

Prognosticating Risk Factors of Undernutrition in Under Five Years Children in Nigeria Using Ensemble Technique

*¹Mohammed Ali Kawo ²Muhammad Nuruddeen Idris ³Kabiru Haruna ⁴Firdausi Umar Kaita

¹Department of Computer Science, Federal University Gusau, Zamfara, Nigeria

²Department of Computer Science, Federal University Gusau, Zamfara, Nigeria

³Department of Computer Science, Federal University Gusau, Zamfara, Nigeria

⁴Department of Computer Science, Federal University Gusau, Zamfara, Nigeria

Corresponding Author Email Address: ali_mak2001@yahoo.com, mohammedali@fugusau.edu.ng

Received on: October, 2025

Revised and Accepted on: November, 2025

Published on: December, 2025

ABSTRACT

Insufficient nutritional intake and nutrient imbalances is referred to as malnutrition, which signifies a universal health matter that has a substantial consequence on children, leading to short-term and long-term consequences for their growth, development and general comfort. This research utilized the potentials of stack ensemble techniques in enhancing the prediction of Child undernutrition in Nigeria. The study exploits three key indicators such as: underweight, stunting and wasting from Nigerian Multiple Indicator Cluster Survey (MICS) 2021 dataset. Four different machine learning classifiers were employed namely: decision tree (DT), random forest (RF), k-nearest neighbors (KNN) and the stack ensemble technique meta classifier that is concatenated with logistic regression (LR) to form stack ensemble technique in other to improve the prognostic risk factors of children under the age of five, considering their nutritional status. Based on their ability to predict outcomes, these classifiers were assessed and contrasted using standard machine learning evaluation metrics like. Accuracy, precision, recall and F1-score respectively. The effectiveness of machine learning techniques, specifically the Stack Ensemble technique, in prognosticating and comprehending the factors that donate undernutrition in Nigerian children is demonstrated by this study. The results offer important information for developing policies and focused actions.

Keywords: Stack Ensemble, imbalance, Machine learning, Evaluation metrics, Undernutrition.

1.0 INTRODUCTION

A high rate of mortality and overall burden of disease are caused by undernutrition, which is a serious global public health concern (Rice *et al.*, 2000). It affects children under five and is particularly common in low- and middle-income countries. Despite a decline in global rates, undernutrition among children remains a major problem in sub-Saharan Africa (Black *et al.*, 2013). Every year, almost 12 million children under the age of five pass away; the majority of these youngsters live in developing countries (WHO, 2020). More than half of these deaths are caused by acute respiratory infections, malaria, measles, and diarrhea; each of these ailments can be avoided or managed with low-cost interventions.

Despite its prevalence in developing nations, malnutrition rarely ranks among the leading causes of death (Steyn *et al.*, 2020). A portion of this can be attributed to the traditional reporting and interpretation of death cause data. In a number of countries, mortality statistics are based on records that only specify one primary cause of demise. According to estimates from the World Health Organization, there were 1.9 billion overweight individuals and 462 million underweight persons worldwide. Furthermore, it was found that 38.3 million children were overweight or obese, 144 million children

were stunted, and 47 million children under the age of five were wasted, with 14.3 million children severely wasted (Indiran *et al.*, 2021). Every year, 2.6 million children worldwide lose their lives due to malnutrition, with undernutrition being the cause of 45% of these fatalities.

A child's long-term survival and health are negatively impacted by starvation during childhood in several ways. (Survey *et al.*, 2021). In the latest indication of Demographic and Health Survey (DHS), children under the age of five (Report, 2019), 37% of them are stunted, meaning they are low in height for their age, and 7% are severely wasted or undernourished, which means they are underweight in relation to their height. Policymakers in Nigeria ought to be quite concerned about these implications of malnutrition.

As of 2018, Nigeria had the largest population in Africa, with about 196 million people living there (Keeley *et al.*, 2019). Population density is rising as a result of the high birth rate of 5.4 children per woman and the population's 2.6 percent yearly growth rate. By 2050, Nigeria is projected to have an incredible 401 million people living in it, making it the third most populous nation in the world, after India and China. Nigeria's urban population

is expanding at a pace of 4.7%, which is among the fastest in the world (Keeley *et al.*, 2019). In relations to the improvement made toward reaching the Sustainable Development Goals, Nigeria is now placed 159th out of 162 countries, according to the SDG (2019) report.

In the domains of epidemiology and public health, child malnutrition is a serious global concern. Malnutrition in children less than five have been the focus of numerous research projects. Only the identification of dietary risk variables using conventional models, like logistic regression, was the focus of previous research (Quang Dat *et al.*, 2018; Rahman & Rahman, 2019).

Consequently, a robust predictive model based on the important risk characteristics identified has to be provided in order to forecast malnutrition. Currently, the capacity of machine learning (ML) to forecast various types of biological and medical data has created a lot of attention. In recent years, machine learning has become more extensively used in public health. Researchers utilized specific machine learning studies to predict various domains, including Undernutrition (Anku & Duah, 2024), neonatal mortality Agu *et al.*, 2019), diabetes (Gayawan *et al.*, 2020), low birth weight (Ayogu *et al.*, 2018), and so forth. Additionally, some research was conducted on machine learning with different classifiers for the prediction of underweight, stunted, and wasting (Bitew *et al.*, 2021). In the meantime, numerous studies have looked into the regional variations and contributing factors to undernutrition in children under five using the conventional analytical technique (Gayawan *et al.*, 2020). Most of these studies were limited to certain areas that do not accurately represent the entire country, for instance, rural areas found in the Northern and Southern regions (Ayogu *et al.*, 2018). The few studies Stunting was the main focus of the few research that show how undernutrition varies geographically among children in Nigeria; other signs of undernutrition, like underweight and wasting, were mainly ignored (Agu *et al.*, 2019; Ayogu *et al.*, 2018; Nadine *et al.*, 2019).

Throughout the last 20 years, the field of machine learning has developed tremendously from a lab curiosity to a useful technology with broad commercial use (Kourou *et al.*, 2015). The development of helpful software for speech recognition and computer vision, diagnostic systems, applications such as robot control and natural language processing has made machine learning the go to approach in artificial intelligence (AI) (Jordan & Mitchell, 2020).

In respect to the findings, a machine learning technique called stacked ensemble method combines several models, or base learners, to enhance overall predictive performance. A higher-level model (meta-learner) that generates the final prediction uses the predictions of the base models, which have been trained on the data, as input features. Stacking was first conceptualized as training several base classifiers with a single first training set, and then having each base classifier create a new training set in order to train the meta-classifier (Feng *et al.*, 2023).

An ensemble machine learning technique is a high-powered technique in which classifying new examples involves combining individual decisions in some way, usually by weighted or unweighted voting (Dietterich, 2000). Research on how to build an effective ensembles classifier has been one of the busiest areas of supervised learning research. The primary finding is that ensembles always reveal significant higher accuracy compared to individual classifiers.

In terms of resolving issues with classification and regression with an enhanced reduction on bias and variance to boost the accuracy of models, ensemble techniques have proven to have greater predictive potential than individual base learners (Odegua, 2020). A wide range of machine learning techniques have been applied in medical research (Kino *et al.*, 2021), (Chandrashekar & Sahin, 2014), (Theses, 2023). To predict the state of illnesses like malaria and cancer, for instance, machine learning techniques like artificial neural networks, support vector machines, and random forests have been used (Theobald, 2017). In distinct to other machine learning methods like linear discriminant analysis (LRA), logistic regression (LR), K-Nearest Neighbors (k-NN), and support vector machines (SVM), the Random Forest (RF) approach performed better.

Additionally, a study conducted in Nigeria that employed Bayesian Additive Regression Trees shown that maternal education defined a ten years of schooling or more reduces the risk of severe undernutrition in children (Nadine *et al.*, 2019). There is more opportunity for research in this field, though, as a scoping analysis (Kino *et al.*, 2021) discovered that machine learning techniques were only applied in a small portion of the massive number of social factors of health studies published each year, which makes it possible to carry out additional research on this. As a result, we used ensemble machine learning techniques in this research which is not commonly used in previous researches to prognosticate the factors leading to child undernutrition in Nigeria. There isn't much information that connects ensemble

techniques to undernutrition in Nigerian children under the age of five, as far as we are aware.

The aim of this feasibility study was to showcase the effectiveness of ensemble techniques in enhancing the prediction of undernutrition in Nigerian children under five, while also identifying associated risk factors.

2.0 METHODOLOGY

2.1 Sources of Dataset

In order to execute the suggested ensemble method, a set of data from the Nigeria Multiple Indicator Cluster Survey (MICS6) were used in this investigation. There was a MICS survey conducted in 2021, where the sample for children under the age of five was 31,103-point data. For the purpose of data collection, Nigeria was partitioned into six distinct geopolitical zones. The most important sampling units in each region were the rural and urban areas. Following that, a two-stage sampling procedure was utilized in order to select the families that would carry out the interviews. The estimations were subsequently employed to assess the state of women and children in terms of their safety, education, and health. Furthermore, it serves as a mechanism for generating statistics to monitor the progress of both local and global objectives aimed at enhancing the welfare of women and children, including the Sustainable Development Goals (SDGs).

Moreover, MICS serves as a crucial repository of data regarding the progress and improvement of the welfare of men, women, and children in Nigeria, while also playing a vital role in the advancement of statistical frameworks.

2.2 Study Variables and Measurements

The primary research goals of the study are the under-five underweight, stunting, and wasting status. The Z-scores of three anthropometric measurements were used to determine the nutritional status as follows: A kid is deemed stunted if their height-for-age Z-score (HAZ) is less than two standard deviations ($-2SD$) from the median, and wasting if their weight-for-age Z-score is fewer than two standard deviations ($-2SD$) from the median. If a child's weight-for-height Z-score is less than or equal to two standard deviations ($-2SD$) from the median, they are deemed underweight.

All three undernutrition indices were included in this study to predict the factors that lead to undernutrition in children. Binary coding was used in this regard, with 1 denoting that the results were either stunted, wasted, or underweight, and 0 denoting that none of these conditions were met. Numerous factors were examined in Nigeria as possible risk factors for children's undernutrition. We reduced the amount of missing data in the machine learning methods while incorporating as many MICS

elements as possible. Age, sex, geopolitical zone, residential area, length of the child, mother's age at birth, weight of the child, feverish or diarrheal child, health insurance, mother's educational attainment, ethnicity, breastfeeding, prenatal care, combined wealth score, and primary source of drinking water are some of these factors. The elements that affected the algorithms' performance were essentially included in the analysis, while those with more than 50% missing data were excluded.

2.3 Analytical Evaluation

We used Jupyter notebook (version 6.5.4 IDE) from Python distributor known as Anaconda to handle and analyze the dataset, we also used the sklearn package for importing most of the machine learning and ensemble libraries for data scrubbing. Furthermore, we used four well known machine learning (ML) algorithms such Decision Tree (DT), Random Forest (RF), K-Nearest Neighbors (KNN), and logistic regression as the meta classifier to prognosticate the factors that contribute to malnutrition in under five children in Nigeria. We then contrasted each algorithm's results using evaluation measures.

2.4 Feature Selection

A critical preprocessing step in machine learning is the feature selection (FS). Eliminating superfluous features, reducing calculation time, and improving prediction model performance are all made easy with FS. Controlling the dataset's noise and outliers while maintaining consistency are the fundamental objectives of feature selection. We use the filter method's chi-square to evaluate feature importance in relation to its prediction potential. In filter approaches, instead of employing learning models, features are selected in light of the data's features. This method is efficient and quick. Two steps are usually involved in a filter technique (Chandrashekar & Sahin, 2014). Step one involves ranking features based on a set of criteria. In step two, the features that have the highest rankings are chosen. The characteristics of the features that includes their addictions on the class label, correlation with each other, and dependence on each other, are measured by these criteria. The four machine learning (ML) classifiers have been used to forecast malnutrition in children by using key risk factors.

2.5 Synopsis of Ensemble Techniques

The art of ensemble learning involves bringing together a variety of learners, or individual classifiers to increase the model's stability and prediction capacity (Dietterich, 2000). Several learners are trained to tackle the same problem using the ensemble technique of machine learning. In standard machine learning techniques, one

hypothesis is learned using training data. In contrast, an ensemble technique builds a set of hypotheses from two or more algorithms and learning is done on the resulting set (Theses, 2023).

According to Woods *et al.* (1997), the original concept behind stacking was to use an initial training set to train several base classifiers, each of which would then provide a new training set to train the meta-classifier. The stacking schematic diagram is displayed in Figure 1. The stacking algorithm's first layer has a variety of base classifier classes and the correlation between different base classifier algorithms is low. The meta-classifier is linear regression (LR) algorithm. Cross-validation methods were used to train and test the samples. The training set is used for hyper-parameter tuning based on the 5-fold cross-validation. The training set is cross-validated where the base classifier is trained and predicted. The prediction result of each base classifier is used as a feature of the meta-classifier training data, and the label of the new sample remains the same as the original one. After training all the base classifiers, a new data set is generated for the training of our meta-classifier. Similarly, the test set serves directly as input to each trained base classifier to get the prediction and the predicted results is used as a feature of the new sample for the meta-classifier. The 5-fold cross validation technique is used in other to avoid some randomness and bias due to sample date.

meta-classifier is part of the second module. While the third module is dedicated to data feature processing, which carries out new feature for final prediction.

3.0 Selection of individual classifier and meta-classifier

The stacking approach is a methodology that combines machine learning algorithms with theoretical understanding of statistics to improve prediction accuracy by combining many algorithms and by utilizing their benefits. For the construction of the stacking ensemble model, the first step is to find the individual base classifiers with respect to accuracy and diversity in terms of performance. In this research, an excellent prediction models with vigorous performance were selected to obtain the best ensemble model. By doing so, our proposed model is able to explore the intrinsic connections existing between data from different perspectives and structural perceptions, and then combine their respective algorithmic values to build predictive model. Therefore, choosing a less correlation classifiers can take the advantages of different machine algorithms and achieve worse in term of classification. In view of these, three trending classifiers was utilized in this research such as Decision Tree (DT), K-Nearest Neighbors (KNN) and Random Forest (RF) as base learner and Logistic Regression (LR) as Meta-Stack ensemble technique to discover the features in Nigerian under five children dataset and predict nutritional indicators.

A synopsis of the three machine learning classifiers that was used to trained each nutritional indicator is given below: k-Nearest Neighbors (KNN), Decision Tree (DT), Random Forest (RF), and Stack Ensemble Technique (STE). For this research, the classifiers were trained to produce features that prognosticate nutritional indicators in Nigerian under five children. The datasets were cross validated using K-fold cross validation techniques respectively.

3.1 Overview of Machine Learning System

With the invention of machine learning algorithms a subset of artificial intelligent that mimic the human intelligence was described by (Theobald, 2017) as the ability to learn without being explicitly programmed. According to (Theobald, 2017) machine learning entails three simple process.

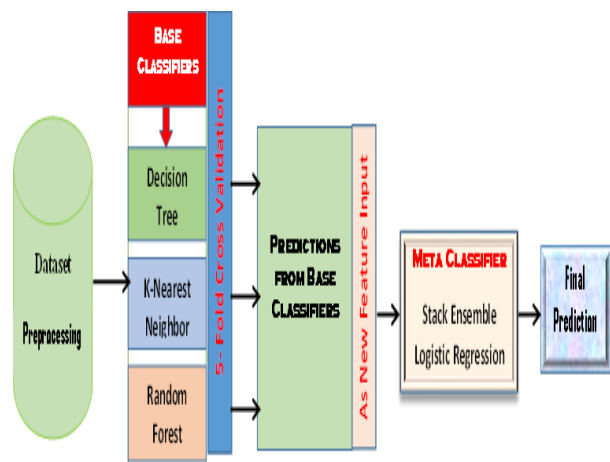


Fig.1: Stack ensemble techniques

Based on the research work, we propose stack ensemble machine learning technique for prognosticating risk factors of undernutrition in less than five years children in Nigeria in order to further improve the prediction accuracy. In this technique, there are three modules. The first module is the classifier integration module, in which an integrated model is built on the basis of the stacking ensemble model. The selection of the base classifier and

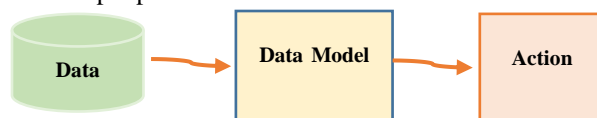


Fig.1: A Simple machine learning process



Machine learning integrates a lot of learning strategies which including supervised, semi-supervised and unsupervised learning. The labeled classes of each learning strategy differentiate its techniques (Osisanwo *et al.*, 2017). Classification and regression are the two widely used forms of machine learning. In supervised learning the input labels are provided (Jijo & Abdulazeez, 2021). Unlike the unsupervised learning where the input labels are not given.

3.1.1 Random Forest (RF)

Random forest (RF), a supervised machine learning technique, classifies data by utilizing the insights from many decision trees. Random forests fit several uncorrelated classification trees and utilize the average of all individual trees to assure a higher prediction accuracy than individual trees, in contrast to decision trees that use only one tree to make predictions. Random Forest chooses a random sample (with replacement) of the training dataset for each decision tree in the RF. This procedure is referred to as bagging or bootstrap aggregation. Moreover, node splitting in the RF model is based on a random set of attributes for each tree. Consequently, the traits of one tree may be different from those of another. The bagging process and variations in feature selection for each tree have an impact on the durability of RF forecast accuracy. (Chaudhary *et al.*, 2016)

3.1.2 K-Nearest Neighbors (KNN)

The KNN method examines the nearest neighbors that must to be scrutinized in order to elucidate the class of a sample data by measuring the nearest neighbors with respect to a value of k (Hassanat *et al.*, 2014). The whole dataset is separated into sample data points and training data in the technique. The distance between each sample point and every training point is computed, and the point with the least distance is referred to as the nearest neighbors (Soofi & Awan, 2017).

3.1.3 Decision Tree (DT)

The Decision Tree (DT) consists of a set of trees-structured decision tests working in a divide-and conquer mode (Zhou, 2025). Decision tree algorithm is generally recursive in nature. It can equally predict on any kind of data, be it text, graphics, continuous data or categorical variable data; they can all be used as independent variables or predictors. Decision Tree (DT) generates a regression tree on the output variable that is continuous and a classification tree for the output variable is distinct or categorical.

3.1.4 Logistic Regression (LR)

The Logistic Regression (LR) (Ghosh *et al.*, 2020) falls into the category of linear model, therefore it has a lot of similarities with multiple linear regression analysis. It is a type of machine learning algorithm used for classification to predict categorical or discrete outcome. With two results: yes/no, 1/0 or true or false. Usually use in calculating mathematical likelihood of an events fitting into a certain class or not (Nusinovici *et al.*, 2020).

3.2 Performance Evaluation

Confusion matrix is used to evaluate the model's accuracy on the dataset. Table 1 below depict the possible outcomes that will be expected from any binary classification predictions (Flach, 2007). For assessing the effectiveness of all machine learning models and algorithms the confusion matrix is a highly used matrix.

Table 1: Confusion Matrix

Confusion Matrix		Predicted Classes	
		Negative	Positive
Actual Classes	Negative	TN	FP
	Positive	FN	TP

Accuracy

$$= \frac{\text{(number of accurate predictions)}}{\text{(total number of predictions)}} \quad (1.1)$$

$$\text{Precision} = \frac{\text{(true positive)}}{\text{(true positive)} + \text{(false positive)}} \quad (1.2)$$

Recall

$$= \frac{\text{(true positive)}}{\text{(true positive)} + \text{(false negative)}} \quad (1.3)$$

$$\text{F1 score} = 2 \times \frac{\text{(precision} \times \text{recall)}}{\text{precision} + \text{recall}} \quad (1.4)$$

4.0 Results and Discussion

4.1 Predictive Algorithms for Child Undernutrition Indicators

Underweight

The Stack ensemble techniques outperformed all other classifiers in the underweight predictive aspect with a significant accuracy score of 81%, a precision score of 83%, recall score of 96%, and F1-score of 89% respectively. Table 2 below displays the confusion matrix of the scores while figure 2 below also depict the bar chart representation of the result for visualization.

Table 2: Confusion Matrix for Underweight prediction on malnutrition dataset

Classifiers	Accuracy	Precision	Recall	F1-score
Decision Tree	73%	84%	82%	83%
Random Forest	80%	82%	95%	89%
K-N-Neighbors	78%	82%	93%	87%
Stack Ensemble	81%	83%	96%	89%

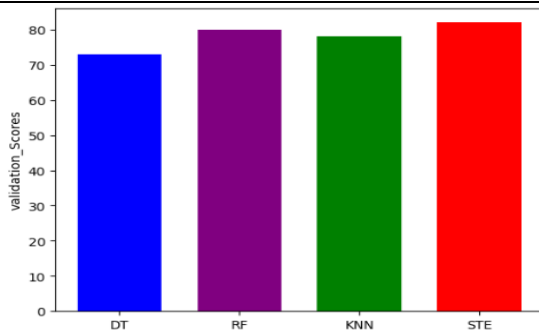


Figure 2: Bar chart for Underweight prediction on malnutrition Dataset

4.1.2 Predictive Algorithms for Child Undernutrition Indicators

Stunting

The consequence of insufficient nourishment during pregnancy and initial life leads to stunting. Offspring with stunt may never fully grow to their full stature and their brains may remain undeveloped. These offspring begin their life with a clear drawback that influence their adulthood. They struggle while mingling with their peers, financially and even academically. They even encounter difficulties in relating with in their societies. The Stack ensemble techniques also performed tremendously in accuracy score of under-fives stunting prediction with a significant score of 92%. However, Random Forest and K-N-Neighbors model came next in term of accuracy score of about 89% respectively. Table 3 below, illustrates the clear confusion matrix that encompasses the precision, recall and F1-score of each classifier. And figure 3 below also shows the bar chart of the stunt predicted result.

Table 3: Confusion Matrix for Stunt prediction on malnutrition dataset

Classifiers	Accuracy	Precision	Recall	F1-score
Decision Tree	81.33%	90%	88%	89%
Random Forest	89%	90%	99%	94%
K-N-Neighbors	89.24%	90%	98%	94%
Stack Ensemble	92%	93%	99%	96%

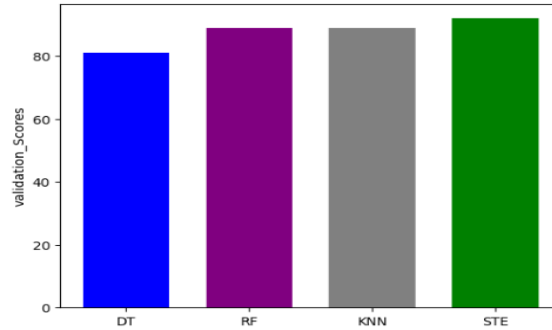


Figure 3: Bar chart for Stunted prediction on malnutrition Dataset

4.1.3 Predictive Algorithms for Child Undernutrition Indicators

Wasting

The hypothetically incurable outcome of insufficient dietary consumption in a child is known as wasting. Children that are in any way involved in wasting are more exposed to lasting delay in developing, with conceded immune deficiency. When the wasting becomes too peril, the offsprings are likely to die. In other to rescue a child with severe wasting, early treatment and proper caring must be held in time. Notwithstanding, the Stack ensemble technique has proven to be an enhanced model for machine learning prediction, with an accuracy score of 90% for wasting in under-five children. Table 4 below displays the confusion matrix for clear indication. While figure 4 below depict the bar chart for clearer visualization of the result.

Table 4: Confusion Matrix for Wasting prediction on malnutrition dataset

Classifiers	Accuracy	Precision	Recall	F1-score
Decision Tree	77%	88%	85%	87%
Random Forest	86%	87%	98%	92%
K-N-Neighbors	84.51%	87%	96%	91%
Stack Ensemble	90%	94%	96%	97%

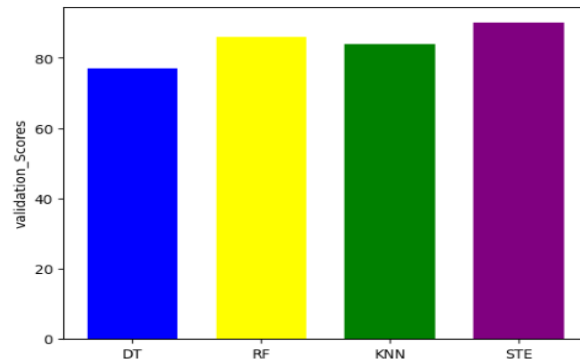


Figure 3: Bar chart for Wasted prediction on malnutrition Dataset

4.2 Results

4.2.1 Summative Analysis of Results

To ascertain the research authenticity the analysis of the result is foreseen, where 31,103 children under the age of five was used as the study sample, the stunted detected is 36.63% while 19.20% were designated as underweight and 6.10% were wasted. Figure 5 below, displays a pie chart for visualizing the information in a clearer manner. This indicates that 61.93% are malnutritional in nature and only 38.07% are estimated to be normal out of the children under the age of five. Figure 5 below, displays a pie chart that illustrate a clear visualized format for elucidation.

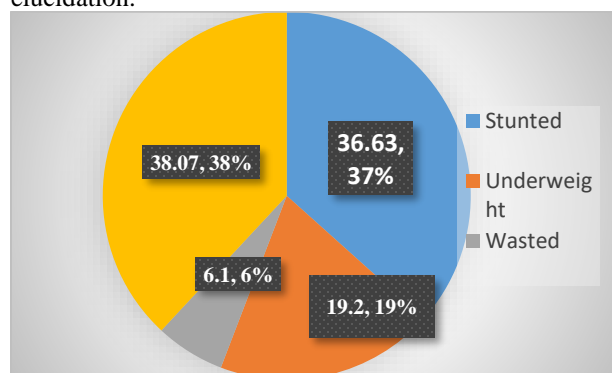


Figure 5: Pie chart for Analysis of malnutrition Dataset Result

4.2.2 The significant Factors that Influence Malnutrition in Children

With the accuracy results obtained from the confusion matrixes above, the summary in table 2, 3 and 4 respectively indicate the outstanding performance of Stack ensemble technique by surpassing all other classifiers in all the malnutrition factors such as underweight, stunting and wasting in term of accuracy score that was predicted. The consequence of the evaluated result from each classifier hinge on variable related to perils such as residential area, breastfeeding, source of water for drinking, family status and the educational background of the parents. All these are substantial peril that was added to the model during prediction of the three factors.

4.3 Discussion

The dataset retrieved for this research is from Nigeria MICS 2021 that contains 31,103 children of under-five ages. Where 61.93% of the children indicate risk of malnutrition be it either underweight, stunting or wasting that was instigated from the expressive predicted results. The

The findings show that all three nutritional indicators exhibit a notable geographical variation in Nigeria. The northwest region has the greatest rate of stunting

prevalence such as Kano, Zamfara, Kebbi, Kaduna, Niger, Bauchi and Jigawa and those residing in the northeast like Adamawa, Yobe and Borno. The south-east regions of Osun and Anambra have the lowest frequency. Once more, the Northwest has the highest prevalence of wasting, while the Southeast has the lowest prevalence. Furthermore, children that are underweight are more prevalent in the Northwest than in the Southwest. There is evidence of similar regional factors in stunting, wasting, and underweight in Nigeria (Ayogu *et al.*, 2018). The training dataset was used on machine learning algorithms to develop a bivariate analysis that demonstrated a significant performance. Food insecurity has a significant effect on wasting, as this study has demonstrated, and stunting is significantly impacted by the frequency and variety of meals that children eat (Motbainor, 2015). In regions such as Jigawa, Kano, Zamfara, Niger, Kaduna, Bauchi, Katsina, Sokoto, Yobe, Adamawa, and Borno, food insecurity is prevalent. Research has also revealed a decline in adult caloric intake in some of these regions in recent years (Olutosin, 2023). Additionally, it has been demonstrated that in these regions, where insurgency and banditry are more common, meal frequency is recurrently decreased, which results in extensive urban migration. These significant regional variations in nutrition indices have a substantial effect on the nutritional status of the nation's population of children under five (Nwozor & Olanrewaju, 2020). When comparing the predictive algorithm to Random Forest (RF), Decision Tree (DT), and K-Nearest Neighbors (KNN), the Stack Ensemble approach has the highest predicting accuracy for all malnutrition outcomes. In order to address health issues including childhood mortality (Bizzego *et al.*, 2021) and childhood asthma (Sejal & Adnan, 2019), a variety of machine learning algorithms have been applied, and they have demonstrated the ability to generate precise and trustworthy predictions. Each of the undernutrition factors has distinct underlying variables. The findings of this study are corroborated by the ample literature (Martins *et al.*, 2011), (Pasricha *et al.*, 2010) regarding the importance of a child's age in predicting their state of undernutrition. This suggests that birth weight may also have a substantial role in predicting a child's nutritional status, as small babies are far more likely to experience undernutrition (Austin & Ntenda, 2019).

The findings of the research affect how well Stack Ensemble technique of machine learning work in population health research. Comparably, several studies have confirmed the usefulness of machine learning (ML) for population health research and policy decision making in a number of areas, including defining cancer disease treatment (Kourou *et al.*, 2015), (Bardhi & Zapira, 2021),

under-five child mortality (Bitew *et al.*, 2020), (Demsash, 2023), undernutrition (Anku & Duah, 2024), (Bitew *et al.*, 2021), women's health (Brand *et al.*, 2024), and malnutrition among women (Turjo & Rahman, 2024). The increasing application of machine learning to forecast population health outcomes is highlighted by this (Emily *et al.*, 2021), (Ahmed *et al.*, 2021).

4.4 Essential Results and Inputs

This research offers important insights on childhood undernutrition in Nigeria, presenting a number of important conclusions and advances in the field: In Nigeria, the frequency of underweight, stunting, and wasting among children under five varies significantly by region, according to our data. These regional differences emphasize the necessity of focused interventions that are adapted to certain areas. Second, we showed how well the Stack Ensemble method predicts childhood undernutrition when compared to other machine learning methods like Decision Tree (DT), Random Forest (RF), and K-Nearest Neighbors (KNN). This discovery adds to the increasing amount of proof that modern analytical techniques are beneficial when used in public health studies. Key factors that contribute to undernutrition were also discovered by our research, including the mother's educational attainment, the residential area, the child's age, the birth size, the household wealth index, and the source of drinking water. These characteristics are crucial because they are consistent across the three undernutrition indicators (underweight, stunting, and wasting). Additionally, we offered a thorough analysis based on a sizable, representative sample of 31,103 children by using the data from the Nigerian Multiple Indicator Cluster Survey (MICS). Our findings are more dependable and broadly applicable thanks to this large dataset. Our research advances the methodology for studying complicated public health concerns by showcasing the successful use of the Stack Ensemble technique of machine learning to assess and forecast childhood undernutrition in a developing country environment.

4.5 Consequences and Suggestions

We suggest the following conclusions and suggestions in light of our findings:

- i. Economic Interventions: Considering the significant impact of the household wealth index on the nutrition of children, we advise creating focused initiatives to reduce poverty. These could include skill-building programs, microfinance projects, and conditional cash transfers with the goal of raising household economic standing.
- ii. Drinking Water Sources: Given the substantial influence that water sources have on undernutrition,

there is an urgent need to increase access to clean water. We advise giving water infrastructure projects first priority, especially in underserved and rural areas. Programs to guarantee access to potable water and encourage good sanitation habits should be put in place by policymakers.

iii. Healthcare Access and Early Childhood Interventions: Promoting early childhood nutrition while strengthening healthcare infrastructure is what we advise. The implementation of comprehensive growth monitoring programs, the provision of nutritional supplements, and training caregivers on appropriate feeding techniques are all part of this, particularly for children under 32 months who demonstrated heightened susceptibility.

iv. Maternal Health and Education: Our results demonstrate how important maternal factors are to the nutrition of children. Strengthening maternal health services, such as prenatal care and nutritional assistance for expectant and nursing mothers, is something we advise. Additionally, spending money on women's education could have a big impact on the health of children, especially in areas with low female literacy rates.

v. Interdisciplinary Collaboration: Encourage collaborations among legislators, nutritionists, data scientists, and public health specialists to create all-encompassing, evidence-based approaches to combating undernutrition in children.

vi. Data-Driven Policy Formulation: We highly advise incorporating data analytics and machine learning into the creation of public health policies. Data scientists and policymakers should work together to use predictive model insights to make better, evidence-based decisions.

vii. Ongoing Monitoring and Evaluation: Set up procedures for continuing data gathering and analysis so that you can monitor results and modify interventions as necessary. This will guarantee that strategies continue to be adaptable to the shifting dynamics of undernutrition in various geographical areas.

Nigeria can significantly improve child nutrition outcomes and move closer to related Sustainable Development Goals by putting these recommendations into practice.

4.6 Limitations

Some potentially important characteristics, like specific dietary information or the availability of local food, might not have been included in the analysis because of data limitations, even though the MICS data was thorough.

The study's conclusions, although typical of Nigeria, could not be entirely applicable to other nations or situations with distinct socioeconomic circumstances or medical systems. When it comes to effective prediction, the Stack Ensemble machine learning technique might be harder to understand than conventional statistical techniques, which could limit its direct usage in policymaking without more context. Lastly, possible seasonal differences in nutritional status were not taken into consideration in this study, which may have an impact on the prevalence and contributing factors of undernutrition in various parts of Nigeria throughout the year.

5.0 CONCLUSION

Using machine learning approaches, this research has offered a thorough examination of childhood undernutrition in Nigeria, identifying important drivers and forecasting results. The Nigerian Multiple Indicator Cluster Survey data is used, together with the Stack Ensemble technique and other machine learning algorithms, to compare and identify the most promising strategy for improved prediction. We now have a better understanding of the nation's complicated child feeding situation. Our study clearly showed that, when compared to other machine learning methods, the Stack Ensemble technique performed better in predicting childhood stunting (92%), wasting (90%), and underweight (81%). This conclusion highlights the potential of data-driven techniques in solving important health concerns in developing nations, in addition to contributing to the methodological improvement in public health research. Access to clean water, household affluence, child age, birth weight, geographic zone, state, ethnicity, mother's educational attainment, and residential location are some of the drivers of childhood undernutrition that the study discovered. All three of the undernutrition indicators showed constant significance for these factors, underscoring the complexity of the issue and the necessity of all-encompassing solutions. Our findings have applications for politicians, medical professionals, and development practitioners that go beyond scholarly curiosity. Stakeholders may create more impactful initiatives, distribute resources more effectively, and track results more precisely by utilizing these data-driven insights.

In summary, this research marks a substantial advancement in our knowledge of and response to undernutrition among children in Nigeria. In order to address complicated social issues, it highlights the importance of combining cutting edge analytical techniques with public health research. To effectively tackle infant undernutrition going forward, it is

imperative that we convert these findings into workable policies and programs, continuously improve our strategies based on ongoing data analysis, and encourage interdisciplinary cooperation. In the end, the effectiveness of these initiatives will be evaluated not only by statistical gains but also by the noticeable improvement in the health, happiness, and prospects for the future of Nigerian children. This research lays the groundwork for such advancements, supporting the overarching mission of guaranteeing healthy lives and advancing everyone's well-being in accordance with international sustainable development goals.

6.0 ACKNOWLEDGMENT AND FUNDING

The writers express their appreciation to Tertiary Education Trust Fund (TETFund) and the Federal University of Gusau, Nigeria for the sponsorship of this research titled: Prognosticating Risk Factors of Undernutrition in Under Five Years Children in Nigeria Using Ensemble Technique, under the Institutional Based Research Grant (Ref ID: TETF/DR&D/CE/UNI/GUSAU/IBR/2024/VOL.1).

7.0 ACCESSIBILITY OF DATA AND RESOURCES

The data gathered and processed for this study are not publicly available in order to protect the privacy of individual responses. The public can access survey datasets in Nigeria National Bureau of Statistics (NBS) for legitimate research reasons only. The registration process must be finished in order to obtain authorization for the use of the dataset.

Writers' Contributions

Conceptualization: Mohammed Ali Kawo, Muhammad Nuruddeen Idris.

Data preprocessing: Mohammed Ali Kawo, Kabiru Haruna, Firdausi Umar Kaita

Formal analysis: Mohammed Ali Kawo, Muhammad Nuruddeen Idris.

Methodology: Mohammed Ali Kawo, Muhammad Nuruddeen Idris.

Writing – original draft: Mohammed Ali Kawo, Muhammad Nuruddeen Idris

Writing – review & editing: Mohammed Ali Kawo, Muhammad Nuruddeen Idris, Firdausi Umar Kaita, Kabiru Haruna

REFERENCES

- Ahmad, F., Almuayqil, S. N., Humayun, M., Naseem, S., & Khan, W. A. (2021). Prediction of COVID-19 cases using machine learning for effective public health management. *Computers, Materials & Continua*. <https://doi.org/10.32604/cmc.2021.013067>
- Agu, N., Emechebe, N., Yusuf, K., Falope, O., & Kirby, R. S. (2019). Predictors of early childhood undernutrition in Nigeria: the role of maternal autonomy. *Public Health Nutrition*, 22(12), 2279–

2289. <https://doi.org/10.1017/S1368980019000818>
- 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 February), 1–16. <https://doi.org/10.1371/journal.pone.0296625>
- Austin, P., & Ntenda, M. (2019). Association of low birth weight with undernutrition in preschool-aged children in Malawi. *Nutrition Journal*, 18, 1–15. <https://doi.org/10.1186/s12937-019-0477-8>
- Ayogu, R. N. B., Afiayeni, I. C., Madukwe, E. U., & Udentia, E. A. (2018). Prevalence and predictors of under-nutrition among school children in a rural South-eastern Nigerian community : a cross sectional study. *BMC Public Health*, 1–9.
- Bitew, F. H., Sparks, C. S., & Nyarko, S. H. (2021). Machine learning algorithms for predicting undernutrition among under-five children in Ethiopia. 25(2), 269–280. <https://doi.org/10.1017/S1368980021004262>
- Bitew, F. H., Nyarko, S. H., Potter, L., & Sparks, C. S. (2020). Machine learning approach for predicting under-five mortality determinants in Ethiopia: Evidence from the 2016 Ethiopian Demographic and Health Survey. *Genus*. <https://doi.org/10.1186/s41118-020-00106-2>
- Bizzego, A., et al. (2021). Predictors of contemporary under-5 child mortality in low- and middle-income countries: A machine learning approach. *International Journal of Environmental Research and Public Health*, 18(3), 1–12. <https://doi.org/10.3390/ijerph18031315>
- Bardhi, O., & Zapirain, B. G. (2021). Machine learning techniques applied to electronic healthcare records to predict cancer patient survivability. *Computers, Materials & Continua*. <https://doi.org/10.32604/cmc.2021.015326>
- Brand, M., et al. (2024). Revolutionizing women's health: A comprehensive review of artificial intelligence advancements in gynecology. *Journal of Clinical Medicine*, 13(4). <https://doi.org/10.3390/jcm13041061>
- Demsash, A. W. (2023). Using best performance machine learning algorithm to predict child death before celebrating their fifth birthday. *Informatics in Medicine Unlocked*, 40(June), 101298. <https://doi.org/10.1016/j.imu.2023.101298>
- Emily, J. M., et al. (2021). Application of machine learning approaches to administrative claims data to predict clinical outcomes in medical and surgical patient populations. *PLoS One*, 1–14. <https://doi.org/10.1371/journal.pone.0252585>
- Feng, W., Gou, J., Fan, Z., & Chen, X. (2023). An ensemble machine learning approach for classification tasks using feature generation. *Connection Science*, 35(1). <https://doi.org/10.1080/09540091.2023.2231168>
- Gayawan, E., Adebayo, S. B., & Waldmann, E. (2020). Modeling the spatial variability in the spread and correlation of childhood malnutrition in Nigeria. *WILEY-Statistics in Medicine*, July 2017, 1–22. <https://doi.org/10.1002/sim.8077>
- Ghosh, S., Hazra, A., & Raj, A. (2020). A Comparative Study of Different Classification Techniques for Sentiment Analysis. *International Journal of Synthetic Emotions*, 11(1), 49–57. <https://doi.org/10.4018/ijse.20200101.0a>
- Indiran, G., Selvandran, R., Ramprakash, K., & Doudou, N. (2021). Malnutrition in children under the age of 5 years in a primary health care setting. *South African Family Practice*. <https://doi.org/https://doi.org/10.4102/safp.v63i1.5416>
- Jijo, B. T., & Abdulazeez, A. M. (2021). Classification Based on Decision Tree Algorithm for Machine Learning. *Journal of Applied Science and Technology Trends*, 02(01), 20–28. <https://doi.org/10.38094/jastt20165>
- Jordan, M. I., & Mitchell, T. M. (2020). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255–260. <https://doi.org/10.1126/science.aaa8415>
- Keeley, B., Chief, E., Little, C., Vrolijk, K., Analyst, D., Wauchope, S., Specialist, P., Al, A., Alnaqshbandi, I., Editor, A., Perellon, C., Editor, S., Reboul-, A., Young, U., Declerck, F., Demaio, A., Harris, J., Hollis, J., Mccoll, K., ... Rudert, C. (2019). *Children , Food and Nutrition*.
- Kino, S., Hsu, Y., Shiba, K., Chien, Y., Mita, C., Kawachi, I., & Daoud, A. (2021). SSM - Population Health A scoping review on the use of machine learning in research on social determinants of health : Trends and research prospects. *SSM - Population Health*, 15, 100836. <https://doi.org/10.1016/j.ssmph.2021.100836>
- Nadine, K., Hans, E., Davia, S., & Adel, D. (2019). The influence of maternal agency on severe child undernutrition in conflict-ridden Nigeria : Modeling heterogeneous treatment effects with machine learning. *PLoS ONE*, 1–16.
- Nusinovici, S., Tham, Y. C., Yu, M., Yan, C., Shu, D., Ting, W., Li, J., Sabanayagam, C., Wong, T. Y., & Cheng, C. (2020). Logistic regression was as good as machine learning for predicting major chronic diseases. *Journal of Clinical Epidemiology*. <https://doi.org/10.1016/j.jclinepi.2020.03.002>



- Nwozor, A., & Olanrewaju, J. S. (2020). The ECOWAS agricultural policy and the quest for food security: assessing Nigeria's implementation strategies. *DEVELOPMENT STUDIES RESEARCH*. <https://doi.org/10.1080/21665095.2020.1785904>
- Odegua, R. O. (2020). An Empirical Study of Ensemble Techniques (Bagging, Boosting and Stacking). *Deep Learning IndabaX*, 12(10), 1683.
- Olotusin, A. O., Ridwan, M., & Oluwaseun, A. O. (2023). Investigating and quantifying food insecurity in Nigeria: A systematic review. *MDPI*, 1–38.
- Osisanwo, F. Y., Akinsola, J. E. T., Awodele, O., Hinmikaiye, J. O., Olakanmi, O., & Akinjobi, J. (2017). Supervised Machine Learning Algorithms: Classification and Comparison. *International Journal of Computer Trends and Technology*, June. <https://doi.org/10.14445/22312803/IJCTT-V48P126>
- Quang Dat, T., Nguyen Huong Giang, L., Thi Tuong Loan, N., & Van Toan, V. (2018). The prevalence of malnutrition based on anthropometry among primary schoolchildren in Binh Dinh province, Vietnam in 2016. *AIMS Public Health*, 5(3), 203–216. <https://doi.org/10.3934/publichealth.2018.3.203>
- Rahman, S., & Rahman, A. (2019). Prevalence of undernutrition in Bangladeshi children. *Journal of Biosocial Science* (2019), 1 of 14. <https://doi.org/10.1017/S0021932019000683>
- Report, K. I. (2019). Nigeria Demographic and Health Survey Key Indicators Report. May.
- Sejal, S., & Adnan, C. (2019). Childhood asthma: Advances using machine learning and mechanistic studies. *American Journal of Respiratory and Critical Care Medicine*, 199(4), 414–422. <https://doi.org/10.1164/rccm.201810-1956CI>
- Soofi, A. A., & Awan, A. (2017). Classification Techniques in Machine Learning: Applications and Issues. *Journal of Basic & Applied Sciences*, 459–465.
- Steyn, N. P., Rp, A., & Dsc, W. (2020). Nutritional status and food security in Sub-Saharan Africa: Predictions for 2020. *Asia Pacific J Clin Nutr*, 9(December 1999), 1–6.
- Survey, H., International, I. C. F., Goals, S. D., Survey, H., International, I. C. F., Bank, W., Bank, W., Bank, C., & Buhari, P. M. (2021). Nigeria: Nutrition Profile. USAID, May.
- Turjo, E. A., & Rahman, M. H. (2024). Assessing risk factors for malnutrition among women in Bangladesh and forecasting malnutrition using machine learning approaches. *BMC Nutrition*, 10(1), 1–25. <https://doi.org/10.1186/s40795-023-00808-8>
- Theobald, O. (2017). Machine Learning For Absolute Beginners.
- Theses, E. (2023). Sentiment Analysis using Ensemble Learning.
- WHO. (2020). Children: improving survival and well-being. In <https://www.who.int/news-room/fact-sheets/detail/children-reducing-mortality>.
- Zhou, Z.-H. (2025). Ensemble methods: Foundations and algorithms (2nd ed.). Boca Raton, FL: CRC Press. <https://doi.org/10.1201/9781032960609>