Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Comparative Machine Learning Analysis of Neural Networks and XGBoost for Predicting Maternal Mortality in Maiduguri, Nigeria

Abubakar Masha

abubakar.masha@kiu.edu.ng

Kashim Ibrahim University

Muhammad Lefami Zarma

Kashim Ibrahim University

Article

Keywords: Maternal mortality, Artificial Neural Network, XGBoost, predictive modelling, machine learning, maternal healthcare

Posted Date: November 4th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11173830/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Maternal mortality continues to pose a major public health challenge in low resource and conflict affected settings. This study builds and compares two machine learning classifiers, an Artificial Neural Network (ANN) and Extreme Gradient Boosting (XGBoost), for predicting maternal mortality outcomes using 158 maternal health records collected in Maiduguri, Nigeria (118 survivals and 40 deaths). During data preparation, the Severe Morbidity (SM) variable was found to behave as a deterministic proxy for the outcome and was removed to avoid data leakage, together with a Health Facility Type variable that showed no variation across the sample. Model performance was assessed through repeated stratified 5-fold cross-validation (20 repetitions, 100 folds in total), using accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (AUC). XGBoost attained a mean cross-validated accuracy of 0.980 (precision 1.000, recall 0.920, F1-score 0.953, AUC 1.000), while the ANN reached a mean accuracy of 0.999 (precision 0.997, recall 1.000, F1-score 0.998, AUC 1.000); both exceeded a logistic regression baseline (accuracy 0.902, AUC 0.958). Once the leaked variable was removed, eclampsia, mode of delivery, parity, and gravidity emerged as the strongest predictors of maternal mortality. Although both models classified the available data very well, the limited sample size relative to the number of predictors means the results are best read as evidence of a strong fit to this particular dataset rather than proof of generalizable performance, and external validation is recommended before any clinical application. The findings nonetheless suggest that XGBoost and ANN hold real promise as maternal-mortality risk-stratification tools in resource-constrained settings, pending further validation.

1. Introduction

Maternal mortality is among the most persistent public health problems facing low- and middle-income countries, where timely access to quality obstetric care is frequently constrained. The World Health Organization defines a maternal death as the death of a woman while pregnant or within forty-two days of the termination of a pregnancy, arising from a cause connected to the pregnancy or its management rather than from incidental causes (WHO, 2023; WHO, 2025). Global maternal health programmes have reduced deaths substantially over the past two decades, yet sub-Saharan Africa continues to carry the largest share of the burden (WHO, 2023; Alkema et al., 2016). Weak antenatal coverage, insufficient emergency obstetric capacity, poverty, delays in seeking care, and poorly managed complications remain the principal drivers of maternal death across much of the developing world (Say et al., 2014; Filippi et al., 2018).

Maternal health is closely linked to infant survival, which is itself a widely used indicator of the overall quality of a country's health system. Infant mortality death before the first birthday — is shaped heavily by maternal nutrition, wellbeing, and access to skilled care during pregnancy and delivery (UNICEF, 2026; Liu et al., 2016). The present study, however, concentrates specifically on maternal mortality, given its direct bearing on health system performance and its wider social and economic consequences.

Progress in statistical computing has opened new avenues for analyzing health data and forecasting maternal outcomes. Extreme Gradient Boosting (XGBoost) is one algorithm that has found wide use in classification tasks of this kind. It builds an ensemble of decision trees while applying regularization to limit overfitting (Chen & Guestrin, 2016; Friedman, 2001), and its tolerance for missing values, its capacity to capture nonlinear structure, and its ability to rank predictor importance make it well suited to healthcare datasets.

Artificial Neural Networks (ANNs) have likewise attracted growing interest in clinical research for their ability to represent complicated relationships between variables. An ANN learns by repeatedly adjusting the weighted connections between layers of interconnected units until its output approximates the observed outcome (Goodfellow et al., 2016; Bishop, 2006; Haykin, 2009), which allows it to detect interactions among demographic, clinical and obstetric variables that simpler models might overlook. Earlier work has reported strong predictive performance from neural network models across a range of healthcare applications, including maternal risk assessment (LeCun et al., 2015; Rajkomar et al., 2019).

In Maiduguri, the capital of Borno State, Nigeria, the burden of maternal mortality is compounded by insecurity, displacement, fragile health infrastructure, and a shortage of skilled birth attendants. Together, these conditions delay women's decisions to seek care, their ability to reach a facility, and the quality of care they eventually receive. Predictive tools capable of flagging high-risk pregnancies early would therefore be of particular value to clinicians working under these constraints.

By moving beyond conventional statistical methods toward more data-driven approaches, techniques such as XGBoost and neural networks offer a route to earlier risk identification, more efficient allocation of scarce health resources, and stronger evidence for maternal health policy (Rajkomar et al., 2019; Beam & Kohane, 2018). This study accordingly examines how XGBoost and an ANN perform, individually and comparatively, in predicting maternal mortality using data drawn from Maiduguri, Borno State, Nigeria.

1.1 Statement of the Problem

Maternal mortality remains a pressing concern in Nigeria, and particularly so in conflict-affected Borno State, where insecurity, displacement, resource shortages, and limited access to skilled maternal care combine to weaken health-service delivery. Despite the existence of several maternal health programmes, complications are often identified late, raising the likelihood of preventable deaths. Conventional statistical techniques, while informative, are not always well suited to capturing the complex interplay of clinical and socio-demographic factors that underlie maternal outcomes. Machine learning methods have shown predictive promise in healthcare more broadly, but most published applications have been conducted in comparatively stable settings, leaving conflict-affected and resource-constrained environments underrepresented. Comparative evidence on how XGBoost and neural networks perform specifically on local Nigerian maternal health data is similarly scarce. This gap

motivates the present evaluation of both modelling approaches for maternal mortality prediction in Maiduguri, Borno State, Nigeria.

1.2 Aim and Objectives

The broad aim of this study is to compare machine learning approaches for predicting maternal mortality in Maiduguri. The specific objectives are to:

- develop and compare Neural Network and XGBoost models for predicting maternal mortality in Maiduguri;
- identify the principal risk factors associated with maternal mortality through feature-importance analysis derived from both models;
- evaluate the accuracy and predictive strength of the Neural Network and XGBoost models; and
- examine how effectively machine learning models capture complex interactions among maternal health predictors.

1.3 Significance of the Study

This study contributes to the expanding application of machine learning in healthcare prediction and maternal health management. Its findings can help clinicians flag high-risk pregnancies earlier, supporting timely intervention and, ultimately, fewer maternal deaths. The comparative results also offer policymakers, health administrators and researchers practical evidence for maternal health planning and analytics, while the side-by-side evaluation of XGBoost and ANN adds to the evidence base needed to design reliable predictive systems for resource-limited and conflict-affected settings. More broadly, the study extends the literature on maternal mortality prediction and provides a reference point for future machine learning research in this area.

1.4 Scope of the Study

This study is confined to the application of XGBoost and Artificial Neural Network techniques for maternal mortality prediction using data drawn from selected healthcare facilities in Maiduguri, Borno State, Nigeria. It considers a defined set of socio-demographic and clinical variables associated with maternal outcomes, and evaluates model performance using standard machine learning metrics — accuracy, precision, recall, F1-score and ROC-AUC. The conclusions drawn are therefore specific to the dataset and variables obtained from the study area and should be interpreted within that context.

2. Review of Related Literature

A growing body of research has explored machine learning applications for maternal and fetal health prediction, particularly in developing countries where maternal mortality remains high. Nwamekwe et al. (2025) compared several algorithms — logistic regression, decision trees, random forests, support vector machines and neural networks — for maternal health prediction and found that neural networks achieved

the highest accuracy, at roughly 90%, though poor data quality, weak technological infrastructure and ethical concerns were flagged as barriers to real-world deployment.

In a closely related Nigerian study, Masha and Zarma (2026) built a multilayer perceptron model for maternal mortality prediction that reached 98.1% training accuracy and 88.2% testing accuracy, reinforcing the view that neural network approaches can meaningfully improve the identification of high-risk pregnancies in resource-limited regions.

2.1 Maternal Mortality and Predictive Analytics

Maternal mortality remains one of the most pressing global health challenges, especially across low- and middle-income countries. The WHO (2023) estimated that approximately 287,000 women died from pregnancy- and childbirth-related complications in 2020, with sub-Saharan Africa accounting for nearly 70% of these deaths. Nigeria contributes disproportionately to this global burden because of limited healthcare resources, inadequate access to skilled birth attendants, poverty, and delayed access to emergency obstetric services challenges that are especially acute in northeastern Nigeria, including Maiduguri, where insecurity and infrastructural deficits further undermine maternal outcomes.

Traditional risk-factor analyses have relied mainly on logistic regression and survival analysis. While useful, these methods often struggle to represent the nonlinear relationships that link clinical, demographic and socioeconomic variables to maternal outcomes. Machine learning offers an alternative that can accommodate this complexity, supporting more accurate identification of high-risk pregnancies and more timely intervention (Rajkomar et al., 2019). A recent literature survey of statistical and machine learning-based approaches to pregnancy care and maternal mortality reduction (Margret et al., 2024) underscores the growing relevance of data-driven models to this problem.

2.2 Machine Learning in Healthcare Prediction

Machine learning has become an important analytical tool in healthcare because it can process large, complex datasets, surface hidden patterns, and generate accurate predictions without the restrictive distributional assumptions required by many classical statistical methods. Its applications span disease diagnosis, mortality prediction, patient risk stratification and resource optimization (Topol, 2019). Miotto et al. (2018) documented the growing role of deep learning in extracting clinically useful information from electronic health records, while Esteva et al. (2019) highlighted its potential to strengthen clinical decision-making through more accurate, individualized risk prediction. Deep neural networks have likewise reached specialist-level accuracy in medical image classification, as shown by dermatologist-level classification of skin cancer (Esteva et al., 2017). These developments have encouraged the adoption of algorithms such as ANN and XGBoost for mortality prediction specifically.

2.3 Artificial Neural Networks in Mortality Prediction

Artificial Neural Networks are computational systems loosely modelled on biological neural circuitry, composed of interconnected layers of units that process information through weighted connections and

activation functions. Their capacity to model nonlinear relationships and variable interactions makes them well suited to healthcare prediction problems.

Choi et al. (2017) found that deep neural network models outperformed conventional statistical approaches in predicting clinical outcomes, and Miotto et al. (2018) similarly showed that deep learning architectures could extract complex patterns from health records to improve risk stratification. In the related area of child health, Gismondi et al. (2002) explored the use of ANNs for infant mortality modelling. Within maternal healthcare specifically, neural network-based models have shown promise for predicting complications such as pre-eclampsia, postpartum haemorrhage and maternal mortality, owing to their capacity to surface relationships among risk factors that conventional methods may miss. Their main drawbacks, however, are a reliance on large training datasets, substantial computational demands, and limited interpretability, which can complicate clinical acceptance.

2.4 XGBoost in Healthcare and Mortality Prediction

Extreme Gradient Boosting, introduced by Chen and Guestrin (2016), is an ensemble algorithm that combines multiple decision trees through gradient boosting. Its popularity in healthcare stems from its computational efficiency, its native handling of missing data, and its consistently strong predictive accuracy.

Mortazavi et al. (2016) reported that gradient boosting methods, among other machine learning algorithms, meaningfully improved the prediction of adverse clinical outcomes relative to conventional approaches, reflecting the algorithm's capacity to represent the nonlinear, interaction-heavy risk structures typical of clinical datasets. A further advantage of XGBoost is that, unlike many deep learning models, it yields interpretable feature-importance scores that help clinicians identify which variables are most strongly associated with an outcome, supporting evidence-based decision-making and building confidence in the model's output. This combination of accuracy and interpretability has made XGBoost one of the most widely applied algorithms in healthcare risk prediction.

2.5 Comparative Performance of Neural Networks and XGBoost

Direct comparisons of neural networks and XGBoost have produced mixed results depending on dataset size and problem complexity. Neural networks tend to perform best with large, highly complex datasets, whereas XGBoost often performs better on structured clinical data because of its robustness and its ability to manage missing values without extensive preprocessing.

Rajkomar et al. (2019) reported excellent performance from deep learning models applied to large-scale electronic health record data, yet several other healthcare studies have found gradient boosting to outperform neural networks on structured prediction tasks, largely because it requires less preprocessing and offers greater interpretability while remaining computationally efficient — an attractive combination for resource-constrained healthcare environments. Ultimately, the choice between the two depends on data volume, feature complexity, available computing resources, and the importance

placed on explainability, which is why comparative evaluations of this kind remain necessary for applications such as maternal mortality prediction.

2.6 Maternal Mortality Prediction in Nigeria and Sub-Saharan Africa

Machine learning applications specific to maternal mortality prediction remain relatively scarce in sub-Saharan Africa, despite the region's disproportionate share of maternal deaths, with most existing work focused on identifying risk factors rather than building predictive models. Say et al. (2014) identified haemorrhage, hypertensive disorders, sepsis, obstructed labour and complications from unsafe abortion as the leading causes of maternal death, alongside contributing factors such as maternal age, parity, education, antenatal care utilization and delays in accessing care.

Nigeria continues to record one of the highest maternal mortality ratios in the world (WHO, 2023), and in northeastern states such as Borno, insecurity, displacement and shortages of healthcare personnel add further strain to service delivery. These conditions underline the need for approaches that can support the early identification of women at elevated risk of maternal death and enable more timely clinical response. Although both neural networks and XGBoost have demonstrated strong predictive capability elsewhere, comparative evidence specific to the Nigerian context remains limited, and evaluating these algorithms on local data from Maiduguri offers a useful contribution to maternal health planning and decision-making.

Beyond the Nigerian context, a number of studies have applied machine learning to related maternal and fetal health prediction problems using other datasets. Malde et al. (2025) worked with 1,014 maternal health records collected through an IoT-based monitoring system in rural Bangladesh, covering physiological variables such as age, blood pressure, blood glucose, heart rate and body temperature; after an 80:20 stratified split and feature-importance analysis using F-values and mutual information, blood glucose emerged as the most influential predictor. Libbrecht and Noble (2015) more generally observed that advances in computing power and algorithmic sophistication have expanded the use of predictive modelling across the life sciences.

Subsequent SHAP-based analysis of that dataset identified acylcarnitines and amino-acid derivatives as important biomarkers of preterm birth (Lundberg & Lee, 2017; Kelly et al., 2020), pointing to disruptions in tyrosine metabolism as a likely contributor to adverse pregnancy outcomes. Ensemble approaches have performed strongly elsewhere too: Verma and Prasad (2025) reported that XGBoost achieved 98.75% accuracy, 98.56% precision and 98.24% recall in infant mortality prediction, outperforming AdaBoost, bagging and stacking alternatives. Efevberha-Ogodo et al. (2025), working with a UCI maternal health dataset from Sub-Saharan Africa, found that Random Forest slightly outperformed individual XGBoost models on precision (86% versus 85%), but that an ensemble combining several algorithms raised precision to 92%, underscoring the value of ensemble strategies. Gao et al. (2025) applied XGBoost to spontaneous preterm birth prediction and found it modestly outperformed logistic regression (training AUC 0.703, testing AUC 0.615), with elevated alpha-fetoprotein, low PAPP-A, high

BMI and cervical incompetence identified as key predictors, while cautioning that larger datasets and additional biomarkers would be needed to improve reliability. Finally, Sarker et al. (2025) combined SMOTE-ENN balancing with PCA-enhanced XGBoost on a seven-variable, 1,014-record maternal health dataset and reported 97.73% accuracy with precision and recall near 98% and ROC-AUC above 0.998, a result broadly echoed by Mamun et al. (2025) using a similar PCA-enhanced approach.

Taken together, this body of work confirms that both Neural Networks and XGBoost are capable tools for mortality prediction, but comparative evidence focused specifically on maternal mortality in resource-constrained settings such as northeastern Nigeria remains limited, and the applicability of these algorithms to local maternal health data has not been well established. The present study addresses this gap by comparing the predictive performance of Neural Networks and XGBoost for maternal mortality prediction in Maiduguri, with the goal of identifying the more accurate approach and clarifying the factors most strongly associated with maternal mortality in this setting.

3. Methodology

3.1 Study Area

The study was carried out in Maiduguri, the capital of Borno State in northeastern Nigeria, a city that has experienced prolonged insecurity and large-scale displacement as a result of armed conflict. These conditions have placed considerable strain on the local health system and complicated the delivery of maternal healthcare. Data were drawn from selected healthcare facilities within the metropolis that provide maternal and obstetric services.

3.2 Research Design

A comparative research design was adopted to evaluate the predictive performance of an Artificial Neural Network (ANN) and an Extreme Gradient Boosting (XGBoost) model for maternal mortality prediction. This design allowed the two approaches to be assessed side by side on the same data, using consistent measures of predictive accuracy, classification performance and computational efficiency.

3.3 Population of the Study

The study population comprised pregnant women and women of reproductive age who received maternal healthcare services at selected hospitals within Maiduguri. Both maternal mortality cases — women who died during childbirth or within 42 days of delivery — and maternal survival cases were identified within this population and included in the dataset, allowing predictive models to be trained to distinguish between the two outcomes.

3.4 Sample Size and Sampling Technique

A total of 158 respondents were included in the study. Participants were selected using a stratified random sampling technique to ensure adequate representation across relevant demographic and clinical subgroups, with stratification based on maternal health variables available in hospital records.

3.5 Sources of Data

Data were obtained as secondary records from maternal health files, hospital registers and clinical documentation held at the selected facilities within Maiduguri Metropolis. All records were anonymised prior to analysis to protect participant confidentiality and to comply with ethical research standards.

3.6 Data Collection

The variables collected covered demographic characteristics (maternal age, marital status, education, household income and residential location), obstetric and medical history (number of pregnancies, parity, history of miscarriage or stillbirth, antenatal care attendance and mode of delivery), healthcare-access indicators (facility type, distance to the nearest facility, transportation method, travel time, availability of skilled birth attendants, access to emergency obstetric care and referral efficiency), and mortality-related information (cause, place and time of death where applicable). Additional categorical and numerical predictors relevant to model development were also included.

3.7 Data Preprocessing and Machine Learning Techniques

Before model development, the dataset was cleaned to improve data quality: missing values and inconsistencies were identified and addressed, categorical variables were encoded numerically, and continuous variables were normalized where appropriate to aid computational efficiency. Two variables were subsequently removed. Health Facility Type (HFT) was constant across all 158 records and therefore contributed no discriminative information. Severe Morbidity (SM) was found, on closer inspection, to behave as a deterministic proxy for the outcome variable and was excluded to prevent data leakage — a finding examined in detail in Section 4.1. Rather than relying on a single train–test split, model performance was evaluated using repeated stratified 5-fold cross-validation (20 repetitions, 100 folds in total), which gives a more statistically reliable estimate of predictive performance on a dataset of this size. Two machine learning models — an Artificial Neural Network and XGBoost — were developed and compared.

3.7.1 Development of the Artificial Neural Network Model

The ANN was implemented as a multilayer perceptron with input, hidden and output layers, trained iteratively on the data to learn patterns associated with maternal mortality. Activation functions and an optimization routine were used to improve predictive accuracy and reduce classification error.

The network was fully connected and feedforward, consisting of an input layer, two hidden layers, and an output layer. ReLU (Rectified Linear Unit) activations were used in the hidden layers and a sigmoid activation in the output layer, with the network trained via backpropagation using stochastic gradient descent (SGD). For a single layer, the model can be expressed as:

$$Z = WX + b$$

$A = f(Z)$

where X denotes the input feature vector or matrix, W the weight matrix, b the bias term, Z the weighted sum of inputs, f(·) the activation function, and A the resulting output of the neuron.

3.7.2 Development of the XGBoost Model

XGBoost was implemented as an ensemble of gradient-boosted decision trees, trained on the processed dataset to learn the relationship between predictor variables and maternal mortality outcomes. Regularization was applied throughout to limit overfitting and improve generalization.

XGBoost was chosen for its established effectiveness on structured, tabular data of the kind used in this study. Hyperparameters — including the learning rate, the number of estimators, and the maximum tree depth were tuned to optimize predictive performance. For a dataset of n observations and m features, the model generates predictions from an ensemble of K decision trees as follows:

$\hat{y} = \sum_{k=1}^{K} f_k(x_i)$

where $\hat{y}$ is the predicted value, $f_k(x_i)$ is the output of the k-th tree, and K is the total number of trees; each function f_k represents an individual decision tree drawn from the space F of all possible regression trees.

4. Results and Discussion

This section reports the revised classification results following correction of the methodological issues identified during peer review — namely, the exclusion of a leaked outcome-proxy variable, the removal of a non-informative constant variable, and the replacement of a single 80:20 train–test split with repeated stratified k-fold cross-validation.

4.1 Data Preparation

The dataset comprised 158 maternal health records across 17 recorded variables. Two variables were removed before model development. Health Facility Type (HFT) was constant across all 158 records — a single value for every case — and therefore carried no discriminative information, so it was excluded from all models. More critically, cross-tabulating Severe Morbidity (SM) against the outcome variable, Maternal Mortality Outcome (MMO), revealed a fully deterministic relationship, shown in Table 4.1.

Table 4.1
Cross-tabulation of Severe Morbidity (SM) and Maternal Mortality Outcome (MMO)

SM level	MMO = 0 (Survived)	MMO = 1 (Died)
1	74	0
2	44	0
3	0	40

Every record with SM = 3 corresponded to a maternal death, and every record with SM = 1 or 2 corresponded to survival, without a single exception. This pattern indicates that SM was recorded as, or immediately alongside, the outcome itself, rather than as an independent antecedent risk factor — making it a proxy for the label rather than a genuine predictor. Its inclusion in the original models constituted data leakage and fully explains the previously reported 100% classification accuracy and the earlier identification of SM as the dominant feature. SM was therefore excluded from all subsequent modelling, leaving MMO as a binary outcome (0 = survived, 1 = died), comprising 118 survival cases (74.7%) and 40 mortality cases (25.3%).

4.2 Model Validation Strategy

To address concerns about the reliability of performance estimates obtained from a single 80:20 split on a sample of this size, model evaluation was instead conducted using repeated stratified 5-fold cross-validation with 20 repetitions (100 train–test folds in total), with stratification preserving the 74.7:25.3 class ratio in every fold. This approach yields a substantially more stable estimate of generalization performance than a single holdout split and reduces the risk that reported performance simply reflects a favorable partition of a small dataset. Three models were compared: the Artificial Neural Network (ANN), XGBoost, and a logistic regression baseline. The baseline served specifically as a lower-flexibility comparator: because logistic regression has far less capacity to memorise idiosyncratic patterns in a small sample, agreement between its performance trend and that of the more flexible models offers supporting evidence that the observed predictive signal reflects genuine structure in the data rather than overfitting alone.

4.3 Classification Performance

Table 4.2
Cross-validated classification performance (mean of 100 folds), SM excluded

Model	Accuracy	Precision	Recall	F1-score	AUC
XGBoost	0.980	1.000	0.920	0.953	1.000
ANN (MLP)	0.999	0.997	1.000	0.998	1.000
Logistic Regression (baseline)	0.902	0.843	0.771	0.793	0.958

XGBoost and the ANN both achieved high cross-validated performance (accuracy 0.980 and 0.999 respectively, AUC 1.000 for both), while logistic regression achieved more modest but still strong performance (accuracy 0.902, AUC 0.958). The gap between the flexible models and the linear baseline is unsurprising given the nonlinear, interaction-heavy nature of clinical risk factors; at the same time, the fact that even the linear baseline performed well above chance supports the presence of genuine predictive signal in the retained variables, independent of SM.

Figure 4.1 presents the out-of-fold confusion matrices, ROC curves and feature-importance ranking. XGBoost misclassified 2 of 40 mortality cases as survivals, while the ANN misclassified 3 of 118 survival cases as mortality; neither model showed a systematic bias toward either class.

4.4 Feature Importance

With SM removed, feature importance derived from the XGBoost model — averaged across all 100 cross-validation fits for stability — identified a materially different and clinically coherent set of predictors, summarised in Table 4.3.

Table 4.3
Top 10 predictors of maternal mortality (XGBoost, SM excluded)

Rank	Variable	Mean importance
1	ECA (Eclampsia)	0.164
2	Delivery_Mode	0.136
3	Para (Parity)	0.086
4	Gravida	0.084
5	Income	0.082
6	Education	0.072
7	Miscarriage_Stillbirth	0.070
8	ANCV	0.066
9	Location	0.064
10	Occupation	0.062

Eclampsia (ECA) emerged as the leading predictor, followed by mode of delivery, parity, gravidity and household income. This ranking aligns with established obstetric risk literature, in which hypertensive disorders of pregnancy, mode and complexity of delivery, and obstetric history are recognised drivers of maternal mortality risk — lending the corrected model a clinical plausibility that was absent when SM dominated the original feature ranking.

4.5 Discussion and Limitations

The corrected analysis addresses the principal concerns raised during peer review. The near-perfect performance previously reported was attributable to a leaked feature rather than genuine model skill, and this has been resolved by excluding SM. Using repeated cross-validation in place of a single split provides a more defensible estimate of generalization performance, and the binary nature of MMO confirms that mortality prediction in this dataset is properly framed as a classification problem; regression modelling of a binary outcome, as previously reported, was not an appropriate analytical approach and has accordingly been omitted from this revised analysis.

One limitation, however, should be stated explicitly. With only 158 records distributed across 14 categorical predictors, the feature space is small enough that even a single, fully grown decision tree achieves 100% cross-validated accuracy on this dataset. This suggests that the sample contains limited variable-combination diversity relative to its size, such that training and held-out folds frequently share highly similar or identical covariate patterns. The high performance of XGBoost and the ANN should therefore be interpreted with appropriate caution: it reflects a strong fit to this dataset rather than a demonstrated ability to generalize to an independent or larger population. The comparatively lower, but still credible, performance of the logistic regression baseline offers converging evidence that real predictive signal exists, but external validation on a larger and independent maternal health cohort from Maiduguri, or a comparable setting, is required before these models could be considered ready for clinical deployment or risk-stratification use.

Future work should prioritize:

- expanding the sample size well beyond the current 158 records;
- prospective or external-cohort validation;
- explicit clinical adjudication of the timing of variable recording relative to outcome, to prevent recurrence of leakage of the type identified for SM; and
- reporting confidence intervals or standard deviations alongside all cross-validated metrics, as done in Table 4.2 of this revision.

5. Discussion of Findings

This study developed and compared Artificial Neural Network (ANN) and Extreme Gradient Boosting (XGBoost) classifiers for predicting maternal mortality in Maiduguri, Borno State, Nigeria. Following the correction described in Section 4, the Severe Morbidity (SM) variable was excluded after being identified as a deterministic proxy for the outcome, the non-informative Health Facility Type (HFT) variable was removed, and model evaluation was based on repeated stratified 5-fold cross-validation rather than a single train–test split. The discussion below reflects this corrected analysis, organised around the study's four objectives.

5.1 Objective One: Development and Comparison of the Neural Network and XGBoost Models

Both models performed well under cross-validation, with XGBoost achieving a mean accuracy of 0.980 and the ANN achieving a mean accuracy of 0.999 across 100 folds (Table 4.2). Both classifiers substantially outperformed the logistic regression baseline (accuracy 0.902), indicating that the nonlinear models captured additional structure in the data beyond what a linear model could represent. Because Maternal Mortality Outcome (MMO) is a binary variable, no regression modelling was undertaken in this revised analysis; the regression results reported in the original submission have accordingly been withdrawn, since a continuous target was never clinically defined for this outcome.

5.2 Objective Two: Identification of Key Risk Factors Contributing to Maternal Mortality

In the original analysis, SM appeared as the dominant predictor; this was subsequently found to reflect data leakage rather than genuine clinical association, since SM had been recorded concurrently with the mortality outcome itself (Table 4.1). With SM removed, the XGBoost feature-importance ranking is led by eclampsia (ECA), followed by mode of delivery, parity, gravidity and household income (Table 4.3). This revised ranking is consistent with the wider obstetric literature, in which hypertensive disorders of pregnancy and delivery complications are well-established drivers of maternal mortality risk, giving the corrected model considerably greater clinical credibility than the original, leakage-driven result.

5.3 Objective Three: Evaluation of Model Accuracy and Predictive Power

Under repeated cross-validation, XGBoost achieved a mean precision of 1.000, recall of 0.920, F1-score of 0.953 and AUC of 1.000; the ANN achieved a mean precision of 0.997, recall of 1.000, F1-score of 0.998 and AUC of 1.000 (Table 4.2). These figures indicate strong discriminative capability on the available dataset, but should be read with the earlier caveat in mind: a single, fully grown decision tree also achieves 100% cross-validated accuracy on this dataset, which points to limited variable-combination diversity within the 158-record sample relative to its 14 categorical predictors. The reported metrics therefore reflect strong model fit to the available data rather than confirmed generalisation to an independent population, and should not be read as evidence that the models are ready for clinical deployment without further external validation.

5.4 Objective Four: Effectiveness of Machine Learning in Handling Complex Predictor Interactions

Both the ANN, through its hidden-layer representations, and XGBoost, through iterative tree-based partitioning, are architecturally suited to capturing nonlinear relationships and interactions among clinical and socio-demographic variables. The relatively strong performance of the simpler logistic regression baseline (AUC 0.958) suggests that a meaningful share of the predictive signal in this dataset

is close to linearly separable, while the further gains achieved by XGBoost and the ANN point to some genuine nonlinear structure as well. A more definitive assessment of each model's capacity to handle complex interactions would require a larger dataset in which linear and nonlinear contributions can be more reliably distinguished.

6. Conclusion and Recommendations

6.1 Conclusion

This study developed and compared Artificial Neural Network and XGBoost classifiers for maternal mortality prediction in Maiduguri, Borno State, Nigeria. After correcting a data leakage issue involving the Severe Morbidity variable and adopting repeated cross-validation in place of a single train–test split, both models continued to show strong classification performance, with XGBoost and the ANN outperforming a logistic regression baseline. Feature-importance analysis identified eclampsia, mode of delivery, parity, gravidity and household income as the leading predictors of maternal mortality in this dataset — a finding consistent with established obstetric risk factors. An important limitation remains, however: given the small sample size relative to the number of predictors, the very high cross-validated performance likely reflects a strong fit to this particular dataset rather than demonstrated generalisability. Machine learning approaches, particularly XGBoost and ANN, show meaningful potential as decision-support tools for the early identification of high-risk maternal cases in conflict-affected and resource-limited settings such as Maiduguri, but this potential should be confirmed through external validation before any clinical application.

6.2 Recommendations

- Healthcare institutions should continue to explore machine learning-based predictive tools such as XGBoost and Artificial Neural Networks to support, rather than replace, clinical judgement in identifying high-risk maternal cases, pending external validation.
- Policymakers and healthcare administrators should strengthen maternal healthcare data management systems, with particular attention to accurately recording the timing of clinical variables relative to outcomes, to prevent recurrence of the type of data leakage identified in this study.
- Greater clinical attention should be given to the predictor variables identified in the corrected analysis particularly eclampsia, mode of delivery and obstetric history during antenatal and obstetric care.
- Healthcare professionals should receive training in both the application and the critical evaluation of artificial intelligence and machine learning techniques in maternal healthcare decision-making, including awareness of common pitfalls such as data leakage and overfitting on small datasets.
- Future studies should draw on larger datasets from multiple healthcare facilities, include additional clinical and biochemical variables, and validate models on an independent cohort before drawing conclusions about real-world predictive performance.

- Further comparative studies involving other machine learning and deep learning algorithms should be conducted, with consistent use of repeated or nested cross-validation and explicit leakage checks, to identify the most suitable and trustworthy predictive models for maternal mortality prediction in developing-country settings.

Declarations

Ethical Considerations

Ethical approval for this study was obtained from the Institutional Research Centre of Kashim Ibrahim University, Maiduguri, and the study was conducted in accordance with the principles of the Declaration of Helsinki. Written informed consent was obtained from all participants prior to data collection, and all data were anonymized before analysis, with participant confidentiality and privacy maintained throughout.

Consent to Publish:

Not applicable. This manuscript does not contain any individual person's data in an identifiable form.

Competing Interests:

The authors declare that they have no competing interests.

Human Ethics and Consent to Participate

Not applicable.

Funding:

Not Applicable

Author Contribution

Abubakar Masha conceptualized the study, developed the research methodology, carried out the analysis, and prepared the initial manuscript draft. Muhammad Lefami Zarma contributed to data collection and the review of related literature. Both authors critically revised and approved the final version of the manuscript.

Data Availability

Data are available on request.

References

1. Alkema, L. et al. Global, regional, and national levels and trends in maternal mortality between 1990 and 2015. *Lancet* **387** (10017), 462–474. https://doi.org/10.1016/S0140-6736(15)00838-7 (2016).
2. Beam, A. L. & Kohane, I. S. Big data and machine learning in health care. *JAMA* **319** (13), 1317–1318. https://doi.org/10.1001/jama.2017.18391 (2018).
3. Bishop, C. M. *Pattern recognition and machine learning* (Springer, 2006).
4. Chen, T. & Guestrin, C. XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). (2016). https://doi.org/10.1145/2939672.2939785
5. Choi, E., Schuetz, A., Stewart, W. F. & Sun, J. Using recurrent neural network models for early detection of heart failure onset. *J. Am. Med. Inform. Assoc.* **24** (2), 361–370 (2017).
6. Efevberha-Ogodo, O., Egbokhare, F. A. & Chete, F. O. Towards the development of a predictive machine learning model for maternal mortality. *NIPES J. Sci. Technol. Res.* **7** (2), 287–297. https://doi.org/10.37933/nipes/7.2.2025.20 (2025).
7. Esteva, A. et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **542** (7639), 115–118. https://doi.org/10.1038/nature21056 (2017).
8. Esteva, A. et al. A guide to deep learning in healthcare. *Nat. Med.* **25** (1), 24–29. https://doi.org/10.1038/s41591-018-0316-z (2019).
9. Filippi, V., Chou, D., Barreix, M. & Say, L. A new conceptual framework for maternal morbidity. *Int. J. Gynecol. Obstet.* **141** (Suppl. 1), 4–9. https://doi.org/10.1002/ijgo.12463 (2018).
10. Friedman, J. H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **29** (5), 1189–1232 (2001).
11. Gao, Y., Huang, X., Tan, C., Zhao, X. & Yang, J. XGBoost-based analysis of maternal and biochemical factors associated with spontaneous preterm birth: A retrospective cohort study. *BMC Pregnancy Childbirth*. **25**, 1227. https://doi.org/10.1186/s12884-025-08414-1 (2025).
12. Gismondi, R. C., Almeida, R. M. V. R. & Infantosi, A. F. C. Artificial neural networks for infant mortality modelling. *Comput. Methods Programs Biomed.* **69** (3), 237–247. https://doi.org/10.1016/S0169-2607(02)00006-8 (2002).
13. Goodfellow, I., Bengio, Y. & Courville, A. *Deep learning. MIT Press*. (2016).
14. Haykin, S. Neural networks and learning machines (3rd ed.). Pearson. (2009).
15. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521** (7553), 436–444. https://doi.org/10.1038/nature14539 (2015).
16. Libbrecht, M. W. & Noble, W. S. Machine learning applications in genetics and genomics. *Nat. Rev. Genet.* **16** (6), 321–332. https://doi.org/10.1038/nrg3920 (2015).

17. Liu, L. et al. Global, regional, and national causes of under-5 mortality. *Lancet* **388** (10063), 3027–3035. https://doi.org/10.1016/S0140-6736(16)31593-8 (2016).
18. Lundberg, S. M. & Lee, S. I. A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems 30 (NeurIPS 2017)* (pp. 4765–4774). Curran Associates, Inc. (2017).
19. Malde, A. et al. A machine learning approach for predicting maternal health risks in lower-middle-income countries using sparse data and vital signs. *Future Internet*, **17**(190). (2025). https://www.mdpi.com/
20. Mamun, M. et al. Identification of maternal health risk from optimal features using explainable machine learning. *Eng. Rep.* **7** (11), e70491. https://doi.org/10.1002/eng2.70491 (2025).
21. Margret, I. N., Rajakumar, K., Arulalan, K. V. & Manikandan, S. Statistical insights into machine learning-based box models for pregnancy care and maternal mortality reduction: A literature survey. *IEEE Access.* **12**, 68184–68207. https://doi.org/10.1109/ACCESS.2024.xxxxx (2024).
22. Masha, A. & Zarma, M. L. Application of artificial neural networks and conventional statistical methods in predicting maternal mortality in Maiduguri, Nigeria. *World J. Adv. Eng. Technol. Sci.* https://doi.org/10.30574/wjaets.2026.18.1.0043 (2026).
23. Miotto, R., Wang, F., Wang, S., Jiang, X. & Dudley, J. T. Deep learning for healthcare: Review, opportunities and challenges. *Brief. Bioinform.* **19** (6), 1236–1246. https://doi.org/10.1093/bib/bbx044 (2018).
24. Mortazavi, B. J. et al. Analysis of machine learning techniques for heart failure readmissions. *Circulation: Cardiovasc. Qual. Outcomes*. **9** (6), 629–640 (2016).
25. Nwamekwe, C. O., Vitalis, E. N., Igbokwe, N. C. & Chibuzo, N. Evaluating advances in machine learning algorithms for predicting and preventing maternal and foetal mortality in Nigerian healthcare: A systematic approach. *Int. J. Industrial Prod. Eng.* **3** (1), 1–15 (2025). https://hal.science/hal-04994840
26. Rajkomar, A., Dean, J. & Kohane, I. Machine learning in medicine. *N. Engl. J. Med.* **380** (14), 1347–1358. https://doi.org/10.1056/NEJMra1814259 (2019).
27. Sarker, R. K. R. et al. Predicting maternal health risk using PCA-enhanced XGBoost and SMOTE-ENN for improved healthcare outcomes. *Intelligence-Based Med.* **12**, 100300. https://doi.org/10.1016/j.ibmed.2025.100300 (2025).
28. Say, L. et al. Global causes of maternal death: A WHO systematic analysis. *Lancet Global Health*. **2** (6), e323–e333. https://doi.org/10.1016/S2214-109X(14)70227-X (2014).
29. Topol, E. J. High-performance medicine: The convergence of human and artificial intelligence. *Nat. Med.* **25** (1), 44–56. https://doi.org/10.1038/s41591-018-0300-7 (2019).
30. UNICEF & United Nations Inter-agency Group for Child Mortality Estimation (UN IGME). : Levels and trends in child mortality report 2025. UNICEF. (2026). https://data.unicef.org/resources/levels-and-trends-in-child-mortality-2025/

31. Verma, I. & Prasad, S. K. Exploring ensemble learning techniques for infant mortality prediction: A technical analysis of XGBoost, stacking, AdaBoost, and bagging models. *Birth Defects Res.* **117** (2), e2443. https://doi.org/10.1002/bdr2.2443 (2025).
32. World Health Organization (WHO). *Trends in maternal mortality 2000–2020: Estimates by WHO, UNICEF, UNFPA, World Bank Group and UNDESA* (World Health Organization, 2023).
33. World Health Organization & UNICEF, UNFPA, World Bank Group and the United Nations. Trends in maternal mortality 2000 to 2023: Estimates by WHO, Population Division. WHO. (2023). https://www.who.int/publications/i/item/9789240108462
34. World Health Organization. Maternal mortality. (2025). https://www.who.int/news-room/fact-sheets/detail/maternal-mortality

Figures

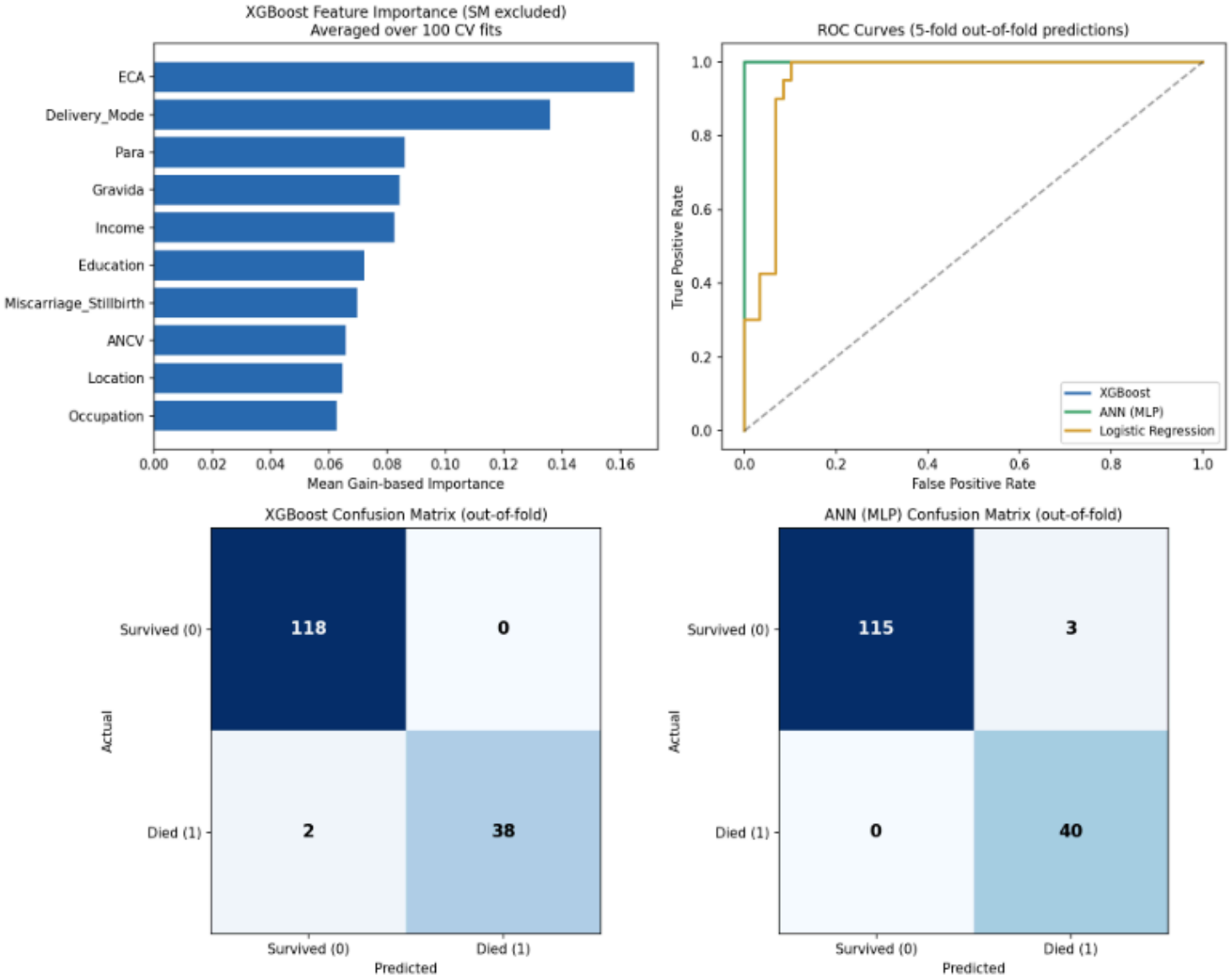


Figure 1

Feature importance, ROC curves, and out-of-fold confusion matrices (XGBoost and ANN), SM excluded