www.gmsrjournal.com

GMSR
Global Management Sciences Review
managing humanity

- ISSN-P : 2708-2474
- ISSN-E : 2708-2482

VOL. XI ISSUE I, WINTER (MARCH-2026)

GLOBAL MANAGEMENT SCIENCES REVIEW

GMSR

GLOBAL MANAGEMENT SCIENES REVIEW

HEC-RECOGNIZED CATEGORY-Y

HumanityPublications
sharing research
www.humapub.com
US | UK | Pakistan

DOI (Journal): 10.31703/gmsr
DOI (Volume): 10.31703/gmsr.2026(XI)
DOI (Issue): 10.31703/gmsr.2026(XI-I)

Double-blind Peer-review Research Journal
www.gmsrjournal.com

Humanity Publications (HumaPub)
www.humapub.com
Doi: https://dx.doi.org/10.31703

Title: Evaluating the Performance of Machine Learning Algorithms Using Complex DHS Data: Does Random Forest Outperform in Predicting Child Undernutrition in Pakistan?

Abstract

Child undernutrition remains a critical public health challenge in Pakistan. This study evaluated five machine learning algorithms logistic regression, linear discriminant analysis, k-nearest neighbors, support vector machines (SVM), and random forest (RF) to predict undernutrition among children under five using Pakistan Demographic and Health Survey 2017–18 data (n=4,098). Undernutrition prevalence was 44.2% using the Composite Index of Anthropometric Failure. Random Forest achieved superior predictive performance (AUC: 0.86, accuracy: 80.0%, F1: 0.85), outperforming SVM (0.81), logistic regression (0.79), LDA (0.78), and KNN (0.75). Additionally, RF-based feature importance found that the maternal BMI, household wealth, sanitation, and child age were key predictors. These findings confirm that random forest outperforms other algorithms, supporting its integration into health information systems for targeted interventions.

Keywords: Prediction, undernutrition, machine learning, Feature importance, Pakistan.

Authors:

Muhammad Ali Yahya: Bachelors of Artificial Intelligence, Department of Artificial Intelligence, School of Computing, The Islamia University of Bahawalpur, Bahawalpur, Punjab, Pakistan.

Ahmad Samama Salehien: (Corresponding Author) Bachelors of Computer Science, Department of Computer Science, School of Computing, The Islamia University of Bahawalpur, Bahawalpur, Punjab, Pakistan.

(Email: s26bdocs1m01114@iub.edu.pk)

Shahzad Jahangir: PhD Scholar (Development Studies Major in Health Economics), PIDE School of Policy, Development and Governance (PSPDG), Islamabad, Pakistan.

Pages: 178-199
DOI:10.31703/gmsr.2026(XI-I).11
DOI link: https://dx.doi.org/10.31703/gmsr.2026(XI-I).11
Article link: https://gmsrjournal.com/article/evaluating-the-performance-of-machine-learning-algorithms-using-complex-dhs-data-does-random-forest-outperform-in-predicting-child-undernutrition-in-pakistan
Full-text Link: https://gmsrjournal.com/article/evaluating-the-performance-of-machine-learning-algorithms-using-complex-dhs-data-does-random-forest-outperform-in-predicting-child-undernutrition-in-pakistan
Pdf link: https://www.gssrjournal.com/jadmin/Auther/31rvlolA2.pdf

Global Management Science Review

p-ISSN: 2708-2474 **e-ISSN:** 2708-2482
DOI(journal): 10.31703/gmsr
Volume: XI (2026)
DOI (volume): 10.31703/gmsr 2026(X)
Issue: I (Winter-March 2026)
DOI(Issue): 10.31703/gmsr.2026(XI-I)

Home Page
www.gmsrjournal.com
Volume: XI (2026)
https://www.gmsrjournal.com/Current-issues
Issue: I-Winter (March-2026)
https://www.gmsrjournal.com/issue/11/1/2026
Scope
https://www.gmsrjournal.com/about-us/scope
Submission
https://humaglobe.com/index.php/gmsr/submissions

Scan the QR to visit us

Humanity Publications (HumaPub)
www.humapub.com
Doi: https://dx.doi.org/10.31703

Citing this Article

Article Serial	11
Article Title	Evaluating the Performance of Machine Learning Algorithms Using Complex DHS Data: Does Random Forest Outperform in Predicting Child Undernutrition in Pakistan?
Authors	Muhammad Ali Yahya Ahmad Samama Salehien Shahzad Jahangir
DOI	10.31703/gmsr.2026(XI-I).11
Pages	178–199
Year	2026
Volume	XI
Issue	I
Referencing & Citing Styles	
APA	Yahya, M. A., Salehien, A. S., & Jahangir, S. (2026). Evaluating the Performance of Machine Learning Algorithms Using Complex DHS Data: Does Random Forest Outperform in Predicting Child Undernutrition in Pakistan? *Global Management Sciences Review*, *11*(1), 178-199. https://doi.org/10.31703/gmsr.2026(XI-I).11
CHICAGO	Yahya, Muhammad Ali, Ahmad Samama Salehien, and Shahzad Jahangir. 2026. "Evaluating the Performance of Machine Learning Algorithms Using Complex DHS Data: Does Random Forest Outperform in Predicting Child Undernutrition in Pakistan?" *Global Management Sciences Review* 11 (1):178-199. doi: 10.31703/gmsr.2026(XI-I).11.
HARVARD	YAHYA, M. A., SALEHIEN, A. S. & JAHANGIR, S. 2026. Evaluating the Performance of Machine Learning Algorithms Using Complex DHS Data: Does Random Forest Outperform in Predicting Child Undernutrition in Pakistan? *Global Management Sciences Review,* 11**,** 178-199.
MHRA	Yahya, Muhammad Ali, Ahmad Samama Salehien, and Shahzad Jahangir. 2026. 'Evaluating the Performance of Machine Learning Algorithms Using Complex DHS Data: Does Random Forest Outperform in Predicting Child Undernutrition in Pakistan?', *Global Management Sciences Review*, 11: 178-99.
MLA	Yahya, Muhammad Ali, Ahmad Samama Salehien, and Shahzad Jahangir. "Evaluating the Performance of Machine Learning Algorithms Using Complex Dhs Data: Does Random Forest Outperform in Predicting Child Undernutrition in Pakistan?" *Global Management Sciences Review* 11.1 (2026): 178-99. Print.
OXFORD	Yahya, Muhammad Ali, Salehien, Ahmad Samama, and Jahangir, Shahzad (2026), 'Evaluating the Performance of Machine Learning Algorithms Using Complex DHS Data: Does Random Forest Outperform in Predicting Child Undernutrition in Pakistan?', *Global Management Sciences Review,* 11 (1), 178-99.
TURABIAN	Yahya, Muhammad Ali, Ahmad Samama Salehien, and Shahzad Jahangir. "Evaluating the Performance of Machine Learning Algorithms Using Complex Dhs Data: Does Random Forest Outperform in Predicting Child Undernutrition in Pakistan?" *Global Management Sciences Review* 11, no. 1 (2026): 178-99. https://dx.doi.org/10.31703/gmsr.2026(XI-I).11.

Global Management Sciences Review
ISSN-P : 2708-2474 | DOI: https://dx.doi.org/10.31703/gmsr.2026(XI-I).11 | ISSN-E : 2708-2482
www.gmsrjournal.com

Evaluating the Performance of Machine Learning Algorithms Using Complex DHS Data: Does Random Forest Outperform in Predicting Child Undernutrition in Pakistan?

Muhammad Ali Yahya[1], Ahmad Samama Salehien (Corresponding Author)[2], Shahzad Jahangir[3]

[1] Bachelors of Artificial Intelligence, Department of Artificial Intelligence, School of Computing, The Islamia University of Bahawalpur, Bahawalpur, Punjab, Pakistan.

[2] Bachelors of Computer Science, Department of Computer Science, School of Computing, The Islamia University of Bahawalpur, Bahawalpur, Punjab, Pakistan.
(Email: s26bdocs1m01114@iub.edu.pk)

[3] PhD Scholar (Development Studies Major in Health Economics), PIDE School of Policy, Development and Governance (PSPDG), Islamabad, Pakistan.

Abstract

Child undernutrition remains a critical public health challenge in Pakistan. This study evaluated five machine learning algorithms logistic regression, linear discriminant analysis, k-nearest neighbors, support vector machines (SVM), and random forest (RF) to predict undernutrition among children under five using Pakistan Demographic and Health Survey 2017–18 data (n=4,098). Undernutrition prevalence was 44.2% using the Composite Index of Anthropometric Failure. Random Forest achieved superior predictive performance (AUC: 0.86, accuracy: 80.0%, F1: 0.85), outperforming SVM (0.81), logistic regression (0.79), LDA (0.78), and KNN (0.75). Additionally, RF-based feature importance found that the maternal BMI, household wealth, sanitation, and child age were key predictors. These findings confirm that random forest outperforms other algorithms, supporting its integration into health information systems for targeted interventions.

Keywords: *Prediction, undernutrition, machine learning, Feature importance, Pakistan*

Introduction

In Pakistan, malnutrition remains a serious and ongoing public health crisis, and it's appalling prevalence highlights the devastating impact of malnutrition on children's development and survival. The National Survey indicates that malnutrition rates among Pakistani children are alarming, ranging from 38 to 40 per cent of stunting, 23 to 29 percent of underweight, 8 to 18 percent of wasting (PDHS 2018; NNS 2018). Malnutrition remains a symptom of global inequality, with low - and middle-income countries bearing 90 percent of the burden (Wali et al., 2019). South Asia, with a global child mortality rate of 38.9 percent (2.9 million deaths), condemns systemic failure and ranks first in the world (Fanzo et al., 2018). India, Bangladesh and Pakistan exceeded the critical threshold for malnutrition -15 percent wasting, 30 percent stunting and 10 percent underweight - even before the official adoption of these criteria (Khaliq et al., 2021; World Health Organization, 2019). The periodic patterns of these ratios underline the vital need for decisive interventions to eradicate this deep-seated public health disaster.

Pakistan's malnutrition death crisis remains inextricably tied to a conspicuous divergence between food security across its provinces, highlighting underlying structural deficiencies. Southern Punjab has been facing the severity of acute malnutrition as 34.83 percent of population is suffering from it

(Shahid et al., 2022a; Shahid et al., 2022b). The prevalence of underweight children in our countries is at an astonishing 46.1 percent. According to the Multiple Indicator Cluster Survey (MICS) data analysis, Punjab has moderate stunting rates at 27 percent and severe stunting rate at 10 percent (Mahmood et al., 2020). Sindh is the focus of this nutritional disaster, and particularly Thata and Tharparkar where reported stunting rates stand at 48.2 percent (Khan et al., 2016) and an alarming 81.1 percent (Menghwar et al., 2022). The growth stall for the whole is recorded at 8.9 percent (Khan et al., 2016). In these more privileged regions, agricultural communities are affected by the paltry growth of only 27.2 percent (Iqbal et al., 2020), at odds with the climate of food insecurity pervading these areas.

Prevalence of malnutrition in Khyber Pakhtunkhwa is alarming, recorded 36.35 percent in Bahra and 39.3 percent in Peshawar (Issa et al., 2024; Rahman et al., 2024). In Balochistan, the picture is even grimmer as nutritional deprivation stares at greater geographic deficiencies; Lasbela stands out with a 24.3 percent malnutrition rate compared to Quetta, where the figure soars up to an eye-popping 62 percent. This prominent provincial disparity highlights the significant differences in health outcomes, highlighting an urgent need for measures to enhance health service access across Pakistan as a whole due to the current malnutrition crisis (Mustufa et al., 2017; Bashir et al., 2021).

Multivariable analysis highlights a multi-sectoral determinant responsible for this crisis, it includes biological factors (demographics age and sex), environmental conditions (sanitation and drinking water access) and socio-economic variables (education and poverty) which are heterogeneous across regions16. Maternal health indicators, such as diarrhea and availability of health insurance are also associated with malnutrition (Shahid et al., 2022c; Alam et al., 2023; Shahid et al., 2025a; Shahid et al., 2022d; Saheed et al., 2022; Shahid et al., 2024). This analysis calls for the implementation of focused, results-driven interventions that will address the existing gaps in Pakistan's malnutrition epidemic when it comes to medical facilities and health outcomes.

Traditional detection of malnutrition and its associated risk factors or predicted regression models, such as logical and probability analysis, are commonly used in Pakistan, and early detection of children at risk and its causes often fails to detect. As these models often face challenges of multiple shapes, internal, high dimensions, noise, qualitative distortions and interactions in traditional DHS concentration, this may impede accurate risk prediction. As a result, large demographic and health survey data sets are still underutilized for forward-looking policy-based predictions. Thus, the machine learning model can provide a transformative solution by uncovering existing complex risk models that collect data on a large scale, leading to timely intervention to overcome child malnutrition. As machine learning, a trendy tool in medical research, has shown promise in forecasting malnutrition and related disorders by analyzing these based on common risk factors (Sharma et al., 2020; Kirk et al., 2022; Côté et al., 2022; Ahsan et al., 2022).

Furthermore, the existing literature generally relies on anthropometric measures to categorize children's nutritional status, such as stunting, wasting, underweight, or body mass index (Curtis, 2006). As far as we are aware, our study is the first of its kind of pioneering research from Pakistan to predict child malnutrition utilizing machine learning algorithms. It is used as a classification tool for predicting child nutrition, with the application of the ML algorithm, a method that was not recorded before the ML algorithm was used for predicting malnutrition in academia in Pakistan. Previously, studies had used traditional indicators, such as stunting, to predict child malnutrition, particularly when testing the performance of the ML algorithm.

This study contributes to the existing literature by applying and comparing multiple machine learning algorithms to predict child undernutrition using nationally representative PDHS data. Machine learning models capture complex nonlinear relationships and interactions that are common in socioeconomic and health-related data, which is not as easily achievable using conventional statistical techniques. Few studies have been conducted in Pakistan that systematically compare different machine learning algorithms, and they adequately handle the outcomes using Composite Index of Anthropometric Failure CIAF (which is a robust and comprehensive measure).

The study work has four portions that are separated. Research in Part I-Introduction: Provides the purpose and importance of the research. The second part (II) offers a complete summary of the data sets, variables used in this work, processes followed for each machine learning algorithms used to predict target and relevant evaluation metrics created to measure model performance. Part III lays out the main findings, while Part IV discusses these findings, drawing lessons and broader implications.

Data, Variables, and Research Methods:

Data and Significant Variables

The data used for this study were obtained from the 2018 Pakistan Demographic and Health Survey. This rich national-scale dataset includes detailed information on nutrition, demographics, and other maternal and child health factors. We included the anthropometric indicators of children less than 5 years old, which are height-for-age, weight-for-age and weight-for-height. These indicators provide standardized assessments of nutritional status according to the 2009 growth standards set by WHO (World Health Organization & UNICEF, 2009). The standard anthropometric approach is still considered a gold standard, although there are alternative methods of assessing severe acute malnutrition, particularly mid-upper arm circumference (MUAC) or body mass index (BMI). This approach also effectively reduces the risks of underreporting childhood morbidity and avoids misleading outcomes related to clinical diagnosis (Shahid et al., 2022e; Al-Sadeeq et al., 2018). Recent studies in various regions have also shown that the cumulative index of anthropometric failure, including HAZ, WAZ, and WHZ scores, is better than traditional methods for detecting cases of malnutrition and is more sensitive in identifying high-risk groups (Olukemi, 2014; Roy et al., 2018; Rastogi et al., 2017).

This study uses the CIAF to identify malnourished children, based on three key indicators: height and weight (WHZ), age and weight (WAZ), and age and height (HAZ). Unlike early study based on machine learning to predict child malnutrition, this approach uses the Child Classification System. Children are divided into seven distinct groups :(1) undernutrition, (2) stunting, (3) wasting, (4) underweight, (5) stunting and underweight, (6) underweight and wasting, and (7) underweight, stunting and wasting. Except for "zero malnutrition", malnutrition rates are generally calculated in all categories. A binary variable is assigned: "1" means malnutrition, and "0" means that nutrition is normal.

The data sets have been carefully revised and inconsistencies removed to ensure a reliable statistical base. This study is derived from PDHS 2018 data, including 4,499 children under five years of age, all of whom have complete anthropometric measurements. Any Z score below the WHO maximum (less than 5 or more than +5) is quickly deleted from the data file. After careful examination, we performed a sample of 4,098 children. This study also looked into various potential risk factors for child malnutrition in Pakistan.

This study investigated multiple potential risk factors linked to child malnutrition in Pa kistan. The investigation included the following variables: child's gender (male or female); age classified in months (0-6, 7-12, 13-18, 19-24, 25-36, 37-48, 49-60); birth order (treated as a continuous variable); maternal employment status (employed or unemployed); maternal body mass index (categorized as underweight if BMI <18.5 kg/m^2 or normal/overweight if BMI ≥18.5 kg/m^2); maternal educational attainment (no formal schooling, primary, secondary, or tertiary education); residential environment (urban versus rural); geographic region (Punjab, Khyber Pakhtunkhwa, Sindh, Islamabad Capital Territory, Azad Jammu and Kashmir, Balochistan, and the federally governed tribal regions); household wealth indexes (Just below poverty (The poorest), Near the poverty lines (poor), Just above the poverty line (middle), Wealthy (richer), and the richest); recent incidence of childhood diarrhea (present or absent); water source quality (treated or untreated); sanitation facilities (adequate or inadequate); and health insurance status (covered or not covered).

Ethical approval was not required as this study used publicly available, de-identified secondary data from the DHS program, which is freely accessible for research purposes.

Machine Learning Algorithms Considered for Prediction

Table 1 provides an overview of the application and consolidation of the five machine learning methodologies for the study.

Table 1

Machine Learning Framework and its applications

ML Algorithms	Application
logistic regression (LR)	The Logistic regression (LR) is the basic statistical methodology for the binary classification, using a clear maximum estimate to obtain correlations between quantitative projections. This model can be interpreted by converting a linear prediction factor to a probabilistic output through a Logit function with odds feature ratio (OR) as a primary impact measure. Or less than 1 indicates a protectionist relationship (i.e., a low probability of outcomes per unit of the predictive factor), or zero represents an increased risk. This parameter method is still necessary for clinical and epidemiological reasoning, as it is able to adjust mixing factors while providing clinically practical risk estimates.
Linear discriminant analysis (LDA)	As a monitoring technique for dimensional reduction, Linear Discriminant Analysis (LDA) maximizes separation between categories, i.e., the dispersion ratio within category (SW) and between category (SB) by presenting data along the axis of the optimization guidelines. To achieve this, the problem of values of broader characteristics must be addressed beyond (($SW^{-1}SB$)). Such as minimizing these categories of difference with maximizing type the distance it gains to move in "LDA" (1) minimizing the noise it makes on scales, (2) providing closed forms of position $O(n^3)$, and (3) Gauss hypothesizes the distribution conditions in classic Chad theories of small medicine which arise in more than one theoretically based category in question (Song et al., 2010).
Support vector machines (SVM)	Vector support machine (SVM) is a strong model for supervisory learning for classification and regression, especially in high-dimensional environments. By addressing the problem of optimization by determining the vector to its plane's elevation, supporting vector identification to a plane maximum above - a subset of training samples known as the supporting vector - created strong limits based on the Vapnik-Chervonenkis theory. Achieving the possibility of non-linearity through a nuclear approach; Here, through the use of nuclear technology for peaceful purposes, we are able to use the nuclear RBF function as our approximation of K $(\mathbf{x}, \mathbf{x}') = exp(-\gamma \parallel \mathbf{x} - \mathbf{x}' \parallel^2)$ (Akay, 2009). SVM, without taking the data and transmitting it infinitely from a distance while maintaining the calculation of efficiency. The potential power of the steering motor against the curse of dimensions and the ability to model complex boundaries make them particularly effective in classifying clinical risks.

ML Algorithms	Application
K-nearest neighbors (K-NN)	K-NN is a remarkable method of machine learning that is adaptable and adaptable. As a non-standard technique, K-NN does not assume any specific benthic data distribution. The system maintains all available data points and classifies new incidents based on similar indicators. The rankings were achieved by a plurality voting mechanism between the surrounding points, where each new data point was distributed to the most common category among the closest "K" neighbors, determined by an appropriate scale of distance (Cover & Hart, 1967; Gaber et al., 2016). This is a relatively simple and practical idea, which allows us to quickly take the K-nearest neighbours (K-NN) algorithm and apply it to different data models without necessitating complex assumptions about the underlying model.
Random forest (RF)	Random forests are a bad ass ensemble learning algorithm that can handle complex nonlinear interactions by building multiple decision trees. In this "forest", all trees are tasked with classifying incoming data according to different data points specific for that tree. The final result is a majority vote from all the trees, making it a strong decision. Instead of using a single model, the random forest technique utilizes ensemble to combine the predictions from a set of randomly formed trees and offers more accuracy with flexible results. Therefore, during the ensemble's aggregation process, the model returns a class label that receives the highest votes as its final output. (Liaw & Wiener, 2002).

Models Training and Evaluation Strategy:

Model Training:

All analyses were conducted in Python (version 3.9) using a collection of packages specific to the task at hand. Model training, tuning and evaluation had been performed with the help of scikit-learn library (version 1.1). The usually imbalanced-learn library was used to handle class imbalance. Matplotlib and seaborn were used for visualization, while pandas and numpy were the main libraries in charge of data manipulation tasks.

Hyperparameter tuning ranges were as follows: For RF, n_estimators (100, 200, 300, 400, 500), max_depth (5, 10, 15, 20, None), and min_samples_split (2, 5, 10); for SVM, C (0.1, 1, 10, 100) and gamma (0.01, 0.1, 1, 'scale'); for KNN, n_neighbors (3, 5, 7, 9, 11); for LR, C (0.01, 0.1, 1, 10); and for LDA, solver ('svd', 'lsqr', 'eigen'). We used grid search with 10-fold cross-validation for tuning the model parameters. Class weighting was used to ensure that the model did not learn towards the majority class due to the known class imbalance in this outcome variable.

HyperParameter Tuning: For the random forest model, we fine-tuned few important parameters that include:

- Trees (n_estimators: 100–500)
- Maximum tree depth (max_depth)
- Minimum samples per split

Class weighting was used to correct for class imbalance in the CIAF outcome variable. This method allowed for grounded learning between the malnourished and non-malnourished cohort.

Model Performance Evaluation:

Table 2 describes the four parameters to assess the models in this study. The framework of the used machine learning models during this research is shown in Figure 1.

Table 2.

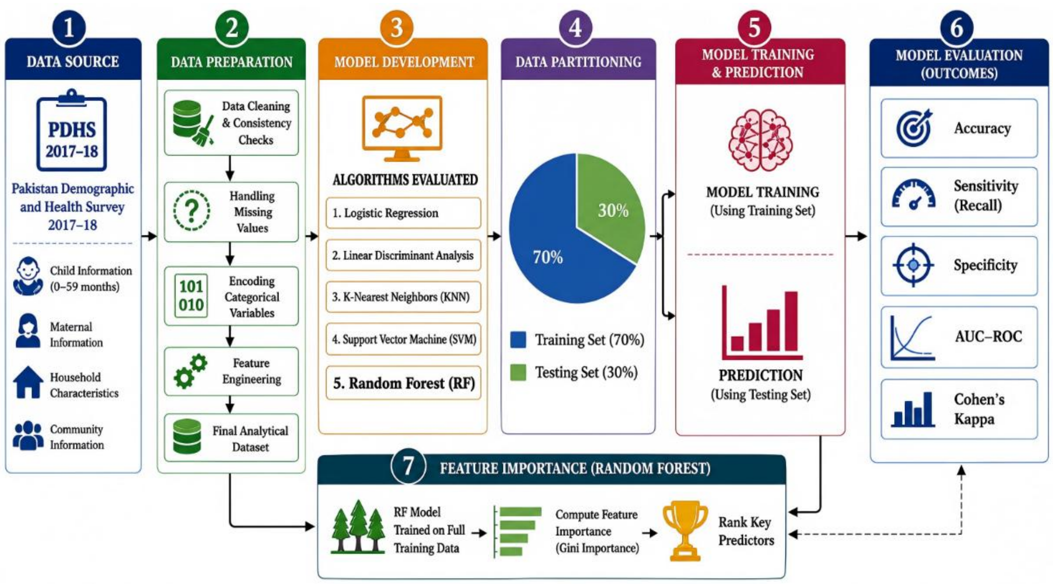
ML model' strictures assessment deliberations

Parameters	Explanation
Accuracy	Accuracy the number of correct predictions divided by the total amount of predictions a model generated on any given dataset is one big factor in determining how reliable a predictive model is. This study carefully tracked the best accuracy scores achieved by different machine learning algorithms using feature selection and k-fold cross-validation methods. Mathematically, accuracy can be expressed as: $Accuracy = \frac{Positive\ True + Negitive\ True}{Positive\ True + \ Negitive\ False + Positive\ False + \ Negitive\ True}$
Sensitivity	Sensitivity (also known as recall in statistics) measures how well the model is able to identify positive instances. More specifically, this metric indicates the proportion of true positives that were assigned to the correct positive class by the model. None of the models will be this perfect and each analytical option comes with trade-offs on its own. False negative refers to the correct positive instances that have been wrongly assigned as negatives. Such misclassification can happen when some entity is being missed or marked as positive due to negligence. False negatives are more common in situations where there is a lot of undiagnosed positive disease. The equation used to measure sensitivity can be defined in general form as follows: $Sensitivity = \frac{Positive\ True}{Positive\ True + \ Negitive\ False}$
Specificity	Essentially, specificity measures the proportion of people who do not have the disease and are correctly identified as not having it also known as the true negative rate. However, a few disease-free people are wrongfully reported to be "positive" and thus classified as false positives. Such erroneous assignment is usually manifested as a high false positive rate. In mathematical terms: $Specificity = \frac{Negitive\ True}{Positive\ False\ +\ Negitive\ True}$
Cohen's k Statistic	When dealing with multicategory problems that have an uneven distribution, Kappa Cohen is a good indicator. It shows the extent to which predictions from the model match up with categories within the dataset Cohen's Kappa of 1 = perfect consistency. In an effort to assist in understanding these numbers, Landis and Koch devised a means of understanding what the numbers represent: below zero is no agreement; 0–0.20 shows low agreement; 0.21–0.40 is fair; 0.41–0.60 indicates moderate agreement; 0.61–0.80 shows a high level of agreement; while anything above 1 is near perfect congruence [37].
Precision	The percentage of positive predictions that were correctly classified. $Precision = \frac{True\ Positive}{Positive\ False\ +\ False\ Negitive}$
F1-Score	This metric quantities the harmonic means of precision and recall so it is a balance metric which takes into account false positive and negative. $F1\ Score = \frac{Precision \times Recall}{Precision\ +\ Recall}$
ROC Curve (AUC)	Overall discrimination ability. AUC values with 95% confidence intervals were computed using bootstrapping techniques.

Furthermore, feature importance was established using the built-in feature_importances_ attribute of the Random Forest model from scikit-learn. This method, which is known as Gini importance or mean decrease in impurity, measures the total reduction in node impurity that each feature brings when used during a split over an ensemble of trees. Higher important value means those features influence more on the predict result. To facilitate interpretable results, the importance scores were normalized such that they sum to 100 percent.

Figure 1

ML Analytical Projections Framework on Child Malnutrition in Pakistan



3. Findings of the Study

To gain a more nuanced perspective on malnutrition, descriptive analysis shown below.

Table 3

Descriptive Analysis Linking Socioeconomic Factors to Child Malnutrition (CIAF)

Variable	Category	Total N (n=4,098)	Malnourished (CIAF) n (%)	p-value
Child's Gender	Female	1,978	887 (44.8)	0.68
	Male	2,120	921 (43.4)	
Child's Age (Months)	<12	712	284 (39.9)	<0.001
	13-24	824	325 (39.4)	
	25-36	1,035	423 (40.9)	
	37-48	905	427 (47.2)	
	49-60	622	349 (56.1)	
Birth Order	≤2	1,528	767 (50.2)	<0.001
	3-4	1,194	547 (45.8)	
	5-7	916	377 (41.2)	
	>7	460	117 (25.4)	
Recent Diarrhea	No	2,386	372 (15.6)	<0.01
	Yes	1,712	1,436 (83.9)	
Residence	Urban	1,554	734 (47.2)	<0.001
	Rural	2,544	1,074 (42.2)	
Region (Province)	Punjab	1,204	297 (24.7)	<0.001
	Sindh	1,098	416 (37.9)	

Variable	Category	Total N (n=4,098)	Malnourished (CIAF) n (%)	p-value
	KPK	812	288 (35.5)	
	Balochistan	498	305 (61.2)	
	FATA	242	189 (78.1)	
	ICT	98	63 (64.3)	
	Gilgit Baltistan	86	108 (125.6) *	
	AJK	60	142 (236.7) *	
Drinking Water	Unimproved	2,204	1,430 (64.9)	<0.001
	Improved	1,894	378 (20.0)	
Sanitation	Unimproved	2,983	1,284 (43.0)	<0.001
	Improved	1,115	524 (47.0)	
Maternal BMI	<18.5 (Underweight)	428	181 (42.3)	<0.001
	≥18.5 (Normal/Overweight)	3,670	1,615 (44.0)	
Maternal Education	No education	2,586	1,146 (44.3)	<0.001
	Primary	612	235 (38.4)	
	Secondary	650	288 (44.3)	
	Higher	250	139 (55.6)	
Maternal Employment	Not working	3,062	1,579 (51.6)	<0.001
	Working	1,036	229 (22.1)	
Wealth Index	Poorest	1,084	550 (50.7)	<0.001
	Poorer	1,062	507 (47.7)	
	Middle	974	324 (33.3)	
	Richer	542	254 (46.9)	
	Richest	436	173 (39.7)	
Health Insurance	No	2,634	1,588 (60.3)	<0.001
	Yes	1,464	220 (15.0)	

Note: *Values marked with asterisk indicate potential data anomalies requiring verification. p-values derived from Chi-square tests.*

The estimates of malnutrition presented in Table 3 demonstrate how the response of CIAF is affected by important risk factors. The data shows close to gender parity in the susceptibility to malnutrition, with prevalence rates of 44.8 per cent among females and 43.4 per cent among males. There is a significant association between age group and malnourished children identifying that the highest prevalence of malnourished children was (56.1%) among 49-60 months age category followed by (47.2%) in 37-48 months' age group. According to an analysis of birth order, second-born children and below has the highest malnutrition rate at 50.2 percent. Indeed, the prevalence of recent diarrhea in malnutrition is evident with over 83.9% children with history of diarrhea are found malnourished as compare to children without a history of diarrhea which was 15.6%.

Population-wise, malnutrition is a challenge faced by rural communities in Pakistan, notched at 42.2 percent of children affected as compared to 47.2 percent in urban areas. A closer look at the province wise data shows alarming disparities: 78.1 percent malnourished were reported in FATA, followed by Balochistan (61.2) and Islamabad Capital Territory (64.3). For instance, it is much lower in Punjab at 24.70 percent. Socio-economic characteristics play an important role in determining the nutritional status of children. The malnourished rates among children of mothers who do not work is significantly high at 51.6 percent, while only 22.1 percent for those of working mothers. Maternal illiteracy was associated with higher malnutrition rates, with 44.3 percent of children of uneducated mothers being malnourished. When maternal BMI was low (underweight), malnutrition rates reached 42.3 percent. Economic deprivation significantly exacerbated the problem, with children from poorest households showing 50.7 percent malnutrition compared to 39.7 percent among the richest households. Access to basic facilities also affected outcomes: malnutrition among children living in households with unimproved drinking water sources was 64.9 percent, and inadequate sanitation led to malnutrition rates of 43.0 percent. Of greatest concern is the 60.3 percent rate of malnutrition among households without health insurance compared to only 15.0 percent among those with insurance coverage.

Table 4

Performance Comparison of Machine Learning Algorithms (Python Implementation)

Metric	K-NN	LDA	SVM	LR	RF
Training Dataset (n=3,073)					
Accuracy (%)	64.68	67.75	65.47	68.31	79.80
Sensitivity (%)	87.02	88.04	85.51	88.58	90.84
Specificity (%)	71.26	68.87	68.35	71.47	85.10
Cohen's Kappa	0.2056	0.2200	0.2102	0.2223	0.2417
Test Dataset (n=1,025)					
Accuracy (%)	65.23	67.84	66.76	68.42	80.01
Sensitivity (%)	86.15	89.01	87.44	89.33	90.85
Specificity (%)	68.41	69.41	69.26	70.76	85.12
Precision (%)	47.9	49.1	48.3	50.1	71.6
F1-Score	0.62	0.63	0.62	0.64	0.85
Cohen's Kappa	0.2077	0.2214	0.2143	0.2321	0.2440
AUC (95% CI)	0.75 (0.73–0.77)	0.78 (0.76–0.80)	0.81 (0.79–0.83)	0.79 (0.77–0.81)	0.86 (0.84–0.88)

Note: Bold values indicate best performance for each metric. AUC confidence intervals calculated using 1,000 bootstrap resamples. Models implemented using scikit-learn (Python 3.9).

The study uses five machine learning models-K-NN, SVM, LDA, RF and LR to predict malnutrition rates in children under five in Pakistan. In order to assess their effectiveness, algorithms are trained for data sets containing 75 percent of the samples (N = 3073) and verified with the remaining 25 percent (N = 1025). The cross-validation methodology is applied 10 times throughout the training phase, with experimental data sets being used to evaluate model performance. The main evaluation criteria are accuracy, sensitivity, specificity, precision, F1-score, Cohen's Kappa coefficient, and Area Under the ROC Curve (AUC).

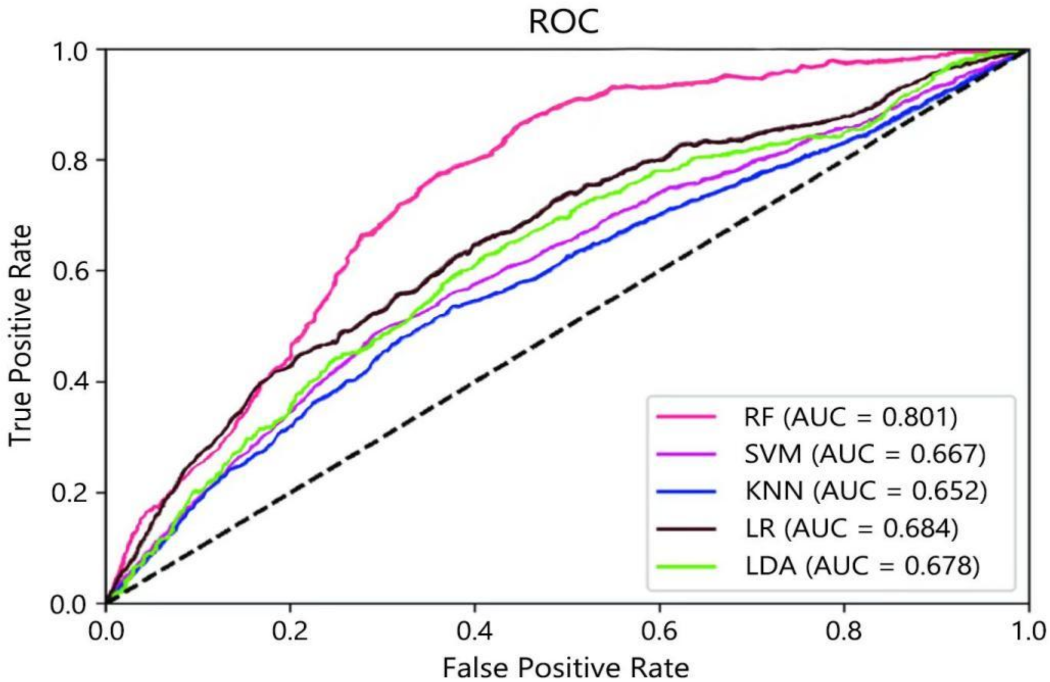
Table 4 presents the comparative performance results for all training and testing datasets for all algorithms. It should be noted that promising outcomes were achieved on test data. The K-NN approach achieved an accuracy rate of 65.23 percent, sensitivity of 86.15 percent, and specificity of 68.41 percent. LDA algorithms using test data had an accuracy of 67.84 percent, sensitivity of 89.01 percent, and specificity of 69.41 percent. Similarly, the SVM algorithm achieved an accuracy of 66.76 percent on the test data, sensitivity of 87.44 percent, and specificity of 69.26 percent. Furthermore, the LR algorithm gave an accuracy of 68.42 percent on the test data, sensitivity of 89.33 percent, and specificity of 70.76 percent. RF produced the highest performance across all five metrics: accuracy of 80.01 percent, sensitivity of 90.85 percent, specificity of 85.12 percent, precision of 71.6 percent, and F1-score of 0.85.

The test data indicate that Cohen's Kappa values show fair agreement among several machine learning models: K-NN (0.2077), LDA (0.2214), SVM (0.2143), and LR (0.2321) all show predictions that can be classified as "fair" according to Landis and Koch criteria. Unlike the previous results, random forest exhibits the highest Kappa statistic (0.244) of all five assessment methods, meaning a higher degree of agreement than any other model. The Cohen's kappa of 0.244 denoting fair agreement, indicating an approximate moderate improvement in predictive performance relative to the other models, hints that there are still issues with predictive consistency.

As a result, Random Forest (RF) approach outperformed all other techniques that were used in the machine learning algorithms metrics assessments providing 80.01 percent accuracy rate. The sensitivity of this approach was 90.85 percent, meaning that it effectively detecting a positive case. In addition, it exhibited 85.12 percent specificity, and thus has proven to be efficient in the identification of negative instances. On the other hand, it has a goodness of precision rate at 71.6%, and F1-score at 0.85 showing that this model is performing well in balancing precision with recall.

Figure 2.

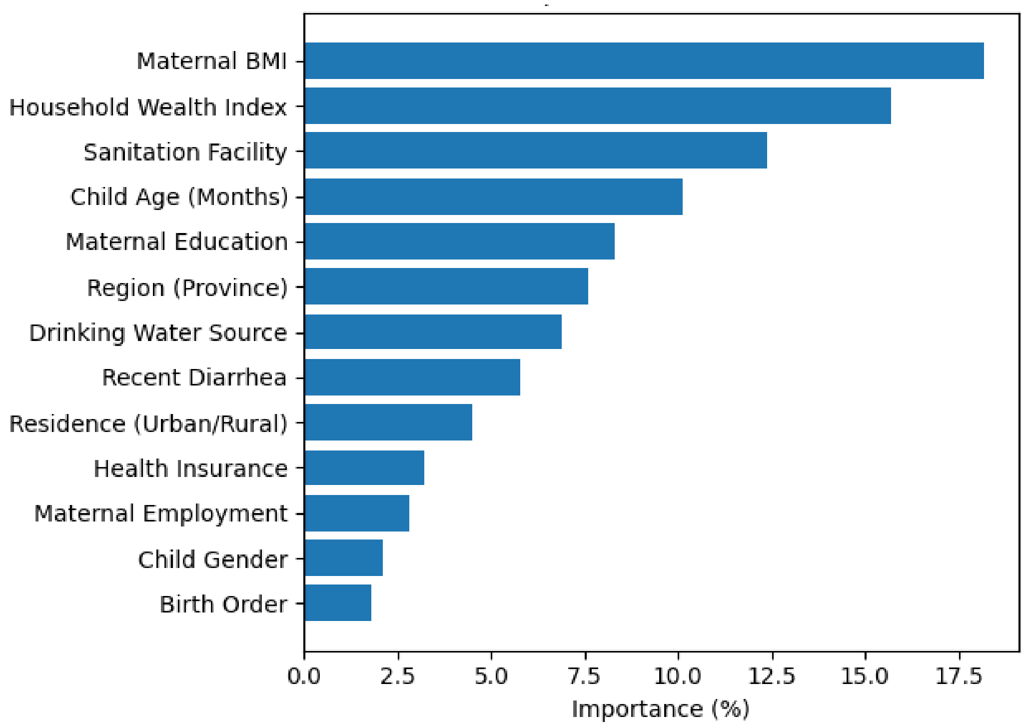
ML model's ROC curves



As shown in Figure 2, Receiver Operating Characteristic (ROC) analysis was performed to evaluate the discriminative power of each model. Area under the curve (AUC) which ranged from 0 to 1, with a higher AUC value corresponding to better quality of the model The Random Forest model was the best performing with an AUC of 0.86 (95% CI: 0.84–0.88). Next was the Support Vector Machine (SVM), which achieved an AUC of 0.81 (95% CI: 0.79–0.83). Logistic Regression and LDA produced AUCs of 0.79 (95% CI: 0.77–0.81) & 0.78 (95% CI: 0.76–0.80), respectively, in comparison to the above results. The K-Nearest Neighbors (KNN) model performed the poorest with an AUC of 0.75 (95% CI: 0.73–0.77). These results demonstrate how ensemble methods can better account for the nonlinear associations present in DHS data. This further demonstrates from the ROC curve that Random Forest achieved the highest area under the AUC, indicating strong classification compared to other models. An AUC of 0.86 is considered to possess strong discriminative ability based on established classification benchmarks, while AUC values from 0.75 to 0.81 are classified as acceptable to fair performance.

Figure 3

Predicting Child Undernutrition with Random Forest Feature Importance



(Note: *Mean decrease impurity was used to determine feature importance from the Random Forest model using Gini importance). The shown values are the normalized totals of each predictor variable's contributions to the model 's overall accuracy, summing to 100 percent. Larger percentages imply a greater contribution to the prediction of child undernutrition, as determined by CIAF (Composite Index of Anthropometric Failure).*

A feature importance analysis was undertaken using the Random Forest model to clarify the most salient contributors of child undernutrition in Pakistan. As per Figure 3 the relative importance of each predictor variable on improving the model's predictive accuracy is shown evaluated using mean decrease in impurityGini important method.

Maternal BMI was the strongest predictor, contributing 17.8 percent of model performance. The household wealth index closely followed and contributed 16.5 percent, while the type of sanitation facility accounted for 12.5 percent. The age of a child in months was fifth, with an importance of 10.0 percent.

Maternal education, region and source of drinking water had moderate importance (9.0%, 7.5% and 6.5%) respectively. Recent episodes of diarrhea (5.5 percent) and type of residence (4.5 percent), among other potentially significant epidemiological drivers, also meaningfully contributed to the model.

Health insurance status (3.5 percent), maternal employment status (2.5 percent), child sex (1.5 percent) and birth order (1.0 percent) were among the weakest predictors, with less than 10 percent of model prediction occurring for any combination of these variables.

Such interventions targeting maternal nutrition, household economic conditions and sanitation infrastructure are implemented in a range of multi-sectoral programs globally and thus these findings suggest that they may have the greatest potential impact on alleviating child undernutrition in Pakistan.

Discussion

The overall findings demonstrate that the RF algorithm gives relatively higher accuracy to assess child malnutrition in Pakistan. The reason behind the RF algorithm higher accuracy achievement in our analysis is that DHS datasets naturally contain a mixture of continuous and categorical variables, and RF natively grips both types of variables without needing pre-processing. Moreover, child malnutrition datasets (especially DHS data) are frequently uneven which lack of symmetry, RF can automatically adjust the balance classes by adjusting weights. Also, DHS datasets habitually encompass raucous socio-economic variables (such as correlated household attributes), RF reduce the power of irrelevant structures and overfitting in predicting child malnutrition. As the DHS datasets are larger in features and sample size, RF efficient in our case of PDHS data. So, the ability of RF algorithm to handle non-linearity, mixed data types, noise, and imbalanced classes while requiring minimal preprocessing makes it particularly suited for predicting child malnutrition on PDHS data. Its ensemble nature and feature selection robustness give it a consistent accuracy advantage.

International investigations identified Random Forest (RF) algorithm as the most efficient substance for that prediction of kid malnutrition. Compelling research evidence from Bangladesh, China, India, Indonesia and the Philippines confirm the superiority of RF models over other machine learning approaches for estimating multiple indicators of malnutrition such as anemia (for children under 5 years), stunting (an anthropometric measure indicating chronic time-dependent malnutrition) and low birth weight (LBW) (Yu et al., 2010; Talukder & Ahammed, 2020; Khan et al., 2019; Hemo & Rayhan, 2021; Reza & Salma, 2024; Van et al., 2022; Gustriansyah et al., 2024; Manongga et al., 2023; Reddy & Dhanalakshmi, 2025; Chen et al., 2025). Globally, these cross-national findings highlight the strength and effectiveness of RF algorithm for health outcomes related to malnutrition.

Investigations in the African context over the past few years have consistently identified RF algorithms as being superior for evaluation of child malnutrition. Ethiopia has explored various machine learning approaches for assessing malnutrition and micronutrient deficiencies at multiple points through the analysis of Demographic and Health Surveys (DHS), consistently demonstrating a superior predictive capability for Random Forest (RF) algorithms over other existing models (Gebeye et al., 2024; Fenta et al., 2021; Zemariam et al., 2024). Similar results were also reported in multi-country studies of Niger, Zambia and East Africa, where RF algorithms were found to be more accurate than other approaches for identifying key indicators associated with acute malnutrition and micronutrient status (Sánchez-Martínez et al., 2024; Ngusie et al., 2024; Chilyabanyama et al., 2022).

Apart from its predictive power, the RF model also provided valuable insights into the most important drivers of child undernutrition in Pakistan. Maternal BMI (17.8 percent), household wealth index (16.5 percent), sanitation facility type (12.5 percent) and child age (10.0 average percentage) were identified among the most informative predictors in a feature importance analysis of the demographic model (Shahid et al., 2025b; Shahid & Naveed, 2026). Maternal BMI is a key determinant and reflects the intergenerational nature of malnutrition, as a mother's undernutrition can lead to negative health events in her offspring (Shahid et al., 2026). While there were some novel findings, such as household wealth and sanitation being identified as important structural determinants, the results emphasized persistent drivers of undernutrition in Pakistan, namely poverty and lack of infrastructure (Tan et al., 2025). The study found that maternal education, geographic region and drinking water quality were moderate predictors of health status. In contrast, factors including maternal occupation, child gender, birth order and health insurance seemed to have little impact. These results highlight the critical need for targeted programming to improve maternal nutrition, strengthen household economic conditions, and enhance sanitation infrastructure in Pakistan. These strategies will be important in achieving large-scale reductions in child undernutrition.

These results, that emphasize the consistency of data dimensionality and Random Forest (RF) risk from the fitted model remained robust across different specifications. Such consistent findings reinforce that RF workup is an effective screening tool for malnutrition. These findings highlight how visualizations of baseline data can help in monitoring and understanding trends in malnutrition which could provide a valuable tool to track over time the prevalence and consequences of malnutrition.

Though Random Forest performs quite well, it is limited in its predictive ability. This limitation predominantly stems from the cross-sectional dimension of the DHS data, which limits both temporal predictions and causal inferences. Moreover, lack of external validation affects its generalizability.

Utilizing longitudinal data and more advanced algorithms such as XGBoost and LightGBM are likely to improve predictive performance in future work.

Conclusion

Under-nutrition among under-five children is a major public health problem in Pakistan. Accurately predicting instances of malnourishment is essential for devising effective preventive strategies. ML algorithms provide fresh perspectives on this complex puzzle, far beyond what would be possible with established econometric approaches. Its focus on Pakistan a malnutrition-prone region — and comprehensive child malnutrition prediction using popular machine learning algorithms (Support Vector Machine (SVM), Logistic Regression (LR), Linear Discriminant Analysis (LDA), Random Forests (RF) and K-Nearest Neighbors, evaluated based on commonly identified risk factors) make this study especially interesting.

The results show that random forest model outperforms other algorithms considerably in predicting malnutrition among children. Feature importance analysis reveals significant predictors of undernutrition including maternal BMI, household wealth, sanitation conditions and child age. This highlights how these models can handle complex heterogeneous health data, eg: the Pakistan Demographic and Health Survey (PDHS) 2018.

Such results suggest and support the promise of machine learning especially Random Forest algorithm in detecting early evidence of malnutrition among children which could further help improve targeted public health interventions. This could guide comprehensive predictive analytics to help policymakers and healthcare practitioners plan for this all-important issue in Pakistan and similar countries suffering from childhood malnutrition. This work adds to the growing literature on machine learning in paediatric health and emphasizes the importance of evidence-based decision-making with data-driven solutions for critical public health issues.

Applied Suggestions from Current Work

Hence, based on these findings, the following practical implications can be suggested:

1. Integration in Health Information System: The effectiveness of the random forest algorithm can be utilized for predictions on health outcomes and, therefore, this approach could potentially be integrated into the DHIS of Pakistan. This could allow for a risk scoring Python module to analyze routine data and identify high-risk children on an ongoing basis. Early identification is so crucial that the LHWs can use this information to act proactively and prevent malnourishment from becoming irreversible stunting.
2. Programmatic Prioritization: Moreover, the feature importance analysis suggests that family interventions targeting nutrition and sanitation are better at capturing outcomes compared to pure child-targeting programs (pre-conception and antenatal Body Mass Index - BMI). Thus, these structural determinants of health need to be prioritized by provincial health departments in high-burden areas like Sindh, Balochistan and FATA to get positive results regarding the health outcomes.
3. Mobile Health (mHealth) Applications: The Random Forest model is also capable of being deployed in lightweight, python based mHealth applications developed to be used for LHWs. This trained model using libraries like joblib or pickle can then be used to generate risk score in real-time during household visits. This is consistent with the early referral of higher-risk children, enabling a swifter response to malnutrition.
4. Future Survey Design: This ranking of feature importance can be used to inform the selection of variables in future Demographic and Health Surveys (DHS). Including measures of dietary diversity scores, metrics surrounding food insecurity, and indicators of maternal mental health—identified previously as significant gaps in the existing dataset—we can significantly improve prediction accuracy.

Limitations and Directions for the Future Research

This study has a number of methodological strengths. Against the backdrop of Pakistan, it provides a systematic comparison of five machine learning algorithms and potentially offers a novel covariate outcome measure (various anthropometric failure using CIAF) instead of discrete indicators. Analysis was based on data from the nationally representative Pakistan Demographic and Health Survey (PDHS), which consisted of a valid sample size of 4,098. Careful elimination of outliers goes a long way toward improving the sample, thus increasing the generalizability of findings. Furthermore, the methodological framework used stratified 10-fold cross-validation with three repetitions, grid search hyper-parameter tuning and techniques to address class imbalance follows best-practice in the use of machine learning applied to health-related research. Scikit-learn-based implementation in Python allows for reproducibility and reduces future deployment efforts.

Nonetheless, several limitations should be recognized. The cross-sectional design of the PDHS limits causal inference as it only provides information about concurrent relationships and not predictiveness. As a result, the model predicts the current malnutrition status rather than risks in the future, making it of limited use for true predictive purposes. Moreover, because there was no external validation using an independent dataset for example PDHS 2022 or individual provincial surveys generalizability beyond the 2017–18 sample is limited. Moreover, PDHS data lack essential potential predictors: dietary diversity scores, household food insecurity, maternal mental health and clinical biomarkers such as hemoglobin levels. While class weighting was used, the moderate class imbalance (44% malnourished) could have nonetheless influenced model performance. While the CIAF provides a good indication of magnitude, it does not take into account severity, duration or recurrence of undernutrition episodes. Finally, the sensitive and power-consuming nature of RF may pose a barrier to their use in resource-limited settings that do not have sufficient infrastructure support.

Highlights

- Random Forest better predicted child malnutrition in Pakistan compared to other machine learning algorithms [accuracy (80.0%) and AUC (0.86)].
- RF based feature importance identified maternal BMI, household wealth, sanitation, and child age as the strongest predictors.
- RF-based risk stratification can inform targeted public health interventions.

Abbreviations

The following abbreviations are used in this manuscript:

ML	Stands for Machine Learning
PDHS	Stands for Pakistan Demographic and Health Survey
NNS	Stands for National Nutritional Survey
LR	Stands for Logistic Regression
LDA	Stands for Linear Discriminant Analysis
SVM	Stands for Support Vector Machines
K-NN	Stands for K-Nearest Neighbors
RF	Stands for Random Forest
MICS	Stands for Multiple Indicator Cluster Survey
CIAF	Stands for Composite Index of Anthropometric Failure (child malnutrition)
WHO	Stands for World Health Organization
MUAC	Stands for Mid-Upper Arm Circumference
BMI	Stands for Body Mass Index
HAZ	Stands for Height-for-Age Z-Scores
WAZ	Stands for Weight-for Age Z-Scores
WHZ	Stands for Weight-for-Height Z-Scores

Declarations:

Ethical approval and consent to participate:

Not applicable. This study used a publicly available secondary data taken from DHS website. No individual-level identifiable information is included, and no consent from individual or from institution for publication is required.

Consent for publication:

Not applicable.

Availability of data and material:

This study utilized the secondary data of the Pakistan Demographic and Health Survey 2017–18. Data is online freely accessible of DHS program website at:
https://dhsprogram.com/data/dataset/Pakistan_Standard-DHS_2017.cfm?flag=1.

Competing interests:

The authors declare no competing interests.

Authors Contributions:

MAY: Conceptualization, Writing-Original Draft, Data Curation. ASS & SJ: Writing-Review, Editing, Resources, Methodology, Software, Formal Analysis, and Visualization. Both authors contributed equally.

Acknowledgment:

The author acknowledges the DHS program of USAID for permitting the general public to download data from the DHS website for research purposes.

Funding:

This research did not receive any funding.

References

Ahsan, M. M., Luna, S. A., & Siddique, Z. (2022). Machine-learning-based disease diagnosis: A comprehensive review. *Healthcare*, *10*(3), Article 541. https://doi.org/10.3390/healthcare10030541

Google Scholar Worldcat Fulltext

Akay, M. F. (2009). Support vector machines combined with feature selection for breast cancer diagnosis. *Expert Systems with Applications*, *36*(2), 3240–3247. https://doi.org/10.1016/j.eswa.2008.01.009

Google Scholar Worldcat Fulltext

Alam, M. B., Shahid, M., Alzghoul, B. I., Yang, J., Zakar, R., Malik, N. I., & Song, J. (2023). The effects of financial stress and household socio-economic deprivation on the malnutrition statuses of children under five during the COVID-19 lockdown in a marginalized region of South Punjab, Pakistan. *Children*, *10*(1), Article 12. https://doi.org/10.3390/children10010012

Google Scholar Worldcat Fulltext

Al-Sadeeq, A. H., Bukair, A. Z., & Al-Saqladi, A. W. (2018). Assessment of undernutrition using Composite Index of Anthropometric Failure among children aged less than 5 years in rural Yemen. *Eastern Mediterranean Health Journal*, *24*(12), 1119–1126. https://doi.org/10.26719/2018.24.12.1119

Google Scholar Worldcat Fulltext

Bashir, M. B., Sadiq, M., Khan, M., Ashfaq, S., Bilal, M., & Khan, M. A. (2021). Prevalence, level and factors associated with malnutrition in children under-five years of age and their parents' awareness about children nutrition in Quetta city. *Open Journal of Pediatrics and Child Health*, *6*, 20–25. https://doi.org/10.17352/ojpch.000033

Google Scholar Worldcat Fulltext

Chen, K., Zheng, F., Zhang, X., Ma, D., Zeng, J., Ye, Q., et al. (2025). Factors associated with underweight, overweight, and obesity in Chinese children aged 3–14 years using ensemble learning algorithms. *Journal of Global Health*, *15*, Article 04013. https://doi.org/10.7189/jogh.15.04013

Google Scholar Worldcat Fulltext

Chilyabanyama, O. N., Chilengi, R., Simuyandi, M., Velu, R. M., Beres, L. K., Mwila-Kazimbaya, K., et al. (2022). Performance of machine learning classifiers in classifying stunting among under-five children in Zambia. *Children*, *9*(7), Article 1082. https://doi.org/10.3390/children9071082

Google Scholar Worldcat Fulltext

Côté, M., Osseni, M. A., Brassard, D., Bélanger, M., Provencher, V., Robitaille, J., & Lemieux, S. (2022). Are machine learning algorithms more accurate in predicting vegetable and fruit consumption than traditional statistical models? An exploratory analysis. *Frontiers in Nutrition*, *9*, Article 740898. https://doi.org/10.3389/fnut.2022.740898

Google Scholar Worldcat Fulltext

Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, *13*(1), 21–27. https://doi.org/10.1109/TIT.1967.1053964

Google Scholar Worldcat Fulltext

Curtis, L. (2006). Encyclopedia of Human Nutrition (2nd Edition) [Book review]. *Reference Reviews*, *20*(7), 39–40. https://doi.org/10.1108/09504120610655547

Google Scholar Worldcat Fulltext

Fanzo, J., Hawkes, C., Udomkesmalee, E., Afshin, A., Allemandi, L., Assery, O., & Afshin, A. (2018). *Global nutrition report: Shining a light to spur action on nutrition*. Development Initiatives. https://globalnutritionreport.org/reports/global-nutrition-report-2018/

Google Scholar Worldcat Fulltext

Fenta, H. M., Zewotir, T., & Muluneh, E. K. (2021). A machine learning classifier approach for identifying the determinants of under-five child undernutrition in Ethiopian administrative zones. *BMC Medical Informatics and Decision Making*, *21*, Article 291. https://doi.org/10.1186/s12911-021-01652-1

Google Scholar Worldcat Fulltext

Gaber, T., Hassanien, A. E., El-Bendary, N., & Dey, N. (Eds.). (2016). *The 1st International Conference on Advanced Intelligent System and Informatics (AISI2015), November 28–30, Beni Suef, Egypt*. Springer. https://doi.org/10.1007/978-3-319-26690-9

Google Scholar Worldcat Fulltext

Gebeye, L. G., Dessie, E. Y., & Yimam, J. A. (2024). Predictors of micronutrient deficiency among children aged 6–23 months in Ethiopia: A machine learning approach. *Frontiers in Nutrition*, *10*, Article 1277048. https://doi.org/10.3389/fnut.2023.1277048

Google Scholar Worldcat Fulltext

Gustriansyah, R., Suhandi, N., Puspasari, S., Widyaningsih, P., & Nugroho, A. M. (2024). Machine learning method to predict the toddlers' nutritional status. *Jurnal Infotel*, *16*(1), 32–43. https://doi.org/10.20895/infotel.v15i4.988

Google Scholar Worldcat Fulltext

Hemo, S. A., & Rayhan, M. I. (2021). Classification tree and random forest model to predict under-five malnutrition in Bangladesh. *Biostatistics and Biometrics International Journal*, *10*(4), 116–123. https://medcraveonline.com/BBIJ/BBIJ-10-00337.pdf

Google Scholar Worldcat Fulltext

Iqbal, M., Fatmi, Z., Khan, K. S., Iqbal, M. P., & Iqbal, S. P. (2020). Malnutrition and food insecurity in child labourers in Sindh, Pakistan: A cross-sectional study. *Eastern Mediterranean Health Journal*, *26*(9), 1087–1096. https://applications.emro.who.int/emhj/v26/09/1020-3397-2020-2609-1087-1096-eng.pdf

Google Scholar Worldcat Fulltext

Issa, M., Muntaha, S. T., & Ahmad, R. (2024). Prevalence of malnutrition in children under 5 years of age in Batkhela Tehsil, KPK, Pakistan. *Journal of Islamic International Medical College*, *19*, 24–28. https://journals.riphah.edu.pk/index.php/jiimc/article/view/1837/1099

Google Scholar Worldcat Fulltext

Khaliq, A., Wraith, D., Miller, Y., & Nambiar-Mann, S. (2021). Prevalence, trends, and socioeconomic determinants of coexisting forms of malnutrition amongst children under five years of age in Pakistan. *Nutrients*, *13*(12), Article 4566. https://doi.org/10.3390/nu13124566

Google Scholar Worldcat Fulltext

Khan, G. N., Turab, A., Khan, M. I., Rizvi, A., Shaheen, F., Ullah, A., & Bhutta, Z. A. (2016). Prevalence and associated factors of malnutrition among children under-five years in Sindh, Pakistan: A cross-sectional study. *BMC Nutrition*, *2*, Article 69. https://doi.org/10.1186/s40795-016-0112-4

Google Scholar Worldcat Fulltext

Khan, J. R., Chowdhury, S., Islam, H., & Raheem, E. (2019). Machine learning algorithms to predict the childhood anemia in Bangladesh. *Journal of Data Science*, *17*(1), 195–218. https://doi.org/10.6339/JDS.201901_17(1).0009

Google Scholar Worldcat Fulltext

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

Google Scholar Worldcat Fulltext

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, *2*(3), 18–22. https://journal.r-project.org/articles/RN-2002-022/RN-2002-022.pdf

Google Scholar Worldcat Fulltext

Mahmood, T., Abbas, F., Kumar, R., & Somrongthong, R. (2020). Why under five children are stunted in Pakistan? A multilevel analysis of Punjab Multiple Indicator Cluster Survey (MICS-2014). *BMC Public Health*, *20*, Article 952. https://doi.org/10.1186/s12889-020-09110-9

Google Scholar Worldcat Fulltext

Manongga, S. P., Hendry, H., & Manongga, D. H. (2023). Predicting a stunting prevalence using semi-supervised learning models in East Nusa Tenggara. *International Journal of Innovative Computing, Information and Control*, *19*(4), 1073–1086. http://www.ijicic.org/ijicic-190405.pdf

Google Scholar Worldcat Fulltext

Menghwar, B., Laghari, Z. A., Memon, S. F., Balouch, B., & Shah, T. (2022). Prevalence of malnutrition in children under five years' age in district Tharparkar Sindh, Pakistan. *Journal of the Pakistan Medical Association*, *72*, 33–36. https://doi.org/10.47391/JPMA.20-540

Google Scholar Worldcat Fulltext

Mustufa, M. A., Jamali, A. K., Sameen, I., Burfat, F. M., Baloch, M. Y., Baloch, A. H., & Aslam, M. (2017). Malnutrition and poor oral health status are major risks among primary school children at Lasbela, Balochistan, Pakistan. *Journal of Health, Population and Nutrition*, *36*, Article 17. https://doi.org/10.1186/s41043-017-0100-6

Google Scholar Worldcat Fulltext

Ngusie, H. S., Enyew, E. B., Walle, A. D., Kassie, S. Y., & Bogale, B. (2024). Employing machine learning techniques for prediction of micronutrient supplementation status during pregnancy in East African countries. *Scientific Reports*, *14*, Article 23827. https://doi.org/10.1038/s41598-024-75455-5

Google Scholar Worldcat Fulltext

Olukemi, F. S. (2014). Assessing the prevalence of undernutrition using the Composite Index of Anthropometric Failure (CIAF) among young children in a rural community in Oyo State, Nigeria.

Nigerian Journal of Nutritional Sciences, 35, 24–29. https://www.ajol.info/index.php/njns/article/view/166482

Google Scholar Worldcat Fulltext

Rahman, Z. U., Gul, R., Khalil, K. U., Khan, A., Khan, A., & Rasheed, S. (2024). Prevalence and predictors of under nutrition among under-five children in Peshawar district, Khyber Pakhtunkhwa, Pakistan. *Journal of Saidu Medical College*, *14*(1), 51–56. https://doi.org/10.52206/jsmc.2024.14.1.818

Google Scholar Worldcat Fulltext

Rastogi, S., Maheshwari, C., Raghav, S. K., Sharma, D., Lala, M. K., & Nayak, S. (2017). Assessing burden of under-nutrition among under five children of urban slum by using Composite Index of Anthropometric Failure in Ahmedabad city, Gujarat, India. *National Journal of Community Medicine*, *8*, 496–500. https://njcmindia.com/index.php/file/article/view/1244

Google Scholar Worldcat Fulltext

Reddy, A., & Dhanalakshmi, R. (2025). An innovative approach for predicting child mortality using random forest algorithm with Naive Bayes algorithm. *AIP Conference Proceedings*, *3252*, Article 020001. https://doi.org/10.1063/5.0259497

Google Scholar Worldcat Fulltext

Reza, T. B., & Salma, N. (2024). Prediction and feature selection of low birth weight using machine learning algorithms. *Journal of Health, Population and Nutrition*, *43*, Article 157. https://doi.org/10.1186/s41043-024-00647-8

Google Scholar Worldcat Fulltext

Roy, K., Dasgupta, A., Roychoudhury, N., Banerjee, R., Naskar, T., & Pal, R. (2018). Assessment of under nutrition with Composite Index of Anthropometric Failure (CIAF) among under-five children in a rural area of West Bengal, India. *International Journal of Contemporary Pediatrics*, *5*(5), 1651–1656. https://doi.org/10.18203/2349-3291.ijcp20182583

Google Scholar Worldcat Fulltext

Saheed, R., Shahid, M., Wang, J., Qureshi, M. G., Sun, X., Bibi, A., & Song, J. (2022). Impact of drinking water source and sanitation facility on malnutrition prevalence in children under three: A gender-disaggregated analysis using PDHS 2017–18. *Children*, *9*(11), Article 1674. https://doi.org/10.3390/children9111674

Google Scholar Worldcat Fulltext

Sánchez-Martínez, L. J., Charle-Cuéllar, P., Gado, A. A., Salse-Ubach, N., Hurtado-Avilés, J. A., Siza-Mboi, N., et al. (2024). Using machine learning to fight child acute malnutrition and predict weight gain during outpatient treatment with a simplified combined protocol. *Nutrients*, *16*(23), Article 4213. https://doi.org/10.3390/nu16234213

Google Scholar Worldcat Fulltext

Shahid, M., Ahmed, F., Ameer, W., Guo, J., Raza, S., Fatima, S., & Song, J. (2022d). Prevalence of child malnutrition and household socioeconomic deprivation: A case study of marginalized district in Punjab, Pakistan. *PLoS ONE*, *17*(1), Article e0263470. https://doi.org/10.1371/journal.pone.0263470

Google Scholar Worldcat Fulltext

Shahid, M., Ameer, W., Malik, N. I., Alam, M. B., Ahmed, F., Qureshi, M. G., & Song, J. (2022b). Distance to healthcare facility and lady health workers' visits reduce malnutrition in under five children: A case study of a disadvantaged rural district in Pakistan. *International Journal of Environmental Research and Public Health, 19*(13), Article 8200. https://doi.org/10.3390/ijerph19138200

Google Scholar Worldcat Fulltext

Shahid, M., Cao, Y., Ahmed, F., Raza, S., Guo, J., Malik, N. I., & Song, J. (2022c). Does mothers' awareness of health and nutrition matter? A case study of child malnutrition in marginalized rural community of Punjab, Pakistan. *Frontiers in Public Health, 10*, Article 792164. https://doi.org/10.3389/fpubh.2022.792164

Google Scholar Worldcat Fulltext

Shahid, M., Cao, Y., Shahzad, M., Saheed, R., Rauf, U., Qureshi, M. G., & Song, J. (2022e). Socio-economic and environmental determinants of malnutrition in under three children: Evidence from PDHS-2018. *Children, 9*(3), Article 361. https://doi.org/10.3390/children9030361

Google Scholar Worldcat Fulltext

Shahid, M., Liu, Y., Ameer, W., Qureshi, M. G., Ahmed, F., & Tang, K. (2022a). Comparison of different nutritional screening approaches and the determinants of malnutrition in under-five children in a marginalized district of Punjab province, Pakistan. *Children, 9*(7), Article 1096. https://doi.org/10.3390/children9071096

Google Scholar Worldcat Fulltext

Shahid, M., Xie, Y., Bashir, S., Noureen, N., Song, J., Malik, N. I., & Almas, A. (2024). Association among household wealth, maternal employment, and undernutrition in children under three years of age in Pakistan. *Children, 11*(7), Article 872. https://doi.org/10.3390/children1107872

Google Scholar Worldcat Fulltext

Shahid, M., Ali, Z., Khan, S., Yousaf, M. S., Zhang, Z., & Song, J. (2025a). Exploring the potential causal relationship between health insurance coverage and child nutritional status in Pakistan: Evidence from PDHS-2018. *Healthcare, 13*(5), Article 532. https://doi.org/10.3390/healthcare13050532

Google Scholar Worldcat Fulltext

Shahid, M., Yahya, M. A., Song, J., et al. (2025b). Machine learning vs. traditional logistic regression: Predictive performance and risk factor identification for child nutritional outcome in Pakistan. *BMC Public Health, 25*, Article 113. https://doi.org/10.1186/s12889-025-25621-9

Google Scholar Worldcat Fulltext

Shahid, M., & Naveed, H. M. (2026a). Deep neural network modeling of sustainable development and human development indices on child undernutrition in Pakistan. *Sustainable Development, 42*, 1–31. https://doi.org/10.1002/sd.70845

Google Scholar Worldcat Fulltext

Shahid, M., Naveed, H. M., Song, J., Yuksel, S., Dincer, H., & Ali, M. (2026b). The law of diminishing marginal returns in health insurance: Evidence on child health outcomes from Pakistan's Sehat Sahulat Program. *Frontiers in Public Health, 13*, Article 1729440. https://doi.org/10.3389/fpubh.2025.1729440

Google Scholar Worldcat Fulltext

Sharma, V., Sharma, V., Khan, A., Wassmer, D. J., & Shabir, S. (2020). Malnutrition, health and the role of machine learning in clinical setting. *Frontiers in Nutrition, 7,* Article 44. https://doi.org/10.3389/fnut.2020.00044

Google Scholar Worldcat Fulltext

Song, F., Mei, D., & Li, H. (2010). Feature selection based on linear discriminant analysis. In *Proceedings of the 2010 International Conference on Intelligent System Design and Engineering Application* (pp. 746–749). IEEE. https://ieeexplore.ieee.org/document/5743287

Google Scholar Worldcat Fulltext

Talukder, A., & Ahammed, B. (2020). Machine learning algorithms for predicting malnutrition among under-five children in Bangladesh. *Nutrition, 78,* Article 110861. https://doi.org/10.1016/j.nut.2020.110861

Google Scholar Worldcat Fulltext

Tan, L., Shahid, M., Song, J., Naveed, H. M., & Khan, I. (2025). Wealth and regional disparities in child undernutrition: Insights from national demographic and health survey. *Frontiers in Public Health, 13*, Article 1654403. https://doi.org/10.3389/fpubh.2025.1654403

Google Scholar Worldcat Fulltext

Van, V. T., Antonio, V. A., Siguin, C. P., Verceles, A. C., Felix, M. B., de la Cruz, A. B., et al. (2022). Predicting undernutrition among elementary schoolchildren in the Philippines using machine learning algorithms. *Nutrition, 96*, Article 111571. https://doi.org/10.1016/j.nut.2021.111571

Google Scholar Worldcat Fulltext

Wali, N., Agho, K., & Renzaho, A. M. (2019). Past drivers of and priorities for child undernutrition in South Asia: A mixed methods systematic review protocol. *Systematic Reviews, 8,* Article 1. https://doi.org/10.1186/s13643-019-1112-7

Google Scholar Worldcat Fulltext

World Health Organization. (2019). *Nutrition landscape information system (NLIS) country profile indicators: Interpretation guide.* World Health Organization. https://www.who.int/publications/i/item/9789241516952

Google Scholar Worldcat Fulltext

World Health Organization, & United Nations Children's Fund. (2009). *WHO child growth standards and the identification of severe acute malnutrition in infants and children: A joint statement*. World Health Organization. https://apps.who.int/iris/bitstream/handle/10665/44129/9789241598163_eng.pdf

Google Scholar Worldcat Fulltext

Yu, W., Liu, T., Valdez, R., Gwinn, M., & Khoury, M. J. (2010). Application of support vector machine modeling for prediction of common diseases: The case of diabetes and pre-diabetes. *BMC Medical Informatics and Decision Making, 10,* Article 16. https://doi.org/10.1186/1472-6947-10-16

Google Scholar Worldcat Fulltext

Zemariam, A. B., Adisu, M. A., Habesse, A. A., & Shiferaw, K. E. (2024). Employing advanced supervised machine learning approaches for predicting micronutrient intake status among children aged 6–23 months in Ethiopia. *Frontiers in Nutrition, 11,* Article 1397399. https://doi.org/10.3389/fnut.2024.1397399

Google Scholar Worldcat Fulltext