JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 13(1), MARCH, 2025
ISSN: 2277-0011; Journal homepage: www.atbuftejoste.com.ng

Predicting Maternal Health Risks Using Machine Learning: A Comparative Study of Classification Algorithms

[1]Ajayi Ore-Ofe, [2]Elvis Obi, [3]Kamorudeen Hassan Taiwo, [1]Ibrahim Ibrahim
[1]Computer Engineering Department, Ahmadu Bello University, Zaria.
[2]Electronics &Telecommunication Engineering Department, Ahmadu Bello University, Zaria,
[3]Department of Family Medicine, Ahmadu Bello University Teaching Hospital, Zaria

ABSTRACT

Maternal health is a critical aspect of public healthcare, with complications during pregnancy and childbirth contributing to high mortality rates worldwide. This study explores the potential of machine learning techniques in predicting maternal health risks using physiological indicators such as age, blood pressure, blood sugar levels, body temperature, and heart rate. A dataset from Kaggle was preprocessed to ensure quality, followed by exploratory data analysis to identify key patterns. Three machine learning models: Random Forest, Support Vector Machines, and K-Nearest Neighbors were trained and evaluated. The Random Forest classifier emerged as the best-performing model, achieving an accuracy of 80.79%, along with superior precision, recall, and F1-score. Performance evaluation using confusion matrices confirmed its effectiveness in distinguishing maternal risk levels. These findings highlight the potential of machine learning for early risk assessment, supporting timely medical interventions. While the study demonstrates promising results, further improvements such as model optimization, dataset expansion, and integration into real-world healthcare systems are necessary to maximize impact. This research contributes to the growing field of AI-driven maternal healthcare, paving the way for predictive tools that could enhance clinical decision-making and improve maternal health outcomes.

ARTICLE INFO
Article History
Received: October, 2024
Received in revised form: December, 2024
Accepted: February, 2025
Published online: March, 2025

KEYWORDS
Maternal Health Risks, Machine Learning, Comparative Study, Classification Algorithms

INTRODUCTION

Maternal health encompasses the physical, mental, and emotional well-being of a mother throughout pregnancy, childbirth, and the postpartum period. The rates of maternal morbidity and mortality during pregnancy serve as crucial indicators of healthcare accessibility and the availability of maternal medical resources. Various pregnancy-related complications, including hypertension, diabetes, haemorrhaging, and preterm birth, are among the primary causes of maternal mortality. Identifying pregnancy-related risks early is essential to preventing adverse outcomes and ensuring timely medical intervention. Machine learning techniques play a significant role in assessing maternal health risks. By leveraging machine learning algorithms to analyze health data and associated risk factors, it is possible to monitor and predict a pregnant woman's risk level. Consequently, machine learning-based models are considered valuable tools in mitigating maternal mortality by addressing complications that arise due to evolving risk factors [1].

Several medical conditions, such as advanced maternal age, blood disorders, and irregular heart rhythms, can contribute to complications during pregnancy. Understanding these health risks can help reduce the likelihood of such complications. Research on maternal health highlights key factors to monitor during pregnancy, including maternal age, heart rate,

Corresponding author: Ajayi Ore-Ofe
ajayioreofe@gmail.com
Computer Engineering Department, Ahmadu Bello University, Zaria.

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 13(1), MARCH, 2025
ISSN: 2277-0011; Journal homepage: www.atbuftejoste.com.ng

systolic and diastolic blood pressure, as well as overall blood pressure. By considering these factors, the timely identification of potential risks through machine learning algorithms can play a crucial role in reducing maternal and neonatal mortality rates, ultimately safeguarding the health of pregnant individuals [2].

Medical complications such as essential hypertension, pre-eclampsia, renal and autoimmune disorders, maternal diabetes, and thyroid conditions can pose significant health risks to pregnant women [7–10]. Additionally, pregnancy-related issues like prolonged gestation, vaginal bleeding, reduced fetal movements, and prolonged rupture of membranes may endanger fetal well-being. Other factors, including intrauterine growth restriction, fetal infections, and multiple pregnancies, further contribute to fetal health risks. Consequently, these conditions can lead to neurodevelopmental impairments in infancy, such as non-ambulatory cerebral palsy, developmental delays, auditory and visual impairments, and fetal distress, which may result in neonatal morbidity or mortality [3].
Recognizing pregnancies at high risk is essential for safeguarding both maternal and fetal health, yet it remains a complex challenge due to the diverse and intricate factors involved.

Maternal health risk during pregnancy pertains to the probability of a woman encountering adverse health conditions during gestation or childbirth. Based on identified risk factors that contribute to pregnancy complications, pregnancies can be categorized as low, moderate, or high risk [4]. An analysis of global maternal mortality reveals that approximately two-thirds (200,000) of maternal deaths occur in sub-Saharan Africa, while South Asia accounts for 19% (57,000) of such deaths. Various maternal health factors, including age and blood-related disorders such as hypertension, hypotension, blood glucose fluctuations, body temperature irregularities, and abnormal heart rates, significantly contribute to pregnancy complications. These complications can lead to pregnancy loss and, in severe cases, the death of both the mother and child. Managing these health factors often requires specialized medical interventions. Early prediction of potential health risks can enable medical professionals to take timely and appropriate measures, thereby reducing the likelihood of maternal and neonatal mortality [5].

LITERATURE REVIEW

Maternal healthcare utilization plays a crucial role in reducing the risks associated with diseases, excessive bleeding, and fatalities resulting from pregnancy and childbirth complications. Both antenatal and delivery care are essential for ensuring the health and well-being of mothers and infants. Effective maternal healthcare can significantly prevent negative pregnancy outcomes by either proactive intervention or efficient management. Healthcare professionals typically recommend that expectant mothers begin antenatal visits early in pregnancy and continue with regular check-ups to mitigate potential pregnancy-related risks. Additionally, childbirth should occur in a sanitary environment equipped with the necessary medical tools and attended by trained healthcare providers to minimize infection risks and ensure that any arising complications are managed appropriately.

The postnatal period, which immediately follows childbirth, is also critical for both mother and newborn, as it carries a heightened risk of mortality. Consequently, adequate postnatal care is vital in lowering mortality rates and enhancing overall quality of life [6]. Maternal mortality represents only a fraction of the overall challenges associated with maternal health. Many women who encounter life-threatening complications may not survive, while others narrowly avoid death through critical obstetric interventions. Studies indicate that for every maternal death resulting from pregnancy-related complications, approximately 20 to 30 women experience severe, life-threatening conditions that have lasting physical and psychological effects [7]. Maternal health literacy (MHL) refers to a mother's capacity to obtain, comprehend, evaluate, and utilize information related to maternal and child health, playing a crucial role in lowering maternal and child mortality rates [8].

Corresponding author: Ajayi Ore-Ofe
ajayioreofe@gmail.com
Computer Engineering Department, Ahmadu Bello University, Zaria.

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 13(1), MARCH, 2025
ISSN: 2277-0011; Journal homepage: www.atbuftejoste.com.ng

Machine learning is a branch of computer science that focuses on developing algorithms capable of learning from experience, allowing them to enhance their performance progressively [9]. Machine learning broadly encompasses the development of predictive models from data and the discovery of meaningful patterns within datasets. This field aims to replicate the human capacity for pattern recognition through computational methods in an objective manner. Machine learning proves particularly valuable when dealing with datasets that are either too extensive—comprising numerous data points—or highly complex, featuring a large number of attributes, making manual analysis impractical. Additionally, it is instrumental in automating data analysis, ensuring a reproducible and efficient workflow [10].

Recent advancements in machine learning have significantly enhanced the capabilities of healthcare professionals by assisting in decision-making, identifying medical trends, and improving disease prediction models. In large medical institutions, machine learning techniques have also been utilized to enhance the management of electronic health records, detect anomalies in blood samples, organs, and bones through medical imaging, and support robot-assisted surgical procedures, ultimately improving overall healthcare efficiency [11]. Machine learning has recently been applied across various medical fields, including heart failure management, clinical decision support in medicine, and medical imaging [12].

METHODOLOGY

The research design adopted for this study is an experimental approach aimed at analyzing maternal health risks using machine learning techniques. The dataset, sourced from Kaggle, comprises various physiological parameters of expectant mothers, including age, blood pressure, blood sugar levels, body temperature, and heart rate. These variables were systematically examined to determine their correlation with maternal health risks. The study follows a structured pipeline involving data preprocessing, visualization, model development, and evaluation. Each stage was carefully designed to ensure data quality and reliable predictive modelling outcomes. The experimental nature of the study allowed for a comparative analysis of different machine learning algorithms, identifying the most effective model for maternal health risk prediction.

The dataset used in this study contains 1,014 instances with seven attributes: age, systolic blood pressure, diastolic blood pressure, blood sugar level, body temperature, heart rate, and risk level as shown in Figure 1 below. The risk level, which serves as the target variable, is categorical with three classifications: low risk, mid risk, and high risk. Each feature within the dataset plays a crucial role in determining maternal health conditions. The dataset was assessed to understand the distribution of values, frequency of occurrences, and presence of any anomalies. Statistical summaries were generated to provide an overview of data trends and variability. Understanding the dataset characteristics was a key step in ensuring the effectiveness of machine learning models in making accurate predictions.

	Age	SystolicBP	DiastolicBP	BS	BodyTemp	HeartRate	RiskLevel
0	25	130	80	15.0	98.0	86	high risk
1	35	140	90	13.0	98.0	70	high risk
2	29	90	70	8.0	100.0	80	high risk
3	30	140	85	7.0	98.0	70	high risk
4	35	120	60	6.1	98.0	76	low risk

Figure 1. Dataset used for study

Corresponding author: Ajayi Ore-Ofe
ajayioreofe@gmail.com
Computer Engineering Department, Ahmadu Bello University, Zaria.

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 13(1), MARCH, 2025
ISSN: 2277-0011; Journal homepage: www.atbuftejoste.com.ng

Before analysis, the dataset was checked for missing values, inconsistencies, and outliers. The dataset was found to be complete with no missing entries. However, exploratory data analysis was conducted to identify any abnormalities that could impact model performance. The categorical target variable was encoded using label encoding to facilitate machine learning model training. The numerical features were standardized using the StandardScaler method to ensure uniform scaling. This step was essential in preventing certain features from having a dominant influence on model training. Furthermore, potential data redundancy was assessed to remove any duplicated observations that could bias the training process. By ensuring a clean and well-prepared dataset, the integrity of the machine learning models was maintained.

Visualization techniques, including heatmaps and pairplots, were employed to explore relationships between variables. Heatmaps, shown in Figure 2, were used to identify correlations among independent variables, while pairplots provided a scatter matrix to observe variable distributions and interactions. These visualizations helped in identifying trends, patterns, and dependencies within the dataset. For instance, it was observed that systolic and diastolic blood pressure showed a moderate positive correlation, reinforcing the importance of these features in risk classification. The use of visual analytics provided an intuitive understanding of how different features contribute to the overall maternal health risk assessment.

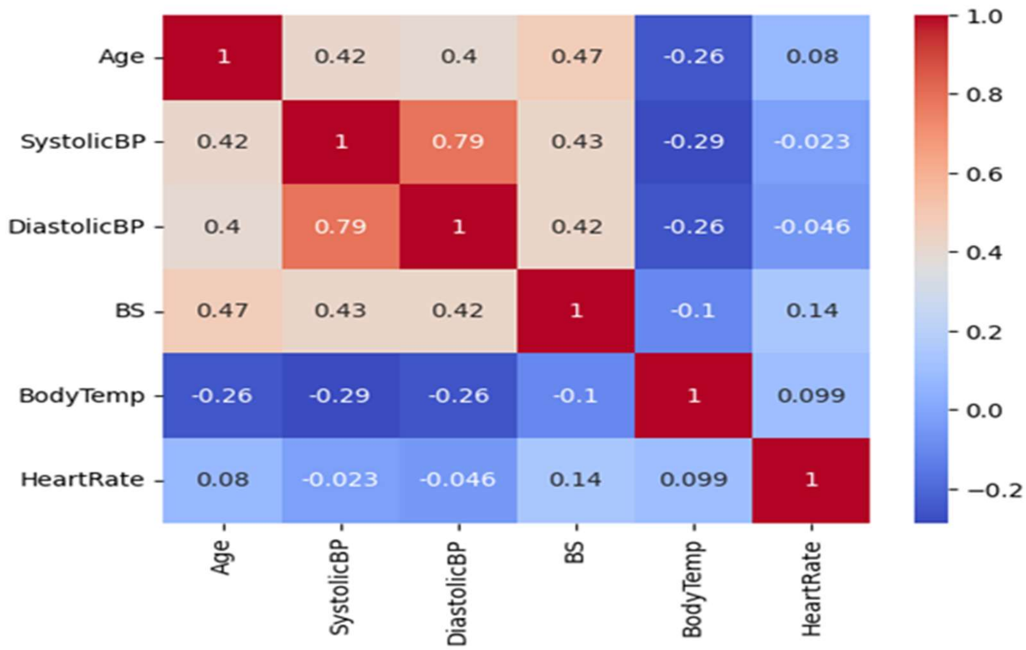

Figure 2. Heatmaps showing correlations between variables

Three classification algorithms were selected for model training: Random Forest, Support Vector Machine (SVM), and K-Nearest Neighbors (KNN). These algorithms were chosen due to their efficiency in handling categorical classification tasks. Random Forest was selected due to its ability to handle complex relationships and reduce overfitting through ensemble learning. SVM was chosen for its robustness in high-dimensional spaces, while KNN was included for its intuitive approach to classification. By employing these diverse algorithms, a comprehensive comparison was made to determine the best-performing model for maternal

Corresponding author: Ajayi Ore-Ofe
ajayioreofe@gmail.com
Computer Engineering Department, Ahmadu Bello University, Zaria.

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 13(1), MARCH, 2025
ISSN: 2277-0011; Journal homepage: www.atbuftejoste.com.ng

health risk prediction. The dataset was split into training and testing sets in an 80-20 ratio. The training set was used to fit the models, while the test set was reserved for evaluating their performance. Each model was trained using its default hyperparameters, followed by hyperparameter tuning for improved accuracy. Training involved feeding the models with labeled data, enabling them to learn patterns and relationships. The training process was monitored to ensure convergence, and performance metrics were tracked throughout iterations. A thorough model training process was essential to achieving reliable predictive outcomes.

A grid search approach was used to optimize model parameters. For Random Forest, the number of estimators was tuned, while for SVM, the kernel type and regularization parameter were adjusted. The number of neighbors for KNN was also optimized. This tuning process was crucial in enhancing model accuracy, as hyperparameters directly influence decision boundaries and classification performance. Various combinations were tested, and the optimal parameters were selected based on cross-validation scores.

Hyperparameter tuning significantly improved classification outcomes, ensuring each model operated at its best potential. The models were evaluated based on accuracy, precision, recall, and F1-score. Confusion matrices were generated to provide insight into classification performance, highlighting false positives, false negatives, true positives, and true negatives. These metrics were instrumental in determining model reliability. High accuracy alone was not sufficient; a balanced precision-recall trade-off was considered to ensure the models effectively identified all risk categories. The confusion matrix further revealed the models' ability to minimize misclassifications, guiding the selection of the best-performing model.

RESULTS AND DISCUSSION

The trained models were tested on the test dataset, and their performances were compared based on evaluation metrics. The accuracy scores for the three models as shown in Figure 3 were recorded as: 80.79% for Random Forest, 67.98% for SVM, 65.02% for KNN.

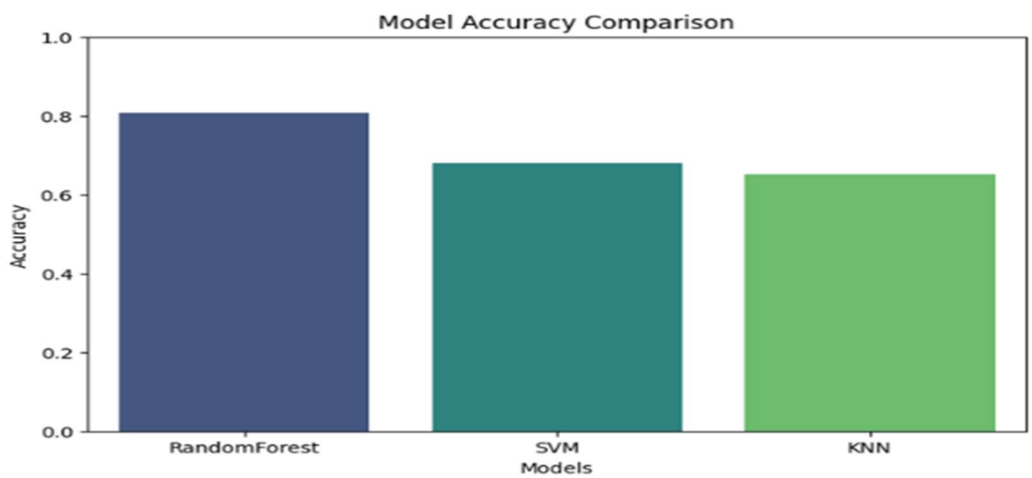

Figure 3. Accuracy scores for each model

Among the three models, Random Forest exhibited the highest accuracy, making it the most reliable for predicting maternal health risks. The performance gap between Random Forest and the other models indicates its robustness in handling structured medical data. This superior performance can be attributed to Random Forest's ability to reduce variance through ensemble learning, leading to better generalization. The SVM and KNN models demonstrated moderate accuracy, but their performance suggests potential improvements

Corresponding author: Ajayi Ore-Ofe
ajayioreofe@gmail.com
Computer Engineering Department, Ahmadu Bello University, Zaria.

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 13(1), MARCH, 2025
ISSN: 2277-0011; Journal homepage: www.atbuftejoste.com.ng

through advanced feature selection and hyperparameter tuning.

The classification reports shown in Table 1 detailed the precision, recall, and F1-scores for each model. Random Forest demonstrated balanced precision and recall across all risk categories, whereas SVM and KNN had lower recall values for certain categories. This balance in classification performance ensured that the model minimized false negatives, which is critical in medical applications were failing to detect high-risk cases could have severe consequences.

Table 1. Classification reports

Model	Precision	Recall	F1-Score
RandomForest	0.814527	0.807882	0.808429
SVM	0.700964	0.679803	0.657794
KNN	0.649654	0.650246	0.642666

The confusion matrices for all models were analyzed to understand classification errors. Random Forest had the lowest number of misclassifications compared to SVM and KNN. A visual representation of the best model's confusion matrix is included in Figure 4.

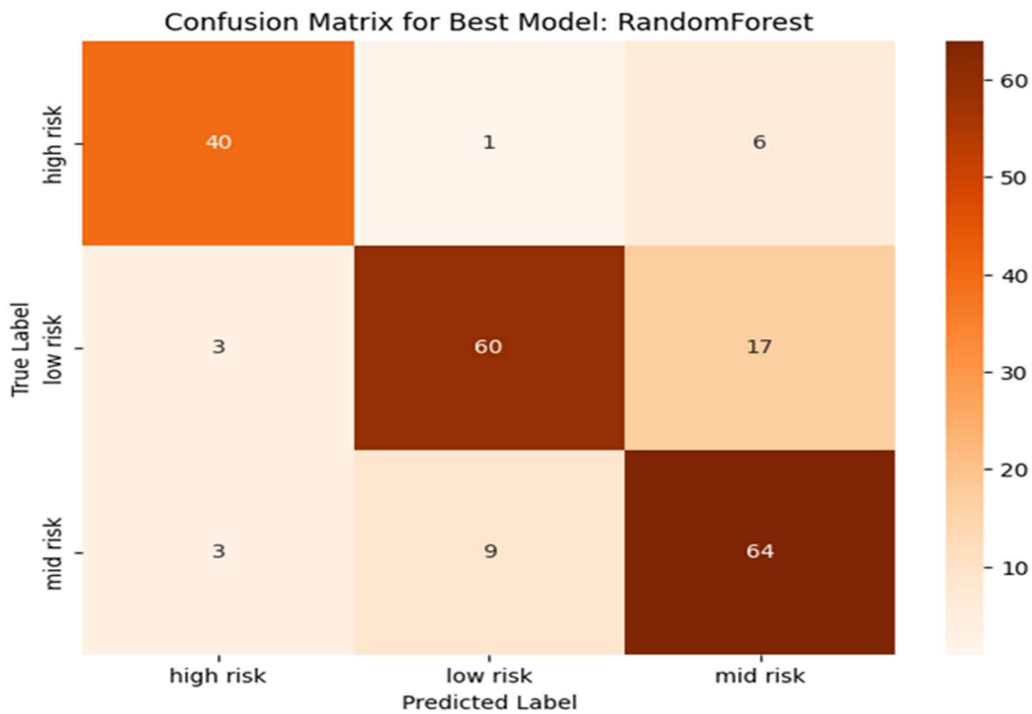

Figure 4. Confusion matrices for the best model

Feature importance was examined for the Random Forest model, highlighting the most influential variables in risk prediction as shown in Figure 5. Systolic blood pressure and blood sugar levels emerged as key determinants of maternal health risks. These findings provide valuable insights into which features healthcare practitioners should prioritize in maternal risk assessments.

Corresponding author: Ajayi Ore-Ofe
ajayioreofe@gmail.com
Computer Engineering Department, Ahmadu Bello University, Zaria.

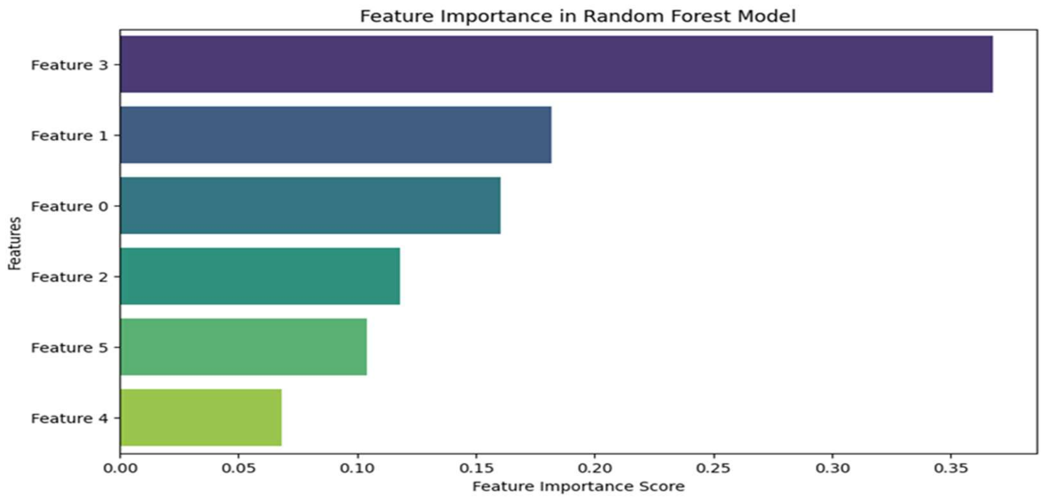

Figure 5. Important feature analysis results

Feature scaling significantly improved model accuracy by ensuring that no feature disproportionately influenced the model. The standardization process helped maintain consistency across all variables. The results align with previous studies where ensemble methods like Random Forest outperform linear classifiers such as SVM and instance-based classifiers like KNN. This highlights the advantage of decision trees in capturing complex patterns within medical datasets. The findings suggest that machine learning can be effectively applied in maternal health risk assessment, allowing early detection of high-risk cases. The ability to classify patients into different risk categories can assist healthcare professionals in prioritizing cases that require urgent attention.

CONCLUSION AND LIMITATION

The study has successfully examined the application of machine learning techniques in predicting maternal health risks based on multiple physiological parameters. By leveraging data-driven approaches, it has been demonstrated that certain machine learning algorithms, particularly Random Forest, provide a reliable means of classifying maternal health risk levels. The research incorporated various steps, including data preprocessing, feature engineering, model selection, performance evaluation, and visualization, ensuring a comprehensive analysis.

The dataset utilized contained critical attributes such as age, blood pressure, blood sugar levels, body temperature, and heart rate, all of which played a significant role in determining maternal risk categories. Through rigorous model training and evaluation, different machine learning classifiers were tested, including Random Forest, Support Vector Machines, and K-Nearest Neighbors. Among them, the Random Forest classifier demonstrated the highest accuracy and overall better performance in terms of precision, recall, and F1-score. The confusion matrix and other performance metrics further reinforced the effectiveness of this approach. Data visualization techniques were also employed to understand the distribution and correlation of the variables, enhancing interpretability and aiding in model refinement. However, certain limitations must be acknowledged.

The dataset, while informative, may not fully capture the diversity of maternal health conditions across different populations, and expanding it to include more diverse demographic and regional representations could improve model generalizability. Additionally, while the study focused on traditional machine learning models, exploring deep learning techniques and advanced ensemble methods could further optimize performance. Another limitation is the lack of integration with clinical systems, which would

Corresponding author: Ajayi Ore-Ofe
ajayioreofe@gmail.com
Computer Engineering Department, Ahmadu Bello University, Zaria.

enable real-time risk assessments for healthcare providers. The study also did not develop a user-friendly application for practical implementation, which could facilitate accessibility for both medical professionals and expectant mothers. Finally, while the models demonstrated strong performance, further validation through clinical trials and collaboration with healthcare professionals is necessary to confirm their real-world applicability and alignment with medical best practices.

REFERENCES

[1] Mutlu, H. B., Durmaz, F., Yücel, N., Cengil, E., & Yıldırım, M. (2023). Prediction of maternal health risk with traditional machine learning methods. *Naturengs*, 4(1), 16-23.

[2] Raihen, M. N., & Akter, S. (2024). Comparative Assessment of Several Effective Machine Learning Classification Methods for Maternal Health Risk. *Computational Journal of Mathematical and Statistical Sciences*, *3*(1), 161-176.

[3] Alam, M. T., Khan, M. A. I., Dola, N. N., Tazin, T., Khan, M. M., Albraikan, A. A., & Almalki, F. A. (2022). [Retracted] Comparative Analysis of Different Efficient Machine Learning Methods for Fetal Health Classification. *Applied Bionics and Biomechanics*, *2022*(1), 6321884.

[4] Togunwa, T. O., Babatunde, A. O., & Abdullah, K. U. R. (2023). Deep hybrid model for maternal health risk classification in pregnancy: synergy of ANN and random forest. *Frontiers in Artificial Intelligence*, *6*, 1213436.

[5] Raza, A., Siddiqui, H. U. R., Munir, K., Almutairi, M., Rustam, F., & Ashraf, I. (2022). Ensemble learning-based feature engineering to analyze maternal health during pregnancy and health risk prediction. *Plos one*, *17*(11), e0276525.

[6] Samuel, O., Zewotir, T., & North, D. (2021). Decomposing the urban–rural inequalities in the utilisation of maternal health care services: evidence from 27 selected countries in Sub-Saharan Africa. *Reproductive Health*, *18*, 1-12.

[7] Beyene, T., Chojenta, C., Smith, R., & Loxton, D. (2022). Severe maternal outcomes and quality of maternal health care in South Ethiopia. *International Journal of Women's Health*, 119-130.

[8] Phommachanh, S., Essink, D. R., Wright, P. E., Broerse, J. E., & Mayxay, M. (2021). Maternal health literacy on mother and child health care: A community cluster survey in two southern provinces in Laos. *Plos one*, *16*(3), e0244181.

[9] Surden, H. (2021). Machine learning and law: An overview. *Research Handbook on Big Data Law*, 171-184.

[10] Greener, J. G., Kandathil, S. M., Moffat, L., & Jones, D. T. (2022). A guide to machine learning for biologists. *Nature reviews Molecular cell biology*, *23*(1), 40-55.

[11] Habehh, H., & Gohel, S. (2021). Machine learning in healthcare. *Current genomics*, *22*(4), 291.

[12] Alanazi, A. (2022). Using machine learning for healthcare challenges and opportunities. *Informatics in Medicine Unlocked*, *30*, 100924.

Corresponding author: Ajayi Ore-Ofe
ajayioreofe@gmail.com
Computer Engineering Department, Ahmadu Bello University, Zaria.