Proc. of the International Conference on Computing and Advances in Information Technology (ICCAIT 2023)
21-23 November 2023, Ahmadu Bello University, Zaria, Nigeria

Machine Learning in Predicting Maternal Health Risks: A Comparative Analysis of Some Algorithms

Andra Umoru
Department of Computer Science,
University of Maiduguri,
Maiduguri, Nigeria.
andraumoru@unimaid.edu.ng

Iliyas Ibrahim Iliyas
Department of Computer Science,
University of Maiduguri,
Maiduguri, Nigeria.
iliyasibrahimiliyas@unimaid.edu.ng

Maryam Daniel Mshelbila
Department of Microbiology
Modibbo Adama University,
Yola, Nigeria.
mmshelbila@gmail.com

Patience Iliya Bwanu
Department of Mathematics,
Federal University Kashere
Kashere, Nigeria.
iliyapatiencebwanu@gmail.com

Abstract*— Recently, Machine learning has found applicability in almost all fields of human endeavours. Healthcare delivery is one of such fields, and more specifically, the field of Maternal Health Risks. This study aims to evaluate and compare the performance of three (3) supervised machine learning algorithms (Logistic regression, Support Vector Machine, and K-Nearest Neighbour) in predicting maternal health risks. Early detection of maternal health risks is crucial in today's world of healthcare delivery. This will help in mitigating the maternal death rate. The dataset was obtained from the UCI machine learning repository. The data set has seven (7) features: they are Age, Systolic Blood Pressure as SystolicBP, Diastolic BP as DiastolicBP, Blood Sugar as BS, Body Temperature as BodyTemp, HeartRate, and RiskLevel (the label). The performance of the selected algorithms was measured with respect to the following metrics: accuracy, precision, recall, F1 scores, and the AUC. The results revealed that the K-NN model recorded the highest accuracy of 78%, precision: 78%, recall: 78%, F1 score: 78%, and 0.93 AUC score, respectively; the SVM also recorded 76% accuracy, precision: 77%, recall: 76%, F1 score: 76%, and 0.88 AUC score, whereas the logistic regression recorded 60% accuracy, 62% precision, 59% recall, 60% F1 score, and 0.74 as AUC score, respectively. Although these algorithms did not achieve state-of-the-art accuracy performance, the study underscores the potential of Machine Learning in Healthcare delivery, especially in the domain of Maternal health.

Keywords*— Machine Learning, Metrics, Classification, Maternal Health Risk, Comparative, Healthcare, Morbidity and Mortality.

I. INTRODUCTION

As a scientific field of study, Machine Learning (ML) relies on training computers to learn data patterns [1][2]; by using these data, the machine makes decisions and forecasts the future [3]. In recent times, advanced methods like Machine Learning and Deep Learning (DL) have found much applicability in healthcare in identifying patterns within complex datasets and predicting patient outcomes with a significant level of accuracy [4]. These methods are increasingly used in maternal health risk to predict future risks and results [1]. ML and DL models have been applied in predicting different healthcare areas, including stroke and neurovascular medicine [5], reproductive medicine [6], medical imaging of the liver [7], radiology [8], and dentistry [9]. Different authors have also obtained results with significant accuracy in some comparative analysis of some machine learning algorithms; [4] for instance, they compared the performance of five supervised machine learning algorithms, namely Support Vector Machine (SVM), K-Nearest Neighbors, Random Forests, Artificial Neural Networks (ANNs) and Logistic Regression. They obtained results as SVM received accuracy, precision, and F1 scores of 97.14%, 95.65%, and 0.9777, respectively, whereas ANNs obtained the highest scores of 98.57%, 97.82%, and 0.9890, respectively.

Maternal health risks refer to the conditions or factors that potentially endanger women's health during pregnancy, childbirth, and the postpartum period [10][11]. WHO argued that to ensure that women and their unborn children achieve their maximum potential for health and well-being, each stage should carry some positivism [10]. These dangers run the risk of increasing maternal morbidity and mortality. Maternal health risks are influenced by some variables, including age, preexisting medical illnesses, complications during pregnancy, access to prenatal care, socioeconomic level, and lifestyle choices [11]. In the past 20 years, significant progress has been achieved. Despite the progress recorded, 287,000 women died during or shortly after giving birth in 2020. This number is on the high side. This mortality rate can be mitigated or, at best, prevented through prompt intervention by qualified health practitioners working in a supportive atmosphere [10]. We believe that the efforts of these qualified health practitioners can be complemented and supported using some advanced technological tools and methods like Machine Learning.

Machine Learning has shown significant potential in healthcare, especially in maternal health risk predictions. Various factors, from age to socioeconomic status, influence these health risks, leading to maternal morbidity and mortality. The vast data available in healthcare allows machine learning to offer priceless insights and recommendations, revolutionizing patient care. In maternal health, these algorithms have predicted events like preterm births and improved treatment decisions. This study aims to harness this potential to identify optimal algorithms for predicting maternal health risks.

This study examines four machine learning algorithms to predict maternal health risks using a selected dataset obtained from the UCI Machine Learning Repository. We aim to identify the most accurate and efficient algorithm for predicting maternal health risks. This can go a long way in contributing to the body of knowledge in both the field of machine learning and healthcare, more specifically, maternal health. These four machine learning algorithms achieved relatively significant accuracy in predicting the three classes of maternal health risks. Furthermore, this study is segmented into the following sections: review of related works, materials, and methods. The following section discusses the theoretical concept of the four selected algorithms. Also, the metrics for performance measurement are described, the experimental setup, results obtained and their analysis, and finally, the conclusion and area of further study were captured in the last section.

II. Background of the Study

A. Maternal Health and Risk Factors

Maternal health continues to demand attention in the field of public health due to the global prevalence of maternal morbidity and mortality rates [10]. Researchers have explored numerous facets of this problem, particularly socioeconomic issues, physical health hazards, and mental health difficulties for pregnant, preterm birth, and postpartum women [10][12]. Maternal health outcomes are highly influenced by socioeconomic status. In India, a study by [13] found an inverse association between socioeconomic class and the rate of maternal death, highlighting the necessity for policies to close economic gaps. Similar findings were made by [14], who found that American women with poor socioeconomic status were more likely to have severe obstetric complications. This data points to the importance of socioeconomic factors in determining outcomes for maternal health.

Gestational diabetes mellitus (GDM), which poses physical health hazards, has been designated as a major issue for maternal health. It's linked to long-term health hazards for both mother and child and unfavourable pregnancy outcomes [15]. Preeclampsia and other hypertension problems have also been identified as major contributors to maternal and fetal morbidity and mortality [16]. The complexity of pregnancy-related health issues and the need for early detection and therapy are both demonstrated by these two potential dangers. A wave of psychological difficulties is frequently brought on by pregnancy and the postpartum period, which harms the mental health of mothers. According to [17], anxiety and depression are common psychological disorders in pregnant and postpartum women. According to [18], factors like a personal or family history of mental health disorders, labour complications, and a lack of social support all raised the likelihood of postpartum depression. The literary work by [19] also looked at how risk factors for maternal health interact with one another. According to [19]'s research, addressing one risk factor in isolation could not have a meaningful impact on maternal health outcomes since maternal health hazards are multi-dimensional. To improve maternal health care, they advocate for a comprehensive strategy that takes socioeconomic, physical, and mental health factors into account.

B. Machine Learning in Healthcare

Within the last decade, the usage of machine learning in the healthcare industry has increased. The marriage between Machine learning and the healthcare sector has recorded significant success because of the enormous amount of data available. Machine learning algorithms can examine complex medical data and draw insightful conclusions that can enhance patient care. Risk prediction is one of the main uses of machine learning in healthcare. Machine learning algorithms can predict a person's likelihood of contracting a specific disease by examining medical records, genetic data, and lifestyle factors. This can help medical professionals put preventive measures in place for high-risk patients. For instance, machine learning has been used to forecast the risk of diabetes, cardiovascular disease, and numerous cancer types [20].

In addition to risk prediction, machine learning has been applied in healthcare decision-making. A patient's medical history and current condition can be analysed using machine learning algorithms to recommend potential diagnosis or treatment. This has the potential to increase healthcare delivery's effectiveness and efficiency greatly. Machine learning has also been utilized in various ways in the field of maternal health. Examining maternal health records and genetic data, for instance, has been utilized to forecast the likelihood of preterm birth [21][22][23]. Additionally, machine learning has been applied to enhance maternal health decision-making. It can assist medical professionals in determining the best course of action for treating problems like gestational diabetes [24].

C. Machine Learning in Maternal Health Risk

In 2022, [28], in a review paper titled *Machine Learning to Predict Pregnancy Outcomes: a Systematic Review, Synthesizing Framework and Future Research Agenda*, systematically investigated and explored the state-of-the-art views, identified future research scopes and also highlighted the limitations of studies focusing on the use of Machine Learning (ML) in predicting pregnancy outcomes. They provided an overview of ML's application in predicting pregnancy outcomes and its current limitations. They also offered insights into data sources, features, and the publication trend of ML-based pregnancy research. Furthermore, they showcased past studies' diverse objectives and an analysis of the ML algorithms used, setting the stage for future research directions.

Most recently, some studies in 2023 by [29] titled *Machine learning-based maternal health risk prediction model for IoMT framework* employed five (5) ML algorithms (Random Forest, Decision Tree, SVM, KNN, and Logistic Regression), with the Random Forest achieving the best accuracy of 91% among the five (5) models. The research opined that pregnant women with higher body temperatures have an increased health risk, irrespective of their heart rate.

Some traditional ML methods were also employed by [30] in predicting maternal health risk. They used six (6) ML classifiers in their experiment with the Decision Tree, achieving an accuracy of 89.16%, while the KNN recorded the lowest classification accuracy of 68.47%. They concluded their findings by stating that early pregnancy presents risks that depend on various factors, including age, birth history, socioeconomic status, and substance use.

D. Machine Learning Algorithms in focus

For this study, we shall consider three (3) supervised machine learning algorithms. The three (3) algorithms selected for this experiment are as follows: Logistic regression, Support Vector Machine, and K-Nearest Neighbour. A brief description of these algorithms is given below:

- Logistic Regression

To solve a classification issue, logistic regression is performed. Based on the values of the input variables, it provides the binomial result, which indicates the likelihood that an event will occur or not (in terms of 0 and 1) [3]. Logistic regression predicts binary outcomes (True/False, 1/0, Yes/No) from a given set of independent variables [4]. For instance, determining whether a tumour is cancerous (malignant) or benign or if an email is regarded as spam can be considered a binomial result of logistic regression. The results of logistic regression analysis may also be multinomial, such as a prediction of the preferred cuisine among Chinese, Italian, Mexican, etc. There may also be an ordinal result, such as a product rating from 1 to 5. Therefore, the goal of logistic regression is to predict a categorical target variable [3]. The Logistic regression model equations, as presented by [4], is given below:

$$x = c_0 \sum_{i=1}^{n} c_i x_i \quad (1)$$

$$P(x) = \frac{e^x}{1 + e^x} \quad (2)$$

where x is the quantity of the illustration variables xi (i = 1, ... , n) that are present, and ci is the regression coefficient that was obtained with the highest probability when compared to its normal errors. The certain acknowledgements of the variables that define the possibility of an excitation are Δci and P(x). In this study, the threshold was considered equal to or larger than 0.5, or P(x) 0.5, resulting in the classification of a record as an excitement [4].

- Support Vector Machine (SVM)

Support vector machines (SVMs) are speculations of a maximal edge classifier, a natural classifier [4]. SVM can handle both classification and regression tasks [3]. The meaning of the hyperplane, which represents in an n-dimensional space, is accompanied by the maximal edge classifier. The level subspace of the hyperplane has (n 1) dimensions and need not pass through the root. Since drawing hyperplanes in higher dimensions is challenging, (n-1) dimensional level subspace is still employed. Creating an SVM classifier is simple if a separating hyperplane is present. Since the dataset categories cannot be separated using a hyperplane, the feature space must be expanded using a polynomial function of higher order or a Gaussian radial basis function (RBF), sigmoid function, cubic function, or even higher order [4]. According to [4], the following is a description of the p-dimensions hyperplane:

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p \quad (3)$$

where X_1, X_2, …, and X_p are the data points in the sample space of p-dimension and β_0, β_1, β_2, …, and β_p are the hypothetical values as presented by [4].

- K-Nearest Neighbour

The K-nearest neighbour algorithm is another supervised machine learning algorithm that is used in pattern recognition and grouping. In predictive analysis, it is frequently employed. The K-NN method finds the nearest existing data points when new data is introduced [4][25]. KNN does not assume any underlying data distribution and so it is called non-parametric [3]. The interval between data points may be sufficiently influenced by any characteristics that can vary widely. In the training stage, the feature vectors and class labels are saved. K-NNs presumptively represent the data samples in a metric space. In the classification phase, the quantity is initially described by K's neighbours, which are the K training sample's most regular. The computation will then identify K nearby neighbours of the newly sampled data [4]. Let us assume that the number of neighbours is denoted by N in K-NNs, and then N samples are considered using the following distance metric value:

$$Minkowski\ Distance: Dist\ (x,y) = (\sum_{i=1}^{n} |x_i - y_i|^p)^{\frac{1}{p}} \quad (4)$$

Where, if $p = 1$, then it is Manhattan distance; if $p = 2$, then it is Euclidean distance; and if $p = \infty$, then it is Chebyshev distance as presented by [4].

III. METHODOLOGY

A. Study design

This study involves the individual training of each chosen machine learning algorithm, which serves as the first step in the comparative analysis in this study. Each algorithm will be trained using the same dataset to develop a model that can predict maternal health risks depending on the input variables. These models will be tested using a different, unexplored section of the dataset (the test set) after the training is finished. Performance metrics, which include accuracy, precision, recall, F1-score, and AUC-ROC, will be determined by contrasting the predictions provided by the models with the actual results. The machine learning algorithms will be evaluated using these performance metrics. The study seeks to give an objective evaluation of which algorithm(s) performs better in the task of maternal health risk classification by comparing the machine learning algorithms' performance on the same task using the same dataset and evaluation criteria.

B. Datasets description

The dataset chosen for this work was obtained from UCI Machine Learning Data repository. was collected from different hospitals, community clinics, and maternal health cares in the rural areas of Bangladesh through the IoT-based risk monitoring system. This dataset was denoted on 31st December 2020. This dataset comprises thousand and fourteen (1014) instances and seven (7) attributes. The following are the attributes of the dataset: Age, Systolic Blood Pressure as SystolicBP, Diastolic BP as DiastolicBP, Blood Sugar as BS, Body Temperature as BodyTemp, HeartRate, and RiskLevel. There are three (3) RiskLevel; they are Low Risk, Mid Risk, and High Risk. An explorative data analysis of the dataset is given below in Fig. 1.

C. Data Preprocessing

- Loading dataset

The dataset selected for this project is stored as a CSV file. This project was implemented using the Python programming language. The dataset was loaded using the Pandas framework. A missing or null value(s) for each column was checked, and there were no missing or null values from the columns of the dataset.

- Changing categorical labels to numeric labels

The label (RiskLevel) in the dataset is of the categorical form (Low Risk, High Risk, and Middle Risk); for the successful implementation of the classification task, these categorical labels must be changed to numeric labels. The categorical labels conversion was decoded thus: Low Risk = 0, High Risk = 2, and Middle Risk = 1.

- Classification classes and separating features from labels

Furthermore, we created classification classes. This classification class is what will be used in classifying the risk level of the record (row) within the dataset. The risk classes are Low Risk, Mid Risk, and High Risk. Also, we separated features from the labels so that the classification task could be implemented.

- *Splitting the data for training and validation set*

As stated above, the dataset selected for this task contains one thousand and fourteen (1014) instances. Before performing the classification experiment, there is a need to split the data into training and testing sets. For this reason, we split the dataset into 70%, i.e. (709 instances) as the training set and 30% (305 instances) as the testing set. In order to obtain an optimal result, we perform hyperparameter tuning using GridSearchCV from the Sci-

kit learn library (sklearn.model_selection) of the Python programming language.

- *Machine Learning Algorithms experimental setup*

Logistic Regression

For logistic regression, we used the sci-kit-learn Python library's logistic regression function. We set the parameters grid values so that we can use the GridSearchCV that can help us implement the hyperparameters tuning in training the model using the training set.

Support Vector Machine (SVM)

We also use the scikit-learn library containing the SVM function. We also defined parameter grid values for the C values (0.1, 1, 10, 100), kernel: ('linear', 'rbf'), degree (1, 2, 3), and gamma: ('scale', 'auto'). This will permit the

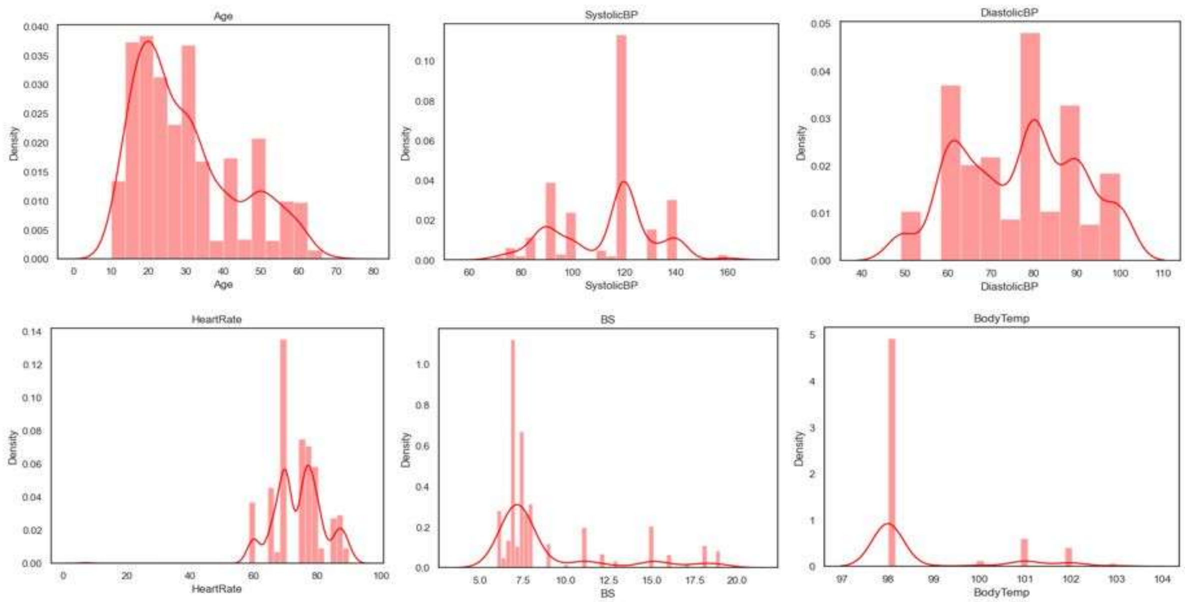

Figure 1: An explorative data analysis of the independent features within the dataset

usage of the GridSearchCV that will perform the hyperparameters tuning with the bit to select the best parameters to be used in training the model using the training set.

K-Nearest Neighbours (KNN)

Just as it was with logistic regression and SVM, we utilized the sci-kit-learn library's KNeighborsClassifier function. Parameters grid values were also defined as follows: 'n_neighbors: [1,3,5,7,9,11,13,15,17,19], weights ('uniform', 'distance'), metric : ('Minkowski, 'Euclidean', 'manhattan'). The GridSearchCV functionality made use of these parameter grid values in choosing the best parameter values to be used in training the model using the training set.

D. Evaluation Metrics

Since our study aims at comparing the performance of machine learning algorithms in predicting maternal health risks, there is a need to define and explain the metrics we will be using in measuring the performance of these algorithms. According to [26], when comparing and evaluating various machine learning or classification models, performance metrics are quite useful. Numerous metrics can be used to evaluate the performance of any multi-class classifier and are beneficial for contrasting the results of two or more different models, examining the behaviour of the same model after adjusting various parameters. It is worthy of note that [26] further argued that the Confusion Matrix serves as the foundation for many metrics because it contains all the necessary data regarding the effectiveness of the algorithm and classification rules. For this reason, we will be briefly describing these metrics of measurement (confusion matrix, accuracy, precision, recall, F1-score, AUC-ROC) below:

- *Confusion Matrix*

The confusion matrix is a cross table that counts the instances between two raters, the true/actual classification, and the predicted classification. According to [26], the rows show the classification, whereas the columns represent model prediction. Because the classes are listed in rows and columns in the same order, the correctly categorized elements are found on the major diagonal, from top left to bottom right, corresponding to the percentage of times the two raters agree. Figure 2 below is a sample confusion matrix according to [26] with dummy data:

		PREDICTED classification				
	Classes	a	b	c	d	Total
ACTUAL classification	a	6	0	1	2	9
	b	3	9	1	1	14
	c	1	0	10	2	13
	d	1	2	1	12	16
	Total	11	11	13	17	**52**

Figure 2: A sample Confusion Matrix with dummy data

- *Accuracy*

Accuracy is one of the most commonly used metrics in model classification. It is described as the percentage of accurate predictions made by the model. It is mathematically determined by dividing the total number of predictions by the sum of true positives and negatives [26]. Furthermore, when classes are balanced or nearly equal numbers of instances in each class, accuracy can serve as a quick and straightforward indicator of model performance. For example, a model with an accuracy of 0.9 produces accurate predictions 90% of the time. However, in situations with imbalanced classes, where one class vastly outnumbers the other (as with this experiment), accuracy can be a misleading indicator. Consider a model created to find an illness that only affects 1% of the population. Despite being unable to accurately identify any cases of the disease, a naive model that predicts "no disease" for all patients would have a 99% accuracy rate. In certain circumstances, different metrics like precision, recall, or the F1 score may offer a more accurate representation of model performance [26]. As mentioned above, we will examine the F1 Score, another crucial model classification metric, because of the imbalanced classes at our disposal. Mathematically, as proposed by [26], accuracy is given by equation 5 below:

$$\text{Accuracy} = \frac{\text{TP}+}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{5}$$

- *F1 Score*

As discussed above, in circumstances of imbalanced classes, when accuracy might not be a reliable indicator of model performance, the F1 score is a metric that aims to balance the precision and recall of a model. According to [26], it is described as the harmonic mean of the metrics precision and recall, which are crucial for model classification in and of themselves. The fraction of accurate positive predictions among all positive predictions generated by the model is known as **precision**, sometimes referred to as the positive predictive value. It is a crucial metric in situations where the cost of false positives is considerable. For instance, a highly precise model in email spam detection would reduce the possibility that crucial communications would be mistakenly classified as spam [26]. Mathematically, precision is given by the equation [26] below:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{6}$$

Recall, often called sensitivity or true positive rate, quantifies the percentage of real positives the model accurately recognized. When the cost of false negatives is high, it is essential. For instance, a high recall is preferred in medical diagnosis models to ensure that the greatest number of positive cases are found for additional research [26]. The equation for the recall as presented by [26] is given as:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{7}$$

The above formula can be understood as follows: how many did the model manage to uncover among everything positive? Even though they may occasionally misidentify some negative examples as positive, a model with high recall does an excellent job of locating all the positive cases in the data. All or a significant portion of the positive points in the data cannot be found by a model with low recall. The F1 score comes in as a single metric that strikes a balance between precision and recall. In addition to making accurate predictions and reducing both types of errors, a model with a high F1 score also excels in recall and accuracy. The F1 score has some limits, much like any measures. It assumes that false positives and negatives cost the same amount, which may not be true in all cases. Therefore, the cost of various sorts of errors should constantly be considered when choosing a measure [26]. F1 score is given by equation 3.4 as presented by [26]:

$$\text{F1 Score} = \frac{2}{\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}} \tag{8}$$

$$= \frac{2 \text{ x Precision x Recall}}{\text{Precision} + \text{Recall}} \tag{9}$$

- *AUC – ROC*

The Area Under the Receiver Operating Characteristic (AUC-ROC) performance measurement is used for classification issues at various threshold values. The true positive rate (sensitivity) vs the false positive rate (1-specificity) for different threshold levels is plotted on the ROC curve. According to [26], the AUC symbolizes the level or measure of separability, demonstrating how well the model can distinguish between classes. At all threshold values, the model accurately distinguishes between positive and negative cases, achieving an AUC of 1, which denotes perfect classification ability. An AUC of 0.5, on the other hand, means that the model has no classification capacity and performs no better than chance. AUC values between 0.5 and 1 indicate that a model can classify data, with higher values suggesting improved performance. The ROC curve is a two-dimensional graph whereby TPR stands for the y-axis and FPR represents the x-axis. Furthermore, the ROC curve is a graph that displays how well a classification model performs across all thresholds. This graph shows the two parameters: True Positive Rate (TPR) and False Positive Rate (FPR). Figure 3, presented by [26], shows a basic ROC Curve showing essential points on the graph.

According to [57], True Positive Rate (TPR) is defined by equation 3.5 below:

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{10}$$

Whereas the False Positive Rate (FPR) is also given by:

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{TN} + \mathrm{FP}} \tag{11}$$

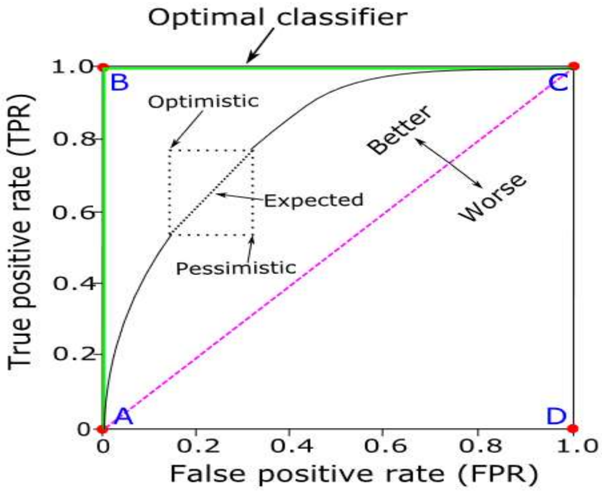

Figure 3: Basic ROC Curve showing important Points

IV. Results and discussion

This section provides the results and discussion of the three (3) machine learning models applied in the prediction of maternal health risks. We shall be presenting the result obtained using each of the models separately and after presenting some tables and plotting comparing the results.

A. Logistic regression results

The experiments were successfully run, and the following results were obtained using the logistic regression model. The confusion matrix obtained from the logistic regression model is presented below:

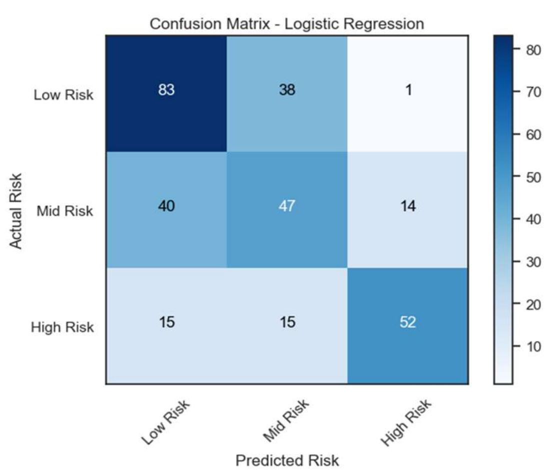

Figure 4: Confusion matrix obtained using LR

From the confusion matrix above in fig. 4, the logistic regression model performance was relatively above average. The model also finds it challenging in correctly classifying the mid risk class as there was a considerably large number of misclassifications more than the correctly place classification.

The bar chart below (Fig. 5) shows the accuracy, precision, recall and F1 score obtained from the logistic regression model. As can be seen from the bar chart, 0.60 accuracy means that 60% of all classifications made by the model are correct. The precision of 0.62 means that when the model predicts the positive class, it is correct 62% of the time; regarding the recall of 0.59, it indicates that the model correctly identifies 59% of all actual positive cases. Considering these scores, it appears the model is performing moderately. In medical diagnosis as it is the situation at hand in this study, false negatives are very costly, a model with a higher recall is preferred in this regard.

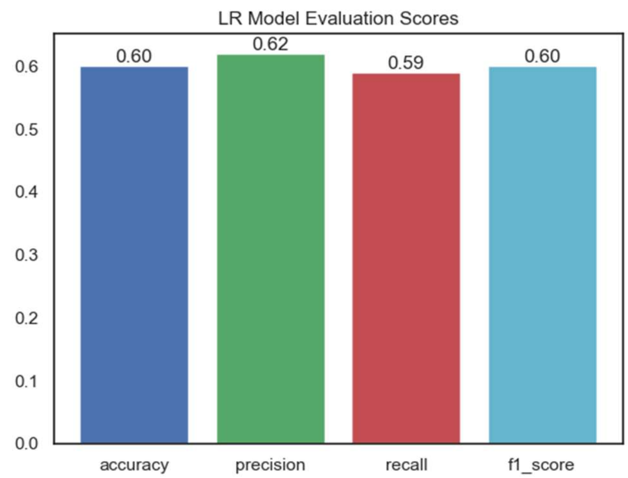

Figure 5: Bar chart showing LR Model Metrics Scores

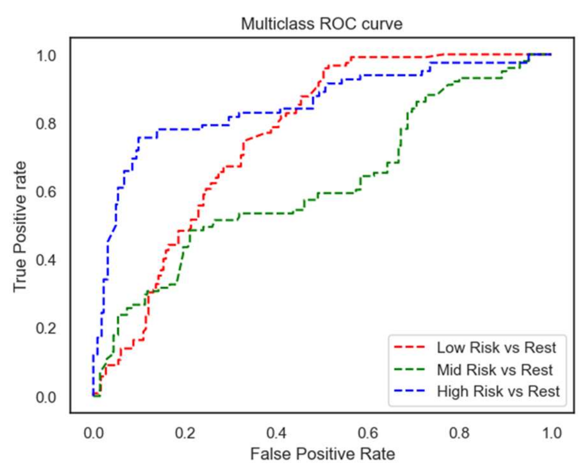

Figure 6: Multiclass ROC Curve showing a class vs the rest

From figure 6 above, we can see a multiclass ROC curve showing the plots for the three risks classes against the rest. The mid risk class recorded the lowest AUC score than both the low risk and the high-risk classes. The logistic regression model recorded an of AUC score of 0.74. In the context of a maternal health risks prediction task, this means that on average, the model has about a 74% chance of correctly distinguishing between positive cases (mothers with health risks) and negative cases (mothers without health risks).

B. SVM Results

The following results were obtained using the Support Vector Machine model. The confusion matrix obtained from the Support Vector Machine model is presented below:

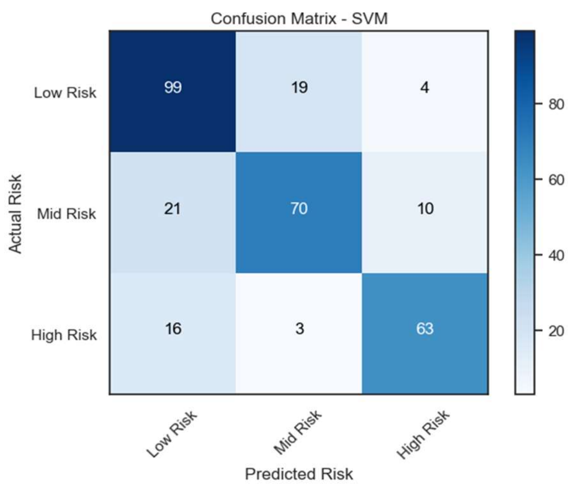

Figure 7: Confusion matrix obtained using SVM

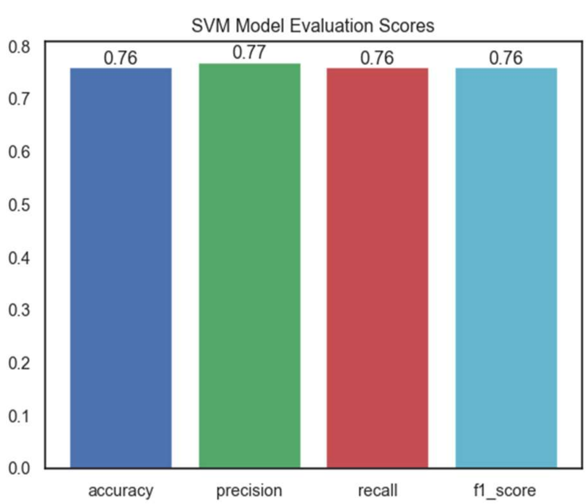

Figure 8: Bar chart showing SVM Model Evaluation Scores

Using the SVM model, from the confusion matrix in fig. 7 above, the SVM model performance was better than that of the logistic regression model. Even though the SVM model also finds it challenging to correctly classify the mid risk class just like the logistic regression model, the number of the correctly classified instances of the mid class is higher than the misclassified ones.

The bar chart above (fig. 8) displays the accuracy, precision, recall and F1 scores obtained using the support vector machine model. Clearly, it can be seen from the bar chart, 0.76 accuracy means that 76% of all classifications made by the model are correct. The precision of 0.77 means that when the model predicts the positive class, it is correct 77% of the time; regarding the recall of 0.76, it indicates that the model correctly identifies 76% of all actual positive cases. Considering these scores, it appears the model is performing moderately. As opined in some research works, a higher recall value is preferred in medical diagnosis, the SVM model classifies the maternal health risks with higher accuracy than the logistic regression model. Since higher values of false negatives are very costly, the SVM model is preferred to the logistic regression model.

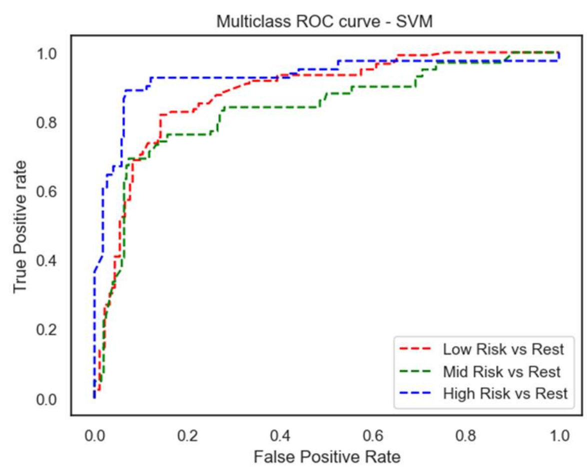

Figure 9: Multiclass ROC Curve showing a class vs the rest using the SVM Model

Fig. 9 above presents the multiclass ROC curve showing the plots for the three risks classes against the rest using the SVM model. Just as we observed with the logistic regression, the mid risk class recorded the lowest AUC score than both the low risk and the high-risk classes. The SVM model has an average AUC score of 0.88. For this maternal health risks prediction task, the model has about a 88% chance of correctly distinguishing between positive cases (mothers with health risks) and negative cases (mothers without health risks).

C. KNN Results

The results obtained using the KNN model are hereby given below:

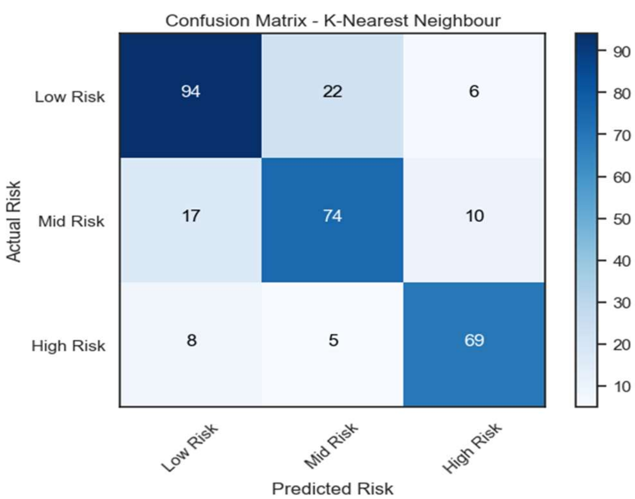

Figure 11: Confusion matrix obtained using KNN

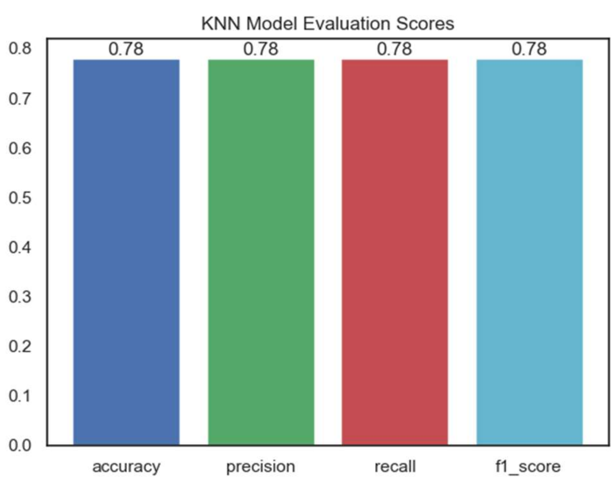

Figure 12: Bar chart showing KNN Model Evaluation Scores

The confusion matrix as presented in (fig. 10) above, K-NN model, proves to be the better of the three (3) machine learning models employed in this comparative study. The K-NN model performed a bit better than the SVM model in correctly predicting the risk classes.

Also, from the bar chart above (fig.11), we can observe that the accuracy, precision, recall and the F1 score are all pegged at 78%. The K-NN model recorded a higher recall score, this suggests that the K-NN model will preferably be recommended for medical diagnosis than the other two models.

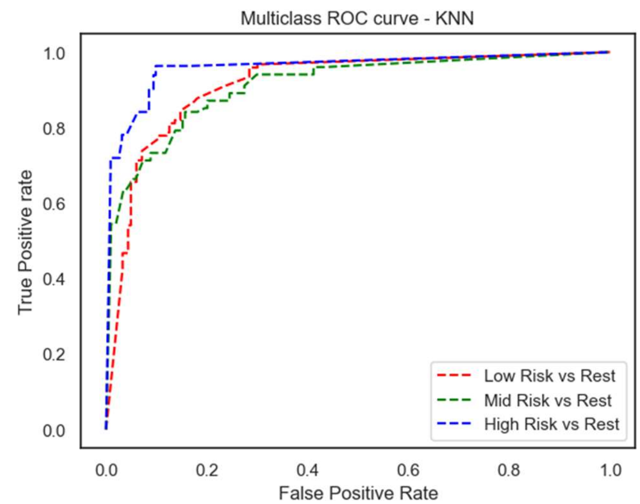

Figure 10: Multiclass ROC Curve showing a class vs the rest using the KNN Model

Fig. 12 above, the plot is from the KNN model. The multiclass ROC curve showing the plots for the three risks classes against the rest. Just as we observed with both the logistic regression and also the SVM models, the mid-risk class recorded the lowest AUC score than both the low-risk and the high-risk classes. The KNN model has an average AUC score of 0.93. This is a good AUC score, more especially for the maternal health risks prediction task; the model has about a 93% chance of correctly distinguishing between positive cases (mothers with health risks) and negative cases (mothers without health risks).

Considering the results obtained, in general, employing machine learning for maternal health risk prediction has enormous potential real-world implications. This encompasses its impact on healthcare outcomes and patient care; some of these areas of potentials includes and not limited to improved health care outcomes (early detection and reduction in maternal and neonatal morbidity rate), optimizing the limited available resources through saving cost and efficient use of resources, enhancing patients' experiences by providing personalized care, and addressing health disparities through equitable care .

D. Comparative Analysis

In this section, we present some explorative results (plots and a table) and discussions about the three (3) machine learning models we implemented.

- *Accuracy comparison*

We begin the comparative analysis by showing a bar chart plot showing the accuracy of the three (3) machine learning models we implemented.

Fig. 13 below presents the prediction accuracy obtained from the three (3) machine learning models we employed. The logistic regression model recorded the least accuracy value of 0.60; this means that the logistic regression model is correctly classification prediction with 60% accuracy. The support vector machine model recorded an accuracy of 76% in correctly predicting maternal health risks, while the K-Nearest Neighbour model scored the highest accuracy of 78% prediction accuracy. We boldly say that the KNN model is performing optimally more than the other two (2) models used in this comparative study.

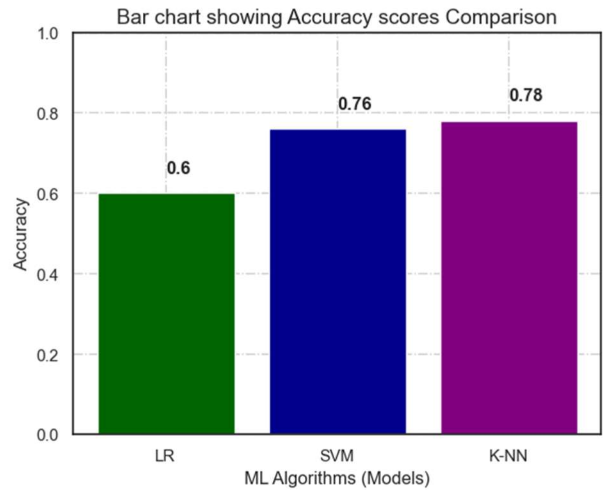

Figure 13: Bar chart showing Accuracy scores Comparison

TABLE 1: METRICS SCORE FOR THE THREE ML MODELS

ML Models	Risk Classes	Accuracy	Precision	Recall	F1 Score	Support
Logistic Regression	Low Risk – 0	0.60	0.60	0.68	0.64	122
	Mid Risk – 1		0.47	0.47	0.47	101
	High Risk – 2		0.78	0.63	0.70	82
Support Vector Machine	Low Risk – 0	0.76	0.73	0.81	0.77	122
	Mid Risk – 1		0.76	0.61	0.73	101
	High Risk – 2		0.82	0.77	0.79	82
K-Nearest Neighbour	Low Risk – 0	0.78	0.79	0.77	0.78	122
	Mid Risk – 1		0.73	0.73	0.73	101
	High Risk – 2		0.81	0.84	0.83	82

Table 1 above shows the metrics scores obtained from the three (3) machine learning models we employed in performing this comparative study in predicting maternal health risks. By taking a cursory look upon table 1 above, we will notice that all the three (3) ML models recorded their best performance in the high-risk class and then it is followed by the low-risk class. This means that the models struggled to some extent in identifying and classifying the mid-risk class. This can be attributed to misclassifying the mid-risk class to either high-risk class or the low-risk class.

- *Precision comparison*

From Figure 14 below, the model's prediction of the three classes of maternal health risks shows reasonable precision. In more specific terms, an average of 78% chance of correctly identifying the instances as belonging to the three risk classes when it performs prediction was recorded by the K-NN model. The K-NN model displayed a good outing more than the other two models. On the other hand, the logistic regression model has the lowest average precision score of 62% chances while the SVM model obtained a 77% chance of correctly identifying the three risk classes.

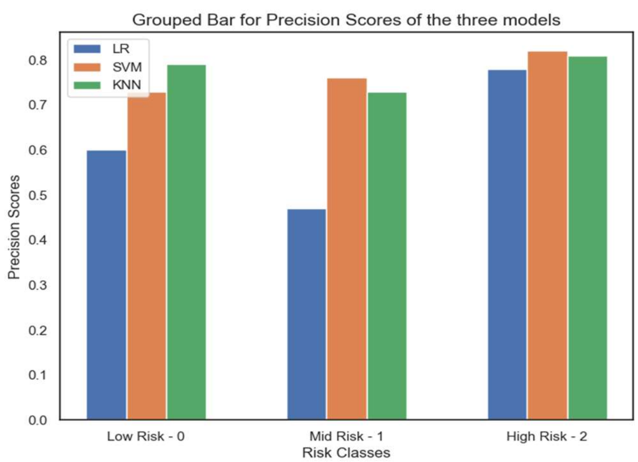

Figure 14: Group bar chart showing precision scores obtained from the three ML Models

Figure 15 above presents a group bar chart showing the recall score of the three (3) machine learning models on the three risk classes. The logistic regression model recorded the least average recall score of 59%. For low-risk class, the recall score of 0.68 implies that the model correctly identifies 68% of the actual positive instances of the class. 47% for the mid-risk class and 63% for the high-risk class. Using the SVM model, the average recall score is 0.73; this means that it is identifying the three classes correctly at a 73% rate. The following percentage were recorded by the SVM model: 81%, 61%, and 77% for the low-risk class, mid-risk, and high-risk classes

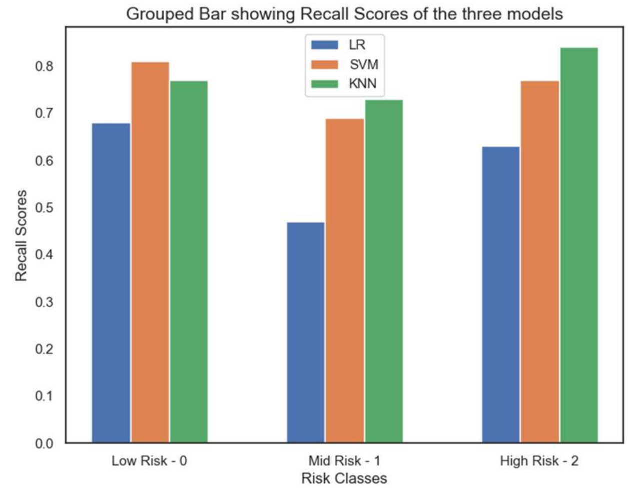

Figure 15:Group bar chart showing recall scores obtained from the three ML Models

respectively. The K-NN model recorded the highest average recall score of 78% among the three (3) models. The low-risk, mid-risk, and high-risk classes recorded the following recall scores: 77%, 73%, and 84%, respectively. These model's recall scores demonstrate a respectable performance in predicting the three classes of maternal health risks. However, there's still a need for improved performance.

- *F1 Scores comparison*

Finally, the F1 scores obtained by the three (3) machine learning models are given figure 16 below. The F1 score is described as the harmonic mean of recall and precision; it shows the balance that exists between these two metrics. It indicates a balanced perspective on the model's performance across precision and recall for each maternal health risk category. For the logistic regression model, the low-risk, mid-risk, and high-risk classes recorded 64%, 47%, and 70% F1 scores, respectively in identifying the true instances of the risk classes while keeping the false classification at a minimum. Using the SVM, the low-risk class recorded 77%, the mid-class obtained 73%, and the high-class was at 79%. The K- NN, as can be seen, performed better than both the SVM and the logistic regression model.

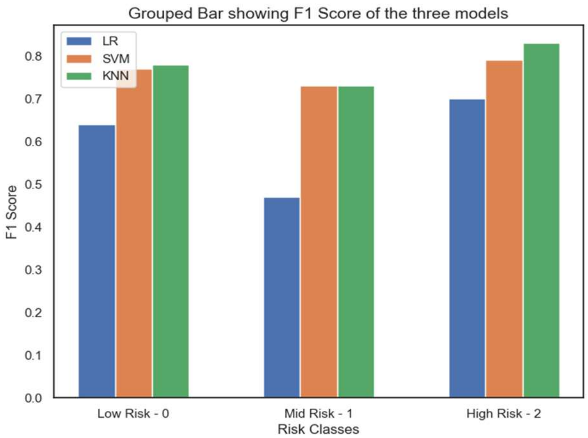

Figure 16: Group bar chart showing F1 scores obtained from the three ML Models

- *Summary of Key Findings*

Considering the results obtained by each of the three (3) algorithms used. We can observe that the KNN obtained the best accuracy recall and the F1 scores in detecting and classifying the risk classes whereas the SVM stood out in achieving the best precision score when considering the three (3) risk classes. The LR models recorded the least performance score of the three algorithms used in this experiment as a result of the non-linearity of the relationship between independent and dependent variables, the nature of the data. The best performance recorded by the KNN boils down to the distribution of the data in which they are closely knitted together in small patches and as a result, it finds ease in locating local neighbours.

When it comes to time and resources constraints, the SVM will not be the best choice for implementation in situations where time and resources are in limited supply. This is because of the quadratic programming the SVM performs to find the hyperplane that maximizes the margin between two classes; also, the kernel computations where n x n (where n is the number of training samples), the usage of significant portion of the training data support vectors that are stored in memory and as a result rendering the SVM model memory-intensive in nature. On the other hand, the Logistic Regression models has better interpretability because of its meaningful and clear coefficients interpretation between features and the outcomes; it has probabilistic outcome of the model by providing the probability of the class label belonging to a particular class.

Summarily, SVM appears to be the most suitable choice for real-world applications in maternal health risk prediction as a result of its high accuracy score. However, in situations where limited computational resources is a factor, the Logistic Regression is recommended due to its efficiency and interpretability, even if it means sacrificing some bit of accuracy. The interpretability of Logistic Regression can be leveraged for patient counseling. By understanding the factors contributing to risk predictions, healthcare providers can offer timely, informed, targeted advice and interventions.

Finally, each algorithm has its strengths and weaknesses, understanding these nuances is crucial for their effective application in maternal healthcare. By choosing the most appropriate algorithm based on specific needs (medical diagnosis) and constraints (time, resources), we can harness the power of machine learning to significantly enhance maternal health outcomes and patient care.

V. Conclusion

In conclusion, it was observed that from this comparative experimental analysis for predicting maternal health risks using the three (3) selected machine learning models. The results show that the logistics regression models recorded the lowest accuracy score of classifying the maternal health risks at 60%, the Support Vector Machine had 76% per cent classification accuracy while the K-Nearest Neighbour (K-NN) model scored 78% classification accuracy. From the foregoing, the K-NN model is the best among the models selected for this comparative study. Although the classification accuracy scores obtained from these models are not high, it can go a long way, providing some insights into maternal health risks endeavours. With more improved classification accuracy scores and ethical modalities and consideration in place, this study will bring about increased efficiency and effectiveness of healthcare delivery through analysis of patients' medical history and current condition to suggest possible closer medical attention, diagnosis, or treatment.

Finally, the practical significance of ML in maternal health risk prediction is vast. It can potentially transform maternal healthcare by making it more proactive, personalized, and efficient. Also, these findings can help in early detection and intervention, healthcare policy making, research and development, and reducing health disparities. However, it's important to approach the integration of ML into healthcare with caution by ensuring that these models are transparent, validated, and free from biases. Collaboration between clinicians, data scientists, patients, and policymakers will be crucial to realize the full potential of ML in improving maternal healthcare. On this note, it is worthy to say that this comparative study provides some guiding light in the interdisciplinary integration between machine learning and healthcare.

VI. Further Work

There are quite a couple number of machine learning algorithms available, this study considers comparing only three

(3) supervised machine learning algorithms (LR, SVM, and the K-NN). Further work into this interdisciplinary comparative study with other machine learning algorithms should be on the card with the view of exploring the performance of these algorithms in predicting maternal health risks. Even though we performed hyperparameter tuning using the GridSearchCV, further hyperparameter tuning with more grid parameters values can be employed, this will most likely bring about more classification accuracy than the ones we obtained using the three (3) selected machine learning algorithms. This worthy of note that the dataset has fewer number of instances, getting a larger dataset or performing data augmentation on the dataset used for this experiment can also help in training and testing the model, this we believe can improve the models' learning and prediction capabilities.

Further research works on areas where ML can further enhance Maternal Health Risk can be considered in the following areas such as predicting Preterm Birth Prediction, Postpartum Depression Prediction, Early Prediction of High-Risk Pregnancies, and Optimizing Prenatal Care will go a long way in enhancing maternal health risk prediction. Finally, Maternal Health Risk is a very sensitive healthcare delivery area, having a more robust, accurate and effective machine learning model(s) can bring about building an application or technology in the future that will complement the service of the skilled health workers we have in the world today. This will mitigate the maternal death rates we have been recording in the world today.

REFERENCES

[1] R. C. Deo, "Machine Learning in Medicine," *Circulation*, vol. 132, no. 20, pp. 1920–1930, Nov. 2015, doi: 10.1161/CIRCULATIONAHA.115.001593.

[2] "novel clinical risk prediction model for sudden cardiac death in hypertrophic cardiomyopathy (HCM Risk-SCD) | European Heart Journal | Oxford Academic." https://academic.oup.com/eurheartj/article/35/30/2010/467191 (accessed Jul. 28, 2023).

[3] S. Ray, "A Quick Review of Machine Learning Algorithms," in *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, Feb. 2019, pp. 35–39. doi: 10.1109/COMITCon.2019.8862451.

[4] Md. M. Islam, Md. R. Haque, H. Iqbal, Md. M. Hasan, M. Hasan, and M. N. Kabir, "Breast Cancer Prediction: A Comparative Study Using Machine Learning Techniques," *SN COMPUT. SCI.*, vol. 1, no. 5, p. 290, Sep. 2020, doi: 10.1007/s42979-020-00305-w.

[5] H. Saber, M. Somai, G. B. Rajah, F. Scalzo, and D. S. Liebeskind, "Predictive analytics and machine learning in stroke and neurovascular medicine," *Neurological Research*, vol. 41, no. 8, pp. 681–690, Aug. 2019, doi: 10.1080/01616412.2019.1609159.

[6] R. Wang *et al.*, "Artificial intelligence in reproductive medicine," *Reproduction*, vol. 158, no. 4, pp. R139–R154, Oct. 2019, doi: 10.1530/REP-18-0523.

[7] L.-Q. Zhou *et al.*, "Artificial intelligence in medical imaging of the liver," *World J Gastroenterol*, vol. 25, no. 6, pp. 672–682, Feb. 2019, doi: 10.3748/wjg.v25.i6.672.

[8] A. Hosny, C. Parmar, J. Quackenbush, L. H. Schwartz, and H. J. W. L. Aerts, "Artificial intelligence in radiology," *Nat Rev Cancer*, vol. 18, no. 8, pp. 500–510, Aug. 2018, doi: 10.1038/s41568-018-0016-5.

[9] M. Vodanović, M. Subašić, D. Milošević, and I. Savić Pavičin, "Artificial Intelligence in Medicine and Dentistry," *Acta Stomatol Croat*, vol. 57, no. 1, pp. 70–84, Mar. 2023, doi: 10.15644/asc57/1/8.

[10] "Maternal health." https://www.who.int/health-topics/maternal-health (accessed Jul. 28, 2023).

[11] N. J. Kassebaum *et al.*, "Global, regional, and national levels and causes of maternal mortality during 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013," *The Lancet*, vol. 384, no. 9947, pp. 980–1004, Sep. 2014, doi: 10.1016/S0140-6736(14)60696-6.

[12] N. S. Prema and M. P. Pushpalatha, "Machine Learning Approach for Preterm Birth Prediction Based on Maternal Chronic Conditions," in *Emerging Research in Electronics, Computer Science and Technology*, V. Sridhar, M. C. Padma, and K. A. R. Rao, Eds., in Lecture Notes in Electrical Engineering. Singapore: Springer, 2019, pp. 581–588. doi: 10.1007/978-981-13-5802-9_52.

[13] S. Goli, Moradhvaj, A. Rammohan, Shruti, and J. Pradhan, "High Spending on Maternity Care in India: What Are the Factors Explaining It?," *PLoS One*, vol. 11, no. 6, p. e0156437, Jun. 2016, doi: 10.1371/journal.pone.0156437.

[14] K. B. Kozhimannil, R. R. Hardeman, L. B. Attanasio, C. Blauer-Peterson, and M. O'Brien, "Doula Care, Birth Outcomes, and Costs Among Medicaid Beneficiaries," *Am J Public Health*, vol. 103, no. 4, pp. e113–e121, Apr. 2013, doi: 10.2105/AJPH.2012.301201.

[15] E. Chiefari, B. Arcidiacono, D. Foti, and A. Brunetti, "Gestational diabetes mellitus: an updated overview," *Journal of Endocrinological Investigation*, vol. 40, Mar. 2017, doi: 10.1007/s40618-016-0607-5.

[16] L. Duley, "The global impact of pre-eclampsia and eclampsia," *Semin Perinatol*, vol. 33, no. 3, pp. 130–137, Jun. 2009, doi: 10.1053/j.semperi.2009.02.010.

[17] J. Fisher *et al.*, "Prevalence and determinants of common perinatal mental disorders in women in low- and lower-middle-income countries: a systematic review," *Bull World Health Organ*, vol. 90, no. 2, pp. 139-149H, Feb. 2012, doi: 10.2471/BLT.11.091850.

[18] A. Biaggi, S. Conroy, S. Pawlby, and C. M. Pariante, "Identifying the women at risk of antenatal anxiety and depression: A systematic review," *J Affect Disord*, vol. 191, pp. 62–77, Feb. 2016, doi: 10.1016/j.jad.2015.11.014.

[19] "National Vital Statistics Reports Volume 69, Number 7 July 16, 2020, Infant Mortality in the United States, 2018: Data From the Period Linked Birth/Infant Death File".

[20] S. F. Weng, J. Reps, J. Kai, J. M. Garibaldi, and N. Qureshi, "Can machine-learning improve cardiovascular risk prediction using routine clinical data?," *PLOS ONE*, vol. 12, no. 4, p. e0174944, Apr. 2017, doi: 10.1371/journal.pone.0174944.

[21] "A Machine Learning Approach to Measure and Monitor Physical Activity in Children to Help Fight Overweight and Obesity | SpringerLink." https://link.springer.com/chapter/10.1007/978-3-319-22186-1_67 (accessed Jul. 27, 2023).

[22] D. S. Bogale, T. M. Abuhay, and B. E. Dejene, “Predicting perinatal mortality based on maternal health status and health insurance service using homogeneous ensemble machine learning methods,” *BMC Med Inform Decis Mak*, vol. 22, no. 1, p. 341, Dec. 2022, doi: 10.1186/s12911-022-02084-1.

[23] A. Fredriksson *et al.*, “Machine learning for maternal health: Predicting delivery location in a community health worker program in Zanzibar,” *Frontiers in Digital Health*, vol. 4, 2022, Accessed: Jul. 29, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fdgth.2022.855236

[24] A. Vellido, “The importance of interpretability and visualization in machine learning for applications in medicine and health care,” *Neural Comput & Applic*, vol. 32, no. 24, pp. 18069–18083, Dec. 2020, doi: 10.1007/s00521-019-04051-w.

[25] G. James, D. Witten, T. Hastie, and R. Tibshirani, “Classification,” in *An Introduction to Statistical Learning: with Applications in R*, G. James, D. Witten, T. Hastie, and R. Tibshirani, Eds., in Springer Texts in Statistics. New York, NY: Springer US, 2021, pp. 129–195. doi: 10.1007/978-1-0716-1418-1_4.

[26] M. Grandini, E. Bagli, and G. Visani, “Metrics for Multi-Class Classification: an Overview.” arXiv, Aug. 13, 2020. Accessed: Jul. 29, 2023. [Online]. Available: http://arxiv.org/abs/2008.05756

[27] A. Tharwat, “Classification assessment methods,” *Applied Computing and Informatics*, vol. 17, no. 1, pp. 168–192, Jan. 2020, doi: 10.1016/j.aci.2018.08.003.

[28] Islam et al. "Machine learning to predict pregnancy outcomes: a systematic review, synthesizing framework and future research agenda." BMC Pregnancy and Childbirth, 2022, 22:348. https://doi.org/10.1186/s12884-022-04594-2.

[29] Mondal, S., Nag, A., Barman, A.K., & Karmakar, M. "Machine learning-based maternal health risk prediction model for IoMT framework." International Journal of Experimental Research and Reviews, 2023, Vol. 32. https://doi.org/10.52756/ijerr.2023.v32.012.

[30] H. B. Mutlu, N. Yücel, F. Durmaz, and M. Yildirim, "Prediction of Maternal Health Risk with Traditional Machine Learning Methods," in NATURENGS (MTU Journal of Engineering and Natural Sciences), vol. 4, no. 1, 2023. https://doi.org/10.46572/naturengs.1293185