Machine Learning Research
2025; 10(2): 110-123
http://www.sciencepublishinggroup.com/j/mlr
doi: 10.11648/j.mlr.20251002.12
ISSN: 2637-5672 (Print); ISSN: 2637-5680 (Online)

Comparative Analysis of Machine Learning Algorithms for Predicting Under-Five Mortality: Evidence from Tanzania Demographic and Health Survey

Salyungu Mabula[1, 2, *], Robert Too[2], Gregory Kerich[2]

[1]Department of Mthematics and Statistics, College of Natural and Mathematical Sciences, University of Dodoma, Dodoma, Tanzania
[2]Department of Mathematics, Physics and Computing, School of Science and Aerospace Studies, Moi University, Eldoret, Kenya

Email address:

salyungu.mabula1@gmail.com (Salyungu Mabula), rtookip2000@gmail.com (Robert Too), kerichgregoryy@gmail.com (Gregory Kerich)
*Corresponding author

To cite this article:

Salyungu Mabula, Robert Too, Gregory Kerich. (2025). Comparative Analysis of Machine Learning Algorithms for Predicting Under-Five Mortality: Evidence from Tanzania Demographic and Health Survey. *Machine Learning Research, 10*(2), 110-123.
https://doi.org/10.11648/j.mlr.20251002.12

Received: 9 July 2025; **Accepted:** 24 July 2025; **Published:** 20 August 2025

Abstract: Under-five mortality remains a global health challenge with the rates of 43 deaths per every 1000 live births in Tanzania and 37 deaths per every 1000 live births globally. Although child mortality has significantly declined in the last twenty years, the current rates are far from reaching the anticipated Sustainable Development Goal of atmost 25 deaths per 1000 live births in 2030. This study intended to find the best performing classifier of under-five mortality status by comparing ten supervised machine learning algorithms. These machine learning algorithms are Decision Trees, Random Forest, Support Vector Machines, SMOTE-Based Boosted Random Forest, XGBoost, LightGBM, CatBoost, Logistic Regression, K-Nearest Neighbors and Stacked Ensemble Methods. The class imbalance of the dataset detected in the pre-processing stage was addressed using weighted categorical cross-entropy and SMOTE with a 5-folds cross validation and data splitting ratio of 80% for training set and 20% for testing set. With 20 experiments for each of the nine algorithms, the average results were reported to ensure that the findings were not by chance. Further, the stacking ensemble model was developed integrating six of the best performing algorithms using an inclusion criterion of AUC > 0.97. The findings revealed that ensemble algorithm consistently outperformed the other nine algorithms by achieving 100%, 100%, 99.97% and 99.24% for AUC, Accuracy, F1-Score and MCC respectively. This implies that stacking ensemble can uncover more insights than the individual algorithms in predicting under-five mortality status. This study recommends designing policies on under-five mortality that integrate insights from the stacking ensemble algorithm which shows the highest predictive performance.

Keywords: Under-five, Mortality, Modeling, Machine Learning, Prediction, Classification

1. Introduction

For the last 20 years since 2005, the global health community, governments and private sector have embarked on and witnessed coordinated efforts to address under-five mortality as one of the priority health challenge. This effort is in line with the focus of achieving Sustainable Development Goal (SDG) 3 of reducing deaths related to under-five children. As a result, significant stride has been made to reduce under-five mortality at global level from 76 deaths per 1000 live births in 2000 to 37 deaths per 1000 live births in 2022 [1, 2], Sub-Saharan Africa from 150 deaths per 1000 live births in 2000 to 76 deaths per 1000 live births in 2019 [2] and Tanzania from 147 deaths per 1000 live births in 1999 to 43 deaths per 1000 births in 2022 [3, 4]. These statistics indicate that Sub-Saharan Africa, and Tanzania in particular, is still above global average and far from achieving the expected SDG 3 target 2 of atmost 25 deaths per 1000 live births by the year 2030. In addition, the United Nation Inter-Agency Group for Child Mortality Estimation (UN IGME) projected that, if the current trends are not checked, 35 million under-five children will die in 2030 [5], with more than 80% of

this burden being distributed mostly in Sub-Saharan Africa and partly in Southern Asia alone [2]. Even within Sub-Saharan Africa, substantial disparities of under-five mortality between Sub-Saharan Africa economic regions is higher with, for instance, West and Central Africa leading at an average of 94.7 deaths per 1000 live births and in Eastern and Southern Africa the average was 55.5 deaths per 1000 live births in 2019 [2]. The leading countries are Nigeria, Central African Republic, Chad, Sierra Leone, Lesotho and Djibouti which all had more than 100 deaths per 1000 live births in 2019. This makes an under-five child in Sub-Saharan Africa, according to [5], being 18 times more risky compared to children of the same age in Australia and New Zealand. In light of this viewpoints, we focus on developing supervised machine learning models for predicting under-five mortality status in Tanzania, a country that has been battling health challenges associated with under-five mortality since early days of her independence in 1960s. Among the three main barriers to Tanzania progress and development are diseases, illiteracy and poverty [6].

For a long time, health policies have been informed by traditional statistical and epidemiological methods like simple multivariate linear regression [7-9]. Although these methods have immensely contributed to evidence of predicting health outcomes, including under-five mortality [8, 10, 11], more sophisticated and advanced hybrid approaches need to be used to gain deeper insights and the benefits of both traditional and modern data science based methods [6, 12] and [13]. Among the proponents of using modern data science methods like machine learning while making balance with traditional statistical approaches include [14-16] and [13].

In the last 15 years since 2010, more sophisticated Artificial Intelligence (AI) based techniques, machine learning methods inclusive, have been widely applied to uncover the deeper structures of complex data, which have helped in addressing the limitations of traditional methods [8, 17-19]. Unfortunately, in Sub-Saharan Africa, AI-based techniques like supervised machine learning techniques are inadequately utilized in predicting under-five mortality status [20, 21], with a severe under-utilization in the health sector in Tanzania [22]. In this study, we developed ten supervised machine learning algorithms namely Decision Tree (DT), Random Forest, Support Vector Machines, SMOTE-Based Boosted Random Forest, XGBoosting, LightGBM, CatBoosting, Logistic Regression, K-Nearest Neighbors and Ensemble Methods. We then compared their predictive performance on under-five mortality status in Tanzania, for the purpose of contributing evidence on their predictive effectiveness. The main contribution of this paper is in the following aspects;

(i) We contribute knowledge by developing a machine learning algorithm that effectively classifies under-five mortality status at individual level using current Demographic and Health Survey (DHS) dataset. In this case, stacking ensemble algorithm outperformed other algorithms achieving the highest predictive ability in all the four performance metrics used, with 100%, 100%, 99.97% and 99.24% for AUC, Accuracy, F1-Score and MCC respectively.

(ii) We also advance the application of sophisticated methods in public health research by emphasizing and motivating the interaction of statistics, data science and public health to promote interdisciplinary approaches in addressing health challenges. By bridging the gap between statistical analysis and public health practices, researchers can better inform evidence-based policies and interventions.

.

2. Machine Learning and Under-five Mortality Related Works

Numerous theoretical models provide foundation in the study of childhood mortality as well as machine learning. Scholarly research for a long time has used traditional statistical models to predict childhood mortality, with a view toward developing public health interventions to reduce mortality in general. For the past 15 years since 2010, machine learning techniques have emerged with the potential to remove limitations of traditional statistical methods by providing more flexible and non-linear algorithms that can capture complex interactions in data without making pre-conceived assumptions.

As revealed by [23], decision trees have been employed in modeling health outcomes in different regions of Africa, including Sub-Saharan Africa. The study applied decision trees to predict infant mortality in more than 10 countries using DHS data. It was found that, decision tree was the best predictive algorithm followed by random forest. Another study by [17] used random forest to analyze secondary data from the Kilimanjaro Christian Medical Centre (KCMC) Medical Birth Registry (MBR) to predict perinatal deaths. The findings indicate no statistical difference in predictive performance between random forest and the other four machine learning algorithms that were used, with random forest, however, surpassing all algorithms in terms of net benefits. Dealing with a similar problem of predicting neonatal mortality from MBR of Bugando Medical Center (BMC) in Tanzania, [24] used Generalized Linear Models (GLMs) and decision trees without applying random forest, but they underscore predictive accuracy of random forest. In the same line, [25] indicate that random forest outperformed logistic regression in predictive performance when they studied 556 benchmark machine learning datasets consistent to [26] who used cross-sectional design for adult patients clinical data in China. Other studies that indicate random forest to outperform other machine learning models are [27] and [28]. In their studies, [29] in Rwanda and [20, 21] in Sub-Saharan African countries have demonstrated that SVM is potentially powerful in predicting various health outcomes. Although there are fewer studies that uses SVM in Tanzania context like [30], the potential of applying this algorithm in different health outcomes and using comprehensive dataset like Tanzania DHS is promising, given its success in other contexts. Furthermore, several

other scholars used various machine learning algorithms in their mortality related studies including [31] who used LightGBM in studying maternal health risks and [32] who applied LightGBM on mortality prediction in healthcare data; reference [33] who used Gradient Boosting-based approach in studying COVID-19 related deaths and [34] who predicted Paediatric Mortality using XGBoosting and Spectral clustering in Sub-Saharan Africa and South Asia.

Another study by [35] which was conducted in Nigeria with the titled "Application of machine learning methods for predicting Under-five Mortality using DHS dataset", applied KNN among other methods. They commended about its effectiveness on both regression and classification, with similar views from [36]. However, there are numerous studies too which indicate KNN method as inferior to some of the predictive machine learning algorithms like LR, RF and SVM [37-40].

Ensemble algorithms have been indicated to address the weakness of individual algorithms by training multiple, homogeneous or heterogeneous, machine learning algorithms in order to solve the same problem, which otherwise a single algorithm would not have produced high predictive power [41]. The foundation of ensemble methods is the weakness of the other machine learning algorithms. In several instances, the combination of multiple weak learners creates a more robust, accurate, powerful and reliable algorithm. The process of developing an ensemble algorithm has two steps. Initially, the base (weak) learners are developed to their best accuracy and diversity, and then these weak learners are combined to a single model that is powerful than each of the individual learners. The mathematical expression of the ensemble methods can involve bootstrap aggregation (Bagging) if the goal is to reduce variance and prevent overfitting, Boosting if the goal is to correct bias or stacking if the goal is to combine multiple strong algorithms. The ensemble approach has been applied in numerous contexts including studies on mortality by [42-44], and they have indicated potentials for enhancing prediction of health outcome like childhood mortality.

Although the application of machine learning has indicated potential of addressing health outcomes, it still remains underutilized in low and middle-income countries like Tanzania [22]. Most of the existing studies on childhood mortality in Tanzania have mainly applied traditional statistical methods, and those which apply machine learning methods are still limited in scope. For example, disparities in rural-urban factors in healthcare access or economic status can vary significantly in impact and interaction. Despite richness in metrics, Tanzania DHS dataset has not been fully tapped for predictive analytics in machine learning research. In this study, we sought to address the existing gaps by developing predictive machine learning algorithms within Tanzania's health sector context, hence providing insights for policymakers to inform planning and projections, and ultimately contributing to reduction of childhood mortality.

Most of the reviewed literature addressed the under-five mortality using different techniques in both developed and developing countries. These techniques produced inconsistent predictive results in terms of both best performing algorithms and the evaluation metrics. Further evidence indicates that most of the previous literature review did not apply stacking ensemble technique. Those which applied ensemble methods did not integrate more than 5 base learners. Furthermore, to our best knowledge, no study in Tanzania employed stacking ensemble, specifically in predicting under-five mortality status using the latest DHS 2022 data.

3. Materials and Methods

3.1. Data Source and Data Management

The current study used Tanzania DHS 2022 dataset, specifically the Kids Under 5 (KR) file, that were sourced from Tanzania National Bureau of Statistics (NBS) through permission of DHS program, ICF. The data has a nationally representative sample of 10783 under-five children as study participants whose information were obtained retrospectively from women of age 15-49 years as sampling units. Sampling was based on 629 clusters with 26 households each, which had a total of 16312 households.

3.2. Data Management and Variable Selection

Initially, the KR file had 1324 variables with missingness of between 0% and 99.9% that was identified using SPSS. Since higher missingness can introduce significant biases and lead to loss of data integrity [45-48], the first step was to exclude all variables with missingness of more than 10% following [46], where only 392 variables were retained. High missingness can be handled properly by advanced model-based imputation, which was not the objective of this study. After implementing the Little's MCAR test [46, 48], the remaining 392 variables revealed that the missingness was completely at random. The data of the remaining 392 features was explored and cleaned. The cleaning process involved; first, imputing all missing values using mode approach of KNN technique with 5-folds cross-validation in Python environment, and secondly, re-coding features that had a long list of categories. The features were further subjected to dimensionality reduction using Principal Component Analysis (PCA), with factor loading of 0.5 being a cut point, from which 28 features were extracted. After this stage, the 28 features were used to develop the ten supervised machine learning algorithms. All features were coded as categorical, where the outcome feature, under-five mortality status, was coded as binary outcome with 1 = dead if the under-five child had died before celebrating the fifth birthday preceding the 2022 Tanzania Demographic and Health Survey, and 0 otherwise.

3.3. Data Analysis and Algorithmic Experimentation

We conducted the analysis using HP ProBook 445 Generation 9, U88 version 01.17.00, with a Windows 10Pro 64-bit system and an AMD Ryzen 7 5825U processor, 16CPUs of 2.0GHz and 16.384GB RAM. The selection of training set

and testing set ratio of each of the nine algorithms (excluding ensemble) was experimented in Python with stratified samples using 5-folds cross-validation. We found the optimal split ratio for DT, RSK, LightGBM, XGBoosting and CatBoosting was 90% to 10%; KNN and BRF were optimal at 85% by 15% while Random Forest was optimal at 70% by 30%. Since DHS is a large dataset, we ruled out to use a conventional split ratio of 80% for training set by 20% for testing set in order to make a reliable comparison of algorithms performance following [49, 50]. After deciding the training and testing sets, the optimal hyper-parameters of each algorithm were determined by fine-tuning several values using GridSearch method. The high imbalance of the dataset was addressed using weighted categorical cross-entropy of [0:1.0, 1:26.7198] in eight of the algorithms except SMOTE-Based BRF where Synthetic Minority Over-Sampling Technique (SMOTE) was used to inherently address the class imbalance. We experimented with each of the nine algorithms using 20 runs and reporting their average results. The purpose was to ensure that the results reported were not by chance. We used a 5-folds cross-validation in each algorithm to determine the optimal parameters. The last algorithm we developed, also using 20 runs, was an ensemble of the six best performing algorithms namely Decision Tree, CatBoosting, LightGBM, XGBoosting, Random Forest and Logistic Regression, which were selected using an inclusion criterion of $AUC > 0.97$. Finally, the decision on the best algorithm was done based on four performance metrics namely Receiver Operating Characteristic curves (ROC-AUC), Accuracy (ACC), F1-Score, and Mathew Correlation Coefficient (MCC).

3.4. Model Estimation Techniques

In this study we developed ten supervised machine learning algorithms for predicting under-five mortality status in Tanzania. The algorithms are Decision Tree, Random Forest, RBF Kernel in SVM, SMOTE-Based Boosted Random Forest, XGBoosting, LightGBM, CatBoosting, Logistic Regression, K-Nearest Neighbors and Ensemble Methods. The algorithms were chosen because all can perform classification task, consistent with the requirement of a binary outcome variable in machine learning modeling.

3.4.1. Decision Trees

A decision tree is defined as a predictive model which can be used to represent both classifiers and regression models [51]. When a decision tree is used for classification, it is most commonly referred to as a classification tree. It generally has a structure with three components namely (i) root nodes that represent a decision based on a feature, (ii) branches which is an outcome of the decision and (iii) leaf nodes that represent the final decision. A decision Tree recursively splits the feature space using entropy or Gin impurity to classify instances. Its mathematical representation with respect to Information Gain is presented as follows;

$$A^* = \arg\max_A \left(H(S) - \sum_{v \in A} \frac{|S_v|}{|S|} H(S_v) \right) \tag{1}$$

Whereby; A^* is the optimal feature chosen for splitting at a node; $\arg\max_A$ finds the feature A^* that maximizes Information Gain, and $H(S)$ represents the entropy (uncertainty) of the dataset S.

3.4.2. Random Forest

In their study, [52] point out that a random forest is one of several ensemble methods which operates fast and simpler but yet effective in implementing variety of domains. This algorithm constructs much simpler decision trees and implement a majority vote (for categorical outcomes) or by averaging (for continuous outcomes) across them in a classification stage [53, 54]. The major significance of voting and averaging strategies lays in their ability to correct undesirable property of decision trees to overfitting training data [55]. The mathematical representation of random forest is as follows;

$$f(X) = \arg\max_i \sum_{t=1}^{T} 1\left(f_t(X) = i\right) \tag{2}$$

Whereby; $f(X)$ represent the final predicted class for input X; T represents the total number of decision trees in the Random Forest; $f_t(X)$ represents the prediction from the t-th individual tree; $1\ f_t(X) = i$ is an indicator function that returns 1 if tree t predicts class i, otherwise 0, and $\arg\max_i$ selects the class i with the highest votes among all trees. By aggregating predictions from all trees, the final classification decision becomes more reliable and generalizable.

3.4.3. Support Vector Machines

The Radial Basis Function (RBF) kernel is one of the most commonly used kernel functions in Support Vector Machines. It works by transforming the input space into a higher-dimensional features space in which it allows SVM to find a non-linear decision boundary. The performance of the algorithm is greatly influenced by the settings of the RBF kernel [6], which captures an intricate relationship between input features and under-five mortality status.

$$K(x, x') = \exp\left(-\gamma \|x - x'\|^2\right) \tag{3}$$

Whereby $K(x, x')$ measures similarity between two data points x and x'; $\|x - x'\|^2$ represents the squared Euclidean distance between x and x'; γ is a positive parameter that controls the spread of the kernel function, and $\exp(\cdot)$ Is exponential function that ensures values remain in the range $(0, 1]$.

3.4.4. SMOTE-Based BRF

In order to overcome higher class imbalance, the SMOTE-Based Balanced Random Forest (SMOTE-Based BRF) works as an improved variant of Balanced Random Forest (BRF) that incorporates the Synthetic Minority Over-sampling Technique [56]. By creating artificial samples for the minority class, it improves classification performance and ensures that there more balanced decision trees in the Random Forest model.

The method enhances generalization and reduces prediction bias by combining SMOTE and BRF [57]. The mathematical representation of the model is as follows;

$$f(X) = \arg\max_i \sum_{t=1}^{T} 1(f_t(S_{SMOTE}, X) = i) \quad (4)$$

Whereby; $f(X)$ is the final predicted class for input X; T is the total number of Decision Trees in the Balanced Random Forest; S_{SMOTE} is the modified dataset where the minority class is synthetically oversampled using SMOTE; $f_t(S_{SMOTE}, X)$ is the prediction from the t-th tree trained on the balanced dataset; $1f_t(S_{SMOTE}, X) = i)$ is the indicator function that equals 1 if tree t predicts class i, otherwise 0 and $\arg\max_i$ is the function that selects the class i with the highest votes across all trees.

3.4.5. Gradient Boosting

The general Gradient Boosting algorithm is an ensemble learning techniques that is used to increase predicted accuracy by iteratively building several weak learners, which most of the time are homogeneous decision trees [31]. The common applied versions of Gradient Boosting that are used for classification purposes include XGBoosting, LightGBM and CatBoosting [33]. The algorithm fits new models to the residual errors of earlier algorithms and iteratively minimizes the loss function. The typical formula for updating the model using Gradient Boosting is give as;

$$F_m(X) = F_{m-1}(X) + \gamma h_m(X) \quad (5)$$

whereby; $F_m(X)$ is the boosted model after m iterations; $F_{m-1}(X)$ is the previous model before adding the new weak learner; γ is the learning rate controlling the contribution of each new tree and $h_m(X)$ is the new weak learner trained on residual errors.

(i) *XGBoost*
XGBoost is an optimized version of Gradient Boosting that includes regularization, parallel computing and tree pruning to improve efficiency and predictive accuracy. The function of XGBoost is defined as;

$$\mathcal{L}(\theta) = \sum_{i=1}^{n} \ell(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \quad (6)$$

whereby; $\mathcal{L}(\theta)$ is the total objective function; $\ell(y_i, \hat{y}_i)$ is the loss function measuring prediction error; $\Omega(f_k)$ is the regularization term to control model complexity; n is the number of training samples and K is the number of trees in the ensemble.

(ii) *LightGBM*
LightGBM is a sophisticated boosting approach that has been tuned to enhance efficiency and speed. In contrast, to conventional boosting, it handles big datasets with less computing by using leaf-wise growth and histogram-based splitting. The functional form of the algorithm is defined by;

$$\text{Gain} = \frac{1}{2}\left(\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda}\right) \quad (7)$$

whereby; G_L, G_R are the gradient sums for left and right child nodes; H_L, H_R are the Hessian sums for left and right child nodes and λ is the regularization parameter to prevent overfitting.

(iii) *CatBoost*
On the other hand, using gradient estimation and ordered boosting, CatBoost is designed to handle categorical data well. It is an open-source machine learning library that is particularly effective for classification tasks. CatBoost incorporates advanced techniques to prevent overfitting and improve model performance as well as generalization. The functional form of the algorithm is as presented below.

$$\mathcal{L}(\theta) = \sum_{i=1}^{n} \ell(y_i, \hat{y}_i) + \beta \sum_{k=1}^{K} \|\nabla \ell(y_i, \hat{y}_i)\| \quad (8)$$

whereby; $\mathcal{L}(\theta)$ is the objective function; $\ell(y_i, \hat{y}_i)$ is the loss function measuring prediction error; β is the regularization parameter controlling gradient smoothness and $\nabla \ell(y_i, \hat{y}_i)$ is the gradient of the loss function.

3.4.6. Logistic Regression

Logistic regression is a standard statistical classification method that has been widely used in modeling dichotomous outcomes in health and other fields [37]. In this study, binary logistic regression was used to predict the mortality status of children under-five. The objective was to evaluate how well the logistic regression model predicted child death compared to other algorithms using a set of independent factors. The probability of under-five mortality is given by;

$$\pi = \frac{e^{\alpha + \sum_{j=1}^{k} \beta_j x_{ij} + \mu}}{1 + e^{\alpha + \sum_{j=1}^{k} \beta_j x_{ij} + \mu}} \quad (9)$$

The binary logit model for this case is defined using the following equation.

$$y(x) = \ln \frac{\pi}{1 - \pi} = \alpha + \sum_{j=1}^{k} \beta_j x_{ij} + \mu \quad (10)$$

whereby; x_i is the vector of independent features (socioeconomic, demographic, and health-related factors); β_j are coefficients of the explanatory features; $y(x)$ is the under-five mortality status feature, and μ is the residual term capturing unexplained variation.

3.4.7. K-Nearest Neighbors

The Nearest Neighbors classifiers, especially k-Nearest Neighbors, are credited for their simplicity along with their effectiveness in predicting health outcomes [35, 39, 40]. In this study, the algorithm was used to classify under-five mortality

status where each observation was classified based on the majority class among its K closest neighbors in the under-five mortality status. The classification function is defined as;

$$y(X) = \arg\max_i \sum_{k=1}^{K} 1\left(y_k = i\right) \tag{11}$$

whereby; $y(X)$ is the predicted under-five mortality status for input X; K is the number of nearest neighbors considered for classification; y_k is under-five mortality status of the k-th nearest neighbor; $1(y_k = i)$ is the indicator function that returns 1 if neighbor k belongs to class i, otherwise 0, and $\arg\max_i$ selects the class i that appears most frequently among neighbors.

The closeness between instances is measured using Euclidean distance defined by;

$$d(X, X') = \sqrt{\sum_{j=1}^{n}(X_j - X'_j)^2} \tag{12}$$

whereby; X, X' are feature vectors representing different children; X_j, X'_j are values of the j-th feature for observations X and X', and n is the total number of features used for classification.

3.4.8. *Stacking Ensemble*

The foundation of ensemble methods is the weakness of the other machine learning algorithms. Ensemble approach is used to train multiple machine learning models in order to solve the same problem, which otherwise a single model would not have produced high predictive power [41]. The combination of multiple weak learners creates a more robust, accurate and reliable model. The three primary ensemble techniques are Bagging, Boosting, and Stacking. In this study, we developed a stacking ensemble learning algorithm of the six best performing algorithms namely CatBoost, LightGBM, XGBoost, Random Forest and Decision Trees. To define the stacking ensemble learning algorithm, we used an inclusion criterion of AUC > 0.97. The stacking algorithms is defined by;

$$f_{\text{final}}(X) = g\left(f_1(X), f_2(X), \ldots, f_T(X)\right) \tag{13}$$

whereby; $f_{\text{final}}(X)$ is the final stacked prediction algorithm; $g(\cdot)$ is the meta-learner that combines outputs from base algorithms; $f_t(X)$ is the prediction from the t-th base classifier; and T is the total number of base algorithms used in stacking.

3.4.9. *Performance Metrics for Algorithms*

The predictive performance of each algorithm was evaluated using four metrics namely Accuracy, Methew Correlation Coefficient (MCC), F1-Score, Cross-Validation Score (CV-Score) and AUC. While AUC is obtained by evaluating the area below the ROC curve, the computation of other performance metrics derives from True Positive (TP), True Negative (TN), False Negative (FN) and False Positive (FP) shown in equations (14) below.

$$\begin{aligned} \text{Accuracy} &= \frac{TP + TN}{TP + TN + FP + FN} \\ \text{F1-Score} &= \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \\ \text{MCC} &= \frac{(TP \times TN) - (FP \times FN)}{D} \end{aligned} \tag{14}$$

Whereby;

$$D = \sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}$$

4. Results and Discussion

In this section, we present results and analyze the comparative performance of the supervised machine learning algorithms of the previous studies.

Pre-processing of data indicated that distribution of the response outcome was highly imbalanced with under-five children alive being 96.4% of the total sample while 3.6% of them had already died by the time the survey was conducted (see Figure 1). This data challenge has a tendency of favoring the larger class and ignoring the small ones in most of the standard classifiers, which may result in either over or under-fitting, and consequently leading to poor predictive performance. Since many machine learning algorithms are affected by class imbalances [58], weighted categorical cross-entropy and SMOTE techniques were used to address the problem.

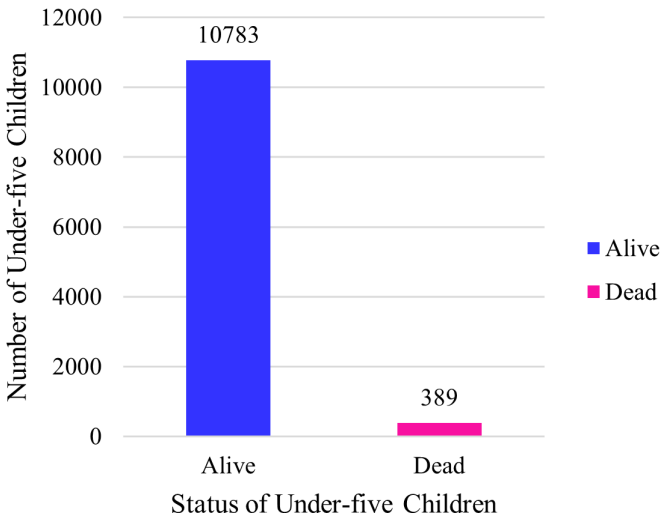

Figure 1. *Distribution of Under-five Mortality for Tanzania 2022 DHS Data.*

4.1. Evaluation of Algorithms Performance

Table 1. *Comparison of Predictive Performance of Ten ML Algorithms in 20 Runs Using 2022 Tanzania DHS Data*

Type of ML Model	F1-Score (Alive)	F1-Score (Died)	Accuracy	MCC	CV-Score	AUC
Decision Trees	0.9982	0.9534	0.9965	0.9524	0.9968	0.9987
Random Forest	0.9958	0.8901	0.9919	0.8864	0.9917	0.9520
RBF Kernel in SVM	0.9971	0.9287	0.9945	0.9280	0.9938	0.9928
SMOTE-Based BRF	0.9825	0.4973	0.9657	0.5208	0.9936	0.7256
XGBoost	0.9973	0.9318	0.9948	0.9303	0.9942	0.9862
LightGBM	0.9969	0.9214	0.9941	0.9191	0.9940	0.9741
CatBoost	0.9969	0.9244	0.9941	0.9238	0.9940	0.9983
Logistic Regression	0.9856	0.7266	0.9726	0.7448	0.9684	0.9858
K-Nearest Negbhors	0.9829	0.5587	0.9666	0.5708	0.9665	0.7437
Stacking Ensemble	0.9997	0.9982	1.0000	0.9924	1.0000	1.0000

Table 1[1] reveals the comparative classification of algorithms in terms of accuracy, Mathew correlation coefficient (MCC), F1-score, Cross validation score and Area under the curve which have been computed as averages of 20 experimental runs. It is apparent from Table 1 and Figure 2 that, Ensemble model outperformed all the other algorithms in predicting under five mortality as demonstrated by the highest scores of accuracy (100%), MCC (99.24%), F1-score for children who were alive (99.97%), F1-score for children who died (99.82%), CV score (100%), and AUC (100%).

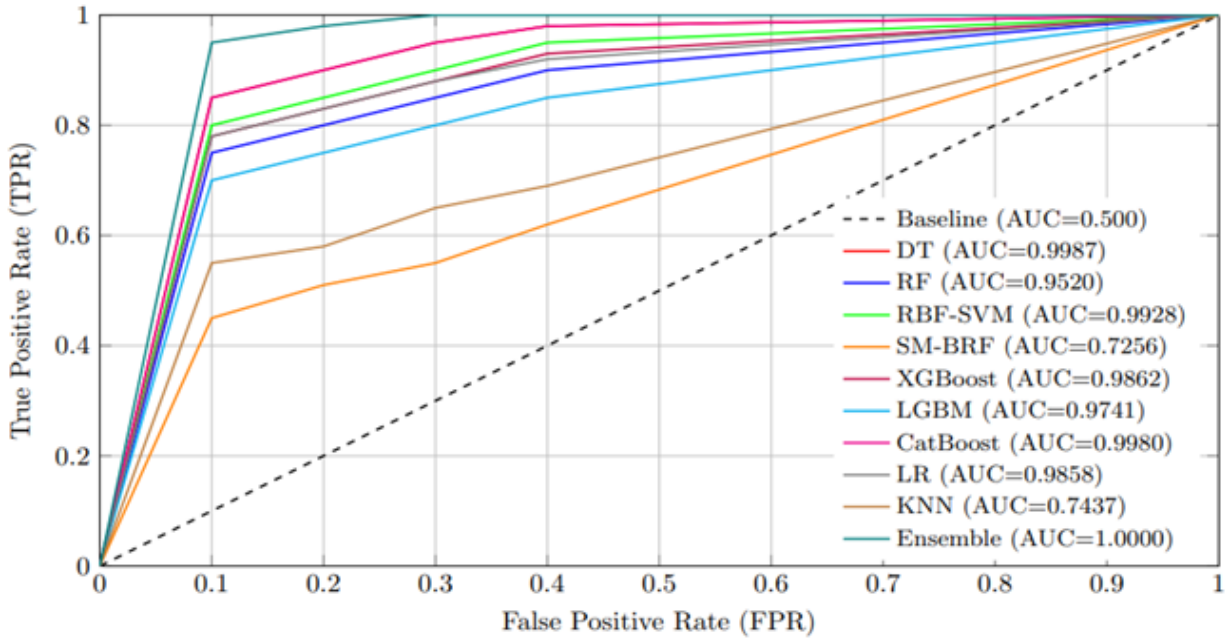

Figure 2. *ROC Curves and AUC of the Ten Classifiers.*

Figure 2 indicates the AUC performance in terms of average from 20 runs of each of the ten predictive classifiers. It is noted that, the algorithms reveal different power of prediction for under-five mortality in Tanzania using DHS data. This result consistently proves that, Ensemble method performs relatively better than other algorithms with an AUC value of 100% followed by Decision Tree, CatBoost and XGBoot with AUC of 99.87%, 99.83% and 99.28% respectively. This implies that, Ensemble algorithm has an ability to capture the complex distribution of under-five mortality in Tanzania, that could guide both policy and decision making.

To validate and further evaluate the robustness of ensemble method as the best predictive algorithm, we plotted the heatmap of its confusion matrices along with the Chi-square metrics. The heatmap was obtained using 20 experiments that utilized random samples from that DHS dataset. The purpose

[1] Stacking ensemble outperformed the other nine ML algorithms in all the evaluation metrics used. The model was validated for near perfect classification through cross-validation for both leakage and overfitting

[2] The confusion matrix in Figure 3 is a result of averages of confusion matrices in all 20 experiments of stacking ensemble

was to assess the predicted results of under-five mortality status using the best classifier. As indicated in Figure 3 [2], ensemble algorithm correctly predicts 2079 under-five children to be alive whereas 78 are correctly predicted to be dead. On the other hand, no under-five children were misclassified, suggesting that Ensemble algorithm perfectly classifies the status of under-five mortality. To ensure that the perfect classification is not due to data leakage or overfitting, further validation using learning curves of train and test sets was conducted as indicated in Figure 4. These results indicate that, the stacking ensemble algorithm is more effective in classifying under-five mortality patterns in Tanzania. The Chi-square test for stacking ensemble indicates significant p-values at conventional levels of 1%, 5% and 10% with Chi-square value of 1913.65 and *p*-value less than 0.000. It is worthy to note from Table 1 that, based on the AUC results, the predictive performance of Ensemble method (100%), Decision Tree (99.87%), CatBoost (99.83%), XGBoost (98.62%) and Logistic Regression (98.58%) have slight differences, with ensemble algorithm leading. This justifies why this study used 20 experiments for each classifier so that the results obtained are not by chance based on the sample selected.

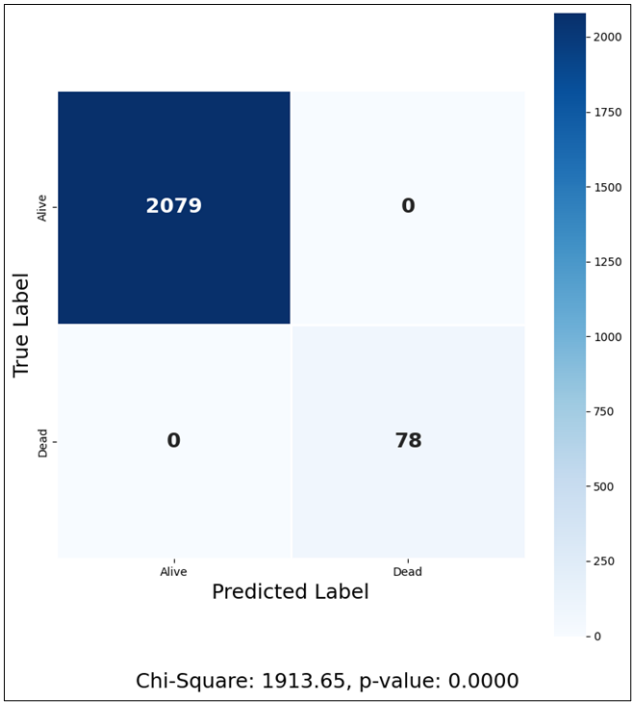

Figure 3. *Heatmap of Confusion Matrices Showing Classification of Under-five Children Using Stacking Ensemble.*

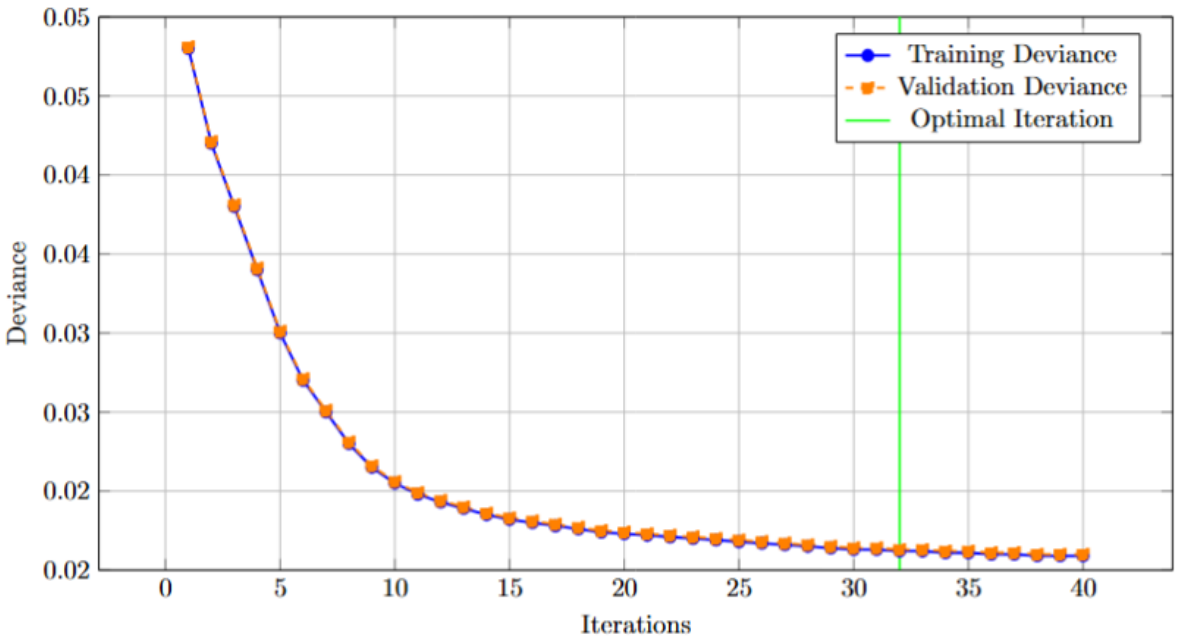

Figure 4. *Learning Curves for Stacking Ensemble Showing Alignment of Train and Validation Sets.*

Figure 4 presents the training and validation deviance for stacking ensemble. Both the training and validation curves start with deviance of 0.05 and drop rapidly aligning perfectly before stabilizing within 30 to 35 iterations. This suggests that stacking ensemble model is learning effectively and reached a solid generalization at the optimal iteration of 32. This results consistently indicate that the perfect classification of stacking ensemble model is not due to data leakage or overfitting.

4.2. Comparative Analysis

4.2.1. Data Characteristics

We also contrasted the current proposed research findings to previous related studies in terms of data characteristics, classification performance and model effectiveness. In this

study, we used the latest Tanzania DHS dataset that was collected in 2022. The data exhibits high imbalance, of which 10394 under-five children were alive while 389 of them had died in the five years preceding the survey. Weighted categorical cross-entropy and SMOTE techniques were used to address the challenge. Previous studies which used DHS data for predicting under-five mortality in different contexts include [35] (in Nigeria), [29] (in Rwanda), [23] (in 10 Sub-Saharan countries), [20] (in Zimbabwe) and [43] (in Ethiopia). Other studies did not use DHS data like [59] (in Brazil, they used Birth and Death Registry of one City) and [17] (in Tanzania, they used Medical Birth Registry from KCMC Hospital). Among these studies, [17, 29, 35] had highly imbalanced data with ratios of about 1:13.5, 1:14.4 and 1:27.0, which were addressed using SMOTE, Random Oversampling and SMOTE techniques respectively. Unlike these studies, we used weighted categorical cross-entropy to assign higher penalties to errors on the minority class so that the models treat both classes fairly while maintaining originality of the data.

4.2.2. Predictive and Classification Performance

The classification performance in Table 1 indicate that Ensemble method outperforms the other nine algorithms. Its AUC (100%) is the highest and perfect along with the other metrics like F1-score (99.97%, 99.82%), Accuracy (100%) MCC (99.24%) and CV-score (100%) being higher, which no other algorithm achieved. The next top candidates in predictive performance are Decision Tree, CatBoost, XGBoost and Logistic Regression. Previous studies also compared several classifiers in predicting child mortality. Based on AUC metric, in [35], Random Forest (96.0%) outperformed other 6 algorithms but they did not use Decision Tree; in [29], Random Forest (84.2%) outperformed other 3 algorithms followed by Decision Tree (83.0%); in [23], an ensemble of several Gradient Boosting classifiers performed better (83.4%) than individual weak learners in Tanzania; and in [20], XGBoost outperformed other 3 algorithms based on Accuracy (81.0%), F1-score (84.0%) and F1-score (89.0%). The least performing algorithms are SMOTE-Based BRF and KNN which scored AUC = 72.56% and AUC = 74.37% respectively, consistent with findings pointed out by [37-40].

The validation of Ensemble algorithm was done using heatmap of the confusion matrices as presented in Figure 3. The Ensemble model perfectly classifies 2079 under-five children to be alive whereas 78 are perfectly classified to be dead, with no any misclassifications. Ensemble algorithm demonstrates a perfect predictive accuracy making it a highly potential model for classifying under-five mortality patterns in Tanzania. These results further indicate that, classification algorithms have performed differently in different contexts. The differences may be brought about by the inherent cultural and environmental characteristics from where the data are collected.

4.3. Limitations

In this study, we focused on supervised machine learning algorithms including heterogeneous ensemble approach. Several studies addressing classification problems using machine learning have suggested using hybrid of supervised and unsupervised machine learning algorithms [60-63]. Future research should apply advanced and hybrid methods like causal inference in order to exploit both statistical modeling and machine learning prediction. In addition, Explainable AI (XAI) as well as self- and semi-supervised methods are relevant because they utilize hybrid frameworks with potential to give more insights in under-five mortality.

5. Conclusions

This study developed ten supervised machine learning classifiers for predicting under-five mortality status using 2022 TDHS dataset. These classifiers were evaluated using four performance metrics namely AUC, Accuracy, F1-Score and MCC. These performance metrics were used to identify the best performing algorithm through comparing their values. To ensure the robustness of results from highly class imbalanced data, 5 cross-validation and weighted categorical cross-entropy techniques were applied along with 20 experimental runs for each classifier. Among the ten classifiers developed, stacking ensemble outperformed the other nine classifiers in all metrics, followed by Decision Tree, CatBoost and XGBoost which performed relatively close to stacking ensemble. On the other hand, SMOTE-Based BRF performed poorly relative to other algorithms. This poor performance was possibly because SMOTE-Based BRF introduces synthetic minority samples that may distort feature distribution, leading to synthetic bias that may compromise model stability. High sensitivity to noise of KNN algorithm may have compromised its ability to discriminate between classes due to high-dimensionality of the dataset, leading to poor class separation. This study recommends that, when designing policy on under-five mortality; the government, policy practitioners and other stakeholders should integrate insights from stacking ensemble classifier, which demonstrates the highest predictive performance.

ORCID

0009-0002-1026-1194 (Salyungu Mabula)
0009-0002-3067-7373 (Robert Too)
0000-0003-2485-9373 (Gregory Kerich)

Abbreviations

DHS	Demographic and Health Survey
SDG	Sustainable Development Goals
ROC	Receiver Characteristic Curve
AUC	Area Under the Curve

AUC	Area Under the Curve
ACC	Accuracy
MCC	Mathew Correlation Coefficient
ICF	Inner City Fund International
MCAR	Missing Completely at Random
DT	Decision Tree
RF	Random Forest
SVM	Support Vector Machines
BRF	Boosted Random Forest
XGBoost	Extreme Gradient Boosting
LGBM	Light Gradient Boosting Machine
CatBoost	Categorical Boosting
LR	Logistic Regression
KNN	K-Nearest Neighbors
EM	Ensemble Methods
KCMC	Kilimanjaro Christian Medical Center
BMC	Bugando Medical Center
MBR	Medical Birth Registry
PCA	Principal Component Analysis
AI	Artificial Intelligence
XAI	Explainable AI

Author Contributions

Salyungu Mabula: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing - original draft, Writing - review & editing

Robert Too: Conceptualization, Supervision, Validation, Writing - review & editing

Gregory Kerich: Conceptualization, Supervision, Validation, Writing - review & editing

Informed Consent

No human or animal subjects were involved.

Funding

This research received no external funding.

Data Availability Statement

Restrictions apply to the availability of these data. Data were obtained from Tanzania Bureau of Statistics (NBS) via DHS Archive-ICF and are available at https://dhsprogram.com/data/available-datasets.cfm with the permission for access from DHS Program.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] United Nations. (2024). Progress towards the sustainable development goals: Report of the secretary-general. Available at: https://coilink.org/20.500.12592/mw6mh58

[2] Sharrow, D., Hug, L. and You, D. (2022). Global, Regional, and National Trends in Under-5 Mortality between 1990 and 2019 with Scenario-Based Projections Until 2030: A Systematic Analysis by the UN Inter-agency Group for Child Mortality Estimation. *The Lancet Global Health*, 10(2): 195-206. Available at: www.thelancet.com/lancetgh

[3] Mwanga, M. K., Mirau, S. S., Tchuenche, J. M and Mbalawata, I. S. (2025). Bayesian Prediction of Under-five Mortality Rates for Tanzania. *Franklin Open*, 10(2025): 100221. Available at: https://doi.org/10.1016/j.fraope.2025.100221

[4] Mwijalilege, S. A, Kadigi, M. L. and Kibiki, C. (2025). Comparing ARFIMA and ARIMA Models in Forecasting under Five Mortality Rate in Tanzania. *Asian Journal of Probability and Statistics*, 27(1): 107-121. Available at: https://doi.org/10.9734/ajpas/2025/v27i1707

[5] UNICEF and WHO. (2024). Levels and Trends Child Mortality-Report 2023: Estimates Developed by the United Nations Inter-Agency Group for Child Mortality Estimation. Available at: https://coilink.org/20.500.12592/9w0w08s

[6] Sende, N. B., Saha, S., Ruganzu, L. and Kar, S. (2025). Prediction of Multidimensional Poverty Status with Machine Learning Classification at Household Level: Empirical Evidence from Tanzania. Available at: https://ieeexplore.ieee.org/document/10869458

[7] Olorunsogo, T. O., Ogugua, J. O., Mounde, M., Maduka, C. P. and Omotayo, O. (2024). Epidemiological Statistical Methods: A Comparative Review of their Implementation in Public Health Studies in the USA and Africa. *World Journal of Advanced Research and Reviews*, 21(1): 1479-1495. Available at: https://doi.org/10.30574/wjarr.2024.21.1.0178

[8] Mwanga, M. K., Mirau, S. S., Tchuenche, J. M and Mbalawata, I. S. (2024). Fuzzy Bayesian Inference for Under-five Mortality Data. *Franklin Open*, 8(2024) 100163. Available at: https://doi.org/10.1016/j.fraope.2024.100163

[9] Singh, J. P. (2022). Statistical Methods in Public Health, in Healthcare System Management: Methods and Techniques. *Springer*. Available at: https://link.springer.com/chapter/10.1007/978-981-19-3076-8_4

[10] Ashwini, C., Bose, S. R., Padmavathy, M. D., Raj, C. and Ajay J. C. (2024). Predicting Child Mortality With Diverse Regression Algorithms Using a Machine Learning Approach. Advancing Intelligent Networks Through Distributed Optimization. *Book Chapter 17*, 329-352. *IGI Global*. Available at: https://doi.org/10.4018/979-8-3693-3739-4.ch017

[11] Adebanji, A. O., Asare, C. and Gyamerah, S. A. (2024). Predictive Analysis on the Factors Associated with Birth Outcomes: A Machine Learning Perspective. *International Journal of Medical Informatics*, 189(2024): 105529. Available at: https://doi.org/10.1016/j.ijmedinf.2024.105529

[12] Pfaffenlehner, M., Behrens, M., Zoller, D., et al. (2025). Methodological Challenges using Routine Clinical Care Data for Real-World Evidence: A Rapid Review Utilizing a Systematic Literature Search and Focus Group Discussion. *BMC Medical Research Methodology,* 25(2025): 8. Available at: https://doi.org/10.1186/s12874-024-02440-x

[13] Nayebirad, S., Hassanzadeh., Vahdani, A. M., et al. (2025). Comparison of machine learning models with conventional statistical methods for prediction of percutaneous coronary intervention outcomes: A systematic review and meta-analysis. *BMC Cardiovascular Disorders*. 25(2025): 310. Available at: https://doi.org/10.1186/s12872-025-04746-0

[14] Scrutinio, D., Amitrano, F., Guida, P., et al. (2025). Prediction of Mortality in Heart Failure by Machine Learning. Comparison with Statistical Modeling. *European Journal of Internal Medicine.* Available at: https://doi.org/10.1016/j.ejim.2025.01.020

[15] Hamdoun, S. H., Abed, M. Q., Salman, S. M., Al-Bayati, H. N. A. and Balina, O. (2024). The Intersection of Statistics and Machine Learning: A Comprehensive Analysis. *Journal of Ecohumanism,* 3(5): 406-421. Available at: https://www.ceeol.com/search/article-detail?id=1273674

[16] Satty, A., Khamis, G. S. M., Mohamed, Z. M., et al. (2025). Statistical Insights into Machine Learning Models for Predicting Under-five Mortality: An Analysis from Multiple Indicator Cluster Survey (MICS). *IEEE Access,* 13(2025): 45312-45320. Available at: https://ieeexplore.ieee.org/abstract/document/10916650

[17] Mboya, I. B., Mahande, M. J., Mohamed, M., Obure, J. and Mwambi, H. G. (2020). Prediction of Perinatal Death using Machine Learning Models: A Birth Registry-Based Cohort Study in Northern Tanzania. *BMJ open,* 10(10): e040132. Available at: https://bmjopen.bmj.com/content/10/10/e040132.abstract

[18] Ogbo, F. A., Ezeh, O. K., Awosemo, A. O., et al. (2019). Determinants of Trends in Neonatal, Post-neonatal, Infant, Child and Under-five Mortalities in Tanzania from 2004 to 2016. *BMC public health,* 19(2019): 1-12. Available at: https://doi.org/10.1186/s12889-019-7547-x

[19] Alanazi, B. S. (2025). A Comparative Study of Traditional Statistical Methods and Machine Learning Techniques for Improved Predictive Models. *International Journal of Analysis and Applications,* 23(2025): 18. Available at: https://doi.org/10.28924/2291-8639-23-2025-18

[20] Mbunge, E., Fashoto, S. G., Muchemwa, B., et al. (2023). Application of Machine Learning Techniques for Predicting Child Mortality and Identifying Associated Risk Factors. *in 2023 Conference on Information Communications Technology and Society (ICTAS).* Durban, South Africa, 2023, pp. 1-5. Available at: https://ieeexplore.ieee.org/abstract/document/10082734

[21] Mbunge, E., Millham, R. C., Sibiya, M. N. and Takavarasha, S. (2022). Application of Machine Learning Models to Predict Malaria using Malaria Cases and Environmental Risk Factors. in 2022 Conference on Information Communications Technology and Society (ICTAS). Durban, South Africa, 2022, pp. 1-5. Available at: https://ieeexplore.ieee.org/abstract/document/9744657

[22] Sukums, F., Mzurikwao, D., Sabas, D., et al. (2023). The Use of Artificial Intelligence-based Innovations in the Health Sector in Tanzania: A Scoping Review. *Health Policy and Technology,* 12(2023): 100728. Available at: https://doi.org/10.1016/j.hlpt.2023.100728

[23] Ogallo, W., Speakman, S., Akinwande, V., et al. (2020). Identifying factors associated with neonatal mortality in sub-saharan africa using machine learning. *in AMIA Annual Symposium Proceedings,* 2020: 963. Available at: https://pmc.ncbi.nlm.nih.gov/articles/PMC8075462/

[24] Kovacs, D., Msanga, D. R., Mshana, S. E., Bilal, M., Oravcova, K. and Mathews, L. (2021). Developing practical clinical tools for predicting neonatal mortality at a neonatal intensive care unit in Tanzania. *BMC pediatrics,* 21(2021): 1-10. Available at: https://link.springer.com/article/10.1186/s12887-021-03012-4

[25] Levy, J. J. and OaMalley, A. J. (2020). Do not dismiss logistic regression: The case for sensible extraction of interactions in the era of machine learning. *BMC medical research methodology,* 20(2020): 171. Available at: https://link.springer.com/article/10.1186/s12874-020-01046-3

[26] Wu, H., Liao, B., Ji, T., Ma, K., Luo, Y., and Zhang, S. (2025). Comparison between traditional logistic regression and machine learning for predicting mortality in adult sepsis patients. *Frontiers in Medicine,* 11(2025): 1496869. Available at: https://doi.org/10.3389/fmed.2024.1496869

[27] Yang, Y., Tang, J., Ma, L., Wu, F. and Guan, X. (2025). A systematic comparison of short-term and long-term mortality prediction in acute myocardial infarction using machine learning models. *BMC Medical Informatics and Decision Making,* 25(2025): 208. Available at: https://doi.org/10.1186/s12911-025-03052-1

[28] Lee, H. and Tsoi, P. (2025). Feature-enhanced machine learning for all-cause mortality prediction in healthcare data. *arXiv preprint arXiv:2503.21241.* Available at: https://arxiv.org/abs/2503.21241

[29] Mfateneza, E., Rutayisire, P. C., Bacyaza, E., Musafiri, S. and Mpabuka, G. (2022). Application of Machine Learning Methods for Predicting Infant Mortality in Rwanda: Analysis of Rwanda Demographic Health Survey 2014-15 Dataset. *BMC Pregnancy and Childbirth,* 22(2022): 388. Available at: https://doi.org/10.1186/s12884-022-04699-8

[30] Lawrence, S. L. (2021). Predicting Stunting Status among Children Under-five Years: The Case Study of Tanzania. *Unpublished Ph.D. Dissertation,* University of Rwanda. Available at: http://dr.ur.ac.rw/handle/123456789/1389

[31] Noviandy, T. R., Naiggolan, S. I., Rahan, R., Firmansyah, I. and Idroes, R. (2023). Maternal health risk detection using light gradient boosting machine approach. *Infolitika Journal of Data Science,* 1(2): 48-55. Available at: https://doi.org/10.60084/ijds.v1i2.123

[32] Silva, G. F. D. S., Wichmann, M., da Silva Junior, F. C. and Chavegatto Flho, A. D. P. (2025). Development and Evaluation of Machine Learning Training Strategies for Neonatal Mortality Prediction using Multi-country Data. *Scientific Reports,* 15(2025):24278. Available at: https://pubmed.ncbi.nlm.nih.gov/40624031

[33] Keser, S. B. and Keskin, K. (2023). A Gradient Boosting-based Mortality Prediction Model for COVID-19 Patients. *Neural Computing and Applications,* 35(33): 23997-24013. Available at: https://doi.org/10.1007/s00521-023-08997-w

[34] Diallo, A. H., Shahid, A. S. M. S. B., Khan, A. F., et al. (2023). Characterising Paediatric Mortality During and After Acute Illness in Sub-Saharan Africa and South Asia: A Secondary Analysis of the CHAIN cohort using a Machine Learning Approach. Available at: https://doi.org/10.1016/j.eclinm.2023.101838

[35] Samuel, O., Zewotir, T. and North, T. (2024). Application of machine learning methods for predicting underfive mortality: Analysis of nigerian demographic health survey 2018 dataset. *BMC Medical Informatics and Decision Making,* 24(2024): 88. Available at: https://doi.org/10.1186/s12911-024-02476-5

[36] Kurniawan, M., Yuliastuti, G. E., Rachman, A., Budi, A. P. and Zaqiah, H. N. (2024). Implementing K-Nearest Neighbors (K-NN) Algorithm and Backward Elimination on Cardiotocography Datasets. *International Journal on Informatics Visualization,* 8(3): 1239-1245. Available at: http://dx.doi.org/10.62527/joiv.8.3.1996

[37] Bitew, F. H., Nyarko, S. H., Potter, L. and Sparks, C. S. (2020). Machine Learning Approach for Predicting Under-five Mortality Determinants in Ethiopia: Evidence from the 2016 Ethiopian Demographic and Health Survey. *Genus,* 76(2020): 1-16. Available at: https://doi.org/10.1186/s41118-020-00106-2

[38] Mollalo, A., Vahedi, B., Bhattarai, S., Hopkins, L. C. and Banik, S. (2020). Predicting the Hotspots of Age-adjusted Mortality Rates of Lower Respiratory Infection Across the Continental United States: Integration of GIS, Spatial Statistics and Machine Learning Algorithms. *International Journal of Medical Informatics,* 142(2020): 104248. Available at: https://doi.org/10.1016/j.ijmedinf.2020.104248

[39] Tedese, Z. B., Nigatu, A. M., Yehuala, T. Z. and Sebastian, Y. (2024). Prediction of Incomplete Immunization Among Under-five Children in East Africa from Recent Demographic and Health Surveys: A Machine Learning Approach. *Scientific Reports,* 14(2024): 11529. Available at: https://doi.org/10.1038/s41598-024-62641-8

[40] Yehuala, T. Z., Derseh, N. M., Tewelgne, M. F. and Wubante, S. M. (2024). Exploring Machine Learning Algorithms to Predict Diarrhea Disease and Identify its Determinants among Under-five Years Children in East Africa. *Journal of Epidemiology and Global Health,* 14(3): 1089-1099. Available at: https://link.springer.com/article/10.1007/s44197-024-00259-9

[41] Abdulhafedh, A. (2022). Comparison between Common Statistical Modeling Techniques used in Research, Including: Discriminant Analysis vs Logistic Regression, Ridge Regression vs Lasso, and Decision Tree vs Random Forest. *Open Access Library Journal,* 9(2): 1-19. Available at: https://doi.org/10.4236/oalib.1108414

[42] Bizzego, A., Gabrieli, G., Bornstein, M. H., et al. (2021). Predictors of Contemporary Under-5 Child Mortality in Low-and Middle-income Countries: A Machine Learning Approach. *International Journal of Environmental Research and Public Health,* 18(3): 1315. Available at: https://doi.org/10.3390/ijerph18031315

[43] Dereje, T., Abuhay, T. M., Letta, A. and Alelign, M. (2021). Investigate Risk Factors and Predict Neonatal and Infant Mortality Based on Maternal Determinants using Homogeneous Ensemble Methods. In *2021 International Conference on Information and Communication Technology for Development for Africa (ICT4DA)*. Bahir Dar, Ethiopia, 2021, pp. 18-23. Available at: https://ieeexplore.ieee.org/abstract/document/9671271

[44] Santos, H., Eilertson, K., Lambert, B., Hauryski, S., Patel, M. and Ferrari, M. (2021). Ensemble Model Estimates of the Global Burden of Measles Morbidity and Mortality from 2000 to 2019: A Modeling Study. *MedRxiv*. Available at: https://doi.org/10.1101/2021.08.31.21262916

[45] Lee, S., Kim, Y., Ji, B. and Kim, Y. (2025). Addressing Missing Data in Slope Displacement Monitoring: Comparative Analysis of Advanced Imputation Methods. *Buildings,* 15(2): 236. Available at: https://doi.org/10.3390/buildings15020236

[46] Zang, L. and Xiong, F. (2025). Harnessing Machine Learning to Address High Levels of Missing Data in CrossNational Studies: From Bias to Precision in Public Service Research. *Journal of Comparative Policy Analysis: Research and Practice,* pp. 1-21. Available at: https://doi.org/10.1080/13876988.2025.2478930

[47] Popovich, D. (2025). How to Treat Missing Data in Survey Research. *Journal of Marketing Theory and Practice,* 33(1): 43-59. Available at: https://doi.org/10.1080/10696679.2024.2376052

[48] Camillieri, G. (2024). Missing Data and Imputation. In: Lyman, S., Ayeni, O. R., Koh, J. L., Nakamura, N., Karlsson, J. (eds) Introduction to Surgical Trials. *Springer*, Cham. Available at: https://doi.org/10.1007/978-3-031-77563-5_16

[49] Wu, Z., Zhu, M., Kang, Y., et al. Do we Need Different Machine Learning Algorithms for QSAR Modeling? A Comprehensive Assessment of 16 Machine Learning Algorithms on 14 QSAR Datasets. *Briefings in Bioinformatics,* 22(4): bbaa321. Available at: https://doi.org/10.1093/bib/bbaa321

[50] Zakariaee, S. S, Naderi, N., Ebrahimi, M. and Kazemi-Arpanahi, H. (2023). Comparing Machine Learning Algorithms to Predict COVID-19 Mortality Using a Dataset Including Chest Computed Tomography Severity Score Data. *Scientific Reports,* 13(2023): 11343. Available at: https://doi.org/10.1038/s41598-023-38133-6

[51] Moslehi, S., Rabiei, N., Soltania, A. R. and Mamani, M. (2022). Application of Machine Learning Models Based on Decision Trees in Classifying the Factors Affecting Mortality of COVID-19 Patients in Hamadan, Iran. *BMC Medical Informatics and Decision Making,* 22(2022): 192. Available at: https://doi.org/10.1186/s12911-022-01939-x

[52] Raihan, M., Saha, P. K., Gupta, R. D., et al. (2024). A deep learning and machine learning approach to predict neonatal death in the context of sao paulo. *International Journal of Public Health Science,* 13(1): 179-190. Available at: https://doi.org/10.11591/ijphs.v13i1.22577

[53] Leo, J., Luhanga, E. and Michael, K. (2019). Machine Learning Model for Imbalanced Cholera Dataset in Tanzania. *The Scientific World Journal,* 2019(1): 9397578. Available at: https://doi.org/10.1155/2019/9397578

[54] Ramosaj, B. and Pauly, M. (2019). Consistent Estimation of Residual Variance with Random Forest out-of-bag Errors. *Statistics & Probability Letters,* 151(2019): 49-57. Available at: https://doi.org/10.1016/j.spl.2019.03.017

[55] Caie, P. D., Dimitiou, N. and Arandjelovic, O. (2025). Precision Medicine in Digital Pathology via Image Analysis and Machine Learning. *Elsevier,* pp. 233-257. Available at: https://doi.org/10.1016/B978-0-323-95359-7.00012-1

[56] Imani, M., Beikmohammadi, A. and Arabnia, H. R. (2025). Comprehensive Analysis of Random Forest and Xgboost Performance with SMOTE, ADASYN, and GNUS under Varying Imbalance Levels. *Technologies,* 13(3): 88. Available at: https://doi.org/10.3390/technologies13030088

[57] Newaz, A., Mohosheu, M. S., Al Noman, M. A., and Jabid, T. (2024). IBRF: Improved Balanced Random Forest Classifier. In *2024 IEEE 35th Conference of Open Innovations Association (FRUCT),* pp. 501-508. Available at: https://ieeexplore.ieee.org/document/10516372

[58] Alsharkawi, A., Al-Fatyani, M., Dawas, M., Saadeh, H. and Alyaman, M. (2021). Poverty Classification using Machine Learning: The Case of Jordan. *Sustainability,* 13(3): 1412. Available at: https://doi.org/10.3390/su13031412

[59] Raihan, M., Saha, P. K., Gupta, R. D., et al. (2025). A Deep Learning and Machine Learning Approach to Predict Neonatal Death in the Context of Sao Paulo. *arXiv preprint arXiv:2506.16929.*

[60] Sende, N. B., Saha, S. and Uwimbabazi, L. F. R. (2025). Spatial Distribution of Poverty Clusters and its Prediction Algorithms: A Visual Analytics Approach to Understanding the Disparities of Poverty Across Zones. In *in IEEE Access,* 13(2025): 96302-96316. Available at: https://ieeexplore.ieee.org/document/11020683

[61] Agrawal, S., Gupta, G. K., Gopalakrishna, P. K., Balasubramaniam, V. S., Goel, L. and Mahadik, S. (2024). Hybrid Machine Learning Models: Combining Strengths of Supervised and Unsupervised Learning Approaches. In *2024 7th International Conference on Contemporary Computing and Informatics (IC3I), Greater Noida, India,* 7(2024): 1056-1061. Available at: https://ieeexplore.ieee.org/abstract/document/10829140

[62] Yakovyna, V., Shakhovska, N. and Szpakowska, A. (2024). A novel hybrid supervised and unsupervised hierarchical ensemble for covid-19 cases and mortality prediction. *Scientific Reports,* 14(2024): 9782. Available at: https://doi.org/10.1038/s41598-024-60637-y

[63] Rodriguez, A., Mendoza, D., Ascuntar, J. and Jaimes, F. (2021). Supervised Classification Techniques for Prediction of Mortality in Adult Patients with Sepsis. *The American Journal of Emergency Medicine,* 45(2021): 392-397. Available at: https://doi.org/10.1016/j.ajem.2020.09.013