Applied Computer Science 2025, vol. 21, no. 3, pp. 87–98
https://doi.org/10.35784/acs_7313

Submitted: 2025-02-18 | Revised: 2025-04-24 | Accepted: 2025-05-06

Keywords: maternal health risk, maternal mortality, machine learning, ensemble model life

Oghenevbaire EFEVBERHA-OGODO [1*], *Francisca A. EGBOKHARE* [1], *Fidelis O. CHETE* [1]

[1] University of Benin, Benin City, Nigeria, oghenevbaire.efevberha-ogodo@physci.uniben.edu, fegbokhare@uniben.edu, fidelis.chete@uniben.edu

[*] Corresponding author: oghenevbaire.efevberha-ogodo@physci.uniben.edu

An ensemble model for maternal health risk classification in Delta State, Nigeria

Abstract

Maternal mortality remains a critical challenge in Sub-Saharan Africa, with Nigeria ranking among the countries with the highest rates. The loss of women in their reproductive years destabilizes families causing emotional trauma, places additional strain on healthcare systems, and has profound economic and national developmental consequences. As a result, one of the United Nations Sustainable Development goals (SDGs) is targetted at reducing maternal mortality and morbidity at all cost. This study explores the application of Artificial Intelligence (AI) in healthcare through the development of a predictive ensemble model to classify maternal health risks as identifying high risk pregnancies can inform timely clinical decision making that mitigates maternal mortality. Maternal health dataset was sourced from three (3) health centers in Delta State, Nigeria.. Nine supervised machine learning classifiers were utilized, including Linear Support Vector Machine, Gaussian Naïve Bayes, Multilayer Perceptron, Decision Tree, Random Forest, Gradient Boosting Decision Tree, Extreme Gradient Boosting, Light Gradient Boosting Machine, and Categorical Boosting. To enhance predictive performance, the classifiers were combined in an ensemble model. Results showed that the Gradient Boosting Decision Tree achieved the highest accuracy at 90% before upsampling and Random Forest achieved an accuracy of 97% at upsampling. The lowest-performing classifier was Linear Support Vector Machine before and after upsampling. The ensemble model surpassed all individual classifiers, achieving 98% accuracy and precision and over 1% increase in accuracy after upsampling. This study highlights the potential of AI-driven predictive models to optimize healthcare resources and improve maternal health outcomes in Delta State, Nigeria.

1. INTRODUCTION

According to World Health Organisation (2019), maternal mortality occurs when a woman dies as a result of complications of pregnancy during pregnancy or within 6 weeks of termination or delivery. Deaths due to accidents or unintentional injuries are not recorded as maternal mortality. The maternal mortality ratio (MMR), on the other hand, is measured as the number of maternal deaths per 100,000 live births recorded during the same period (World Health Organisation, 2023). It is one of the United Nations (UN) Millennium Development Goals (MDGs) in the area of health to promote global development. Maternal and child mortality is covered under Sustainable Development Goal (SDG) 3, which is to "ensure healthy lives and promote well-being for all at all ages". In Nigeria and Africa, maternal and infant mortality rates during pregnancy and childbirth are significantly higher than in Europe and other continents (Ogallo et al., 2020). According to World Health Organisation (2019), the maternal mortality rate in Nigeria in 2017 can be estimated at 917 per 100,000 live births. This is in sharp contrast to the SDG target of at least 70 deaths per 100,000 live births (Olonade et al., 2019). While this is not unrelated to the systems of governance unique to a developing country like Nigeria, these numbers make maternal mortality an important area of research. There are several factors that can lead to maternal mortality, including severe bleeding (hemorrhage) after childbirth, infections during pregnancy, hypertensive disorders such as pre-eclampsia, sepsis, risky abortions, and obstructed labor (Mboya et al., 2023). Especially in Nigeria, there are other external factors such as inadequate data on maternal health, shortage of health workers, cultural beliefs and access to affordable health facilities (Izugbara et al., 2016). According to Rosser et al. (2022), health worker density in sub-Saharan Africa is 13 doctors and 91 nurses/midwives per 100,000 people-far below the World Health Organization (WHO) recommendation of 1 doctor per 600 people. With the ongoing brain drain caused by the mass emigration of Nigerian health workers, this situation is expected to worsen.

However, most maternal deaths are preventable if appropriate and timely medical care is available (Olonade et al., 2019). In maternal health care, early identification of women at risk for pregnancy complications is critical to preventing adverse outcomes (Mustamin et al., 2023). This requires the ability to analyze complex medical data, reports, and diagnostic images with high accuracy and speed. Often, anomalies or risk factors are not immediately apparent to the human eye. In these situations, machine learning (ML) becomes critical, as it can quickly process large datasets, uncover hidden patterns, and identify intricate relationships within the data that may go unnoticed by humans (Jayatilake & Ganegoda, 2021). Machine learning techniques are able to detect subtle indicators more accurately and with fewer errors than traditional statistical methods (James & Osubor, 2025). By using ML techniques, healthcare providers can make more informed, data-driven decisions, ultimately ensuring better outcomes for both mothers and babies during pregnancy (Nirmala & Kambili, 2023).

Moreover, the growing volume and complexity of medical data-from electronic health records to diagnostic imaging and real-time monitoring-have made traditional statistical approaches increasingly inadequate (Mutlu et al., 2023). In contrast, ML models excel at handling high-dimensional, heterogeneous datasets, revealing nonlinear and latent relationships that often elude traditional methods (James & Osubor, 2025). These capabilities make ML highly effective for clinical decision support, early diagnosis, patient risk stratification, and outcome prediction (Mboya et al., 2020). By leveraging historical and real-time patient data, machine learning enables the development of intelligent systems that assist clinicians in diagnosing conditions, predicting risks, and recommending treatments (Machoron et al., 2022). In maternal healthcare, ML-based decision support systems can analyze a wide range of factors-from demographics and clinical history to socioeconomic variables-to help classify risk levels, guide resource allocation, and ultimately reduce preventable maternal mortality (Dawodi et al., 2020).

Researchers have increasingly used machine learning (ML) models to predict maternal, neonatal, and child mortality. Some studies have focused on specific causes of maternal mortality, including neonatal infections, pre-eclampsia, sepsis, and cesarean sections (Gomez-Jemes et al., 2022; Kopanitsa et al., 2021; Ogallo et al., 2020). Others have used standardized or real-world data sets collected at maternal health centers to develop predictive models of maternal outcomes (Nirmala & Kambili, 2023; Mboya et al., 2020; Togunwa et al., 2023). The results of these models have been used to raise awareness and provide pregnant women with relevant information to support safer pregnancies and successful live births (Machoron et al., 2022). According to Gomez-Jemes et al. (2022), ML techniques can significantly improve maternal care by closely monitoring the health status of mothers and, consequently, their infants during pregnancy. Despite ongoing ethical and legal concerns regarding the privacy of patient data, ML holds tremendous potential for the future of healthcare, especially in remote areas where medical facilities are scarce or nonexistent (Wahl et al., 2018).

Applications of machine learning have led to significant improvements in the management of pregnancy-related conditions such as postpartum hemorrhage, pre-eclampsia, intrauterine growth restriction (IGR), preterm birth, gestational diabetes, and overall maternal health (Gomez-Jemes et al., 2022). For example, Machoron et al. (2022) applied semi-supervised algorithms to improve supervised learning models for predicting high-risk pregnancies in the Philippines. Six supervised ML algorithms-Decision Tree, Random Forest, Support Vector Machine, Multilayer Perceptron, Naïve Bayes, and K-Nearest Neighbors-were evaluated. The Decision Tree model achieved the best performance with an accuracy of 93.7%. Further application of a semi-supervised approach improved the accuracy of the model to 97.01%, emphasizing that the strength and quality of the input data are crucial for successful prediction models. Similarly, Togunwa et al. (2023) proposed a hybrid model integrating Artificial Neural Networks (ANN) with Random Forest (RF) to improve classification accuracy for maternal health risk prediction. Using a standardized dataset from the UCI Machine Learning Repository and a 75/25 training and testing split, their model achieved an impressive accuracy of 95%, which was further evaluated using F1 score, precision, and recall metrics.

In another study, Nirmala and Kambili (2023) developed an ensemble prediction model to identify pregnant women at risk of maternal mortality due to pre-eclampsia in Bangladesh and India. Using a standardized dataset from the University of California (UC) Machine Learning Repository (Ahmed, 2023), the authors categorized risk levels based on various patient characteristics. Machine learning algorithms including Random Forest, K-Nearest Neighbors (KNN), Neural Networks, and Naive Bayes were applied, with Random Forest achieving the highest individual accuracy. A stacked ensemble model was developed to improve predictive performance. Overall, the use of machine learning-based predictive models in maternal health improves the ability of health professionals to provide more efficient, timely, and quality care, helping to optimize the use of scarce resources and support informed clinical decision making.

The objectives of this study include collecting data from health centers in Delta State, Nigeria, to classify and accurately predict high-risk pregnancies that may lead to maternal mortality by leveraging the use of ensemble learning to increase model accuracy. This can potentially optimize healthcare resources by targeting high-risk pregnant women, facilitate timely clinical decision making, and consequently improve maternal health outcomes.

2. METHODOLOGY

The maternal data set was obtained from three (3) different medical centers in Delta State, Nigeria. Nigeria, a country in sub-Saharan Africa, is located in the western region of the continent. It consists of 36 states and the Federal Capital Territory (O'Neil, 2023). Delta State, one of Nigeria's oil-producing states, is located in the southern region of the country and was officially created in August 1997. The state has a population of over 4 million people and its capital city, Asaba, is located in the northern part of the state (United Nations Country Office in Nigeria, 2015). Table 1 provides a list of the three senatorial districts and the local government areas under them.

Tab. 1. The list of local government areas in Delta state and their Senatorial districts (O'Neil, 2023).

Delta Senatorial Districts	Local Government Areas
DELTA-CENTRAL	Sapele, Ethiope West, Ethiope East, Okpe, Ughelli North, Ughelli South , Udu,Uvwie
DELTA-NORTH	Aniocha North, Ukwuani, Aniocha South, Ika North East, Ika South, Ndokwa East, Ndokwa West, Oshimili North and Oshimili South
DELTA-SOUTH	Bomadi, Burutu, Isoko South, Isoko North, , Warri North, Warri South and Warri South West, Patani

The centers were selected from each senatorial district in the state, all administered by the state government. Ethical approval for data collection was obtained from the appropriate research ethics committee: Health Research Ethics Committee of Delta State University Teaching Hospital (HREC DELSUTH), Oghara (Delta Central), The Ethics and Research Committee of Central Hospital, Warri (Delta South) and Model Health Care Center, Oshimili South (Delta North) representing the three senatorial districts of Delta State. The data were collected from the antenatal records of the selected health facilities by research assistants in accordance with the guidelines and regulations of the Ethics Committee. The data were anonymized to ensure that there was no breach of confidentiality. Patient consent was not required. Data were entered manually into Microsoft Excel spreadsheets because all health centers did not have electronic health record (EHR) systems. Each patient's information was maintained in a serialized physical file stored in cabinets, which were retrieved for reference and updates during clinic visits.

The data was localized and the characteristics captured include Maternal Age (age of pregnant woman), Gestational Age (number of weeks pregnant), Systolic Blood Pressure as Systolic BP, Diastolic Blood Pressure as Diastolic BP, Parity (number of times a woman has given birth), Weight (the woman's weight at the antenatal visit), Height (the woman's height at the visit), Body Mass Index (derived from height and weight), Urine Analysis (to show the presence of glucose or protein that may be harmful to the pregnant woman), and Maternal Health Risk Level as Risk Level (derived from medical consultation). Table 2 shows a description of the data set used to develop the model.

2.1. Data preprocessing

In 1581 cases, there were some missing values in the height characteristics of the patients. This was filled with the highest occurring height value (mode value) in the dataset (Togunwa et al., 2023).

	MATERNAL AGE	GESTATIONAL AGE	PARITY	WEIGHT	HEIGHT	BMI	SYSTOLIC BP	DIASTOLIC BP	URINE ANALYSIS	RISK LEVEL
0	25	41.0	0	67.0	1.71	22.913033	180	90	NEGATIVE	High
1	31	21.0	0	76.0	1.61	29.319856	130	80	NEGATIVE	Mid
2	40	16.0	0	90.0	1.65	33.057851	140	90	NEGATIVE	High
3	30	19.0	0	57.0	1.33	32.223416	130	80	NEGATIVE	Mid
4	33	25.0	3	78.0	1.61	30.091432	120	60	NEGATIVE	Mid

Fig. 1. Excerpt of some instances of the dataset showing the features and their corresponding risk level

Tab. 2. Description of dataset

Variable Name	Role	Type	Description
Maternal Age	Feature	Integer	Age in years when a woman is pregnant.
Gestational Age	Feature	Integer	Number of weeks of pregnancy
Parity	Feature	Integer	Number of times a woman has given birth
Weight	Feature		Weight of Patient at ante-natal visit in kg
Height	Feature		Height of patient at ante-natal visit in cm
Body Mass Index (BMI)	Derived Feature	Float	Derived from weight and height of patient
Systolic BP	Feature	Integer	Upper value of Blood Pressure in mmHg, a significant feature in maternal health.
Diastolic BP	Feature	Integer	Lower value of Blood Pressure in mmHg, a significant feature in maternal health.
Urine Analysis	Feature	Integer	The Presence of Glucose and/or Protein in Urine
Risk Level	Target	Categorical	Predicted Risk Intensity Level of pregnant woman

The data was then classified into three (3) categories based on specific characteristics. The categories are as follows: Low(representing women at low risk of maternal mortality), Mid (representing women at moderate risk of maternal mortality), and High(representing women at significant risk of maternal mortality). Table 3 shows the categorization process, which was conducted under the supervision of a health professional specializing in women's reproductive health to ensure accuracy and reliability. According to World Health Organisation (2019) and Mashrafi et al. (2024), the clinical features present in the dataset are sufficient to accurately categorize the level of maternal health risk.

Tab. 3. Categorization of risk levels based on clinical features

	Low risk	Mid risk	High risk
Maternal Age	16-39 (years)	40-45 (years)	>45 years
Systolic Blood Pressure	<140 (mmhg)	140-160 (mmhg)	>160 (mmhg)
Diastolic Blood Pressure	<90 (mmhg)	90-110 (mmhg)	>110 (mmhg)
Parity	2-4	0	>4
BMI	18.5-24.9 (Normal)	25-29.9 (Overweight)	>30 (Obesity) or <18.5 (Underweight)
Gestational Age (at booking)	Less than 16 weeks	16–26 weeks	More than 26 weeks
Urinalysis	Absence of Protein and/or Glucose	Trace values of Protein and/or Glucose	High levels of Protein and/or Glucose

Data cleaning procedures performed on the dataset included the removal of invalid entries, such as instances where characters or strings were inadvertently entered in place of numeric values. In addition, outlier detection was performed using the interquartile range (IQR) method to identify and handle anomalous data points. Data transformation was performed to prepare the dataset for effective analysis and modeling. The target variable representing the level of maternal health risk (originally categorized as Low Risk, Medium Risk, and High Risk) was not immediately suitable for most machine learning algorithms, which typically require numeric input. To address this, the categorical risk levels were encoded into numerical values-highrisk, medium risk, and low risk were mapped to 0, 1, and 2, respectively. This encoding facilitated computational processing and model compatibility, as shown in Table 3.

Tab. 4. Risk levels and corresponding numeric values

Target Variable	Numeric value
Low Risk	2
Mid Risk	1
High Risk	0

A stratified 10-fold cross-validation, a variation of the standard K-fold validation technique, was used to assess model performance. In this approach, the data set was divided into ten (10) equally sized folds. For each iteration, the model was trained on nine (9) folds and validated on the remaining one (1) fold, repeating this process nine more times to ensure that each instance in the dataset is used exactly once for both training and validation. The stratification ensures that the distribution of the target variable (maternal health risk levels) is maintained proportionally across all folds, thereby minimizing sampling bias. This method improves the reliability of model evaluation by providing a more representative estimate of performance and reducing the risk of overfitting, especially given the relatively small size of the dataset (Togunwa et al., 2023).

2.2. Data upsampling (random oversmapling)

In data analysis and machine learning, unbalanced datasets, where some classes have significantly fewer instances than others, can lead to biased models that perform poorly on the minority classes. In such cases, data upsampling is a widely used method to address class imbalance (Okpono et al., 2024). Itis an oversampling technique that increases the number of samples of the minority class by simply duplicating existing samples of the same minority class until all classes are balanced (Mashrafi et al., 2024).

Tab. 5. Target variables before and after upsampling

	Classes	Before Sampling	After Sampling
1	Low Risk	844	844
2	Mid Risk	540	844
3	High Risk	192	844

Prior to upsampling, the dataset exhibited class imbalance, with the High Risk class containing only 192 samples and the Mid Risk class containing 540 samples, compared to 844 samples in the Low Risk class. This imbalance could bias the model toward predicting the majority class ("Low Risk") more frequently, thereby reducing its ability to accurately classify minority classes. To address this, upsampling was applied, resulting in all three classes, High Risk, Medium Risk, and Low Risk - being equally represented with 844 samples each.

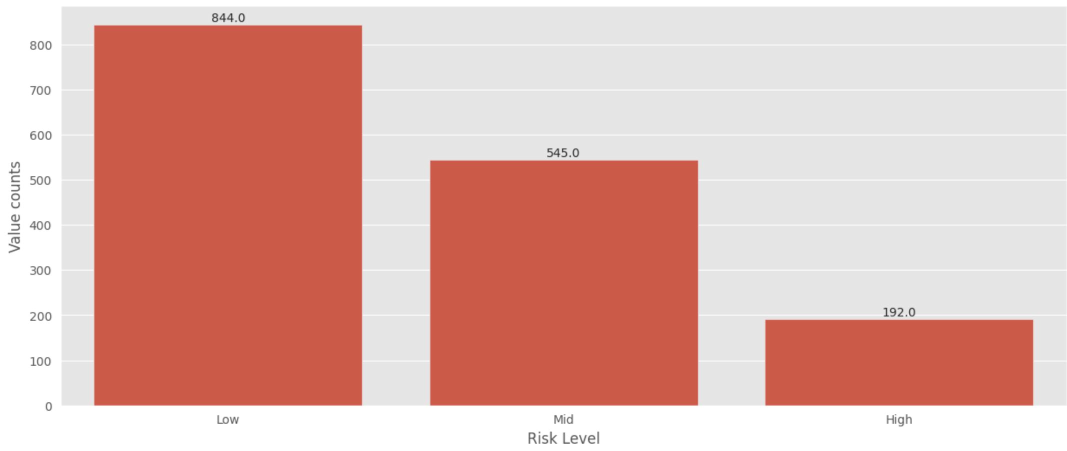

Fig. 2. Statistical distribution of risk level in dataset before upsampling

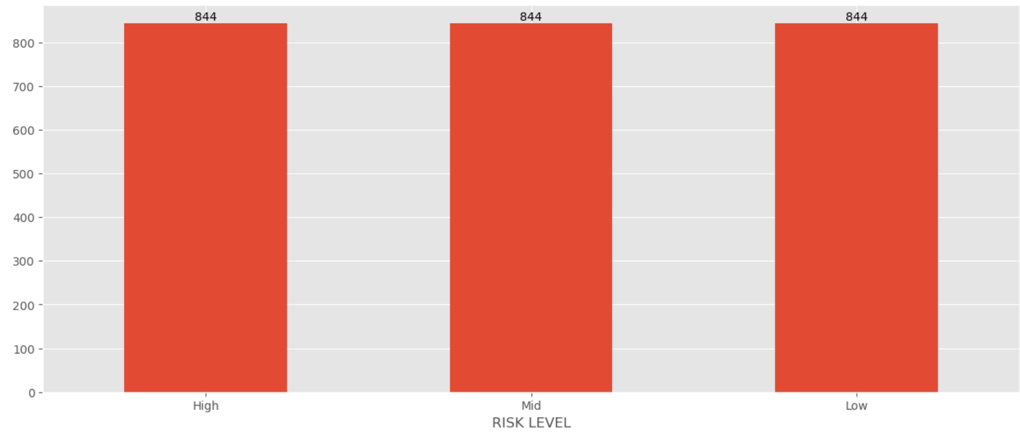

Fig. 3. Statistical distribution of risk level in dataset after upsampling

2.3. Model architecture

Figure 4 illustrates the design of the model. The data set was preprocessed to prepare it for modeling. The data was split into two subsets, with 90% used for training and 10% reserved for testing. To prevent over-fitting and ensure that all data samples were adequately represented in the model's predictions, the stratified k-fold sampling technique was applied. This approach maintained the proportion of different risk categories across each fold, thereby increasing the robustness and reliability of the model.

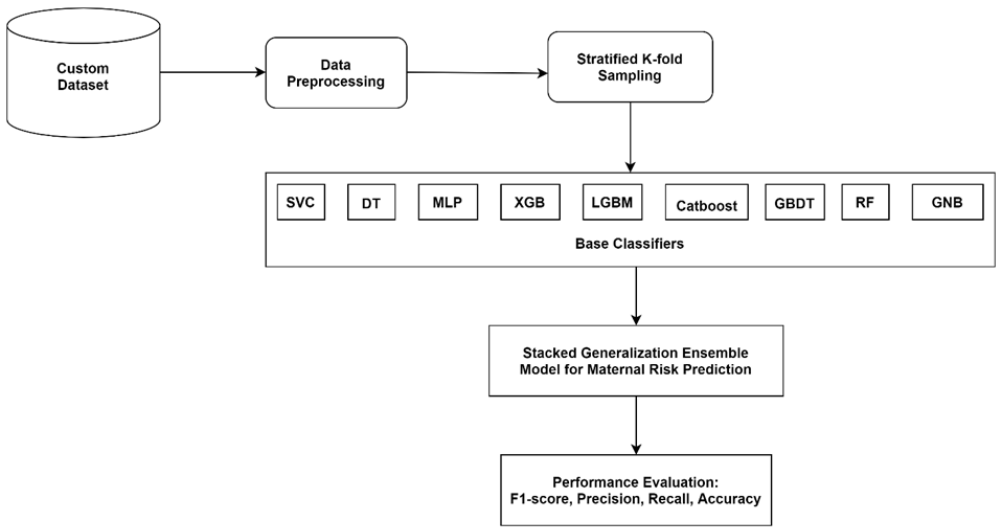

Fig. 4. Ensemble model design

The design includes a representation of all selected classifiers, including Random Forest (RF), Gradient Boosted Decision Tree (GBDT), Support Vector Classifier (SVC) (a linear variant of Support Vector Machine), Decision Tree (DT), Multilayer Perceptron (MLP) (a type of Artificial Neural Network), Extreme Gradient Boosting (XGB), Gaussian Naïve Bayes (GNB), Light Gradient Boosting Machine (LightGBM), and Categorical Boosting (CatBoost).To improve the overall performance beyond that of any single classifier, an ensemble model was constructed using these nine classifiers as base learners. The performance of the developed model was evaluated using key metrics including F1 score, precision, accuracy and recall.

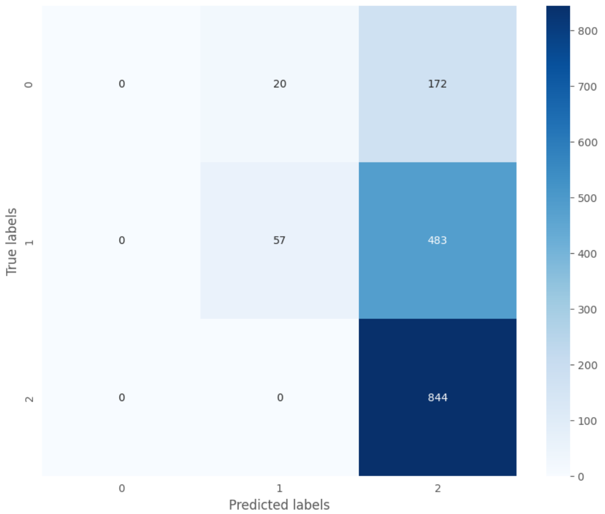

Fig. 5. Confusion matrix showing performance of the model before upsampling

Figure 5 shows the confusion matrix for the model's performance on the unbalanced dataset. The diagonal values represent correct predictions for each class of maternal health risk. The model achieved perfect accuracy for the low-risk class, correctly classifying all 844 cases. However, it completely failed to identify any high-risk cases, misclassifying 172 instances as low-risk and 20 instances as medium-risk. For the medium risk class, only 57 out of 540 cases were correctly predicted, with the remainder likely misclassified as low risk. These results highlight a significant bias toward the majority (low-risk) class, a common problem in unbalanced datasets. The inability of the model to accurately detect high-risk cases is of particular concern in clinical applications, as it may lead to misdiagnosis and increased patient risk. In addition, this biased performance skews the overall interpretation of model effectiveness and underscores the need to address class imbalance during training.

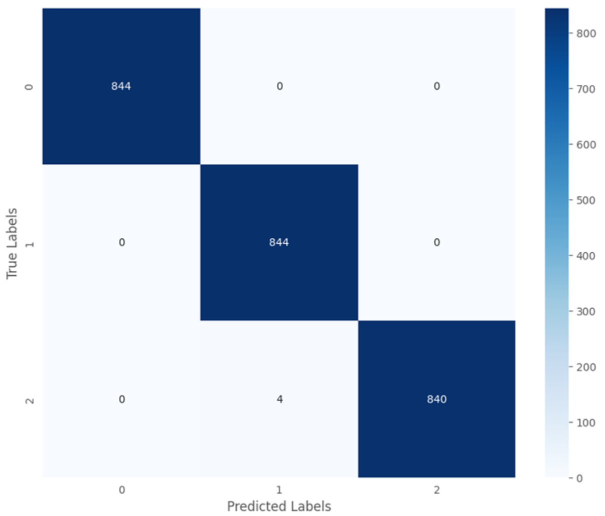

Fig. 6. Confusion matrix showing performance of the model after upsmapling

The confusion matrix in Figure 6 shows that there were no misclassifications at the high and low risk levels. Low risk pregnancies were correctly classified 840 times out of 844 instances. This model has over a 90% rate of correctly classified instances. The higher the accuracy of the model, the better its performance.

To better understand the behavior of the model and the predictive value of clinical features, the feature importance scores and SHAP (SHapley Additive Explanations) scores were examined in Figures 5 and 6 below. For feature importance, BMI is identified as the most important predictor, followed by maternal age, systolic blood pressure, and gestational age. These features consistently contribute high value to the learning process of the model. Diastolic blood pressure, weight, height, and parity also showed significant influence, while urine analysis and the artificial binary feature Non-Negative Urine Analysis showed minimal importance, indicating limited impact on the model's predictive decisions.

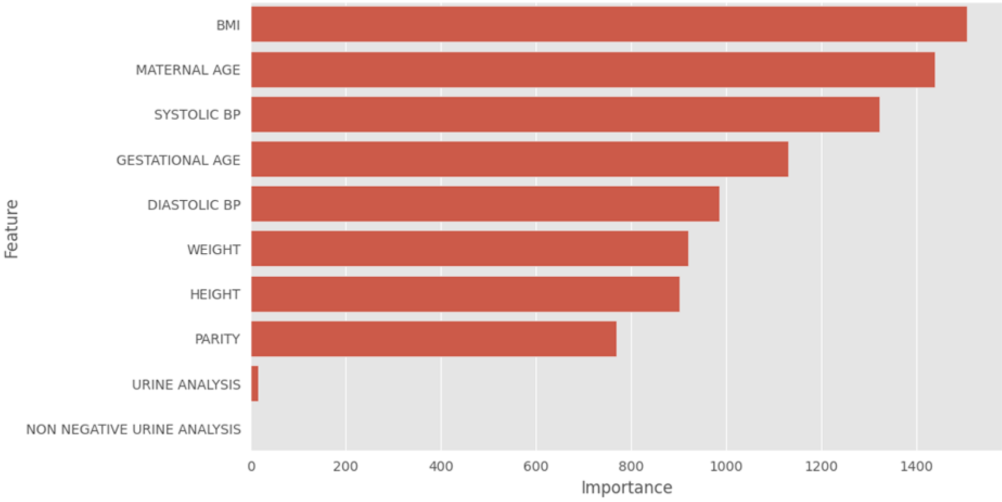

Fig. 7. Feature importance scores of the predictive model

For feature engineering, a new binary feature, Non-Negative Urine Analysis, was developed from the original Urine Analysis variable to improve model interpretability and risk classification. The rationale for creating this feature was to simplify the representation of complex urine test results into a clinically meaningful risk indicator. Incorporating this binary indicator allows for easier integration into machine learning models and aligns with clinical practice, where any abnormal urine result typically triggers closer monitoring and intervention. BMI is identified as the most important predictor, followed by maternal age, systolic blood pressure and gestational age. These characteristics consistently contribute high value to the learning process of the model. Diastolic blood pressure, weight, height, and parity also showed significant influence, while urine analysis and the artificial binary feature Non-Negative Urine Analysis showed minimal importance, indicating limited impact on the model's predictive decisions.

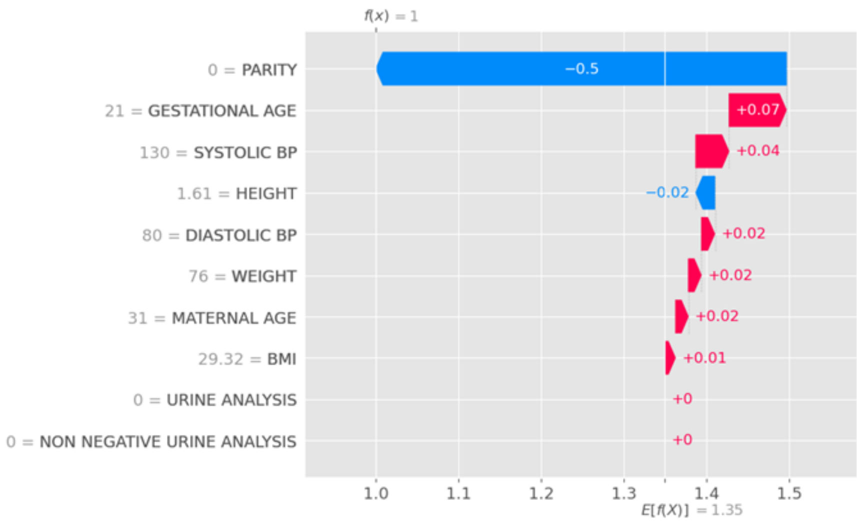

Fig. 8. SHapely additive explanations (SHAP) plot illustrating the contribution of data set features in the prediction model

The feature importance rankings in Figure 7 are consistent with the Shapely Additive Explanations (SHAP) results in Figure 8 from individual predictions, where features such as parity, gestational age, and systolic blood pressure showed significant directional influence (positive or negative) on the model output. Specifically, the SHAP analysis showed that Parity had the strongest negative contribution in one case, while

Gestational Age and Systolic BP pushed the prediction higher. Meanwhile, urine analysis and non-negative urine analysis had SHAP values close to zero, reinforcing their limited interpretive and predictive relevance.

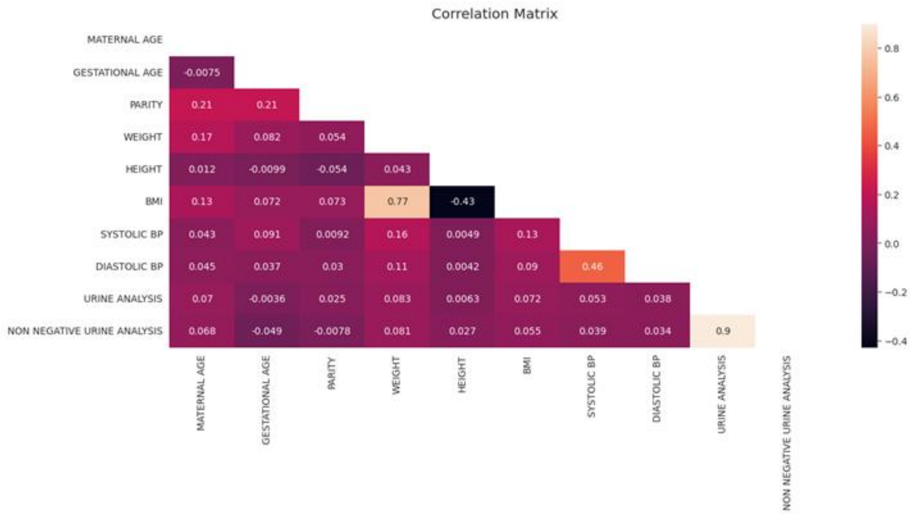

Fig. 9. Correlation analysis of dataset features

The Pearson correlation coefficient (r) in the correlation matrix provides insight into the strength and direction of linear relationships between clinical features in the maternal health dataset. The highest positive correlation is observed between urine analysis and non-negative urine analysis (r = 0.90), indicating a strong association and potential overlap in the information conveyed by these features. Similarly, BMI and weight are highly correlated (r = 0.77), consistent with the fact that BMI is calculated from weight and height. There is a moderate correlation between systolic and diastolic blood pressure (r = 0.46), reflecting their physiological relationship. In contrast, characteristics such as maternal age, parity, and gestational age have low or negligible correlations with other variables (ranging from approximately -0.05 to 0.21), suggesting that they contribute to the dataset independently and with minimal redundancy.

3. RESULTS PRESENTATION

This section presents the results of training and testing individual classifiers and the ensemble model. Table 6 summarizes the training results and Table 7 gives an overview of the testing results. Table 8 shows the performance of the ensemble model before upsampling. Table 9 shows the performance of the ensemble model after upsampling.

Tab. 6. Summary of training performance results of individual classifiers

Classifiers	F1_Score	Precision	Recall	Accuracy
Support Vector Classification (SVC)	0.308959	0.395001	0.369074	0.564936
Gaussian Naïve Bayes	0.675902	0.711995	0.71662	0.685548
Multilayer Perceptron (MLP)	0.629312	0.77701	0.619302	0.737177
Decision Tree	1	1	1	1
Random Forest	1	1	1	1
Gradient Boosting Decision Trees (GBDT)	0.947148	0.958796	0.936851	0.955654
Extreme Gradient Boosting (XGB)	1	1	1	1
Light Gradient Boosting Machine (LightGBM)	1	1	1	1
Categorical Boosting (CatBoost)	0.999212	0.999387	0.99904	0.999013

Tab. 7. Summary of testing performance results before upsampling of individual classifiers

Classifier	F1_Score	Precision	Recall	Accuracy
Support Vector Classification (SVC)	0.289605	0.338474	0.352927	0.541244
Gaussian Naïve Bayes	0.655454	0.707669	0.697896	0.670261
Multilayer Perceptron (MLP)	0.607645	0.698405	0.607162	0.720765
Decision Tree	0.789993	0.797832	0.790753	0.82553
Random Forest	0.842325	0.86714	0.834894	0.886427
Gradient Boosting Decision Trees (GBDT)	0.85694	0.873463	0.851552	0.897226
Extreme Gradient Boosting (XGB)	0.840721	0.85875	0.835387	0.879444
Light Gradient Boosting Machine (LightGBM)	0.842733	0.864576	0.834895	0.884528
Categorical Boosting (CatBoost)	0.839259	0.860744	0.83322	0.883254

Tab. 8. Performance results for stacked ensemble model before upsampling

Performance Metric	Result
F1_Score	0.975485
Precision	0.975137
Recall	0.977912
Accuracy	0.982166

Tab. 9. Performance results for stacked ensemble model after upsampling

Performance Metric	Result
F1_Score	0.998422
Precision	0.998485
Recall	0.998431
Accuracy	0.998419

4. DISCUSSION OF RESULTS

The training results in Table 6 showed that Decision Tree, Random Forest, Extreme Gradient Boosting, and Light Gradient Boosting Machine all achieved perfect scores (1.0) on all performance metrics evaluated. This indicates that these tree-based models have likely memorized the training data, a phenomenon common when working with relatively small datasets such as the 1,581 samples used in this study (Okpono et al., 2024). As shown in Table 7, the test results showed a slight drop in performance. This slight drop is expected, but can be partially attributed to the fact that the tests were not conducted on a truly independent dataset that the models had not seen before, which may have led to optimistic performance estimates (Mutlu et al., 2023).

In Table 7, the pre-upsampling test results showed that Gradient Boosting Decision Tree was the best performing algorithm with an accuracy of 89.72% and an F1 score of 85.62%. Support Vector Classification was the worst performing algorithm with an accuracy of 54.12% and an F1 score of 28.96%. The comparison between Table 8 and Table 9 shows that the up-mapping performed on the data significantly improved the performance of the ensemble.

As shown in Table 6 and Table 7, these results suggest that using bagged and boosted decision tree algorithms (such as Random Forest, Extreme Gradient Boosting, Light Gradient Boosting Machine, and Categorical Boosting) with the dataset used can improve model predictions of maternal mortality (Mienye & Sun, 2022; Kalirane, 2024).

The ensemble model outperformed all the individual classifiers, achieving an accuracy of 98.21% and an F1 score of 97.55%. These values are higher than those of the previously considered Gradient Boosting Decision Tree (accuracy of 89.72% and F1 score of 85.62%). This result demonstrates that the prediction of maternal mortality can be significantly improved by stacking individual classifiers into an ensemble model (Mienye & Sun, 2022; Khadka, 2024). The upsampled data as an ensemble performed significantly better than the ensemble without upsampling, achieving an accuracy and F1 score of 99.84% each.

5. CONCLUSIONS AND FUTURE WORKS

Maternal mortality remains a critical challenge in sub-Saharan Africa, with Nigeria among the countries with the highest rates. If Nigeria is to achieve the Sustainable Development Goals (SDGs) by 2030, there is a need to scale up interventions to reduce maternal mortality. In this study, machine learning was used to classify maternal health risk during pregnancy using nine (9) basic classifiers, namely, Linear Support Vector Machine, Gaussian Naïve Bayes, Multilayer Perceptron, Decision Tree, Random Forest, Gradient Boosting Decision Tree, Extreme Gradient Boosting, Light Gradient Boosting Machine, and Categorical Boosting. Accurate classification of maternal health risk can lead to individualized maternal care (especially for high-risk pregnancies), optimize the use of scarce resources, and reduce adverse pregnancy outcomes. Successful implementation and use of machine learning models should include knowledge of the site of application, therefore real-world data were collected at three (3) different health centers in Delta State during antenatal visits. Standardized data sets such as those used in a number of studies (Ahmed, 2023; Nirmala & Kambili, 2023; Mutlu et al., 2023; Mondal et al., 2023; Mustamin et al., 2023; Mashrafi et al., 2024) may not be a true reflection of the Nigerian population. This is not the case in this study, as real-world data were collected from local health centers and pre-processed for model development.

The accuracy results of the ensemble model are significantly higher than any of the individual classifiers considered. Implementation of the ensemble model can potentially optimize healthcare resources by targeting high-risk pregnant women, facilitate timely clinical decision making, and consequently improve maternal health outcomes. This model can also be adapted to other states in Nigeria. Future research will focus on expanding the dataset (and features) and implementing the model in real world health care settings.

Funding

No external funding was received for this research.

Acknowledgments

The authors would like to acknowledge the Delta State Government health centers for access to the Maternal Health data in selected health centers in the state, which formed the dataset used for this work. The authors also acknowledge Dr. Uniamikogbo Eugene Jafe for his valuable contribution in categorizing maternal health (pregnancy) risk into various levels.

Conflicts of interest

The authors declare no conflict of interest.

REFERENCES

Ahmed, M. (2020). *Maternal health risk*. UCI Machine Learning Repository. https://doi.org/10.24432/C5DP5D

Dawodi, M., Wada, T., & Baktash. J. A. (2020). Applicability of ICT, data mining and machine learning to reduce maternal mortality and morbidity: Case study Afghanistan. *International Information Institute*, *23*(1), 33-45.

Gomez-Jemes, L., Oprescu, A. M., Chimenea-Toscano, A., Lutgardo, G., & Romero-Tornero, M. (2022). Machine learning to predict pre-eclampsia and uterine growth restriction in pregnant women. *Electronics*, *11*(19), 3240. https://doi.org/10.3390/electronics11193240

Izugbara, C. O., Wekesah, F. M. & Adedini, S. A. (2016). Maternal health in Nigeria: A situation update. *African Population and Health Research Centre*. http://dx.doi.org/10.13140/RG.2.1.1291.9924

James, I., & Osubor, V. (2025). Machine learning evidence towards the eradication of malria burden: A scoping review. *Applied Computer Science*, *21*(1), 46-69. https://doi.org/10.35784/acs_6873

Jayatilake, S. M., & Ganegoda, G. (2021). Involvement of machine learning tools in healthcare decision making. *Hindawi Journal of Healthcare Engineering*, *2021*(1), 679512. https://doi.org/10.1155/2021/6679512

Khadka, N. (2024). *How stacking technique boost's machine learning model's performance*. Retrieved September 23, 2024 from https://dataaspirant.com/stacking-technique

Kopanitsa, G., Metsker, O., Paskochev, D. &, Greschischeva, S. (2021). Identification of risk factors and prediction of sepsis in pregnancy using machine learning. *Procedia Computer Science*, *193*, 393-401. https://doi.org/10.1016/j.procs.2021.10.040

Machoron, J. J. E., Inbaraj, C. N., & Jeng, X. A. (2022). A semi-supervised machine learning approach in predicting high-risk pregnancies in the Philippines. *Diagnostics*, *12*(11), 2782. https://doi.ord/10.3390/diagnostics/2112782

Mashrafi, S. S. A., Tafakori, L., & Abdollahian, M. (2024). Predicting maternal risk levels using machine learning models. *BMC Pregnancy and Child Health*, *24*, 820. https://doi.org/10.1186/s12884-024-07030-9

Kalirane, M. (2024). *Bagging, boosting and stacking: Ensemble learning in ml models*. Analitics Vidhya. https://www.analyticsvidhya.com/blog/2023/01/ensemble-learning-methods-bagging-boosting-and-stacking/

Mboya, I. B., Mahande, M. J., Mohanad, M., Obure, J., & Mwambi, H. G. (2020). Prediction of perinatal death using machine learning models: A birth registry-based cohort study in northern tanzania. *British Medical Journal Open*, *10*(10), e040132. https://doi.org/10.1136/bmjopen-2020-040132

Mienye, D. I., & Sun, Y. (2022). A survey of ensemble learning: Concepts, algorithms, applications and prospects. *IEEE Access*, *10*, 99129-99149. https://doi.org/10.1109/ACCESS.2022.3207287

Mondal, S., Nag, A., Barman, A. K., & Karmakar, M. (2023). Machine-learning based maternal health risk prediction for IoMT framework. *International Journal of Experimental Research and Review*, *32*, 145-159. https://doi.org/10.52756/ijerr.2023.v32.012

Mustamin, N. F., Aziz, F., Firmansyah, & Ishak, P. (2023). Classification of maternal health risk using three models Naïve Base method. *Indonesian Journal of Computing and Cybernetics Systems*,17(4), 395-404. https://doi.org/10.22146/ijccs.84242

Mutlu, H. B., Yucel, N., Durmaz, F., Cengil, E., & Yildirim, M. (2023). Prediction of maternal health risk with traditional machine learning methods. *MTU Journal of Engineering and Natural Sciences*, *4*(1), 16-23. https://doi.org/10.46572/naturengs.1293185

Nirmala, & Kambili, R. S. (2023). Maternal risk level prediction using ensemble model. *International Journal of Creative Research Thoughts*, 11(3),916-920.

O'Neil, A. (2023). *Total population in Nigeria from 2011 to 2021, by Gender (in millions)*. Statista. https://www.statista.com/statistics/967908/total-population-of-nigeria-by-gender/#:~:text=This%20statistic%20shows%20the%20total,to%20approximately%20107.83%20million%20inhabitants

Ogallo, W., Speakman, S., Akinwande, V., Vashrey, R. K., Walcott-Bryant, A., Wayua, C., Weldermarian, K., Mershon, C., & Orobaton, N. (2020). Identifying factors associated with neonatal mortality in sub-saharan Africa using machine learning. *medRxiv*. https://doi.org/10.1101/2020.10.14.20212225

Okpono, J., Asedegbega, J., & Sanyaolu, T. O. (2024). Enhancing the predictive accuracy of machine learning models for project management challenges: A comparative study of advanced techniques for handling imbalanced data. *Technix International Journal for Engineering Research*, *11*(9), 99129-99137. http://dx.doi.org/10.5281/zenodo.13817083

Olonade, O., Olawande, T. I., Alabi, O. J. & Imhonopi, D. (2019). Maternal mortality and maternal healthcare in nigeria: Implications for socio-economic development. *Open Access Macedonian Journal of Medical Sciences*, *7*(5), 849-855. https://doi.org/10.3889/oamjms.2019.041

Rosser, J. I., Aluri, K. Z., Kempinsky, A., Shannon, R., & Bendavid, E. (2022). The effect of healthcare worker density on maternal health service utilization in sub-saharan Africa. *The American journal of tropical medicine and hygiene*, *106*(3), 939-944. https://doi.org/10.4269/ajtmh.21-0727

Togunwa, T. O., Babatunde, A. O., & Abdullah, K. 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. https://doi.org/10.3389/frai.2023.1213436

United Nations Country Office in Nigeria.(2015). Delta State Development Performance. Health Sector Report, 1991 - 2013 https://www.undp.org/sites/g/files/zskgke326/files/migration/ng/UNDP_NG_DeltaState_Health_2015.pdf

Wahl, B., Cossy-Gantner, A., Germann, S., & Schwalbe, N. R. (2018). Artificial intelligence (AI) and global health: How can AI contribute to health in resource-poor settings?. *British Medical Journal Global Health 2018*, *3*, e000798. https://doi.org/10.1136/bmjgh-2018-000798

World Health Organisation. (2019). *Maternal Health*. http://www.who.int/maternal-health/en/

World Health Organisation. (2023). *Trends in maternal mortality 2000 to 2017: Estimates by WHO, UNICEF, UNFPA, World Bank Group and the United Nations Population Division: Executive summary*. http://apps.who.int/iris/handle/10665/327596