World Nutrition, September 2026;17(3): 108-123
DOI: https://doi.org/10.26596/wn.2026173108-123

WORLD NUTRITION
The Official Journal of the World Public Health Nutrition Association
Journal homepage: https://www.worldnutritionjournal.org
WPHNA
ISSN: 2041-9775

Research Article

Predicting early childhood stunting and associated factors in Kenya using machine learning

Bernard Thuo[1,*]

[1] Department of Computer Science and Digital Technologies, University of East London, London, United Kingdom

OPEN ACCESS

***Corresponding Author**
Thuo B.
thuobernarde@gmail.com
ORCID: 0009-0006-4279-1981

Article History
Received: April 23, 2026
Reviewed: August 06, 2026
Revised: September 09, 2026
Accepted: September 27, 2026
Published: September 30, 2026

Citation
Thuo B. (2026). Predicting early childhood stunting and associated factors in Kenya using machine learning. *World Nutrition*, 17(3), September, 108–123. https://doi.org/10.26596/wn.2026173108-123

Academic Editor
Ted Greiner, PhD

Copyright © 2026 Thuo B. Published by WPHNA. This is an open-access article distributed under the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Highlights/Key Messages

- Machine learning models can predict early childhood stunting using survey data
- Simpler models performed strongly, rivalling advanced machine learning algorithms
- Key stunting-associated factors that emerged matched known biological and social factors
- The choice of the model and threshold tuning influenced predictive performance

Background
Stunting in the first two years of life remains a major public health concern in Kenya, affecting about one in five children and leading to long-term consequences on cognitive development and economic productivity. Traditional, largely reactive approaches that rely on linear statistical models have a limited ability to capture the complex, non-linear determinants of stunting. Given that stunting arises from an interplay of immediate, underlying, and structural determinants, machine learning (ML) methods offer a promising alternative. ML techniques can leverage large, heterogeneous datasets to uncover hidden patterns and non-linear interactions among associated factors, potentially enabling earlier and more accurate identification of children with higher probabilities of stunting.

Objective
This study applied supervised machine learning techniques to predict early childhood stunting among children aged 0-23 months and identify associated factors using nationally representative data from the 2022 Kenya Demographic and Health Survey

Methods
Three models—logistic regression (LR), balanced random forest (BRF), and categorical boosting (CatBoost)—were developed and evaluated using metrics aligned with programmatic priorities, including recall, area under the receiver operating characteristic curve (AUC), F1-score, and Brier Score. To address class imbalance, balanced class weights were applied, and models were trained at a threshold of 0.4 to optimize sensitivity.

Results
CatBoost achieved the highest overall performance (AUC = 70.8%, recall = 85.8%), while LR remained highly competitive (AUC = 71.5%, Recall = 77.2%), outperforming BRF. Across all three models, consistent predictors included the child's age, sex, birth size, twin status, maternal stature, household wealth, and geographic location, reflecting the multifactorial nature of stunting and aligning with existing literature.

Conclusions
Simpler models, when combined with domain-informed feature engineering, could perform comparably to advanced algorithms while supporting transparency, which is essential for policy-relevant decision-making. The findings also provide methodological support for integrating predictive modelling into nutrition research and highlight opportunities to improve stunting prevention strategies.

Keywords: machine learning, stunting, early childhood, predictive algorithms, malnutrition, risk factors

Introduction

Stunting is an increasingly prevalent and persistent global public health concern. WHO defines stunting in children under five years as a chronic form of impaired linear growth characterized by a height-for-age Z-score (HAZ) more than two standard deviations below the WHO 2006 growth standards. When the HAZ is more than three standard deviations below the WHO median, this is classified as severe stunting (Tanaka et al. 2019). Globally, 150 million children under five years are stunted, with 100 million of them residing in South Asia and sub-Saharan Africa (UNICEF, 2025). Two out of every five (43%) stunted children live in Africa. In East Africa, at least 31.2% of children under five years are stunted (UNICEF, 2025). The period from a child's conception to the second birthday is widely recognized as a critical "window of opportunity" during which targeted interventions can effectively prevent or reverse stunting (Kinshela et al. 2020; Georgiadis & Penny, 2017).

In Kenya, about one in every five children is stunted. Although stunting decreased from 28.6% in 2012 to 18.4% in 2022, it remains unequally distributed across the country. Counties such as Kilifi, West Pokot, Samburu, and Meru continue to report high stunting prevalences – 37%, 34%, 31%, and 25%, respectively (KNBS, 2023). Prevalence also varies significantly by children's age, sex, household wealth status, and type of residence. Generally, stunting rises steadily within the complementary feeding window in children aged 6-23 months; each additional month of a young child's life tends to increase stunting odds by 1% (Okutse & Athiany, 2025). Boys are more likely to be stunted than girls (Tamir et al. 2024). One in every four children from the poorest households experiences stunting, compared with one in every eleven children in the wealthiest households. Rural communities face a persistently higher stunting burden than urban areas due to low socioeconomic development, poor sanitation, and limited access to health services (Masit et al. 2024).

The World Health Assembly goal of reducing stunting by 40% by 2025 was embedded in the Sustainable Development Goal (SDG) 2.2 as an intermediate milestone toward ending all forms of malnutrition by 2030. Currently, only 28% of countries, including Kenya, are on track to halve stunting (UNICEF, 2025). Kenya's annual average reduction rate of 4.3% exceeded the targeted 4% for the period 2012 –2025 (Aggarwal & Mishra, 2025).

Kenya's progress in reducing stunting reflects the integration of nutrition into National (Big 4 Agenda and Vision 2030) and County development plans, reinforced by policies such as the National Food and Nutrition Security Policy and Nutrition Action Plans. These frameworks institutionalized accountability, mainstreamed nutrition across governance structures, and facilitated the enforcement of food fortification standards. Complementary strategies, such as the 'Inua Jamii' social protection programs, private sector engagement, and improved infant feeding practices, have also enhanced household resilience and nutrition outcomes (Aggarwal & Mishra, 2025).

Building on these gains, Kenya's recent policy shifts—such as the Food and Nutrition Security and Resilience Programme and the Food Systems Transformation Pathway—signal a renewed commitment to integrating food, health, and resilience systems (Breisinger et al., 2023). However, persistent barriers, including uneven county-level implementation, financing gaps, and climate-induced shocks, threaten progress and widen inequities (Aggarwal & Mishra, 2025). Leveraging data and digital technologies offers a critical pathway to address these challenges by enabling real-time monitoring, predictive analytics, and targeted resource allocation, hence accelerating progress toward the 2030 nutrition targets.

Machine Learning (ML), a subset of Artificial Intelligence (AI), offers a unique analytical opportunity to understand these complex multidimensional factors contributing to stunting. ML uses statistical models and algorithms to analyse and learn patterns from a dataset and make predictions without necessarily using explicit programming instructions (Rajula et al. 2020). In nutrition research, ML has been utilized for analyzing complex datasets, uncovering hidden patterns, and improving predictive accuracy in areas such as nutrient profiling, malnutrition detection, treatment optimization, and surveillance (Colmenarejo, 2020; Yehuala et al., 2025; Atalell et al., 2023; Cai et al., 2024; Qasrawi et al., 2024). Unlike traditional statistical methods that rely on predefined assumptions and linear relationships, ML algorithms can model nonlinear interactions and automatically rank predictive factors, making them well-suited for large, diverse datasets influenced by multifactorial determinants (Anku & Duah, 2024; Islam et al., 2024; Shen et al., 2023; Ayele et al., 2025).

The Kenya Demographic and Health Survey (KDHS), part of the global DHS Program, has been conducted since 1988–1989, with the most recent round completed in 2022. KDHS is implemented approximately every five years and provides nationally representative data on health, nutrition, and socio-demographic characteristics, including detailed anthropometric measurements for children under five. Its standardized methodology, large sample size, and comprehensive coverage of household, maternal, and child-level variables offer a unique opportunity to apply predictive modelling techniques to uncover patterns and relationships influencing early childhood outcomes, such as stunting.

This study is anchored in the UNICEF Conceptual Framework on Maternal and Child Nutrition (Black et al. 2020), which classifies the determinants of malnutrition into three interlinked levels — immediate (child-level dietary and health factors), underlying (household, caregiving, and maternal factors), and basic (socioeconomic and structural factors). This framework has been previously applied to stunting research in Rwanda (Kalinda et al. 2023) and to machine learning in nutrition more broadly (Kirk et al. 2022).

Given the persistent burden of early childhood stunting in Kenya and the limitations of conventional approaches, this study aims to develop and evaluate three ML models, logistic regression (LR), balanced random forest (BRF), and categorical boosting (CatBoost), using KDHS 2022 data to predict stunting among children aged 0–23 months, identify the key factors associated with early childhood stunting, and compare the overall performance of the three predictive models. By identifying key associated factors and assessing model performance, the study seeks to inform targeted interventions and support evidence-based policy decisions.

Materials and Methods

Research Design

The study employed a cross-sectional analytical design to predict and identify key factors associated with early childhood stunting in Kenya (children 0-23 months of age). Rao et al.'s (2025) meta-analysis showed that cross-sectional ML models effectively predict malnutrition outcomes across diverse settings. This design integrated traditional epidemiological analysis with modern machine learning approaches to provide both explanatory and predictive insights. Stunting was defined as the proportion of children with HAZ below −2 at the time of the survey, rather than an incidence rate, because incidence would require longitudinal data to track the rate of new cases (Rothman et al. 2008). Similarly, consistent with the existing cross-sectional stunting literature (Danaei et al. 2016; Vaivada et al. 2020), predictor variables were described as associated factors, rather than risk factors, and model predictions as classification of current stunting status from concurrently measured characteristics, rather than prospective forecasting of future stunting (Rao et al. 2025).

Data Source

The study used the KDHS 2022 Kids Recode (KR) dataset. KDHS 2022 used a stratified two-stage cluster sampling design, selecting 1692 (666 urban and 1026 rural) clusters at the first stage and 25 households per cluster at the second stage. The KR dataset contained health and nutrition information on all children aged 0-59 months born to the interviewed women in the five years preceding the survey. This dataset also contained maternal characteristics data derived from the women's questionnaire and linked household data. The KR dataset contained 17,847 child records, of which 7,540 were aged 0-23 months. The final dataset used for the present study comprised 7383 children, as the rest were dropped due to missing outcome variable (stunting/HAZ) data.

Outcome Variable

The outcome variable was stunting status, defined as a binary classification (0=not stunted; 1=stunted), with stunting defined as a height-for-age z-score below -2 standard deviations of the WHO child growth standards.

Predictor Variables

Predictor variables included child-level, maternal, household, and geographical factors. Child-level variables included age, sex, birth weight, size at birth, feeding practices, recent illnesses, immunization, nutrient supplementation history, and anthropometric data. Maternal variables included demographic characteristics (including husband/partner's age, education, and occupation), age at first birth, maternal height, body mass index, parity, dietary habits, antenatal care (ANC), delivery, and postnatal care history. Household variables included wealth index, residence type, household size, drinking water source, cooking fuel, sanitation facility, mosquito net use, and time needed to access the nearest health services. Lastly, geographical factors were represented by the county of residence.

Understanding the Data

Exploratory data analysis (EDA) was undertaken to understand the underlying structure, distribution, and relationships within the KDHS 2022 dataset before modelling. EDA enables the identification of data quality issues such as missing values, outliers, and inconsistencies, while providing insights into variable distributions and potential multicollinearity (Ekbote et al. 2023).

Weighted descriptive statistics were computed using DHS sampling weights to ensure population-representative results, while unweighted descriptive statistics and prevalence estimates reflected the actual sample size analysed. Bivariate associations between the factors and stunting outcomes were assessed using unweighted Pearson chi-square tests (categorical factors) and two-sample t-tests (continuous factors, with Welch's correction applied where variances differed), performed in Stata (version 17). This also informed feature selection and preprocessing strategies. Machine learning models were later trained on unweighted data, as the focus was on predictive performance rather than population inference (see Figure 1 for the predictive modelling workflow).

Data Preprocessing and Feature Selection

To prepare the dataset for modelling, several steps were implemented. First, five variables were dropped, as they had more than 50% missing values (see Table 1). Secondly, predictor variables were classified as numerical or categorical using a heuristic based on data type and the number of unique values. The county variable was treated as categorical and mapped from numeric codes to county names before modelling to improve interpretability. Thirdly, two preprocessing pipelines were constructed: for continuous variables, a numeric transformer was applied to impute the missing values using the median, and the imputed values were then standardized using the z-score scaling approach – this ensured all numerical features were on a comparable scale. For categorical variables, missing values were imputed with the most frequent category, and one-hot encoding was applied to convert each category into a binary indicator variable, making the dataset suitable for LR and BRF models, which require numerical inputs.

Table 1. Variable names and proportions of each that were missing values

Variable	Description (per the KDHS 2022)	Missing (%)
c_ARI	A child had a cough and short, rapid breathing in the two weeks preceding the survey	94.93%
c_malaria	A positive malaria rapid test result – test carried out during the survey	80.16%
c_evervaccinated	Child received at least one dose of any vaccine	78.25%
c_ebf	Child received only breast milk in the 24 hours preceding the survey	74.63%
c_birthweightkg	Birth weight in kilograms to 1 decimal – either recalled or recorded	58.66%

Measurement limitations in the KDHS 2022 dataset resulted in the exclusion of the breastfeeding variables from the models. KDHS captures breastfeeding practices only through current-status (24-hour recall) indicators, consistent with 2021 WHO/UNICEF guidance, and hence lacks retrospective recall data (Croft et al., 2023). The current-status exclusive breastfeeding variable had missing data for almost 75% of the children in the sample. KDHS captures a breastfeeding duration (in months) variable, but for children who are still breastfeeding at the time of the survey, this is calculated as the interval between a child's date of birth and date of interview – assigning a child's age as completed duration (ICF, 2018; KNBS, 2023). Because most children in the present study's analytic sample (0-23 months) were still breastfeeding at the time of the interview, this variable would introduce censoring bias in the models and was therefore excluded as a continuous predictor.

Feature engineering entailed constructing composite indices by summing conceptually related variables to better capture multidimensional aspects and underlying patterns relevant to predicting stunting. Table 2 shows the variables included in each of the seven composite indices utilized in the study. These composite indices were constructed as formative measures, combining conceptually related but distinct indicators that represent different points of exposure within a broader determinant domain rather than repeated indicators of one underlying construct. Consistent with the UNICEF Conceptual Framework (Black et al. 2020), child-level indices (Feeding Practices, Child Health Services, Illness Burden, Age Birthweight Interaction) operationalized immediate determinants; the Maternal Care Index and household-level indices (WaSH, Wealth-Residence) operationalized underlying determinants; and county of residence operationalized basic/structural determinants.

Feeding practice indicators followed the WHO/UNICEF (2021) Indicators for Assessing Infant and Young Child Feeding Practices. Minimum Dietary Diversity (MDD) reflected whether a child aged 6–23 months had consumed foods from at least five of eight defined food groups during the previous day. Minimum Meal Frequency (MMF) reflected whether a child had been fed solid, semi-solid, or soft foods (or milk feeds, for non-breastfed children) at least the recommended number of times for their age and breastfeeding status. Minimum Acceptable Diet (MAD) combined both measures, requiring a child to meet MDD and MMF together, with an additional minimum-milk-feeds requirement for non-breastfed children. Unhealthy food consumption, assessed separately for children and mothers, reflected whether a sugary beverage, a processed or salty snack food, or, for children, an absence of any fruit or vegetable intake, had been consumed during the previous day.

Maternal dietary diversity, assessed using the Minimum Dietary Diversity for Women (MDD-W) indicator, was met when a mother consumed foods from at least five of ten defined food groups during the previous day (FAO, 2021).

Table 2. Manual feature-engineered composite indices included in the predictive models.

Composite Index	Included Features
Maternal Care Index	Number of antenatal care (ANC) visits; ANC services received; ANC offered by skilled personnel; health facility-based delivery; Birth delivery supported by a skilled attendant; postnatal care services received
Child Health Services Index	Vitamin A Supplementation in the last 6 months; Dewormed in the last 6 months; Iron Supplementation in the last 12 months; Multiple Micronutrient Powder Supplementation in the last 12 months
Illness Burden Index	Diarrhoea in the last 2 weeks; Fever in the last 2 weeks
Feeding Practices Index	Met Minimum Meal Frequency (MMF), Minimum Dietary Diversity (MDD), and Minimum Acceptable Diet (MAD)
Water Sanitation and Hygiene (WaSH) Index	Type of cooking fuel, Source of drinking water, Type of sanitation facility
Wealth-Residence Index	urban vs rural residence, and household wealth index
Age-Birthweight Interaction Index	Child's age in months, Reported size at Birth

ANC services received *= summary of seven recommended antenatal care content items reported as received: blood pressure measurement, urine sample collection, blood sample collection, foetal heartbeat monitoring, and counselling on diet, breastfeeding, and vaginal bleeding.*

PNC services received *= summary of nine care practices reported as received within the first two days after birth: examination of the umbilical cord, temperature measurement, counselling on danger signs requiring medical attention, breastfeeding counselling, breastfeeding observation, blood pressure measurement, discussion of vaginal bleeding, family planning counselling, and guidance on cord cleaning.*

Feature selection used a hybrid model that combined domain expertise, statistical methods, and multi-model ML approaches. Recursive feature elimination was conducted by assessing feature relevance using a combination of regression coefficients from the LR model, one-hot encoded, and mean permutation importances from the BRF model, predictive value change, SHAP (SHapley Additive exPlanation) values, and loss function variation from the CatBoost model. Features that consistently showed strong predictive power across these methods were retained, reducing the risk of overreliance on a single algorithm. The selected features were validated using domain expertise and existing nutritional epidemiological literature. A total of 25 features were validated and included in the ML modelling.

Train-Test-Split

As illustrated in Figure 1, the study used a stratified 70:30 train–test split approach – 70% (5168 children) of the data was used for training and 30% (2215 children) for testing. Stratification was applied to mitigate class imbalance, ensuring the proportion of stunted and non-stunted children was preserved in both subsets.

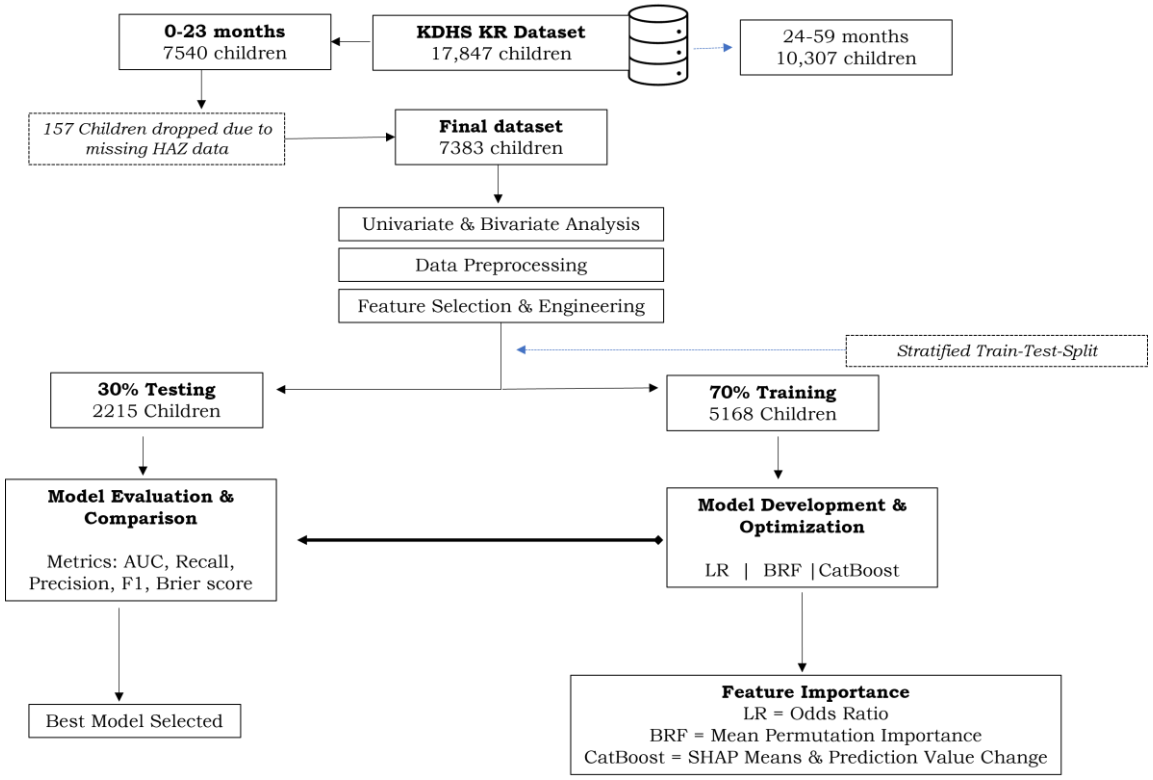


Figure 1. Predictive modelling workflow for early childhood stunting using KDHS 2022

Model Development, Training, Hyperparameter Tuning, and Evaluation

AUC-ROC (Area Under the Receiver Operating Characteristic Curve) assessed each model's ability to discriminate between stunted and non-stunted children. An AUC of 0.5 indicates a model lacks discriminative power, whereas a value closer to 1.0 indicates stronger discriminative power. To assess the quality of predicted probabilities, the Brier score quantified the mean squared difference between these predicted probabilities and the observed outcomes. Low Brier scores indicate better model calibration. Precision captured the number of false positives (non-stunted children misclassified as stunted) and answered, "*Of all children the model predicted to be stunted, how many were actually stunted?*" Recall (equivalent to sensitivity in the clinical and epidemiological literature) captured the number of true positives (stunted children correctly classified as stunted) and answered, "*Of all truly stunted children, how many did the model manage to identify as stunted?*" The F1 score harmonized the mean of precision and recall, with a higher value indicating better model performance in both recall and precision. Recall and precision are used throughout when reporting model output and table values, consistent with standard machine learning terminology, while sensitivity and specificity are used interchangeably in narrative discussion of the study's public health implications. Confusion matrices were used to tabulate the model predictions by summarizing the numbers of True Positives (TP), False Positives (FP), True Negatives (TN – non-stunted children correctly classified as non-stunted), and False Negatives (FN – stunted children incorrectly classified as non-stunted).

To predict early childhood stunting, three supervised ML classification algorithms were developed: LR, BRF, and CatBoost. These models were selected based on their applicability and proven effectiveness in similar studies (Anku & Duah, 2024; Islam et al. 2024; Shen et al. 2023; Ayele et al. 2025).

Logistic Regression is a commonly used algorithm for binary classification. It models the log-odds of the target class as a linear combination of input features and predicts the probability of a binary outcome using the sigmoid function (Birk et al. 2021). In this study, LR was implemented with balanced class weights to address imbalance between the stunted and non-stunted classes. Hyperparameter tuning was performed using a grid search strategy across penalty types and regularization strengths. To preserve the distribution of the dependent variables across folds, a stratified ten-fold cross-validation was also applied. The optimal configuration (C = 0.1, penalty = l2, solver = lbfgs) yielded the highest mean cross-validated area under the receiver operating characteristic curve (AUC) of 0.702 and was subsequently adopted for the final logistic regression model.

BRF is an ensemble ML algorithm that handles class imbalance by constructing each decision tree on a bootstrap sample that is balanced across classes, thus enhancing the detection of the minority classes. The BRF model combines predictions from multiple decision trees, resulting in strong predictive performance (Ayele et al. 2025). Hyperparameter

tuning was done using a randomized search strategy with 25 candidate set parameters, and a ten-fold stratified cross-validation schema. As with the LR model, the best parameter set (n_estimators = 500, max_depth = None, min_samples_split = 2, min_samples_leaf = 2, max_features = 'sqrt') achieved the highest mean cross-validated AUC (0.712) and was used in the final BRF model.

CatBoost is a gradient boosting algorithm that natively handles categorical data. To prevent overfitting and target-outcome leakage, the model uses ordered boosting and target-based encoding. CatBoost builds an additive model of decision trees trained sequentially, allowing it to handle complex interactions (Bitew et al. 2022). CatBoost's hyperparameter tuning was done using systematic grid search with stratified 3-fold cross-validation. The model used a lower fold count compared to the other two models, as it is computationally intensive, and has inbuilt internal regularization mechanisms such as early stopping and subsampling, which reduce variance and hence would still perform optimally with fewer folds. To handle class imbalances, balanced class weights were used. CatBoost's best parameters (iterations = 200, depth = 4, learning_rate = 0.05, l2_leaf_reg = 7, bagging_temperature = 0, border_count = 64, loss_function = Logloss, eval_metric = AUC, class_weights = [1, 5.75], random_seed = 42) had the highest cross-validated AUC of 0.718, and were used in the final predictive model.

All the models were trained at a specific threshold (0.4), rather than the conventional 0.5, to account for the class imbalances and to optimize model sensitivity(recall). Given the study's public health objective of minimising missed stunting cases, this lower threshold increased detection of true positives (improving sensitivity), though at the expense of reduced specificity due to more false-positive classifications (Kirk et al. 2022). Table 3 presents a threshold sensitivity analysis comparing precision, recall, and F1-score at thresholds of 0.3, 0.4, and 0.5.

Table 1. Threshold sensitivity analysis

Model	Threshold	Recall	Precision	F1-Score	True Negative	False Positive	False Negative	True Positive
LR	0.3	90.7%	21.2%	34.3%	525	1304	36	350
	0.4	77.2%	24.2%	36.9%	897	932	88	298
	0.5	64.0%	29.6%	40.4%	1240	589	139	247
BRF	0.3	95.3%	19.1%	31.9%	274	1555	18	368
	0.4	83.9%	21.9%	34.7%	672	1157	62	324
	0.5	65.5%	25.4%	36.6%	1087	742	133	253
CatBoost	0.3	95.1%	19.6%	32.4%	319	1510	19	367
	0.4	85.8%	23.1%	36.4%	725	1104	55	331
	0.5	70.2%	26.9%	39.0%	1094	735	115	271

At a threshold of 0.3, recall exceeded 90% across all models (LR: 90.7%; BRF: 95.3%; CatBoost: 95.1%), but the models substantially over-identified non-stunted children as stunted, incorrectly flagging 71.3% (LR), 85.0% (BRF), and 82.6% (CatBoost) of truly non-stunted children as false positives. In a public health setting, this threshold would impose considerable resource demands for follow-up assessment and diagnosis. At a threshold of 0.5, false positives fell drastically, but recall dropped to between 64.0% and 70.2% across models, suggesting that close to a third or more of truly stunted children would be missed entirely, undermining the models' core public health objective. At a threshold of 0.4, the models achieved a more sustainable balance: recall remained meaningfully higher than at 0.5 – CatBoost recall rose from 70.2% to 85.8% - without the extreme resource demand for follow-up screening and diagnosis confirmation observed at 0.3. AUC and Brier scores are threshold-invariant — they do not change with threshold calibration — and were therefore not included in Table 3.

All model development, training, tuning, testing, and evaluation used unweighted data since the present study focused on individual-level classification rather than population-level inference. As a result, feature importance rankings and predicted probabilities reflected the characteristics of the analytic sample rather than the DHS-weighted national population and may have slightly overrepresented patterns that were more prevalent in the oversampled strata due to DHS's two-stage cluster sampling design. All analyses were conducted in Python version 3.10.19 using libraries such as scikit-learn, seaborn, pandas, numpy, joblib, matplotlib, imblearn, catboost, shap, and statsmodel.api.

Feature Importance

Factors associated with stunting were ranked using feature-importance approaches aligned with each model. LR used odds ratios, which quantified the change in odds of stunting per unit increase in a predictor feature while holding other variables constant (Tadesse et al. 2025). BRF's permutation importance captured a feature's contribution by measuring the decrease in predictive power when its values were randomly permuted (Tadesse et al. 2025; Ayele et al. 2025). CatBoost utilized two complementary approaches: Prediction value change assessed the average change in model predictions for each feature across all trees, while Shapley Additive exPlanations (SHAP) provided an easier interpretation of how each feature increased or decreased the model's predicted probability of stunting (Begashaw et al. 2025; Shen et al. 2023). Given the study's cross-sectional design, feature importance rankings were predictive rather than causal and should not be interpreted as evidence of causal relationships.

Ethical Considerations

The study used anonymized, publicly available secondary data and did not involve any direct interaction with human subjects. As such, it was exempted from formal ethical review. The research adhered to responsible AI practices, ensuring fairness, transparency, and accountability in model development and evaluation.

Results

Characteristics of the Study Population

Child characteristics

The analytical sample (the unweighted column in Table 4) included 7383 children aged 0-23 months, with a mean age of 11.2 months (SD = 6.74). Regarding birth characteristics, the mean birth weight was 3.2kg (SD = 0.60), Unhealthy food consumption (such as sugary snacks or drinks) was reported for 52.0% of children.

The prevalence of diarrhea, fever, and acute respiratory infection (ARI) in the two weeks preceding the survey was 21.0%, 19.9%, and 43.3%, respectively. Immunization coverage was generally high, with 83.1% of children reported to have ever received at least one vaccine.

Anthropometric measurements showed that the mean weight and height of the children were 8.4kg (SD = 2.11) and 70.9cm (SD = 98.93), respectively. The prevalence of stunting was 17.4%, wasting was 6.2%, and underweight was 10.4%. Table 4 below summarizes the child characteristics. Note that the unweighted column in this and the following two tables (Table 5 & Table 6) is based on the sample used in the present study. The weighted values column from the original survey indicates how closely related these two samples are.

Table 4. Frequency and percentage distribution of child characteristics

Characteristic	Category	Unweighted N (%)[a]	Weighted %
Sex (N=7383)	Male	3792 (51.4%)	50.3%
	Female	3591 (48.6%)	49.7%
Age Group (N=7383)	0-5 months	1873 (25.4%)	25.1%
	6 – 23 months	5510 (74.6%)	74.9%
Birth Size (N=3641)	Very large	75 (2.1%)	2.0%
	Larger than average	488 (13.4%)	14.4%
	Average	2595 (71.3%)	70.9%
	Smaller than average	408 (11.2%)	11.1%
	Very small	75 (2.1%)	1.5%
Underweight (N=7380)	Normal *(-2 ≤ WAZ ≤ 2)*	6498 (88.1%)	89.8%
	Moderate *(-3 ≤ WAZ < -2)*	609 (8.3%)	6.8%
	Severe *(WAZ < -3)*	156 (2.1%)	1.7%
	Overweight *(WAZ > 2)*	117 (1.6%)	1.7%
Wasting (N=7358)	Normal *(-2 ≤ WHZ ≤ 2)*	6903 (93.8%)	95.6%
	Moderate *(-3 ≤ WHZ < -2)*	384 (5.2%)	3.7%
	Severe *(WHZ < -3)*	71 (1.0%)	0.7%
Stunting (N=7383)	Normal *(-2 ≤ HAZ ≤ 2)*	6098 (82.6%)	82.7%
	Moderate *(-3 ≤ HAZ < -2)*	958 (13.0%)	13.1%
	Severe *(HAZ < -3)*	327 (4.4%)	4.2%
	Early initiation of breastfeeding (N=3823)[b]	2468 (64.6%)	61.1%
	Exclusive breastfeeding (N=1873)[c]	870 (46.5%)	46.6%
	Minimum dietary diversity (N=5510)[d]	869 (15.8%)	18.5%
	Minimum meal frequency (N=5510)[d]	1782 (32.3%)	35.4%
	Minimum acceptable diet (N=5510)[d]	686 (12.5%)	15.2%
	Unhealthy food consumption (N=5510)	2867 (52.0%)	51.3%
	Diarrhoea for the last 2 weeks (N= 7381)	1546 (21.0%)	21.3%
	Fever for the last 2 weeks (N=7382)	1465 (19.9%)	20.2%
	Cough for the last 2 weeks (N=7383)	1820 (24.7%)	26.9%
	ARI for the last 2 weeks (N=374)	162 (43.3%)	37.3%
	Ever vaccinated (N =1606)	1334 (83.1%)	88.0%
	Fully immunized (N =3511)	1594 (45.4%)	52.0%
	Vitamin A supplementation in the last 6 months (N=7374)	4134 (56.1%)	60.3%
	Dewormed in the last 6 months (N =7371)	2094 (28.4%)	31.1%
	Iron supplementation in the last 12 months (N=7374)	1264 (17.1%)	16.6%
	Multiple Micronutrient Powder in the last 12 months (N=7369)	944 (12.8%)	11.4%

WAZ = weight-for-age z-score; WHZ = weight-for-height z-score; HAZ = height-for-age z-score; ARI = acute respiratory infection; Birth size = the child's size at birth based on the mother's recall

[a]The sample being used for analyses in the present study

[b]Data on timing of breastfeeding initiation were missing for 3560 children (48.2% of the analytical sample). Among children with data available, 2468 were put to breast immediately after birth and 1355 initiated breastfeeding after 1 hour.

[c]Only reported for infants <6 months of age. It is biased downward because older infants (> 6 months) who may have been exclusively breastfed were also excluded.

[d]Based on a 24-hour recall; only for children 6-23 months of age; defined in the methods section.

Maternal Characteristics

Mothers of the sampled children had a mean age of 27.8 years (SD = 6.49) and had spent, on average, 7.9 years in school (SD = 4.97). Most mothers were Christian (75.3%), married (83.2%), and predominantly from the Kikuyu, Luhya, and Kalenjin ethnic groups. The average maternal body mass index (BMI) for non-pregnant and non-postpartum women aged 20–49 years was 23.3 kg/m² (SD = 5.02), and the average age at 1st birth was 19.9 years (SD = 3.76). Mothers attended their 1st ANC visit around 4.5 months (SD = 1.64) on average. Iron supplementation for at least 90 days during pregnancy was reported by 89.6% of mothers, while only 40.0% met the Minimum Dietary Diversity (MDD), indicating low consumption of diverse food groups. The sampled mothers' partners had a mean age of 35.7 years (SD = 9.47). Table 5 below further outlines the overall maternal characteristics.

Table 5. Frequency and percentage distribution of maternal characteristics

Characteristic	Category	Unweighted N (%)[a]	Weighted %
Education level completed[b] (N=7383)	No education	1551 (21.0%)	9.4%
	Primary level	2449 (33.2%)	35.7%
	Secondary level	2351 (31.8%)	36.9%
	Higher level	1032 (14.0%)	18.0%
Religion (N=7383)[b]	No religion	106 (1.4%)	1.3%
	Christians	5556 (75.3%)	86.6%
	Muslims	1510 (20.5%)	9.2%
	Traditionalists	51 (0.7%)	0.4%
	Other religions	160 (2.2%)	2.5%
Marital status (N=7383)	Single	745 (10.1%)	11.3%
	Married	6139 (83.2%)	81.7%
	Widowed	86 (1.2%)	1.2%
	Divorced	70 (1.0%)	0.6%
	Separated	343 (4.7%)	5.3%
Employment (N=7376)	Not working	4050 (54.9%)	47.9%
	Working	3326 (45.1%)	52.1%
Earning more than husband/partner (N=2007)	More than husband	175 (8.7%)	7.9%
	Less than husband	1568 (78.1%)	79.6%
	About the same	264 (13.2%)	12.5%
Husband/Partner education Level (N=7383)[b]	No education	1263 (20.8%)	9.2%
	Primary level	1945 (32.0%)	34.4%
	Secondary level	1698 (27.9%)	32.2%
	Higher level	1173 (19.3%)	24.2%
Husband/Partner employment (N=6124)	Not working	1172 (19.1%)	11.5%
	Working	4952 (80.9%)	88.5%
Maternal body mass index (N=3843)	Underweight (<18.5kg/m2)	551 (14.3%)	9.7%
	Normal (18.5-24.9 kg/m2)	2120 (55.2%)	54.8%
	Overweight/Obese (≥25kg/m2)	1172 (30.5%)	35.5%
Number of antenatal care visits (N=7131)	Less than 4 visits	2732 (38.3%)	34.0%
	4 visits or more	4399 (61.7%)	66.1%
Integrated antenatal care services (N=7383)		4264 (57.8%)	61.7%
Facility-based delivery (N=7382)		5953 (80.6%)	88.2%
Caesarean section delivery (N=7383)		962 (13.0%)	16.1%
Birth delivery by a skilled attendant (N=7383)		6119 (82.9%)	89.4%
Kangaroo Mother Care (N=3691)		2369 (64.2%)	67.0%
Integrated postnatal care (N=7383)		1314 (17.8%)	19.4%
Deworming during pregnancy (N=3719)		1090 (29.3%)	28.3%
Minimum dietary diversity (N=7383)		2956 (40.0%)	46.3%
Unhealthy food consumption (N=7383)		5588 (75.7%)	77.0%
Cigarette/Tobacco use (N= 7383)		3626 (49.1%)	49.0%

[a]The sample being used for analyses in the present study

[b]Weighted and unweighted estimates diverge for religion and education because KDHS 2022 allocated clusters across counties to ensure comparable precision for county-level reporting, rather than in proportion to county population size (KNBS, 2023); this over-represents smaller counties —most of which have distinct religious composition and lower education levels — in the unweighted sample, an effect corrected by weighting.

Household Characteristics

Table 6 shows the overall household characteristics of our analytical sample. Households varied in size, with a mean household size of 5.9 members (SD = 2.59). On average, it took a household 41.8 minutes (SD = 49.22) to reach the nearest healthcare facility.

Table 6. Frequency and distribution of household factors

Characteristic	Category	Unweighted N (%)[a]	Weighted %
Residence type (N=7383)	Urban	2439 (33.0%)	35.1%
	Rural	4944 (67.0%)	64.%
Wealth index (N=7383)	Poorest	2406 (32.6%)	22.9%
	Poorer	1300 (17.6%)	18.6%
	Middle	1299 (17.6%)	18.0%
	Richer	1405 (19.0%)	20.6%
	Richest	973 (13.2%)	19.9%
Cooking fuel (N=7168)	Clean fuel	1114 (15.5%)	25.7%
	Polluting fuel[b]	6054 (84.5%)	74.3%
Source of drinking water (N=7181)	Improved water source	5207 (72.5%)	76.5%
	Unimproved water source	1974 (27.5%)	23.5%
Sanitation facility (N=7181)	Open defecation	1273 (17.7%)	9.8%
	Improved sanitation facility	2246 (31.2%)	35.4%
	Unimproved sanitation facility	3662 (51.0%)	54.8%
Female-headed households (N=7383)		2071 (28.1%)	27.3%
Mosquito nets used for children under 5 years (N=7376)		4862 (66.0%)	67.4%

[a]The sample being used for analyses in the present study

[b]polluting cooking fuel = household's primary cooking fuel is kerosene, charcoal, wood, straw/shrubs/grass, or agricultural crop residue; clean fuel comprises electricity, natural gas, liquefied petroleum gas (LPG), biogas or solar.

Early Childhood Stunting Prevalence

In Kenya, nearly two in every ten children (17.4%) were stunted, with 13.0% moderately stunted and 4.4% severely stunted. Kilifi (37.6%), Samburu (33.3%), and West Pokot (29.1%) had the highest stunting prevalence, highlighting persistent nutritional vulnerability in arid, semi-arid, and coastal communities. At the household level, children in rural areas exhibited higher stunting prevalence (19.8%) than those in urban households (12.6%; p<0.001). Stunting prevalence differed significantly by household wealth (p<0.001), ranging from 8.8% among the richest households to 23.7% among the poorest. Use of clean cooking fuel (10.4%), improved sanitation (13.9%), improved drinking water (16.0%), and mosquito net use among children (16.0%) were all significantly (p<0.001) associated with lower stunting prevalence.

Maternal characteristics also showed strong associations. Stunting prevalence differed significantly by maternal education (p<0.001), ranging from 10.8% among mothers with education beyond secondary to 20.5% among mothers with primary education. Stunting prevalence was higher among children of underweight mothers (24.5%) compared to those of normal weight or overweight mothers (p<0.001). Lower stunting prevalence was associated with iron supplementation during pregnancy (16.7%; p=0.038), at least four ANC visits (15.8%; p<0.001), facility-based delivery (16.2%; p<0.001), and skilled birth attendance (16.5%; p<0.001). Lower stunting prevalence was also associated with meeting the Minimum Dietary Diversity among mothers (14.6%; p<0.001) and non-use of tobacco (16.5%; p=0.030). Maternal occupation, unhealthy food consumption, and self-

reported health status were not significantly associated with stunting.

Stunting prevalence also differed significantly by birth size (p<0.001), ranging from 11.3% among infants larger than average to 30.7% among very small infants. Table 7 below summarizes the associations between stunting and various factors.

Table 7. Unweighted Pearson chi-square test associations between child, maternal, paternal, and household characteristics and early childhood stunting

Factor	Category	N	% Stunted (Unweighted)	Chi-square (df)	p-value
Child characteristics					
Twin	Singleton	7189	16.7%	102.41 (1)	0.0000 ***
	Twin	194	44.9%		
Sex	Male	3792	20.2%	41.19 (1)	0.0000 ***
	Female	3591	14.5%		
Age group	0-5 months	1873	10.8%	74.67 (1)	0.0000 ***
	6-23 months	5510	19.6%		
Birth size[a]	Very large	75	12.0%	71.60 (4)	0.0000 ***
	Larger than average	488	11.3%		
	Average	2595	16.1%		
	Smaller than average	408	29.9%		
	Very small	75	30.7%		
Early initiation of breastfeeding[b]	No	1355	17.8%	0.72 (1)	0.3976
	Within 1 hour after birth	2468	16.7%		
Breastfeeding (age < 6 months)[c]	Exclusive	870	10.0%	1.02(1)	0.3114
	Non-exclusive	1003	11.6%		
Immunization	Missed 1 or more doses	1917	25.8%	15.09 (1)	0.0000 ***
	Fully immunized	1594	20.2%		
Diarrhea (in the last 2 weeks)	No	5835	17.4%	0.03 (1)	0.8594
	Yes	1546	17.6%		
Fever (in the last 2 weeks)	No	5917	17.5%	0.33 (1)	0.5630
	Yes	1465	16.9%		
Cough (in the last 2 weeks)	No	5563	17.7%	1.18 (1)	0.2768
	Yes	1820	16.5%		
Acute respiratory infection (in the last 2 weeks)	No	212	20.3%	1.94 (1)	0.1637
	Yes	162	14.2%		
Minimum Dietary Diversity (6-23months)	Not adequate	4641	20.3%	8.95 (1)	0.0028 **
	Adequate	869	15.9%		
Minimum Meal Frequency (6-23months)	Not adequate	3728	20.5%	5.19 (1)	0.0227 *
	Adequate	1782	17.9%		
Minimum Acceptable Diet (6-23months)	Not adequate	4824	20.2%	6.18 (1)	0.0129 *
	Adequate	686	16.0%		
Unhealthy food consumption (6-23months)	No	2643	20.1%	0.51 (1)	0.4763
	Yes	2867	19.3%		
Maternal characteristics					
Education level	No education	1551	19.1%	53.72 (3)	0.0000 ***
	Primary level	2449	20.5%		
	Secondary level	2351	16.0%		
	Higher than secondary	1032	10.8%		
Occupation	Non-employed	4050	17.2%	0.27 (1)	0.6032
	Employed	3326	17.7%		
Self-reported health status	Very good	1829	18.1%	2.03 (4)	0.7309
	Good	4333	17.0%		
	Moderate	1102	17.4%		
	Bad	112	19.6%		
	Very bad	7	28.6%		
Body mass index	Underweight	551	24.5%	53.90 (2)	0.00000 ***
	Normal	2120	18.4%		
	Overweight/Obese	1172	11.1%		
Minimum Dietary Diversity (maternal)	5 or more food groups	4427	19.3%	26.38 (1)	0.0000 ***
	Less than 5 food groups	2956	14.6%		
Unhealthy foods consumption (maternal)	No	1795	17.0%	0.24 (1)	0.6206
	Yes	5588	17.5%		
	Non-user	3757	16.5%	4.72 (1)	0.0297 *

Factor	Category	N	% Stunted (Unweighted)	Chi-square (df)	p-value
Cigarette & Tobacco use	User	3626	18.4%		
Iron supplementation (pregnancy)	No	745	19.7%	4.29 (1)	0.0383 *
	Yes	6392	16.7%		
Number of antenatal care visits	Less than 4 visits	2732	18.9%	10.92 (1)	0.0000 ***
	4 or more visits	4399	15.8%		
Place of delivery	Home or other	1429	22.5%	32.14 (1)	0.000 ***
	Health facility	5953	16.2%		
Attendance at delivery	Unskilled attendant	1264	21.8%	19.72 (1)	0.000 ***
	Skilled attendant	6119	16.5%		
Paternal characteristics					
Father's education level	No education	1263	20.9%	69.45 (3)	0.000 ***
	Primary level	1945	20.0%		
	Secondary level	1698	15.0%		
	Higher than secondary	1173	10.2%		
Father's occupation	Non-employed	1172	19.5%	7.24 (1)	0.0071 **
	Employed	4952	16.2%		
Household factors					
Residence	Urban	2439	12.6%	59.31 (1)	0.000 ***
	Rural	4944	19.8%		
Wealth index	Poorest	2406	23.7%	151.39 (4)	0.000 ***
	Poorer	1300	19.7%		
	Middle	1299	15.7%		
	Richer	1405	12.0%		
	Richest	973	8.8%		
Cooking fuel	Clean fuel	1114	10.4%	44.79 (1)	0.000 ***
	Polluting fuel[d]	6054	18.8%		
Drinking water	Improved source	5207	16.0%	26.31 (1)	0.000 ***
	Unimproved source	1974	21.2%		
Sanitation facility	Open defecation	1273	23.1%	48.56 (2)	0.000 ***
	Improved facility	2246	13.9%		
	Unimproved facility	3662	17.6%		
Head of household	Male-headed	5312	17.1%	1.20 (1)	0.2729
	Female-headed	2071	18.2%		
Mosquito nets use for children under 5	No	2514	20.2%	20.12 (1)	0.000 ***
	Yes	4862	16.0%		

p < 0.05 * p < 0.01 ** p < 0.001 ***

[a]Birth size = the child's size at birth based on the mother's recall

[b]Data on timing of breastfeeding initiation were missing for 3560 children (48.2% of the analytical sample). Among children with data available, 2468 were put to breast immediately after birth and 1355 initiated breastfeeding after 1 hour.

[c]Only reported for infants <6 months of age. It is biased downward because older infants (> 6 months) who may have been exclusively breastfed were also excluded.

[d]polluting cooking fuel = household's primary cooking fuel is kerosene, charcoal, wood, straw/shrubs/grass, or agricultural crop residue; clean fuel comprises electricity, natural gas, liquefied petroleum gas (LPG), biogas or solar.

Comparative analysis revealed that children with a higher probability of stunting were, on average, older (mean difference: +2.78 months; p<0.001), lived in larger households (+0.34 members; p<0.001), and reported longer travel time to the nearest health facility (+8.5 minutes; p<0.001). Mothers of the stunted children had a lower age at first birth (-0.64 years; p<0.001) and higher parity (+0.27 births; p<0.001). Current maternal age and paternal age showed no significant differences.

Table 8. Comparison of child, maternal, paternal, and household factors by stunting

Variable	Non-stunted (N)	Stunted (N)	Non-stunted Mean (SE)	Stunted Mean (SE)	Mean Difference (SE)	95% CI for Difference	t-statistic (df)	p-value
Child age (months)	6098	1285	10.67 (0.09)	13.45 (0.19)	+2.78 (0.21) ***	[2.38, 3.18]	13.60 (7381)	0.000
Household size	6098	1285	5.86 (0.03)	6.20 (0.08)	+0.34 (0.08) ***	[0.18, 0.50]	4.27 (7381)	0.000
Time (in minutes) to the nearest health facility	3193	658	40.37 (0.83)	48.88 (2.25)	+8.51 (2.40) ***	[3.81, 13.22]	3.54 (3849)	0.000
Maternal age (years)	6098	1285	27.84 (0.08)	27.58 (0.18)	-0.26 (0.20)	[-0.65, 0.12]	-1.32 (7381)	0.1867
Maternal age at 1st birth (years)	6098	1285	20.05 (0.05)	19.40 (0.10)	-0.64 (0.11) ***	[-0.86, -0.42]	-5.58 (7381)	0.000
Paternal age (years)	5100	1039	35.59 (0.13)	36.14 (0.32)	+0.55 (0.34)	[-0.12, 1.22]	1.61 (6137)	0.1076
Parity (number of children born)	6098	1285	3.16 (0.03)	3.43 (0.06)	+0.27 (0.07) ***	[0.13, 0.40]	3.82 (7381)	0.000

Unweighted two-sample t-tests (Welch's correction applied where variances differed, per Levene's test), run on the same unweighted analytic sample used for model development. p < 0.05 * p < 0.01 ** p < 0.001 ***.

Machine learning models performance

The Logistic Regression model achieved moderate discrimination (AUC = 0.715) and demonstrated the best calibration, with the lowest Brier score (0.210). At the selected threshold (0.4), the model correctly identified 77.2% of stunted children, with a precision of 24.2% and an F1-score of 0.369. These results indicated stable performance and reliable probability estimates, supporting LR as a strong baseline model.

The BRF model achieved a recall of 83.9%, indicating improved sensitivity to stunted cases. However, BRF had the lowest AUC (0.680) among the models and a higher Brier score (0.236). The confusion matrix revealed a high number of false positives, indicating that the model overestimated the stunting status (See Table 9 on the confusion matrices for the evaluated models).

Table 9. Confusion matrices for the evaluated models, presenting the number of cases in each category (threshold = 0.4)

Model	Non-stunted correctly classified (TN)	Non-stunted misclassified as stunted (FP)	Stunted misclassified as non-stunted (FN)	Stunted correctly classified (TP)
Logistic Regression	897 (49.0%)	932 (51.0%)	88 (22.8%)	298 (77.2%)
Balanced Random Forest	672 (36.7%)	1157 (63.3%)	62 (16.1%)	324 (83.9%)
CatBoost	725 (39.6%)	1104 (60.4%)	55 (14.3%)	331 (85.8%)

TN = True Negatives, FP = False Positives, FN = False Negatives, TP = True Positives

The CatBoost model demonstrated a favourable balance between discrimination and sensitivity. It achieved an AUC of 0.708, comparable to Logistic Regression, and the highest recall (85.8%), indicating superior detection of stunted children. Precision (23.1%) and F1-score (0.364) were similar to LR, although calibration was weaker. These results reflected CatBoost's ability to capture complex, non-linear relationships among predictors while maintaining a strong minority-class detection. Table 10 below summarizes the comparative performance of the three models.

Table 10. The performance metrics comparison for the three predictive models

Model	AUC	Recall	Precision	F1-score	Brier Score
Logistic Regression	71.5%	77.2%	24.2%	36.9%	21.0%
Balanced Random Forest	68.0%	83.9%	21.9%	34.7%	23.6%
CatBoost	70.8%	85.8%	23.1%	36.4%	22.9%

Figure 2 below compares the models' performance. Although all three models show moderate discriminatory ability (AUC = 0.68 - 0.71), LR and CatBoost perform comparably and outperform the BRF across most false positive rates. The BRF's lower ROC curve reflects its limited ability to distinguish stunted from non-stunted children, despite improved sensitivity.

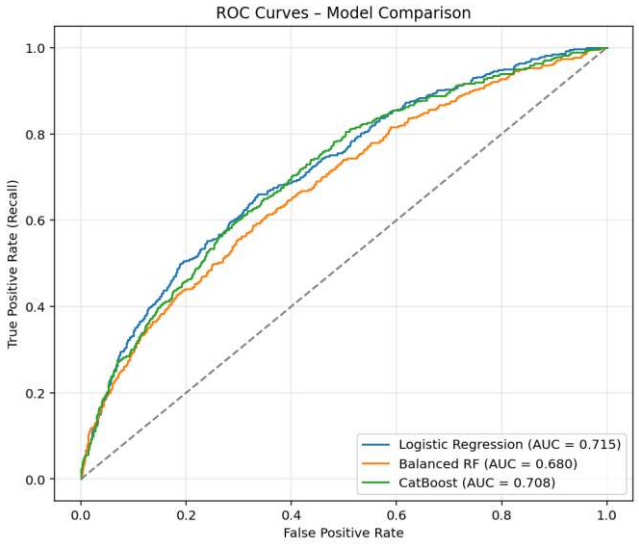


Figure 2. Receiver Operating Characteristic curves comparing the performance of Logistic Regression, Balanced Random Forest, and Categorical Boosting models in predicting stunted children

The difference between the two best-performing models was more evident in minority-class detection. CatBoost achieved the highest recall, correctly identifying over 85% of stunted children, compared to 77% for LR (See Table 10). In contrast, LR offered better calibration and interpretability. CatBoost was selected as the final model because of its superiority in sensitivity. This aligned with the study's public health objective of minimizing missed cases of early childhood stunting. LR can be recommended as a benchmark model for comparative evaluation.

Feature importance analysis

To identify the key factors associated with early childhood stunting in Kenya, feature importance was analysed across the three models, revealing both distinct and overlapping sets of influential variables.

The LR model provided interpretable odds ratios. Older infants *(OR = 1.67, 95% C.I [1.52, 1.83], p<.001)*, and children born as twins *(OR = 4.03, 95% C.I [2.56, 6.35], p<.001),* male sex *(OR = 0.63, 95% C.I [0.53, 0.73], p<.001)* and those born smaller than average *(OR = 2.74, 95% C.I [1.04, 7.21], p<.041)* or very small *(OR = 3.34, 95% C.I[1.07, 10.46], p=.038)* were associated with higher odds of stunting. Higher household wealth *(OR = 0.75, 95% C.I [0.70, 0.80], p<.001)* was associated with reduced odds of stunting, while residence in Kilifi County *(OR = 3.45, 95% C.I [1.26, 4.77], p=.008)* was associated with elevated stunting prevalence.

For the BRF model, permutation importance based on mean decrease in AUC was computed, and child's age,

maternal height, and twin status emerged as key factors associated with stunting. Other features ranked highly included child's age*size at birth interaction and size at birth (independently), as illustrated in Figure 3.

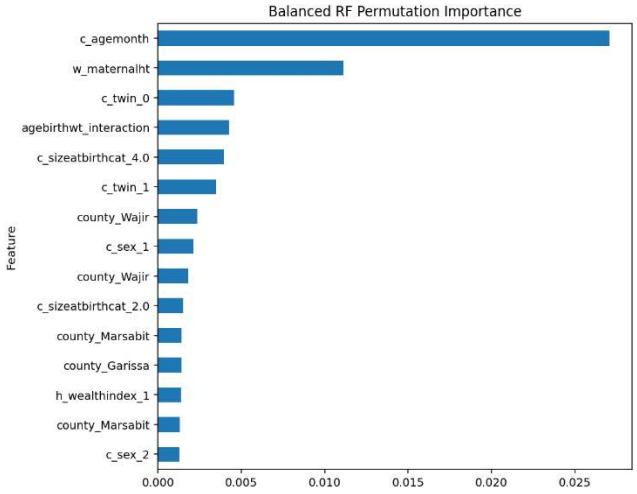


Figure 3. Balanced Random Forest feature importance ranking by mean Area Under Receiver Operating Characteristics permutations

Note: c_sex:1 = male, 2 = female; c_twin:0 = single birth, 1= twin; c_sizeatbirthcat:1= very large, 2= larger than average, 3= average, 4 = smaller than average, 5 = very small; h_wealthindex: 1 = poorest, 2 = poorer, 3 = middle, 4 = richer, 5 = richest.

c=child, w=woman,h=household, cat=category

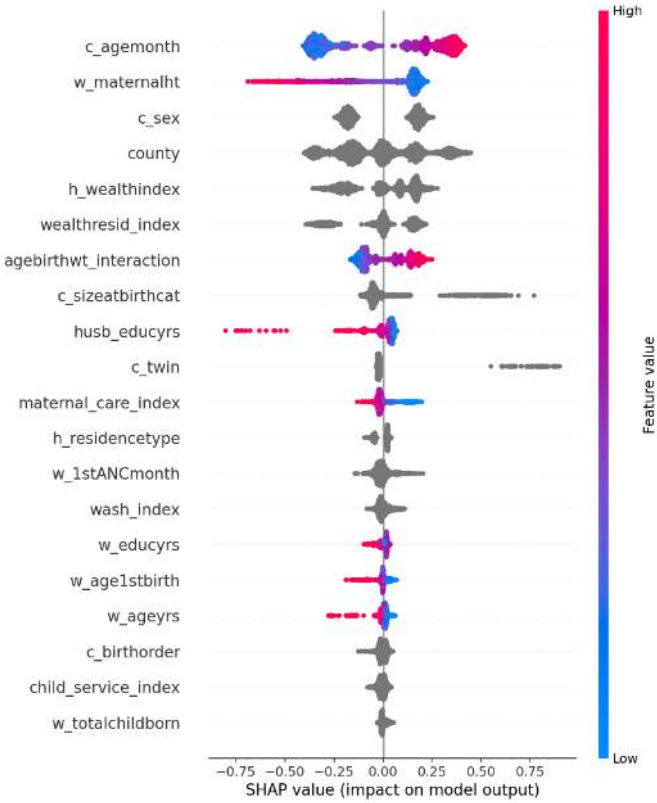


Figure 3. Shapley Additive exPlanations summary plot for global feature importance ranking in descending order of importance

CatBoost's Prediction Values Change ranked child's age, county, maternal height, child's age*size at birth interaction, household wealth, twin status, and child's sex as the most influential factors. In Figure 4, SHAP analysis confirmed these findings. Older child age and small birth size increased the predicted probability of stunting, while maternal height was generally protective, with taller mothers associated with lower predicted probabilities of stunting. County-level SHAP dispersion also pointed to geographical variations in stunting prevalence.

Overall, across all models, the most consistent factors associated with stunting included twin status, child age, birth size, age*birth size, maternal height, household wealth, child sex, and county of residence.

Discussion

This study aimed to develop machine learning models to predict early childhood stunting among Kenyan children aged 0–23 months, identify key associated factors, and compare model performance. Predictive performance largely depended on the data structure, feature selection, and evaluation metrics.

Among the three models tested, Categorical Boosting (CatBoost) slightly outperformed Logistic Regression (LR), while Balanced Random Forest (BRF) lagged. CatBoost's superior performance is likely due to its boosted algorithm's sequential error correction, native handling of categorical variables, and ability to capture non-linear relationships often missed by generalized linear models (Rao et al. 2025; Haile & Legisso, 2024; Shen et al. 2023). Interestingly, LR surpassed BRF in both discrimination and calibration, challenging assumptions that complex ensemble methods always outperform simpler models. Evidence suggests that for public health datasets characterized by monotonic additive relationships, optimized LR models remain competitive (Birk et al. 2021; Levy & O'Malley, 2020). In the present study, LR's strong performance was aided by manual feature engineering that incorporated variable interactions, informed by existing literature and domain expertise, enabling it to approximate non-linear effects. Conversely, BRF's weaker performance may reflect class imbalance, which introduces noise and diminishes signal – a limitation echoed in prior research on ensemble approaches (Rao et al. 2025; Theodore Armand et al. 2024).

This study prioritized recall, AUC, F1 score, calibration, and confusion matrix over accuracy, aligning with its goal of correctly identifying stunted children (avoiding false negatives). As Kirk et al. (2022) emphasize, evaluation metrics should be chosen based on the modelling objective rather than arbitrarily. Accuracy is particularly unreliable in imbalanced datasets, where the minority class (in the present study, stunted children) is often the primary focus, because it often masks poor detection of that class (Khan & Khan, 2024). The study's model evaluations highlighted clear trade-offs between sensitivity (recall) and specificity (false positive rates). Whereas this study used a fixed threshold (0.4) in favour of higher sensitivity, threshold selection should be tailored to the intended implementation context –

an optimal threshold balances available resources, programmatic priorities, and relative costs of missing stunted children compared to incorrectly identifying healthy children as stunted (Kirk et al. 2022).

The study identified key factors associated with early childhood stunting, with importance rankings varying slightly across models. This was an expected outcome, given differences in how models capture interactions. Across all three models, factors consistently identified included the child's age, sex, birth size, twin status, maternal stature, household wealth, and geographic location. Stunting prevalence was higher among children with small birth size, boys, older children, and twins. Consistent with the previous research, stunting is a cumulative condition that increases with age and peaks between 6–23 months, with its initiation coinciding with the introduction of complementary feeding; hence the need to emphasize diverse and adequate diets (Tamir et al. 2024; Mengesha et al. 2021). The higher stunting prevalence observed among boys in this study aligns with evidence proposing that boys' higher metabolic demands may make them biologically more vulnerable to nutrient deficits (Yani et al. 2023; Afolabi & Palamuleni, 2021). The association between small birth size and stunting agrees with prior evidence linking impaired intrauterine growth to maternal undernutrition and inadequate antenatal care, with growth deficits reported to persist after birth (Danaei et al. 2016). Similarly, the elevated association observed among twins is consistent with reported patterns of in-utero resource competition, low birth weight, and poor immunity—factors that have been proposed to increase susceptibility to illness and perpetuate the vicious cycle between infections and malnutrition (Tamir et al. 2024; Elema et al. 2025; Danaei et al. 2016).

Short maternal stature was strongly associated with stunting, consistent with evidence that maternal height may limit foetal and postnatal growth potential in children (Rahman & Akter, 2025; Htet et al. 2023). Similarly, low household wealth may serve as a proxy for broader deprivation and has been associated in prior literature with poverty, food security, healthcare access, sanitation, and caregiving practices – factors proposed to influence stunting outcomes (Vaivada et al. 2020; Okutse & Athiany, 2025). Geographic disparities in stunting - with higher prevalences in Kilifi – highlight the need for context-specific interventions (Masit et al. 2024). The current study also identified moderately predictive factors such as husband's age and education, maternal age and education, age at first birth, and timing of first ANC visit. These factors align with the intergenerational, behavioral, and health system pathways highlighted in the literature, through which maternal resources and health-seeking practices may relate to child growth outcomes (Rahman & Akter, 2025; Habimana et al. 2023; Danaei et al. 2016).

The study findings have practical implications for nutrition programming and child health surveillance in Kenya. First, several of the key factors identified by the models, such as child age, sex, and birth weight, antenatal care variables, growth monitoring indicators, and geographical location, are already captured through routine health information systems such as Kenya Health Information System (KHIS) and the Kenya Electronic Community Health Information System (eCHIS). This presents an opportunity to leverage the predictive models to support early identification of children with higher stunting probabilities. LR models may produce more false-positive classifications (healthy children misclassified as stunted) but offer a transparent, easily interpretable scoring approach with well-calibrated probability estimates, which may matter more. Conversely, CatBoost could identify a higher proportion of truly stunted children, which may be preferable where minimizing missed cases is the priority, provided model explanation tools are available to support interpretation. Second, other factors associated with stunting that the models identified, such as the household wealth index, may impose additional data collection burdens and reduce feasibility. Furthermore, the geographical variations in stunting observed in this study suggest that the models are likely to require county-specific validation and recalibration before deployment. Lastly, any tool built from these models should only be used as decision support aids, complementing, but not replacing, human clinical judgment and programmatic expertise.

Limitations of the Study

The study faced several limitations that should be considered when interpreting the findings. First, the analysis was restricted to variables available in the KDHS 2022 dataset. Although childhood stunting is influenced by a broad range of factors, relevant determinants not captured in the dataset - such as household food security shocks and intra-household resource allocation - were not included in the models and this may have introduced residual confounding into the identified associations. Second, all models were internally validated using stratified train-test splitting and k-fold cross-validation within the KDHS 2022 sample. No external validation was conducted using an independent dataset or a different KDHS round, so the generalisability of model performance to other time periods, populations, or settings remains untested. Third, since models were trained on unweighted data to prioritise individual-level classification, predicted probabilities and feature importance rankings reflect the composition of the analytic sample rather than the DHS-weighted national population. Given the two-stage cluster design, this may modestly bias rankings toward patterns more prevalent in oversampled clusters. Fourth, reliance on secondary cross-sectional survey data might have introduced potential sources of bias, such as measurement errors in anthropometric indicators and recall bias in self-reported variables. Lastly, the study's cross-sectional design limited the ability to infer causal relationships between identified key factors and stunting outcomes, as predictors and outcomes were measured at a single point in time; the models therefore estimated associations rather than causal effects. Future studies using longitudinal data would be needed to establish temporal sequencing. Moreover, the study could not investigate how

stunting risk changes over time, and hence, long-term patterns and interactions contributing to chronic undernutrition were not fully captured.

Conclusions

This study demonstrates the potential of machine learning to predict early childhood stunting and identify associated factors using nationally representative survey data. By comparing Logistic Regression (LR), Balanced Random Forest (BRF), and Categorical Boosting (CatBoost) models, the analysis showed that predictive performance depends not only on algorithm complexity but also on data structure, threshold calibration, and the rigor of evaluation metrics. While CatBoost achieved the highest overall performance, optimized LR remained highly competitive, outperforming BRF and challenging the assumption that more complex ensemble methods always outperform simpler models. Consistent predictors—including child age, sex, birth size, twin status, maternal stature, household wealth, and geographic location—aligned with established evidence, highlighting the multifactorial nature of stunting. Overall, the findings suggest that strong predictive performance can still be achieved with simpler models. Leveraging domain expertise in feature selection, model design, and evaluation enables simpler models like LR to perform comparably to more complex machine learning approaches, supporting their use in policy and programmatic decision-making.

Future studies should continue to prioritize evaluation metrics that align with the study's objectives rather than relying on arbitrary choices. In nutrition-related data modelling, class weighting is preferable for handling class imbalance, as it preserves the original data structure and reflects real-world conditions. Stunting is a chronic condition; hence, relying only on cross-sectional surveys limits the ability to capture long-term determinants. Lastly, future work should integrate richer data sources like agricultural production, climate patterns, and market access to capture the broader drivers of malnutrition.

Author Contributions

BT: Conceptualisation, Methodology, Software, Formal Analysis, Data Curation, Visualisation, Writing - Original Draft, Writing - Review & Editing. The authors have read and approved the final version of the paper and its submission and have consented to publication.

Declaration of Generative AI and AI-Assisted Technologies in Scientific Writing

During the preparation of this manuscript, the author used Microsoft Copilot to assist with language editing and grammar correction. All AI-generated suggestions were reviewed and revised as appropriate by the author. The author assumes full responsibility for the content of this publication.

Acknowledgements

None

Data Availability Statement

The KDHS data is publicly available and can be accessed from the Kenya National Data Archive (KeNADA).

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors

Conflict of Interest

The author declares that there is no conflict of interest regarding the publication of this paper.

Publisher's Note: *All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher. The publisher remains neutral with regard to jurisdictional claims.*

REFERENCES

Afolabi, R. F., & Palamuleni, M. E. (2021). Exploring determinants of under-5 stunting in Malawi using a generalised linear mixed model. *South African Journal of Child Health*, 15(1), 18-24. https://hdl.handle.net/10520/ejc-m_sajch_v15_n1_a5

Aggarwal, A., & Mishra, N. (2025). Progress on Selected World Health Assembly Nutrition Targets in 12 countries. *Global Alliance for Improved Nutrition*. https://doi.org/10.36072/wp.52

Anku, E. K., & Duah, H. O. (2024). Predicting and identifying factors associated with undernutrition among children under five years in Ghana using machine learning algorithms. *Plos One*, 19(2), e0296625. https://doi.org/10.1371/journal.pone.0296625

Atalell, K. A., Techane, M. A., Terefe, B., & Tamir, T. T. (2023). Mapping stunted children in Ethiopia using two decades of data between 2000 and 2019. A geospatial analysis through the Bayesian approach. *Journal of Health, Population and Nutrition*, 42(1), 113. https://doi.org/10.1186/s41043-023-00412-3

Ayele, M. K., Baye, G. A., Yesuf, S. H., Engda, A. A., & Mitiku, E. T. (2025). Predicting stunting status among under-five children in Ethiopia using ensemble machine learning algorithms. *Scientific Reports*, 15(1), 27907. https://doi.org/10.1038/s41598-025-03206-1

Begashaw, G. B., Zewotir, T., & Fenta, H. M. (2025). A deep learning approach for classifying and predicting children's nutritional status in Ethiopia using LSTM-FC neural networks. *BioData Mining*, 18(1), 11. https://doi.org/10.1186/s13040-025-00425-0

Birk, N., Matsuzaki, M., Fung, T. T., Li, Y., Batis, C., Stampfer, M. J., ... & Lake, E. (2021). Exploration of machine learning and statistical techniques in development of a low-cost screening method featuring the global diet quality score for detecting prediabetes in rural India. *The Journal of Nutrition*, 151, 110S-118S. https://doi.org/10.1093/jn/nxab281

Black, M. M., Lutter, C. K., & Trude, A. C. (2020). All children surviving and thriving: re-envisioning UNICEF's conceptual framework of malnutrition. *The Lancet Global Health*, 8(6), 766-767. https://doi.org/10.1016/S2214-109X(20)30122-4

Breisinger, C., Keenan, M., Mbuthia, J., & Njuki, J. (2023). *Food systems transformation in Kenya: Lessons from the past and policy options for the future*. International Food Policy Research Institute. https://doi.org/10.2499/9780896294561

Cai, C., Li, H., Zhang, L., Li, J., Duan, S., Fang, Z., ... & Zeng, Z. (2024). Machine learning identification of nutrient intake variations across age groups in metabolic syndrome and healthy populations. *Nutrients*, 16(11), 1659. https://doi.org/10.3390/nu16111659

Colmenarejo, G. (2020). Machine learning models to predict childhood and adolescent obesity: a review. *Nutrients*, 12(8), 2466. https://doi.org/10.3390/nu12082466

Croft, T.N., Allen, C.K., & Zachary, B. W. (2023). Guide to DHS statistics: DHS-8. ICF. Retrieved from https://dhsprogram.com/pubs/pdf/DHSG1/Guide_to_DHS_Statistics_DHS-8.pdf

Danaei, G., Andrews, K. G., Sudfeld, C. R., Fink, G., McCoy, D. C., Peet, E., ... & Fawzi, W. W. (2016). Risk factors for childhood stunting in 137 developing countries: a comparative risk assessment analysis at global, regional, and country levels. *PLoS Medicine,* 13(11), e1002164. https://doi.org/10.1371/journal.pmed.1002164

Elema, T. B., Nega, F., Ali, J. M., Edae, G., & Boneya, D. J. (2025). Magnitude and determinants of stunting among children under five years of age in Ethiopia: a systematic review and meta-analysis. *Frontiers in Pediatrics*, 13, 1499921. https://doi.org/10.3389/fped.2025.1499921

Ekbote, N., Dhanshetti, P., & Sakhrekar, S. (2023). Techniques of Exploratory Data Analysis. *Madhya Pradesh Journal of Social Sciences*, 28(2).

FAO. (2021). Minimum dietary diversity for women. https://doi.org/10.4060/cb3434en

Georgiadis, A., & Penny, M. E. (2017). Child undernutrition: opportunities beyond the first 1000 days. *The Lancet Public Health*, *2*(9), e399. https://doi.org/10.1016/s2468-2667(17)30154-8

Habimana, J. D. D., Uwase, A., Korukire, N., Jewett, S., Umugwaneza, M., Rugema, L., & Munyanshongore, C. (2023). Prevalence and correlates of stunting among children aged 6–23 months from poor households in Rwanda. *International Journal of Environmental Research and Public Health*, 20(5), 4068. https://doi.org/10.3390/ijerph20054068

Haile, Y. T., & Legisso, T. Z. (2024). Trends and determinants of child stunting in Ethiopia: a longitudinal analysis of sociodemographic factors and regional disparities. *BMC Public Health*, 24(1), 3391. https://doi.org/10.1186/s12889-024-20962-3

Htet, M. K., Do, T. T., Wah, T., Zin, T., Hmone, M. P., Raihana, S., ... & Dibley, M. J. (2023). Socio-economic and agricultural factors associated with stunting of under 5-year children: findings from surveys in mountains, dry zone and delta regions of rural Myanmar (2016–2017). *Public Health Nutrition*, 26(8), 1644-1657. https://doi.org/10.1017/S1368980023001076

ICF.(2018). *Demographic and health surveys standard recode manual for DHS-7*. The Demographic and Health SurveysProgram. Retrieved from https://dhsprogram.com/pubs/pdf/DHSG4/Recode7_DHS_10Sep2018_DHSG4.pdf

Islam, M. M., Kibria, N. M. S. J., Kumar, S., Roy, D. C., & Karim, M. R. (2024). Prediction of undernutrition and identification of its influencing predictors among under-five children in Bangladesh using explainable machine learning algorithms. *PloS One*, 19(12), e0315393. https://doi.org/10.1371/journal.pone.0315393

Kalinda, C., Phri, M., Qambayot, M. A., Ishimwe, M. C. S., Gebremariam, A., Bekele, A., & Wong, R. (2023). Socio-demographic and environmental determinants of under-5 stunting in Rwanda: Evidence from a multisectoral study. *Frontiers in Public Health*, 11, 1107300. https://doi.org/10.3389/fpubh.2023.1107300

Khan, N., & Khan, S. (2024). Comparing Models and Performance Metrics for Lung Cancer Prediction using Machine Learning Approaches. *Sir Syed University Research Journal of Engineering & Technology,* 14(2), 29–33. https://doi.org/10.33317/ssurj.629

Kirk, D., Kok, E., Tufano, M., Tekinerdogan, B., Feskens, E. J., & Camps, G. (2022). Machine learning in nutrition research. *Advances in Nutrition*, 13(6), 2573-2589. https://doi.org/10.1093/advances/nmac103

KNBS, Ministry of Health, ICF. (2023). Kenya Demographic and Health Survey 2022. Demographic and Health Survey, Kenya National Bureau of Statistics, Nairobi. Retrieved from https://dhsprogram.com/pubs/pdf/FR380/FR380.pdf

Levy, J. J., & O'Malley, A. J. (2020). Don't dismiss logistic regression: the case for sensible extraction of interactions in the era of machine learning. *BMC Medical Research Methodology*, 20(1), 171. https://doi.org/10.1186/s12874-020-01046-3

Masit, J., Malenje, B. M., & Imboga, H. (2024). Spatial patterns and risk factors of stunting among under-five children in Kenya: a multilevel and spatial analysis. *International Journal of Data Science and Analysis*, 10(3), 49-60. https://doi.org/10.11648/j.ijdsa.20241003.12

Mengesha, A., Hailu, S., Birhane, M., & Belay, M. M. (2021). The prevalence of stunting and associated factors among children under five years of age in southern Ethiopia: community based cross-sectional study. *Annals of Global Health*, 87(1), 111. https://doi.org/10.5334/aogh.3432

Okutse, A. O., & Athiany, H. (2025). Socioeconomic disparities in child malnutrition: trends, determinants, and policy implications from the Kenya demographic and health survey (2014-2022). *BMC Public Health*, 25(1), 295. https://doi.org/10.1186/s12889-024-21037-z

Qasrawi, R., Sgahir, S., Nemer, M., Halaikah, M., Badrasawi, M., Amro, M., ... & Al-Jawaldeh, A. (2024). Machine learning approach for predicting the impact of food insecurity on nutrient consumption and malnutrition in children aged 6 months to 5 years. *Children*, 11(7), 810. https://doi.org/10.3390/children11070810

Rahman, M. T., & Akter, S. (2025). Changing Determinants of Child Wasting: Insights from the Prevalence of Stunting, Wasting, and Malnutrition and Their Determinants—An Analysis of BDHS 2022 Data. https://doi.org/10.21203/rs.3.rs-6737252/v1

Rajula, H. S. R., Verlato, G., Manchia, M., Antonucci, N., & Fanos, V. (2020). Comparison of conventional statistical methods with machine learning in medicine: diagnosis, drug development, and treatment. *Medicina*, 56(9), 455. https://doi.org/10.3390/medicina56090455

Rao, B., Rashid, M., Hasan, M. G., & Thunga, G. (2025). Machine Learning in Predicting Child Malnutrition: A Meta-Analysis of Demographic and Health Surveys Data. *International Journal of Environmental Research and Public Health*, 22(3), 449. https://doi.org/10.3390/ijerph22030449

Rothman, K. J., Greenland, S., & Lash, T. L. (2008). *Modern epidemiology* (3rd ed.). Lippincott Williams & Wilkins.

Tadesse, G. A., Ferguson, L., Robinson, C., Kuria, S., Wanyonyi, H., Murage, S., ... & Dilkina, B. (2025). Forecasting acute childhood malnutrition in Kenya using machine learning and diverse sets of indicators. *PLoS One*, 20(5), e0322959. https://doi.org/10.1371/journal.pone.0322959

Tamir, T. T., Gezhegn, S. A., Dagnew, D. T., Mekonenne, A. T., Aweke, G. T., & Lakew, A. M. (2024). Prevalence of childhood stunting and determinants in low and lower-middle income African countries: Evidence from standard demographic and health survey. *PloS One*, 19(4), e0302212. https://doi.org/10.1371/journal.pone.0302212

Tanaka, J., Yoshizawa, K., Hirayama, K., Karama, M., Wanjihia, V., Changoma, M. S., & Kaneko, S. (2019). Relationship between dietary patterns and stunting in preschool children: a cohort analysis from Kwale, Kenya. *Public Health*, 173, 58-68. https://doi.org/10.1016/j.puhe.2019.05.013

Theodore Armand, T. P., Nfor, K. A., Kim, J. I., & Kim, H. C. (2024). Applications of artificial intelligence, machine learning, and deep learning in nutrition: A systematic review. *Nutrients*, 16(7), 1073. https://doi.org/10.3390/nu16071073

United Nations Children's Fund (2025). *Levels and trends in child malnutrition child malnutrition: UNICEF/WHO/World Bank Group Joint Child Malnutrition Estimates: Key findings of the 2025 edition*. United Nations Children's Fund. https://www.who.int/publications/i/item/9789240112308

Vaivada, T., Akseer, N., Akseer, S., Somaskandan, A., Stefopulos, M., & Bhutta, Z. A. (2020). Stunting in childhood: an overview of global burden, trends, determinants, and drivers of decline. *The American journal of clinical nutrition*, 112, 777-791. https://doi.org/10.1093/ajcn/nqaa159

Yani, D. I., Rahayuwati, L., Sari, C. W. M., Komariah, M., & Fauziah, S. R. (2023). Family household characteristics and stunting: an update scoping review. *Nutrients*, 15(1), 233. https://doi.org/10.3390/nu15010233

Yehuala, T. Z., Baykemagn, N., & Terefe, B. (2025). Machine learning algorithms to predict feeding practices during diarrheal disease and its determinants among under-five children in East Africa. *Frontiers in Public Health*, 13, 1513922. https://doi.org/10.3389/fpubh.2025.1513922