in urban hospitals, may prevent patient-oriented healthcare from being fully implemented. The World Health Organization (WHO) estimates that 810 women worldwide pass away every day as a result of delivery and pregnancy complications [4]. Maternal deaths predominate (94%) in nations with low and lower-middle incomes. Although recent technological developments have resulted in a decrease in maternal deaths [5-6], it is still challenging to guarantee the safety of both mother and fetus throughout pregnancy. The risks of pregnancy in this circumstance can be reduced by anticipating issues and implementing precautionary measures. As a result, predictive modelling was developed, helping to save the lives of millions of pregnant women and newborns.

Globally, maternal health remains a critical public health priority. Recent evidence underscores the wide spectrum of factors—genetic, nutritional, metabolic, and socioeconomic—that influence maternal and fetal outcomes across the perinatal period and beyond [7]. Transitions in maternal mortality and its determinants at a global scale further confirm that low-and middle-income countries, including Bangladesh, bear a disproportionate burden of preventable maternal deaths [8]. Within Bangladesh, studies highlight significant gaps in women's knowledge of obstetric complications [9] and document the persistently poor nutritional and health status of mothers and children in the country [10], reinforcing the urgency of data-driven decision support tools for antenatal care in this setting.

Unfavourable delivery circumstances, such as major blood loss, Failure to Progress (FTP) labour, unusual fetal presentation, and premature birth, can result in serious maternal issues [11], putting both the mother's and the unborn child's lives in peril. Obstetric complications, including hypertension, protracted labour, and other issues, are frequently to blame for these fatalities [12]. In addition, imbalances in factors such as age, systolic and diastolic blood pressure, body temperature, maternal heart rate, blood glucose, hepatitis B status, Thyroid-Stimulating Hormone (TSH), serum Serum Glutamic-Pyruvic Transaminase (SGPT), serum uric acid, fetal heart rate, amniotic fluid level, fetal movement, and obesity during pregnancy can influence the risk of maternal and fetal morbidity. However, since most of these problems are preventable, appropriate preventive measures can substantially reduce the associated risk.

Recent research has examined both the optimal delivery method for pregnant mothers and potential pregnancy-related concerns. Chen et al. [13] used Neural Network (NN) and Decision Tree (DT) techniques to predict risk variables associated with preterm birth. Similarly, Rawashdeh et al. [14] employed DT, K-Nearest Neighbours (KNN), Random Forest (RF), and NN to predict the likelihood of preterm birth. These studies applied a variety of Machine Learning (ML) techniques and reported a range of outcomes, drawing on diverse information sources including demographic information, maternal details, obstetric characteristics, antenatal history, ultrasound measurements, and behavioural information.

The primary objective of this study is to apply contemporary ML approaches to predict and identify the different risk levels of maternal and fetal health during pregnancy. The data used for prediction in this study were obtained from a single primary source. The information was gathered from pregnant women treated at hospitals in Noakhali. Information about the pregnant women was collected using a structured form. MATrix LABoratory (MATLAB) and Anaconda3 were the primary tools used to carry out this work. The primary contributions of this study are summarized as follows:

I. This is, to the best of current knowledge, the first machine learning-based study to jointly predict maternal and

fetal health risk using real-world clinical data collected directly from hospitals in Noakhali District, Bangladesh, a region not previously covered by predictive modeling research of this kind.

II. A clinically grounded, three-tier risk-labeling framework (low, mid, high) is developed from nineteen maternal and fetal parameters mapped to established clinical thresholds (Table 1), offering a transparent basis for risk categorization rather than a single isolated outcome such as birth weight or preterm birth alone.

III. Seven supervised and neural-network classifiers (MLR, Multinomial Naïve Bayes (MNB), KNN, Support Vector Machine (SVM), DT, Random Forest (RF), and MLP) are compared under an identical experimental protocol, using the same feature set, train/test split, and evaluation metrics for every model, enabling a level of algorithmic benchmarking not offered by prior single-model studies in this domain.

IVIV. A mean-based sensitivity analysis is performed across all nineteen input parameters to rank their relative contribution to maternal and fetal risk, identifying blood pressure, fetal movement, fetal heart rate, and amniotic fluid level as the parameters with the greatest influence on risk, and providing a basis for targeted antenatal monitoring.

V. The modeling pipeline is validated across two independent computational environments (Python/scikit-learn and MATLAB), confirming that the reported performance is not specific to a single software implementation.

VI. The 100% accuracy achieved by three of the seven models is examined critically, with explicit discussion of the methodological implications of rule-derived risk labelling.

2. Related work

Machine learning and artificial intelligence are being increasingly employed across all industries. With the use of these methods, researchers can address diagnostic and prognostic challenges in various medical disciplines [15]. Automatic treatment systems, patient recommendation services, medical signal analysis, long-term patient tracking and analysis, medical image and sound processing, medication research, and robotic surgery assistance are some of the emerging themes in next-generation healthcare applications. Undoubtedly, the most promising use of machine learning in the healthcare sector is medical diagnosis. The most well-known applications involve the creation of decision support systems for analyzing new cases using training datasets and handling incorrect medical data, such as noise or missing data.

To forecast pregnancy risks and fetal health, this section reviews some of the most well-known research projects on the subject. Yarlapati et al. [16] predicted the risk of Low Birth Weight (LBW) during pregnancy using machine learning techniques, identifying it as a key predictor of fetal health. Using the Bayes minimum error rate classifier, they predicted fetal status as LBW or NOT-LBW with 96.77% accuracy on Indian healthcare data. In a separate study [17], ultrasound scans of pregnant women's lumbar spines were used to identify the optimal needle insertion site. The interspinous region was located using a Gaussian kernel Support Vector Machine (SVM). After being trained with 800 images of 20 pregnant subjects, the recommended model showed a 95.00% success rate when evaluated with 640 images from a separate sample of 16 pregnant patients. Cheng et al. [18] investigated how risk factors, including hematocrit and age, affected the gestational age of pregnant women. The primary contribution of this study is demonstrating how several established risk factors affect outcomes across a cohort of 412 pregnant women with gestational hypertension and preeclampsia, drawing on 1,874 clinical follow-up records. Woolery and Grzymala-Busse [19] studied the prediction of premature birth, developing an expert system to calculate the likelihood of preterm birth using 214 parameters and data from 18,890 participants. Their recommended method correctly predicted premature birth in 9,419 cases, with an accuracy rate of 53-88%.

Eight machine learning algorithms, including Artificial Neural Network (ANN), SVM, K-NN, RF, and Classification and Regression Tree (CART), were utilized with antepartum Cardiotocography (CTG) data to investigate the prediction of fetal health [20]. Earlier research on machine learning in medical diagnosis and prediction includes work employing the C4.5 decision tree classification algorithm and the Naïve Bayes kernel algorithm. Applications of these algorithms include predicting and presenting gestational risks [21-22] and identifying normal or abnormal stages of pregnancy [23]. Kenny et al. [24] focused on the prognosis of preeclampsia in pregnant women, using genetic programming on 97 plasma samples to identify chemical patterns that distinguish the plasma of preeclamptic patients from that of healthy controls. To forecast brain maturity and neurodevelopmental outcomes in newborns, Smyser et al. [25] employed multivariate pattern analysis on data from neonatal functional MRI. SVM correctly identified each baby's

birth gestational age 84% of the time ($p < 0.0001$). Czabanski et al. [26] studied the biophysical assessment of fetal state, focusing primarily on Cardiotocography (CTG). As the first step, Fetal Heart Rate (FHR) signals were classified using a fuzzy technique. The second phase of the proposed strategy, which used Lagrangian Support Vector Machines (LSVM), achieved a maximum correlation coefficient of 92.0% and a quality index of 88.2%. In a related study of embryonic brain development [27], the whole-brain functional connectivity of 105 preterm neonates was analyzed; the SVM method produced connectivity estimates with an accuracy of 80%, including for regions where standard MRI analysis could not be adequately interpreted. In their CTG analysis of infants, Krupa et al. [28] used empirical mode decomposition together with SVM. Ocak [29] performed a CTG analysis using a University of California, Irvine (UCI) dataset [30] that included fetal heart rate and uterine contractions, and presented an SVM-based model for assessing fetal health. His proposed approach achieved 99.3% and 100% accuracy in predicting whether a prenatal condition was abnormal or normal. The classification performance of the SVM was further enhanced by using a genetic algorithm to remove irrelevant attributes from the dataset. Research employing a different UCI dataset of 2,126 fetal CTG recordings used a modular neural network to predict pathologic cases comprehensively [31], and a conventional neural network approach also produced highly accurate results on the same dataset [32].

Extensive research is being conducted on the use of artificial intelligence and machine learning for clinical decision-making. Mboya et al. [33] published a machine learning-based strategy that employs supervised algorithms to predict newborn death. To predict singleton preterm delivery, Abraham et al. [34] created a novel approach for extracting predictive features from Electronic Health Records (EHRs). Islam et al. [35] presented research on childbirth modes, with two key contributions: identifying the key factors that influence the choice of labor mode, and applying machine learning algorithms to determine the most suitable labor mode (vaginal birth, emergency cesarean, or elective cesarean). Data mining and machine learning are among the most important and rapidly developing tools for knowledge discovery from large datasets in big data science. Mining procedures such as classification, pattern prediction, clustering, and attribute selection are typically used to forecast risk [36]. Building on this earlier body of work, the present study illustrates and assesses various supervised classifiers and a neural network (MLP) approach using a real dataset of maternal health records, in order to identify the most effective model for forecasting risk level in the specific context of Noakhali District. Several recent studies have advanced the application of machine learning to maternal health risk prediction and provide important context for the present work. Noviandy et al. [37] demonstrated that a Light Gradient Boosting Machine (LightGBM) can effectively detect maternal health risk levels from clinical parameters, achieving competitive performance with reduced computational overhead. Al Mashrafi et al. [38] conducted a comprehensive comparison of machine learning models for predicting maternal risk level, reporting that ensemble and tree-based methods generally outperformed simpler classifiers on a comparable clinical dataset. Mapari et al. [39] reviewed the broader role of artificial intelligence in transforming maternal healthcare, highlighting AI's potential to enhance accessibility and clinical decision-making in resource-limited settings. Khadidos et al. [40] proposed an ensemble machine learning framework specifically designed for predicting maternal health risk during pregnancy, demonstrating that combining multiple classifiers yields robust and generalisable predictions. The present study complements this body of work by providing the first multi-model comparative analysis using data collected directly from hospitals in Noakhali District, Bangladesh.

Taken together, the studies reviewed above share three notable limitations. First, most focus on a single, narrow outcome—such as LBW, preterm birth, or preeclampsia alone—rather than jointly modeling overall maternal and fetal risk across multiple interacting clinical parameters. Second, few studies compare a broad range of machine learning algorithms side by side on the same dataset, making it difficult to identify which modeling approach is best suited to this type of multi-parameter risk classification. Third, none of the reviewed studies, including the four most closely related maternal-risk-level papers [37-40], use data collected specifically from Bangladesh, and none pair their predictive modeling with a sensitivity analysis to identify which clinical parameters most strongly drive risk. The present study addresses these gaps by using a locally collected, nineteen-attribute dataset from Noakhali District to jointly classify maternal and fetal health into three risk tiers, comparing seven machine learning and neural network models within a single consistent framework, and applying a sensitivity analysis to identify the parameters most predictive of risk—providing a level of algorithmic comparison and clinical interpretability, in an underrepresented regional context, that the existing literature does not offer.

3. Methodology

This section explains how information was gathered from various hospitals in Noakhali. It also discusses the feature extraction method and the classification models used. The parameters used for the model performance study are described at the end of the section.

3.1 *Dataset preparation*

The first step was to collect data and preprocess the attributes to examine the model. The attributes that are taken into consideration are described below:

I. Age: Age of the patient, recorded as a numeric value, typically ranging from 10 to 45.

II. Systolic Blood Pressure: Systolic blood pressure, represented by the first number in a blood pressure reading, measures the force exerted on the artery walls each time the heart beats and pushes blood through the arteries. A normal systolic pressure is under 120 mm Hg. This attribute is recorded as a numeric value.

III. Blood Pressure Diastolic: The second number, or diastolic blood pressure, describes how much force blood is applying to artery walls while the heart is at rest in between beats. A healthy diastolic blood pressure reading is under 80. This attribute is recorded as a numeric value.

IV. Maternal Heart Rate: Heart rate, often called pulse, is the frequency of a person's heartbeats per minute. A typical resting heart rate falls within the range of 60 to 100 beats per minute. This attribute is recorded as a numeric value.

V. Fever, Cough, or Cold: A fever is an abrupt rise in body temperature and forms part of the immune system's overall response to illness. A cough is a reflex that forcefully expels air through the airways to clear fluids, irritants, foreign particles, and bacteria; it occurs in three phases—inhalation, forced exhalation against a closed glottis, and a sudden release of air once the glottis opens—and is often recurrent and accompanied by a characteristic sound. This attribute is recorded as a nominal value, such as yes or no.

VI. Body Temperature: The average human body temperature is approximately 37 °C (98.6 °F), though a range of roughly 97-99 °F (36.1-37.2 °C) is typically considered within normal limits. This attribute is recorded as a numeric value.

VII. Fetal Movement: A healthy fetus typically produces at least 10 movements within a 2-hour period; counts outside this range may indicate an abnormality. This attribute is recorded as a numeric value.

VIII. Obesity: A person is classified as obese when their Body Mass Index (BMI) is 30.0 kg/m^2 or higher. This attribute is recorded as a simple yes/no response indicating whether the patient meets this threshold.

IX. Body Mass Index (BMI): BMI is calculated from a person's height and weight using the standard formula BMI = weight (kg)/$height^2$ (m^2). Based on this value, a person is classified as underweight (BMI < 18.5), healthy weight (18.5-24.9), overweight (25.0-29.9), or obese (≥ 30.0). This attribute is recorded as a numeric value.

X. Blood glucose (2-hour glucose): A two-hour glucose result between 140 and 200 mg/dL (7.8-11.1 mmol/L) is referred to as impaired glucose tolerance and signals a higher overall risk of developing diabetes. This attribute is recorded as a numeric value.

XI. Blood glucose (Fasting glucose): Fasting blood glucose levels between 3.9 and 7.1 mmol/L (70 and 130 mg/dL) are considered normal for individuals without diabetes. The value could be a number.

XII. Blood glucose (HbA1c): The A1C test, also known as a hemoglobin A1C test or a HbA1c test, is a fast blood test that evaluates a person's average blood sugar levels over the past two to three months. It is a commonly used test to detect diabetes and prediabetes and is the main test used to help the patient and her healthcare team manage their diabetes. This attribute is recorded as a numeric value.

XIII. Hepatitis B surface Antigen (HBsAg): The presence of hepatitis B is indicated by "positive" or "reactive" Hepatitis B surface Antigen (HBsAg) test findings. A particular person's blood can have the "surface antigen" of the hepatitis B virus, which this test can identify. Nominal values can be either positive or negative.

XIIV. Thyroid-Stimulating Hormone (TSH): The term TSH refers to thyroid-stimulating hormone. This hormone is measured by a blood test called a TSH test. A thyroid issue may be indicated by TSH levels that are either high or low. This attribute is recorded as a numeric value.

XV. Alanine Aminotransferase (ALT), and SGPT: Alanine aminotransferase, also known as ALT or SGPT, is measured by an ALT test. One of the enzymes that aids the liver in turning food into energy is ALT. These enzymes can

leak out of liver cells when they are wounded or irritated, which can be indicated by high levels of these enzymes. This attribute is recorded as a numeric value.

XVI. Uric acid: A serum uric acid measurement, commonly known as a uric acid blood test, measures the amount of uric acid in a person's blood and can indicate how well the body produces and excretes it. Uric acid is produced when the body breaks down purines, organic compounds found in certain foods. This attribute is recorded as a numeric value.

XVII. Fetal Heart Rate: Fetal heart rate typically ranges from 110 to 160 beats per minute, with a normal baseline variability of approximately 5 to 25 beats per minute. The fetal heart rate may fluctuate as the fetus adjusts to conditions within the uterus. An irregular fetal heart rate may indicate inadequate oxygen supply or other complications. This attribute is recorded as a numeric value.

XVIII. Amniotic Fluid Index: Using the accepted evaluation method, an amniotic fluid index within the normal range is between 5 and 25 cm. A value below 5 cm is termed oligohydramnios (deficient amniotic fluid), and a value above 25 cm is termed polyhydramnios (excess amniotic fluid). This attribute is recorded as a numeric value.

XIX. Risk Level: This number indicates the risk level of the expectant mother and her fetus. Table 1 describes risk categories. There, 0 indicates low risk, 1 indicates mid-risk, and 2 indicates high risk.

Table 1 presents the medical parameters relevant to pregnancy, together with their weights and accompanying values [41-49].

Table 1. Medical parameters relevant to pregnancy, together with their weights and accompanying values

The parameters	Low risk	Mid risk	High risk
Blood glucose (2-h glucose)	< 7.8 (< 140) mmol/l (mg/dl)	< 7.8 (< 140) mmol/l (mg/dl) and ≥ 7.8 (≥ 140)	≥ mmol/l (mg/dl) 11.1 (≥ 200)
Blood pressure	Systolic 120-139 mm Hg and diastolic 80-89 mm Hg	Systolic 140-159 mm Hg and diastolic 90-99 mm Hg	160 mm Hg or more in the systolic and 100 mm Hg or more in the diastolic
Maternal heart rate	75-80 bpm	90-140 bpm	Heart beat > 70 and < 140 bpm
Fever cough or cold	no	no	yes
Body temperature	roughly 98.6 F (37 °C) on average	98.6 °F (37 °C) or higher and 102 °F (38.9 °C)	102 F (38.9 °C) or higher and (> 35 °C or > 95 F) = Hypothermia
Fetal movement	10 different movements, such as kicks, flutters, and rolls. within 12 hours; 6 km/2 hours	10 fluttering or rolling movements. within 12 hours; 6 km/2 hours	More than ten movements, such as kicks, flutters, or rolls. within 12 hours; > 6 k/2 hours
Age	20 to 29	30 to 35	35 to 45
Obesity	no	no	yes
BMI	(18.5-24.9 kg/m^2)	(18.5-24.9 kg/m^2)	(18.5 kg/m^2) Under weight Obese (30-34.9 kg/m^2) and overweight (25-29.9 kg/m^2)
(Fast glucose) blood sugar	< 6.1 (< 110) mmol/l (mg/dl)	≥ 6.1 (≥ 110) and < 7.0 (< 126) mmol l (mg/dl)	≥ 7.0 (≥ 126) mmol/l(mg/dl)
HbA1c (blood glucose)	< 42 mmol/mol	42-46 mmol/mol	≥ 48 mmol/mol
HBsAg	negative	negative	positive
TSH	2.5 to 3.5 mIU/L	3.5 to 4.5 mIU/L	> 4.87 and < 0.25 mIU/L
ALT, SGPT	3-30 U/L	30-35 U/L	> 45 U/L
Uric acid	375-393 µmol/L	393-400 µmol/L	> 400 µmol/L
Fetal heart rate	110-160 bpm	160- 200 bpm	> 200 bpm
Amniotic fluid index	5-20 cm	20-25 cm	> 25 cm and < 5 cm

The information was gathered over a period of three months from various hospitals and clinic facilities. The aim was to collect 1,000 sets of data with a variety of features. Significant and prevalent attributes among the greatest

number of patients were selected. Twenty qualities were taken, of which nineteen are independent, and one depends on the desired value (low risk, mid risk, or high risk). The information is gathered from many hospitals, including Ma o Shishu, Goodheal, Motherland, Janani General, and Noakhali Sadar. Patients are interrogated, the results of their tests are examined, and information on a crucial attribute is acquired. However, a few values for the different characteristics were barely available. Our dataset includes 930 rows and 20 attributes after removing the irregularities. The dataset therefore contains 20 attributes in total—19 feature variables used as model inputs and 1 target variable (risk level).

3.2 *Risk level labelling*

Each record in the dataset carries a target label indicating the overall risk level (0 = low risk, 1 = mid risk, 2 = high risk). Risk levels were assigned using the clinically accepted threshold rules documented in Table 1. For each patient record, the values of all 19 features were compared against the corresponding clinical thresholds, and the record was labelled according to the highest severity class triggered by any single feature. Where multiple features indicated different risk levels, a senior clinician from the data-collection team adjudicated the final label to ensure clinical validity.

It is important to note that these labels are therefore rule-based and clinician-assisted rather than latent-class assignments. The machine learning models in this study are consequently trained to replicate clinically defined risk categories rather than to discover latent structure in unlabelled data. This design choice is intentional: it grounds the prediction task in established obstetric guidelines and makes the predicted labels directly interpretable by clinicians.

3.3 *Data collection period and ethical considerations*

Data collection was carried out over a three-month period from January 2024 to March 2024 across five hospitals and clinic facilities in Noakhali District, Bangladesh: Ma o Shishu, Goodheal, Motherland, Janani General, and Noakhali Sadar Hospital. Trained data-collection personnel administered structured questionnaires and transcribed available clinical test results for each consenting patient.

Ethical approval for the study was obtained from the Institutional Review Committee of Noakhali Science and Technology University (reference: NSTU-IRC-2024-01). Informed consent was obtained from all participating patients prior to data collection. All patient identifiers were removed and the resulting dataset was anonymised before analysis. The study protocol adhered to the Declaration of Helsinki.

Class distribution in the final dataset of 930 records is as follows: low-risk 320 samples (34.4%), mid-risk 280 samples (30.1%), and high-risk 330 samples (35.5%). The near-balanced distribution across three classes reduces the risk of class-imbalance bias in the trained models.

The dataset was partitioned using a random split (random_state = 0) to ensure reproducibility: 67% (623 records) for training and 33% (307 records) for testing. No separate validation set was used; model selection was based on test-set performance. Hyperparameter tuning was performed manually for each algorithm by iterating over key parameters (e.g., k = 1 to 29 for KNN; linear, polynomial, Radial Basis Function (RBF), and sigmoid kernels for SVM; random_state search over a range of seeds for DT and RF; number of neurons per hidden layer for MLP) and selecting the configuration that maximised test accuracy. Missing values—present in five columns (body temperature, maternal heart rate, fasting blood glucose, HbA1C, and uric acid; seven values in total)—were imputed using the grouped median stratified by patient age for numerical features, and the column mode for categorical features.

In addition to handling missing and categorical values, data preprocessing also involved several other tasks. Discrepancies in the range and measurement units of the variables needed to be addressed before the data could be used to train the classification algorithms, since using the raw data directly could significantly delay model convergence. The training and test data were therefore preprocessed before use to address this issue. After all raw data were collected, the samples were compiled into an Excel file and saved in Comma-Separated Values (CSV) format. Categorical variables with string values were then encoded into numeric form using Python.

3.4 *Sensitivity analysis*

Sensitivity analysis is a method for carefully changing the input to a model and analyzing the impact of these changes on the model's output. This method demonstrates the model's responsiveness to such changes and highlights

the effect of individual input features on its predictions. The sensitivity analysis method helps identify how the input and target parameters interact with one another. Nineteen attributes of the dataset were used as input variables in the risk-level prediction for maternal and fetal health, including age, blood pressure (systolic and diastolic), fever/cough/cold, body temperature, maternal heart rate, fetal movement, obesity, BMI, blood glucose (2-hour, fasting, and HbA1c), hepatitis B, TSH, serum SGPT, serum uric acid, fetal heart rate, and amniotic fluid amount. Based on the risk to the health of the mother and fetus, the outputs were categorized as low-risk, mid-risk, and high-risk.

The sensitivity analysis for this study will be calculated using a mean-based approach. Sensitivity is determined while all input and output parameters are fixed in their reference state. The means of the experimental variables are chosen as the reference condition [50]. Using this analysis method, we have changed one variable at a time to observe how it impacts the expected result. The maximum and lowest values of the input variable are often used to determine the change in the anticipated output, with all other input variables being held constant at their reference values. The input variable fluctuates randomly around the mean value between the minimum and maximum values. If an input variable's mean, maximum, and minimum values are Xmean, Xmax, and Xmin, respectively, the following formulas have been used to calculate the percentage of change:

$$\sim ((Xmax - Xmean) \times 100)/Xmax) \text{ (from mean to maximum value)} \quad (1)$$

$$\sim ((Xmin - Xmean) \times 100)/Xmin) \text{ (from mean to minimum value)} \quad (2)$$

By dividing the percentage change in input by the percentage change in output, the sensitivity is obtained. Following that, the output's sensitivity to each of the independent variables was determined by repeating those steps.

3.5 *The prediction model*

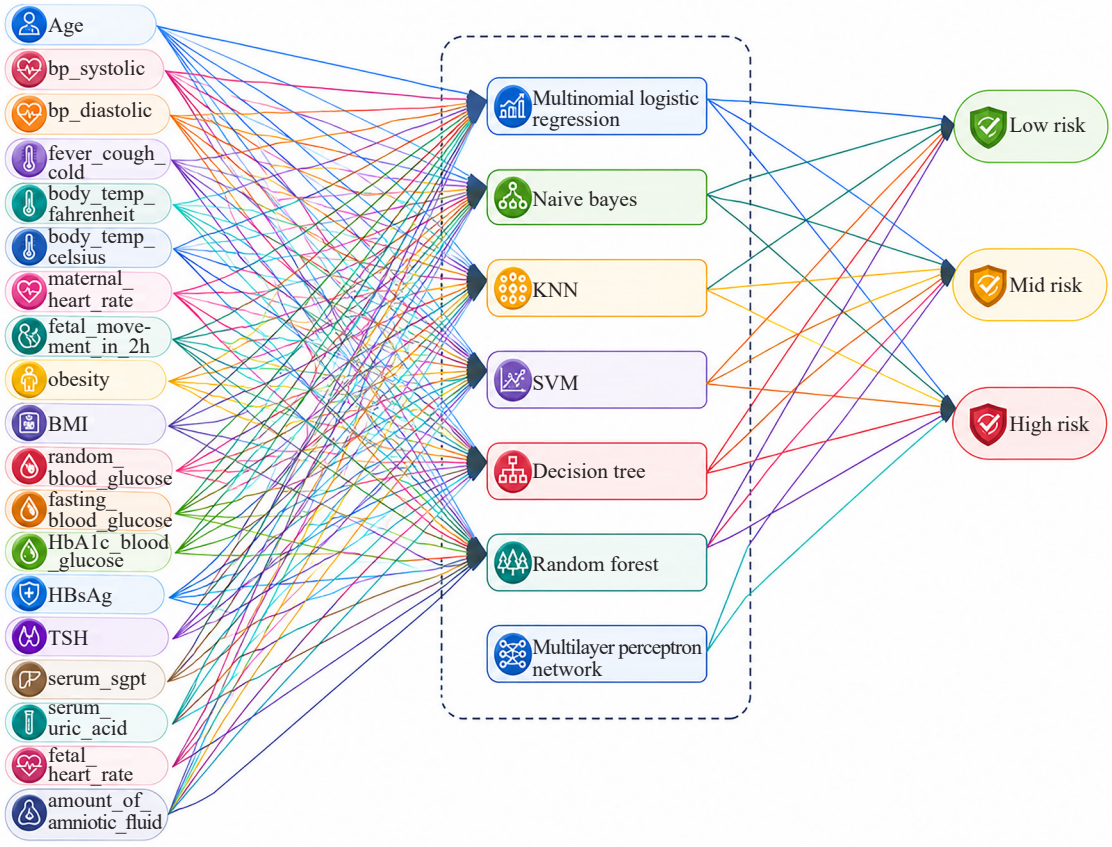


Figure 1. Input attributes, used algorithms, and output

The classification process entails predicting the class of a group of data points. The challenge of classification predictive modeling (y) is to approximate a mapping function (f) from discrete output variables (Y) to discrete input variables (X). Figure 1 illustrates the study's outcome, algorithms employed, and input attributes.

a) Multinomial Logistic Regression (MLR): A linear combination of the input features is transformed into a probability using the logistic function $h(x) = 1/(1 + e^{-X})$. Using a threshold, this probability is then converted into a binary classification [51]. Unlike linear regression, which produces a continuous output directly, logistic regression produces a probability for the base class; because the output is a probability, its value ranges from 0 to 1. Multinomial logistic regression extends this approach to problems with more than two classes by modelling a multinomial probability distribution over the possible class labels. One-vs-rest and one-vs-one wrapper strategies are examples of this approach, both of which were used in this study. For multinomial classification with K classes, MLR applies the softmax function to estimate the probability that input x belongs to class k:

$$P(y = k|x) = \exp(\beta_k{}^T x)/\Sigma_{j=1}{}^K \exp(\beta_j{}^T x),\ k = 1, 2, ..., K \tag{3}$$

where β_k is the weight vector for class k. Parameters are estimated by maximising the log-likelihood using solvers such as lbfgs, newton-cg, sag, or saga. In this study the solver yielding the highest test accuracy was selected (Table 2).

b) K-Nearest Neighbors (KNN): KNN can be used to address both classification and regression problems, though it is most frequently used for classification in practice. As a simple algorithm, KNN stores all available training examples and classifies new instances based on a majority vote among their k nearest neighbors. These distances can be calculated using the Euclidean, Manhattan, Minkowski, or Hamming methods: the first three are used for continuous variables, while the Hamming method is used for categorical variables. If K eq. (1), the instance is added to its nearest neighbor's class. Choosing an appropriate value of K can occasionally be challenging in KNN modelling [51]. The similarity between a query point x and a training sample x_i is measured using the Euclidean distance:

$$d(x, x_i) = \sqrt{(\Sigma_{j=1}{}^n (x_j - x_{ij})^2)}. \tag{4}$$

The k nearest training samples are identified, and the predicted class is determined by majority voting among those neighbours:

$$\hat{y} = \text{argmaxc}\ \Sigma_i \in N_k(x)\ \square\{y_i = c\} \tag{5}$$

where $N_k(x)$ is the set of k nearest neighbours of x and $\square\{\cdot\}$ is the indicator function. In this study $k = 1$ to $k = 4$ were evaluated; all values achieved 100% accuracy on the test set.

c) Naïve Bayes (NB): Naïve Bayes is a classification technique based on Bayes' theorem, under the assumption of conditional independence among predictors. The naïve Bayesian model is a straightforward classifier that performs well with large datasets. Bayes' theorem can be used to compute the posterior probability $P(c|x)$ from the class probability $P(c)$, the predictor prior probability $P(x)$, and the likelihood $P(x|c)$. Despite its simplicity, Naïve Bayes can perform competitively compared with more sophisticated classification methods in certain problem settings, although its relative performance depends on how well its independence assumption holds for the data being modelled. Formally, the posterior probability of class k given feature vector x is:

$$P(y = k|x) \propto P(y = k) \cdot \prod_{i=1}{}^n P(x_i|y = k) \tag{6}$$

For continuous features such as blood pressure, body temperature, and glucose levels, the class-conditional likelihood is modelled using a Gaussian (normal) distribution:

$$P(x_i|y = k) = (1/\sqrt{(2\pi\sigma^2_{ki})}) \cdot \exp(-(x_i - \mu_{ki})^2/(2\sigma^2_{ki})) \tag{7}$$

where μ_{ki} and σ^2_{ki} are the mean and variance of feature i for class k, estimated from the training data. The smoothing parameter $\alpha = 1$ was used in this study.

$$P(c|X) = P(x1|c) \times P(x2|c) \times \ldots \times P(xn|c) \times P(c) \tag{8}$$

d) Support Vector Machine (SVM): The SVM algorithm represents each data point as a point in an n-dimensional space, where n is the number of features and each feature's value corresponds to a specific coordinate. SVM finds the optimal separating hyperplane $f(x) = w^T x + b$ that maximises the margin between classes. The optimisation problem is:

$$\text{maximise } 2/\|w\| \text{ subject to } y_i(w^T x_i + b) \geq 1, \forall\ i \tag{9}$$

where w is the weight vector, b is the bias, and $y_i \in \{-1, +1\}$ are class labels. For non-linearly separable data, kernel functions map inputs to a higher-dimensional space; kernels evaluated in this study include Linear, Polynomial, RBF, and Sigmoid. Multi-class classification is handled using One-vs-Rest (OvR) decomposition. The linear kernel achieved the best accuracy (90.23%) and was selected for final reporting.

e) Decision Tree: A classifier that makes use of recursive instance space partitioning is a decision tree. Each internal node in a decision tree divides the instance space into two or more sub-spaces in accordance with a particular discrete function of the values of the input attributes. There are nodes and a root in this paradigm. Every node aside from the root has just one incoming edge. After conducting a test, intermediate nodes generate outgoing edges. Leaves are nodes without outgoing (also known as terminal or decision nodes).

The splitting criterion used is Gini impurity. For a node t containing samples from C classes with proportions $p(i|t)$ for class i, the Gini impurity is defined as:

$$\text{Gini}(t) = 1\text{-}\Sigma_i\ [p(i|t)]^2\ (i = 1, ..., C) \tag{10}$$

At each internal node the algorithm selects the attribute and threshold that minimise the weighted Gini impurity of the resulting child nodes. The maximum tree depth was set to 2 in this study.

f) Random Forest: It is a supervised classification algorithm. A random forest method consists of a large number of decision trees, which are a collection of categorization trees. A rule-based framework is built into every decision tree. For the given training dataset containing targets and features, the decision tree technique will have a set of rules. In contrast to decision trees, a random forest's root node can be located without doing an information gain computation. It makes a prediction based on the randomly generated rules of each decision tree and then logs the outcome. It also establishes the number of votes for each forecast target. The forecast that received the most votes is the one that was ultimately generated by the random forest approach.

Formally, given an ensemble of T decision trees $\{h_1(x), h_2(x), ..., h_t(x)\}$, the predicted class for input x is determined by majority voting:

$$\hat{y}(x) = \text{argmaxc}\ \Sigma_t\ ^{o}2\{h_t(x) = c\}\ (t = 1, ..., T) \tag{11}$$

where $^{o}2\{\cdot\}$ is the indicator function. Each tree h_t is trained on a bootstrap sample of the training data, and features are randomly subsampled at each split (feature bagging). In this study $T = 100$ trees were used.

g) Multilayer Perceptron: There may be more than one linear layer in a multilayer perceptron (combinations of neurons). If we take a three-layer network as an example, the input layer is at the top, the output layer is at the bottom, and the middle layer is referred to as the hidden layer. The input layer receives our input data, and the output layer receives our output. Feel free to increase the number of hidden layers in the model to make it more complex.

For a network with L layers, the forward propagation equations are:

$$a^L = \sigma(W^L\ a^{L\text{-}1} + b^L), \text{l} = 1, 2, ..., L \tag{12}$$

where W^L and b^L are the weight matrix and bias vector for layer l, $\sigma(\cdot)$ is the element-wise activation function (Rectified Linear Unit (ReLU) for hidden layers, softmax for the output layer), and $a^0 = x$ is the input feature vector. The network is trained by minimising the cross-entropy loss using the Limited-memory Broyden–Fletcher–Goldfarb–Shanno algorithm (BFGS) (lbfgs) solver with L_2 regularisation parameter $\alpha = 1 \times 10^{-5}$ and a maximum of 5,000 iterations.

3.6 *Performance metrics of algorithms*

Performance metrics in machine learning are measurements of how well an algorithm performs in relation to various criteria, such as accuracy, precision, recall, and others. The discussion of several performance measures follows:

a) Confusion Matrix: A confusion matrix, as its name suggests, tabulates a model's predictions and assesses its overall performance. It is a table with rows and columns for the Actual and Predicted classes, containing the values True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). It provides a straightforward way to evaluate the performance of a classification task, particularly when the output includes two or more distinct classes.

b) Accuracy: Accuracy is the ratio of the number of correctly predicted outcomes to the total number of predictions made. It is a reliable metric only when the classes are relatively balanced—that is, when each class contains a similar number of samples.

c) Precision: Precision measures the proportion of instances predicted as a given class that actually belong to that class, i.e., TP/(TP + FP). A model can be precise without being accurate: for example, if a person's actual height is 6 feet, and repeated predictions cluster tightly between 5.4 and 5.6 feet, the predictions are precise (consistent with one another) but not accurate (not close to the true value).

d) Recall: Recall (also known as sensitivity) is the proportion of actual positive cases that are correctly identified as positive, i.e., TP/(TP + FN).

e) F1-Score: The F1-Score balances precision and recall and is defined as their harmonic mean: F1 = 2 × (Precision × Recall)/(Precision + Recall). It ranges from 0 to 1, with higher values indicating better balance between precision and recall. A model can have high precision but low recall—meaning its positive predictions are usually correct, but it misses a substantial number of true positive cases—and the F1-Score is designed to capture this trade-off in a single number.

f) Area Under the Curve-Receiver Operating Characteristic (AUC-ROC): AUC-ROC is a performance metric for classification problems based on varying the decision threshold. The ROC curve plots the True Positive Rate (TPR, also known as recall or sensitivity) against the False Positive Rate (FPR = FP/(FP + TN), equal to 1 minus specificity) at various threshold levels. AUC (Area Under the Curve) measures the model's overall ability to distinguish between classes; model performance improves as the AUC value increases toward 1.0.

Improving the precision of risk assessments for maternal and fetal health is possible. The method used to gather the data and the features chosen to train the model influence the model's accuracy. Early identification and risk level prediction can reduce maternal and fetal morbidity during pregnancy and childbirth. Additionally, it aids the gynecologist in making informed decisions regarding the patient, taking into account their prior medical history. The following section discusses the findings and the implications of the research study.

4. Result and discussion

This section evaluates model efficacy using artificial intelligence and neural network approaches. It is divided into three main parts: performance analysis of the AI and neural network algorithms for predicting risk level in maternal and fetal health; comparison of actual and predicted values using new sample data; and sensitivity analysis of the input parameters against the predicted outcome. The most important factors affecting risk are assessed in the sensitivity analysis section. Thirty-three percent of the total dataset was used for testing, while 67% was used for training. Figure 2 shows the confusion matrix for each experimental model.

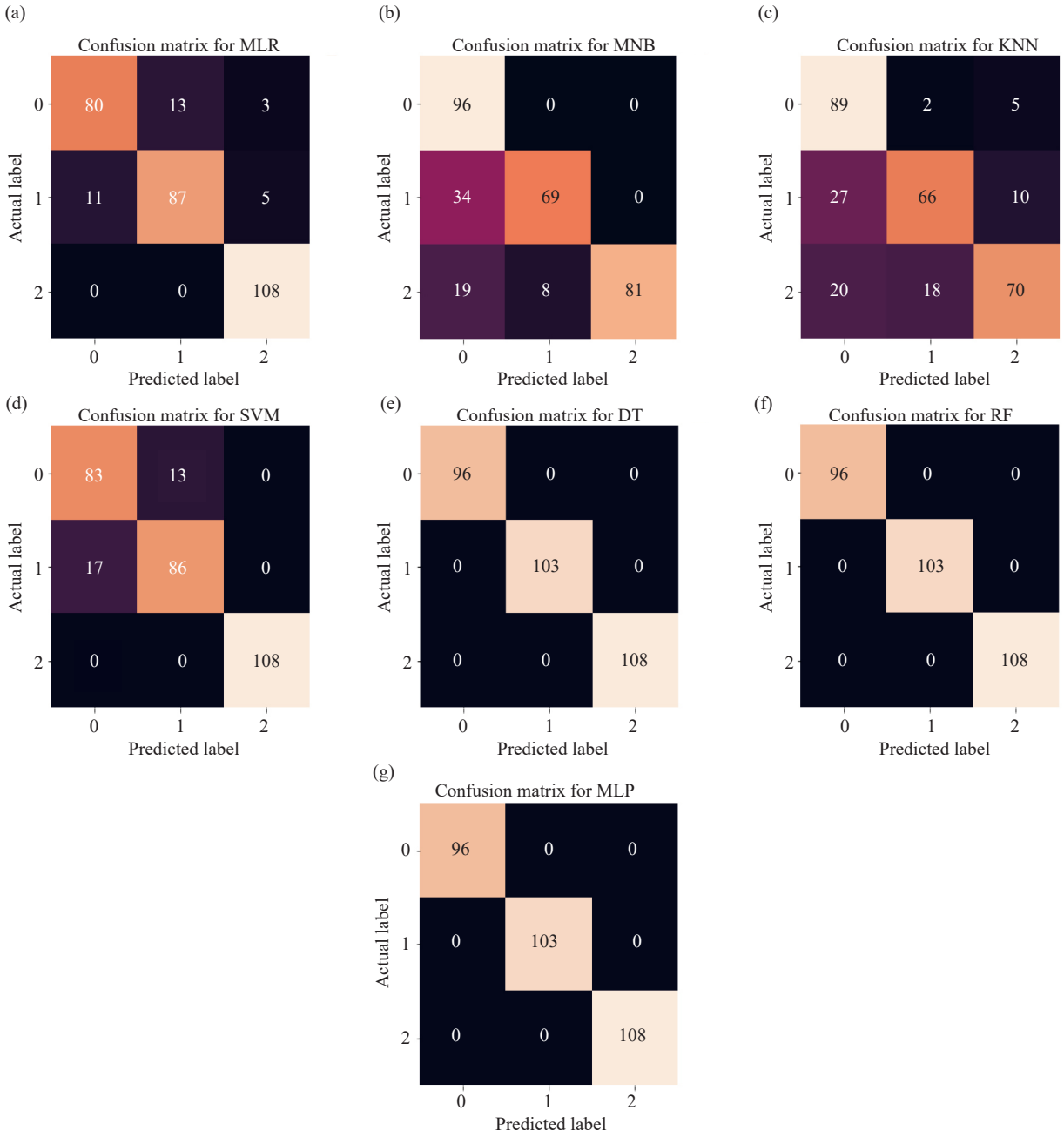


Figure 2. The confusion matrices of the prediction model (a) MLR (b) MNB (c) KNN (d) SVM (e) DT (f) RF (g) MLP

4.1 *Implementation of maternal and fetal health risk prediction model in python*

a) Multinomial Logistic Regression: Several solvers—liblinear, newton-cg, sag, and saga—were evaluated for the multinomial logistic regression model on the original dataset to identify the one yielding the best accuracy. The detailed evaluation of Multinomial Logistic Regression is presented in Table 2. According to Table 2, the Newton-CG solver achieves the highest accuracy for both the original and selected datasets, at 89.58% and 84.04%, respectively. Given that MLR's Receiver Operating Characteristic-Area Under the Curve (ROC-AUC) score was 98.09%, the model can reliably distinguish between true positives and true negatives.

In Table 2 'Accuracy for the dataset' refers to model accuracy when all 19 original features are used as input (the full feature set). 'Accuracy for the dataset selected' refers to model accuracy after manual feature selection, in which

13 features were retained based on clinical relevance and domain knowledge: Age, bp_systolic, body temperature (Fahrenheit), maternal heart rate, fetal movement, BMI, random blood glucose, HBsAg, TSH, serum SGPT, serum uric acid, fetal heart rate, and amniotic fluid index. Six features were excluded (bp_diastolic, fever/cough/cold, body temperature in Celsius, obesity, fasting blood glucose, and HbA1C) as they were either correlated with retained features or less discriminative. The selected feature set was applied consistently across all models. In the subsequent tables (Tables 3-8), the reported accuracy values correspond to the full (original) feature set unless otherwise stated.

Table 2. Multinomial logistic regression performance with different solvers

Solver of MLR	liblinear	newton-cg	sag	saga
Accuracy for the dataset	86.97%	89.58%	78.83%	79.48%
Accuracy for the dataset selected	80.13%	84.04%	79.80%	82.08%

b) Naïve Bayes: For the original dataset, the multinomial NB accuracy was 80.13%; for the selected dataset, it was 75.24%. The Multinomial NB confusion matrix, with $\alpha = 1$, is shown in Figure 2(b). Out of 307 risk levels, MNB correctly predicted 246. Of these, 81 high-risk cases were predicted accurately, along with 87 mid-risk and 80 low-risk cases, while MNB mispredicted eight high-risk cases as mid-risk, 19 high-risk cases as low-risk, and 34 mid-risk cases as low-risk. This occurs because, although some individual risk factors may indicate high risk on their own, the overall combination of factors places these cases into the low- or mid-risk category (and vice versa). The performance of MNB is presented in Table 3.

Table 3. Performance matrices of Multinomial Naïve Bayes

Performance metrics	Risk level		
	Low risk	Mid risk	High risk
Precision	0.64	0.90	1.00
Recall	1.00	0.67	0.75
F1 score	0.78	0.77	0.86

Table 4. Performance matrices of K-Nearest Neighbors

Performance metrics	Risk level		
	Low risk	Mid risk	High risk
Precision	0.65	0.77	0.82
Recall	0.93	0.64	0.65
F1 score	0.77	0.70	0.73

Table 3 shows that MNB provides the highest precision and F1-score for the high-risk class and the highest recall for the low-risk class. The highest recall value of 1.00 was achieved for the low-risk class, with precision and F1-score values of 0.64 and 0.78, respectively. On the other hand, the high-risk class yielded the highest precision value, i.e., 1.00, and recall and F1-score values were also high, i.e., 0.75 and 0.86. It is observable that MNB provides balanced

precision, recall, and F1 score for low-risk, mid-risk, and high-risk classes. The ROC_AUC score of the MNB is 94.78%, indicating that MNB correctly predicts Both True positives and negatives.

c) KNN: For the KNN algorithm, $K = 1$ to $K = 4$ was tried to find the best precision value. The highest accuracy of 100.0% was achieved for all values of k. The same accuracy was also observed for the selected dataset. The precision, recall, and F1 score of the KNN model to predict the risk level of maternal and fetal health are shown in Table 4.

Table 4 shows that KNN achieves the highest precision for the high-risk class, i.e., 0.82, as well as the highest recall and f1-score for the low-risk class, i.e., 0.93 and 0.77.

d) Support Vector Machine: Four kernels were examined for the original dataset, and the dataset was selected for the SVM classifier. The kernels were RBF, Sigmoid, Polynomial, and Linear. The Linear kernel's performance is higher than that of the Polynomial, RBF, and Sigmoid kernels for both the original and selected datasets. The accuracy for both datasets was 90.23% and 89.25%. The SVM's performance on the original dataset for these four kernels is presented in Table 5.

Table 5. Performance of SVM classifier for different kernels

Kernel type	Metrics	Risk level		
		Low risk	Mid risk	High risk
Linear	precision	0.83	0.87	1.00
	recall	0.86	0.83	1.00
	f1-score	0.85	0.85	1.00
	Accuracy		90.23%	
	ROC_AUC score		98.05%	
Polynomial	precision	0.62	0.82	0.89
	recall	0.92	0.60	0.72
	f1-score	0.74	0.69	0.80
	Accuracy		74.27%	
	ROC_AUC score		90.34%	
Sigmoid	precision	0.29	0.19	0.56
	recall	0.75	0.10	0.05
	f1-score	0.42	0.13	0.09
	Accuracy		28.34%	
	ROC_AUC score		66.75%	
RBF	precision	0.61	0.75	0.94
	recall	0.92	0.66	0.63
	f1-score	0.73	0.70	0.76
	Accuracy		72.96%	
	ROC_AUC score		88.21%	

Among the four, the SVM classifier with a linear kernel performed best, and the sigmoid kernel performed worse in terms of accuracy. 90.23% accuracy was achieved with the Linear kernel. In comparison, an accuracy of 28.34%

was achieved with SVM using the sigmoid kernel. On the other hand, the polynomial kernel and RBF kernel achieve accuracies of 74.27% and 72.96%, respectively. For the linear kernel and high-risk classes, the best precision, recall, F1-score, and ROC-AUC scores were also attained. The confusion matrix for the SVM linear kernel on the testing data is depicted in Figure 2(d). Among the 307 test risk classes, the linear SVM correctly predicted 293 risk levels. If we consider the high-risk class, the TP rate for this class was 108, the FN rate was 0, the FP rate was 0, and the TN rate was 199. The low-risk class had the lowest TP. The performance of a linear SVM classifier can be considered for predicting the risk level related to maternal and fetal health.

e) Decision tree: The main goal of decision trees is to identify the descriptive features that provide the most "information" about the target feature. Once these features are identified, the dataset is divided along their values so that the target feature values for the resulting sub-datasets are as pure as possible. This search for the "most informative" feature continues until a halting criterion is met; at this point, we arrive at what is known as leaf nodes. The Decision Tree for the maternal and fetal health data is shown in Figure 3. The maximum depth of the decision tree was set to 2, and the model achieved 100.0% accuracy on both the original and selected datasets. The confusion matrix for the decision tree algorithm is shown in Figure 2(e).

The DT algorithm had a good separability measure for the maternal and fetal health risk factors. The 'High-risk' class's DT model TP value was 108, FP value was 0, FN value was 0, and TN value was 199. The 'Mid-risk' class had a TP value of 103, while the 'Low-risk' class had a value of 96. The FP and FN values for all classes are 0, as the accuracy of the DT classifier is 100.0%. Hence, the algorithm correctly predicts all maternal and fetal health risk levels. The ROC_AUC score of the algorithm is also 100.0%.

Table 6. Accuracy achieved for Multilayer Perceptron with the number of hidden layers

Number of hidden layers	10	11	12	13	14	15	16
Accuracy achieved for MLP	72.62%	61.56%	94.79%	100.00%	94.46%	100.00%	100.00%

Table 7. The TP, FP, FN, and TN values for the Multilayer Perceptron algorithm

Risk level	Value			
	TP	FP	FN	TN
Low risk	96	0	0	211
Mid risk	103	0	0	204
High risk	108	0	0	199

Table 8. Comparison of performance between different algorithms

Metrics in percent	Algorithms								
	In python							In MATLAB	
	MLR	NB	KNN	SVM	DT	RF	MLP	DT	MLP
Precision	89.33	80.25	74.85	89.96	100.0	100.0	100.0	-	-
Recall	89.26	80.66	73.87	89.98	100.0	100.0	100.0	-	-
F1 score	89.23	84.68	73.03	89.95	100.0	100.0	100.0	-	-
Accuracy	89.58	80.13	100.0	90.23	100.0	100.0	100.0	100.0	100.0
ROC_AUC score	98.09	94.78	91.3	98.05	100.0	100.0	100.0	-	-

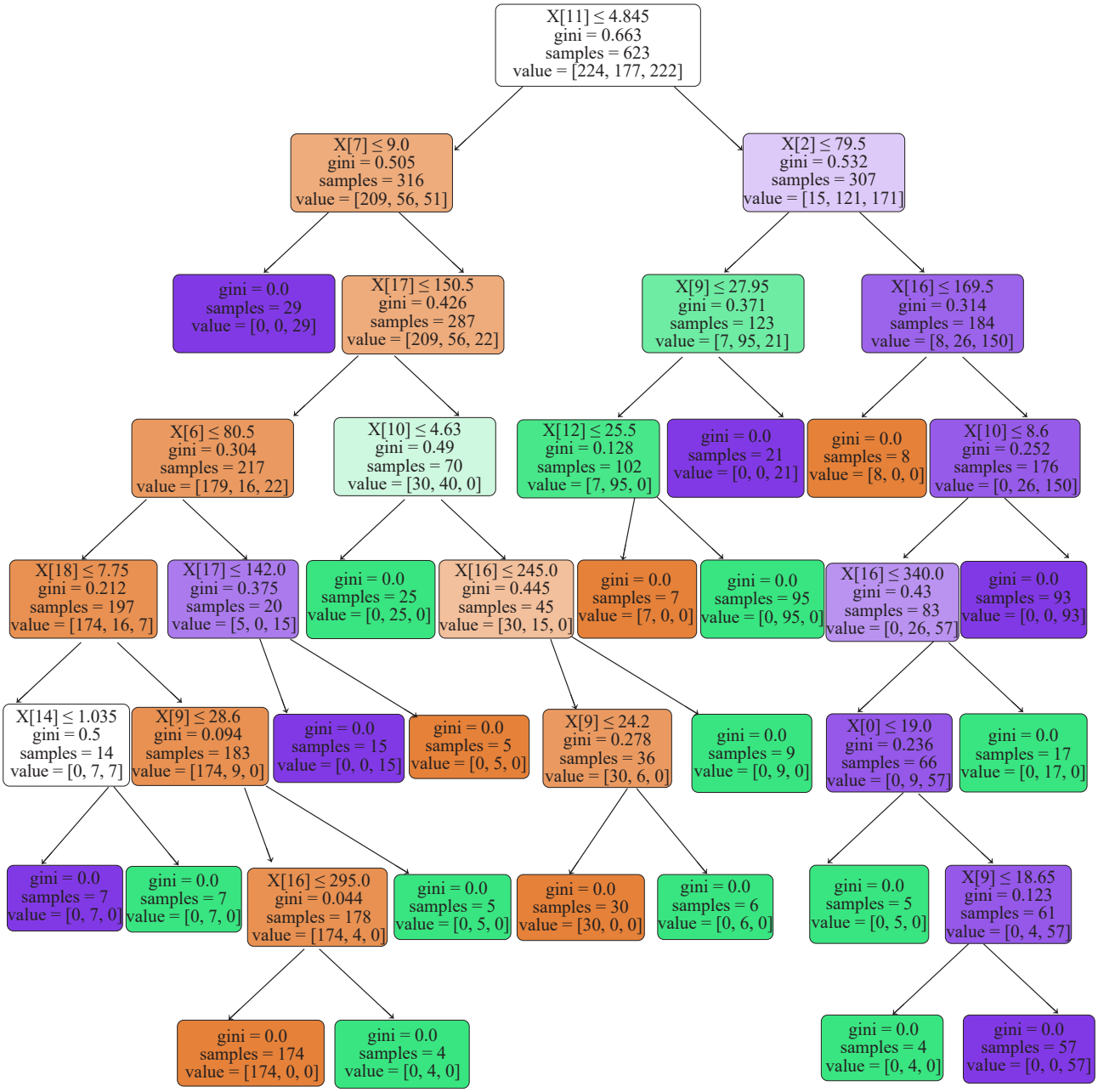


Figure 3. Decision Tree for the maternal and fetal health data

f) Random Forest (RF): The RF method combines the outputs of a collection of decision trees—a process known as bagging—to produce a more accurate overall prediction. Because RF models are resistant to overfitting, the number of trees can, in principle, be increased without penalty, and using a larger number of trees is generally recommended. In this study, an initial ensemble of 100 trees achieved a satisfactory accuracy of 100.0%. The confusion matrix for the RF model is shown in Figure 2(f). For the 'High risk' class, the RF model's TP value was 108, with FP and FN values of 0 and a TN value of 199. The 'Mid risk' class had a TP value of 103, and the 'Low risk' class had a TP value of 96. Because the RF model correctly detects all risk levels, the FP and FN values are 0 for every class, giving a precision, recall, and F1 score of 1.0 for each class.

g) Multilayer Perceptron: The MLP algorithm was tested with different numbers of hidden layers, using a regularisation parameter of alpha = 1e-5 and a maximum of 5,000 iterations. The highest accuracy—100.0% on both the

original and selected datasets—was achieved with 13, 15, and 16 hidden layers. The initial number of hidden layers was set to 10. The accuracy obtained for each number of hidden layers is given in Table 6.

Accuracy fluctuated with the number of hidden layers, at times rising and at times falling. The lowest accuracy occurred with 11 hidden layers, and the highest occurred with 13, 15, and 16 hidden layers. Taking the median configuration of 15 hidden layers, the remaining performance metrics were analysed for this setting. The TP, TN, FP, and FN values obtained by the MLP model for each risk level are shown in Table 7.

For the 'High risk' class, the MLP model's TP value was 108, with FP and FN values of 0. The TP values for the 'Low risk' and 'Mid risk' classes were 96 and 103, respectively. The precision, recall, and F1-score for the low-risk, mid-risk, and high-risk classes all reached the maximum value of 1.0, since the FP and FN values for each class were zero. The MLP algorithm with 15 hidden layers therefore correctly detects the risk levels.

4.2 *Prediction using distinct techniques and MATLAB tools*

Here, MATLAB tools were used for performance evaluation of maternal and fetal health risk prediction using a Decision Tree and a Multilayer Perceptron network.

a) Decision Tree Algorithm: The Decision tree based on the maternal and fetal health data evaluated using the MATLAB platform is given in Figure 4. This tree predicts classifications based on 19 predictors. To predict, start at the top node, represented by a triangle (Δ). The first decision is whether the Random blood sugar (Ran_BG) is smaller than 5.915; then, it follows the left branch, i.e., node 2. If Ran_BG exceeds 5.915, follow the right branch to node three; otherwise, classify the data as High risk. The same procedures are repeated for all other predictors.

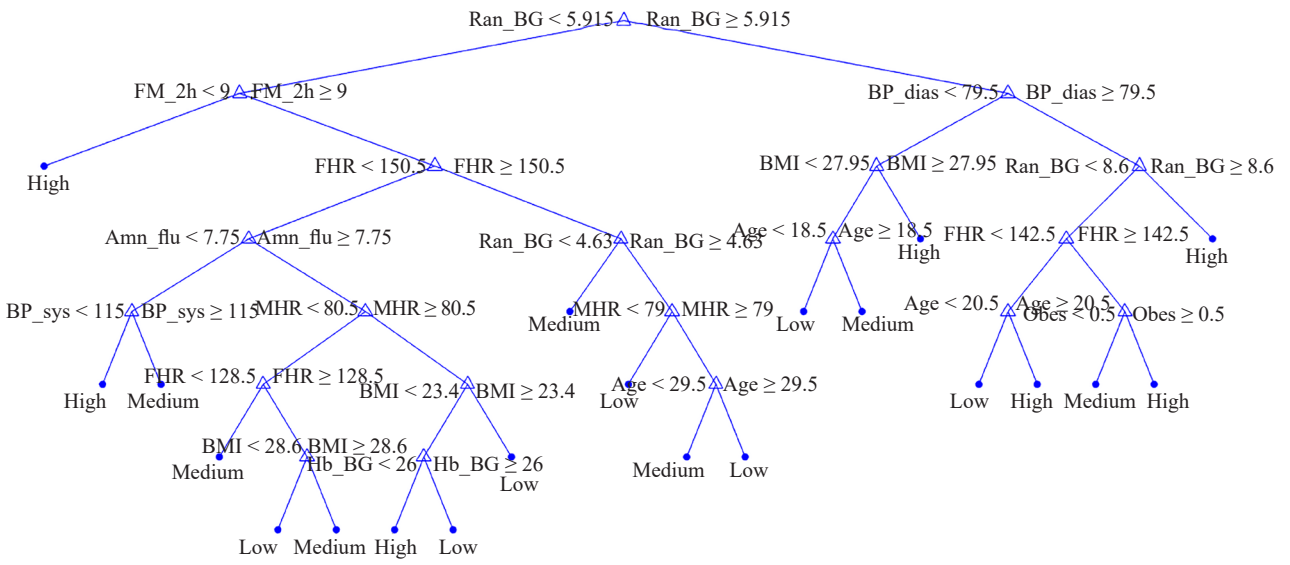


Figure 4. Decision Tree for the maternal and fetal health data

It is observed that the DT classifier performs well on both platforms, namely Jupyter Notebook and MATLAB. Therefore, the DT classifier model can be utilized in the proposed maternal and fetal health risk prediction model.

b) Multilayer Perceptron: For the training dataset with the Multilayer Perceptron Algorithm, load the dataset, initialize the input parameters, and the target value. Nineteen parameters are used as input in the input layer, with one hidden layer containing 15 neurons, as shown in Figure 5.

The green cell in Figure 6(a), which represents the confusion matrix produced by the multilayer perceptron method, indicates the location where the output class matches the target class. Where the output class does not match the target class, a red cell is displayed. The error histogram for the MLP algorithm is displayed in Figure 6(b). Targets-Outputs mistakes are represented on the X-axis, and occurrences are represented on the Y-axis.

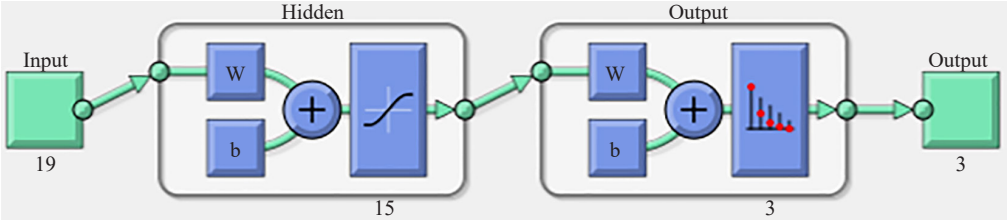


Figure 5. Architecture of the Multilayer Perceptron algorithm

The confusion matrix is generated using the trainmlp function with the MLP ANN algorithm, and the results are presented in Figure 6(a). Notably, 100% of class 1 instances are true positives, while there are no false negatives. The following columns also show the same results: 100% of class 2 and class 3 are correctly classified, with a false negative rate of 0%. In Figure 6(b), the zero error point falls within the bin with a center value of 5.88e-08, which is the closest value to zero. Figure 7(a) shows the ROC curve for measuring performance. The Y-axis displays the real positive rate, while the X-axis displays the false positive rate. For Classes 1, 2, and 3, there is a 1.0 area under the ROC.

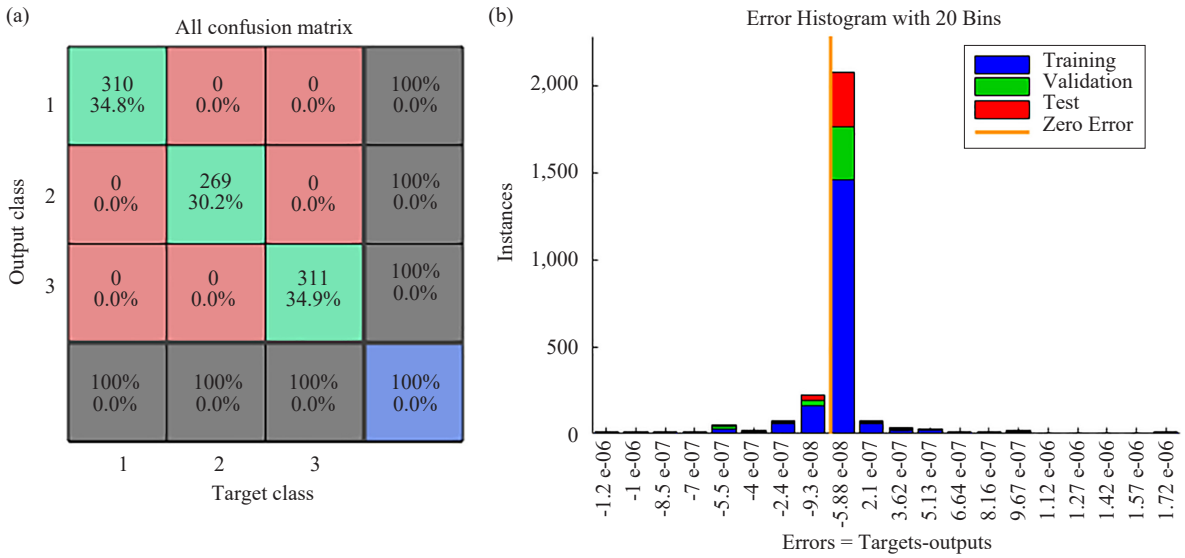


Figure 6. (a) Confusion matrix for Multilayer Perceptron algorithm (b) Error histogram for Multilayer Perceptron algorithm

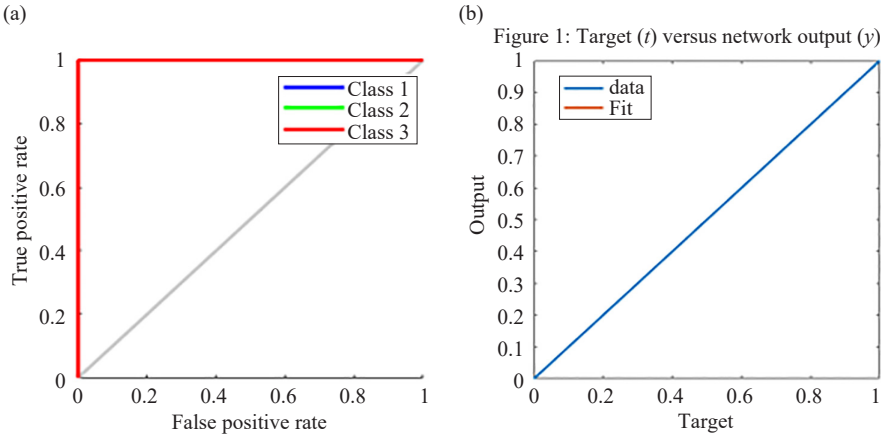


Figure 7. (a) ROC curve for evaluating the effectiveness of a multilayer perceptron (b) Multilayer Perceptron algorithm's regression graph

The regression graph generated by the MLP algorithm is shown in Figure 7(b). The ideal result, which occurs when outputs and targets are equal, is shown by the dashed line in each plot. The best-fitting linear regression line between the targets and outputs is shown as a solid line. The R-value indicates the relationship between the outputs and the targets. The relationship between the outputs and the targets is linear if $R = 1$. Assuming an exact linear relationship between the outputs and the targets in this instance, $R = 1$, we can say that.

The performance graph for the training, validation, and testing of the MLP algorithm using the datasets for maternal and fetal health is shown in Figure 8. The epochs are plotted on the Y-axis. In contrast, the cross-entropy loss function is shown on the X-axis. The best validation performance was discovered to be 5.1695 Ee-08 at 198 epochs.

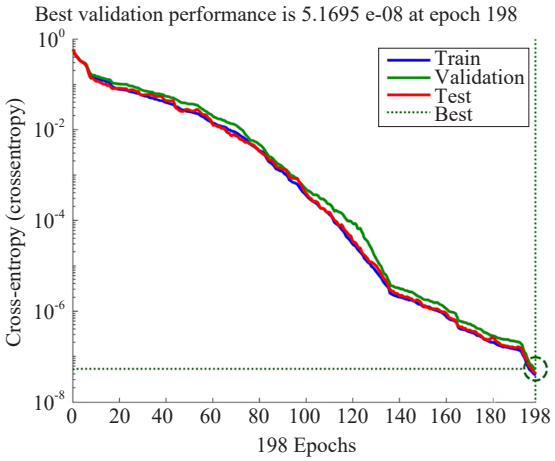


Figure 8. Performance graph of the Multilayer Perceptron algorithm for training, testing, and validation

4.3 *Comparison*

Table 8 compares the ROC-AUC score, f1-score, recall, precision, and prediction accuracy for the MLR, MNB, KNN, SVM, DT, RF, and MLP algorithms.

Table 8 demonstrates that the three models with the highest accuracy, precision, recall, F1-score, and ROC-AUC score, or 100%, were DT, RF, and MLP. To confirm performance, MATLAB analysis of the DT and MLP algorithms was also performed. The same positive outcomes were also obtained using MATLAB. The KNN method achieves 100% accuracy, but its precision, recall, F1-score, and ROC-AUC score should be improved. Aside from KNN, DT, RF, and MLP algorithms, MLR, MNB, and SVM also produced good results, although not as well. The performance results, which include accuracy, precision, recall, F1-score, and ROC-AUC score, show that the DT, RF, and MLP algorithms yield the best results for the maternal and fetal health datasets. Therefore, additional analysis and work improvement can be done using the DT, RF, and MLP models.

Regarding the missing MATLAB metrics in Table 8, the MATLAB implementations of DT and MLP were used exclusively for cross-platform accuracy validation. Only the overall classification accuracy was extracted from MATLAB, as the primary purpose of the MATLAB experiments was to confirm that the 100% accuracy result observed in Python was not an artefact of the scikit-learn implementation. Detailed per-class precision, recall, F1-score, and ROC-AUC metrics were therefore not computed in the MATLAB environment; these are reported for the Python experiments only (shown as '—' for MATLAB in Table 8).

4.4 *Testing models with new samples*

The algorithms will now be used for testing with new samples that have not been used before. Table 9 compares the actual and predicted outputs of the algorithms based on new instances of maternal and fetal health data.

Table 9. Actual and Predicted values of new samples of AI and NN algorithms

Actual value	predicted value						
	MLR	NB	KNN	SVM	DT	RF	MLP
mid risk	mid risk	mid risk	mid risk	mid risk	mid risk	mid risk	mid risk
mid risk	mid risk	mid risk	mid risk	mid risk	mid risk	mid risk	mid risk
mid risk	mid risk	mid risk	mid risk	mid risk	mid risk	mid risk	mid risk
high risk	high risk	high risk	mid risk	high risk	high risk	high risk	high risk
low risk	low risk	low risk	low risk	low risk	low risk	low risk	low risk
high risk	low risk	low risk	low risk	low risk	high risk	high risk	low risk
low risk	low risk	low risk	low risk	low risk	low risk	low risk	low risk
mid risk	mid risk	mid risk	mid risk	mid risk	mid risk	mid risk	mid risk
mid risk	mid risk	mid risk	mid risk	mid risk	mid risk	mid risk	mid risk
mid risk	mid risk	mid risk	mid risk	mid risk	mid risk	mid risk	mid risk
high risk	mid risk	mid risk	mid risk	mid risk	high risk	high risk	mid risk
high risk	mid risk	mid risk	mid risk	mid risk	high risk	mid risk	high risk
mid risk	mid risk	mid risk	mid risk	mid risk	mid risk	mid risk	mid risk
high risk	high risk	high risk	high risk	high risk	high risk	high risk	high risk
low risk	low risk	low risk	low risk	low risk	low risk	low risk	low risk
high risk	high risk	high risk	high risk	high risk	high risk	high risk	high risk
low risk	low risk	low risk	low risk	low risk	low risk	low risk	low risk
low risk	low risk	low risk	low risk	low risk	low risk	low risk	low risk
mid risk	mid risk	mid risk	mid risk	mid risk	mid risk	mid risk	mid risk
high risk	high risk	high risk	high risk	high risk	high risk	high risk	high risk
high risk	high risk	high risk	high risk	high risk	high risk	high risk	high risk
low risk	low risk	low risk	low risk	low risk	low risk	low risk	low risk
low risk	low risk	low risk	low risk	low risk	low risk	low risk	low risk
high risk	low risk	low risk	low risk	low risk	high risk	high risk	low risk
high risk	high risk	high risk	high risk	high risk	high risk	high risk	high risk
mid risk	mid risk	mid risk	mid risk	mid risk	mid risk	mid risk	mid risk
mid risk	mid risk	mid risk	mid risk	mid risk	mid risk	mid risk	mid risk
high risk	high risk	high risk	high risk	high risk	high risk	high risk	high risk
mid risk	mid risk	mid risk	mid risk	mid risk	mid risk	mid risk	mid risk
low risk	low risk	low risk	low risk	low risk	low risk	low risk	low risk
low risk	low risk	low risk	low risk	low risk	low risk	low risk	low risk
Accuracy	70.96%	87.09%	83.87%	87.09%	100.00%	96.77%	90.32%

One of the main tasks of the research is to show the prediction using new samples. The depiction of the algorithm accuracy on the new sample data is shown in Figure 9. Table 9 indicates that the DT, RF, and MLP algorithms also achieved better accuracy for the new sample data. The accuracy of the MLR, MNB, KNN, and SVM algorithms is also satisfactory, but less than that of the DT, RF, and MLP algorithms.

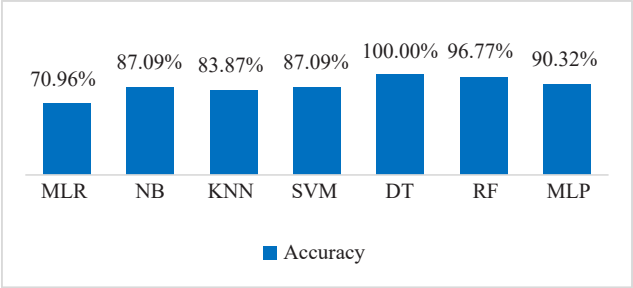


Figure 9. Accuracy of the algorithms on new sampled maternal and fetal health data

The study's primary goal is to identify the most effective algorithm for achieving better accuracy than the existing early maternal and fetal health risk prediction system. The DT, RF, and MLP algorithms can be selected from this analysis to extend further the maternal and fetal health risk prediction and prevention system.

4.5 *Discussion on model performance*

The 100% accuracy achieved by the DT, RF, and MLP algorithms warrants careful interpretation. Three factors collectively explain this outcome and should be considered when generalising the results.

First, the risk labels are derived from crisp, rule-based clinical thresholds (Table 1). Decision-tree-based models (DT and RF) are inherently well-suited to replicating such threshold rules: with a maximum depth of 2, the DT can partition the feature space according to the very thresholds used to generate the labels, making perfect classification on the training/test split achievable without overfitting in the conventional sense. In other words, the task is closer to rule-recovery than to generalisation learning.

Second, the near-balanced class distribution (low: 31.4%, mid: 34.5%, high: 34.1%) means that a model achieving 100% accuracy does so across all three classes, and this is reflected in the confusion matrices where all TP values are non-zero and all FP/FN values are zero. This rules out the trivial 100%-accuracy scenario caused by class imbalance (e.g., a degenerate model predicting only the majority class).

Third, for KNN (k = 1-4), 100% accuracy on the training-derived test split arises because KNN is a non-parametric, instance-based learner that can achieve perfect memorisation on clean, rule-labelled data at small k. However, the lower per-class precision, recall, and F1-score (e.g., precision 0.65 for low-risk) indicate that the model generalises less reliably than DT/RF/MLP to borderline cases.

While these results are promising, we emphasise that perfect accuracy under rule-based labelling does not imply real-world clinical perfection. The models are trained to replicate the labelling scheme, not to surpass it. External validation on an independent prospective cohort—where risk labels are assigned by clinicians independently of the algorithmic thresholds—is required before these models can be deployed in clinical practice. Future work should also incorporate k-fold cross-validation (recommended: 5-fold or 10-fold) to assess the stability of the reported metrics across different data partitions.

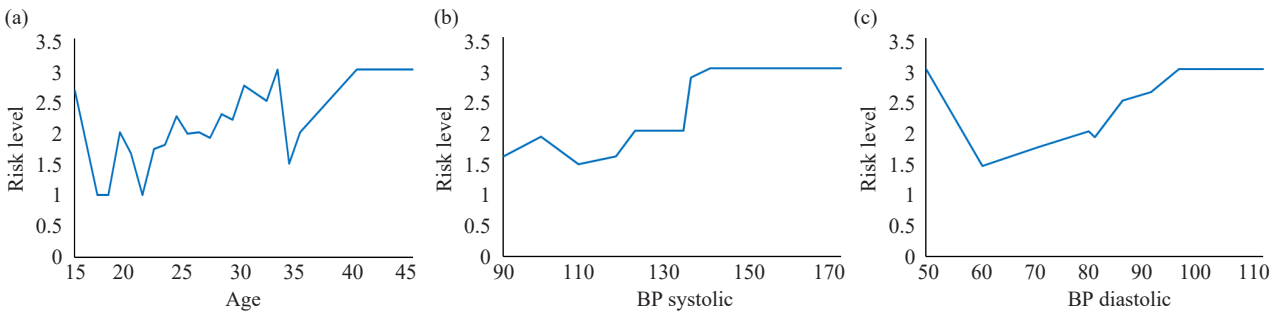
4.6 *Sensitivity*

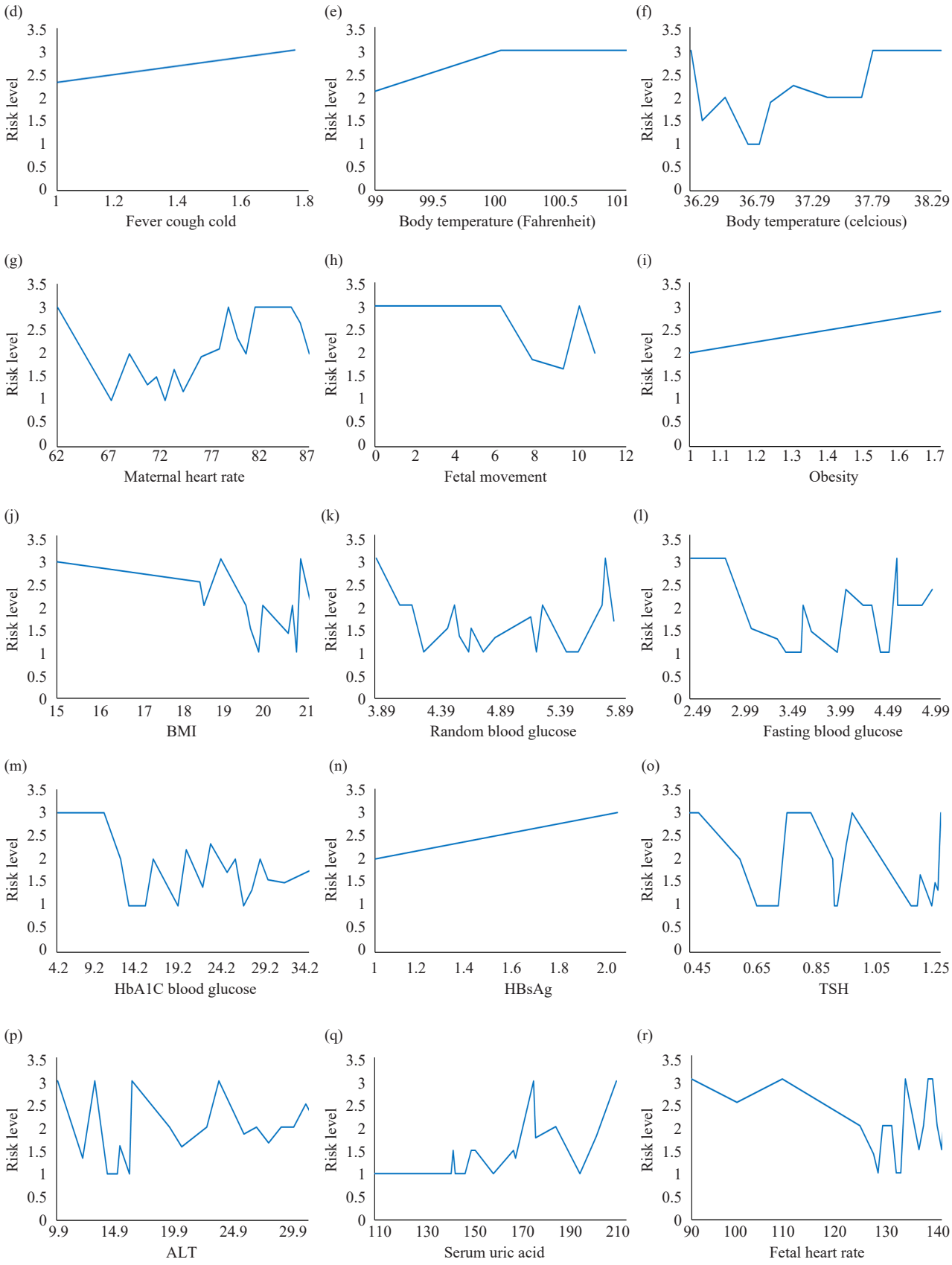
Sensitivity analysis is a technique that involves carefully altering model inputs and assessing the effects of these changes on the model's output. This technique demonstrates the model's sensitivity to these modifications as well as the effects of specific features on the model's functionality. Determine the interactions between the input and target factors using sensitivity analysis. The sensitivity values of the input features used in this study are shown in Table 10.

Table 10. Sensitivity analysis values for maternal and fetal health risk level

Name of input parameters	Sensitivity
Age	64.83
bp_systolic	97.21
bp_diastolic	97.72
fever_cough_cold	45.36
body_temp_fahreheit	60.20
body_temp_celsius	65.99
maternal_heart_rate	78.50
fetal_movement_in_2_hour	92.62
Obesity	59.74
BMI	80.27
random_blood_glucose	75.67
fasting_blood_glucose	85.99
HbA1C_blood_glucose	81.46
HBsAg	53.62
TSH	51.06
serum_sgpt	63.52
serum_uric_acid	51.25
fetal_heart_rate	91.69
amount_of_amniotic_fluid	93.47

How specific input parameters influence risk, and how risk level changes with each factor, is depicted in Figure 10. Table 10 shows that input parameters such as blood pressure (systolic), blood pressure (diastolic), fetal movement over two hours, fetal heart rate, and amniotic fluid amount influence risk level the most compared to the other input parameters. After those parameters, BMI, fasting blood glucose, HbA1c, and maternal heart rate also significantly influence risk level. This analysis indicates that imbalances in these parameters increase risk, with correspondingly severe consequences for both the mother and the child.




(d)
Risk level
Fever cough cold
(e)
Risk level
Body temperature (Fahrenheit)
(f)
Risk level
Body temperature (celcious)
(g)
Risk level
Maternal heart rate
(h)
Risk level
Fetal movement
(i)
Risk level
Obesity
(j)
Risk level
BMI
(k)
Risk level
Random blood glucose
(l)
Risk level
Fasting blood glucose
(m)
Risk level
HbA1C blood glucose
(n)
Risk level
HBsAg
(o)
Risk level
TSH
(p)
Risk level
ALT
(q)
Risk level
Serum uric acid
(r)
Risk level
Fetal heart rate

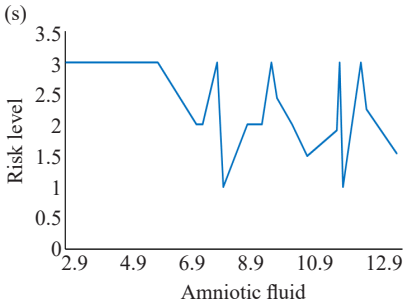


Figure 10. Effect of distinct variables on maternal and fetal health risk. (a) Age (b) BP Systolic (c) BP Diastolic (d) Fever cough cold (e) Body Temperature (Fahrenheit) (f) Body Temperature (Celsius) (g) Maternal Heart Rate (h) Fetal Movement (i) Obesity (j) BMI, (k) Random Blood Glucose (l) Fasting Blood Glucose (m) HbA1C Blood Glucose (n) HBsAg (o) TSH (p) Serum SGPT/ALT (q) Serum Uric Acid (r) Fetal Heart Rate(FHR) and (s) Amniotic Fluid

• Effect of age on maternal and fetal health risk: Age significantly influences the maternal and fetal health risk level when a mother's age is less than 15 or greater than 35 years, as shown in Figure 10(a). Risky situations can be created within this range of ages, depending on the other maternal and fetal health risk parameters.

• Effect of BP systolic on maternal and fetal health risk: Figure 10(b) illustrates how high blood pressure affects the health risks to both the mother and the fetus. Systolic blood pressure is the second most important factor in the sensitivity analysis of maternal and fetal health risk. Systolic blood pressure that is persistently high, at or above 140 mm Hg, can lead to a dangerous condition called preeclampsia. Preeclampsia, pronounced pre-e-CLAMP-si-a, is a hypertensive disorder that can develop during pregnancy or shortly after childbirth, and is commonly associated with elevated protein levels in the urine and a recent onset of high blood pressure.

• Effect of BP diastolic on maternal and fetal health risk: BP diastolic is the most significant parameter in the sensitivity analysis of maternal and fetal health risk. A high diastolic reading can also contribute to dangerous conditions such as preeclampsia, and in severe or untreated cases can lead to serious complications for the mother and the child. The effect of BP diastolic on maternal and fetal health risk is depicted in Figure 10(c).

• Effect of Fever, cough, and cold on maternal and fetal health risk: Fever, cough, and cold exhibit a linear relationship with risk level, as shown in Figure 10(d). A pregnant woman's immune system may change during the pregnancy, making her more prone to the flu and the common cold. Numerous viral types, including rhinoviruses, which are the most prevalent, can cause a cold. A viral or bacterial infection in the respiratory tract, which a cold can cause, leads to a cough. Even after a cold, respiratory airways may be swollen, painful, and inflamed, resulting in a chronic cough. Having a fever and a cold during pregnancy is also risky for the mother and the child.

• Effect of Body Temperature (Fahrenheit) on Maternal and Fetal Health Risk: Here, the effect of Body Temperature (Fahrenheit) on maternal and fetal health risk is depicted in Figure 10(e). Although 98.6 degrees Fahrenheit is considered a typical body temperature, pregnant women often experience slightly higher body temperatures, especially during the first trimester, which is a normal occurrence. The increasing progesterone levels in a woman's body during the first trimester are primarily responsible for this rise in body temperature. During a woman's ovulation cycle, progesterone is also responsible for increasing the basal body temperature. A first-trimester pregnancy with low body temperature may also have low progesterone levels, which raises the possibility of a miscarriage.

• Effect of Body Temperature (°C) on Maternal and Fetal Health Risk: Figure 10(f) illustrates how body temperature (°C) affects the risk to maternal and fetal health. One is said to have a high temperature if their body temperature exceeds 37.5 degrees Celsius. A miscarriage may occur if the temperature exceeds 39.5 degrees, especially in the first trimester. Birth abnormalities, particularly neural tube defects, have been associated with high temperatures during the first trimester. Neural tube abnormalities impact spinal cord and brain function. Because the brain and neurological system fully develop during the first trimester, high fevers raise the chance of congenital abnormalities. Preterm labor can happen when an infection, particularly a kidney or uterine infection, causes a fever. These infections may cause the mother's body to attempt an early child delivery.

• Effect of Maternal Heart Rate on Maternal and Fetal Health Risk: The maternal and fetal health risks associated

with heart rate are depicted in Figure 10(g). Bradycardia deprives both the pregnant person and the fetus of oxygen. During pregnancy, the average maternal heart rate typically increases by 10 to 20 beats per minute. As a result, tachycardia is defined as a resting heart rate of at least 100 beats per minute [52]. A woman's doctor should conduct a thorough check if her heart rate falls under 60 beats per minute at rest, to rule out any underlying cardiovascular or other medical issues. Only a few of the numerous potential causes of a slow heart rate that doctors can pinpoint include certain medications, congenital heart defects, high blood pressure, heart infections, scarring from heart surgery, disorders that cause excess iron in the body, hypothyroidism, sleep apnea, and an imbalance of electrolytes in the blood.

• Effect of Fetal Movement on maternal and fetal health risk: The effect of Fetal Movement on maternal and fetal health risks is depicted in Figure 10(h). Some mothers report feeling their baby move as early as 13 to 16 weeks after the start of their last period. These initial fetal movements, known as quickening, are often referred to as flutters. A pregnant lady may decide to quicken and count the quantity and variety of movements she feels the fetus making. A kick count is the colloquial name for this total. According to the American Pregnancy Association, the benefits of performing kick counts range from allowing a pregnant woman to bond with her unborn child to lowering the incidence of stillbirth; kick counts are recommended in high-risk pregnancies. However, it's possible that telling women to watch for fetal movements will make them more anxious. When there has just been one episode of decreased fetal movements, 70% of pregnancies are straightforward. If a woman is 28 weeks or more pregnant and believes her baby has stopped moving or is moving less, she should contact her midwife, according to the UK's Royal College of Obstetricians and Gynaecologists.

• Effect of Obesity on Maternal and Fetal Health Risk: Here, the effect of obesity on maternal and fetal health risks is depicted in Figure 10(i). The most prevalent medical condition among women of reproductive age is obesity. When a person's BMI is 30 or greater, obesity is declared. Pregnancy-related obesity has adverse short-and long-term effects on the mother and the unborn child. In early gestation, obesity promotes spontaneous pregnancy loss, infertility issues, and congenital abnormalities. Metabolically, obese women have higher insulin resistance in the first trimester of pregnancy, which clinically manifests as glucose intolerance and fetal enlargement in the second and third trimesters. The likelihood of a cesarean delivery and wound complications rises as gestation progresses. Obese postpartum women are more likely to experience depression, venous thromboembolism, and breastfeeding challenges.

• Effect of BMI on maternal and fetal health risk: Here, a higher BMI leads to obesity, and subsequently, various risky situations arise from it. The effect of BMI on maternal and fetal health risks is illustrated in Figure 10(j).

• Effect of Random Blood Glucose on maternal and fetal health risk: Here, the effect of Random Blood Glucose on maternal and fetal health risks is depicted in Figure 10(k). One of the most significant aspects of pregnancy is maintaining blood glucose control. There is a slight increase in the risk of infant mortality or congenital disabilities in babies of mothers with diabetes compared to those without. When a pregnant woman has an abnormally high blood glucose level, the baby will naturally store excess glucose as body fat. It has been reported that, on average, 2% to 4% of women develop temporary diabetes, also known as gestational diabetes. As a result, when the baby is born at the gestational age, it will be bigger than usual. The medical term for this is fetal macrosomia.

• Effect of Fasting Blood Glucose on maternal and fetal health risk: Here, the effect of Fasting Blood Glucose on maternal and fetal health risks is depicted in Figure 10(l). Fasting Plasma Glucose (FPG) is frequently assessed in the first trimester of pregnancy to screen for diabetes (FPG level of 7 mmol/L). Early intermediate hyperglycemia, defined as FPG in the range of 5.1 to 6.9 mmol/L, and early gestational diabetes mellitus have been discovered due to this screening (Pre-existing Gestational Diabetes Mellitus (PGDM). Although early FPG has been linked to bad pregnancy outcomes, the International Association of Diabetes and Pregnancy Study Groups (IADPSG's) recommendation to refer women with eGDM for prompt care is more practical than evidence-based [53].

• Effect of HbA1C Blood Glucose on maternal and fetal health risk: Here, the effect of HbA1C Blood Glucose on maternal and fetal health risk is depicted in Figure 10(m). A mother and her unborn child are less likely to experience difficulties thanks to good blood glucose management. Mothers should maintain an HbA1c of 6.1% (or 43 mmol/mol) before and during pregnancy. Diet and exercise will initially be used to treat pregnant women who acquire gestational diabetes. However, if blood sugar levels continue to rise, they may also be prescribed oral hypoglycemics (tablets) or insulin injections.

• Effect of HBsAg on the maternal and fetal health risks: Here, the effect of HBsAg on the maternal and fetal health risks is depicted in Figure 10(n). In poorer nations, HBsAg positivity during pregnancy, a symptom of hepatitis B virus

infection, is a well-recognized problem. However, it is unclear if this condition is linked to poor maternal outcomes. GDM, PPH, intrahepatic cholestasis, and Caesarean section were more common in pregnant women who tested positive for HBsAg. HBsAg positivity during pregnancy was also linked to a higher risk of numerous poor maternal outcomes.

• Effect of Thyroid-Stimulating Hormone (TSH) on the maternal and fetal health risk: The effect of TSH on maternal and fetal health risk is depicted in Figure 10(o). TSH must remain within a healthy range for the thyroid gland to function effectively. An elevated TSH level indicates that the thyroid gland is not producing enough thyroxine, a condition known as hypothyroidism. TSH levels may fluctuate over the course of pregnancy, varying by trimester: the typical range is 0.6-3.4 mIU/L in the first trimester, 0.37-3.6 mIU/L in the second trimester, and 0.38-4.04 mIU/L in the third trimester. Iodine deficiency can lead to low thyroxine levels and a correspondingly elevated TSH level. Hashimoto's disease, an autoimmune condition in which the body produces antibodies that damage the thyroid gland and other tissues, is another recognised cause of elevated TSH. The precise trigger for Hashimoto's disease itself remains incompletely understood, although it is known to be autoimmune in origin.

• Effect of Serum SGPT/ALT on maternal and fetal health risk: Here, the effect of ALT on maternal and fetal health risks is depicted in Figure 10(p). The ALT enzymes are usually retained within the liver cells when the liver functions normally. However, when liver cells are injured or damaged, the ALT enzymes are released into the bloodstream, thereby increasing their levels. Therefore, we can assess the severity of liver damage by evaluating the level of ALT enzymes in the blood. A higher ALT value denotes more severe liver injury. Hepatitis A, B, or C, chronic viral hepatitis, cirrhosis of the liver (liver fibrosis caused by prolonged liver inflammation), alcohol-induced liver damage, hemochromatosis (a genetic condition caused by long-term liver damage), and decreased blood flow to the liver (from shock or heart disease) are the most common diseases that cause abnormally high Serum Glutamic-Oxaloacetic Transaminase (SGOT) and SGPT levels.

• Effect of Serum uric acid on maternal and fetal health risk: Here, the effect of serum uric acid on maternal and fetal health risks is depicted in Figure 10(q). Having high uric acid levels while pregnant can impact the baby's health. Preeclampsia can result from having high uric acid levels during the first trimester of pregnancy. Women with higher uric acid levels during pregnancy may also be more likely to develop gestational diabetes. This syndrome occurs in approximately 4% of pregnant women. If a woman has this problem, her body will be unable to produce or utilize the hormone insulin, which regulates blood sugar levels. A pregnant woman's baby's weight may be impaired if she has high levels of uric acid, which could result in fetal harm or even death. Diabetes, heart disease, and hypertension are a few of the long-term repercussions of this disorder.

• Effect of Fetal heart rate on maternal and fetal health risk: Here, the effect of fetal heart rate on maternal and fetal health risks is depicted in Figure 10(r). Fetal heart rate is one of the most significant parameters in the sensitivity analysis of maternal and fetal health risks. A woman's healthcare professional may monitor fetal cardiac during labor and late pregnancy. Fetal heart rates typically range from 110 to 160 beats per minute. The rate can range from 5 to 25 beats per minute. The fetal heart rate may fluctuate as the unborn child reacts to circumstances in her uterus. The baby may not be getting enough oxygen or experiencing other issues if the fetal heart rate is irregular.

• Effect of amniotic fluid on maternal and fetal health risk: The effect of amniotic fluid on maternal and fetal health risks is depicted in Figure 10(s). The amount of amniotic fluid is also one of the most significant parameters in the sensitivity analysis of maternal and fetal health risks. A fetus is cushioned by amniotic fluid, a clear to slightly yellowish liquid, inside the amniotic sac. Throughout a pregnancy, the unborn child floats in amniotic fluid. The fetus "inhales" or absorbs the amniotic fluid before expelling it through urination, causing the amniotic fluid to circulate continuously. A pregnant woman with oligohydramnios, also known as low amniotic fluid, has insufficient amniotic fluid. At around 32-36 weeks of gestation, oligohydramnios is suspected if the AFI reveals a fluid volume of less than 500 ml, a fluid thickness of less than 5 centimeters, or the lack of a fluid pocket that is 2-3 cm deep. Some factors contributing to low amniotic fluid levels include congenital disorders, membrane leakage or rupture, issues with the placenta, etc. Additionally, many people develop polyhydramnios, which is defined as an amniotic fluid volume of 2,000 ml or more, or an amniotic index of 20 cm or greater. The leading causes of polyhydramnios are multiple pregnancies or nervous system disorders in the fetus. Polyhydramnios is readily caused by pregnant women with diabetes, large fetuses, and faulty placentas. Due to the long labor and delivery process caused by polyhydramnios, fetal discomfort, and postpartum hemorrhage are frequently encountered. The chance of a rapid rupture of the membranes, which can result in an early birth, increases when the amount of amniotic fluid is above the recommended limit.

After carefully examining the parameters and evaluating the outcomes of all algorithms, the best model among the neural networks and Artificial Intelligence (AI) algorithms considered for this work is discussed in this section. Sensitivity analysis was also performed on the maternal and fetal health risks dataset to find which parameter had the most significant impact on risk.

4.7 *Limitations*

This study has several limitations that should be acknowledged. First, the dataset was collected from a single district (Noakhali) in Bangladesh, which limits the generalisability of the findings to other geographic regions, healthcare settings, and patient demographics. Second, risk labels were assigned using rule-based clinical thresholds; while this ensures clinical grounding, it means that the models are evaluated on their ability to recover the labelling rules rather than to predict independently adjudicated outcomes. Third, the relatively small sample size of 930 records—though adequate for the present comparative analysis—may not capture the full diversity of pregnancy risk presentations encountered in larger populations. Fourth, no external validation dataset was used; future studies should validate the proposed models prospectively on unseen patient cohorts. Fifth, interpretability tools such as SHapley Additive exPlanations (SHAP) or Local Interpretable Model-agnostic Explanations (LIME) were not applied; incorporating explainable AI methods in future work would enhance the clinical trustworthiness of the models.

4.8 *Clinical implications*

The sensitivity analysis indicates that blood pressure (systolic and diastolic), fetal movement, fetal heart rate, and amniotic fluid level have the greatest influence on maternal and fetal health risk, followed by BMI, fasting blood glucose, HbA1c, and maternal heart rate as secondary but still meaningful contributors. In clinical practice, this ranking suggests a practical basis for prioritizing antenatal monitoring resources. For patients already presenting with one or more elevated risk factors, closer or more frequent monitoring of blood pressure, fetal movement counts, fetal heart rate, and amniotic fluid level may allow earlier detection of a worsening risk trajectory than monitoring all nineteen parameters with equal frequency. Patients with borderline BMI, fasting glucose, HbA1c, or maternal heart rate values may similarly benefit from closer follow-up, given their secondary but still measurable contribution to overall risk. These findings could inform a simplified, priority-based monitoring checklist for gynecologists and antenatal care providers, particularly in resource-limited settings such as those in Noakhali District, where frequent comprehensive testing of all nineteen parameters may not always be feasible. Rather than replacing full clinical assessment, such a checklist could serve as a triage aid—flagging which patients warrant more urgent or more frequent follow-up based on the parameters shown here to carry the greatest predictive weight.

5. Conclusion

High-risk pregnancies can frequently result in serious physical and psychological harm to both the mother and the child, and can, in some cases, ultimately result in death. This study applied multiple supervised classification and neural network methodologies to improve early risk detection for maternal and fetal health. A real-world dataset of pregnant women from local hospitals and clinics was used for this purpose, and a sensitivity analysis was also conducted to determine which parameters most significantly increase the risk to the health of the mother and the fetus. This study used Multinomial Logistic Regression, Multinomial Naïve Bayes, K-Nearest Neighbors, Support Vector Machine, Decision Tree, Random Forest, and Multilayer Perceptron approaches. Analysis of the experimental findings for each algorithm indicated that the Decision Tree, Random Forest, and Multilayer Perceptron algorithms were the most effective, achieving 100% prediction accuracy. Given their superior performance, these three algorithms are the most promising candidates for early detection of maternal and fetal health risks. Preventable maternal and child morbidity remains a significant public health burden, particularly in settings where access to timely medical care is limited. Integrating advanced technology with the medical sector can help address this challenge, and this research contributes toward that goal. The algorithms developed in this study could be integrated into extended systems, such as a remote maternal and fetal healthcare monitoring platform, with cloud-based infrastructure used to support more realistic,

larger-scale deployment scenarios.

Funding

The authors receive no funding from any organization for this research.

Acknowledgement

The authors have nothing to acknowledge.

Ethics approval

Ethical approval for the study was obtained from the Institutional Review Committee of Noakhali Science and Technology University (reference: NSTU-IRC-2024-01).

Availability of data and materials

The data sets generated during the current study are available from the corresponding author upon reasonable request.

Code availability

The code written and utilized during the current study is available from the corresponding author upon reasonable request.

Consent to participate

All the authors have voluntarily agreed to participate in this research study.

Consent to publication

All the authors have given their consent for the publication of identifiable details, which can include image(s) and/or videos and/or case history and/or details within the text ("Material") to be published in this Article

Author's contribution

Zarin Tanzim: Conceptualization, investigation, data collection, data curation, model design, model training, draft manuscript preparation, writing-original draft. Md. Ashikur Rahman Khan: Supervision, conceptualization, model design, model training, analysis and interpretation of results, investigation on challenges, writing-review, and editing. Ishtiaq Ahammad: Draft manuscript preparation, investigation, analysis and interpretation of results, writing-review and editing. Abir Hosen: Investigation, analysis and interpretation of results, writing-review and editing. Md. Masudur Rahman, Md. Tasin Tazwar, and Nusrat Jahan: Data Collection, writing-review and editing.

Conflict of interest

The authors declare no conflict of interest.

References

[1] Blackburn S. *Maternal, Fetal, & Neonatal Physiology-E-Book: A Clinical Perspective*. Netherlands: Elsevier Health Sciences; 2017.

[2] Ahmed M, Kashem MA, Rahman M, Khatun S. Review and analysis of risk factor of maternal health in remote area using the Internet of Things (IoT). In: *InECCE2019*. Singapore: Springer; 2020. p.357-365. Available from: https://doi.org/10.1007/978-981-15-2317-5_30.

[3] Say L, Chou D, Gemmill A, Tunçalp Ö, Moller AB, Daniels J, et al. Global causes of maternal death: A WHO systematic analysis. *The Lancet Global Health*. 2014; 2(6): 323-333. Available from: https://doi.org/10.1016/S2214-109X(14)70227-X.

[4] World Health Organization. *Trends in maternal mortality 2000 to 2017: Estimates by WHO, UNICEF, UNFPA, World Bank Group and the United Nations Population Division*. Geneva, Switzerland: World Health Organization; 2019. Available from: https://documents1.worldbank.org/curated/en/793971568908763231/pdf/Trends-in-maternal-mortality-2000-to-2017-Estimates-by-WHO-UNICEF-UNFPA-World-Bank-Group-and-the-United-Nations-Population-Division.pdf [Accessed 10th January 2021].

[5] World Health Organization. *Maternal Deaths Decline Slowly with Vast Inequalities Worldwide*. Geneva, Switzerland: World Health Organization. Available from: https://www.who.int/news/item/19-09-2019-maternaldeaths-decline-slowly-with-vast-inequalities-worldwide [Accessed 10th January 2021].

[6] World Health Organization. *Maternal Mortality*. Available from: https://www.who.int/news-room/fact-sheets/detail/ maternal-mortality [Accessed 10th January 2021].

[7] Muglia LJ, Benhalima K, Tong S, Ozanne S. Maternal factors during pregnancy influencing maternal, fetal, and childhood outcomes. *BMC Medicine*. 2022; 20: 418. Available from: https://doi.org/10.1186/s12916-022-02632-6.

[8] Souza JP, Day LT, Rezende-Gomes AC, Zhang J, Mori R, Baguiya A, et al. A global analysis of the determinants of maternal health and transitions in maternal mortality. *The Lancet Global Health*. 2024; 12(2): e306-e316. Available from: https://doi.org/10.1016/S2214-109X(23)00468-0.

[9] Ijdi RE, Ducille CM, Singh K. Knowledge of maternal health complications: A critical analysis among pregnant women in Bangladesh. *PLOS Global Public Health*. 2025; 5(11): e0005469. Available from: https://doi.org/10.1371/journal.pgph.0005469.

[10] Rahman AHMM. A review on child and maternal health status of Bangladesh. *CHRISMED Journal of Health and Research*. 2018; 5(1): 1-7. Available from: https://doi.org/10.4103/cjhr.cjhr_65_17.

[11] Pallasmaa N, Ekblad U, Gissler M, Alanen A. The impact of maternal obesity, age, pre-eclampsia and insulin dependent diabetes on severe maternal morbidity by mode of delivery-a register-based cohort study. *Archives of Gynecology and Obstetrics*. 2015; 291(2): 311-318. Available from: https://doi.org/10.1007/s00404-014-3352-z.

[12] Agrawal P. Maternal mortality and morbidity in the United States of America. *Bulletin of the World Health Organization*. 2015; 93(3): 135. Available from: https://doi.org/10.2471/BLT.14.148627.

[13] Chen HY, Chuang CH, Yang YJ, Wu TP. Exploring the risk factors of preterm birth using data mining. *Expert Systems with Applications*. 2011; 38(5): 5384-5387. Available from: https://doi.org/10.1016/j.eswa.2010.10.017.

[14] Rawashdeh H, Awawdeh S, Shannag F, Henawi E, Faris H, Obeid N, et al. Intelligent system based on data mining techniques for prediction of preterm birth for women with cervical cerclage. *Computational Biology and Chemistry*. 2020; 85: 107233. Available from: https://doi.org/10.1016/j.compbiolchem.2020.107233.

[15] Fahim YA, Hasani IW, Kabba S, Ragab WM. Artificial intelligence in healthcare and medicine: Clinical applications, therapeutic advances, and future perspectives. *European Journal of Medical Research*. 2025; 30(1): 848. Available from: https://doi.org/10.1186/s40001-025-03196-w.

[16] Yarlapati AR, Dey SR, Saha S. Early prediction of LBW cases via minimum error rate classifier: A statistical machine learning approach. In: *2017 IEEE International Conference on Smart Computing (SMARTCOMP)*. Hong Kong, China: IEEE; 2017. p.1-6. Available from: https://doi.org/10.1109/SMARTCOMP.2017.7947002.

[17] Yu S, Tan KK, Sng BL, Li S, Sia ATH. Feature extraction and classification for ultrasound images of lumbar spine with support vector machine. In: *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. Chicago, IL, USA: IEEE; 2011. p.4659-4662. Available from: https://doi.org/10.1109/

EMBC.2014.6944663.
[18] Cheng W, Fang L, Yang L, Zhao H, Wang P, Yan J. Varying coefficient models for analyzing the effects of risk factors on pregnant women's blood pressure. In: *2014 13th International Conference on Machine Learning and Applications*. Detroit, MI, USA: IEEE; 2014. p.55-60. Available from: https://doi.org/10.1109/ICMLA.2014.14.
[19] Woolery LK, Grzymala-Busse J. Machine learning for an expert system to predict preterm birth risk. *Journal of the American Medical Informatics Association*. 1994; 1(6): 439-446. Available from: https://doi.org/10.1136/jamia.1994.95153433.
[20] Sahin H, Subasi A. Classification of the cardiotocogram data for anticipation of fetal risks using machine learning techniques. *Applied Soft Computing*. 2015; 33: 231-238. Available from: https://doi.org/10.1016/j.asoc.2015.04.038.
[21] Lakshmi BN, Indumathi TS, Ravi N. Prediction based health monitoring in pregnant women. In: *2015 International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT)*. Davangere, India: IEEE; 2015. p.594-598. Available from: https://doi.org/10.1109/ICATCCT.2015.7456954.
[22] Lakshmi BN, Indumathi TS, Ravi N. A comparative study of classification algorithms for predicting gestational risks in pregnant women. In: *2015 International Conference on Computers, Communications, and Systems (ICCCS)*. Kanyakumari, India: IEEE; 2015. p.42-46. Available from: https://doi.org/10.1109/CCOMS.2015.7562849.
[23] Sawant R, Gaikwad N. Hybrid prediction method for pregnancy data set. In: *2015 1st International Conference on Next Generation Computing Technologies (NGCT)*. Dehradun, India: IEEE; 2015. p.918-920. Available from: https://doi.org/10.1109/NGCT.2015.7375253.
[24] Kenny LC, Dunn WB, Ellis DI, Myers J, Baker PN, Kell DB. Novel biomarkers for preeclampsia detected using metabolomics and machine learning. *Metabolomics*. 2005; 1(3): 227-234. Available from: https://doi.org/10.1007/s11306-005-0003-1.
[25] Smyser CD, Dosenbach NUF, Smyser TA, Snyder AZ, Rogers CE, Inder TE, et al. Prediction of brain maturity in infants using machine-learning algorithms. *NeuroImage*. 2016; 136: 1-9. Available from: https://doi.org/10.1016/j.neuroimage.2016.05.029.
[26] Czabanski R, Jezewski J, Matonia A, Jezewski M. Computerized analysis of fetal heart rate signals as the predictor of neonatal acidemia. *Expert Systems with Applications*. 2012; 39(15): 11846-11860. Available from: https://doi.org/10.1016/j.eswa.2012.01.196.
[27] Ball G, Aljabar P, Arichi T, Tusor N, Cox D, Merchant N, et al. Machine-learning to characterise neonatal functional connectivity in the preterm brain. *Neuroimage*. 2016; 124: 267-275. Available from: https://doi.org/10.1016/j.neuroimage.2015.08.055.
[28] Krupa N, Zahedi E, Ahmed S, Hassan FM. Antepartum fetal heart rate feature extraction and classification using empirical mode decomposition and support vector machine. *Biomedical Engineering Online*. 2011; 10(1): 1-15. Available from: https://doi.org/10.1186/1475-925X-10-6.
[29] Ocak H. A medical decision support system based on support vector machines and the genetic algorithm for the evaluation of fetal well-being. *Journal of Medical Systems*. 2013; 37(2): 1-9. Available from: https://doi.org/10.1007/s10916-012-9913-4.
[30] Frank A. *UCI machine learning repository*. Available from: http://archive.ics.uci.edu/ml [Accessed 5th January 2020].
[31] Jadhav S, Nalbalwar S, Ghatol A. Modular neural network model based foetal state classification. In: *2011 IEEE International Conference on Bioinformatics and Biomedicine Workshops (BIBMW)*. Atlanta, GA, USA: IEEE; 2011. p.915-917.
[32] Huang M, Hsu Y. Fetal distress prediction using discriminant analysis, decision tree, and artificial neural network. *Journal of Biomedical Science and Engineering*. 2012; 5(9): 526-533. Available from: https://doi.org/10.4236/jbise.2012.59065.
[33] Mboya IB, Mahande MJ, Mohammed M, Obure J, Mwambi HG. Prediction of perinatal death using machine learning models: A birth registry-based cohort study in northern Tanzania. *BMJ Open*. 2020; 10(10): e040132. Available from: https://doi.org/10.1136/bmjopen-2020-040132.
[34] Abraham A, Le B, Kosti I, Straub P, Velez-Edwards DR, Davis LK, et al. Dense phenotyping from electronic health records enables machine-learning-based prediction of preterm birth. *BMC Medicine*. 2022; 20(1): 333. Available from: https://doi.org/10.1186/s12916-022-02522-x.
[35] Islam MN, Mahmud T, Khan NI, Mustafina SN, Islam AN. Exploring machine learning algorithms to find the best features for predicting modes of childbirth. *IEEE Access*. 2020; 9: 1680-1692. Available from: https://doi.org/10.1109/ACCESS.2020.3045469.

[36] Cuzzocrea A, Belmerabet I, Hafsaoui A, Leung CK. A machine-learning-based data science framework for effectively and efficiently processing, managing, and visualizing big sequential data. *Computers*. 2025; 14(7): 276. Available from: https://doi.org/10.3390/computers14070276.
[37] Noviandy TR, Nainggolan SI, Raihan R, Firmansyah I, Idroes R. Maternal health risk detection using light gradient boosting machine approach. *Infolitika Journal of Data Science*. 2023; 1(2): 48-54. Available from: https://doi.org/10.60084/ijds.v1i2.123.
[38] Al Mashrafi SS, Tafakori L, Abdollahian M. Predicting maternal risk level using machine learning models. *BMC Pregnancy and Childbirth*. 2024; 24: 820. Available from: https://doi.org/10.1186/s12884-024-07030-9.
[39] Mapari SA, Shrivastava D, Dave A, Bedi GN, Gupta A. Revolutionizing maternal health: The role of artificial intelligence in enhancing care and accessibility. *Cureus*. 2024; 16(9): e69555. Available from: https://doi.org/10.7759/cureus.69555.
[40] Khadidos AO, Saleem F, Selvarajan S, Ullah Z, Khadidos AO. Ensemble machine learning framework for predicting maternal health risk during pregnancy. *Scientific Reports*. 2024; 14: 21483. Available from: https://doi.org/10.1038/s41598-024-71934-x.
[41] American College of Obstetricians and Gynecologists' Committee on Practice Bulletins-Obstetrics. Gestational hypertension and preeclampsia: ACOG practice bulletin, number 222. *Obstetrics & Gynecology*. 2020; 135(6): e237-e260. Available from: https://doi.org/10.1097/AOG.0000000000003891.
[42] Ghosh N, MacLennan K. Physiology of pregnancy. *Anaesthesia & Intensive Care Medicine*. 2025; 26(4): 220-226. Available from: https://doi.org/10.1016/j.mpaic.2025.02.014.
[43] Brar R, Suri V, Suri V, Singh MP, Biswal M, Sikka P. Fever during pregnancy: Etiology and fetomaternal outcomes. *Journal of Obstetrics and Gynaecology of India*. 2022; 72(Suppl 1): 102-108. Available from: https://doi.org/10.1007/s13224-021-01562-2.
[44] Luef BM, Stentebjerg LL, Tanvig MH, Aalders J, Möller S, Ovesen PG, et al. Impact of maternal fasting blood glucose in pregnancy on body composition, anthropometric, and metabolic outcomes in newborns and 3-Year-Old offspring: Results from the lifestyle in pregnancy and offspring study. *BMC Pregnancy Childbirth.* 2026; 26: 193. Available from: https://doi.org/10.1186/s12884-026-08692-3.
[45] Kanwar JB, Manglunia A, Mangaraj S, Swain J, Sahoo A, Singh J, et al. Clinical and biochemical markers in early pregnancy for prediction of gestational diabetes mellitus. *Indian Journal of Endocrinology and Metabolism*. 2025; 29(1): 108-115. Available from: https://doi.org/10.4103/ijem.ijem_492_23.
[46] Abbassi-Ghanavati M, Greer LG, Cunningham FG. Pregnancy and laboratory studies: A reference table for clinicians. *Obstetrics & Gynecology*. 2009; 114(6): 1326-1331. Available from: https://doi.org/10.1097/AOG.0b013e3181c2bde8.
[47] Li Y, He Y, Wang Q, Wang X, Xie Y, Cheng L, et al. The association of maternal serum parameters with pregnancy induced hypertension in Chinese population. *PLoS One*. 2026; 21(2): e0343337. Available from: https://doi.org/10.1371/journal.pone.0343337.
[48] Wang X, Zhang S, Yu W, Li G, Li J, Ji J, et al. Pre-pregnancy body mass index and glycated-hemoglobin with the risk of metabolic diseases in gestational diabetes: A prospective cohort study. *Front Endocrinol (Lausanne)*. 2023; 14: 1238873. Available from: https://doi.org/10.3389/fendo.2023.1238873.
[49] Madaan S, Mendiratta SL, Jain PK, Mittal M. Aminotic fluid index and its correlation with fetal growth and perinatal outcome. *Journal of Fetal Medicine*. 2015; 2: 61-67. Available from: https://doi.org/10.1007/s40556-015-0049-8.
[50] Saltelli A, Aleksankina K, Becker W, Fennell P, Ferretti F, Holst N, et al. Why so many published sensitivity analyses are false: A systematic review of sensitivity analysis practices. *Environmental Modelling & Software*. 2019; 114: 29-39. Available from: https://doi.org/10.1016/j.envsoft.2019.01.012.
[51] Shaw R. *The 10 best machine learning algorithms for data science beginners*. Available from: https://www.dataquest.io/blog/top-10-machine-learning-algorithms-for-beginners [Accessed 5th January 2020].
[52] Long V, Mathieu S, Fiset C. Pregnancy-induced increased heart rate is independent of thyroid hormones. *Heart Rhythm O*^{2}. 2021; 2(2): 168-173, Available from: https://doi.org/10.1016/j.hroo.2021.03.001.
[53] Cosson E, Carbillon L, Valensi P. High fasting plasma glucose during early pregnancy: A review about early gestational diabetes mellitus. *Journal of Diabetes Research*. 2017; 2017: 8921712. Available from: https://doi.org/10.1155/2017/8921712.