Predictive Data Analytics on Global Healthy Diet Cost and Prevalence of Malnutrition

Yash Tripathi[1], Anoop Kumar[2], Sumit Sagar[3], Pratha Saxena[4]
[1,2,3,4]*Department of Computer Applications, Invertis University, Bareilly*
doi.org/10.64643/IJIRTV12I11-207269-459

***Abstract*—Affordability of healthy diets and undernourishment have been ongoing concerns in the world, but there is limited evidence on their predictive understanding on the regional level. This research proposes a data analysis of the dynamics of healthy diet costs and malnutrition prevalence using two FAO databases [5]: the "Healthy Diet Cost" database (more than 170 countries, 2017 -2024, PPP-corrected [8]) and "Prevalence of Undernourishment" database (2000-2022). On the global scale, healthy diet costs significantly exceed the $2.15/day international poverty line (The World Bank updated the international poverty line to $3.00/day in June 2025 based on 2021 PPPs [World Bank, 2025]; however, as this dataset uses 2017 PPP-adjusted values, the 2017 baseline is retained for consistency), as only 10% of people in poor countries can afford a healthy diet [1][10]. Global average healthy diet costs amount to $3.68/day, growing by 42.1%, i.e., from $3.14 in 2017 to $4.46 in 2024, mainly due to increased inflation and logistical problems in 2021-2024 [1][11]. Multiple machine learning techniques were compared; the best one (Gradient Boosting) resulted in R^2 equal to 96.07%. According to the SARIMAX forecast, costs will increase in Africa, Asia, and America over the next five years, while costs in Europe will be relatively low. Although there was a 24.74% decrease in undernourishment between 2000 and 2022 [1][3], regional differences still exist.**

***Index Terms*—Food Affordability, Food Security, Healthy Diet Cost, Nutritional Economics, Prevalence of Malnutrition, Purchasing Power Parity (PPP), Predictive Analytics, Time Series Forecasting (SARIMAX).**

I. INTRODUCTION

The analysis will utilize datasets collected by FAO from over 170 countries between the years 2000 to 2024 regarding cost of diets and malnutrition cases, to create different machine learning models and SARIMAX predictive models, which shall be applied in predicting future trends in the region on the cost of nutrition diet and malnutrition cases over the next five years. The objective is to create recommendations using evidence-based policy implication on ways to alleviate food insecurity among people, especially in the African continent and Asia where costs are higher [1]. The nutritional issues represent one of the most important health concerns, encompassing malnutrition and micronutrient deficiency as defined by WHO [2]. In cases involving impoverished regions like sub-Saharan Africa where more than 20% of individuals were undernourished in 2024, undernutrition leads to low height, poor brain functions, and low production capabilities [1][3][4]. Malnutrition adversely affects productivity, learning skills and the development of the country economically [3]. According to SOFI 2025, any rise in the cost of food by 10% may cause a 2.7%-4.3% increase in childhood wasting [1].

Earlier studies have discussed issues related to the unaffordability of healthy diets and malnutrition separately. The cost thresholds were set by Herforth et al. [10] using the FAO prices. Headey and Alderman [4] found that there was a certain trend in relation to the differences between healthy and unhealthy food costs depending on income and the region. Smith and Haddad [3] found critical factors affecting childhood undernutrition in developing countries. While SOFI 2025 provides a comprehensive perspective on the situation worldwide, it lacks a forecasting approach to estimate the future impacts of the trends in question on malnutrition outcomes in regions. This is the problem addressed in the present study. The present study seeks to consider both the datasets of the FAO on diet costs and malnutrition in over 170 countries (years 2000-2024) together to forecast diet affordability and undernutrition outcomes with machine learning approaches and the SARIMAX model for the next five years.

II. DATA AND METHODOLOGY

This study utilizes two datasets from the Food and Agriculture Organization (FAO) [5]: the prevalence of undernourishment by country (2000–2022) and the PPP-adjusted cost of a healthy diet across 170+ countries (2017–2024). Costs are expressed in PPP-adjusted USD to enable standardized cross-country comparisons, as recommended by the World Bank [8]. Data preprocessing was carried out to clean and standardize the datasets. In case of missing values in both datasets, the imputation procedure involved filling them with medians of numerical variables and modes of categorical variables. Country name encodings were performed using scikit-learn. In EDA part, statistics and visualization were applied to find time series trends, compare by regions, and use scatter matrix plot visualization with help of pandas, matplotlib, and seaborn libraries [6][7]. On healthy diet dataset, five models have been trained using scikit-learn where the data was divided into a ratio of 80% for training and 20% for testing with the usage of 10 features. The selected algorithm used for the model was the Gradient Boosting method because of its higher R^2 value (0.9607) and lower RMSE value ($0.178). Full model specifications and performance comparison are reported in Section 12 and Table III. Finally, when fitting a SARIMAX model (1,1,1) separately per region based on the annual cost data in regard to the diet, and without the influence of seasonality because the data is annual, forecasts on how the costs will perform over the next five years were made, at a 95% confidence level.

A. Dataset – Healthy Diet Cost

- Source: FAOSTAT, Food and Agriculture Organization [5]
- Size: ~1,379 records (countries × years)
- Time Period: 2017–2024
- Target: Daily cost of a healthy diet (PPP-adjusted USD)
- Features: Country, region, year, cost categories (vegetables, fruits), data quality

B. Dataset – Prevalence of Undernourishment

- Source: FAOSTAT, Food and Agriculture Organization [5]
- Size: ~4,455 records (countries × years)
- Time Period: 2000–2022
- Target: Prevalence of Undernourishment (Normalised Rate)
- Features: Country, year, prevalence of undernourishment, percentage

C. Numerical Features Summary — Healthy Diet Cost Dataset

- Daily Cost: Mean = $3.68, Range = $1.70 – $8.39. Distribution is right-skewed, with some countries recording a significantly higher daily cost than the mean.
- Annual Cost: Mean = $1,342.26, Range = $620.50 – $3,062.35.
- Daily Cost of Vegetables: Mean = $0.76, Range = $0.21 – $1.69 (Asia recording the highest regional average).
- Daily Cost of Fruits: Mean = $0.68, Range = $0.17 – $2.50 (Asia recording the highest regional average).
- Daily Cost of Total Food Components: Mean = $1.44, Range = $0.38 – $3.40 (Asia recording the highest regional average).

D. Numerical Features Summary — Prevalence of Undernourishment Dataset

- Prevalence of Undernourishment per Year:

Mean = 11.40%, Range = 2.50% – 70.90% (Africa recording the highest regional prevalence).

- Relative Undernourishment Score per Year [Engineered]:

Mean = 1.77, Range = −10.60 – 61.50 (Africa recording the highest score).

E. Categorical Features Summary — Healthy Diet Cost Dataset

- Country: Most frequent instance is Albania.
- Region: The most represented region is Europe.
- Cost Category: Two categories — High and Medium cost. The majority of records fall under High Cost (1,160 out of 1,368).

F. Categorical Features Summary — Prevalence of Undernourishment Dataset

- Entity: Most frequent instance is Africa.
- Relative Undernourishment Category [Engineered]: Four categories — Lowest, Low, High, and Highest. The majority of records fall under the High category (2,328 out of 4,455).

III. VISUAL ANALYSIS OF PREVALENCE OF UNDERNOURISHMENT ACROSS CONTINENT

As shown in Fig 1, it is evident that there is a lot of difference regarding the percentages of malnutrition amongst people from different continents.

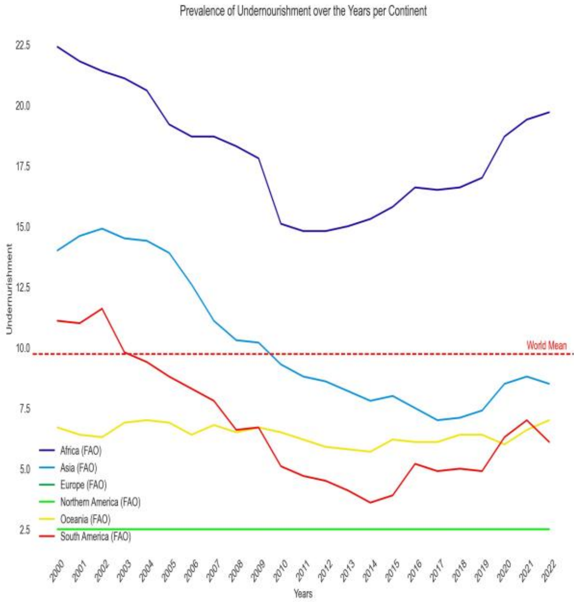


Fig 1: Prevalence of undernourishment by continent, 2000–2022.

Some of the continents that demonstrate high percentages of malnutrition include African and Asian continents, which are always above the average in terms of percentages worldwide. Some examples of continents with low percentages of malnutrition are Europe and Northern America shown in Table I.

TABLE I: SUMMARY OF PREVALENCE RATE

Continents	Undernourishment
Africa	18.06
Asia	10.26
South America	6.80
Oceania	6.41
Northern America	2.50
Europe	2.50

IV. VISUAL ANALYSIS OF PREVALENCE OF UNDERNOURISHMENT IN ASIA, AFRICA

Fig 2(a) and fig 2(b) represent information on the trends in undernourishment in terms of Asian regions, African regions, and income economies.

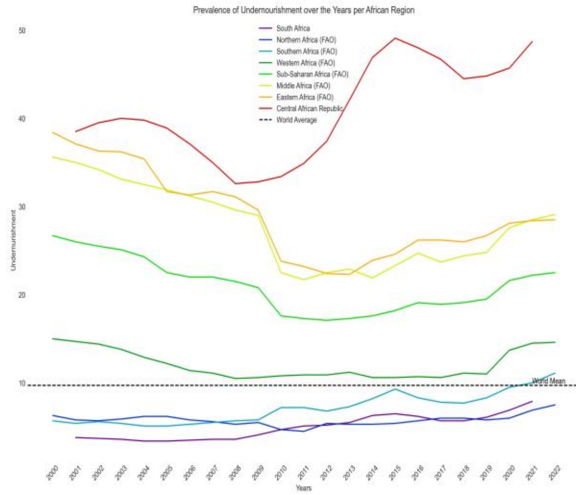


Fig 2(a): Prevalence in African Regions

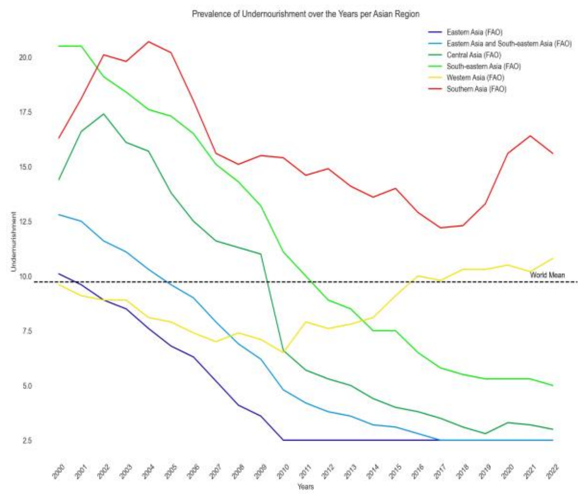


Fig 2(b): Prevalence in Asian Regions

The Central African Republic represents the country with the highest average prevalence rate, followed by Eastern Africa and Middle Africa, where the majority of the sub regions in Africa record values above the global mean value represent in Table II(a).

TABLE II(A): SUMMARY OF PREVALENCE RATE

Regions	Undernourishment
Central African Republic	40.78
Eastern Africa	29.09
Middle Africa	27.83
Sub-Saharan Africa	21.06
Western Africa	12.08
Southern Africa	7.10
Northern Africa	5.77
South Africa	4.98

The Southern Asian region records the highest average value while the Eastern Asian region records the lowest average, with all subregions experiencing a decreasing trend represent in Table II(b).

TABLE II(B). SUMMARY OF PREVALENCE RATE

Regions	Undernourishment
Southern Asia	15.84
South-eastern Asia	11.51
Western Asia	8.71
Central Asia	8.44
Eastern Asia	4.49

Fig 2(c) confirms a clear income-based disparity — high-income economies remain near zero throughout, while low-income economies peaked at ~34% and, despite an initial decline, rose again after 2011, recording a 16.09% increase from 2010 to 2022. This post-2012 reversal across multiple regions is consistent with global food security assessments [1].

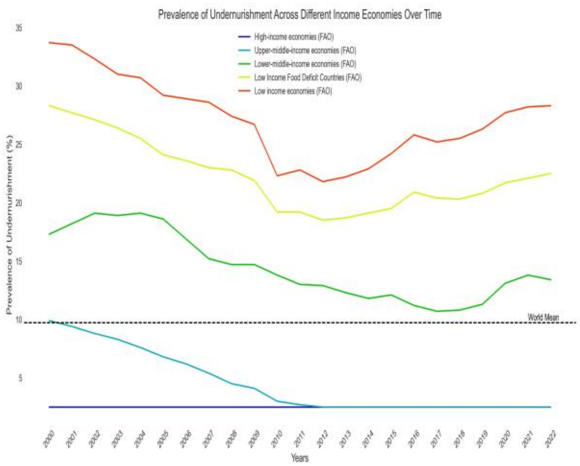
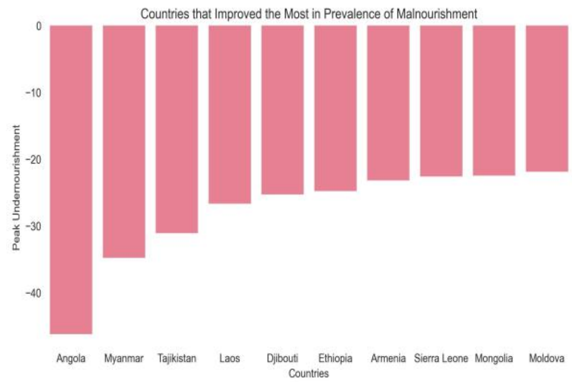


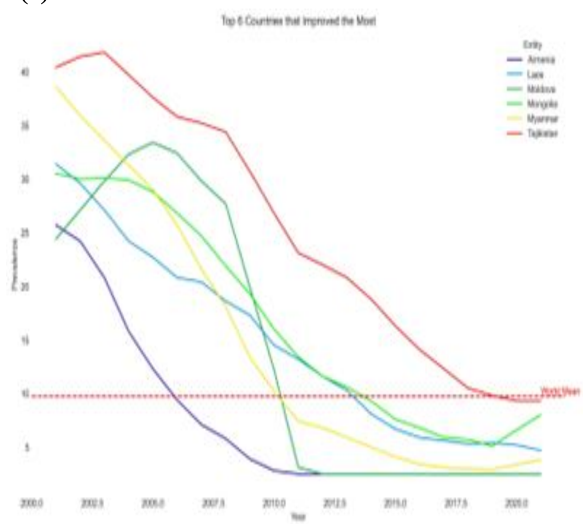
Fig 2(c): Prevalence of Undernourishment in Income Economy Groups.

V. VISUAL ANALYSIS OF COUNTRIES THAT IMPROVED FROM UNDERNOURISHMENT

Fig 3(a) ranks the countries with the greatest reduction in undernourishment prevalence between 2001 and 2022, measured as peak change. Fig 3(b) shows that top improvers such as Armenia, Tajikistan, Laos, Mongolia, and Moldova achieved sustained, consistent declines over the full period — with Moldova reducing prevalence to as low as 2.5%, falling well below the world mean.



3(a)



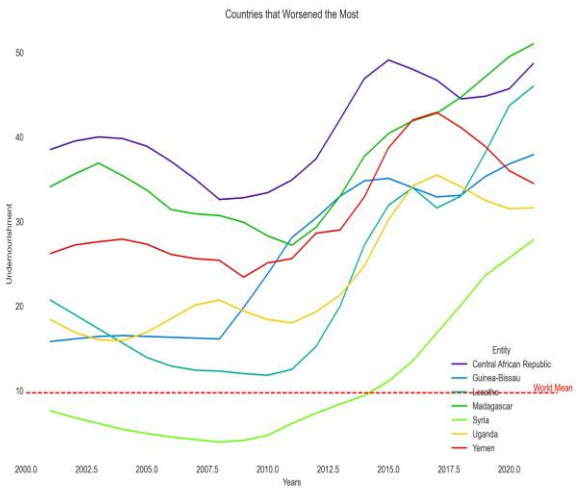
3(b)

Fig 3: Countries that improved the most in undernourishment prevalence (2001–2022): (a) ranked by peak reduction, (b) trend trajectories of the top 6 countries.

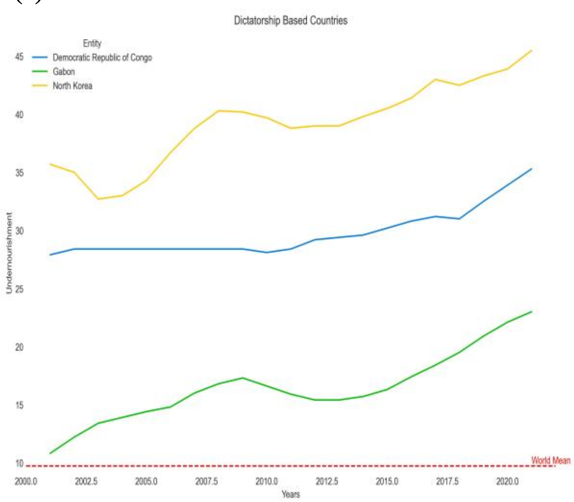
VI. VISUAL ANALYSIS OF COUNTRIES THAT WORSENED FROM UNDERNOURISHMENT

The situation in Syria in 2001 represents a considerable decline in malnutrition, but nations below 20% in 2001 were expected to exceed this fig by 2021. Fig 4(a) shows the most extreme cases, with a fast rise witnessed between 2013 and 2018, especially following the interruption that occurred during the period of 2010-2011 when there was a sharp increase in undernourishment for nations that previously had improved.

Fig 4(b) clearly indicates an increasing pattern of undernourishment for nations that have dictatorial regimes, including the DRC, Gabon, and North Korea [12][13][14][15].



4(a)



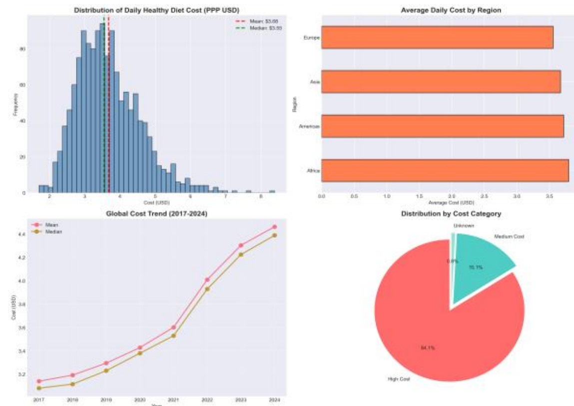
4(b)

Fig 4: Countries with the Highest Increase in Undernourishment Prevalence (2000–2022): (a) Countries That Worsened the Most, (b) Dictatorship-based Countries

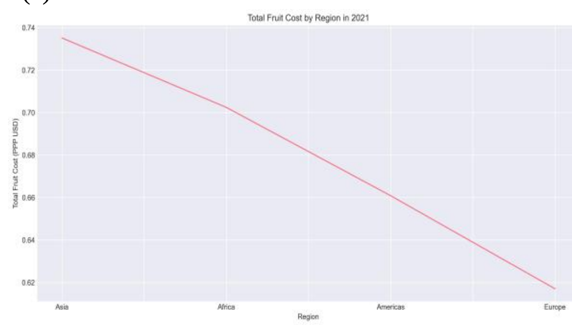
VII. VISUAL ANALYSIS OF HEALTHY DIET COST DATASET

Fig 5(a) shows that the daily healthy diet cost is right-skewed (mean $3.58, median $3.08), with the majority of records falling under the High Cost category (84.1%).

Fig 5(b) shows a consistent upward trend in annual diet cost from 2017 to 2024, with Africa rising the steepest and Europe remaining the lowest throughout, reflecting a global pattern of increasing food costs consistent with FAO and World Bank assessments[1][11].



5(a)



5(b)

Fig 5: Distribution and Regional Trends of Healthy Diet Cost, 2017–2024: (a) Daily Cost Distribution and Regional Averages, (b) Annual Cost by Region.

Table III shows Summary of cost estimates worldwide. The average cost was calculated to be USD 3.68 per day, which equates to an annual expenditure of approximately USD 1,342.26. The reported daily costs varied from USD 1.70 to USD 8.39, indicating a large variation among the analyzed cases. The standard deviation of USD 0.90 also indicates a moderate dispersion around the global average cost.

TABLE III. SUMMARY OF PREVALENCE RATE

Global Average Cost	$3.68/day
Annual Equivalent	$1342.26/year
Cost Range	$1.70 - $8.39
Standard Deviation	$0.90

VIII. VISUAL ANALYSIS OF COST OF VEGETABLES AND FRUITS ACROSS REGIONS

As shown in Fig 6 regional variation in vegetable and fruit costs, Africa records the highest vegetable cost, while Asia leads in fruit cost — averaging $0.74/day, which is 2.07% above the global average and 4.35% above Europe's (Fig 6(b). Europe consistently records the lowest costs for both components across all regions.

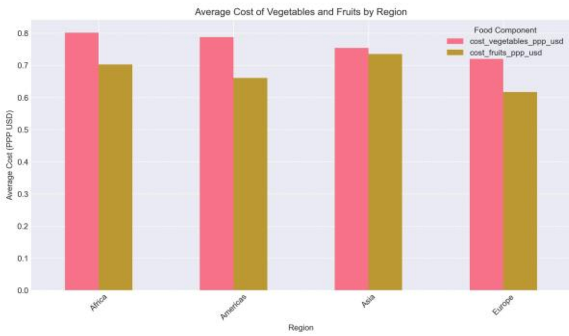


Fig 6: Regional Comparison of Vegetable and Fruit Costs

Below Fig 7 shows a positive relationship between vegetable and fruit costs, with higher vegetable costs generally associated with higher fruit costs and overall diet cost.

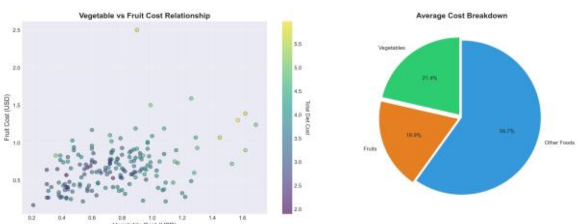


Fig 7: Vegetable and Fruit Cost Analysis: Relationship Between Vegetable and Fruit Costs by Total Diet Cost and Average Daily Diet Cost Breakdown by Food Component.

As shown in Table IV, other foods dominate the average daily diet cost at 59.7%, followed by vegetables at 21.4% and fruits at 18.9%, consistent with evidence that nutritious food components constitute a disproportionate share of diet expenditure [4][10].

TABLE IV. SUMMARY OF FOOD COMPONENTS COST

Average Vegetable Cost	$0.76/day (21.4%)
Average Fruit Cost	$0.68/day (18.9%)
Average Other Foods	$2.13/day (59.7%)

IX. VISUALIZATION OF COUNTRIES BASED ON AFFORDABLE HEALTHY DIET

Fig 8 highlights the cost extremes in 2024. South Sudan records the highest daily healthy diet cost at $8.39/day, followed by Japan ($7.62) and Mongolia ($7.02), reflecting the disproportionate cost burden in conflict-affected and geographically isolated economies. At the other end, Ireland ($2.83/day) and Belgium ($2.81/day) represent the most affordable, consistent with the pattern of lower diet costs in high-income European countries.

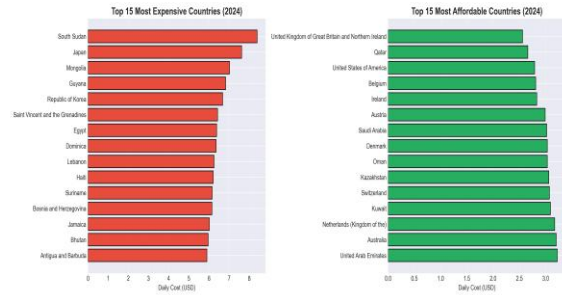


Fig 8: Top 15 Most Expensive and Most Affordable Countries by Daily Healthy Diet Cost, 2024.

X. RELATIONSHIP BETWEEN KEY PRICE ESTIMATES

Fig 9 shows that vegetables, fruit, and total food component costs are strongly positively correlated, collectively driving overall healthy diet cost. The distribution of diet cost is right-skewed, concentrated between $3–5 PPP/day, with high-cost outliers above $6–7 PPP/day likely reflecting import dependency, supply chain constraints, or inflationary pressures in specific countries.

Fig 9: Pair Plot of Key Diet Cost Variables by Region, 2017–2024.

XI. PREDICTIVE MODELLING HEALTHY DIET COST

A. Feature Engineering and Data Preprocessing:

Ten additional features were created as part of preprocessing the data for model training. These features were time-related features (number of years from 2017), label-encoded features (region, country, type of expense), expense per food items (vegetables, fruits, total), one interaction feature for region and year, and total cost per country. Food item costs with any missing values were filled with zeros, implying that there is no recorded expense value. The dataset consisted of 1,379 entries, with a target value range of $1.70-$8.39 per day.

B. Model Training and Comparison:

Five different models were tested using the scikit-learn library for regression: Linear Regression, Ridge Regression with alpha=1.0, Random Forest Regressor (n=100, max_depth=15), XGBoost Regressor (n=100, max_depth=5, learning_rate=0.1), and Gradient Boosting Regressor (n=100, max_depth=5). Gradient Boosting performed well by achieving the highest value of R^2, which was 0.9607. It also recorded the lowest value of RMSE, at $0.178. This was marginally better than Random Forest, whose R^2 and RMSE were 0.96 and $0.190. Fig 11(a) shows the model comparison across all three metrics. Model performance was assessed using RMSE, MAE, and R^2 score on the held-out test set (Table III).

TABLE III: MACHINE LEARNING MODEL PERFORMANCE COMPARISON

Model	R^2 Score	RMSE ($)	MAE ($)
Gradient Boosting	0.96	0.18	0.13
Random Forest	0.96	0.19	0.13
XGBoost	0.95	0.20	0.13
Linear Regression	0.93	0.23	0.17
Ridge Regression	0.93	0.23	0.17

Model Assessment: The Gradient Boosting model exhibits excellent prediction ability, with a mean residual of -$0.003, which indicates that there is no systematic bias in the predictions. As illustrated in Fig 10(a), the model's predictions correlate well with the cost data along the diagonal line.

From the residual chart in Fig 10(b), the dots are randomly dispersed around zero, implying that the model fits well. With a confidence interval of 95%, the prediction values deviate ±$0.350 from the actual values. This deviation is very minimal when compared to the total objective range of $1.70-$8.

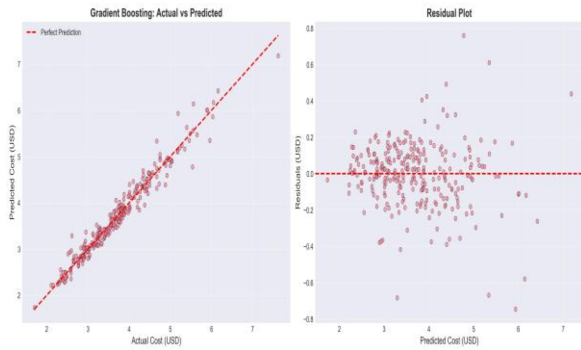


Fig 11(b): Gradient Boosting —actual vs predicted daily healthy diet cost on the test set And Gradient Boosting — residual plot showing prediction error distribution.

XII. KEY FINDINGS

Prevalence of Undernourishment: Global undernourishment declined by 24.74% from 2000 (14.19%) to 2022 (10.68%), reflecting meaningful progress. Africa records the highest regional prevalence (18.06%), followed by Asia (10.26%), while Europe remains the lowest (2.50%).

Within Africa, the Central African Republic leads at 40.78%; within Asia, Southern Asia is highest at 15.84%. Countries such as Tajikistan, Armenia, and Moldova demonstrate that sustained improvement is achievable, with Moldova reducing prevalence to 2.5%. Conversely, a sharp deterioration is observed across several African and Middle Eastern countries after 2010–2011[1][2], with conflict-affected states — including Syria, North Korea, Gabon, and the Democratic Republic of Congo — showing the most severe worsening trends, consistent with evidence linking political instability to food insecurity [12][13][14][15].

Healthy Diet Cost:

Global average daily healthy diet cost stands at $3.68/day, rising 42.1% from $3.14 in 2017 to $4.46 in 2024 [1][11]. Africa, Americas, and Asia show the highest regional averages with considerable within-region variation, as summarized in Table VI.

TABLE VI: REGIONAL AVERAGE DAILY HEALTHY DIET COST

Region	Average Daily Cost	Std. Deviation
Africa	$3.79/day	±$0.80
Americas	$3.72/day	±$1.00
Asia	$3.67/day	±$0.89

Europe records the lowest costs for both vegetables and fruits. Asia leads in fruit cost ($0.74/day, 2.07% above the global average). Other food components (grains, pulses, wheat) constitute the largest share of diet cost (59.7%), followed by vegetables (21.4%) and fruits (18.9%). At the country level, South Sudan records the highest cost ($8.39/day) while the United Kingdom is the most affordable ($2.56/day), with some countries paying three to four times more than others [3].

Regional Disparities:

Regional comparisons indicate that lower-income regions face both higher undernourishment rates and greater difficulty affording healthy diets, reflecting structural inequalities in food systems, income distribution, and agricultural productivity [3]. The exploratory analysis suggests a potential relationship between diet affordability and malnutrition prevalence, though additional socioeconomic factors also contribute to this relationship.

XIII. CONCLUSION AND FUTURE OUTCOMES

This study demonstrates that healthy diet costs have risen significantly across all regions from 2017 to 2024, while undernourishment — though declining globally by 24.74% — remains critically high in Africa and conflict-affected states. These findings corroborate SOFI 2025 and prior work by Smith and Haddad [1][3], confirming that even where food supply is sufficient, economic access remains the primary barrier to nutritional adequacy [10]. The SARIMAX forecasts project continued cost increases through 2029, particularly in Africa, Asia, and the Americas, signaling that food affordability will remain a pressing policy challenge.

The reasons for the crisis include rising food prices, climate problems associated with food production, conflicts leading to damage in food systems, and structural reasons like post-harvest loss and inadequate social safety nets, which increase the vulnerability of developing countries [1][2][11]. Specifically, policy makers should ensure that more emphasis be placed on subsidy programs and improved social protection structures in areas where the cost of living is relatively higher. Further, investments should also be made towards enhancing local food production and climate-resilient farming techniques that can help reduce dependence on imports. Additionally, subsidies should be directed at crop types that are nutritious rather than staple foods. Most importantly, improvements should be made towards transportation and logistics in rural settings – inefficient infrastructure and the absence of cold storage facilities account for losses of highly perishable foods such as fruits and vegetables, which contribute to increased dietary expenses as shown in this research. Future research should incorporate additional socioeconomic indicators — such as income inequality and conflict indices — and explore ensemble forecasting approaches to improve long-term predictive precision.

Increasing trends for annual spending on healthier eating habits will be observed for all regions till 2029 according to SARIMAX model predictions (see Fig 10). This result reflects global trends associated with rising costs of food products [11]. According to the results, the greatest increase will be observed for the continent of Africa, followed by Asia and Americas.

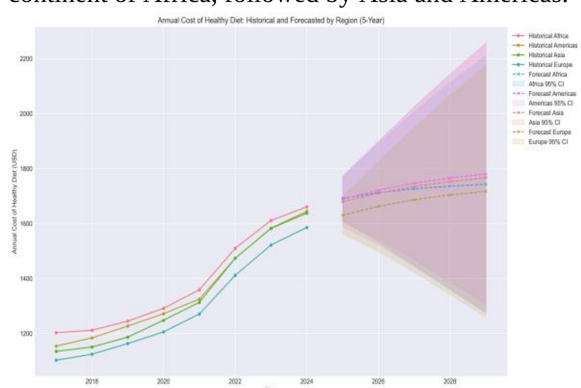


Fig 10: SARIMAX five-year forecast of annual healthy diet cost by region, 2024–2029, with 95% confidence intervals.

REFERENCES

[1] Food and Agriculture Organization of the United Nations (FAO), The State of Food Security and Nutrition in the World 2025 (SOFI 2025), 2025. [Online]. Available: https://openknowledge.fao.org/server/api/core/bit

streams/de95e011-1af9-4b28-9a07-d8ce61f8aa6c/content/state-food-security-and-nutrition-2025/executive-summary.html

[2] World Health Organization (WHO), "Malnutrition – Fact Sheet," 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/malnutrition

[3] L. C. Smith and L. Haddad, "Reducing child undernutrition: Past drivers and priorities for the post-MDG era," World Dev., vol. 68, pp. 180–204, 2015.

[4] D. Headey and H. Alderman, "The relative caloric prices of healthy and unhealthy foods differ systematically across income levels and continents," J. Nutr., vol. 149, no. 11, pp. 2020–2033, 2019.

[5] Food and Agriculture Organization of the United Nations (FAO), FAOSTAT Database. [Online]. Available: https://www.fao.org/faostat/en/#data/CAHD

[6] J. D. Hunter, "Matplotlib: A 2D graphics environment," Comput. Sci. Eng., vol. 9, no. 3, pp. 90–95, 2007, doi: 10.1109/MCSE.2007.55.

[7] Seaborn Development Team, Seaborn, Version 0.11.0, 2020. [Online]. Available: https://seaborn.pydata.org

[8] World Bank, Purchasing Power Parity (PPP). [Online]. Available: https://www.worldbank.org/en/programs/icp

[9] G. van Rossum and F. L. Drake Jr., Python 3 Reference Manual. Scotts Valley, CA, USA: CreateSpace, 2009.

[10] A. Herforth, Y. Bai, A. Venkat, K. Mahrt, A. Ebel, and W. Masters, "Cost and affordability of healthy diets across countries," Lancet Glob. Health, vol. 8, no. 1, pp. e59–e70, 2020.

[11] World Bank, Food Security Update: World Bank Solutions to Food Insecurity, Mar. 2026. [Online]. Available: https://www.worldbank.org/en/topic/agriculture/brief/food-security-update

[12] A. (Alireza) Adkanian and M. Khosravi, "Investigating the roots of hunger in North Korea," Int. J. Soc. Sci. Rev., 2024. [Online]. Available: https://ijssrr.com/journal/article/view/1935

[13] A. de Waal, "Hunger in global war economies: Understanding the decline and return of famines," Disasters, 2024, doi: 10.1111/disa.12661.

[14] Al Jazeera, "UN sounds alarm over rising hunger crisis in eastern DR Congo," Nov. 7, 2025. [Online]. Available: https://www.aljazeera.com/news/2025/11/7/un-sounds-alarm-over-rising-hunger-crisis-in-eastern-dr-congo

[15] Action Against Hunger, "The hungriest countries in the world in 2026," 2026. [Online]. Available: https://www.actionagainsthunger.org.uk/our-impact/stories/the-hungriest-countries-in-the-world