

An Explainable Stacking Ensemble Model for Predicting Childhood Malnutrition among Under-Five Children in Nigeria

Zainab Ahmed*, Bala Modi, Ali Ahmad Aminu, Abdulrashid Isiyaku

Department of Computer Science, Faculty of Science, Gombe State University, Gombe, Nigeria.

*Corresponding Author

DOI: <https://doi.org/10.51584/IJRIAS.2026.11070185>

Received: 04 August 2026; Accepted: 10 August 2026; Published: 19 August 2026

ABSTRACT

Childhood malnutrition remains a major public health challenge in Nigeria, contributing to child morbidity, mortality, impaired cognitive development, and poor long-term socioeconomic outcomes. Although machine learning has shown promise for malnutrition prediction, many models operate as black boxes, limiting their interpretability and adoption in clinical and public health decision-making. This study developed an explainable stacking ensemble model for classifying and predicting childhood malnutrition among children Under-five in Nigeria, using the 2023–24 Nigeria Demographic and Health Survey (NDHS). Data from 9,374 children with valid anthropometric measurements were analysed, with malnutrition status determined using WHO Child Growth Standards based on height-for-age, weight-for-age, and weight-for-height z-scores. Following preprocessing and a multi-method, multi-phenotype feature selection process combining Recursive Feature Elimination, Mutual Information, and Random Forest Feature Importance, twenty-two socioeconomic, demographic, geographic, maternal, and child-health variables were retained for modelling. A stacking ensemble comprising Random Forest, XGBoost, LightGBM, CatBoost, and Support Vector Machine as base learners, with Logistic Regression as meta-learner, was developed using mother-level stratified group cross-validation to prevent clustering-related data leakage. Within-fold SMOTE addressed class imbalance, and Youden's J statistic guided threshold optimisation. Model interpretability was achieved using SHAP, providing both global and individual-level explanations. The stacking ensemble outperformed all individual base learners, achieving a ROC-AUC of 0.750, F1-score of 0.662, recall of 72.6%, precision of 61.0%, and accuracy of 69.9% at the optimal threshold of 0.40. SHAP analysis identified household wealth index, maternal education, child age, and geographic location as the most influential predictors of malnutrition risk. These findings demonstrate that integrating ensemble learning with explainable artificial intelligence provides an accurate, transparent, and practical decision-support framework for early identification of at-risk children, supporting evidence-based nutrition interventions and public health policy in Nigeria.

Keywords: Childhood malnutrition, Stacking ensemble, Explainable artificial intelligence, SHAP, Nigeria Demographic and Health Survey (NDHS).

INTRODUCTION

Childhood malnutrition remains one of the most serious public health challenges worldwide, particularly in low- and middle-income countries, where it continues to threaten child survival, growth, and development. According to the World Health Organization, an estimated 149 million children are stunted, 45 million are wasted, and millions more are underweight globally (WHO, 2023). In Nigeria, childhood malnutrition continues to pose a substantial burden, with stunting, wasting, and underweight disproportionately affecting children from poor households and rural communities (FMoHSW, NPC, & ICF, 2025). Malnutrition during early childhood is associated with increased susceptibility to infectious disease, impaired cognitive development, poor educational attainment, and reduced economic productivity later in life (UNICEF, WHO, & World Bank Group, 2023).

The determinants of childhood malnutrition are multidimensional, involving socioeconomic, demographic, maternal, environmental, and health-related factors. Previous studies have identified household wealth,

maternal education, birth interval, sanitation, feeding practices, and access to healthcare as important predictors of child nutritional status (Ijaiya et al., 2022; Jebero et al., 2023; Tziritidou-Chatzopoulou et al., 2024). These findings suggest that effective malnutrition screening requires analytical approaches capable of capturing complex interactions among multiple risk factors.

Traditionally, statistical models such as logistic regression and multilevel regression have been widely used to investigate the determinants of childhood malnutrition. Studies using Nigeria Demographic and Health Survey (NDHS) data have demonstrated significant associations between malnutrition and socioeconomic as well as maternal characteristics (Akombi et al., 2017; Tesfaw & Fenta, 2021). Although these methods provide interpretable estimates, they are often limited in their ability to model nonlinear relationships and high-dimensional interactions among predictors (Shahid et al., 2025).

Recent advances in machine learning have provided more flexible alternatives for health prediction tasks. Algorithms such as Random Forest, Support Vector Machine, XGBoost, LightGBM, and CatBoost have shown improved predictive performance in malnutrition classification studies (Kawo et al., 2024; Mahmud et al., 2024; Yanto et al., 2024). Ensemble learning methods, which combine multiple base learners, have been particularly successful in improving robustness and generalisation (Nirmani & Kudagamage, 2024). Among ensemble approaches, stacking is especially attractive because it integrates the complementary strengths of diverse algorithms through a meta-learning framework.

Despite their predictive advantages, many machine learning models remain difficult to interpret. Their black-box nature limits their practical use in clinical and public health settings, where practitioners must understand why a child is classified as being at risk before making intervention decisions (Mahmud et al., 2024; Setegn & Dejene, 2025). Explainable Artificial Intelligence (XAI) has emerged as an important solution to this challenge. Techniques such as SHapley Additive exPlanations (SHAP) provide transparent explanations by quantifying the contribution of each predictor to a model's output (Lundberg & Lee, 2017). The integration of SHAP into health prediction models has been shown to improve interpretability and support evidence-based decision-making (Islam et al., 2024; Tamanna et al., 2025).

Although several studies have applied machine learning to childhood malnutrition prediction, important gaps remain. Many studies rely on small or localised datasets, limiting generalisability, while others emphasise predictive accuracy without incorporating explainability mechanisms (Qasrawi et al., 2024; Reusken et al., 2024). Furthermore, few studies have combined multiple state-of-the-art ensemble algorithms within a stacking framework while simultaneously applying SHAP-based interpretation using nationally representative Nigerian data, and safeguards against household- or mother-level clustering when partitioning survey data are rarely reported. This gap reduces the practical value of existing models for large-scale nutrition surveillance and policy planning in Nigeria.

To address these limitations, this study develops an explainable stacking ensemble learning model for classifying and predicting childhood malnutrition among children Under-five years of age in Nigeria using data from the 2023–24 Nigeria Demographic and Health Survey. The proposed framework combines Random Forest, XGBoost, LightGBM, CatBoost, and Support Vector Machine as base learners with Logistic Regression as the meta-learner. SHAP is employed to provide both global and individual-level explanations of model predictions. By integrating predictive accuracy with interpretability, the study aims to provide a transparent and scalable decision-support tool for the early identification of children at risk of malnutrition and for informing targeted nutrition interventions and public health policy in Nigeria. Specifically, the study (i) designs a stacking ensemble model for classifying child malnutrition status based on socioeconomic, demographic, geographic, and health-related predictors; (ii) applies SHAP to interpret the model at global and individual levels and identify children at elevated risk; and (iii) evaluates model performance using accuracy, precision, recall, F1-score, and ROC-AUC.

RELATED WORK

Traditional Statistical Approaches to Childhood Malnutrition Prediction

Childhood malnutrition has traditionally been investigated using statistical models such as logistic regression,

multivariate regression, and structural equation modelling to identify the socioeconomic, demographic, and environmental determinants of nutritional outcomes. These approaches have consistently demonstrated that household wealth, parental education, maternal nutritional status, birth characteristics, sanitation, and access to healthcare are significant predictors of stunting, wasting, and underweight among children Under-five. For example, Tesfaw and Fenta (2021) employed multivariate logistic regression using NDHS data and identified household wealth index and maternal characteristics as major determinants of child malnutrition. Similarly, Akombi et al. (2017) applied multilevel logistic regression to Nigerian DHS data to account for household- and community-level influences on childhood nutritional status. Although these methods provide interpretable statistical relationships, they are often limited in modelling the complex nonlinear interactions that characterise childhood malnutrition (Shahid et al., 2025).

Machine Learning Approaches for Childhood Malnutrition Prediction

The increasing availability of large-scale demographic and health datasets has accelerated the adoption of machine learning (ML) techniques for childhood malnutrition prediction. Unlike traditional regression models, ML algorithms can effectively capture complex nonlinear relationships and interactions among demographic, socioeconomic, maternal, and environmental variables, thereby improving predictive performance. Consequently, several studies have explored the application of supervised learning algorithms such as Decision Trees (DT), Random Forest (RF), Support Vector Machine (SVM), K-Nearest Neighbour (KNN), Naïve Bayes (NB), Artificial Neural Networks (ANN), and Extreme Gradient Boosting (XGBoost) for predicting nutritional outcomes among children Under-five years of age.

Using data from the Nigerian Multiple Indicator Cluster Survey (MICS), Kawo et al. (2024) compared KNN, Random Forest, Decision Tree, and Logistic Regression models for predicting stunting, wasting, and underweight among children. Their findings showed that the KNN model achieved the highest predictive performance, reporting accuracies of 89.65% for stunting, 95.00% for wasting, and 80.04% for underweight, with household wealth, child age, maternal education, and place of residence identified as the most influential predictors. Similarly, Mahmud et al. (2024) evaluated Decision Tree, Random Forest, Support Vector Machine, and Multi-layer Perceptron (MLP) using the 2019 Bangladesh MICS dataset, reporting that Random Forest and MLP achieved predictive accuracies of 98.55% and 98.69% respectively, and highlighted breastfeeding status and weight-for-age as key determinants of nutritional status.

Aryuni et al. (2023) developed a nutritional status prediction model comparing C4.5 Decision Tree, KNN, and Naïve Bayes, in which the C4.5 algorithm achieved the best performance with an accuracy of 89.87% and an F1-score of 91.67%. Wajgi and Wajgi (2022) applied Decision Tree models to classify severe acute malnutrition (SAM), moderate acute malnutrition (MAM), and normal nutritional status using anthropometric measurements, demonstrating the effectiveness of Gini-index-based splitting for improving classification accuracy. Kishore et al. (2023) compared Support Vector Machine, K-Nearest Neighbour, Logistic Regression, Naïve Bayes, and a bilayered Artificial Neural Network for detecting malnutrition among children under two years of age, with SVM and the neural network producing superior classification performance. Yanto et al. (2024) proposed an integrated framework combining K-Means clustering, Pearson correlation analysis, Artificial Neural Networks, and Decision Trees, in which the ANN classifier achieved an accuracy of 99.23%, illustrating the capability of neural-network-based techniques for nutritional status prediction.

Collectively, these studies demonstrate that machine learning algorithms substantially improve the prediction of childhood malnutrition by modelling complex nonlinear relationships that conventional statistical techniques often fail to capture. However, despite their impressive predictive performance, most existing models prioritise accuracy over interpretability, making it difficult for healthcare professionals and policymakers to understand the factors influencing model predictions. This lack of transparency remains a significant limitation to the practical adoption of machine learning in public health and provides a strong motivation for integrating explainable artificial intelligence with advanced ensemble learning approaches.

Ensemble Learning for Healthcare Prediction

Ensemble learning has become one of the most effective approaches for improving predictive performance in healthcare applications. Unlike individual machine learning algorithms, ensemble methods integrate multiple

base learners to exploit their complementary strengths, thereby improving model robustness, reducing variance, and enhancing generalisation. Common ensemble strategies include bagging, boosting, voting, and stacking, each employing a different mechanism for combining individual classifiers. Bagging techniques such as Random Forest reduce prediction variance through bootstrap aggregation, while boosting algorithms, including XGBoost, LightGBM, and CatBoost, iteratively improve weak learners by emphasising previously misclassified observations.

Among these approaches, stacking has received increasing attention because it combines predictions from multiple heterogeneous base learners through a meta-learning algorithm. Rather than relying on a single modelling paradigm, stacking enables different algorithms to capture diverse patterns within complex datasets, allowing the meta-learner to produce more accurate and stable predictions. This characteristic makes stacking particularly suitable for healthcare datasets, where nonlinear relationships, feature interactions, and heterogeneous patient characteristics often reduce the effectiveness of individual classifiers. Recent healthcare studies have demonstrated the effectiveness of stacking ensembles in disease diagnosis, clinical risk prediction, and medical decision support, with similar improvements reported in public health applications involving maternal and child health, where stacking has consistently outperformed individual machine learning models in terms of predictive accuracy and robustness.

Despite these advances, the application of stacking ensemble learning to childhood malnutrition prediction remains relatively limited. Most existing studies have focused on individual machine learning algorithms or simple ensemble techniques such as Random Forest and Gradient Boosting, with comparatively few investigations adopting comprehensive stacking architectures that integrate multiple state-of-the-art classifiers. Furthermore, studies combining stacking ensembles with explainable artificial intelligence techniques remain scarce, particularly when applied to nationally representative demographic and health survey datasets from sub-Saharan Africa. This limitation provides an important motivation for the present study, which integrates heterogeneous ensemble learning with SHAP-based explainability to develop an accurate and interpretable framework for predicting childhood malnutrition.

Explainable Artificial Intelligence in Healthcare Prediction

As machine learning models become increasingly integrated into healthcare decision support systems, model interpretability has become an essential requirement. Explainable Artificial Intelligence (XAI) techniques provide mechanisms for understanding how complex machine learning models generate predictions, thereby improving transparency, accountability, and user trust. Among existing XAI approaches, SHapley Additive exPlanations (SHAP) has emerged as one of the most widely adopted methods because it provides both global feature importance and local explanations for individual predictions (Lundberg & Lee, 2017). Recent studies have successfully combined SHAP with machine learning algorithms to identify key determinants of childhood malnutrition and other health outcomes. Islam et al. (2024) used a feature selection process combined with SHAP analysis to identify parental education, household wealth, maternal body mass index, sanitation, and water source as major determinants of childhood malnutrition in Bangladesh. Tamanna et al. (2025) similarly incorporated SHAP to explain Random Forest predictions for stunting, wasting, and underweight, while Setegn and Dejene (2025) highlighted the importance of integrating advanced explainability techniques with ensemble learning models for nutritional prediction. Although these studies demonstrate the value of explainable machine learning, the application of SHAP within advanced stacking ensemble frameworks remains relatively limited, particularly for nationally representative datasets from sub-Saharan Africa.

Research Gap

The reviewed literature demonstrates that machine learning has substantially improved the prediction of childhood malnutrition compared with conventional statistical approaches. However, several important limitations remain. First, many existing studies rely on localised or relatively small datasets, reducing the generalisability of their findings. Second, although ensemble learning methods have consistently achieved superior predictive performance, relatively few studies have employed comprehensive stacking architectures that integrate multiple state-of-the-art learners within a single predictive framework. Third, the majority of existing models remain black-box systems with limited interpretability, restricting their acceptance by healthcare professionals and policymakers. Fourth, safeguards against household- or mother-level clustering

when partitioning survey data for training and testing are rarely reported, creating a risk of data leakage and optimistic performance estimates. Finally, despite Nigeria's high burden of childhood malnutrition, there remains limited evidence on explainable stacking ensemble models developed using the most recent nationally representative 2023–24 NDHS dataset.

This study addresses these limitations by developing an explainable stacking ensemble framework that integrates Random Forest, XGBoost, LightGBM, CatBoost, and Support Vector Machine with Logistic Regression as the meta-learner. The proposed model combines rigorous multi-method feature selection, mother-level cluster-aware cross-validation, threshold optimisation, and SHAP-based interpretation to provide both high predictive performance and transparent explanations of childhood malnutrition risk. By leveraging the nationally representative 2023–24 NDHS dataset, the study contributes an interpretable and scalable decision-support tool capable of supporting evidence-based nutrition interventions and public health policy in Nigeria.

METHODOLOGY

Research Framework

This study proposes an explainable stacking ensemble learning framework for predicting childhood malnutrition among children Under-five years of age using data from the 2023–24 Nigeria Demographic and Health Survey (NDHS). The methodological framework was designed to ensure systematic data preparation, robust model development, transparent prediction, and comprehensive model evaluation. As illustrated in Figure 1, the proposed framework comprises six sequential stages: data acquisition from the NDHS; data preprocessing to improve data quality through cleaning, transformation, encoding, and handling of missing values; feature engineering and multi-method feature selection to identify the most informative predictors; training of five heterogeneous base learners within a five-fold, mother-level stratified group cross-validation scheme with within-fold SMOTE; combination of base-learner outputs by a Logistic Regression meta-learner with threshold optimisation; and SHAP-based interpretation followed by comprehensive evaluation on a held-out test set.

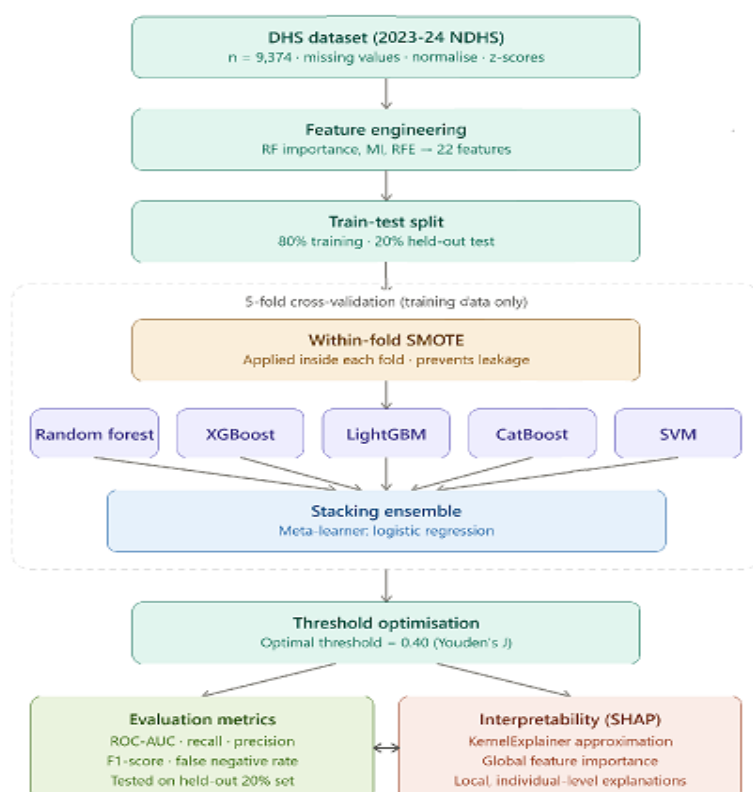


Figure 1: Overall Methodological Framework of the Proposed Explainable Stacking Ensemble Model

Data Source and Study Population

The study utilised data from the 2023–24 Nigeria Demographic and Health Survey (NDHS), implemented by the Federal Ministry of Health and Social Welfare (FMoHSW) and the National Population Commission (NPC), with technical assistance from ICF (FMoHSW, NPC, and ICF, 2025). The NDHS is one of the largest nationally representative health surveys in Nigeria and applies a stratified two-stage cluster sampling technique to ensure adequate representation across the country's geopolitical zones.

The analysis focused on children aged 0–59 months with valid anthropometric measurements. Anthropometric indices were computed according to the WHO Child Growth Standards using height-for-age z-score (HAZ), weight-for-height z-score (WHZ), and weight-for-age z-score (WAZ) (World Health Organization, 2006). Children without all three z-score measurements, or with implausible values outside WHO-recommended ranges, were excluded from the analysis. Following data cleaning and quality assessment, the final analytical dataset comprised 9,374 children aged 0–59 months, drawn from 18,399 unique mothers across all six geopolitical zones of Nigeria.

The outcome variable represented a composite measure of childhood malnutrition, combining three WHO anthropometric indicators: stunting ($HAZ < -2$ SD), wasting ($WHZ < -2$ SD), and underweight ($WAZ < -2$ SD). A child meeting the threshold for at least one indicator was classified as malnourished. These indicators are internationally recognised measures for assessing chronic, acute, and composite forms of childhood malnutrition, respectively, and their combination into a single binary outcome supports the development of a general-purpose screening tool rather than a phenotype-specific model.

Data Preprocessing and Feature Engineering

Data preprocessing was undertaken to improve data quality and ensure compatibility with machine learning algorithms. The preprocessing pipeline included data cleaning, handling of missing values, feature transformation, and categorical encoding. Missing numerical values were imputed using the mean, and missing categorical values using the mode. Categorical variables were transformed into numerical representations using one-hot encoding, while continuous variables such as child age and maternal age were scaled using Min-Max normalisation to improve convergence for scale-sensitive algorithms.

Feature engineering was performed to derive candidate variables relevant to childhood malnutrition prediction, drawn from socioeconomic, demographic, maternal, geographic, environmental, and child-health domains available within the NDHS dataset. These included household wealth index, maternal education, maternal age, child age, birth order, birth size, sex of the child, breastfeeding status, immunisation history, vitamin A supplementation, deworming status, improved sanitation, improved drinking water source, urban–rural residence, geopolitical zone, and recent childhood illnesses (fever, diarrhoea, cough), among others.

The dataset was subsequently partitioned into training (80%) and held-out test (20%) subsets using stratified sampling to preserve the distribution of the outcome classes. Because childhood malnutrition data are inherently imbalanced, the Synthetic Minority Oversampling Technique (SMOTE) was applied exclusively within each training fold during cross-validation, rather than on the full training set prior to splitting, thereby preventing information leakage into validation and test data.

Feature Selection

Selecting the most informative predictor variables is essential for improving model performance, reducing computational complexity, and enhancing interpretability. A multi-method, multi-phenotype feature selection strategy was adopted, applying three complementary techniques; Recursive Feature Elimination (RFE), Mutual Information (MI), and Random Forest Feature Importance (RF-FI) independently to each of the three malnutrition phenotypes (stunting, wasting, and underweight).

Recursive Feature Elimination, implemented using a Random Forest estimator with 10-fold cross-validation at each iteration, iteratively removed the least important features until the optimal subset remained (Guyon et al.,

2002). Mutual Information measured the statistical dependency between each predictor and the outcome, capturing both linear and non-linear relationships (Vergara & Estévez, 2014). Random Forest Feature Importance ranked predictors according to their average contribution to reducing Gini impurity across the ensemble of trees (Breiman & Cutler, 2001; Biau & Scornet, 2016).

For each phenotype, a phenotype-specific consensus feature set was derived by retaining features selected by at least two of the three methods. A global consensus feature set was then constructed by retaining features that were consensus features for at least two of the three phenotypes, ensuring that the final predictor set captured variables broadly relevant across different forms of malnutrition rather than being specific to a single phenotype. This process converged on sixteen features selected by all three methods across all three phenotypes, with an additional six features selected by at least two methods for at least two phenotypes, yielding a final global consensus set of twenty-two predictor variables (Table 1).

Table 1: Global Consensus Feature Set Retained for Model Development

Feature	Stunting	Wasting	Underweight	Frequency
Household wealth index	✓	✓	✓	3/3
Maternal education	✓	✓	✓	3/3
Maternal age	✓	✓	✓	3/3
Child age (months)	✓	✓	✓	3/3
Child sex	✓	✓	✓	3/3
Birth order	✓	✓	✓	3/3
Birth size	✓	✓	✓	3/3
Currently breastfeeding	✓	✓	✓	3/3
Urban residence	✓	✓	✓	3/3
Geopolitical zone (Zone 3)	✓	✓	✓	3/3
Improved sanitation	✓	✓	✓	3/3
Improved water source	✓	✓	✓	3/3
Fever in last 2 weeks	✓	✓	✓	3/3
Deworming (last 6 months)	✓	✓	✓	3/3
Vitamin A supplementation	✓	✓	✓	3/3
Iron supplementation (12m)	✓	✓	✓	3/3
Bed-net use	✓	✓	✓	3/3
Bottle-fed	✗	✓	✓	2/3
Religion (category 3)	✓	✗	✓	2/3
Religion (category 2)	✓	✗	✓	2/3
Polio 3 received	✓	✓	✗	2/3
Geopolitical zone (Zone 2)	✓	✗	✓	2/3

The prominence of socioeconomic indicators such as household wealth index and maternal education is consistent with previous studies identifying poverty and maternal educational attainment as major determinants of childhood malnutrition. Similarly, the recurrence of child age, birth size, breastfeeding practices, and environmental conditions across all three phenotypes reinforces existing evidence that childhood malnutrition results from a complex interaction of biological, socioeconomic, and environmental factors, and provides additional justification for an ensemble learning framework capable of modelling such interactions more effectively than conventional statistical techniques.

Proposed Explainable Stacking Ensemble Architecture

The proposed prediction model employs a heterogeneous stacking ensemble architecture to leverage the

complementary strengths of multiple machine learning algorithms. Five base learners Random Forest (RF), Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), CatBoost, and Support Vector Machine (SVM) were selected for their demonstrated effectiveness in handling nonlinear relationships, heterogeneous data distributions, and high-dimensional healthcare datasets (Chen & Guestrin, 2016; Ke et al., 2017; Prokhorenkova et al., 2018; Cortes & Vapnik, 1995). Each algorithm was tuned via grid search with cross-validation to identify parameter configurations that maximised predictive performance while limiting overfitting.

As illustrated in Figure 2, each base learner independently receives the selected feature subset and generates prediction probabilities through five-fold cross-validation, stratified and grouped by mother identifier so that children from the same mother do not appear in both training and validation folds, preventing clustering-related data leakage. The out-of-fold predicted probabilities produced by the five base learners are concatenated to form a new feature matrix, which serves as input to a Logistic Regression meta-learner. The meta-learner combines the base-learner predictions to generate the final stacked classification decision, exploiting the complementary strengths of each algorithm while mitigating individual weaknesses.

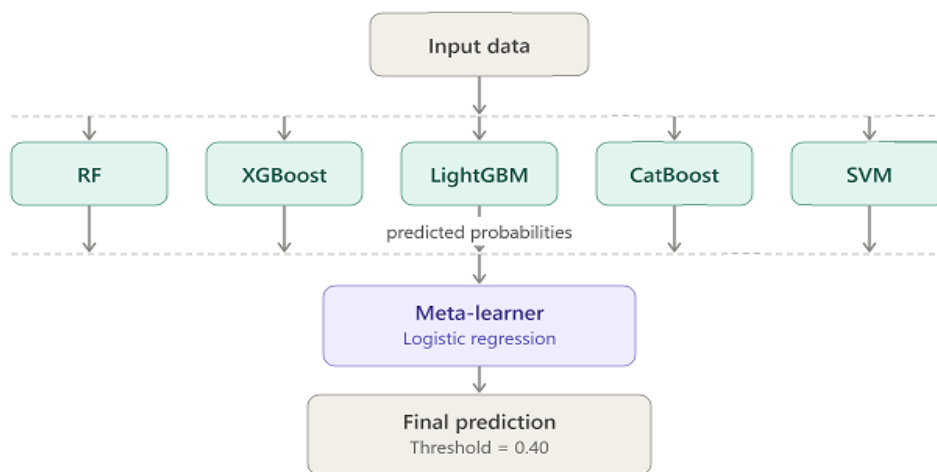


Figure 2: Architecture of the Proposed Explainable Stacking Ensemble Model

Model Training, Cross-Validation, and Threshold Optimisation

The analytical dataset was partitioned into training and held-out test subsets using stratified sampling to preserve the distribution of the target classes. Because childhood malnutrition data are naturally imbalanced, with substantially fewer malnourished than well-nourished children, SMOTE was applied exclusively to the training data within each cross-validation fold, thereby preventing information leakage into validation or test data. Rather than adopting the conventional probability threshold of 0.50, threshold optimisation was performed using Youden's J statistic to determine the decision threshold that achieved the best balance between sensitivity and specificity, with performance evaluated across threshold values ranging from 0.25 to 0.55.

SHAP-Based Model Explainability

To enhance transparency and facilitate interpretation, SHapley Additive exPlanations (SHAP) were employed to explain predictions generated by the stacking ensemble model using a Kernel Explainer approximation (Lundberg & Lee, 2017). SHAP is a game-theoretic explainability technique that assigns each predictor variable a contribution score representing its influence on the model's prediction, computed as:

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|! (|F| - |S| - 1)!}{|F|!} \times [f(S \cup \{i\}) - f(S)]$$

where ϕ_i is the SHAP value for feature i , F is the full set of features, S is a subset of features excluding feature i , and $f(S \cup \{i\}) - f(S)$ is the change in the model's prediction when feature i is added to subset S . Global

explainability was achieved through SHAP summary plots ranking predictor variables by their overall contribution across the dataset, while local explainability was provided through individual-level SHAP explanations illustrating how specific predictor values increased or decreased the predicted malnutrition risk for a given child. The integration of SHAP within the proposed framework enables healthcare professionals and policymakers to understand not only whether a child is predicted to be at risk of malnutrition, but also the specific factors responsible for that prediction, supporting responsible adoption of artificial intelligence within healthcare systems.

Performance Evaluation Metrics

Model performance was evaluated using the confusion matrix and five classification metrics: Accuracy, Precision, Recall (Sensitivity), F1-score, and the Area Under the Receiver Operating Characteristic Curve (ROC-AUC). Where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives respectively:

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN})$$

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$$

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$$

$$\text{F1-score} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$$

ROC-AUC is computed as the area under the curve formed by plotting the True Positive Rate against the False Positive Rate across all classification thresholds. These metrics jointly provide a comprehensive evaluation of predictive power and error trade-offs, which is essential for health-related decision-support systems where the costs of false positives and false negatives differ.

RESULTS AND DISCUSSION

Descriptive Characteristics of the Study Sample

Following data cleaning, preprocessing, and quality assessment, the final analytical dataset comprised 9,374 children aged 0–59 months with complete anthropometric and predictor information, drawn from 18,399 unique mothers across all six geopolitical zones of Nigeria. Table 2 presents the descriptive characteristics of the study population. The sample was near-evenly split by sex (50.5% male, 49.5% female) and predominantly rural (57.4%), consistent with Nigeria's population structure. Maternal education was low, with 36.7% of mothers having no formal education, while household wealth was broadly distributed across all five quintiles. Applying WHO cut-offs, 36.6% of children were stunted, 8.2% were wasted, and 24.7% were underweight; combined into the composite outcome, 43.1% (4,037 children) were classified as malnourished and 56.9% (5,337 children) as not malnourished.

Table 2: Descriptive Characteristics of the Study Sample (n = 9,374)

Characteristic	n	%
Age (months)		
0–5	1,002	10.69
6–11	1,017	10.85
12–23	1,936	20.65
24–35	1,765	18.83
36–47	1,798	19.18
48–59	1,856	19.80
Sex		
Male	4,733	50.49
Female	4,641	49.51

Residential area		
Urban	3,991	42.58
Rural	5,383	57.42
Birth order		
1st	1,879	20.04
2nd	1,804	19.24
3rd	1,648	17.58
4th	1,287	13.73
5th or higher	2,756	29.40
Maternal education		
No education	3,441	36.71
Primary	1,138	12.14
Secondary	3,598	38.38
Higher	1,197	12.77
Household wealth index		
Poorest	2,020	21.55
Poorer	1,703	18.17
Middle	1,841	19.64
Richer	2,071	22.09
Richest	1,739	18.55
Geopolitical zone		
North Central	1,678	17.90
North East	1,643	17.53
North West	2,383	25.42
South East	1,686	17.99
South South	1,105	11.79
South West	879	9.38
Currently breastfeeding		
Yes	4,908	52.36
No	4,466	47.64
Fever in last 2 weeks		
Yes	1,564	16.68
No	7,810	83.32
Diarrhoea in last 2 weeks		
Yes	1,265	13.49
No	8,105	86.46
Improved drinking water source		
Yes	6,690	71.37
No	2,684	28.63
Improved sanitation		
Yes	5,430	57.93
No	3,944	42.07
Malnutrition status (WHO criteria)		
Stunted (HAZ < -2 SD)	3,429	36.58
Wasted (WHZ < -2 SD)	767	8.18
Underweight (WAZ < -2 SD)	2,312	24.66
Malnourished (composite outcome)	4,037	43.07
Not malnourished (composite outcome)	5,337	56.93

Feature Selection Results

As described in Section 3.4, the consensus-based, multi-phenotype feature selection process converged on a final global feature set of twenty-two predictor variables (Table 1), spanning socioeconomic, demographic,

maternal, environmental, and child-health domains. Seventeen features including household wealth index, maternal education, maternal age, child age, birth order, birth size, breastfeeding status, urban residence, geopolitical zone, sanitation, water source, and immunisation-related variables were selected by all three methods (RFE, MI, and RF-FI) across all three malnutrition phenotypes, indicating strong and consistent predictive relevance. An additional five variables were retained on the basis of consensus across at least two of the three phenotypes. By eliminating redundant and less informative variables, the feature selection process reduced computational complexity, improved model generalisation, and enhanced the interpretability of the final prediction framework.

Performance of Individual Machine Learning Models

The predictive performance of the five individual base learners was evaluated on the independent held-out test set prior to constructing the stacking ensemble. Table 3 presents the comparative performance of Random Forest, XGBoost, LightGBM, CatBoost, and Support Vector Machine using Accuracy, Precision, Recall, F1-score, and ROC-AUC.

Table 3: Performance Comparison of Individual Machine Learning Models on the Held-Out Test Set

Model	ROC-AUC	F1-Score	Recall	Precision	Accuracy
CatBoost	0.746	0.626	58.2%	67.6%	70.0%
Random Forest	0.745	0.606	54.9%	67.6%	69.3%
XGBoost	0.734	0.609	57.2%	65.1%	68.4%
LightGBM	0.729	0.602	56.6%	64.3%	67.8%
Support Vector Machine	0.735	0.578	50.8%	67.1%	68.1%

All five algorithms achieved competitive and closely clustered performance (ROC-AUC 0.729–0.746), indicating that the selected feature subset effectively captured the key determinants of childhood malnutrition. Among the individual base learners, CatBoost achieved the highest overall performance, attaining the largest ROC-AUC (0.746) and F1-score (0.626), likely owing to its ordered boosting mechanism and native handling of categorical predictors, which are prevalent in DHS survey data. Support Vector Machine achieved comparatively lower recall (50.8%), reflecting the difficulty of a single kernel-based decision boundary in capturing the heterogeneous, non-linear feature interactions present in the dataset. Critically, no single classifier dominated across all metrics: CatBoost and Random Forest achieved the highest precision (67.6%), while XGBoost and LightGBM achieved comparatively higher recall. This complementary pattern of strengths and weaknesses across algorithms motivated the development of the proposed stacking ensemble, which combines the base learners through a meta-learner in order to exploit their complementary predictive information.

Performance of the Proposed Stacking Ensemble Model

Table 4 compares the performance of the proposed stacking ensemble, evaluated at its Youden's J-optimised decision threshold of 0.40, against the five individual base learners on the same held-out test set.

Table 4: Performance Comparison Between Individual Models and the Proposed Stacking Ensemble

Model	ROC-AUC	F1-Score	Recall	Precision	Accuracy
Random Forest	0.745	0.606	54.9%	67.6%	69.3%
XGBoost	0.734	0.609	57.2%	65.1%	68.4%
LightGBM	0.729	0.602	56.6%	64.3%	67.8%
CatBoost	0.746	0.626	58.2%	67.6%	70.0%
Support Vector Machine	0.735	0.578	50.8%	67.1%	68.1%
Proposed Stacking Ensemble (0.40)	0.750	0.662	72.6%	61.0%	69.9%

The results indicate that the proposed stacking ensemble achieved the best overall discriminative performance among all evaluated models, attaining the highest ROC-AUC (0.750) and F1-score (0.662). Compared with the

strongest individual classifier (CatBoost), the stacking ensemble improved F1-score by 5.8% and recall by 24.7%, at the cost of a moderate reduction in precision. This trade-off reflects the deliberate prioritisation of recall in the design of the ensemble, which is appropriate for a screening application in which the cost of missing a malnourished child (a false negative) is considerably higher than the cost of investigating a false positive. Figure 3 presents the ROC curve for the stacking ensemble, confirming that the model consistently outperforms the random-classifier reference line across the full range of thresholds.

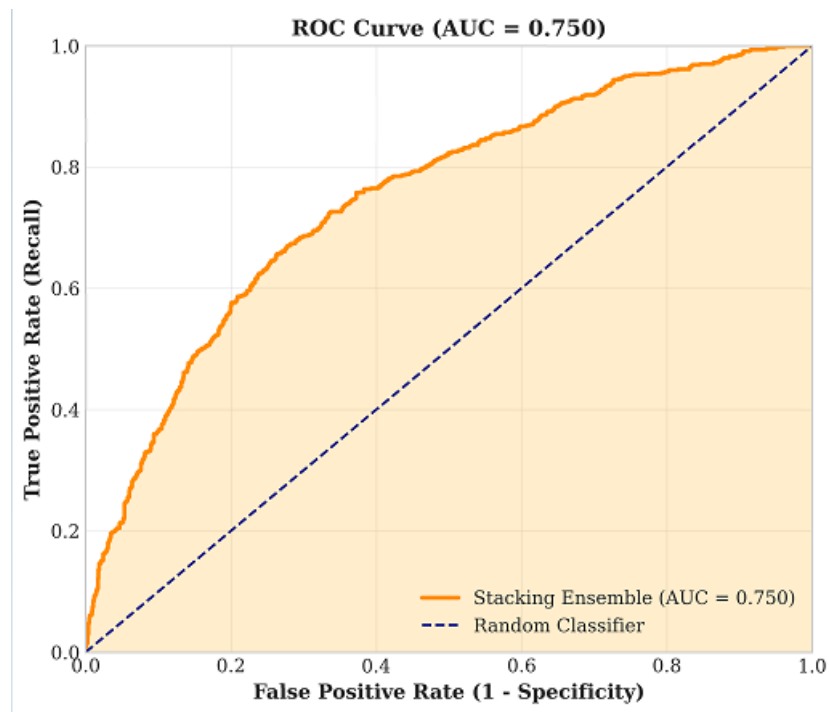


Figure 3: ROC Curve for the Proposed Stacking Ensemble Model (AUC = 0.750)

By combining the complementary strengths of heterogeneous classifiers through a Logistic Regression meta-learner, the ensemble effectively reduced the weaknesses associated with individual algorithms while improving generalisation. This finding is consistent with the wider healthcare literature, where stacking has been repeatedly shown to outperform individual classifiers by exploiting complementary information that would otherwise remain unexploited by standalone algorithms (Islam et al., 2024; Tamanna et al., 2025).

Threshold Optimisation and Cost Analysis

A default classification threshold of 0.50 does not necessarily provide the most appropriate balance between precision and recall for imbalanced healthcare datasets. Threshold optimisation using Youden's J statistic was therefore performed across threshold values from 0.25 to 0.55. Table 5 summarises the resulting trade-offs.

Table 5: Performance of the Proposed Stacking Ensemble at Different Decision Thresholds

Threshold	ROC-AUC	F1-Score	Recall	Precision
0.25	0.739	0.629	96.7%	46.6%
0.30	0.739	0.641	93.4%	48.8%
0.35	0.750	0.661	79.9%	56.4%
0.40 (optimal)	0.750	0.662	72.6%	61.0%
0.45	0.739	0.658	75.7%	58.3%
0.50	0.750	0.618	56.5%	68.2%
0.55	0.750	0.578	47.7%	71.8%

At the lowest threshold tested (0.25), the model achieved near-maximal recall (96.7%), correctly flagging almost all malnourished children, but at the cost of low precision (46.6%) and a correspondingly high rate of

false alarms. At the highest threshold tested (0.55), precision improved to 71.8% but recall declined sharply to 47.7%, meaning that more than half of malnourished children would be missed. The threshold of 0.40 produced the highest F1-score (0.662) and the most favourable overall balance between sensitivity and specificity, and was therefore adopted as the recommended operating threshold for the final model. The consistently high ROC-AUC (0.739–0.750) across all thresholds confirms that the model's underlying discriminative ability is stable and not an artefact of threshold selection; rather, threshold choice determines where along a consistently strong operating curve the model is deployed, and should be guided by the relative costs of false positives and false negatives in the specific screening context.

Confusion Matrix Analysis

Figure 4 presents the confusion matrix for the stacking ensemble at the recommended threshold of 0.40. The model correctly classified 586 of 807 malnourished children in the test set (72.6% sensitivity) and 694 of 1,068 not-malnourished children (65.0% specificity). The relatively high false-positive rate (35.0%) reflects the deliberate recall-oriented calibration of the model: in a screening context, a child incorrectly flagged for further assessment carries a much lower cost than a malnourished child who is missed entirely.

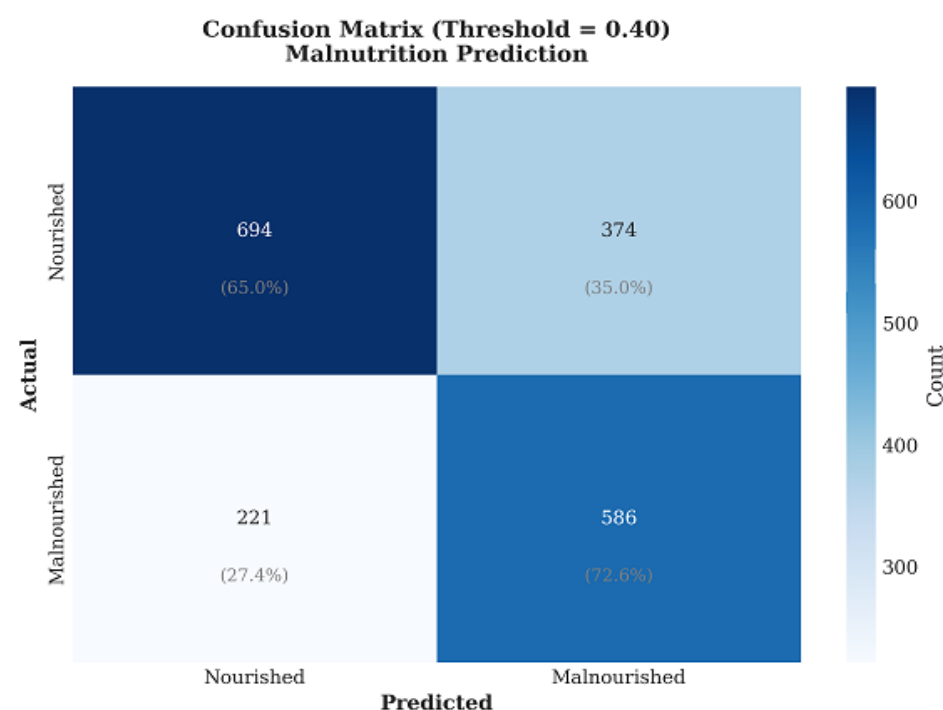


Figure 4: Confusion Matrix for the proposed Stacking Ensemble at Threshold 0.40

Calibration, Clinical Utility, and Geographic Generalisability

Beyond discrimination metrics, calibration analysis, decision-curve analysis (DCA), and leave-one-zone-out validation were performed to evaluate the reliability, clinical utility, and geographic generalisability of the stacking ensemble. The model achieved a Brier score of 0.201, which was lower than the uninformative baseline score of 0.245 derived from the test-set prevalence, indicating better probabilistic performance than the prevalence-based reference. The calibration slope was 1.293 (ideal = 1.0), with an intercept of 0.071 (ideal = 0.0), indicating that the predicted probabilities were somewhat underconfident, with observed risks tending to be more extreme than predictions across the probability range. Decision-curve analysis showed that the model provided greater net benefit than both the "treat-all" and "treat-none" strategies across threshold probabilities of 0.14 to 0.82, encompassing the 0.40 operating threshold selected via Youden's J statistic. These findings support the clinical utility of the model as a screening decision-support tool, while also highlighting the need to account for the observed calibration limitation in practice.

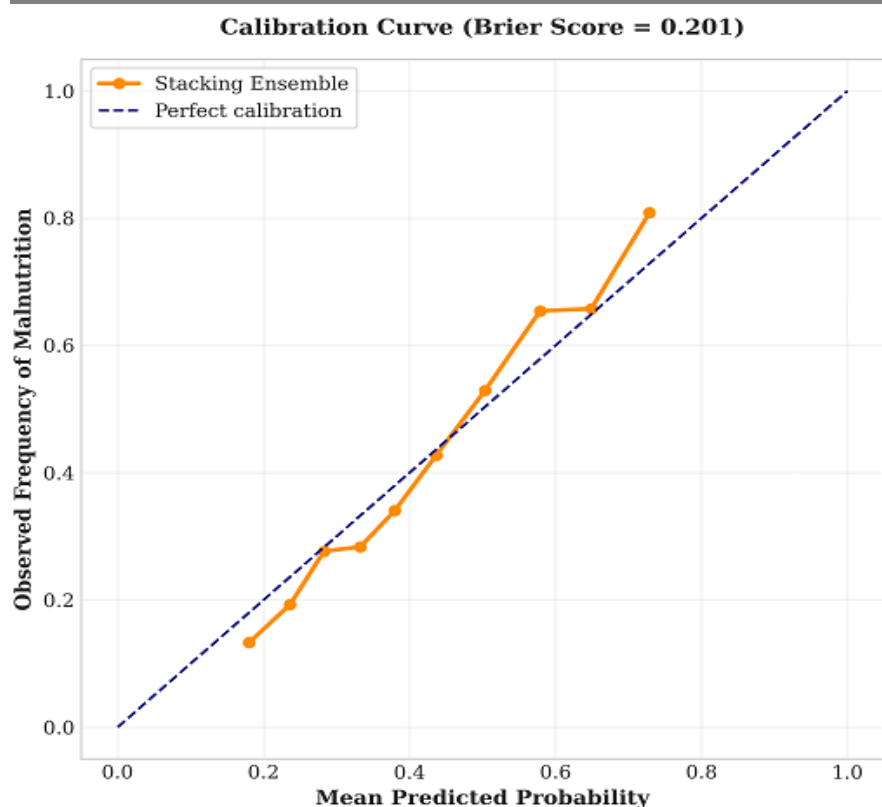


Figure 5: Calibration Curve for the proposed Stacking Ensemble (Brier Score = 0.201)

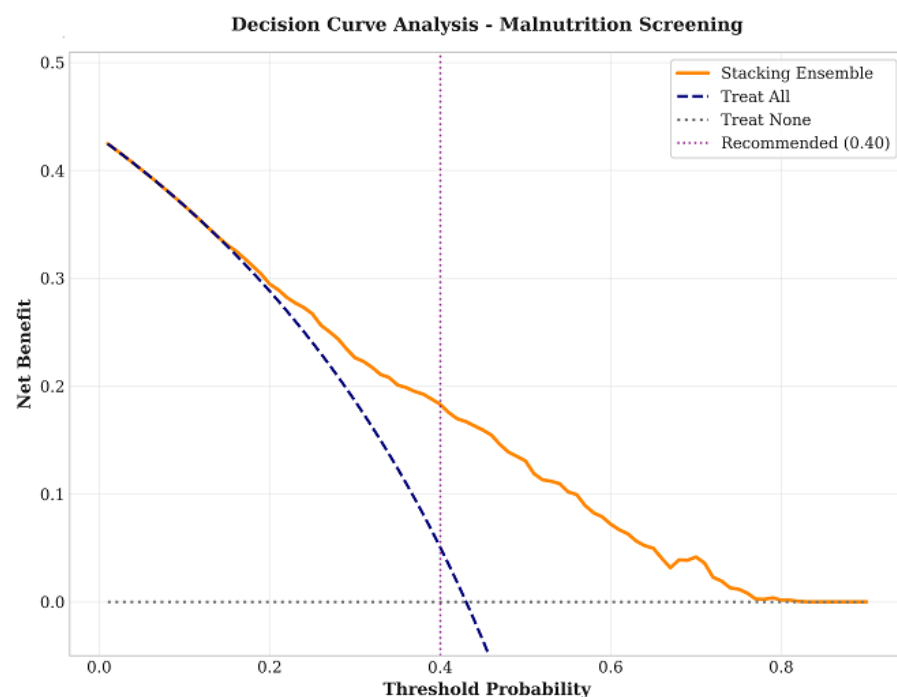


Figure 6: Decision-Curve Analysis of the proposed Stacking Ensemble for Malnutrition Screening

Geographic Generalisability Assessment

To assess geographic generalisability, the stacking ensemble was retrained six times, each time excluding one of Nigeria's six geopolitical zones from training and evaluating exclusively on the held-out zone, using the same 22-feature consensus set, within-fold SMOTE, and 0.40 operating threshold as the main analysis. Mean

ROC-AUC across the six zones was 0.641 (SD = 0.015), compared with 0.750 in the original random train-test split, indicating a meaningful decline in discriminative performance when the model is applied to a geopolitical zone entirely absent from training. F1-score, recall, and precision varied substantially across zones (F1 range: 0.35–0.72), largely reflecting differences in malnutrition prevalence between zones (26.2%–57.6%), while discriminative performance remained comparatively stable across zones (AUC range: 0.622–0.664).

This gap suggests that part of the model's original performance is attributable to having observed data from every zone during training, including zone-specific patterns represented by the zone_2 and zone_3 indicator features retained in the consensus feature set. When a zone is held out entirely, the model cannot learn a zone-specific effect for that unseen zone, which may contribute to the observed performance reduction. This finding tempers the geographic generalisability suggested by the original random-split evaluation and indicates that deployment in a zone not represented in the training data may require zone-specific recalibration or retraining using locally representative data. For practical deployment, this implies that models developed on nationally aggregated data should be validated and, if necessary, recalibrated at the subnational level before use in routine screening programmes.

Table 6: Leave-one-geopolitical-zone-out validation results of the proposed stacking ensemble.

Zone	N	Prevalence	AUC	F1	Recall	Precision	Accuracy
1	1678	0.408224	0.65561	0.598719	0.750365	0.498062	0.589392
2	1643	0.570298	0.664028	0.710428	0.814301	0.630058	0.621424
3	2383	0.576164	0.637075	0.723039	0.859432	0.624008	0.620646
4	1686	0.261566	0.634973	0.402969	0.430839	0.378486	0.666074
5	1105	0.290498	0.635319	0.347328	0.283489	0.448276	0.690498
6	879	0.318544	0.621989	0.369048	0.332143	0.415179	0.638225

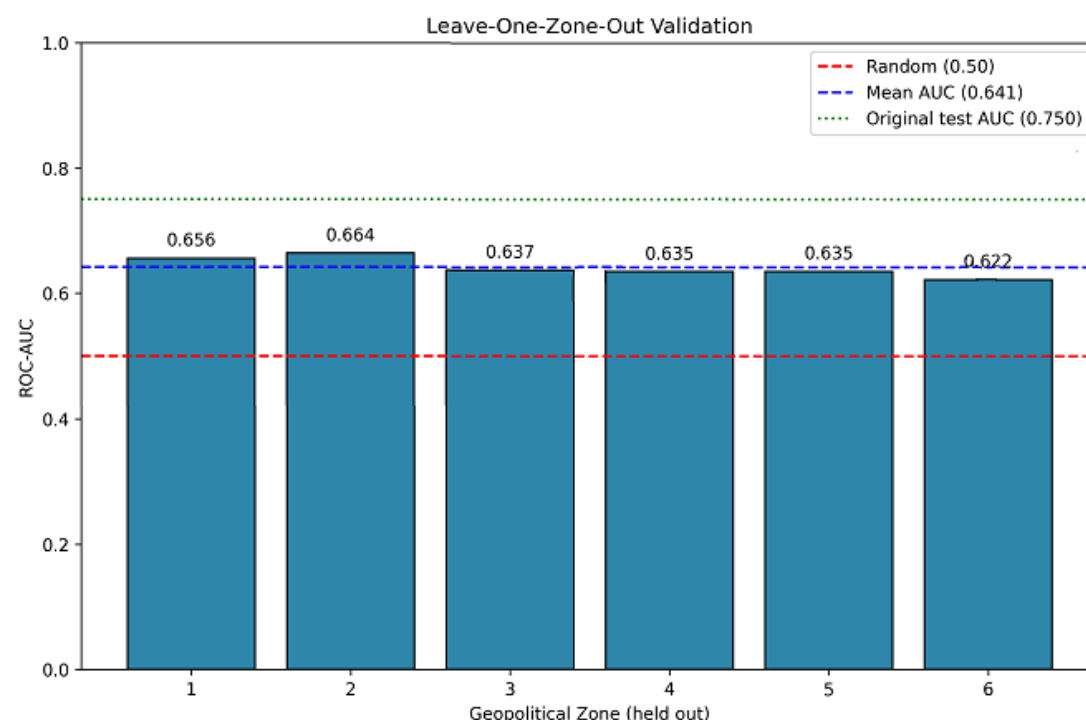


Figure 7: Leave-One-Zone-Out Validation Results by Geopolitical Zone

SHAP-Based Model Explainability

SHAP analysis was applied to the final stacking ensemble to provide both global and local interpretations of model predictions. Figure 8 presents the global SHAP summary, ranking predictors by their mean absolute contribution to the model's output across the entire test set.

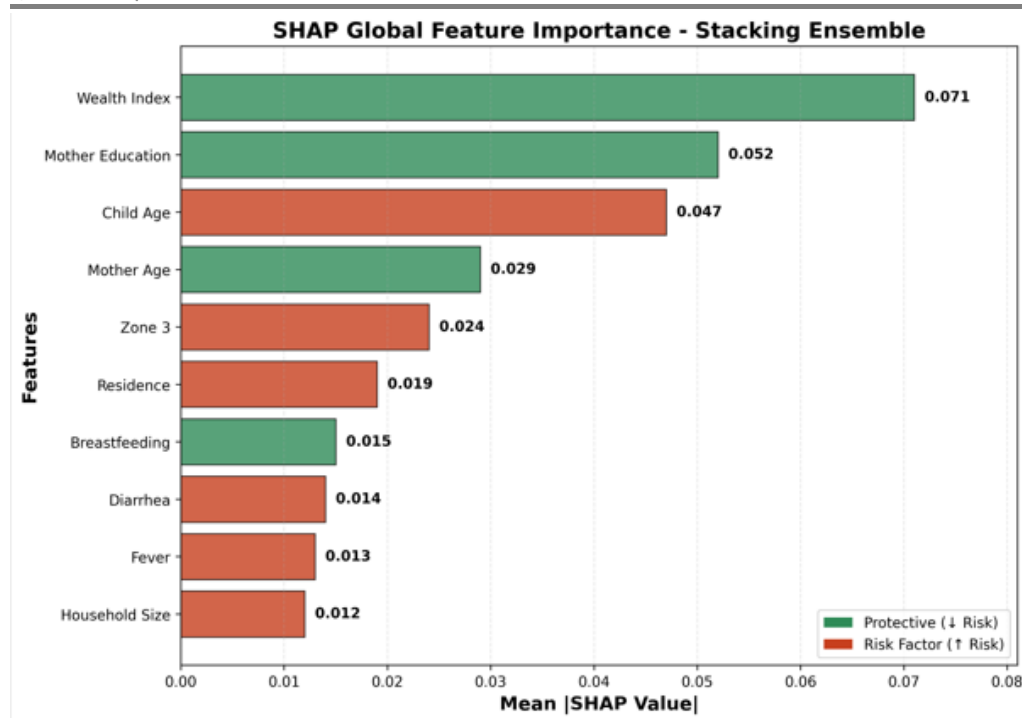


Figure 8: SHAP Global Feature Importance for the Stacking Ensemble Model

Household wealth index emerged as the most influential predictor (mean SHAP = 0.071), with higher wealth quintiles associated with reduced malnutrition risk, reflecting the protective effects of improved food security, sanitation, and healthcare access. Maternal education ranked second (mean SHAP = 0.052), with higher educational attainment showing a strong protective effect, plausibly operating through improved nutritional knowledge, health-seeking behaviour, and household decision-making power. Child age ranked third (mean SHAP = 0.047), with older children showing higher risk, consistent with the cumulative nature of malnutrition as prolonged exposure to inadequate diet and repeated infection compounds over time. Maternal age, geopolitical zone (with Zone 3 showing elevated independent risk), rural residence, breastfeeding status, and recent illness (diarrhoea and fever) rounded out the top ten predictors, summarised in Table 7.

Table 7: Top Ten Predictors of Childhood Malnutrition Risk (Global SHAP)

Rank	Feature	Mean SHAP	Direction
1	Household wealth index	0.071	Protective (higher wealth → lower risk)
2	Maternal education	0.052	Protective (more education → lower risk)
3	Child age (months)	0.047	Risk factor (older child → higher risk)
4	Maternal age	0.029	Protective (older mother → lower risk)
5	Geopolitical zone (Zone 3)	0.024	Risk factor (region-specific)
6	Area of residence (rural)	0.019	Risk factor
7	Breastfeeding status	0.015	Protective
8	Recent diarrhoea	0.014	Risk factor
9	Recent fever	0.013	Risk factor
10	Household size	0.012	Risk factor

To illustrate model interpretability at the individual level, a local SHAP explanation was generated for a representative malnourished child with a predicted risk of 70.5% (Figure 9). Low household wealth (SHAP = +0.061), maternal illiteracy (SHAP = +0.058), residence in the highest-risk geopolitical zone (SHAP = +0.052), and the child's age of 24 months (SHAP = +0.082) combined to drive the elevated prediction, while the mother's young age (19 years, SHAP = +0.015) contributed further. Rather than a single dominant cause, this case illustrates the compound effect of multiple, co-occurring risk factors poverty, low maternal education, geographic disadvantage, and young maternal age consistent with the multidimensional nature of malnutrition documented in the literature.

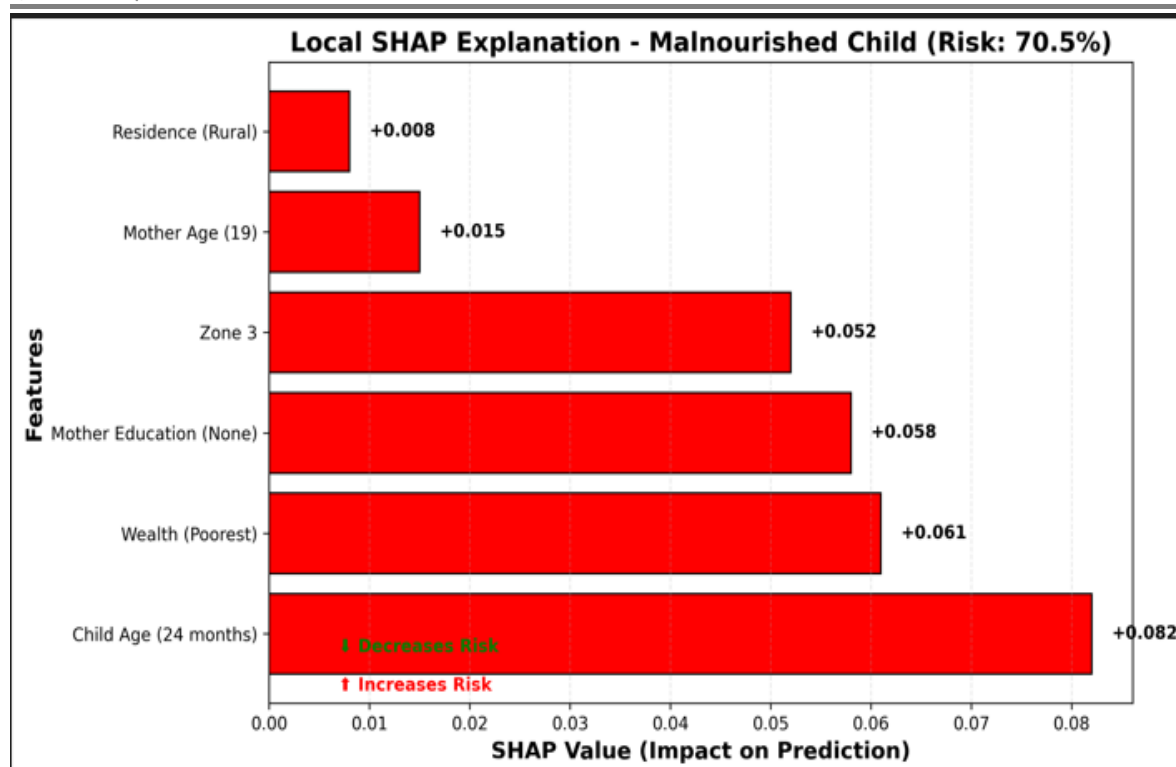


Figure 9: Local SHAP Explanation for a Representative Malnourished Child (Predicted Risk: 70.5%)

Comparative Benchmarking with Prior DHS-Based Studies

Table 8 benchmarks the proposed model against five recently published studies that applied machine learning to DHS derived malnutrition data. Because these studies differ in survey year, country, sample size, and outcome definition (most model a single phenotype, whereas this study models a composite outcome combining stunting, wasting, and underweight), the comparison should not be read as a strict performance ranking. Rather, it highlights methodological safeguards; cluster-aware, mother-level train-test splitting, within-fold SMOTE, and transparent threshold optimisation that are largely absent from the reviewed literature but were implemented in this study.

Table 8: Comparative Benchmarking with Prior Studies Using DHS Data

Study	Country / Survey	n	Outcome	Best Model	Best Performance	XAI	Leakage Safeguards
Mkungudza et al. (2024)	Malawi — MDHS 2015–16	5,149	Stunting	Logistic regression	AUROC 0.64; Sens. 61%	No	None reported
Tamanna et al. (2025)	Bangladesh — BDHS 2022	7,910	Stunting/wasting/underweight (separate)	Random Forest	F1 0.750; Recall 56.3%	Yes — SHAP	SMOTE in CV; cluster split not reported
Sara, Khan & Talukder (2024)	Bangladesh/India/Nepal — BDHS/IDHS/NDHS	8,759/232,920/5,083	Stunting	Random Forest	AUC 66.7–84.3%	No	SMOTE on training set only
Setegn & Dejene (2025)	Ethiopia — EDHS 2011–2019	63,651	Dietary diversity	LightGBM	Accuracy 95.3%; F1 0.953	Yes — Eli5 + LIME	SMOTE on training set
Ngusie et al. (2024)	12 East African countries — DHS	138,426	Micronutrient supplementation	Random Forest	AUC 0.892; Acc. 94.0%	Yes — SHAP	SMOTE compared vs. 3 methods

This study	Nigeria — NDHS 2023–24	9,374	Composite (stunting/wasting /underweight)	Stacking ensemble	ROC-AUC 0.750; F1 0.662; Recall 72.6%	Yes — SHAP (global + local)	Within-fold SMOTE; mother-level cluster-aware split; Youden's J threshold
------------	------------------------	-------	---	-------------------	---------------------------------------	-----------------------------	---

Comparison with Previous Studies

The findings of this study are broadly consistent with previous research demonstrating the superiority of ensemble learning methods for healthcare prediction. As in the studies reviewed in Section II, tree-based and boosting algorithms exhibited stronger predictive performance than simpler classifiers, reflecting their ability to model nonlinear relationships within complex demographic and health survey data. The proposed stacking ensemble further improved predictive performance by integrating multiple heterogeneous classifiers into a unified learning framework, supporting previous findings that stacking exploits the complementary strengths of different algorithms to produce more robust predictions (Islam et al., 2024; Tamanna et al., 2025).

Unlike many earlier malnutrition-prediction studies, which report substantially higher raw accuracy but typically model a single phenotype using smaller, country-specific samples without reporting safeguards against clustering-related data leakage, this study prioritised methodological rigour alongside predictive performance. The SHAP analysis further confirmed the importance of socioeconomic status, maternal education, child age, and geographic location as major determinants of childhood malnutrition, consistent with earlier studies identifying household wealth and maternal characteristics as dominant predictors of nutritional outcomes (Akombi et al., 2017; Ijaiya et al., 2022; Jebero et al., 2023). However, the integration of explainable artificial intelligence within a cluster-aware, heterogeneous stacking framework represents an important methodological advancement over much of the prior literature.

Practical Implications

The proposed explainable stacking ensemble model has important implications for healthcare practice and public health policy in Nigeria. By accurately identifying children at high risk of malnutrition and providing transparent explanations for each prediction, the framework can support community health workers in prioritising vulnerable children for nutritional assessment and referral. Because the model relies exclusively on variables routinely collected during household and community health visits, it is compatible with existing data collection workflows and could be integrated into mobile decision-support tools without requiring additional diagnostic infrastructure.

The dominance of household wealth and maternal education as predictors reinforces the case for addressing malnutrition through cross-sectoral interventions social protection, conditional cash transfers, and investment in girls' education as foundational, long-term strategies, alongside more immediate nutrition-specific programming. The identification of Zone 3 as an independently high-risk region supports the geographic targeting of nutrition resources, while the contributions of breastfeeding, diarrhoea, and fever highlight the value of integrating nutrition promotion with water, sanitation, and hygiene (WASH) and primary healthcare services rather than treating these as separate service domains. Overall, the proposed framework demonstrates that combining heterogeneous ensemble learning with explainable artificial intelligence provides an effective and practical decision-support tool for childhood malnutrition prediction in Nigeria.

CONCLUSION

This study developed and evaluated an explainable stacking ensemble learning framework for predicting childhood malnutrition among children Under-five years of age in Nigeria, using nationally representative data from the 2023–24 NDHS. The proposed framework integrated Random Forest, XGBoost, LightGBM, CatBoost, and Support Vector Machine as heterogeneous base learners with Logistic Regression as the meta-

learner, combined with consensus-based multi-method feature selection, mother-level cluster-aware cross-validation, within-fold SMOTE, Youden's J threshold optimisation, and SHAP-based interpretability.

The stacking ensemble outperformed all individual base learners across the evaluated metrics, achieving ROC-AUC of 0.750, F1-score of 0.662, and a recall of 72.6% at the optimal decision threshold of 0.40, confirming the value of heterogeneous ensemble learning for modelling the complex relationships underlying childhood malnutrition. SHAP-based interpretation identified household wealth index, maternal education, child age, and geographic zone as the leading predictors of malnutrition risk, and demonstrated that model predictions can be explained at both the population and individual child level.

Beyond predictive accuracy, the proposed framework addresses one of the principal limitations of existing artificial intelligence models in healthcare by improving model transparency and interpretability. The resulting tool is transparent, deployable with routinely collected survey variables, and offers a practical basis for the early identification of at-risk children, supporting targeted nutrition interventions and evidence-based policy formulation in Nigeria and other low- and middle-income countries facing similar nutritional challenges.

Limitations

Several limitations should be acknowledged. First, the cross-sectional design limits causal inference. Second, external validation is needed to confirm generalisability beyond Nigeria. Third, while SMOTE was used to address class imbalance, synthetic oversampling may not fully capture the complexity of real-world cases. Fourth, the model relies on structured survey variables and excludes complementary clinical, dietary, and seasonal data that could further improve prediction. Calibration analysis revealed some underconfidence in the predicted probabilities (calibration slope = 1.293), while leave-one-zone-out validation showed reduced performance when the model was applied to unseen geopolitical zones (mean AUC = 0.642 vs. 0.750 in the original random-split evaluation), indicating that geographic generalisability is more limited than the original evaluation suggested. Recalibration or region-specific retraining may therefore be required before deployment in zones not represented in the training data.

Future Research

Future studies should prospectively validate the proposed model in community and clinical settings to determine whether its performance is maintained beyond the survey context in which it was developed. External validation using an independent Nigerian dataset, such as a subsequent wave of the NDHS or another nationally representative Nigerian survey when available, would help assess temporal and population-level generalisability, while validation using DHS datasets from other high-burden countries in sub-Saharan Africa and South Asia could further establish the robustness and transferability of the proposed framework. Given the decline in performance observed during leave-one-zone-out validation, future research should investigate recalibration methods, such as Platt scaling or isotonic regression, as well as zone-specific fine-tuning strategies to improve model performance in regions not represented in the training data. Enriching the model with dietary diversity, micronutrient, and infant feeding practice data, and applying it to longitudinal datasets that follow children from birth through age five, could support the development of models capable of predicting future malnutrition trajectories rather than current nutritional status alone. Future studies should also evaluate whether model performance is consistent across subgroups defined by sex, wealth, geopolitical zone, and maternal education, and should investigate prospective implementation to assess the practical effectiveness of the proposed decision-support system within routine maternal and child healthcare services.

Data Availability

Data for this study were sourced from Demographic and Health surveys (DHS) and are available at <https://www.dhsprogram.com>

Funding

The authors received no specific funding for this research.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this study.

Author Contributions

ZA conceptualized the study, conducted the data analysis, prepared the tables and figures, and drafted the manuscript. BM and AA supervised the study and critically revised the manuscript. AI contributed to the literature review. All authors reviewed and approved the final manuscript

ACKNOWLEDGEMENTS

The authors thank the Demographic and Health Surveys (DHS) Program for making the 2023–24 Nigeria Demographic and Health Survey data available for this analysis.

REFERENCES

1. Akombi, B. J., Agho, K. E., Merom, D., Hall, J. J., & Renzaho, A. M. (2017). Multilevel analysis of factors associated with wasting and underweight among children under-five years in Nigeria. *Nutrients*, 9(1), 44. <https://doi.org/10.3390/nu9010044>
2. Aryuni, M., Kumbangsila, M., Zakiyyah, A. Y., Sano, A. V. D., & Bhatti, F. M. (2023). Comparison of nutritional status prediction models of children under 5 years of age using supervised machine learning. *Proceedings of the 3rd International Conference on Electronics, Biomedical Engineering, and Health Informatics*, 1–10. https://doi.org/10.1007/978-981-99-0248-4_19
3. Biau, G., & Scornet, E. (2016). A random forest guided tour. *HAL*. <https://hal.science/hal-01307105>
4. Breiman, L., & Cutler, A. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
5. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. <https://doi.org/10.1145/2939672.2939785>
6. Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297. <https://doi.org/10.1007/BF00994018>
7. Federal Ministry of Health and Social Welfare of Nigeria (FMoHSW), National Population Commission (NPC) [Nigeria], and ICF. (2025). Nigeria Demographic and Health Survey 2024. FMoHSW, NPC, and ICF.
8. Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine Learning*, 46(1–3), 389–422. <https://doi.org/10.1023/A:1012487302797>
9. Ijaiya, M. A., Anjorin, S., & Uthman, O. A. (2022). Individual and contextual factors associated with childhood malnutrition: A multilevel analysis of the double burden of childhood malnutrition in 27 countries. *Journal of Global Health Science*. <https://doi.org/10.1186/s41256-022-00276-w>
10. 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>
11. Jebero, Z., Moga, F., Gebremichael, B., & Tesfaye, T. (2023). Determinants of acute malnutrition among under-five children in governmental health facilities in Sodo Town, Southern Ethiopia: Unmatched case-control study. *International Journal of Pediatrics*. <https://doi.org/10.1155/2023/3882801>
12. Kawo, M. A., Idris, M. N., Haruna, K., & Kaita, F. U. (2024). Predicting undernutrition risk factors using machine learning techniques in Nigerian under-five children. *International Journal of Computer Science and Mobile Computing*, 13(7), 56–70. <https://doi.org/10.47760/ijcsmc.2024.v13i07.006>
13. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, 30, 3146–3154.

14. Kishore, K. K., Suman, J. V., Mnikyamba, L. I., Polamuri, S. R., & Venkatesh, B. (2023). Prediction of malnutrition in newborn infants using machine learning techniques. Research Square. <https://doi.org/10.21203/rs.3.rs-2958834/v1>
15. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 4765–4774.
16. Mahmud, T., Wara, T. U., & Joy, C. D. (2024). Risk factor identification and classification of malnutrition among under-five children in Bangladesh: Machine learning and statistical approach. arXiv. <https://arxiv.org/abs/2412.05813>
17. Nirmani, H. M. C., & Kudagamage, U. P. (2024). Ensemble approach for early prediction of malnutrition level of children: A case study on children Under-five years old. *Proceedings of the 2024 Moratuwa Engineering Research Conference (MERCon)*.
18. Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: Unbiased boosting with categorical features. *Advances in Neural Information Processing Systems*, 31, 6639–6649.
19. Qasrawi, R., Sgahir, S., Nemer, M., Halaikah, M., Badrasawi, M., Amro, M., Vicuna Polo, S., AbuAl-Halawa, D., Mujahed, D., & Nasreddine, L. (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>
20. Reusken, M., Coffey, C., Cruijssen, F., et al. (2024). Identification of factors associated with acute malnutrition in children under 5 years and forecasting future prevalence: Assessing the potential of statistical and machine learning methods. *BMJ Public Health*, 3, e001460. <https://doi.org/10.1136/bmjph-2024-001460>
21. Setegn, G. M., & Dejene, B. E. (2025). Improving machine learning models through explainable AI for predicting the level of dietary diversity among Ethiopian preschool children. *Italian Journal of Pediatrics*, 51(1), 45. <https://doi.org/10.1186/s13052-025-01345-9>
22. Shahid, M., Yahya, M. A., Song, J., Naveed, H. M., Yuksel, S., Dincer, H., & Ali, M. (2025). Machine learning vs. traditional logistic regression: Predictive performance and risk factor identification for child nutritional outcome in Pakistan. *BMC Public Health*. <https://doi.org/10.1186/s12889-025-25621-9>
23. Tamanna, T., Mahmud, S., Salma, N., Hossain, M. M., & Karim, M. R. (2025). Identifying determinants of malnutrition in under-five children in Bangladesh: Insights from the BDHS-2022 cross-sectional study. *Scientific Reports*, 15, 12345. <https://doi.org/10.1038/s41598-025-99288-y>
24. Tesfaw, L. M., & Fenta, H. M. (2021). Multivariate logistic regression analysis on the association between anthropometric indicators of under-five children in Nigeria: NDHS 2018. *BMC Pediatrics*, 21, Article 194. <https://doi.org/10.1186/s12887-021-02657-5>
25. Tzitiridou-Chatzopoulou, M., Zournatzidou, G., Orovou, E., Lithoxopoulou, M., Drogouti, E., Sklavos, G., Antoniou, E., & Tsakalidis, C. (2024). Evaluating malnutrition practices and mother's education on children failure-to-thrive symptoms using entropy-weight and TOPSIS method. *Children*, 11(8), 903. <https://doi.org/10.3390/children11080903>
26. UNICEF, WHO, & World Bank Group. (2023). Levels and trends in child malnutrition: Key findings of the 2023 edition. World Health Organization. <https://www.who.int/>
27. Vergara, J. R., & Estévez, P. A. (2014). A review of feature selection methods based on mutual information. *Neural Computing and Applications*, 24(1), 175–186. <https://doi.org/10.1007/s00521-013-1368-0>
28. Wajgi, R., & Wajgi, D. (2022). Malnutrition detection in infants using machine learning approach. *AIP Conference Proceedings*. <https://doi.org/10.1063/5.0076876>
29. World Health Organization. (2006). WHO child growth standards: Length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age: Methods and development. WHO.
30. World Health Organization. (2023). Levels and trends in child malnutrition. <https://www.who.int/>
31. Yanto, M., Hadi, F., & Arlis, S. (2024). Determination of children's nutritional status with machine learning classification analysis approach. *Indonesian Journal of Electrical Engineering and Computer Science*, 33(1), 303–313. <https://doi.org/10.11591/ijeecs.v33.i1.pp303-313>