scientific reports

Check for updates

OPEN

Employing machine learning techniques for prediction of micronutrient supplementation status during pregnancy in East African Countries

Habtamu Setegn Ngusie[1]✉, Ermias Bekele Enyew[2], Agmasie Damtew Walle[3], Bayou Tilahun Assaye[4], Mulugeta Desalegn Kasaye[2], Getanew Aschalew Tesfa[5] & Alemu Birara Zemariam[6]

Micronutrient deficiencies, known as "hidden hunger" or "hidden malnutrition," pose a significant health risk to pregnant women, particularly in low-income countries like the East Africa region. This study employed eight advanced machine learning algorithms to predict the status of micronutrient supplementation among pregnant women in 12 East African countries, using recent demographic health survey (DHS) data. The analysis involved 138,426 study samples, and algorithm performance was evaluated using accuracy, area under the ROC curve (AUC), specificity, precision, recall, and F1-score. Among the algorithms tested, the random forest classifier emerged as the top performer in predicting micronutrient supplementation status, exhibiting excellent evaluation scores (AUC = 0.892 and accuracy = 94.0%). By analyzing mean SHAP values and performing association rule mining, we gained valuable insights into the importance of different variables and their combined impact, revealing hidden patterns within the data. Key predictors of micronutrient supplementation were the mother's education level, employment status, number of antenatal care (ANC) visits, access to media, number of children, and religion. By harnessing the power of machine learning algorithms, policymakers and healthcare providers can develop targeted strategies to improve the uptake of micronutrient supplementation. Key intervention components involve enhancing education, strengthening ANC services, and implementing comprehensive media campaigns that emphasize the importance of micronutrient supplementation. It is also crucial to consider cultural and religious sensitivities when designing interventions to ensure their effectiveness and acceptance within the specific population. Furthermore, researchers are encouraged to explore and experiment with various techniques to optimize algorithm performance, leading to the identification of the most effective predictors and enhanced accuracy in predicting micronutrient supplementation status.

Keywords Micronutrient supplementation, Pregnant women, Demographic health survey, Machine learning algorithms

Abbreviations

ADASYN	Adaptively generating minority data
ANC	Antenatal care
AUC	Area under the ROC curve

[1]Department of Health Informatics, School of Public Health, College of Medicine and Health Sciences, Woldia University, PO Box 400, Woldia, Amhara, Ethiopia. [2]Department of Health Informatics, College of Medicine and Health Science, Wollo University, Desie, Ethiopia. [3]Department of Health Informatics, College of Medicine and Health Science, Debre Berhan University, Debre Berhan, Ethiopia. [4]Department of Health Informatics, College of Health Science, Debre Markos University, Debre Markos, Ethiopia. [5]School of Public Health, Dilla University, Dilla, Ethiopia. [6]Department of Pediatrics and Child Health Nursing, School of Nursing, College of Medicine and Health Science, Woldia University, Woldia, Ethiopia. ✉email: habtamuhi3@gmail.com

CI	Confidence interval
DHS	Demographic and Health Survey
FPR	False positive rate
HIV	Human immunodeficiency virus
KNN	K-nearest neighbor
PCA	Principal component analysis
RBF	Radial basis function
RFE	Recursive feature elimination
ROC	Receiver operating characteristic curve
SHAP	Shapley Additive exPlanations
SDGs	Sustainable Development Goals
SMOTE	Synthetic minority over-sampling technique
SVM	Support vector machine
TPR	True positive rate
WHO	World Health Organization
XGBoost	eXtreme Gradient Boosting

Micronutrients, such as vitamins and minerals, are essential for the human body in micrograms or milligrams per day[1,2]. They play a crucial role in bodily functions, and deficiencies in any of these micronutrients can lead to severe, and potentially fatal, illnesses[3]. "Hidden hunger" or "hidden malnutrition," which refers to micronutrient deficiencies, remains a significant global public health challenge, affecting one out of every four people worldwide, totaling more than two billion individuals[4,5]. Common micronutrient deficiencies include vitamin A, folate, iron, iodine, and zinc[6]. Low-income countries in Africa and Asia bear a disproportionate burden of micronutrient deficiencies, with vulnerable groups such as pregnant women and children being particularly affected[7].

Pregnancy-related micronutrient deficiencies have implications for the health of pregnant women and the development of their children[6,7]. Micronutrient supplementation is essential to prevent pregnancy complications and reduce the risk of adverse outcomes[8]. Of particular concern is iron-folic acid deficiency, as it contributes significantly to anemia among pregnant women and is the most prevalent micronutrient deficiency worldwide during pregnancy[7,9]. Globally, anemia affects up to 36% of pregnant women aged 15–49 years and is estimated to be the cause of 22% of maternal deaths[10]. Low-income countries bear the highest burden of anemia, with East Africa accounting for 39% of global pregnant anemia cases[10].

The World Health Organization (WHO), in collaboration with member states and other partners, invests significant efforts to address all forms of malnutrition. These initiatives encompass high-folic acid and iron supplementation and fortification of food with micronutrients[2]. As the reduction of all forms of malnutrition is a key agenda of the Sustainable Development Goals (SDGs) for 2030, WHO member states strive to achieve the target level of reducing micronutrient malnutrition[11]. However, despite these efforts, the risks of death due to micronutrient malnutrition remain high among pregnant women and children in Eastern Africa[10,12].

Existing literature on micronutrient intake and deficiency has identified several potential factors contributing to these deficiencies[14–16]. These factors include educational level[13–21], age[14], sex[22], residency[23], wealth status[13,15,16,18,24], number of children[14], access to health facilities[25], media exposure[14], marital status[26], working status[26], and antenatal care (ANC) visits[18,27,28].

Machine learning technologies have witnessed widespread adoption and achieved remarkable advancements in various fields. However, their application in the realm of public health and medicine has been relatively limited[29,30]. Traditional analytical techniques have predominantly been employed in prior studies investigating the supplementation status of pregnant women in East African countries[31]. However, utilizing machine learning models for predicting micronutrient supplementation can enhance empirical evidence. Machine learning algorithms can capture complex relationships, handle high-dimensional data, adapt to non-linear patterns, and provide robust and efficient analysis of large-scale datasets[32,33]. Therefore, in this study, we employed eight state-of-the-art machine learning algorithms, including association rule mining, to predict the status of micronutrient supplementation using recent DHS data from East African countries.

Method

Data source

This study utilized secondary data from the most recent DHS conducted in 12 East African countries, namely Ethiopia (2016), Kenya (2014), Uganda (2016), Tanzania (2016), Burundi (2017), Rwanda (2015), Madagascar (2009), Mozambique (2011), Zimbabwe (2015), Zambia (2018), Malawi (2016), and Comoros (2012). The data sets from these countries were extracted from the official DHS program database, which can be accessed at https://dhsprogram.com/data/available-datasets.cfm.We obtained ethical approval from the Institutional Review Board for the DHS program to access the data.

The DHS Program has conducted standardized surveys in over 90 countries, collecting representative data on population, health, human immunodeficiency virus (HIV), and nutrition. The surveys used a multi-stage stratified sampling method, selecting participants from households within clusters. Sampling strata were created based on rural and urban sectors, and enumeration areas were chosen using probability proportional to size. Within the selected enumeration areas, households were chosen using equal probability systematic sampling. The study specifically examined women aged 15 to 49 residing in East Africa, focusing on those who had been pregnant within the preceding five years of data collection. The research sample included a total of 138,426 individuals from 12 countries in East Africa. The dataset utilized in the study comprised 13 features[34].

Study variables and measurements

The dependent variable in this study was micronutrient supplementation, defined as the usage of iron folic acid tablets or syrup for at least ninety days or the usage of deworming medicine during a previous pregnancy[35]. Pregnant women meeting these criteria were classified as "supplemented" (coded as 1), while those who did not receive this supplementation were classified as "not supplemented" (coded as 0)[31]. The study considered various independent variables, including place of residence, age group, religion, number of living children, ANC visit, working status, media exposure, marital status, educational status, wealth status, birth interval, and distance from health facility. The selection of these independent variables was based on a comprehensive review of previous literature[13,31].

Data preprocessing

The process of machine learning begins with data pre-processing, which involves modifying or encoding the data to make it suitable for computer interpretation[36]. In our machine learning workflow, we employed a continuous improvement process for our models. This process included selecting and engineering relevant features, balancing the data, splitting the data, model training, model evaluation, model optimization, choosing the top performer model, and deploying the selected model for prediction. Through an iterative approach, we refined our models. Figure 1 provides a visual representation of the steps in our workflow; however, it does not encompass certain tasks that were iteratively performed throughout the process.

Data cleaning

During the data analysis process, we manually examined the data for redundancy and determined that no redundant data were present in our dataset. To handle missing values, we utilized the K-nearest neighbors (KNN) imputation technique[41]. We employed various visualization techniques such as scatter plots, box plots, and histograms to identify outliers. Additionally, we assessed multicollinearity by examining the correlation matrix and considering a correlation value above 0.8 between two pairs of variables as indicative of high correlation[37,38].

Feature engineering

Feature engineering involves identifying, acquiring, and modifying the most relevant characteristics from the available data to construct machine learning models that are more accurate and efficient[39]. We employed one-hot encoding for nominal categorical variables and label encoding for ordinal categorical variables to encode the data[40].

Dimensionality reduction

We employed various techniques for dimensionality reduction in our study, aiming to optimize model performance and reduce the complexity of our dataset. These techniques included univariate selection, recursive feature elimination (RFE), random forest feature elimination, principal component analysis (PCA), lasso regression, and a feature selection method based on Boruta[41].

Through repeated experiments, we found that the Boruta-based feature selection method outperformed other techniques in terms of accuracy and robustness. The Boruta-based feature selection method assesses feature importance by comparing their performance against randomly generated shadow features that simulate noise. Features consistently outperforming the shadow features are deemed significant and incorporated into our predictive model[42].

Data balancing

Data imbalance poses a common challenge in data mining and machine learning, often leading to decreased classification accuracy for instances belonging to the minority class[43]. To tackle this issue, we utilized four data balancing methods: under-sampling, over-sampling, adaptive synthetic sampling (ADASYN), and synthetic minority oversampling technique (SMOTE). Each of these techniques has distinct characteristics and aims to address class imbalance effectively.

Under-sampling involves reducing the number of instances from the majority class to achieve a more balanced dataset. By randomly removing instances from the majority class, under-sampling aims to align the number of instances in the minority class with that in the majority class. This approach prevents classifier bias towards the majority class, but it may result in the loss of potentially valuable information[44–46].

On the other hand, over-sampling increases the number of instances in the minority class by replicating or generating new instances. This technique ensures a balanced dataset by ensuring that the number of instances in the minority class is comparable to that in the majority class. Over-sampling can be achieved through methods such as random duplication, bootstrapping, or synthetic data generation[46,47].

ADASYN extends the SMOTE technique to address its limitation in handling datasets with varying densities within the minority class. ADASYN synthesizes new instances in the minority class by considering the distribution of instances in the feature space. It focuses on generating more synthetic examples for the minority class instances that are harder to classify, thereby adapting the sampling strategy to the local characteristics of the data[48].

SMOTE, on the other hand, is a popular over-sampling technique that creates synthetic instances in the minority class by interpolating between existing instances. It randomly selects a minority class instance and identifies its k nearest neighbors. It then generates synthetic instances by randomly selecting a neighbor and creating a new instance along the line segment between the original instance and the chosen neighbor. SMOTE helps balance the dataset and introduces diversity in the minority class[48,49].

To enhance the performance of our predictive model, we initially trained our machine learning algorithms using unbalanced data. Subsequently, we explored and applied the aforementioned balancing techniques to train

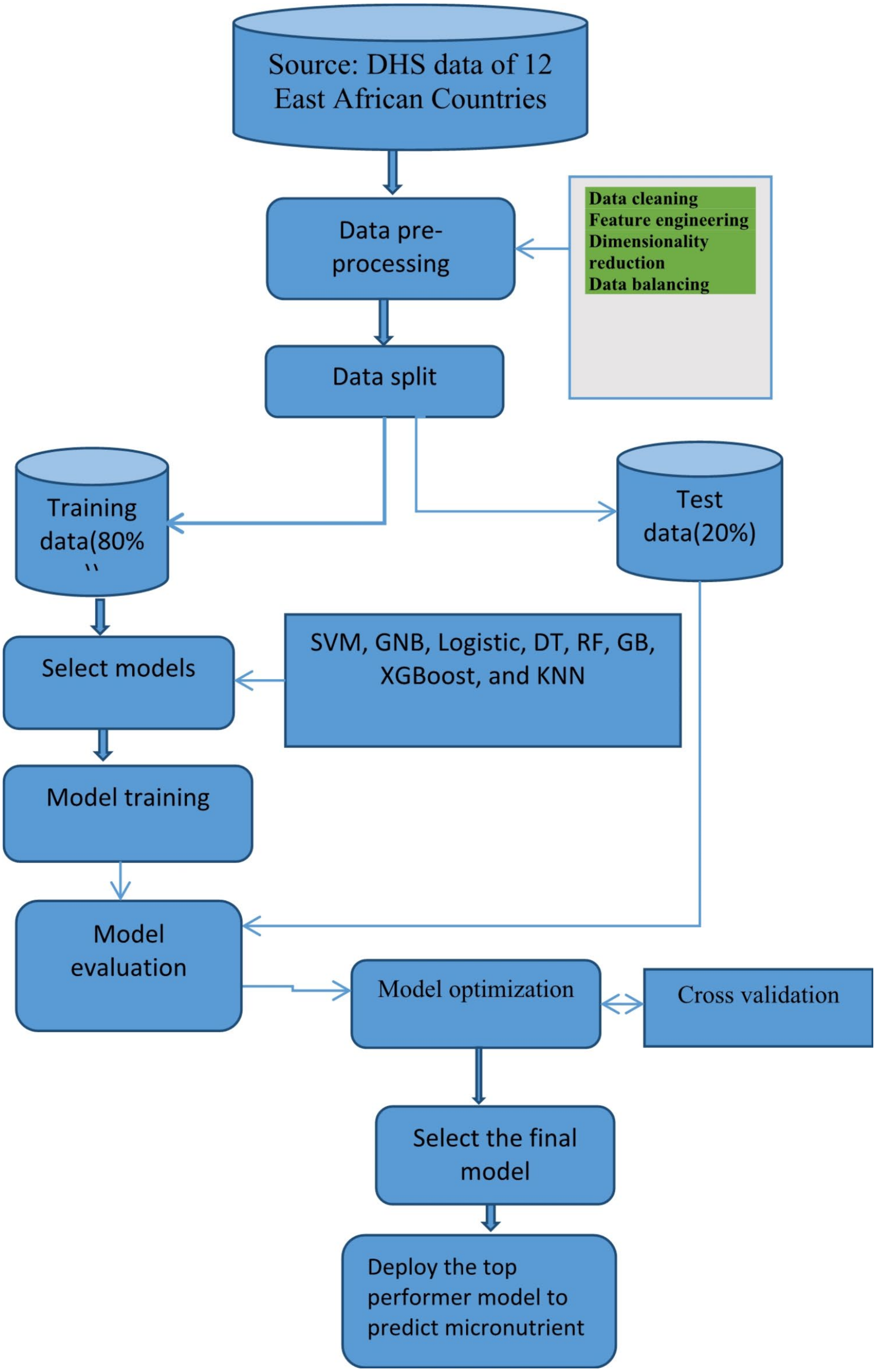

Fig. 1. Study work flow diagram.

the models using balanced datasets. To evaluate the performance of each model across each data balancing technique, we conducted a comprehensive analysis using five performance metrics: accuracy, precision, recall, F1-score, and AUC.

While accuracy is suitable for evaluating models with balanced classes, AUC becomes particularly valuable when dealing with imbalanced datasets or when the relative costs of false positives and false negatives are uncertain. However, it is advisable to consider both accuracy and AUC, along with other relevant metrics, to obtain a comprehensive evaluation of model performance and make informed comparisons between different machine learning algorithms[50–52]. Based on these considerations, we selected the data balancing technique that demonstrated superior performance for the final prediction.

Model selection and development

In our study, the dependent variable, micronutrient supplementation status, required a classification approach, as it was divided into "supplemented" and "not supplemented" categories. To make predictions, we needed to select appropriate classifiers. We utilized the Scikit-learn version 1.3.2 packages in Python, implemented within Jupyter Notebook, to employ machine learning algorithms.

To evaluate the predictive capabilities of machine learning algorithms for predicting micronutrient supplementation status, we employed eight state-of-the-art algorithms. Each algorithm has its unique approach and characteristics. Here are brief definitions for each of the algorithms we utilized:

1. Support Vector Machines (SVM) with Kernel Methods: SVM is a powerful algorithm used for both classification and regression tasks. It constructs a hyperplane or a set of hyperplanes in a high-dimensional space to separate different classes[53].
2. Gaussian Naive Bayes: This algorithm is based on Bayes' theorem, assuming that features are conditionally independent. It is particularly effective when dealing with high-dimensional data[54,55].
3. Logistic Regression: Logistic regression is a statistical model that is used for binary classification. It estimates the probability of an instance belonging to a particular class based on the input features[56,57].
4. Decision Tree Classifier: Decision trees are hierarchical structures that make decisions based on the values of input features. The decision tree classifier uses a tree-like model of decisions and their possible consequences to predict the class label of instances[58].
5. Random Forest Classifier: Random forest is an ensemble learning method that combines multiple decision trees. It generates a set of decision trees and makes predictions by averaging the outputs of individual trees[59].
6. Gradient Boosting Machines: Gradient boosting is another ensemble learning method that combines multiple weak prediction models, typically decision trees, to create a strong predictive model. It trains new models to correct the mistakes made by previous models[60,61].
7. eXtreme Gradient Boosting (XGBoost): XGBoost is an optimized implementation of gradient boosting that provides better performance and scalability. It employs a variety of regularization techniques to prevent overfitting and enhance the overall predictive power[62].
8. KNN: KNN is a non-parametric algorithm that classifies instances based on their similarity to neighboring instances. It assigns a class label to an instance by considering the labels of its k nearest neighbors in the feature space[63].

The selection of these algorithms was based on their suitability for classification tasks and their compatibility with the characteristics of our dataset[64–66].

Model training and evaluation

In order to construct a reliable predictive model within machine learning, it is essential to perform model training and evaluation[67,68]. In this particular study, a straightforward approach was employed, where the data was divided into an 80% training set and a 20% testing set. This division allowed us to assess the performance of each predictive model effectively.

To evaluate the performance of the predictive models, several metrics were utilized, including accuracy, precision, recall, F1-score, and AUC. Each of these metrics provides valuable insights into different aspects of the model's performance.

1. Accuracy: Accuracy measures the overall correctness of the model's predictions. It is calculated as the ratio of the number of correct predictions to the total number of predictions.
2. Precision: Precision evaluates the accuracy of positive predictions made by the model. It quantifies the proportion of true positive predictions out of the total predicted positives. The formula for precision is:

Precision = TP / (TP + FP) where TP represents true positive and FP represents false positive.

3. Recall: Recall, also known as sensitivity or true positive rate, assesses the model's ability to identify all positive instances. It measures the proportion of true positive predictions out of the total actual positives. The formula for recall is:

Recall = TP / (TP + FN) where TP represents true positive and FN represents false negative.

4. F1-score: The F1-score provides a balanced measure of a model's performance by considering both precision and recall. It is the harmonic mean of precision and recall, and it is calculated using the following formula:

F1-score = 2 * (Precision * Recall) / (Precision + Recall).

5. AUC: AUC is a metric calculated from the area under the receiver operating characteristic (ROC) curve. The ROC curve represents the true positive rate (TPR) plotted against the false positive rate (FPR) at various classification thresholds. AUC indicates the algorithm's ability to discriminate between classes, where a higher AUC value suggests better discrimination.

In summary, by utilizing these metrics, we were able to comprehensively evaluate the performance of each predictive model in terms of overall correctness, accurate positive predictions, identification of positive instances, balanced measure, and discriminatory ability[69].

In order to further evaluate the performance of the model, tenfold cross-validation techniques were employed. Prior to this, different k-fold validation techniques, including three-fold, five-fold, and ten-fold, were compared to determine the most suitable approach[70].

The study also conducted a thorough analysis of hyperparameters to refine and improve the model's performance. Grid search, random search, and Bayesian optimization were systematically explored to find the best hyperparameter settings. Comparing the outcomes from these techniques helped identify the configurations that provided the highest performance. In order to improve the accuracy and dependability of the model used in this study, we conducted model calibration. Through fine-tuning the model via calibration, we enhanced its performance in accurately predicting the desired outcome.

In our study, we conducted an extensive comparison of different kernel methods for the SVM model, with the main goal of identifying the most suitable kernel function to maximize the model's performance. We carefully evaluated and compared various kernel functions, including linear, polynomial, radial basis function (RBF), and sigmoid. Through meticulous analysis, our objective was to select the kernel method that produced the most favorable outcomes and achieved optimal performance for the SVM model[71].

Model interpretability

In our comprehensive approach to understanding the data and exploring the factors influencing the prediction of micronutrient supplementation, we employed various techniques including the Apriori algorithm. Firstly, we calculated the mean SHAP (Shapley Additive exPlanations) values to assess the average impact of each feature on the model's predictions, providing insights into the relative significance of different variables. This allowed us to understand the individual contributions of each feature.

Additionally, we utilized the Apriori algorithm, a popular algorithm for association rule mining. By applying the Apriori algorithm, we were able to uncover hidden patterns and relationships among the variables in the dataset. The algorithm allowed us to discover frequent item sets and association rules based on measures such as lift and confidence. Lift helped us determine the strength of the associations between different variables, indicating the degree to which the presence of one variable influences the likelihood of another variable occurring. Confidence, on the other hand, provided us with a measure of reliability or certainty in the association rules, indicating how often the consequent variable appeared when the antecedent variable was present[72–74].

By incorporating mean SHAP values and the Apriori algorithm into our analysis, we gained a deeper understanding of the dataset and the factors influencing the prediction of micronutrient supplementation. These techniques allowed us to uncover concealed patterns and relationships, leading to robust predictions and identification of influential factors. In summary, our approach involved calculating mean SHAP values to determine feature importance and utilizing the Apriori algorithm for association rule mining. This comprehensive methodology provided us with valuable insights into the dataset, enhancing model interpretability and facilitating a better understanding of the factors impacting micronutrient supplementation predictions[75–77].

Results

Descriptive results of the background characteristics

The study encompassed a comprehensive analysis of descriptive and socio-demographic characteristics among a weighted sample of 138,426 pregnant women. Among the participants, the largest proportion, comprising 57,174 (41.30%), fell within the age group of 26 to 34 years. In terms of residence, the majority, accounting for 105,613 (76.30%) of the study participants, hailed from rural areas. Regarding employment status, a significant number of respondents, totaling 101,407 (73.26%), were employed (See Table 1 for more detailed information).

Micronutrient supplementation status in east African countries

According to the specified DHS dataset, the pooled prevalence of micronutrient supplementation status among pregnant women in East Africa was found to be 28.90% (95% CI: 28.68, 29.12). Ethiopia had the lowest rate of micronutrient supplementation among pregnant women, with only 7.80% receiving supplementation. On the other hand, Zambia had the highest prevalence of micronutrient supplementation, with 66.94% of pregnant women receiving supplementation (See Fig. 2 for more detailed information).

Machine learning analysis of micronutrient status

Feature selection

Upon evaluating various feature selection methods, we observed that the Boruta algorithm exhibited strong performance. As depicted in Fig. 3, the algorithm effectively visualized the importance of variables, with significant variables highlighted in green, unimportant variables in red, and tentative variables in yellow. Tentative variables are those that require further investigation[78].

In our comprehensive analysis, we decided to exclude birth interval and marital status from consideration, as the Boruta algorithm deemed them unimportant. No tentative variables were identified. Consequently, we proceeded with utilizing the variables selected by the Boruta algorithm to predict the micronutrient supplementation status and explore data patterns through association rule mining.

Data balancing

In Table 2, various data balancing techniques, such as under-sampling, over-sampling, ADASYN, and SMOTE, were compared. Among the evaluated techniques, SMOTE demonstrated the highest performance, with the

Variable		Frequency	Percent
Residence	Urban	32,813	23.70%
	Rule	105,613	76.30%
Religion	Catholic	35,962	28.06%
	Protestant	44,319	34.58%
	Muslim	12,952	10.11%
	Adventist	11,847	9.24%
	Jehovah	14,498	11.31%
	Tradition animist	5,824	4.54%
	No religion	816	0.64%
	Sect	491	0.38%
	Other	1,440	1.12%
Educational status	No education	34,780	25.13%
	Primary education	72,182	52.14%
	Secondary education	27,016	19.52%
	Higher education	4,448	3.21%
Age (in years)	15–20	14,662	10.59%
	21–25	35,573	25.70%
	26–34	57,174	41.30%
	35–49	31,017	22.41%
Marital status	Single	6,681	4.83%
	Married	117,994	85.24%
	Widowed	2,404	1.74%
	Separated/divorced	11,347	8.20%

Table 1. Individual characteristics of reproductive age group women in east African countries ($n = 138{,}426$).

random forest classifier achieving an AUC of 0.878 and an accuracy of 91%. The results indicated that SMOTE outperformed other data balancing methods.

Development and performance comparisons of machine learning-based models
By utilizing these performance metrics such as accuracy, precision, recall, F1 score, and AUC, we conducted a comprehensive evaluation to determine how effectively the algorithms could predict micronutrient supplementation. Table 3 presents the performance measures of multiple selected algorithms after data balancing and calibrated tuning processes had been applied. Based on the evaluation results, the top three machine learning algorithms for classifying micronutrient supplementation status were found to be the random forest classifier, decision tree classifier, and XGBoost with excellent ROC value (**See** Table 3).

Figure 4 illustrates the ROC curve analysis conducted on selected machine learning algorithms. These algorithms were trained on balanced data using the SMOTE data balancing technique and underwent optimized hyperparameter tuning. Among the different hyperparameter tuning techniques experimented with, grid search proved to be the most suitable for our dataset. The application of hyperparameter tuning resulted in a significant enhancement in the performance of our model. For a detailed comparison, please refer to Figs. 4 and 5, which respectively present the ROC curve analysis before and after hyperparameter tuning.

As depicted in Fig. 5, the final ROC curve for the tuned model showcased that the random forest classifier outperformed all other machine learning algorithms, exhibiting an AUC of 0.892. The decision tree classifier and XGBoost followed closely behind with AUC values of 0.862 and 0.856, respectively, which can be considered excellent. The KNN, Gradient boosting classifier, and SVM achieved reasonably acceptable ROC values of 0.797, 0.739, and 0.721, respectively. However, the logistic regression and Gaussian Naïve Bayes algorithms displayed lower AUC values of 0.683 and 0.651, respectively, indicating suboptimal discrimination.

We also provided a comprehensive comparison of model performance using various performance metrics. In cases where the machine learning algorithms exhibited similar performance and distinguishing the superior algorithm became challenging, we used a comprehensive analysis. For further details, please refer to Fig. 6.

Model interpretability
SHAP value interpretation
Based on the findings presented in Fig. 7, the mean SHAP value report provided insights into the relative importance of different features in the classification model. ANC visits, number of living children, and media exposure emerged as the most influential factors and exhibited high mean SHAP values. This indicated that these features had high significant impact on the model's predictions.

Additionally, religion, working status, education status, age group, and wealth status displayed minimal influence on the classification outcome, as evidenced by their low mean SHAP values. These features contribute less to the model's decision-making process and have limited importance on the model prediction. On the other

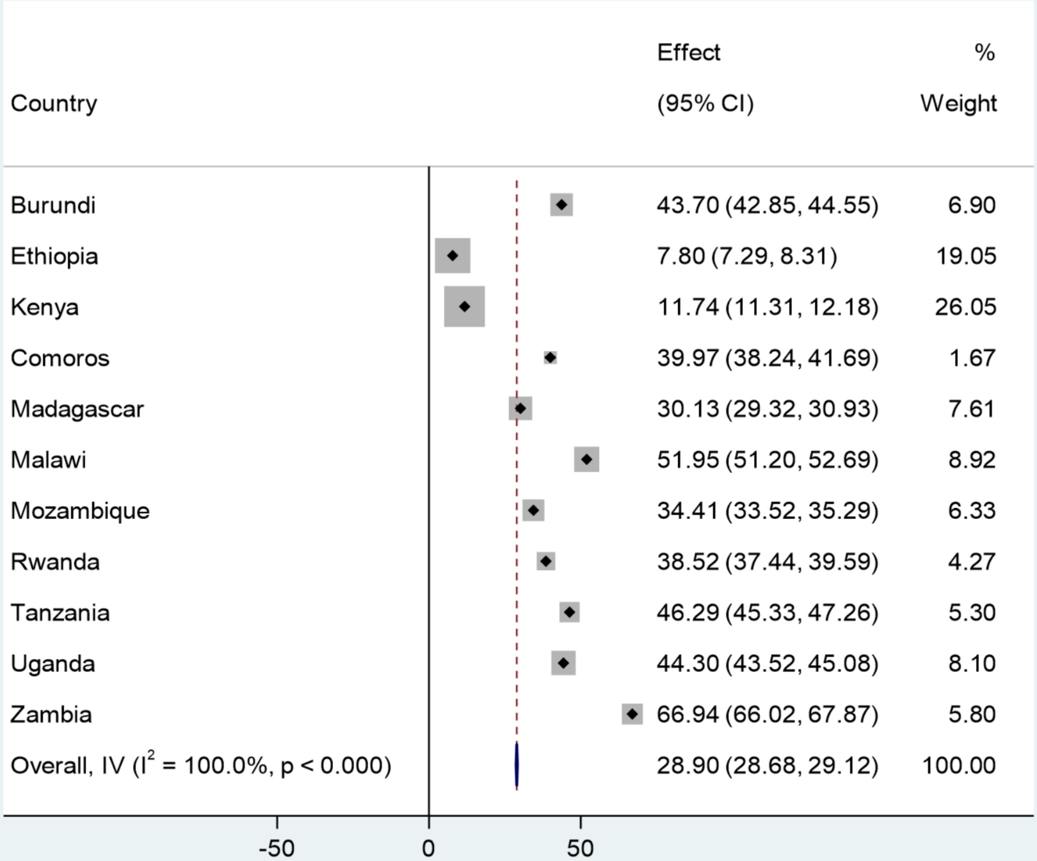

Fig. 2. Prevalence of micronutrient supplementation status among pregnant women in East Africa countries using forest tree plot.

hand, residence and perceived distance were positioned at the bottom of the graph, with a mean SHAP value of zero. This suggests that these features do not contribute to the model prediction.

Association rule mining

Using the Apriori algorithm, our research identified the most influential association rules based on their lift values and confidence. These rules provided valuable insights into the probability of micronutrient supplementation during pregnancy in East African countries. Remarkably, the recurring presence of factors such as maternal education level, employment status, ANC visit, media accessibility, number of living children, and religion in these association rules indicated their consistent association with the likelihood of receiving micronutrient supplementation.

The top five association rules and their corresponding lift values are as follows:

1. 1. If the mother has a secondary education, is employed, has more than four ANC visits, and has media exposure, the probability of being supplemented with micronutrients is 86.3% (Confidence = 0.863 and lift = 2.36).
2. 2. If the mother is in the age group of 21 to 25, has less than three children, has more than four ANC visits, and follows the catholic religion, the probability of being supplemented with micronutrients is 85.9% (Confidence = 0.859 and lift = 2.04).
3. 3. If the mother has less than three children, access to media, follows the Catholic religion, and has higher education, the probability of being supplemented with micronutrients is 81.4% (Confidence = 0.814 and lift = 1.64).
4. 4. If the mother has less than three children, is employed, in the age group of 26 to 34, has higher education, and follows the traditional or animist religion, the probability of being supplemented with micronutrients is 78.4% (Confidence = 0.784 and lift = 1.83).
5. 5. If the mother has four ANC visits, is employed, has access to media, belongs to a high-income wealth status, and has a primary education, the probability of being supplemented with micronutrients is 74.3% (Confidence = 0.743 and lift = 1.25).

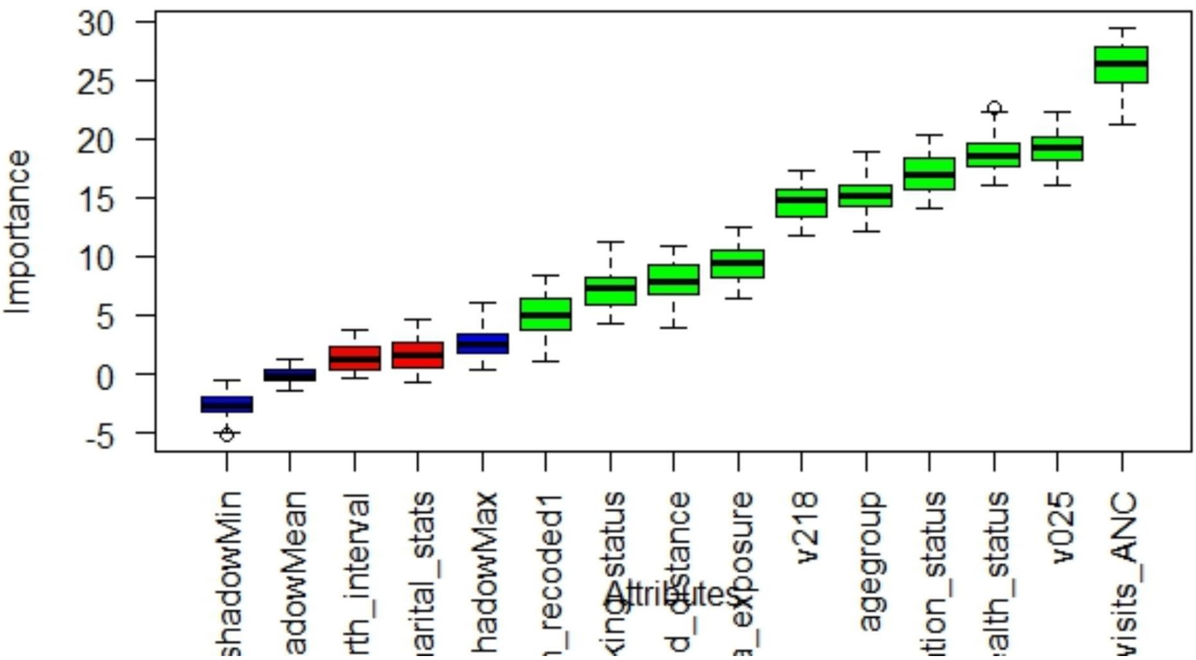

Fig. 3. Feature selection using Boruta algorithm (Note: v218 = number of children, v025 = residence).

Algorithms	Comparison method	Unbalanced data	Under-sampling	Over-sampling	ADASYN	SMOTE
SVM	Accuracy (%)	65.0%	69.0%	74.0%	66.0%	72.0%
	AUC	0.508	0.600	0.529	0.513	0.491
Gaussian naive baye	Accuracy (%)	59.0%	60.0%	67.0%	59.0%	68.0%
	AUC	0.660	0.668	0.691	0.632	0.669
Logistic regression	Accuracy (%)	64.0%	71.0%	**73.0%**	60.0%	72.0%
	AUC	**0.684**	**0.703**	**0.703**	0.657	0.698
Decision tree classifier	Accuracy (%)	56.0%	65.0%	70.0%	77.0%	88.0%
	AUC	0.467	0.504	0.467	0.825	0.829
Random forest classifier	Accuracy (%)	60.0%	71.0%	76.0%	**88.0%**	**91.0%**
	AUC	0.635	0.656	0.635	**0.853**	**0.878**
Gradient boosting classifier	Accuracy (%)	**67.0%**	**76.0%**	73.0%	87.0%	87.0%
	AUC	0.639	0.636	0.639	0.728	0.748
XGBoost	Accuracy (%)	66.0%	69.0%	**73.0%**	76.0%	78.0%
	AUC	0.565	0.594	0.577	0.626	0.637
KNN	Accuracy (%)	61.0%	68.0%	71.0%	69.0%	88.0%
	AUC	0.606	0.632	0.613	0.784	0.815

Table 2. Compression of imbalanced data handling techniques using accuracy and AUC values.

Discussion

The study demonstrated the potential of machine learning algorithms in accurately predicting the status of micronutrient supplementation among pregnant women in East Africa. The random forest classifier, decision tree classifier, and XGBoost were identified as the most effective models for classifying supplementation status. Specifically, the random forest classifier outperformed other algorithms, with an AUC value of 0.892 and an accuracy of 94.0%. Similar studies conducted in Rwanda[33], Zambia[79], Ethiopia[80], Mozambique, and Nigeria[81] also found the random forest model to be superior in predicting various health outcomes.

Models	Accuracy	AUC
SVM	86.0%	0.721
Gaussian naive baye	77.0%	0.651
Logistic regression	74.0%	0.683
Decision tree classifier	92.0%	0.862
Random forest classifier	94.0%	0.892
Gradient boosting classifier	86.0%	0.739
XGBoost	92.0%	0.856
KNN	91.0%	0.797

Table 3. Accuracy and AUC value of the selected machine algorithm after data balancing and tuning.

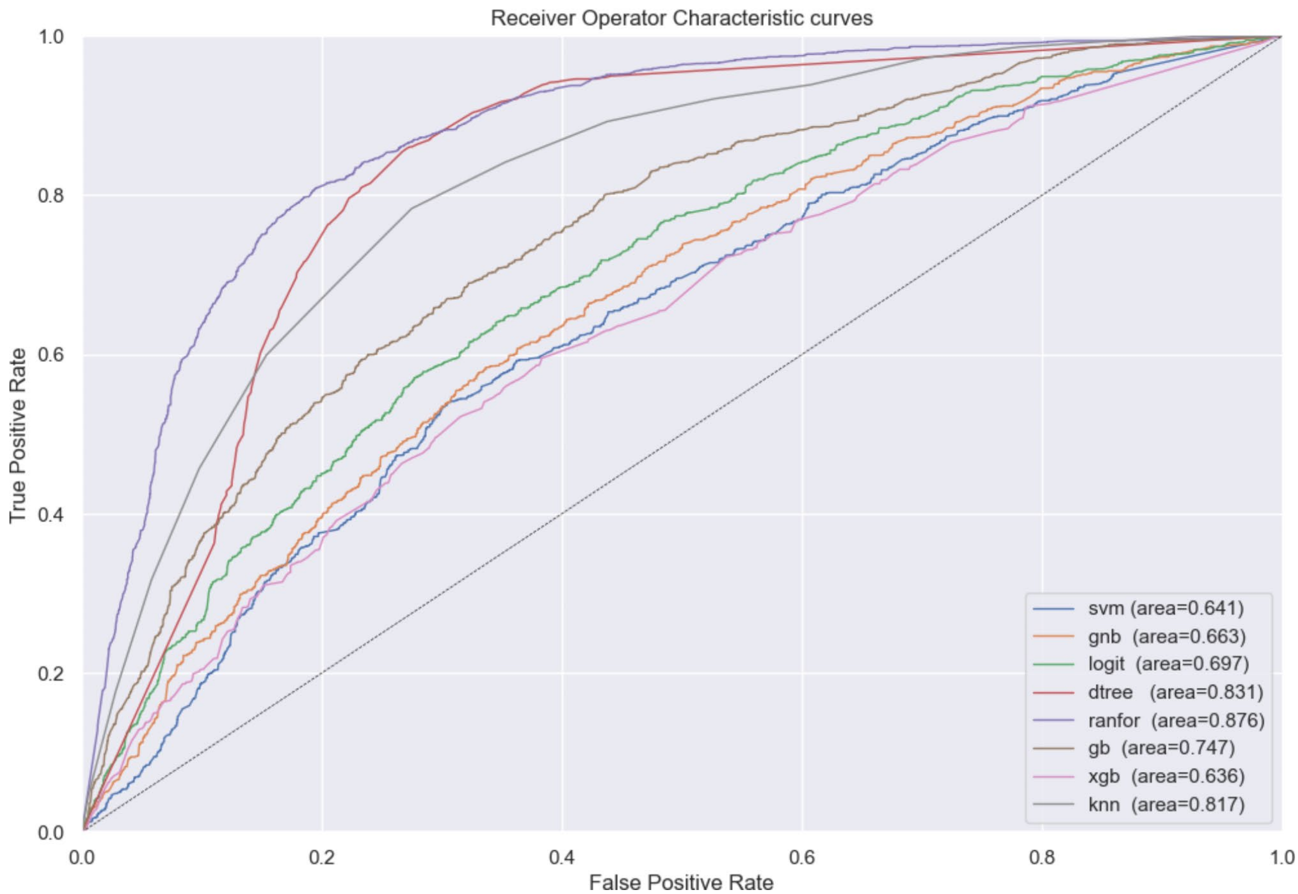

Fig. 4. ROC curve analysis of selected machine learning algorithms with balanced data using SMOTE data balancing technique.

We used association rule mining to identify the top predictors in our data set. Accordingly, we built five top rules using association rule mining and implied that the mother's education level consistently appeared in association rules and had a significant impact on the likelihood of receiving micronutrient supplementation. The justification for this impact was supported by the existing research literature. Similar to our findings studies conducted in Central Ethiopia[21], Kenya[19], and Bangladesh[20], and elsewhere in the world[13–21] noticed the importance of education for enhancing micronutrient supplementation status. One possible justification for this finding could be that education empowers individuals to make informed decisions regarding their health and encourages adherence to prenatal care.

Based on the finding of the association rule mining, ANC visits were a strong predictor that influenced the likelihood of receiving micronutrient supplementation. This finding was supported by global studies highlighting the significance of ANC visits in enhancing supplementation status[18,27,28]. The possible justification could be ANC visits may serve as a platform for healthcare providers to educate pregnant women about the importance of proper nutrition and micronutrient supplementation, addressing misconceptions and providing information on the specific benefits of supplementation through counseling sessions.

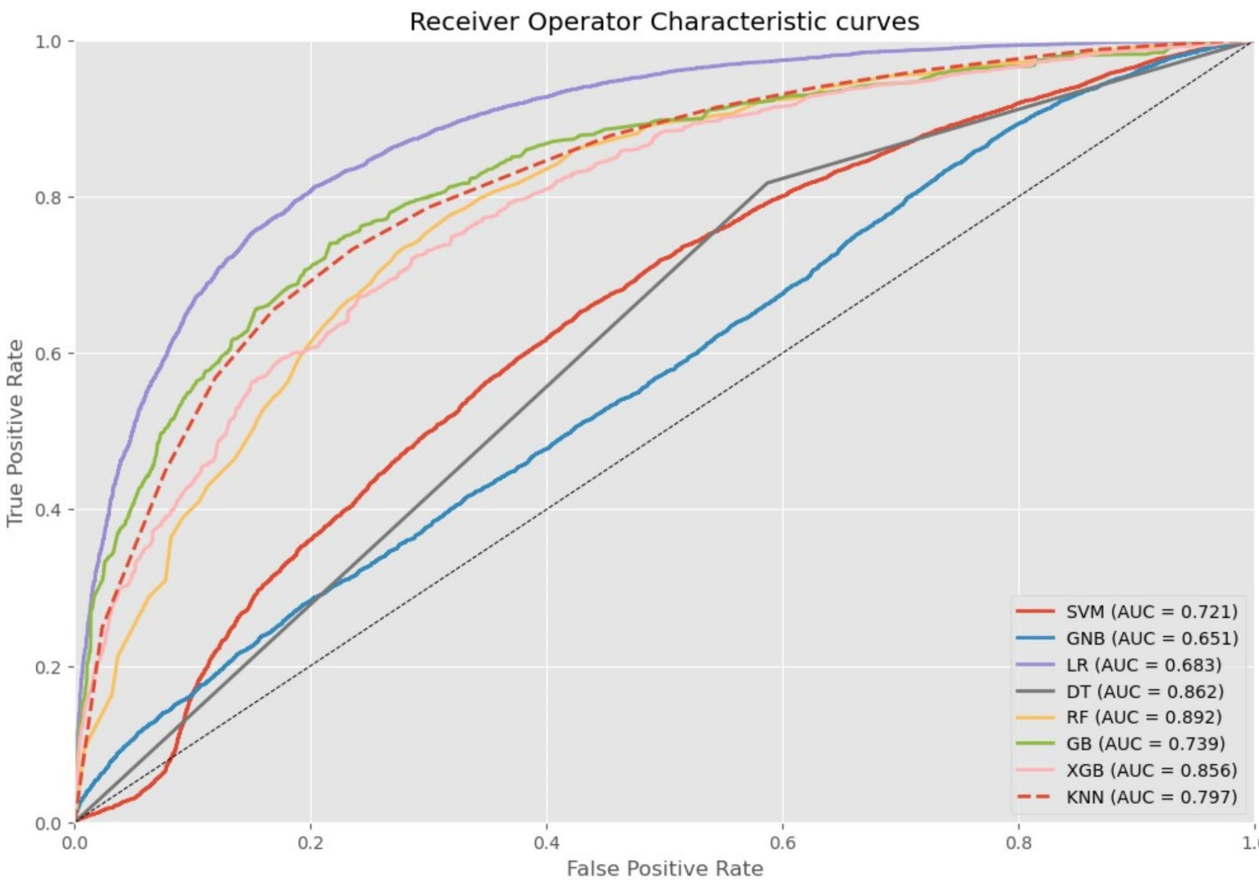

Fig. 5. ROC curve analysis of selected machine learning algorithms after optimized Hyperparameter tuning using balanced data.

Association rule mining consistently demonstrated that religion played a significant role in determining the likelihood of receiving micronutrient supplementation. This finding underscores the importance of considering religious factors in public health interventions and highlights the need for culturally sensitive strategies to address micronutrient deficiencies in diverse religious communities[82].

The association rule mining consistently demonstrated that an increase in the number of children had a negative association with the likelihood of receiving micronutrient supplementation. This association was supported by the literature, which underscores the significance of family size as a determinant of access to and utilization of health services[14]. The possible reason for this finding is that the added responsibilities of raising a larger family may pose challenges for parents, especially busy mothers, in prioritizing and accessing healthcare, including attending antenatal care visits and receiving the necessary micronutrient supplements.

The association rule mining consistently provided evidence that individuals with media exposure had a higher probability of receiving adequate micronutrient supplementation. This finding was further supported by the literature, which emphasized the role of media in influencing health-related behaviors[14]. The justification for this finding could be media platforms provide opportunities to educate individuals about the benefits of adequate nutrition and the availability of supplementation programs.

The consistent findings from association rule mining indicate that individuals who had primary, secondary, and higher levels of education are more likely to receive micronutrient supplementation. Previous studies conducted worldwide supported this relationship[13–21]. This could be attributed to the fact that individuals with higher education levels may possess a deeper understanding of the significance of proper nutrition and the specific advantages of taking micronutrient supplements during pregnancy.

The association rule mining consistently revealed the significant impact of employment status on the likelihood of receiving micronutrient supplementation. This finding was consistent with a study conducted in Northwest Ethiopia[26]. One possible justification is that employed women may have greater exposure to health-related information through workplace wellness programs, employee benefits, or interactions with colleagues.

Strength and limitations of the study

The study demonstrates its strength by employing a comprehensive analysis of predictive capabilities through the use of eight supervised machine-learning algorithms. This approach enhances the reliability and credibility of the findings by revealing hidden patterns and relationships within the data. However, it is important to

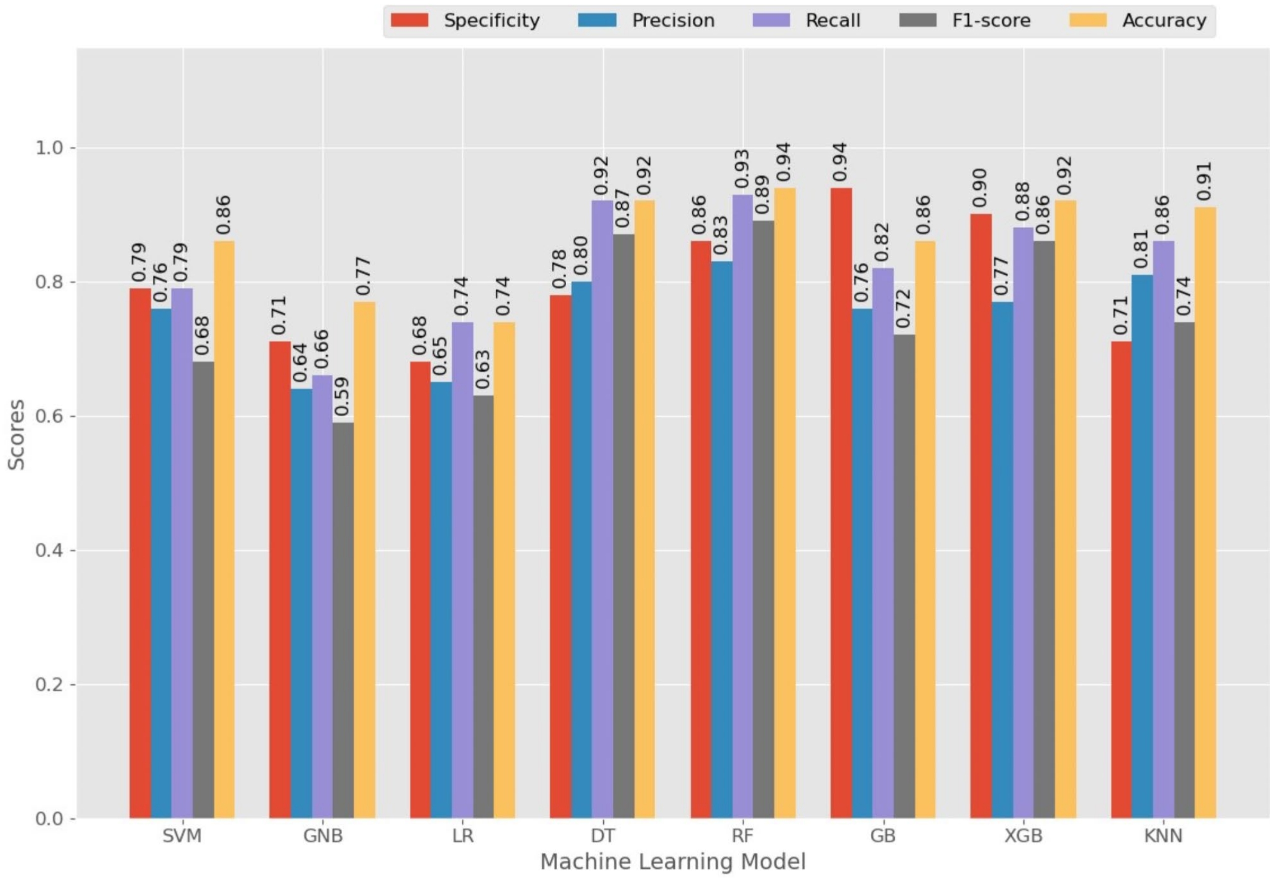

Fig. 6. Overall model performance of selected machine learning algorithms after optimized hyperparameter tuning of a balanced data.

acknowledge certain limitations associated with the study. The retrospective data collection and reliance on secondary data introduce potential drawbacks, including the possibility of incomplete or missing data, which may impact the accuracy and introduce biases in the results.

Additionally, the study only explores four data balancing techniques, which may limit the ability to fully harness the potential power of each machine learning algorithm. Another limitation is the absence of exploring the performance of each algorithm through ensembling, a technique that combines multiple models to enhance predictive accuracy. Lastly, it is worth noting that the study shares the inherent limitation of building association rules solely based on the Apriori algorithm.

Conclusion

The study highlights the effectiveness of machine learning in accurately predicting the status of micronutrient supplementation among pregnant women in East Africa. Notably, the random forest classifier demonstrated exceptional performance, achieving excellent evaluation scores with an AUC of 0.892 and an accuracy of 94.0% in predicting the supplementation status. This emphasizes the potential of machine learning algorithms as valuable tools for policymakers and healthcare providers to develop targeted strategies aimed at improving the uptake of micronutrient supplementation among pregnant women.

Based on the study's findings, several key intervention components are recommended to enhance the utilization of micronutrient supplementation. Strengthening educational initiatives can provide vital information on the importance and benefits of supplement intake. Improving antenatal care services ensures proper screening, monitoring, and guidance for pregnant women regarding their nutritional needs. Comprehensive media campaigns can effectively raise awareness and emphasize the significance of micronutrient supplementation during pregnancy. It is also crucial to consider cultural and religious sensitivities when designing interventions to ensure their acceptance and effectiveness within the specific population.

Furthermore, researchers are encouraged to explore and experiment with various techniques to optimize algorithm performance, leading to the identification of the most effective predictors and enhanced accuracy in predicting micronutrient supplementation status.

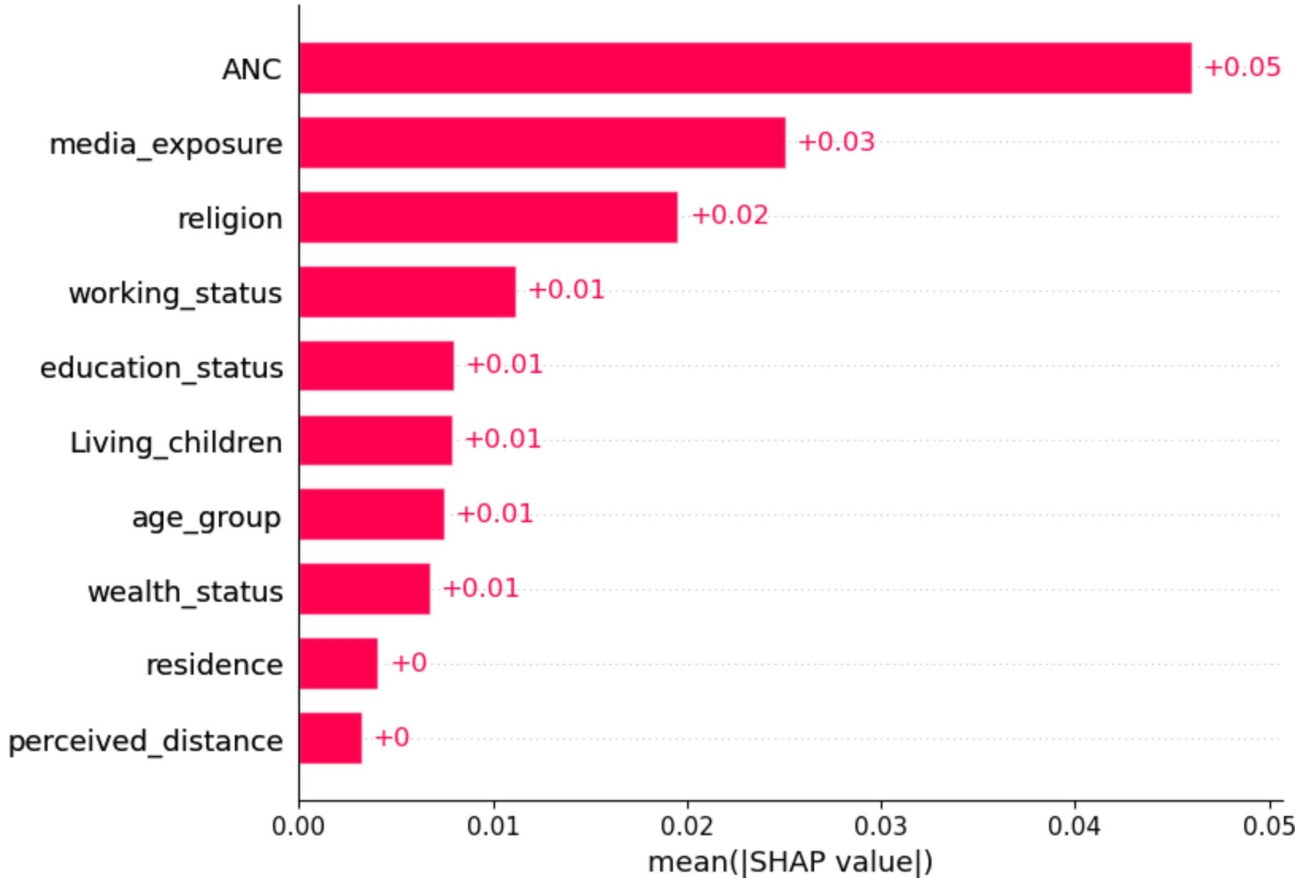

Fig. 7. A mean SHAP value report.

Data availability

To access the data used in the study, it is necessary to log in to the official website of the DHS: http://www.dhsprogram.com.

Received: 20 December 2023; Accepted: 4 October 2024
Published online: 11 October 2024

References

1. Corvallis Minerals Micronutrient Information Center, Oregon State University: Linus Pauling Institute; updated Nov 21, 2023. (2018). https://lpi.oregonstate.edu/mic
2. WHO. Ambition and action in nutrition 2016–2025 Geneva: WHO. (2016). https://www.who.int/health-topics/micronutrients#tab=tab_1
3. Godswill, A. G., Somtochukwu, I. V., Ikechukwu, A. O. & Kate, E. C. Health benefits of micronutrients (vitamins and minerals) and their associated deficiency diseases: a systematic review. *Int. J. Food Sci.* **3** (1), 1–32 (2020).
4. Punia, S. & Kumar, M. *Functionality and Application of Colored Cereals: Nutritional* (Elsevier, 2022).
5. Von Grebmer, K. et al. *Synopsis: 2014 Global Hunger Index: The Challenge of Hidden Hunger* (Intl Food Policy Res Inst, 2014).
6. Bailey, R. L., West, K. P. Jr. & Black, R. E. The epidemiology of global micronutrient deficiencies. *Ann. Nutr. Metab.* **66**(Suppl. 2), 22–33 (2015).
7. Gernand, A. D., Schulze, K. J., Stewart, C. P., West, K. P. Jr. & Christian, P. Micronutrient deficiencies in pregnancy worldwide: health effects and prevention. *Nat. Rev. Endocrinol.* **12**(5), 274–289 (2016).
8. Haider, B. A., Yakoob, M. Y. & Bhutta, Z. A. Effect of multiple micronutrient supplementation during pregnancy on maternal and birth outcomes. *BMC Public. Health.* **11**, 1–9 (2011).
9. Karami, M., Chaleshgar, M., Salari, N., Akbari, H. & Mohammadi, M. Global prevalence of anemia in pregnant women: a comprehensive systematic review and meta-analysis. *Matern. Child Health J.* **26** (7), 1473–1487 (2022).
10. Stevens, G. A. et al. National, regional, and global estimates of anaemia by severity in women and children for 2000–19: a pooled analysis of population-representative data. *Lancet Global Health.* **10** (5), e627–e39 (2022).
11. Redón Lago, A. Hidden hunger in the current world: causes, consequences and solutions to a global public health challenge and a particular look at Spain. (2021).
12. Galani, Y., Orfila, C. & Gong, Y. A review of micronutrient deficiencies and analysis of maize contribution to nutrient requirements of women and children in Eastern and Southern Africa. *Crit. Rev. Food Sci. Nutr.* **62** (6), 1568–1591 (2022).
13. Brunst, K. J. et al. Racial/ethnic and sociodemographic factors associated with micronutrient intakes and inadequacies among pregnant women in an urban US population. *Public Health. Nutr.* **17** (9), 1960–1970 (2014).
14. Serra-Majem, L. et al. Determinants of nutrient intake among children and adolescents: results from the enKid Study. *Ann. Nutr. Metab.* **46**(Suppl. 1), 31–38 (2002).

15. Cembranel, F., Wagner, K. J. P., González-Chica, D. A. & d'Orsi, E. Education and income levels are associated with energy and micronutrient intake. Int. J. Vitam. Nutr. Res. (2019).
16. Liu, F-L. et al. Nutrient intakes of pregnant women and their associated factors in eight cities of China: a cross-sectional study. *Chin. Med. J.* **128** (13), 1778–1786 (2015).
17. Shalini, T. et al. Micronutrient intakes and status assessed by probability approach among the urban adult population of Hyderabad city in South India. *Eur. J. Nutr.* **58**, 3147–3159 (2019).
18. Yeneabat, T. et al. Maternal dietary diversity and micronutrient adequacy during pregnancy and related factors in East Gojjam Zone, Northwest Ethiopia, 2016. *BMC Pregnancy Childbirth.* **19** (1), 1–9 (2019).
19. Kiboi, W., Kimiywe, J. & Chege, P. Determinants of dietary diversity among pregnant women in Laikipia County, Kenya: a cross-sectional study. *BMC Nutr.* **3** (1), 1–8 (2017).
20. Shamim, A. A. et al. Pregnant women diet quality and its sociodemographic determinants in southwestern Bangladesh. *FoodNutr. Bull.* **37** (1), 14–26 (2016).
21. Kebede, A. N., Sahile, A. T. & Kelile, B. C. Dietary diversity and Associated factors among pregnant women in Addis Ababa, Ethiopia, 2021. *Int. J. Public Health* **67**, 1605377 (2022).
22. Thillan, K. et al. Micronutrient status and associated factors of adiposity in primary school children with normal and high body fat in Colombo municipal area, Sri Lanka. *BMC Pediatr.* **21** (1), 1–14 (2021).
23. Gebremedhin, T., Aschalew, A. Y., Tsehay, C. T., Dellie, E. & Atnafu, A. Micronutrient intake status and associated factors among children aged 6–23 months in the emerging regions of Ethiopia: a multilevel analysis of the 2016 Ethiopia demographic and health survey. *PloS One.* **16** (10), e0258954 (2021).
24. Nguyen, P. H. et al. Micronutrient intakes among women of reproductive age in Vietnam. *PLoS One.* **9** (2), e89504 (2014).
25. Tzioumis, E., Kay, M. C., Bentley, M. E. & Adair, L. S. Prevalence and trends in the childhood dual burden of malnutrition in low- and middle-income countries, 1990–2012. *Public Health Nutr.* **19**(8), 1375–1388 (2016).
26. Awoke, M. et al. Micronutrient intake inadequacy and its associated factors among lactating women in Bahir Dar City, Northwest Ethiopia, 2021. *Plos One.* **17** (7), e0271984 (2022).
27. Nguyen, P. H. et al. Integrating nutrition interventions into an existing maternal, neonatal, and child health program increased maternal dietary diversity, micronutrient intake, and exclusive breastfeeding practices in Bangladesh: results of a cluster-randomized program evaluation. *J. Nutr.* **147** (12), 2326–2337 (2017).
28. Haileslassie, K., Mulugeta, A. & Girma, M. Feeding practices, nutritional status and associated factors of lactating women in Samre Woreda, South Eastern Zone of Tigray, Ethiopia. *Nutr. J.* **12**, 1–11 (2013).
29. Sarker, I. H. Machine learning: algorithms, real-world applications and research directions. *SN Comput. Sci.* **2** (3), 160 (2021).
30. Jordan, M. I. & Mitchell, T. M. Machine learning: Trends, perspectives, and prospects. *Science.* **349** (6245), 255–260 (2015).
31. Enyew, E. B. et al. Micronutrient intake and associated factors among pregnant women in East Africa: Multilevel logistic regression analysis. *Plos One.* **18** (4), e0281427 (2023).
32. Tesfaye, B., Atique, S., Azim, T. & Kebede, M. M. Predicting skilled delivery service use in Ethiopia: dual application of logistic regression and machine learning algorithms. *BMC Med. Inf. Decis. Mak.* **19** (1), 1–10 (2019).
33. Mfateneza, E., Rutayisire, P. C., Biracyaza, E., Musafiri, S. & Mpabuka, W. G. Application of machine learning methods for predicting infant mortality in Rwanda: analysis of Rwanda demographic health survey 2014–15 dataset. *BMC Pregnancy Childbirth.* **22** (1), 388 (2022).
34. Guide to DHS Statistics DHS-7 (version 2) [Internet]. USAID. [cited November 24, 2023]. (2020). https://dhsprogram.com/data/Guide-to-DHS-Statistics/index.cfm
35. WHO. World Health Organization recommendations on antenatal care for a positive pregnancy experience Geneva2018 [https://www.who.int/publications/i/item/9789241549912
36. Kadhim, A. I. An evaluation of preprocessing techniques for text classification. *Int. J. Comput. Sci. Inform. Secur. (IJCSIS).* **16** (6), 22–32 (2018).
37. Liu, X. et al. Machine learning-based correlation study between perioperative immunonutritional index and postoperative anastomotic leakage in patients with gastric cancer. *Int. J. Med. Sci.* **19** (7), 1173 (2022).
38. Anand, H. & Vinodchandra, S. (eds) Applying correlation threshold on Apriori algorithm. 2013 IEEE International Conference ON Emerging Trends in Computing, Communication and Nanotechnology (ICECCN); : IEEE. (2013).
39. Zheng, A. & Casari, A. *Feature Engineering for Machine Learning: Principles and Techniques for data Scientists* (O'Reilly Media, Inc., 2018).
40. Al-Shehari, T. & Alsowail, R. A. An insider data leakage detection using one-hot encoding, synthetic minority oversampling and machine learning techniques. *Entropy.* **23** (10), 1258 (2021).
41. Rawat, S., Rawat, A., Kumar, D. & Sabitha, A. S. Application of machine learning and data visualization techniques for decision support in the insurance sector. *Int. J. Inform. Manag. Data Insights* **1**(2), 100012 (2021).
42. Pudjihartono, N., Fadason, T., Kempa-Liehr, A. W. & O'Sullivan, J. M. A review of feature selection methods for machine learning-based disease risk prediction. *Front. Bioinf.* **2**, 927312 (2022).
43. Arafat, M. Y., Hoque, S., Xu, S. & Farid, D. M. Machine learning for mining imbalanced data. (2019).
44. Krawczyk, B. Learning from imbalanced data: open challenges and future directions. *Progress Artif. Intell.* **5** (4), 221–232 (2016).
45. Jadhav, A., Mostafa, S. M., Elmannai, H. & Karim, F. K. An empirical assessment of performance of data balancing techniques in classification task. *Appl. Sci.* **12** (8), 3928 (2022).
46. Bach, M., Werner, A., Żywiec, J. & Pluskiewicz, W. The study of under-and over-sampling methods' utility in analysis of highly imbalanced data on osteoporosis. *Inf. Sci.* **384**, 174–190 (2017).
47. Shelke, M. S., Deshmukh, P. R. & Shandilya, V. K. A review on imbalanced data handling using undersampling and oversampling technique. *Int. J. Recent. Trends Eng. Res.* **3** (4), 444–449 (2017).
48. Brandt, J. & Lanzén, E. A comparative review of SMOTE and ADASYN in imbalanced data classification. (2021).
49. Fernández, A., Garcia, S., Herrera, F. & Chawla, N. V. SMOTE for learning from imbalanced data: progress and challenges, marking the 15-year anniversary. *J. Artif. Intell. Res.* **61**, 863–905 (2018).
50. Carrington, A. M. et al. Deep ROC analysis and AUC as balanced average accuracy, for improved classifier selection, audit and explanation. *IEEE Trans. Pattern Anal. Mach. Intell.* **45** (1), 329–341 (2022).
51. Bekkar, M., Djemaa, H. K. & Alitouche, T. A. Evaluation measures for models assessment over imbalanced data sets. *J. Inf. Eng. Appl.* **3**(10), 156-168 (2013).
52. Japkowicz, N. Assessment metrics for imbalanced learning. Imbalanced learning: Foundations, algorithms, and applications. pp. 187–206. (2013).
53. Scholkopf, B. & Smola, A. J. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and beyond* (MIT Press, 2018).
54. Kamel, H., Abdulah, D. & Al-Tuwaijari, J. M. (eds) Cancer classification using gaussian naive bayes algorithm. 2019 international engineering conference (IEC); IEEE. (2019).
55. Jahromi, A. H. & Taheri, M. (eds) A non-parametric mixture of Gaussian naive Bayes classifiers based on local independent features. 2017 Artificial intelligence and signal processing conference (AISP); IEEE. (2017).
56. Christodoulou, E. et al. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J. Clin. Epidemiol.* **110**, 12–22 (2019).

57. Zou, X., Hu, Y., Tian, Z. & Shen, K. (eds) Logistic regression model optimization and case analysis. 2019 IEEE 7th international conference on computer science and network technology (ICCSNT); IEEE. (2019).
58. Priyanka, K. D. Decision tree classifier: a detailed survey. *Int. J. Inform. Decis. Sci.* **12** (3), 246–269 (2020).
59. Parmar, A., Katariya, R. & Patel, V. (eds) A review on random forest: An ensemble classifier. International conference on intelligent data communication technologies and internet of things (ICICI) 2018; Springer. (2019).
60. Lu, H., Karimireddy, S. P., Ponomareva, N. & Mirrokni, V. (eds) Accelerating gradient boosting machines. International conference on artificial intelligence and statistics; PMLR. (2020).
61. Bentéjac, C., Csörgő, A. & Martínez-Muñoz, G. A comparative analysis of gradient boosting algorithms. *Artif. Intell. Rev.* **54**, 1937–1967 (2021).
62. Chen, T. et al. Xgboost: extreme gradient boosting. *R Package Version 04–2*. **1** (4), 1–4 (2015).
63. Zhang, Z. Introduction to machine learning: k-nearest neighbors. *Ann. Transl. Med.* **4**(11), 382-393 (2016).
64. Ali, M. M. et al. Heart disease prediction using supervised machine learning algorithms: performance analysis and comparison. *Comput. Biol. Med.* **136**, 104672 (2021).
65. Dhall, D., Kaur, R. & Juneja, M. Machine learning: a review of the algorithms and its applications. Proceedings of ICRIC. : Recent Innovations in Computing. 2020:47–63. (2019).
66. Austin, P. C., Tu, J. V., Ho, J. E., Levy, D. & Lee, D. S. Using methods from the data-mining and machine-learning literature for disease classification and prediction: a case study examining classification of heart failure subtypes. *J. Clin. Epidemiol.* **66** (4), 398–407 (2013).
67. Luo, W. et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. *J. Med. Internet. Res.* **18** (12), e323 (2016).
68. Bowles, M. *Machine Learning in Python: Essential Techniques for Predictive Analysis* (Wiley, 2015).
69. Jiang, T., Gradus, J. L. & Rosellini, A. J. Supervised machine learning: a brief primer. *Behav. Ther.* **51** (5), 675–687 (2020).
70. Xu, Y. & Goodacre, R. On splitting training and validation set: a comparative study of cross-validation, bootstrap and systematic sampling for estimating the generalization performance of supervised learning. *J. Anal. Test.* **2** (3), 249–262 (2018).
71. Kung, S. Y. *Kernel Methods and Machine Learning* (Cambridge University Press, 2014).
72. Kumbhare, T.A. and Chobe, S.V. An overview of association rule mining algorithms. *International Journal of Computer Science and Information Technologies.* **5** (1), 927-930 (2014).
73. Hussein, N., Alashqur, A. & Sowan, B. Using the interestingness measure lift to generate association rules. *J. Adv. Comput. Sci. Technol.* **4** (1), 156 (2015).
74. Bao, F., Mao, L., Zhu, Y., Xiao, C. & Xu, C. An improved evaluation methodology for mining association rules. *Axioms*. **11** (1), 17 (2021).
75. Council, N. R. *Frontiers in Massive data Analysis* (National Academies, 2013).
76. Council, N. R. Statistical Analysis of Massive Data Streams: Proceedings of a Workshop: National Academies Press; (2004).
77. Nohara, Y., Matsumoto, K., Soejima, H. & Nakashima, N. Explanation of machine learning models using shapley additive explanation and application for real data in hospital. *Comput. Methods Programs Biomed.* **214**, 106584 (2022).
78. Chen, R-C., Dewi, C., Huang, S-W. & Caraka, R. E. Selecting critical features for data classification based on machine learning methods. *J. Big Data*. **7** (1), 52 (2020).
79. Chilyabanyama, O. N. et al. Performance of machine learning classifiers in classifying stunting among under-five children in Zambia. *Children*. **9** (7), 1082 (2022).
80. Kebede, S. D. et al. Prediction of contraceptive discontinuation among reproductive-age women in Ethiopia using Ethiopian demographic and Health Survey 2016 dataset: a Machine Learning Approach. *BMC Med. Inf. Decis. Mak.* **23** (1), 1–17 (2023).
81. Stockman, J., Friedman, J., Sundberg, J. & Harris, E. Predictive analytics using machine learning to identify ART clients at health system level at greatest risk of treatment interruption in Mozambique and Nigeria. *JAIDS J. Acquir. Immune Defic. Syndr.* **8**(6), 10-97 (2022).
82. Alonso, E. B. The impact of culture, religion and traditional knowledge on food and nutrition security in developing countries. (2015).

Acknowledgements
We would like to thank the measure DHS program for providing the dataset.

Author contributions
HSN, EBE, ADW, and ABZ were involved in the conception, design, data collection, supervision, investigation, data analysis, interpretation, and writing of the manuscript. BT, MDK, and GAT contributed to data extraction, preprocessing, proposal development, validation, manuscript revision, figure preparation, data analysis, visualization, and interpretation. All authors (HSN, EBE, ADW, BT, MDK, GAT, and ABZ) reviewed and approved the final manuscript.

Funding
The author(s) received no specific funding for this work.

Declarations

Competing interests
The authors have declared that no competing interests exist.

Ethics approval and consent to participate
We obtained approval from the DHS Program to access and utilize their data for our study.

Consent for publication
Not applicable.

Additional information
Correspondence and requests for materials should be addressed to H.S.N.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.