Computational Journal of Mathematical and Statistical Sciences
3(1), 161–176
DOI:10.21608/CJMSS.2024.259490.1036
https://cjmss.journals.ekb.eg/

Comparative Assessment of Several Effective Machine Learning Classification Methods for Maternal Health Risk

Md Nurul Raihen[1,*], **Sultana Akter**[2]

[1] Department of Mathematics and Computer Science, Fontbonne University, Saint Louis, MO, 63105, USA.

[2] Department of Statistics, Western Michigan University, Kalamazoo, MI, 49006, USA.

* **Correspondence:** nraihen@fontbonne.edu

Abstract: By analyzing maternal age, heart rate, blood oxygen level, blood pressure, and body temperature, it has the potential to evaluate the risk complexity for certain patients. Early identification and classification of risk variables can successfully prevent pregnancy-related issues by reducing the number of errors. Maternal risk analysis can improve prenatal care, improve mother and baby health, and optimize healthcare resources by identifying misclassified observations using machine learning algorithms such as LDA, QDA, KNN, Decision Tree, Random Forest, Bagging, and Support Vector Machine, all of which have a significant impact on maternity health risk assessment. The split validation technique was applied, using 800 observations for training and 214 for testing. In addition, the most dependable model was determined using a 10-fold cross-validation technique. The suggested model outperforms all others in terms of accuracy and efficiency, with an accuracy score of **86.13%** for the support vector machine using a 10-fold cross validation technique. The purpose of this research is to use machine learning techniques to estimate the level of intensity of maternal health concerns by employing a classification strategy in the risk factor analysis.

Keywords: Classification, Accuracy, Validation, Health, Confusion matrix
Mathematics Subject Classification: 03C45, 55R80.
Received: 30 December 2023; Revised: 23 January 2024; Accepted: 28 January 2024; Online: 9 February 2024.

1. Introduction

During pregnancy, labor, and the postpartum period, maternal health is essential to a woman's overall health. Goal 3 of the United Nations Sustainable Development Agenda is to reduce preventable deaths among children and mothers (SDG 3). Each day, around 6,700 infants and 810 pregnant women lose their lives (WHO, 2019, 2020) [1, 2]. An overview of mortality in a given location can be provided via a variety of metrics, including death rates, death counts, and life expectancy, among others. The most common ways to measure excess mortality are by looking at death rates or death counts [3, 4, 5, 6].

A number of medical conditions, including advanced maternal age, blood disorders, and irregular heartbeat, might cause complications during pregnancy. It is possible to lessen the probability of complications during pregnancy by being aware of certain health risks. According to qualitative research on maternal health, the most important variables to consider while pregnant are age, heart rate, diastolic and systolic blood pressure, and total blood pressure. Considering these considerations, it is possible to minimize maternal and newborn mortality rates by the timely detection of dangers using machine learning algorithms, which in turn protects the health of the pregnant woman [7]. Advanced predictive analysis has the potential to revolutionize healthcare by reducing risk, increasing early disease diagnosis, and decreasing death rates [8]. This is because to the combination of data mining and sophisticated machine learning. Additionally, low-income nations struggle with conventional attitudes and practices, inadequate healthcare, and a lack of knowledge [9].

Variation in the anticipated maternal mortality level has been accounted for by researchers using a number of different death rates, such as crude death rates (CDRs), age-specific death rates, and age-standardized death dates (SDRs). Some of these characteristics are more prevalent in high-income nations than low-income ones, but other risk factors for serious pregnancy problems include medical misdiagnosis, a lack of coordination among providers, and racial/ethnic health disparities [10]. To avoid complications during pregnancy, such as premature birth or death, and to administer the necessary treatments, early detection of pregnancy-related risks is essential. Determining threats to maternal health is a crucial area for machine learning algorithms [11, 12]. Machine learning is the application of computational algorithms to massive datasets containing a myriad of complicated features in order to discover patterns within them. In [13] Worldwide, many pregnant women die from complications related to diseases that develop in their bodies during pregnancy, as reported by the UN Children's Fund and the World Health Organization [10].

Ignoring the dangers to the mother and child's life by significant blood loss, failure to progress (FTP) labor, abnormal fetal presentation, premature birth, and other adverse delivery conditions can lead to serious complications for the mother [14]. There are now more tools available to healthcare systems for the detection and treatment of pregnancy-related health concerns, thanks to the development of ML algorithms and deep learning algorithms. A realistic method of monitoring health conditions associated with pregnancy is crucial in developing nations, where healthcare is scarce in many parts. So, it's best to keep an eye on pregnant women at home. The goal of developing a maternal health monitoring system during pregnancy and delivery is to detect health problems early and intervene promptly; ML can help with this.

The use of models based on machine learning is thought to be effective in reducing maternal mortality rates as a result of complications arising from changes in risk factors [15].To mitigate the risk of maternity health complications for both the mother and child, various machine learning techniques such as Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), K-Nearest Neighbors (KNN), Decision Tree, Random Forest, Bootstrap Aggregating (Bagging), and Support Vector Machine (SVM) are employed to predict maternal health risks. Therefore, in this study, seven different machine learning approaches were applied. Upon comparing the findings produced from different methods, it was revealed that the Support Vector Machine was the most effective way for assessing maternal health risk. The proposed effective model exhibits the highest accuracy rate and efficiency of all, with the highest accuracy score of 86.13% for the support vector machine. Highest accuracy gives the less misclassification information which means that the support vector machine with highest accuracy score identifies the more accurate information of maternal risk.

The paper is divided into seven sections, and each of these sections provides a comprehensive summary of the analysis procedures that were applied. In the second section 2, a comprehensive explanation of the methodology and datasets that were utilized in the research is presented. In the third section 3, important analysis investigations and programming are presented. In the fourth section, the algorithms that were utilized for classification are broken down in detail, and discussion and results are presented. The method of analysis and the findings for forecasting hazards to maternal health based on physiological markers are described in Section 5, along with an explanation of the correctness of the results. The sixth and seventh sections demonstrate the areas in which our research and the potential applications of it are restricted. The fundamental objective of the research is to investigate the ways in which standardization influences the accuracy of risk projections throughout the course of pregnancy. The goal of this article is to establish which of the seven classification methods that were considered for the article is the most effective. The findings demonstrate how much more accurate risk levels may be estimated by employing standardized parameters.

2. Methodology

2.1. Dataset

This project is about classification problems and the data comes from the UCI Machine Learning repository (Maternal Health Risk - UCI Machine Learning Repository). Daffodil International University, Dhaka, Bangladesh is the original source of this dataset. Original data has been collected from different hospitals, community clinics, maternal health cares from the rural areas of Bangladesh through the IoT (Internet of Things) based risk monitoring system in 2020 [9].

The dataset is part of the collection's 'life' category in the UCI Machine Learning repository. The dataset has 1014 observations with 7 variables. Table 1 shows there are 6 quantitative predictors of the dataset (Age, Systolic BP, Diastolic BP, BS, BodyTem, HeartRate) and one qualitative variable (RiskLevel). RiskLevel has been considered with three levels and collected as low risk, mid risk, and high risk where 406 was classified as low-risk level, 336 in mid and 272 was in high-risk level of total observation. There is no missing value in this maternal risk data set.

Table 2 shows the statistical analysis of the age, systolic BP, diastolic BP, blood sugar, body temperature, and heart rate variables.

By Figure 1, Flowchart (a) illustrates the predictor variable from maternal risk data, as well as the explanatory factors of age, body temperature, blood sugar, age, heart rate, systolic BP, diastolic BP, and BS. All variables are quantitative.

Figure 2 shows that the age between **10-70** year has the highest risk level of maternity is low risk, so women within this age range are obtaining in low risk of their pregnancy, and post pregnancy periods. Body temperature has the equal risk for the women. A pregnant Woman with the range of Blood sugar **6-19** has the high level of maternal risk, and all others also show same.

Figure 3 represents that the low risk level has the highest frequency, which is **406**, and high-risk

Table 1. Variable Explanation of maternal risk data

VariableName	Variable Description	Type
Age	Any ages in years when a woman during pregnant	Numeric
SystolicBP	Upper value of Blood Pressure in mmHg	Numeric
DiastolicBP	Lower value of Blood Pressure in mmHg	Numeric
BS	Blood glucose levels are in terms of molar concentration.	Numeric
BodyTemp	Body Temperature	Numeric
HeartRate	A normal resting heart rate	Numeric
RiskLevel	(**Target variable**) Predicted Risk Intensity Level during pregnancy considering the previous attributes	Category(low=0, mid=1, High=2)

Table 2. Description and statistical distribution of the dataset features

	Age	Systolic BP	Diastolic BP	Blood Sugar	Body Temperature	Heart Rate
Count	1014	1014	1014	1014	1014	1014
Mean	29.8718	13.1982	76.4605	8.7259	98.6650	74.3017
Std	13.4744	18.4039	13.8858	3.2935	1.3714	8.0887
Min	10.0000	70.0000	49.0000	6.0000	98.0000	7.0000
Median	26.0000	120.0000	80.0000	8.0000	98.0000	76.0000
max	70.0000	160.0000	100.0000	19.0000	103.0000	90.0000

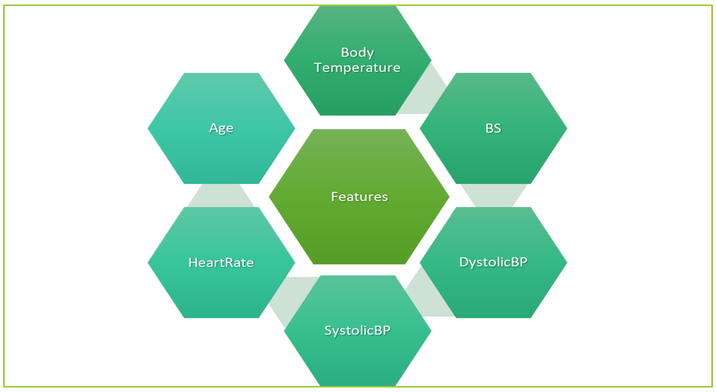
level the lowest frequency of **272** for maternity risk of women.

2.2. Analysis:

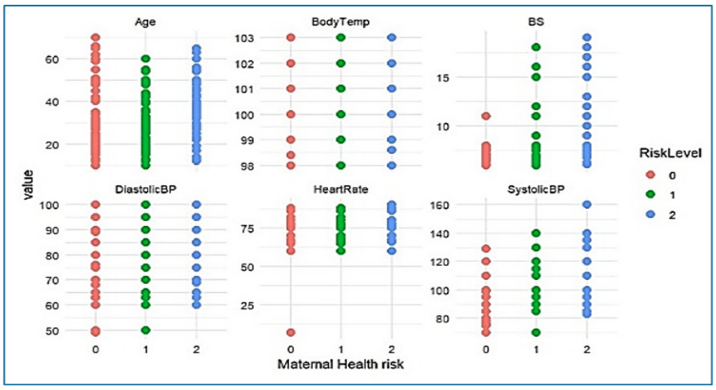
Figure 4 is Flowchart a presentation of the whole process of analysis plan in briefly here.

3. Description of analysis process

1. **Tool**: R programming software.
2. **Data preprocessing**: Checked the missing value and outliers. Luckily, no missing value, and no extreme outlier in this maternal risk data set.
3. Validation Approach: Split validation and Cross validation
4. ***Validation (split train/test):*** Validation of statistical models entails testing the accuracy of a chosen statistical model. The purpose of model validation is to assess how well a trained model performs on a different testing dataset once training has finished. The dataset utilized for training and testing could be different, or it could be a subset of the same dataset.
5. A training data set is a collection of instances that is used throughout the developing process to adjust the parameters.
6. Data that is distinct from the training data set but follows the same distribution as the training data set is called a test data set.



(a) Flowchart: Features of maternal risk data



(b) Flowchart: Frequency representation of each factor of RiskLevel for all predictors.

Figure 1. displays the highest and lowest frequency of maternal risk levels according on age, systolic and diastolic blood pressure, heart rate, and blood pressure. The colors red, green, and blue indicate the highest, middle, and lowest maternal risk levels, respectively.

For this analysis, Train data contains **800** observations, and test data contains **214** observations.

1. ***Cross validation***: It is a sampling technique that includes excluding certain portions of the data during the fitting process. This allows us to assess how well the model predicts the excluded data points, determining whether they are in close proximity or significantly deviate from the predicted values. There are several types of cross validation approaches. But here we used k-Fold cross validation process.
2. K-folds cross validation: Predictive models are evaluated using K-fold cross-validation. K subsets are formed by folding the dataset. Every iteration of training and testing uses a different validation fold. The model's generalizability is estimated by averaging the performance metrics of each fold.

To evaluate all classification models utilized 10-folds cross validation, Figure 5 is a representation of a 10-folds cross validation process.

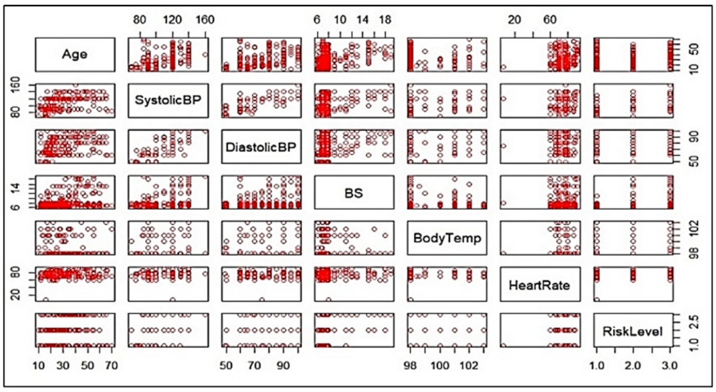


Figure 2. Correlation Plot of Maternal Risk Data

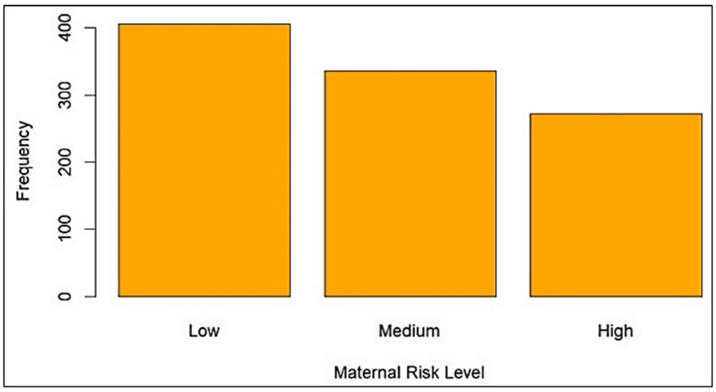


Figure 3. Frequency of each category of RiskLevel

3.1. Classification models

1. ***Linear Discriminant Analysis (LDA)***: This study examined possible risks to mothers' health and used it as a machine learning tool for analyzing datasets. Because it effectively reduces the number of dimensions while retaining the capacity to distinguish classes apart, LDA is a popular classification approach. To use this technique, one must first choose the optimal linear combination of features for classifying occurrences into many categories. In this case, LDA was run with three distinct labels for the target variable (RiskLevel). This research used LDA on the training dataset after fine-tuning its parameters using validation and cross-validation. The simplicity, efficiency, and capacity to disclose classification aspects of LDA make it a powerful tool for dealing with multi-class challenges.
2. ***Quadratic Linear Analysis (QDA)***: It is used to assess maternal health risks. QDA is an LDA variant that separates classes non-linearly. QDA believes each class has its own covariance matrix, unlike LDA. This makes QDA more versatile for datasets without a shared covariance matrix. This study used QDA and cross-validation to optimize parameters. QDA was used in the study because it can simulate more complex feature-class connections and handle datasets with varying covariance patterns. A bigger sample size is needed to estimate covariance matrices accurately.
3. ***K-Nearest Neighbors (KNN)***: This algorithm was used to analyze the dataset in comprehensive maternal health risk research. A simple but powerful classification method used in data mining and machine learning is KNN. Here I utilized k=5 nearest neighbors. This idea is that similar

Figure 4. Decision Making of Maternal Risk Data Analysis

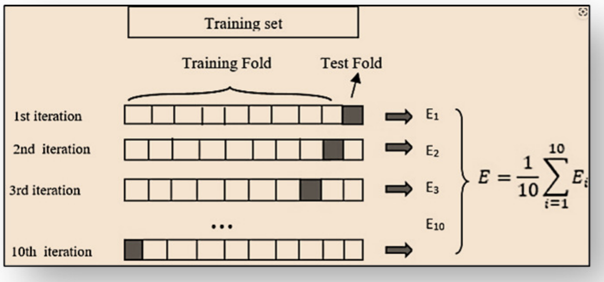
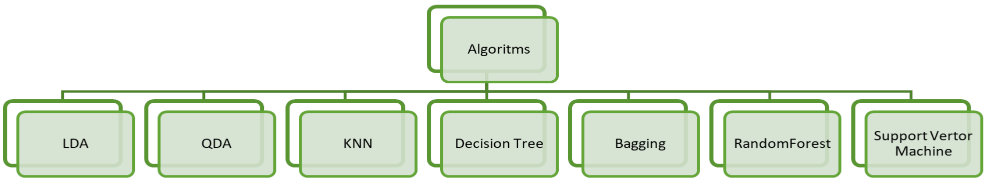


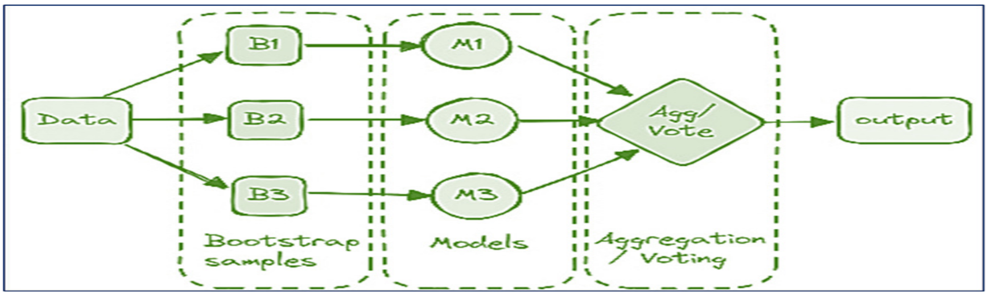
Figure 5. 10 folds cross validation

data points are likely to be close in feature space. This method classifies a fresh sample by the majority class of its 'k' nearest neighbors in the training dataset. This work optimized the critical parameter 'k' using cross-validation. However, the choice of 'k' and the distance measure can affect its performance, and computing distances to all training samples can be computationally costly for large datasets. Here I performed validation and cross validation approaches to fit the KNN classification model.

4. ***Decision Tree***: The decision tree (DT) is a supervised learning method for regression and classification that does not use parameters. The goal is to construct a model capable of predicting the value of a target variable using fundamental decision rules inferred from the characteristics of the data. The fact that they resemble trees is the inspiration for their name. When classifying data, they start at the root of the tree and work their way up through the various branches, each of which represents a potential outcome, until they reach the leaf node, at which point they provide the final binary result.
5. ***Bagging***: It is also known as bootstrap aggregating, which is an ensemble technique that entails training many models separately on randomly selected subsets of the data. The predictions of these models are then combined through voting or averaging. To forecast classification, we record the predicted class from each tree and conduct a majority vote. Total prediction is the most common class. Below provides all steps of bagging in **Figure 7**.

   When using bagging classification trees, we may calculate the cumulative reduction in the Gini





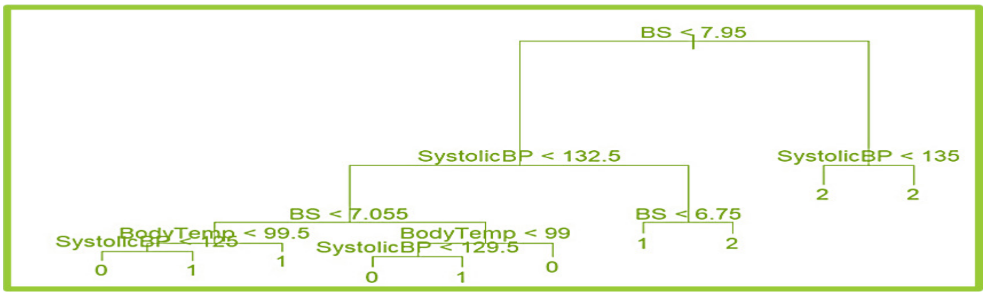


Figure 6. Decision Tree of maternal risk analysis

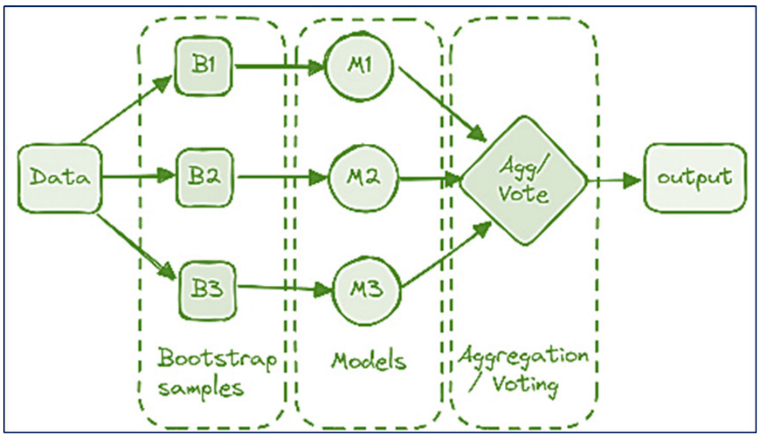


Figure 7. Structure of Bagging method

index caused by splits on a certain predictor. The variables that possess high values are crucial variables.

	MeanDecreaseGini
Age	72.34083
SystolicBP	97.25349
DiastolicBP	33.91138
BS	187.56695
BodyTemp	30.60406
HeartRate	43.58682

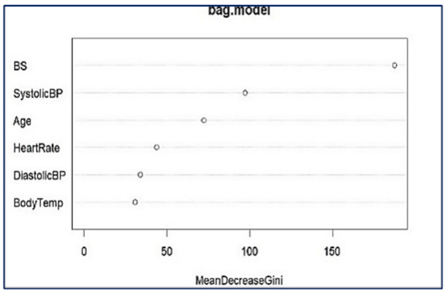


Figure 8. Predicted of Best predictor

In this analysis, the bagging shows in **Figure 8** that BS (Blood sugar) is the most influential predictor then systolicBP then others for maternity risk.

Random Forest: An adaptation of the bagging technique, the random forest algorithm uses a combination of bagging and feature randomness to produce an unconnected network of decision trees. In random forests, only a portion of those traits are chosen. It is standard practice to choose a number of predictors equal to the square root of the total number of predictors when evaluating a split in a tree, as this results in a new set of predictors.

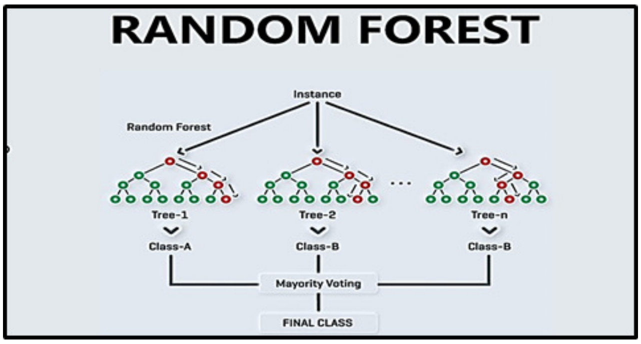


Figure 9. Sample Figure of Random Forest Model

Figure 9 shows that for random forest model we picked subset predictors 3 among of total 6 predictors.

6. ***Support vector Machine:*** One of the most popular machine learning approaches used by data scientists is SVM. SVM is powerful, easy to explain, and versatile. SVM distinguishes classes with a decision boundary. Basic SVM lacks support for multiclass classification. Divide data points into two groups and utilize binary categorization. After dividing the multiclassification problem into binary classification problems, the same idea is employed. Map data points to a high-dimensional space so every two classes are linearly separated. This "One-to-One approach." divides the multiclass problem into binary classification questions. For each pair of classes, binary classification exists. One-to-Rest is another option. That approach splits each group into a binary classifier. This study examines two important SVM parameters, C (cost) and gamma.
7. **C(cost):** The C parameter in SVM regularization parameter regulates the trade-off between low training and testing errors. A smaller C generates a smoother decision boundary (less complex

model) with more training data misclassifications. A bigger C allows the machine to choose more complex decision limits to categorize all training instances accurately.

8. **Gamma**: This parameter indicates how far a single training example influences the feature space, with low values representing 'far' and high values close. Gamma controls the decision boundary shape in RBF kernel SVM.A small gamma value slows decision boundary variation, making it linear. However, a big gamma value bends the decision boundary to match the data, potentially collecting more data nuances but risks overfitting. Here, for analysis for maternal risk, utilizes parameters cost = c (0.1, 1, 10, 100, 1000), gamma = c (0.5, 1, 2, 3, 4).

3.2. Criteria of Chosen the Best Model:

There are many ways which we can use to evaluate our fitted model:

1. *Accuracy Score/Error rate*
2. *Null Accuracy*
3. *Precision*
4. *Recall*
5. *f1 score*
6. *ROC — AUC*

Here, we used accuracy scores for each classification model, and figured out the highest accuracy score (which means *of correct predictions). In order to determine the diagonal elements of this matrix, we used the accuracy score (accurate predictions) that the confusion matrix provided. To explain how well a classification model performs on a set of known test data, confusion matrices are frequently employed. To construct the confusion matrix, four variables are used. There are four potential outcomes: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). Displayed in* Figure 10 is the format of the 3*3 confusion matrix.

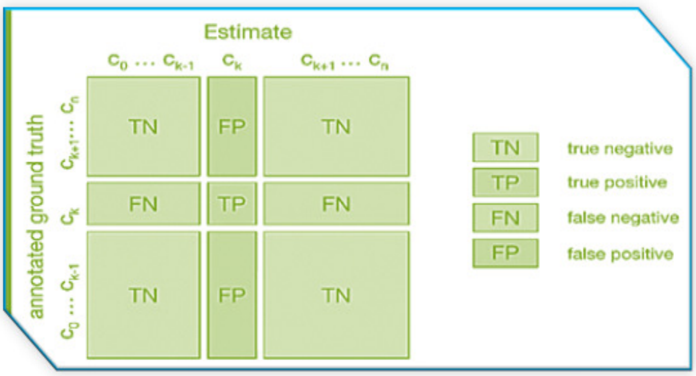


Figure 10. Confusion Matrix

The confusion matrix was used to gather crucial metrics for an appropriate evaluation of the machine learning classifiers. Other measures like True Positive Rate (TPR), False Positive Rate (FPR), and Accuracy Rate (AR) are also considered.

The formula to calculate the accuracy score:

$$\textbf{Accuracy} = \frac{(TN+TP+TN)}{(TN+FP+TN+FN+TP+FN+TN+FP+TN)}$$

*And **error rate** =1- Accuracy Score*

4. Result and Discussion

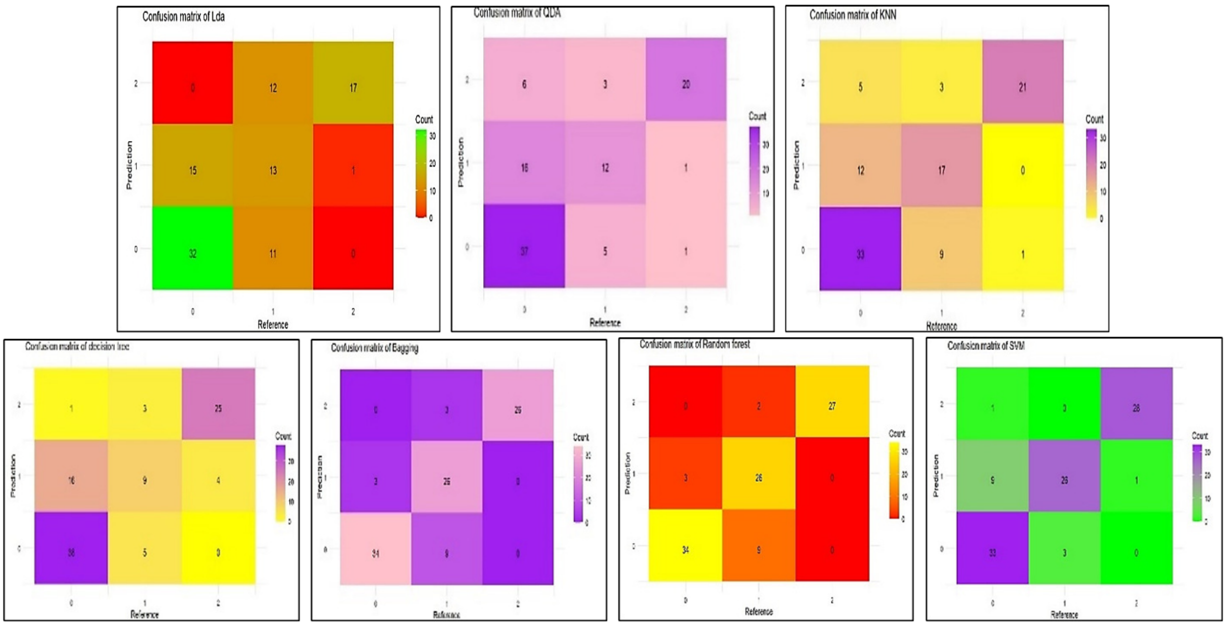


Figure 11. Confusion matrix table of all classification models.

In **Figure 11**, plots of confusion matrix table for all classification models which diagonal elements are indication of the accurate prediction, and non-diagonal elements observe the misclassification observation.

Table 3 displays the classification models together with their corresponding accuracy scores, illustrating the use of the split validation approach. The split validation approach demonstrates that the Bagging model achieves the highest accuracy score of **85.98%**. In this experiment, 800 observations were used as training data and **214** observations were used as test data. All six predictors, namely age, systolic BP, diastolic BP, blood sugar, body temperature, and heart rate, were utilized as parameters for the Bagging model.

From, the Table 3, and Figure 12 determine that to utilize split validation approach the Bagging model has the most appropriate information about the risk of maternity of women.

Table 4 and **Figure 13** implement all results of classification models with their accuracy score. It presents that Support vector machine contains the highest accuracy score of 86.13%, and the lowest accuracy score contains for Linear Discriminant Analysis model.

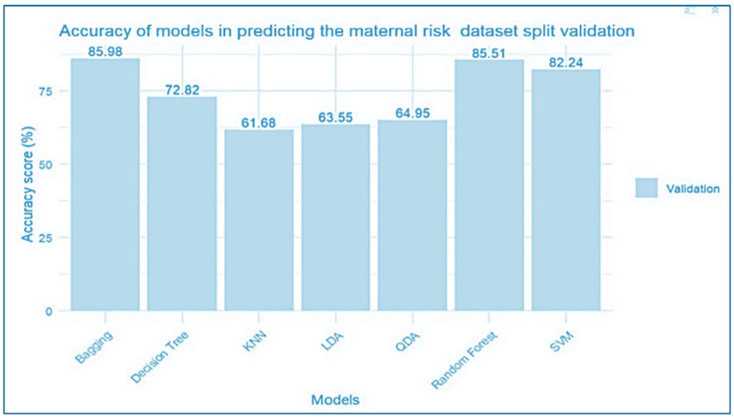


Figure 12. Comparable plots of accuracy score for validation split.

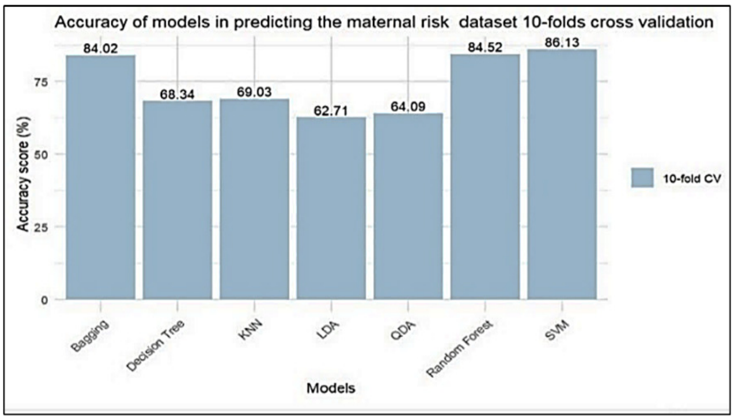


Figure 13. Comparable plots of accuracy score for 10 folds CV.

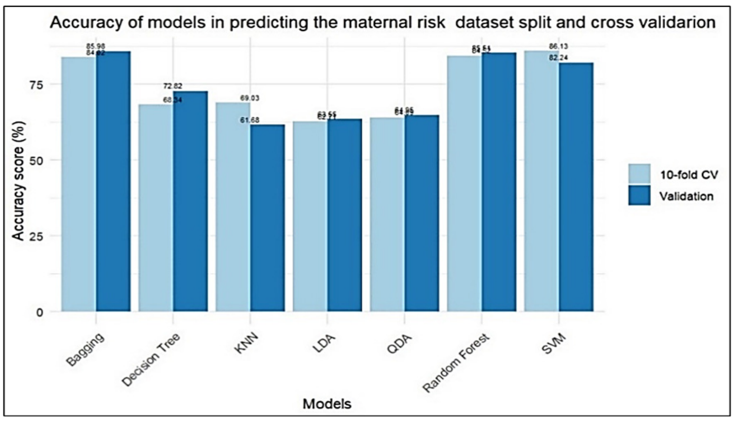


Figure 14. plot of summary results to compare the accuracy score.

Table 3. Result of split validation approach

Result: Validation Approach (Split test and train set)		
Model	Error rate	Accuracy rate
LDA	0.3645	0.6355
QDA	0.3505	0.6495
KNN(k=5)	0.3832	0.6168
Decision Tree	0.2711	0.7289
Random Forest	0.1449	0.8551
Bagging	0.1402	0.8598
Support Vector Machine	0.1776	0.8224

Table 4. Result of 10 folds cross validation approach

Result: Cross Validation Approach (10 folds)		
Model	Error rate	Accuracy rate
LDA	0.3729	0.6271
QDA	0.3591	0.64098
KNN(k=5)	0.3097	0.6903
Decision Tree	0.3166	0.6834
RandomForest	0.1548	0.8452
Bagging	0.1598	0.8402
Support Vector Machine	0.1387	0.8613

Presented in this summary **Table 5** is a comparison of all of the outcomes with their respective accuracy scores for each and every model that was applied in this research. When the Support Vector Machine (SVM) utilized the 10 folds cross validation strategy rather than the split validation method, it is evident that it achieved the highest accuracy score of 86.13% among all of the methods.

Figure 14 depicts the summary results of comparing several models using validation and cross-validation processes. The purpose is to select the appropriate test and train datasets to identify the most accurate model for assessing maternal risk in this study.

Table 6 displays the results of the final selected model, which achieved an accuracy score of **86.13%** using a 10-fold cross-validation strategy. The model was trained using several combinations of cost parameters (0.1, 1, 10, 100, 1000) and gamma parameters (0.5, 1, 2, 3, 4). We employed 10-fold cross-validation to assess our optimal Support Vector Machine model, which had a cost of 1000. It implies that support vector machine gives the most accurate result to analysis of maternal risk of women.

5. Conclusion

Through the enhancement of diagnostic accuracy, the reduction of physician effort, the reduction of expenses, and the provision of comparative analysis for tests that display considerable variability

Table 5. summary table of all models

Methods	Split validation Approach		10 folds cross validation	
Model	Error rate	Accuracy rate	Error rate	Accuracy rate
LDA	0.3645	0.6355	0.3729	0.6271
QDA	0.3505	0.6495	0.3591	0.6409
KNN(k=5)	0.3832	0.6168	0.3097	0.6903
Decision Tree	0.2711	0.7289	0.3166	0.6834
RandomForest	0.1449	0.8551	0.1548	0.8452
Bagging	0.1402	0.8598	0.1598	0.8402
Support Vector Machine	0.1776	0.8224	0.1387	0.8613

Table 6. Best fitted model

Model	Support Vector Machine
Parameters	cost = c(0.1, 1, 10, 100, 1000), gamma = c(0.5, 1, 2, 3, 4) kernel = radial best cost=1000
Validation Approach	10 folds cross validation approach
Accuracy rate	**86.13%**
Error Rate	13.87%

in interpretation between specialists, machine learning has the potential to revolutionize the field of healthcare. In conclusion, the findings of my research led to the creation of a model known as the Support Vector Machine, which represents a classification system for the hazards that are associated with maternal health. In comparison to earlier models, this one achieved a high level of accuracy and performed in a manner that was comparable. The goal of this research was to provide support with decision-making in health management, and it did so by conducting an analysis of the maternal problem scenario for mortality prediction. Due to the fact that the Support Vector Machine algorithm has the highest accuracy score, it is recommended to use the **10-folds** cross validation process in order to select the best cost as 1000 for the purpose of evaluating this model in the context of maternal health risk factor analysis.

Medical history, genetic information, lifestyle variables, and other factors are common components of maternity-related data. This high-dimensional data may be handled using SVM, allowing for the simultaneous consideration of multiple parameters. SVM can forecast future pregnancy problems by examining older data. With this kind of prediction power, early interventions can be made, which in turn improves the health of the mother and child. The unique requirements of maternity care can be met by customizing SVM. Targeted insights can be obtained, for instance, by adjusting it to concentrate on certain issues such as gestational diabetes or pre-eclampsia. Having a model that can effectively adapt to new, unknown data is of the utmost importance in medical applications. Because SVM is resistant to overfitting, it may be trusted to provide predictions or classifications that accurately represent the actual medical situation, not the peculiarities of the training data.

Medical professionals can better prepare for and deal with high-risk pregnancies if they are able to anticipate them in advance. Clinics and hospitals can improve resource allocation, such as concentrating on high-risk cases, with the help of SVM-based forecasts. Individual risk factors for women and children can be identified by SVM, which can contribute to tailored healthcare programs. So, the **SVM** method obtains the most reliable information about maternity related difficulties for the mother and child.

6. Future Application

1. By utilizing this comprehensive approach, it is possible to develop more comprehensive care plans, so benefiting the overall health of both the mother and the fetus.
2. This approach promotes particular patient care and future public health efforts, research, and policymaking.
3. This can improve health care through improving diagnosis by dropping their wrong information.
4. We need to gather more data and see if our approach works on different demographics in the future. In addition to the structured variables utilized in this study, future research can also make use of unstructured data like text or photos. Because of this, optimization is possible, and the risk of adverse pregnancy outcomes may be properly classified.

7. Limitations

1. Due to the fact that maternal data has a direct impact on patient outcomes, so model selection and validation must be more precise.
2. Additional comprehensive data is required in order to make more targeted and practical judgments regarding the level of maternal risk.
3. Establishing the hyperparameter tuning is challenging.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.
Funding: No Funding.
Author contributions: All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

References

1. WHO. (2019). Maternal Mortality. Geneva, Switzerland: World Health Organization.
2. WHO. (2020). Newborns: Improving Survival and Wellbeing. Geneva, Switzerland: World Health Organization.
3. Bilinski, A., & Emanuel, E. J. (2020). COVID-19 and excess all-cause mortality in the US and 18 comparison countries. Jama, 324(20), 2100-2102.

4. Faust, J. S., Krumholz, H. M., Du, C., Mayes, K. D., Lin, Z., Gilman, C., & Walensky, R. P. (2021). All-cause excess mortality and COVID-19–related mortality among US adults aged 25-44 years, March-July 2020. Jama, 325(8), 785-787.

5. Scholey, J. (2021). Robustness and bias of European excess death estimates in 2020 under varying model specifications. MedRxiv, 2021-06.

6. Raihen, M. N., Akter, S., Tabassum, F., Jahan, F., & Begum, S. (2023). A Statistical Analysis of Excess Mortality Mean at Covid-19 in 2020-2021. *Computational Journal of Mathematical and Statistical Sciences*, *2*(2), 223-239.

7. Ahmed, M., & Kashem, M. A. (2020, December). IoT based risk level prediction model for maternal health care in the context of Bangladesh. In *2020 2nd International Conference on Sustainable Technologies for Industry 4.0 (STI)* (pp. 1-6). IEEE.

8. Raihen, M. N., & Akter, S. (2023). Forecasting Breast Cancer: A Study of Classifying Patients' Post-Surgical Survival Rates with Breast Cancer. *Journal of Mathematics and Statistics Studies*, *4*(2), 70-78.

9. Washio, Y., & Humphreys, M. (2018). Maternal behavioral health: Fertile ground for behavior analysis. *Perspectives on Behavior Science*, *41*(2), 637-652.

10. Pawar, L., Malhotra, J., Sharma, A., Arora, D., & Vaidya, D. (2022, August). A Robust Machine Learning Predictive Model for Maternal Health Risk. In *2022 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC)* (pp. 882-888). IEEE.

11. Pawar, L., Malhotra, J., Sharma, A., Arora, D., & Vaidya, D. (2022, August). A Robust Machine Learning Predictive Model for Maternal Health Risk. In *2022 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC)* (pp. 882-888). IEEE.

12. Varshavsky, J., Smith, A., Wang, A., Hom, E., Izano, M., Huang, H., ... & Woodruff, T. J. (2020). Heightened susceptibility: A review of how pregnancy and chemical exposures influence maternal health. *Reproductive toxicology*, *92*, 14-56.

13. Raihen, M. N., & Akter, S. (2024). Prediction modeling using deep learning for the classification of grape-type dried fruits. *International Journal of Mathematics and Computer in Engineering*.

14. Easter, S. R., Bateman, B. T., Sweeney, V. H., Manganaro, K., Lassey, S. C., Gagne, J. J., & Robinson, J. N. (2019). A comorbidity-based screening tool to predict severe maternal morbidity at the time of delivery. *American Journal of Obstetrics and Gynecology*, *221*(3), 271-e1.

15. MUTLU, H. B., DURMAZ, F., YÜCEL, N., CENGİL, E., & YILDIRIM, M. (2023). Prediction of Maternal Health Risk with Traditional Machine Learning Methods. *NATURENGS*, *4*(1), 16-23.