International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

A Ridge-Based Multivariate Logistic Modelling for Assessing Factors Associated with Multifactorial Determinants of Under Five Child Malnutrition Outcomes in Mozambique

Adelino Benedito Nhancale[1], Fredirick Onyango[2], Joyce Otieno[3]

[1]PhD in Applied Statistics (ongoing), MSc in Statistics and Information Management, Lecturer at Eduardo Mondlane University – Faculty of Economics, Mozambique.
[2,3]PhD, Department of Statistics & Actuarial Science, Maseno University, School of Mathematics, Statistics and Actuarial Science, Kenya.

Abstract

In Mozambique, individual and integrated programmes targeting the healthcare, sanitation, socioeconomic sectors, and nutrition still leave child malnutrition, specifically within the under-five category, as a critical public health challenge. This study sought to discern the most important predictors of child malnutrition using DHS-aligned variables designed to guide focused strategies. A Ridge Logistic Regression model was conducted to analyse data from the nationally representative Mozambique Demographic and Health Survey (MDHS2022) to predict a binary composite malnutrition outcome—defined by the presence of at least one of three conditions: stunting, underweight, or wasting, or anthropometric Z-scores (< -2 SD). Independent variables included child-specific and household demographic, maternal, and reproductive health, education, socioeconomic status, water, and sanitation. Even though the model was fitted based on a binary composite outcome, classification metrics were retained, ROC curves, precision-recall curves, and goodness-of-fit diagnostics performed separately for each type of malnutrition. The findings showed the best model performance for wasting (AUC = 0.99; Accuracy = 0.9317), followed by underweight (AUC = 0.86) and stunting (AUC = 0.63). Other significant factors included male gender, child age 24-35 and 48-59 months, low birthweight, rural dwelling, poor sanitation, inadequate shelter, lower BMI z-score, and low standing.

Acute environmental factors were associated with wasting and showed high model stability, while underweight reflected a mix of acute and chronic drivers with moderate predictive performance. Stunting proved harder to classify due to its association with structural deprivation. In any case, utilising Ridge Logistic Regression revealed underlying issues affixed to each form of malnutrition in conjunction with DHS-aligned predictors and advanced the understanding of distinct risk factors, which affirmed the model's utility—even with respect to wasting—for policy and intervention strategy formulation guided by empirical data.

Keywords: Child malnutrition, Ridge Logistic Regression, DHS, Mozambique, stunting, underweight, wasting, public health

Introduction

As stated by Black and other scholars in 2013, along with the 2021 report from UNICEF and the World Health Organization, child malnutrition remains an issue in Mozambique alongside something known as child morbidity. Child malnutrition harms children's physical and mental wellbeing, posing a threat to child development. Mozambique is well known for being impoverished with underdeveloped industries and limited government programmes to help decrease malnutrition. Akombi and others in 2017, along with Victora and others in 2021, show that the lack of resources and insufficient systems to help children leads to malnutrition and other illnesses due to the lack of basic needs. These issues lead to social inequalities and worsen the prospects for future generations even more.

From what we gathered, the situation of malnutrition in Mozambique is complex and multifactorial. Some contributing factors include undernutrition, poor housing, illiteracy among women, insufficient healthcare services, and the lack of clean water and sanitation facilities. Other issues like poverty, maternal care, housing, and nutrition also contribute to the problem (Masibo & Makoka, 2012; Hossain et al., 2019; Cardoso et al., 2006). Moreover, demographic characteristics of children such as their age and sex are known to impact their likelihood of suffering from malnutrition. To these factors must be added the rural-urban divide, where geography limits healthcare access, and consequently, healthcare outcomes are worse for rural children (Kandala et al., 2011).

Understanding the predictors of malnutrition in Mozambique is important both in terms of theory and efficient policy formulation that is well targeted. Most earlier studies relied on simple logistic models or even univariate analyses, which are lacking because of their simplistic treatment of multicollinear high dimensional predictors (Zou & Hastie, 2005; Bhutta et al., 2013; Faye et al., 2019). On the contrary, Ridge Logistic Regression slightly relaxes these traditional bounds by incorporating a regularisation term which stabilises coefficient estimates in the presence of multicollinearity (Hoerl & Kennard, 1970; Hirpa et al., 2020; Namara & Golitko, 2022).

As a standardised framework for child health assessments, this paper adopts the structure and variable arrangements of the Demographic and Health Survey (DHS). The variables are divided into Child Demographic and Biological Characteristics, Maternal and Child Health, Education, Socioeconomic Status and Wealth Proxy, Household, Water and Sanitation. This alignment with internationally set reporting principles facilitates multinational research cooperation and at the same time explains important policy-oriented results (UNICEF, WHO, & World Bank Group, 2021; Alderman & Headey, 2017).

Following this approach, we examine the factors associated with stunting (chronic malnutrition), underweight (intermediate malnutrition), and wasting (acute malnutrition). Our objective is to geographically and chronologically parse each condition, model Diva's discrepancies, as well as recommend programme and policy interventions in the context of Mozambique.

Literature Review

The issue of child malnutrition in low and middle economies, particularly in Sub-Saharan Africa, is well documented. Based on Black et al.'s research, the early malnourishment of children leads to high mortality rates alongside cognitive and developmental potential losses which perpetuate a cycle of impoverished society (2013). UNICEF, WHO, and the World Bank (2021) have consistently reported Mozambique to be among the highest in child stunting and underweight rates. Stunting, as noted by Victora et al., is a reflection of deep-rooted poverty, poor living standards, and a general lack of caretaking (2021). The work of Akombi et al. (2017) and Kandala et al. (2011) identified education of the mother, birthweight, rurality of residence, and sanitation as recurrently determining health outcomes in Africa. Masibo and Makoka (2012) attributed the decline of stunting and underweight in East and Southern Africa to the increased educational attainment of women, which is also linked to greater household income and enhanced nutritional outcomes as suggested by Alderman and Headey (2017). Additionally, Bhutta et al. (2013) showed how combined maternal and child health interventions tailored to be nutrition-sensitive and specific significantly reduce undernutrition.

The Mozambique DHS data has individual-level malnutrition correlates, but most studies seem to centre on stunting, underweight, and wasting, modelling them via sparse predictive modelling. Multicollinearity is a well-known problem affecting traditional logistic regression, resulting in unstable estimates and overfitting (Hoerl & Kennard, 1970). These constraints have led to the successful development of regularisation techniques, such as Ridge regression. Commenting on the application of regularisation, Zou and Hastie (2005) pointed out its usefulness in health research because of the use of interdependent predictors and their associations. Ridge Logistic Regression is beneficial for malnutrition modelling since it reduces overfitting and improves generalisability to new data, as shown by Hirpa et al. (2020), who demonstrated the benefit of Ridge and Lasso over standard logistic regression for predicting undernutrition in Ethiopia. Also, robust and interpretable findings were reported by Namara and Golitko (2022), who used Ridge regression to analyse the risk of malnutrition associated with hygiene practices.

Machine learning and regularisation techniques have been found useful in this regard in recent research. For example, Tiwari et al. (2021) applied sophisticated machine learning techniques to analyse childhood wasting in South Asia, and Gebretsadik et al. (2022) used penalised regression methods to forecast stunting in Ethiopia. In addition to these findings, Goyal and Soni (2021) described the improvement of interpretability in health analytics through the application of Ridge regression, and Ogunjimi et al. (2020) extended these analyses to determinants of maternal health. Faye et al. (2019) also advocate the application of statistical learning for public health nutrition, and Hossain et al. (2019) used multivariate approaches to show the association of household environmental factors with malnutrition in South Asia. Further, Khulu et al. (2021) studied the socio-environmental predictors of stunting in Southern Africa, and Tarekegn et al. (2021) emphasised the significance of predictive modelling for health planning in early childhood.

This study expands on these contributions by using Ridge Logistic Regression principles on DHS data from Mozambique to address malnutrition in the form of stunting, underweight, and wasting while addressing multicollinearity and model overfitting. Accurate consideration of the DHS-aligned predictors along with strong model-building checks enables this model to integrate the myriad factors shaped by biological, maternal, socioeconomic, environmental, and geographic constituents. With this, it fills an important void in the Mozambican literature and strengthens the policy impact of malnutrition modelling in resource-poor countries.

Methodology

The scope of this research focuses specifically on risk factors contributing to childhood malnutrition in Mozambique using a cross-sectional analytical approach. It is based on representative survey data and adheres to the thematic outline utilised by the Demographic and Health Surveys (DHS). This research concentrates on three types of malnutrition: stunting, underweight, and wasting, which are each treated as binary outcome variables.

Data Source and Sample: The data is derived from the last DHS conducted in Mozambique, with 3953 of sample size, which provided comprehensive data on child health as well as maternal, household, and environmental services. The final analytic dataset comprises children under five years old with complete anthropometric measurements and relevant background variables.

Variable Classification: To include the most relevant factors contributing to child malnutrition, explanatory variables were consolidated into six major groups using the standard framework of Demographic and Health Surveys (DHS). Age (captured as < 12, 12–23, 24–35, 36–47 and 48–59 months), sex: male or female, and survival status: alive or deceased are classified as biological and demographic features of a child. Along with birthweight (kg) as an example of a critical early-life health indicator, anthropometric measurements of weight (kg), height (cm), and BMI-for-age Z scores were also included. A mother's weight (kg), childbirth history (number of births in the past five years), and nursing status indicate maternal and reproductive health factors associated with caregiving and maternal welfare. Furthermore, marital status alongside perceived barriers to obtaining health care, in terms of distance from the nearest medical facility, was also included. Education-related factors included years of schooling and the highest qualification attained for the mother or caregiver. These proxy variables reflect health and childcare literacy. The DHS wealth index, a composite measure of household assets divided by urban and rural residence, served to evaluate socioeconomic status and household wealth.

The additional criteria of the respondent's economic activity and possession of particular articles (electricity, radio, television, refrigerator, bicycle, motorcycle, and car/truck) also furthers geographic economic inequalities. Demographic and structural household characteristics were the number of children under five, household size, and type of residence (rural or urban). Geographically, the location also added to the dwelling's description concerning its construction quality, which was measured through materials used for the roof, walls, and floors. Lastly, environmental determinants are assessed by the WASH variables. These included the main source of drinking water, the distance to fetch water, and the type of toilet facility. Collectively, these factors enabled an in-depth analysis of the multifaceted factors contributing to child malnutrition in Mozambique.

Modelling Strategy: Separate Ridge Logistic Regression models were constructed for stunting, underweight, and wasting. All predictors were standardised where necessary. The outcome variables were binary indicators representing whether a child was stunted, underweight, or wasted based on standard anthropometric thresholds (e.g., Z-scores below –2 SD).

Validation and Diagnostics: A deep validation framework was employed to ensure reliable model performance and interpretability. In the evaluation of the model, its performance was diagnosed using classification metrics such as F1-score, recall, accuracy, and precision. In order to visualise the sensitivity versus specificity trade-off, complemented by possible class imbalance in malnutrition data, ROC curves and precision-recall curves were added to these metrics. Calibration plots were utilised to measure the degree in which the model's confidence reliably corresponded to its actual performance by comparing predicted probabilities and observed frequencies. Furthermore, the statistical measures of model fit included deviance, Bayesian Information Criterion (BIC), and Akaike Information Criterion (AIC) all to ensure that the subsequently explained recommend in proportion to the complexity of the model provided optimal fit. As to diagnostic checks for non-normality or heteroscedasticity, these included residuals histograms, while leverage plots along with Cook's distance sought significant observations with potentially excessive influence on the model estimates. To combat overfitting and enhance generalisability, bootstrapping and cross-validation techniques were systematically applied. Through these resampling methods, model stability across data subsets and performance metric confidence interval estimates were achieved. Considering the high prevalence of malnutrition in Mozambique, this mixed-methods framework ensured equal consideration of interpretability and predictive precision to maximise health intervention impact.

Ridge Logistic Regression Framework

Ridge Logistic Regression is an adaptation of logistic regression that incorporates regularisation by adding a penalty term that increases the loss function, reducing the size of the regression coefficients. This is very useful in cases of high dimensions and the existence of multicollinearity between the predictors, which is common with socioeconomic and demographic variables from DHS surveys.

Ridge regression is characterised by retaining all variables wherein instead of removing them, their coefficients are diminished.

With the application of an L2 penalty, Ridge regression drags coefficients towards zero but leaves them slightly above it. While the impact of the parameters is minimised, all remain in the model. This is particularly helpful when dealing with multicollinearity among predictors.

The regularised loss function of Ridge Logistic Regression is expressed as:

$$L(\beta) = -\sum_{i=1}^{n}[y_i log(p_i) + (1 - y_i) log(1 - p_i)] + \lambda \sum_{j=1}^{p} \beta_j^2$$

Where:

- y_i is the observed binary response (1 if malnourished, 0 otherwise).
- $p_i = 1/\left(1 + e^{-x_i{}^T\beta}\right)$ is the predicted probability from the logistic function.
- β_i are the regression coefficients
- λ is the regularisation parameter controlling the strength of the penalty.

The penalty term $\lambda \sum_{j=1}^{p} \beta_j^2$ is essential for reducing model variance and mitigating multicollinearity which improves model stability and generalisability. Ridge regression's properties allow the model to maintain stability even when predictors with strong correlations are removed from the equation.

In this study, hyperparameter optimisation was performed with cross-validation for selecting an appropriate value of λ. As part of model training, all features were centred and scaled to have a mean of zero and unit standard deviation, ensuring coefficients received equal penalties. Interpretation focused on the predicted directional and relative strength of effects using exponentiated coefficients (odds ratio, $OR = e^{\beta_i}$), noting that statistical significance is not evaluated in Ridge due to the imposed shrinkage penalty.

The ability of ridge regression to handle multicollinearity along with a large number of predictors makes it particularly suited for public health modelling where biological, economic, and environmental factors intertwine and are multicollinear. This approach further improves the robustness and accuracy of predictions for various phenotypes of malnutrition.

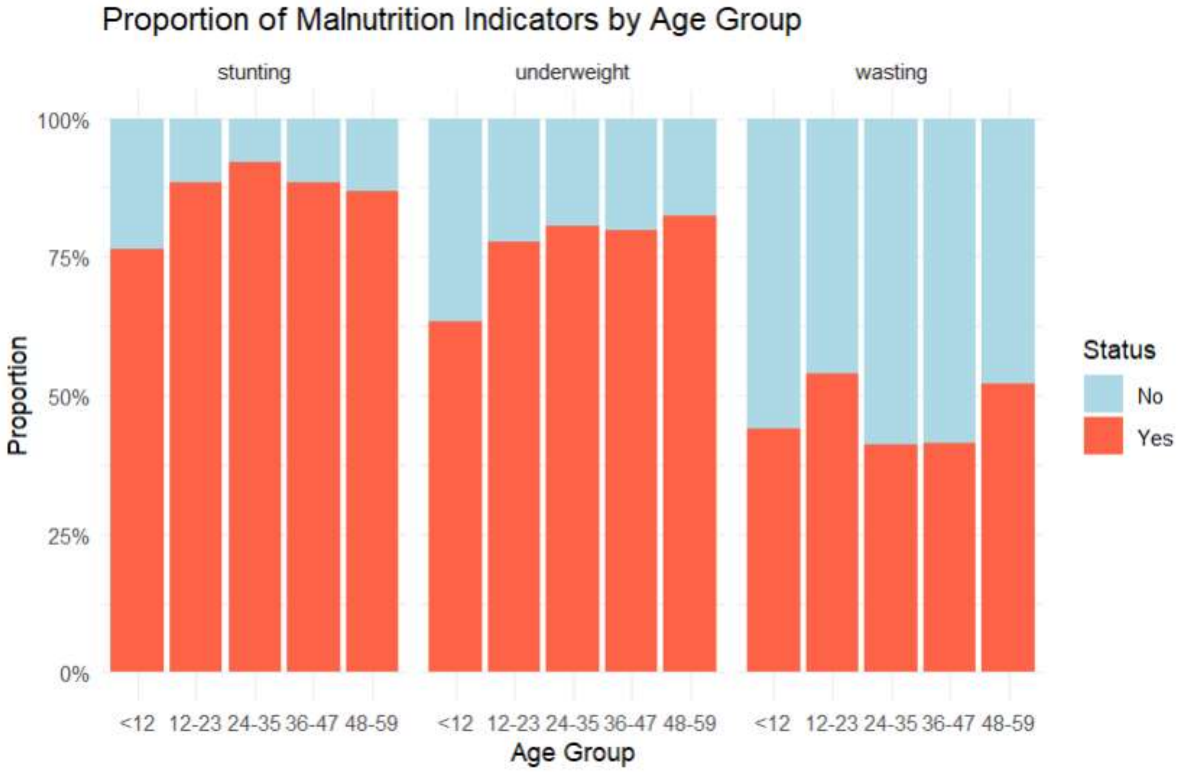
Descriptive Statistics, Data Analysis and Discussion

A probing age-specific examination of children under five revealed distinct patterns of malnutrition. They are visually displayed using a bar chart depicting the prevalence of stunting, underweight, and wasting within each age group. Furthermore, Ridge Logistic Regression models for the classification of children's malnutrition into three categories: stunting, underweight, and wasting are provided together with their corresponding results. The analysis is organised by malnutrition type and includes interpretation of the model coefficients, ORs, OR diagnostics, and classification performance. These interpretations are, however, still related to variable groupings based on the DHS surveys.

International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Descriptive Statistics

Figure 1: Bar chart of the Proportion of Malnutrition Indicators by Age Group



The bar chart “Proportion of Malnutrition Indicators by Age Group” displays the three forms of malnutrition—stunting, underweight and wasting—over six child age periods: <12 months, 12–23, 24–35, 36–47, and 48–59 months. For each category of malnutrition, the bars are separated into two colours where red indicates proportion affected (“Yes”), while blue represents proportion not affected (“No”). The chart indicates the highest prevalence of stunting in all age groups, especially in the 24–35 months where about 85-90% of children were affected. There is also a significantly high prevalence of underweight which increases steadily with age; from approximately 65% in sub 12 months old children to almost 85% in the 48-59 months group. Conversely, wasting shows a more variable pattern with a low nadir in children 24-47 months. These changes suggest that although stunting and underweight are more chronic and persistent conditions, wasting is more fluctuant and reflects more of an acute short-term deficiency of nutrition varying more with age.

Feature	Level	FreqPerc	Coefficient	OddsRatio	FeatureImportance
Child Demographic and Biological Characteristics					
age_group: age_in_months (Reference Category: <12)	<12	897 (22.7%)			
	48-59	715 (18.1%)	1.8189	6.1652	1.8189
	24-35	809 (20.5%)	1.5216	4.5793	1.5216
	36-47	736 (18.6%)	1.3297	3.7800	1.3297
	12-23	796 (20.1%)	0.7320	2.0793	0.7320
alive (Reference Category: 0)	0	64 (1.6%)			
	1	3889 (98.4%)	0.1276	1.1361	0.1276
height			-0.0054	0.9946	0.0054
weight			-0.2202	0.8024	0.2202
sex (Reference Category: 1)	1	1972 (49.9%)			
	2	1981 (50.1%)	-0.3750	0.6873	0.3750
bmi_z_score			-0.4567	0.6334	0.4567
birthweight			-0.7098	0.4918	0.7098
Maternal and Reproductive Health					
maritalstatusrespond (Reference Category: 0)	0	223 (5.6%)			
	1	1252 (31.7%)	0.1256	1.1338	0.1256
	5	328 (8.3%)	0.0857	1.0895	0.0857
	4	86 (2.2%)	0.0225	1.0227	0.0225
	2	1999 (50.6%)	-0.0814	0.9218	0.0814
	3	65 (1.6%)	-0.1025	0.9026	0.1025
hospitaldistance (Reference Category: 1)	1	1716 (43.4%)			
	2	2237 (56.6%)	0.0100	1.0101	0.0100
weightrespondant			-0.0148	0.9853	0.0148
numbirthslastfiveyears			-0.0920	0.9121	0.0920
breastfeedingstill (Reference Category: 0)	0	1816 (45.9%)			
	1	2137 (54.1%)	-0.1877	0.8289	0.1877
Education					
edsingleyears			0.0311	1.0315	0.0311
educattended (Reference Category: 0)	0	1144 (28.9%)			
	2	877 (22.2%)	0.0409	1.0418	0.0409
	1	1860 (47.1%)	-0.1387	0.8705	0.1387
	3	72 (1.8%)	-0.4691	0.6256	0.4691
Socioeconomic Status and Wealth Proxy					
respondentsoccupation			0.0031	1.0031	0.0031
wealthindex (Reference Category: 1)	1	863 (21.8%)			
	5	629 (15.9%)	0.4028	1.4960	0.4028
	4	803 (20.3%)	0.0897	1.0939	0.0897
	2	749 (18.9%)	-0.0260	0.9743	0.0260
	3	909 (23%)	-0.0849	0.9186	0.0849
wealthindurban (Reference Category: 1)	1	870 (22.0%)			
	4	822 (20.8%)	0.1693	1.1844	0.1693
	2	811 (20.5%)	-0.2077	0.8125	0.2077
	5	727 (18.4%)	-0.2682	0.7647	0.2682
	3	723 (18.3%)	-0.3641	0.6948	0.3641
bicycle (Reference Category: 0)	0	2656 (67.2%)			
	1	1297 (32.8%)	0.1153	1.1222	0.1153
elect (Reference Category: 0)	0	2578 (65.2%)			
	1	1375 (34.8%)	0.0932	1.0977	0.0932
tv (Reference Category: 0)	0	2845 (72.0%)			
	1	1108 (28%)	0.0494	1.0506	0.0494
radio (Reference Category: 0)	0	2660 (67.3%)			
	1	1293 (32.7%)	0.0413	1.0421	0.0413
motorcycle (Reference Category: 0)	0	3343 (84.6%)			
	1	610 (15.4%)	-0.0605	0.9413	0.0605
cartruck (Reference Category: 0)	0	3704 (93.7%)			
	1	249 (6.3%)	-0.1853	0.8308	0.1853
refrigerator (Reference Category: 0)	0	3263 (82.5%)			
	1	690 (17.5%)	-0.3907	0.6766	0.3907

CONTINUATION

Feature	Level	FreqPerc	Coefficient	OddsRatio	FeatureImportance
Household Characteristics					
numhh			0.0052 ↑	1.0052	0.0052
numunderfive			-0.0087 ↓	0.9913	0.0087
province (Reference Category: 1)	1	479 (12.1%)			
	6	426 (10.8%)	0.5500 ↑	1.7333	0.5500
	3	540 (13.7%)	0.0822 ↑	1.0857	0.0822
	5	368 (9.3%)	0.0213 ↑	1.0216	0.0213
	2	537 (13.6%)	-0.1003 ↓	0.9045	0.1003
	9	280 (7.1%)	-0.1732 ↓	0.8410	0.1732
	7	391 (9.9%)	-0.2197 ↓	0.8028	0.2197
	10	212 (5.4%)	-0.4069 ↓	0.6657	0.4069
	8	247 (6.2%)	-0.5512 ↓	0.5763	0.5512
	11	166 (4.2%)	-0.5967 ↓	0.5506	0.5967
	4	307 (7.8%)	-0.7742 ↓	0.4611	0.7742
urbanrural (Reference Category: 1)	1	1212 (30.7%)			
	2	2741 (69.3%)	0.0743 ↑	1.0771	0.0743
floormat (Reference Category: 11)	11	1991 (50.4%)			
	97	46 (1.2%)	0.5190 ↑	1.6803	0.5190
	12	528 (13.4%)	0.1662 ↑	1.1808	0.1662
	22	53 (1.3%)	0.1158 ↑	1.1228	0.1158
	33	1213 (30.7%)	0.0481 ↑	1.0493	0.0481
	32	8 (0.2%)	-0.4681 ↓	0.6262	0.4681
	34	91 (2.3%)	-0.6197 ↓	0.5381	0.6197
	21	11 (0.3%)	-1.0258 ↓	0.3585	1.0258
	31	9 (0.2%)	-1.3347 ↓	0.2632	1.3347
	96	3 (0.1%)	-1.5797 ↓	0.2060	1.5797
roofmat (Reference Category: 11)	11	63 (1.6%)			
	96	5 (0.1%)	2.1016 ↑	8.1795	2.1016
	34	2 (0.1%)	1.0995 ↑	3.0025	1.0995
	97	46 (1.2%)	0.4989 ↑	1.6469	0.4989
	22	449 (11.4%)	0.0680 ↑	1.0703	0.0680
	12	1591 (40.2%)	-0.0419 ↓	0.9590	0.0419
	31	1703 (43.1%)	-0.1090 ↓	0.8967	0.1090
	32	52 (1.3%)	-0.1163 ↓	0.8902	0.1163
	33	42 (1.1%)	-0.6203 ↓	0.5378	0.6203
wallsmat (Reference Category: 11)	11	45 (1.1%)			
	96	1 (0%)	2.6027 ↑	13.5004	2.6027
	23	5 (0.1%)	1.3772 ↑	3.9637	1.3772
	97	46 (1.2%)	0.5138 ↑	1.6717	0.5138
	32	513 (13%)	0.1455 ↑	1.1566	0.1455
	21	795 (20.1%)	0.0876 ↑	1.0915	0.0876
	22	417 (10.5%)	0.0633 ↑	1.0654	0.0633
	33	1007 (25.5%)	-0.0119 ↓	0.9882	0.0119
	12	309 (7.8%)	-0.0272 ↓	0.9731	0.0272
	31	748 (18.9%)	-0.1387 ↓	0.8705	0.1387
	24	67 (1.7%)	-0.2021 ↓	0.8170	0.2021
Water and Sanitation					
typetoilet (Reference Category: 11)	11	13 (0.3%)			
	96	1 (0%)	1.4688 ↑	4.3439	1.4688
	14	2 (0.1%)	1.1935 ↑	3.2986	1.1935
	53	11 (0.3%)	0.3531 ↑	1.4235	0.3531
	21	375 (9.5%)	0.2897 ↑	1.3360	0.2897
	22	515 (13%)	0.2097 ↑	1.2333	0.2097
	52	236 (6%)	0.1787 ↑	1.1957	0.1787
	23	1680 (42.5%)	0.0949 ↑	1.0996	0.0949
	31	973 (24.6%)	-0.1632 ↓	0.8494	0.1632
	12	90 (2.3%)	-0.1763 ↓	0.8383	0.1763
	97	40 (1%)	-0.5260 ↓	0.5910	0.5260
	51	14 (0.4%)	-0.6447 ↓	0.5248	0.6447
	54	3 (0.1%)	-0.9974 ↓	0.3688	0.9974
sourcewater (Reference Category: 11)	11	43 (1.1%)			
	51	28 (0.7%)	2.2118 ↑	9.1326	2.2118
	41	11 (0.3%)	1.8302 ↑	6.2350	1.8302
	61	19 (0.5%)	0.3536 ↑	1.4241	0.3536
	43	326 (8.2%)	0.2327 ↑	1.2620	0.2327
	71	17 (0.4%)	0.2023 ↑	1.2242	0.2023
	31	213 (5.4%)	0.1798 ↑	1.1969	0.1798
	42	81 (2%)	0.1759 ↑	1.1924	0.1759
	21	853 (21.6%)	0.0513 ↑	1.0527	0.0513
	32	1037 (26.2%)	0.0255 ↑	1.0258	0.0255
	13	270 (6.8%)	0.0182 ↑	1.0183	0.0182
	12	556 (14.1%)	-0.0136 ↓	0.9865	0.0136
	14	454 (11.5%)	-0.1207 ↓	0.8863	0.1207
	97	40 (1%)	-0.5416 ↓	0.5818	0.5416
	96	5 (0.1%)	-1.9907 ↓	0.1366	1.9907
timesourcewater			-0.0002 ↓	0.9998	0.0002

Table 1: Ridge Logistic Regression Results

From the Ridge Logistic Regression results, it has been possible to understand the different factors associated with malnutrition of children below five years of age in Mozambique, summarised within 6 key thematic DHS groupings. Child Demographic and Biological Characteristics is the most powerful predictor: children aged 48-59 months (coef = 1.8189, OR = 6.1652) suffered most, followed by those aged 24-35 (coef = 1.5216, OR = 4.5793), 36-47 (coef = 1.3297, OR = 3.7800), and 12-23 (coef = 0.7320, OR = 2.0793) weighing less relative to infants under 12 months, indicating the weaning age. Males are more affected than females as shown by a negative coef for sex (-0.3750, OR = 0.6873). Higher BMI Z-score (coef = -0.4567, OR = 0.6334) and birthweight (coef = -0.7098, OR = 0.4918) significantly decrease risk which confirms the protective influence of adequate nutritional status at a given age and at birth. Maternal and reproductive health variables show some contradictory tendencies: current breastfeeding protects (coef = 0.8289) but relatively recent births tend to increase competition for maternal resources, increasing the risk of malnutrition to some degree. Distance from the hospital as perceived by the respondent together with maternal weight shows only modest or unclear influences.

Education has mother's educational level acting as a modest factor: each additional year of education slightly reduces the risk of complications (coef = -0.0311) and the odds of the child not being malnourished increase by roughly 3.15% per year of maternal education. The relative chances of a child not being malnourished compared to children of uneducated mothers, children with educated mothers demonstrate different levels of association with malnutrition. Children whose mothers completed primary education have about 13% lower odds of not being malnourished (OR = 0.8705), suggesting that primary education alone does not provide strong protection. Those whose mothers completed secondary education possess around 4% better odds of not being malnourished (OR = 1.0418) which is a slight improvement compared to the no education group. More striking, children of mothers with higher level education demonstrate roughly 37% lesser odds of not being malnourished (OR = 0.6256), which is counterintuitive. The higher education unexpected outcome may reveal unseen socioeconomic or contextual reasons such as urban stressor burdens, absence due to job engagements, or uneven availability. Wealth and socioeconomic status spatially and strongly influence results for nutritional attainment outcomes.

Children from the wealthiest households stratified by wealth index category 5 or the richest (coef = 0.4028, OR = 1.4960) demonstrate marked improvements. Ownership of assets, specifically refrigerators (coef = -0.3907, OR = 0.6766) and cars or trucks, appears to reduce the odds of malnutrition. On the other hand, bicycle ownership (coef = 0.1153, OR = 1.1222) increases risk marginally, perhaps indicating that the family is better off economically. Regarding Household Characteristics, taking Niassa (1) Province as the reference category and applying the Ridge Logistic Regression model shows the existence of regional variation in the probability of children's malnutrition prevention across the country in Mozambique. Children from Manica (6) appear to have the best results as they have (OR = 1.7333) suggesting that they are about 73% more likely to avoid malnutrition compared to children from Niassa (1). Nampula (3) and Tete (5) show only slight improvements over Niassa (1) with ORs of 1.0857 and 1.0216, respectively, suggesting very limited benefits. On the other hand, Cabo Delgado (2), Gaza (9) and Sofala (7) have children with lower odds of avoiding malnutrition showing 10-20% disadvantage with ORs from 0.9045 to 0.8028. More extreme differences were observed in Maputo Província (10), Inhambane (8), and Cidade de Maputo (11) with Zambézia (4) showing much lower odds—33% to 54%—compared to Niassa.

These results underscore stark differences in a specific region's economic health access and structural inequalities and how these inequalities are biased towards certain provinces over others. Living in rural areas increases the risk of poverty and malnutrition (coef = 0.5149, OR = 1.6711), along with poor

construction materials like mud walls which significantly increase risk (coef = 2.6927, OR = 14.7936). Analysing the variables related to Water and Sanitation facilities for toilet type, water source, and time required to fetch water—within considerable bounds—demonstrates substantial differences in odds of escaping malnutrition among children, with category 11 serving as the reference group for both toilet type and water source. For toilet type, children in households using facilities categorised as 96, 14, and 53 show significantly higher odds of avoiding malnutrition, with odds ratios (ORs) of 4.3439, 3.2986, and 1.4235 respectively, indicating better sanitation conditions than the reference. On the other hand, several categories such as 97, 54, and 51 along with many others for the same dependent variable are associated with lower odds (ORs below 0.65) which indicates poor sanitation and an increase in the risk of malnutrition. For those households on the lower end, the same category holds true in regard to water source. Households in categories 51 and 41 demonstrate exceptionally high odds (ORs = 9.1326 and 6.2350 respectively) relative to the reference.
At the same time, households with water source categories 96, 97, 14, and 12 exhibit ORs below 1, especially category 96 with an OR of 0.1366, which demonstrates greater susceptibility because of inadequate or unsafe water access. Lastly, the time to source water has a coefficient of almost zero (–0.0002, OR = 0.9998), signifying almost no impact on malnutrition in this model.

Classification Performance Metrics for Ridge Logistic Regression in Malnutrition Outcomes

Table 2: Classification Performance Metrics for Stunting, Underweight and Wasting

Outcome	Accuracy	Precision	Recall	Specificity	F1 Score	Balanced Accuracy	Log Loss	Brier Score
stunting	0.8609	0.8615	0.9990	0.0061	0.9252	0.5025	0.3908	0.1171
underweight	0.8339	0.8484	0.9525	0.4520	0.8974	0.7022	0.3780	0.1172
wasting	0.9317	0.9152	0.9401	0.9244	0.9275	0.9323	0.1536	0.0475

Analysing the classification performance metrics of the Ridge Logistic Regression model for the three types of malnutrition — stunting, underweight, and wasting — provides critical insights into the predicting and detecting performance for children in Mozambique. Concerning stunting, the model achieves exceptionally high recall (0.999) with a strong F1 Score of (0.9252), signalling near perfect detection of stunted children. On the other hand, its specificity is extremely low (0.0061) alongside moderate balanced accuracy of 0.5025. This indicates that the model fundamentally misunderstands the classification problem by failing almost entirely to correctly classify non-stunted children, which aligns with predominant structural risk factors like living in substandard housing and lack of safe water and sanitation. This over-sensitivity suggests that the model has some utility in identifying children at chronic risk, but becomes functionally useless because it generates far more false positives than true positives. Conversely, the underweight model's performance appears more balanced across metrics with precision (0.8484), recall (0.9525), and even lower specificity (0.4520) resulting in higher balanced accuracy (0.7022). This indicates a greater ability to differentiate underweight from non-underweight children, which could stem from a wider range of factors such as birth weight, maternal education, and overall family economic status.

The wasting model demonstrates the"stro'gest and most balanced performance: it records the highest accuracy (0.9317), specificity (0.9244), and lowest log loss (0.1536), indicating superb calibration and class separation. This performance is due to wasting's close association with immediate and quantifiable risks such as contaminated water, illness, and recent healthcare use, which the model captures well. Integrating these findings, stunting is most difficult to classify due to its chronic, structurally driven, multifactorial nature, whereas wasting is simplest due to its acute, sharply defined markers. Underweight is positioned between these two extremes, benefiting from a mix of chronic and acute risk influences. Thus, policy actions should prioritise enhanced stunting model accuracy through better class imbalance adjustment or alternative modelling methods, while leveraging the wasting model's high reliability to guide urgent responses for acute malnutrition

Table 3: Goodness of Fit Metrics

Outcome	Null Deviance	Deviance	AIC	BIC	Pseudo R2 (McFadden)	Pseudo R2 (Cox-Snell)
stunting	3,189.20	2,458.62	2,528.62	2,748.50	0.2291	0.0883
underweight	4,331.96	1,980.50	2,050.50	2,270.38	0.5428	0.2573
wasting	5,459.74	1,052.76	1,122.76	1,342.64	0.8072	0.4273

The Ridge Logistic Regression model fit for the three outcomes of malnutrition, namely stunting, underweight, and wasting, provides insight into how each model explains the variation in the data and how well each model fits relative to the classification metrics within the socioeconomic, biological framework discussed earlier. For stunting, the model's null deviance (3189.20) and residual deviance (2458.62) indicate some reduction in unexplained variation, but the Pseudo R^2 values—McFadden (0.2291) and Cox-Snell (0.0883)—suggest weak explanatory power. This limited fit is in line with the high recall but very low specificity described previously; it is as though the model captures broad risk for stunting but has imprecise discrimination ability due to a dominant structural (e.g., poor housing and sanitation) effect predicting presence vs absence of stunting. The AIC (2528.62) and BIC (2748.50) are high as well, reinforcing that the model's complexity does not enhance fit or parsimony. In contrast, the underweight model appears to capture the underlying phenomena more accurately, as the deviance is significantly lower: from 4331.96 to 1980.50, accompanied by a markedly higher McFadden Pseudo R^2 (0.5428) and Cox-Snell R^2 (0.2573), which suggests better balance between complexity and explanatory power.

This corresponds with better classification metrics as well as a more comprehensive set of risk factors (e.g., birth weight, wealth index, maternal education) which likely improve model discriminability. Stunting is associated with lower AIC and BIC values of 2050.50 and 2270.38 respectively demonstrating increased model efficiency. For wasting, the model fit is the strongest: the null deviance of 5459.74 dramatically drops to a residual deviance of 1052.76 with very high McFadden (0.8072) and Cox-Snell (0.4273) Pseudo R^2 values yielding 805% and 42.73% respectively —suggesting that over 80% of the variation in wasting status is explained by the model. Reinforcing this claim is the extremely low AIC of 1122.76 and BIC of 1342.64 making it the most parsimonious and predictive of the three. Earlier findings that wasting, an acute and immediate condition, is best predicted by clearly measurable factors such as water safety, sanitation, recent illness, and other sharp determinants comes to mind. In summary, model fit statistics strongly reflect the classification performance: the wasting model is best-fitting and well-balanced, underweight is moderately strong, and stunting, though sensitive, suffers from poor specificity and weaker explanatory power emphasizing the need for refined modelling strategies to better encapsulate the chronic, structurally driven nature of stunting.

Diagnostic Performance Curves

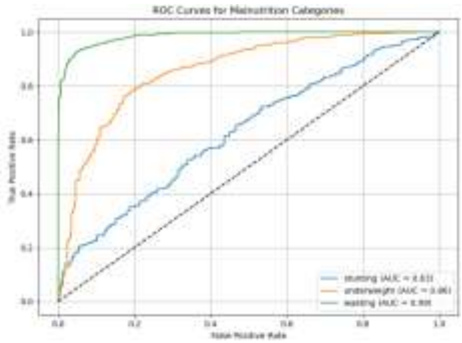


Figure 2: ROC Curves for Malnutrition

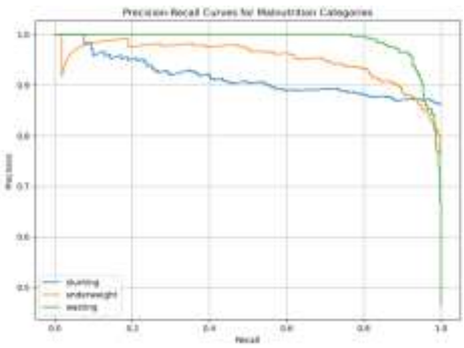


Figure 3: Precision-Recall Curves for Malnutrition

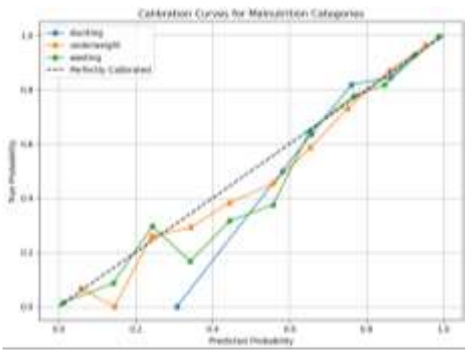


Figure 4: Calibration Curves for Malnutrition

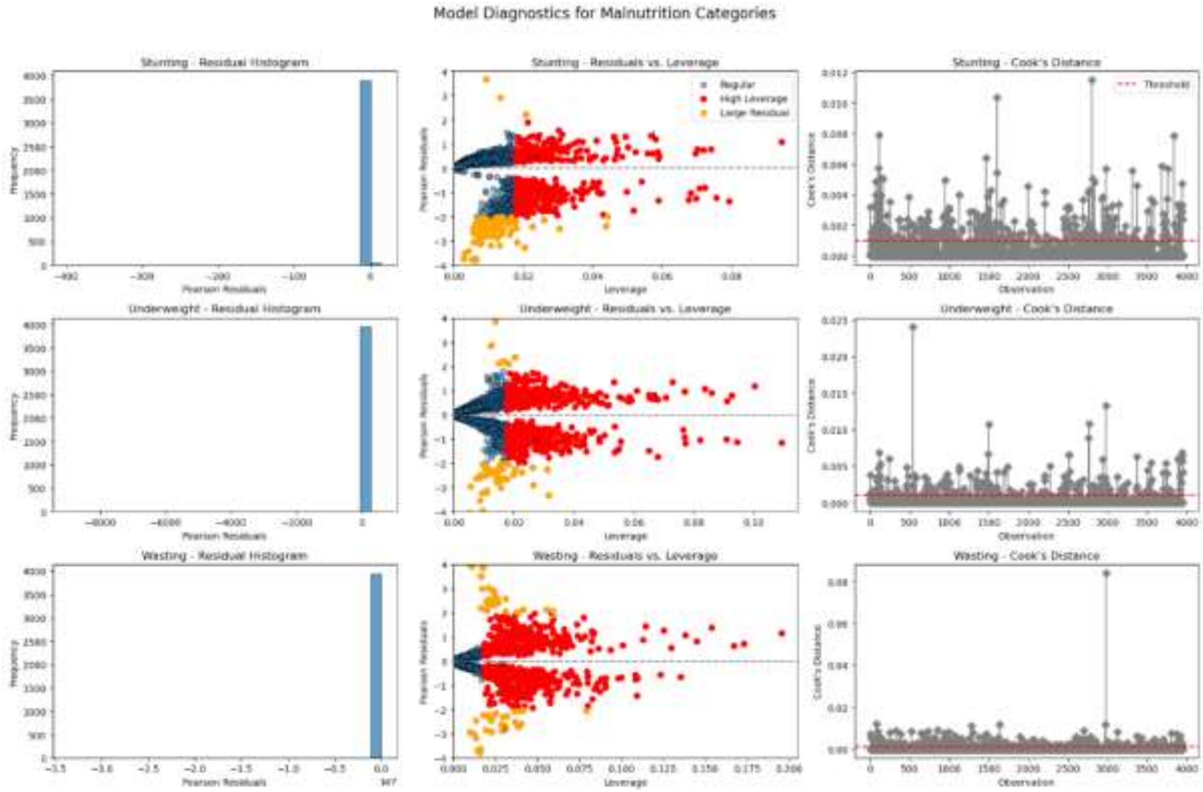
The ROC curve represents the True Positive Rate (Recall) and the False Positive Rate for different classification thresholds. Also, it is important to note that AUC is a single value that measures how well a model distinguishes the different classes and an AUC of 1.0 means perfect discrimination while 0.5 means no discriminative power whatsoever. The ROC curve for wasting is almost perfect, right above the top left corner. This yields an AUC of 0.99 which reflects exceptional discrimination between the wasted and non-

wasted children. The model effectively captures acute risk factors such as poor sanitation, unsafe water access, recent illness, which are observable and highly predictor. This is corroborated by high specificity (0.9244), accuracy (0.9317), and McFadden's pseudo R^2 of 0.8072, indicating a strong model fit. For underweight, the model reasonably succeeds in detecting underweight children, attaining an AUC of 0.86. The curve rises sharply, signifying moderate false-positive error alongside good recall. The condition underweight arises due to the combination of several biological (e.g. low birth weight) as well as socioeconomic (e.g. maternal education, household wealth) determinants. The balanced performance of the model is further shown by a recall of 0.9525, specificity of 0.4520, and balanced accuracy of 0.7022.
In comparison, the stunting model has a weak AUC of 0.63 with a curve that tracks the diagonal, showing low discriminative ability. Stunting represents the chronic undernutrition associated with deep-rooted structural deprivation like enduring poverty, inadequate housing, and low maternal literacy levels. These factors are more universal, observed in both stunted and non-stunted children, which decreases the model's specificity (0.0061) despite near-perfect recall (0.9990). Consequently, its balanced accuracy stays around 0.5025, just slightly above random chance.
With regards to the Precision-Recall Curve Analysis which is evaluated under class imbalance, in the context of healthcare outcomes, PR curves are particularly useful with the majority of cases exhibiting an imbalanced distribution. They assess the ratio of correct positive identifications to total predicted positives (precision), and the ratio of true cases to identified cases (recall). The wasting PR curve is almost flat and high, yielding a precision of near 1.0 even with high recall. This indicates that classification is almost flawless—most predictions of wasting are correct and very few at-risk children are missed. The curve supports the high AUC, high specificity, and low log-loss corroborating the model's efficacy for directing acute malnutrition treatment in settings constrained by poor infrastructure and scant healthcare resources. Underweight PR curve shows robust balanced performance outperforming other models. Precision does not drop below 0.94 with recall approaching 0.90. The model seems to reliably identify true underweight cases while issuing few false positives. These results stem from the dataset that captures a broad risk profile: biological, economic, and educational factors. The model is therefore strengthened achieving an F1-score of 0.8976 and McFadden's R^2 of 0.5428. Finally, the stunting PR curve illustrates the increasing precision-recall gap as a regression of the model's capabilities.
When recall is near 1.0, precision falls below 0.85, revealing an abundance of incorrect positive results. The model's very low specificity and accuracy illustrates an important balance—recall is gained by incorrectly classifying a significant number of children who are healthy as stunted. This tendency highlights the difficulty of predicting stunting by using survey-based indicators that do not reflect the persistent and systemic causes of the condition.
Calibration curves measure the agreement between predicted probabilities and observed outcomes. A model is considered well-calibrated if its probabilities align with the actual risk frequency in a positive directional manner. The calibration curve for wasting is closely aligned with the diagonal, especially between the 0.5 to 1.0 range, which represents excellent calibration. There is a strong confirmation that the model's probabilistic outputs are reliable, as predictions substantively surpass clinical reality, thereby confirming its use in public health decision-making. The underweight calibration curve demonstrates acceptable alignment for central probabilities (0.3–0.9). There is slight risk underestimation at low probabilities, though the model remains reliable. This indicates that the model's risk estimation is robust and may be applicable for population-level screening and prioritisation. As for the stunting calibration curve, it diverges from the ideal line in the mid-probability region. It undershoots the real risk from 0.4–

0.6 and overestimates risk beyond that, resulting in rough calibration. This illustrates poor calibration stemming from the model's prior difficulties with specificity and excessive classification. Such outputs should be considered untrustworthy and must be handled cautiously in practical applications.

Model Diagnostics for Malnutrition Categories

Figure 5: Residual Histogram, Residual vs. Leverage and Cook's Distance for Malnutrition Categories



The chart 5 displays regression diagnostic plots for Ridge Logistic Regression models applied to the three malnutrition outcomes in 5-year old children: stunting, underweight, and wasting. The diagnostic plots are shown in three columns: Pearson residual histograms (left), residuals versus leverage plots (centre), and Cook's distance plots (right). These analyses provide information about the fit of the model, outlier detection, and model influence diagnostics.

Within the three stunted categories, almost all the children showed remaining differences graphed against height, weight, and age which peaked near the vertical zero reference line suggesting that the model fitted the vast majority of children reasonably thus the plotted value of their age. However, a small number of children yield very low residuals especially in weight measuring under 3, then 5 "pounds" away which suggests extreme overexpectation due to non-linear averaging or unbalanced datasets on risk factors not accounted for. Though more extreme stunting outliers appear balanced, moderate free stunting remains prevalent across models.

The behaviour of the model is further pronounced by the residuals versus leverage plots. In every plot, most observations are classified as regular (blue), but a significant proportion of points either have large residuals (red) or high leverage (orange). For stunting and underweight, many red points signal substantial misfit where some observations greatly diverge from what was estimated. These misfits occur at different

levels of leverage, suggesting that the errors do not result solely from extreme values of the predictors but could be more systematic. In contrast, the wasting model has fewer red or orange points, which indicates fewer cases of trouble and greater predictive stability.
The Cook's distance plots give some indication of where important observations may be situated. A red dashed line marks the conventional line of influence (4/n) which is used as a threshold. For stunting and underweight, only a few cases that are clustered around this line and most prominent around 500 and 2500 marks, which suggests those cases are important as they alter the fitted values greatly. In wasting, however, Cook's distance does not change across all observations and is on the lower side, indicating that the model is not affected much by observations that are unique.
The diagnostics confirming the wasting model shows the best performance has least concern with leverage influence issues and residuals are closely clustered. Under-stunting and underweight, on the other hand, seem to display greater difficulty with outlier sensitivity and broader residual fit that suggest possible underfitting, data quality, or missing variable bias. Further work might consider robust feature construction or data scrubbing which targets high residual or high leverage observations to improve model performance, particularly for categorization within stunting and underweight.

Conclusion

The application of Ridge Logistic Regression within this research aimed to explore the comprehensive determinants of childhood malnutrition across Mozambique utilising an extensive set of predictors structured within Demographic and Health Survey (DHS) thematic frameworks. As the results illustrate, the nature of the outcomes of malnutrition is not only limited to stunting and underweight but rather encompasses a diverse spectrum which includes wasting, each with unique modelling challenges and risk profiles.
Wasting, which is a form of acute malnutrition, was predicted with the greatest accuracy, calibration, and overall model fit. This indicates that immediate risk factors such as poor sanitation, unsafe drinking water, and substandard housing strongly corroborate wasting status and are well captured by the existing dataset. The model for underweight reflected an equilibrium of both chronic and acute factors in primary school education, maternal education, household assets, and environmental features. It performed moderately to highly on predictive performance across classification and calibration metrics.
While the model for stunting showed high recall, it struggled with specificity and discriminative ability. This reflects the nature of stunting as a complex chronic systemic condition disproportionately influenced by widespread deprivation that impacts both stunted and non-stunted children. The Ridge model showed that structural indicators such as poor wall material, type of toilet facilities considered unsafe, and living in rural areas were some of the strongest predictors for stunting. However, the broad prevalence of these factors constrained the model's discriminatory ability to classify cases.
These outcomes confirm the efficiency of Ridge Logistic Regression in controlling multicollinearity as well as allowing consistent estimation over many dependent interrelated predictors. The grouping of variables aligned with the DHS proved useful for both analytic coherence and policy translatability. Ultimately, the study emphasises the multifaceted adaptable intervention approach for intervention: acute malnutrition responds to immediate hygiene and infrastructural interventions, whereas stunting requires enduring systemic changes in housing, education, and rural development.

The Policy Recommendations

Addressing child malnutrition in Mozambique requires policymakers to focus on distinct strategies for under-five stunting, underweight, and wasting separately. Community sanitation improvement, water safety promotion, and disease surveillance enhancement are immediate interventions needed during the acute environmental risk phase of wasting. Stunting necessitates long-term investment in rural infrastructure, maternal education, and poverty mitigation. Policies should be constructed using DHS variable groups: For intervention. breastfeeding and maternal care are antenatal, postnatal, and care; education retention of females; asset targeting for cash/food transfers; and infrastructure for non-secure water and sanitation zones with increased need. Moreover, health systems supplemented with national machine learning algorithms, such as Ridge Logistic Regression, can facilitate anticipatory predictive analytics for directed pre-emptive action. Programming and budgeting—subsidy housing and sanitation improvement, literacy investment, and others—should be guided by model output. Last, continuous evaluation through retraining and model updates enhances adaptive policy and need-responsive data-driven governance.

References

1. Akombi, B. J., Agho, K. E., Hall, J. J., Wali, N., Renzaho, A. M., & Merom, D. (2017). Stunting, Wasting And Underweight In Sub-Saharan Africa: A Systematic Review. International Journal of Environmental Research and Public Health, 14(8), 863. https://doi.org/10.3390/ijerph14080863
2. Alderman, H., & Headey, D. D. (2017). How Important Is Parental Education For Child Nutrition? World Development, 94, 448–464. https://doi.org/10.1016/j.worlddev.2017.02.007
3. Bhutta, Z. A., Das, J. K., Rizvi, A., Gaffey, M.F., Walker, N., Horton, S.,…& Black, R. E. (2013). Evidence-based interventions for improvement of maternal and child nutrition. The Lancet, 382(9890), 452–477. https://doi.org/10.1016/S0140-6736(13)60996-4
4. Black, R. E., Victora, C. G., Walker, S. P., Bhutta, Z. A., Christian, P., de Onis, M.,… & Uauy, R. (2013). Maternal and child undernutrition and overweight in low-income and middle-income countries. The Lancet, 382(9890), 427–451. https://doi.org/10.1016/S0140-6736(13)60937-X
5. Cardoso, M. A., Scopel, K. K. G., Muniz, P. T., Villamor, E., & Ferreira, M. U. (2006). Underlying Factors Associated With Anemia In Amazonian Children: A population-Based study. Public Health Nutrition, 9(5), 575–582. https://doi.org/10.1079/PHN2005897
6. Faye, C. M., Fonn, S., & Sartorius, B. (2019). Evaluating child malnutrition determinants with machine learning. International Journal of Environmental Research and Public Health, 16(3), 456. https://doi.org/10.3390/ijerph16030456
7. Gebretsadik, A., Tadesse, T. and Fenta, A. (2022). Penalized regression models for stunting prediction among under-five children in Ethiopia. BMC Pediatrics, 22, 51. https://doi.org/10.1186/s12887-022-03131-4
8. Goyal, A., & Soni, A. (2021). Regularized regression for interpretable healthcare analytics. Health Informatics Journal, 27(1), 146–159. https://doi.org/10.1177/1460458220946082
9. Hirpa, S., Gebretsadik, A., & Nigatu, Y. (2020). Predictive modelling of childhood malnutrition in Ethiopia: a machine learning approach. BMC Medical Informatics and Decision Making, 20, Article 306. https://doi.org/10.1186/s12911-020-01315-5
10. Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12(1), 55–67. https://doi.org/10.1080/00401706.1970.10488634

11. Hossain, M., Mani, K. K., Udoy, S. M. R., & Fahim, S. M. (2019). Environmental determinants of childhood stunting in South Asia: a systematic review. Environmental Health and Preventive Medicine, 24(1), 1–14. https://doi.org/10.1186/s12199-019-0829-3
12. Kandala, N. B., Madungu, T. P., Emina, J. B., Nzita, K. P. D., & Cappuccio, F. P. (2011). Malnutrition among children under-five years of age in the Democratic Republic of Congo (DRC): Does geographic location matter? BMC Public Health, 11, Article 261. https://doi.org/10.1186/1471-2458-11-261
13. Khulu, S., Adebayo, S. B., & Mapoma, C. C. (2021). Socio-environmental determinants of child stunting in southern Africa. Journal of Biosocial Science, 53(4), 588–602. https://doi.org/10.1017/S0021932020000301
14. Masibo, P. K., & Makoka, D. (2012). Trends and determinants of undernutrition among young Kenyan children: Kenya demographic and health survey. African Population Studies, 26(2), 159–173. https://doi.org/10.11564/26-2-242
15. Namara, R. E., & Golitko, M. (2022). Modelling hygiene and sanitation impacts on child nutrition using regularized regression methods. *Journal of Water, Sanitation and Hygiene for Development,* 12(1), 33–43. https://doi.org/10.2166/washdev.2021.095
16. Ogunjimi, L. O., Gbogboade, A. A. & Alabi, O. S. (2020). Determinants of maternal healthcare utilization: A regularized logistic regression approach. BMC Health Services Research 20(1), 1105. https://doi.org/10.1186/s12913-020-05969-1
17. Tarekegn, H., Kim, J. & Bekele, A. (2021). Machine Learning Models for Predicting Child Malnutrition: Evidence from Ethiopia. PLOS ONE, 16(7), e0254293. https://doi.org/10.1371/journal.pone.0254293
18. Tiwari, R., Ausloos, M. & Dey, L. (2021). Machine Learning Techniques for Child Health in South Asia. Computers in Biology and Medicine, 134, 104464. https://doi.org/10.1016/j.compbiomed.2021.104464
19. UNICEF, WHO, & World Bank Group. (2021). Levels And Trends In Child Malnutrition: UNICEF/WHO/World Bank Group Joint Child Malnutrition Estimates: Key findings of the 2021 edition. https://data.unicef.org/resources/jme-report-2021/
20. Victora, C. G., Christian, P., Vidaletti, L. P., Gatica-Domínguez, G. Menon, P. & Black, R. E. (2021). Revisiting Maternal And Child Undernutrition In Low- And Middle-Income Countries: Variable Progress Towards An Unfinished Agenda. The Lancet, 397(10282), 1388–1399. https://doi.org/10.1016/S0140-6736(21)00394-9
21. Zou, H. & Hastie, T. (2005). Regularization And Variable Selection Elastic Net. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67(2), 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x