Unveiling Socioeconomic Factors Shaping Global Food Prices and Security: A Machine Learning Approach

Shan Shan

Sociology Department, Zhejiang University, Hangzhou, China

Asian Journal of Agriculture and Development (AJAD)
Vol. 21 | 20th Anniversary Issue October 2024
Complete Lineup

ABSTRACT

Global concern over food prices and security has intensified due to armed conflicts such as the Russia–Ukraine war, the recent COVID-19 pandemic, and climate change. Traditional analysis of global food prices and their associations with socioeconomic factors has relied on static linear regression models. However, the complexity of socioeconomic factors and their implications extend beyond simple linear relationships. To address this gap, this study aimed to identify critical socioeconomic characteristics and multidimensional relationships influencing food prices and security by incorporating determinants, critical characteristic identification, and comparative model analysis. Machine learning tools were used to uncover the socioeconomic factors influencing global food prices from 2000 to 2022. Four key dimensions of food price security were identified: economic and population metrics, military spending, health spending, and environmental factors. Given the complexity of these dimensions, the support vector regression model's efficiency rendered it most suitable for precise analyses among the models assessed. The findings revealed shifts in the food price index, particularly in relation to military expenditure, healthcare expenditure, and economic contributions. Based on these findings, the research proposes a framework centered around six thematic areas related to (1) governance; (2) health and environment; (3) environment, climate, and military spending; (4) comprehensive analytical tools; (5) collaborative efforts; and (6) resilience and sustainability. This framework enables policymakers to further expand on actionable recommendations.

Keywords: environment and growth, global economics, food price, support vector regression, machine learning
JEL codes: Q01, O13, Q18

Contact Shan Shan shshan@zju.edu.cn https://doi.org/10.37801/ajad2024.21.2.1

INTRODUCTION

Addressing global food price volatility is essential for achieving "No Poverty" and "Zero Hunger," as outlined in UN Sustainable Development Goals (SDGs) 1 and 2, ensuring universal access to affordable and nutritious food. However, amid the rapidly changing global landscape, increasing food prices and security concerns have become significant challenges. These issues stem from a complex interplay of factors, including armed conflicts, large-scale health crises, and the relentless advancement of climate change, necessitating in-depth analysis and innovative solutions. The urgency of addressing these challenges has been underscored by recent global events, such as the Russia–Ukraine conflict and the COVID-19 pandemic. Recent events have increased the vulnerability of the global food systems. Disrupted supply chains, pressure on agricultural productivity, and rising inflation have led to dramatic fluctuations in food prices, thereby posing serious threats to global food security. Consequently, advanced data-driven methods can address the pressing need to examine the complex determinants of food price and security. What factors contribute to this situation, and how do multilevel socioeconomic features influence food security? To address these issues, traditional research methods often use linear regression models. Although these models have significantly contributed to our understanding of food prices and security, they tend to oversimplify multiple interactions between phenomena and their numerous socioeconomic determinants. The intricate nature of different socioeconomic systems often defies simple linear assumptions and necessitates more advanced analytical approaches (Bar-Yam 2004). Machine learning methods, when combined with techniques like feature extractions, have proven highly effective in elucidating the complex interplay of systems, especially in social phenomena (Grimmer, Roberts, and Stewart 2021). These methods enhance data analysis efficiency by identifying the most influential variables. This not only reduces the number of variables but also helps to clarify the dynamics of complex systems (Awan et al. 2019). Echoing the importance and effectiveness of machine learning in data science, Jordan and Mitchell (2015) highlight machine learning's capability to adeptly handle a diverse range of data and problem types across various scenarios, thereby offering a robust, computerized learning experience compared to traditional methods.

Focusing on the SDG food security domain, the Food and Agriculture Organization (FAO) Food Price Index (FFPI)[1] provides insights into the changing prices of food in the global market. This index serves as a valuable tool for monitoring and analyzing trends and fluctuations in food prices, which can significantly impact food security, trade, and agricultural policy. Although several studies have investigated the relationship between specific social dimensions, such as urbanization and food prices, less attention has been devoted to understanding how multilevel social factors influence food prices on a global scale. Existing literature on the volatility of food prices underscores the influence of various factors, including oil prices, global economic activity, fertilizer prices, and geopolitical risks, along with the socioeconomic and political ramifications of fluctuations in food prices. Studies by Zmami and Ben-Salha (2023), Brinkman et al. (2010), and Hendrix and Haggard (2015) identify the key drivers of food price volatility and its adverse effects on nutrition, health, and social stability, particularly among vulnerable populations and across governance types. However, a significant gap remains in understanding the long-term impacts of such volatility on global food prices and security as well as the potential mitigating role of technological advances in agriculture. This gap underscores the need for comprehensive research that integrates economic, social, and political factors to enhance food security strategies and to address the challenges posed by food price volatility.

1 https://www.fao.org/prices/en/

To address this gap in the literature, this study examines the multilevel socioeconomic factors that influence global food prices and security. By leveraging the capabilities of machine learning, this study aims to overcome the limitations of traditional linear models and capture the complex and often nonlinear relationships among various variables influencing global food prices and security.

Specifically, this study focuses primarily on the effects of socioeconomic factors that influence food security and aims to determine the extent to which common machine learning algorithms capture better the complex interactions among the socioeconomic determinants of global food prices and security compared to traditional linear regression models. It also sought to identify the specific socioeconomic factors that have had the greatest influence on global food prices and security in recent decades and proposes approaches that enable the transition of insights derived from machine learning models into actionable policy interventions aimed at stabilizing food prices and enhancing global food security.

The overall goal was to shed light on the complexity inherent in these issues and advocate a comprehensive understanding of data-driven policymaking in this domain. This study aims to address the impact of food price fluctuations and security issues on a global scale, by providing policymakers with robust strategies. Additionally, the methodological framework presented in this study serves as an innovative approach for formulating effective mitigation strategies and emphasizes the need for future research in this area.

The study enables a comprehensive examination of the factors influencing global food prices and security. Moreover, it identifies the correlations among these factors and determines which are most influential on global food prices. It further develops a public policy framework based on these findings to offer actionable recommendations.

METHODS

The study adopted a comprehensive methodological approach, beginning with an analysis of multidimensional social causes using data mining and feature extraction techniques. This process involved reviewing numerous academic papers to identify and categorize key topics into four socioeconomic dimensions. Machine-learning techniques were then employed to quantify and highlight the essential features, followed by a comparative evaluation of popular models for data fitting. This method underscores using advanced computational techniques for detailed social cause analysis.

Data

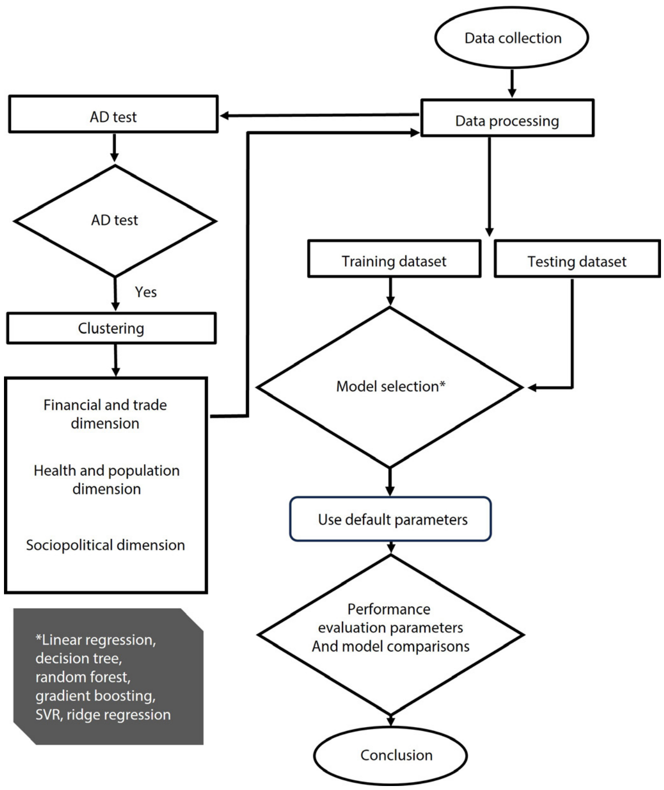
The study focused on analyzing food prices and security indicators to capture the dynamics of global food prices and security. This indicator served as the target variable and was derived from the FFPI (2014–16 = 100).[2] A total of 104 features were selected from the World Bank's World Development Indicators (WDI) database,[3] as described in Appendix Table 1. This database encompasses a spectrum of indicators, including political, military, economic, and health domains. Political and military indicators include statistics such as armed force personnel (total number and percentage of the total labor force) and military expenditure (percentage of gross domestic product [GDP]). Economic indicators include metrics such as foreign direct investment (net outflows and inflows, percentage of GDP); urban population (and its annual growth); contributions of agriculture, forestry, and fisheries (value added as a percentage of GDP); services and trade (value added as a percentage of GDP); GDP per capita (adjusted to constant 2015 USD and its annual growth); manufacturing value added; manufacturing exports (percentage of goods exports); exports of goods

2 https://www.fao.org/worldfoodsituation/foodpricesindex/en/

3 https://datatopics.worldbank.org/world-development-indicators/

and services; goods and services expenses (percentage of total expenditure); and urban population (percentage of total population). Health indicators include life expectancy at birth, lifetime risk of maternal death (both as a percentage and in terms of a "1 in x" rate), health spending measures, such as current health spending (percentage of GDP and per capita in the US and international dollars adjusted for PPP), and domestic general government health expenditure (percentage of GDP). Collectively, these indicators provide a comprehensive overview of global food price, security, and related societal factors.

Figure 1. Research design of the study



Data Splitting and Kernel Density Estimate

The research design shown in Figure 1 involves several key steps. Initially, multidimensional social causes derived from data mining and feature techniques were examined. This facilitated analyzing numerous academic papers, identifying popular topics, and classifying them into four distinct categories representing different socioeconomic dimensions. Subsequently, machine learning techniques were employed to quantify and highlight key features. Finally, the commonly used models were studied to compare the fit of the data.

Within the dataset, 80 percent was allocated for training[4] purposes and the remaining 20 percent was allocated for testing[5] using a fixed random seed of 42 for consistency and model reliability (Shan 2023; Bisong 2019). Ensuring similarity in the distribution of the training and testing sets is crucial for the model to effectively learn relevant patterns during training and generate accurate estimates during testing (Jain, Duin, and Mao 2000). The methodology comprises four main steps: the Anderson-Darling (AD) test, feature clustering, feature selection (F-value test), and identification of the top 30 key features. These steps aim to refine the pool of features and select those that are most influential in predicting food price and security.

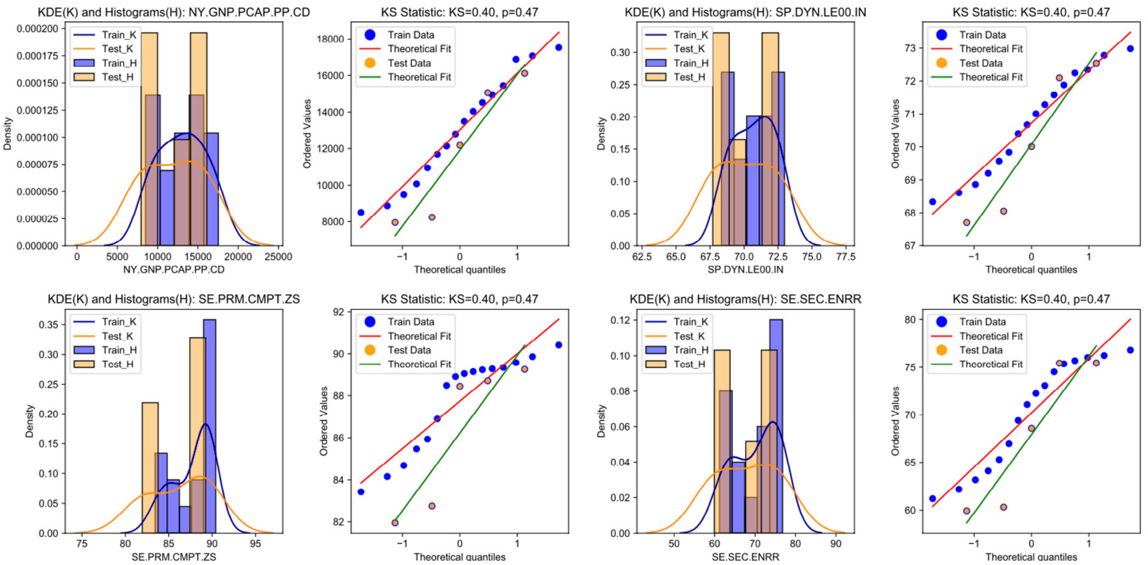
4 Machine learning (ML) model training is the process of teaching a machine learning algorithm to detect patterns and predict outcomes by exposing it to labeled data. This approach starts with random parameters that are repeatedly modified to minimize the discrepancy between its predictions and the training data labels (https://www.hpe.com/emea_europe/en/what-is/ml-model-training.html#:~:text=Machine%20Learning%20(ML)%20model%20training,and%20the%20training%20data%20labels).

5 Once your machine learning model is built (with your training data), you need unseen data to test your model. This data is called testing data, and you can use it to evaluate the performance and progress of your algorithms' training and adjust or optimize it for improved results (https://www.obviously.ai/post/the-difference-between-training-data-vs-test-data-in-machine-learning#:~:text=What%20is%20Testing%20Data%3F,optimize%20it%20for%20improved%20results).

The AD test assessed the data distribution and identified variables conforming to a normal distribution performed on each column of both the training dataset (X_train) and testing dataset (X_test). Using kernel density estimation plots displayed the results visually as histograms and fitted curves, with $p < 0.05$. Additionally, the analysis code included a "failed_tests" function to monitor variables failing to meet the AD testing criteria, particularly those with nonsignificant results ($p \geq 0.05$), indicating a deviation from a normal distribution. Figure 2 provides examples of variables that successfully passed the AD test, which evaluates the data fit to a normal distribution with a significance level set at 0.05. Abbreviations include GNP for gross national product and PPP for purchasing power parity expressed in current international dollars. Life expectancy is the expected lifespan at birth; primary completion rate is the percentage of students finishing primary school; and secondary enrollment ratio is the percentage of secondary education enrollment. The top-left panel displays the probability density functions for "NY.GNP.PCAP.PP.CD," representing GNP per capita in PPP terms, with GNP values on the x-axis and their frequency as a probability density on the y-axis. The top-right panel features a quantile-quantile plot for "SP.DYN.LE00.IN," illustrating life expectancy at birth against a theoretical normal distribution, with theoretical quantiles on the x-axis and empirical life expectancy quantiles on the y-axis. In the middle-left panel, the probability density functions for "SE.PRM.CMPT.ZS" show the primary completion rate percentage on the x-axis and the probability density of these rates on the y-axis. The middle right panel's quantile-quantile plot for "SE.SEC.ENRR" displays the secondary enrollment ratios, with theoretical quantiles on the x-axis and observed ratios on the y-axis. The legend indicates the datasets: the training phase with a blue line for the initial data exposure and the testing phase with an orange line for the model performance assessment. Additional variables from AD exploratory data analysis are presented in Appendix Figure 1.

This comprehensive approach ensures a thorough assessment of the distributional properties of the data. Non-normally distributed variables were transformed prior to further analysis. Subsequently, feature clustering is employed to group the features into clusters based on their relationships and similarities, with each cluster containing highly correlated features.

Figure 2. Assessment of distribution fits across economic and educational metrics: GNP per capita, life expectancy, primary completion, and secondary enrollment ratios



This approach simplifies the model by reducing the number of features without an excessive loss of information. This was accomplished by characterizing each cluster using a single feature that best represented the common characteristics of the group. Feature selection using the F-value test was conducted to measure the degree of linear dependence between the two random variables and identify the most significant features. In this study, the F-value test helped determine the features that contributed the most to the model's ability to predict food prices and security. From a refined pool of variables, the top 30 key features in terms of impact (as determined by the F-value test and feature clustering) were identified, further reducing the feature set to a manageable number without compromising predictive power. These key characteristics were subsequently utilized in the final model to forecast food prices and security. This systematic and robust methodological approach enables a comprehensive analysis of the variables that influence global food prices and security.[6]

Modeling evaluation

The comprehensive methodology utilized six different machine learning models (Table 1) chosen for their ability to address various aspects of the data. The selected models included support vector regression (SVR), ridge regression, linear regression, random forest regression, gradient boosting, and decision-tree regression.

The SVR is effective in high-dimensional spaces and is resilient to overfitting,[7] making it particularly advantageous for handling nonlinear data (Liao, Dai, and Kuosmanen 2024; Roy and Chakraborty 2023). However, the requirements for appropriate kernel and regularization parameter selection, as well as the computational intensity for large datasets, represent well-known limitations of this model (Moradi and Minaei 2020; Ding et al. 2015).

Ridge regression efficiently and effectively handles multicollinearity. However, it may exhibit poor performance when dealing with nonlinear data and may introduce increased complexity owing to regularization (Kigo, Omondi, and Omolo 2023).

Linear regression, although straightforward and efficient for modeling linear relationships, is constrained by its assumption of linearity and sensitivity to outliers, limiting its applicability to modeling complex relationships (Peña and Slate 2006; Rousseeuw and Leroy 2005).

Random forest regression constructs a set of decision trees from randomly selected subsets of the training set and averages their predictions. It was selected for its superior prediction accuracy and reduced risk of overfitting compared with single decision trees (Sun et al. 2024; Sahin 2020; Ali et al. 2012). This approach entails greater computational complexity and reduced interpretability compared with a single decision tree; however, it strikes a balance between interpretability and complexity (Fratello and Tagliaferri 2019; Kirasich, Smith, and Sadler 2018; James et al. 2013).

Gradient boosting can capture complex relationships and patterns; however, although it provides insights into feature importance, it can be computationally intensive for large datasets and is

6 In addressing potential confounding variables, handling the non-normal and/or nonlinear conditional dependencies often seen in predictive models poses a significant challenge. To ensure valid model diagnostics, these dependencies require that no assumptions are made about the conditional distributions of the model output, even in cases of non-normally and nonlinearly dependent predictions (Dinga et al. 2020). While the study utilizes randomized data splitting to ensure unbiased training and testing sets, this method primarily supports model validation rather than directly addressing confounders. A correlation map serves as a valuable diagnostic tool in the preliminary data analysis phase to identify potential relationships. However, it does not inherently adjust for confounding variables.

7 Overfitting means creating a model that matches (*memorizes*) the training set, or the subset of the dataset used to train a model, so closely that the model fails to make correct predictions on new data. An overfit model is analogous to an invention that performs well in the lab but is worthless in the real world (https://developers.google.com/machine-learning/crash-course/overfitting/overfitting).

Table 1. General comparison of six machine learning models used in this study

Machine Learning Model	Mechanics	Pros and Cons
SVR	Fits the best line within a threshold error margin and employs various kernel functions to handle nonlinear relationships.	**Pros:** Effective in high-dimensional spaces and with nonlinear data; robust against overfitting in high-dimensional space. **Cons:** Requires the selection of appropriate kernel and regularization parameters; can be computationally intensive for large datasets.
Ridge regression	Minimizes the sum of squared residuals with an added penalty proportional to the square of the magnitude of the coefficients.	**Pros:** Handles multicollinearity well; simple and computationally efficient. **Cons:** Might not perform well with nonlinear data; model complexity can be increased by introducing regularization.
Linear regression	Attempts to model the relationship between dependent and independent variables by fitting a linear equation to the observed data.	**Pros:** Simple to understand and implement; efficient for problems with linear relationships. **Cons:** Assumes a linear relationship; sensitive to outliers; cannot model complex relationships such as those in nonlinear data.
Random forest regression	Creates a set of decision trees from randomly selected subsets of the training set and averages their predictions.	**Pros:** High predictive accuracy; less prone to overfitting than a single decision tree. **Cons:** Computationally more intensive than a single decision tree; lower interpretability than a single decision tree but better than that of the SVR.
Gradient boosting	Iteratively builds new models that focus on reducing the errors made by the previous model.	**Pros:** Effective for capturing complex relationships and patterns; provides insights into feature importance. **Cons:** Computationally intensive for large datasets; prone to overfitting; requires careful hyperparameter tuning; challenging to interpret owing to its ensemble nature.
Decision tree regression	Splits the dataset into subsets based on feature values; this process is recursively repeated until the tree reaches a predefined depth or purity.	**Pros:** High interpretability; can capture nonlinear relationships. **Cons:** Prone to overfitting, especially with complex datasets; can create overly complex trees.

susceptible to overfitting (Jun 2021). Furthermore, its ensemble nature necessitates tuning hyperparameters and complicates interpretation (Huber et al. 2022; Freeman et al. 2016).

Decision tree regression offers good interpretability and the ability to capture nonlinear relationships. However, it is prone to overfitting and can generate excessively complex trees (Costa and Pedreira 2023; James et al. 2013; Lou, Caruana, and Gehrke 2015).

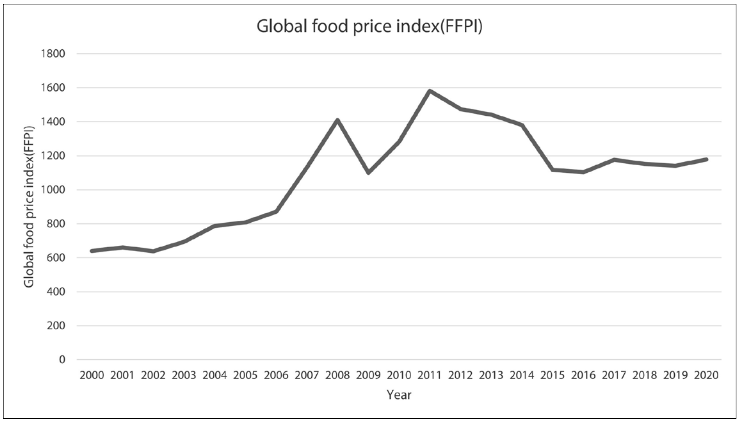
Each model was employed to ensure a robust and comprehensive analysis of the dataset, accounting for its strengths and weaknesses.

RESULTS AND DISCUSSION

Food security includes food availability, food access, utilization (Barrett 2010), and stability that combines access and availability (FAO 2006). Food availability ensures an adequate supply; access implies that people can easily obtain the food they need; utilization denotes sufficient nutrient intake; and stability refers to the consistent ability to access food (FAO 2006; Shemyakina 2022).

Food prices showed a general upward trend between 2000 and 2022 with significant fluctuations, particularly between 2006 and 2011. The period after 2011 was more stable, albeit at higher price

Figure 3. Food prices (derived from the FFPI) from 2000 to 2020



levels, than that in the early 2000s. This temporal pattern highlights the dynamics of food prices, which are influenced by a range of socioeconomic and environmental factors (Figure 3).

Machine learning techniques were employed to extract features and identify four dimensions that significantly affected food prices: economic aspects, social health, political and military factors, and demographic characteristics. The efficiency of the model was evaluated using two methods: random forest and SVR. Using data mining and feature extraction techniques, key variables associated with food prices and security along with their determinants were extracted from an extensive database of previous studies. This process yielded 30 critical features spanning a variety of factors ranging from carbon emission levels and population metrics to GDP per capita, military expenditure, health expenditure, and trade services.

Shifts in Food Prices and Security in Relation to Complex Social Factors

This study investigated the relationships between significant fluctuations in food prices, security, and key socioeconomic variables (Table 2, Appendix Figure 2). Changes in food prices and security are intricately linked to a complex network of social factors encompassing economic, political, and cultural components. Understanding these relationships is crucial for devising effective strategies for managing food price and security. Data mining and machine learning tools facilitated identifying 30 key characteristics that affect food prices and security across all socioeconomic levels (Figure 4). The cell at the intersection of row i and column j shows the correlation between the ith and jth features. Contrasting colors indicate the strength and direction of the correlation: a value close to 1 signifies a strong positive correlation, whereas a value close to −1 indicates a strong negative correlation. A value close to 0 suggests no linear relationship. The clustering algorithm organizes rows and columns to group similar features together, thereby improving interpretability. This heatmap offers a visual representation of the correlation between various socioeconomic indicators. For clarity and ease of interpretation, each row or column corresponds to a specific indicator, denoted by both its abbreviation and full name as listed below:

Column and row labels

Year (year)	The calendar year of the data
FFPI	An index measuring the change in international prices of a basket of food commodities
Total population (SP.POP.TOTL)	The total number of people in a given country or region
Population growth, annual percentage (SP.POP.GROW)	The year-over-year percentage change in total population

Appendix Table 1 includes all abbreviations and full names as necessary for the indicators represented on the heatmap.

Economic Indicators and Their Impacts on Global Food Prices

Economic indicators play a pivotal role in shaping the complex network underlying global food price and security (Table 3). Gross national

Table 2. Summary of exploratory data analysis

Variable	Mean	Median	SD	IQR	CI Lower Bound	CI Upper Bound
Food Price	1,110.02	1,136.14	285.82	450.58	969.97	1,250.06
NY.GNP.ATLS.CD (×10^12) GNI, Atlas method (current USD)	64.89	67.65	18.00	27.73	56.07	73.71
NY.GNP.PCAP.CD GNI per capita, Atlas method (current USD)	9,116.49	9,646.94	1,993.31	3,149.37	8,139.79	10,093.2
NY.GDP.MKTP.CD (×10^12) GDP at market prices (current USD)	65.38	70.25	17.55	27.33	56.78	73.97
NV.AGR.TOTL.ZS Agriculture, value added (% of GDP)	3.76	3.92	0.39	0.70	3.57	3.96
NE.EXP.GNFS.ZS Exports of goods and services (% of GDP)	27.90	28.52	2.26	3.47	26.79	29.01
NE.IMP.GNFS.ZS Imports of goods and services (% of GDP)	27.29	27.85	2.06	3.41	26.28	28.30
TG.VAL.TOTL.GD.ZS Trade (% of GDP)	44.90	45.20	3.71	5.78	43.08	46.72
BM.TRF.PWKR.CD.DT (×10^11) Personal remittances, paid (current USD)	3.12	3.33	1.09	1.93	2.59	3.66
DT.ODA.ODAT.PC.ZS Net ODA received per capita (current USD)	17.96	18.55	4.26	5.17	15.87	20.05
NE.RSB.GNFS.ZS External balance on goods and services (% of GDP)	0.61	0.65	0.25	0.32	0.49	0.74
NY.GDS.TOTL.ZS Gross domestic savings (% of GDP)	25.55	26.08	1.41	2.09	24.85	26.24
NE.CON.PRVT.ZS Household final consumption expenditure (% of GDP)	57.49	57.04	1.46	1.87	56.78	58.21
NY.GDP.MINR.RT.ZS Mineral rents (% of GDP)	0.33	0.29	0.21	0.24	0.23	0.43
BX.TRF.PWKR.DT.GD.ZS Personal remittances, received (% of GDP)	0.64	0.65	0.10	0.15	0.59	0.69
MS.MIL.XPND.CD (×10^12) Military expenditure (current USD)	1.50	1.65	0.36	0.56	1.33	1.68
NY.GDP.PCAP.CD GDP per capita (current USD)	9,188.97	9,881.72	1,933.54	3,095.88	8,241.55	10,136.39
SH.DYN.MORT Mortality rate, under-5 (per 1,000 live births)	52.46	50.45	10.86	17.38	47.14	57.79

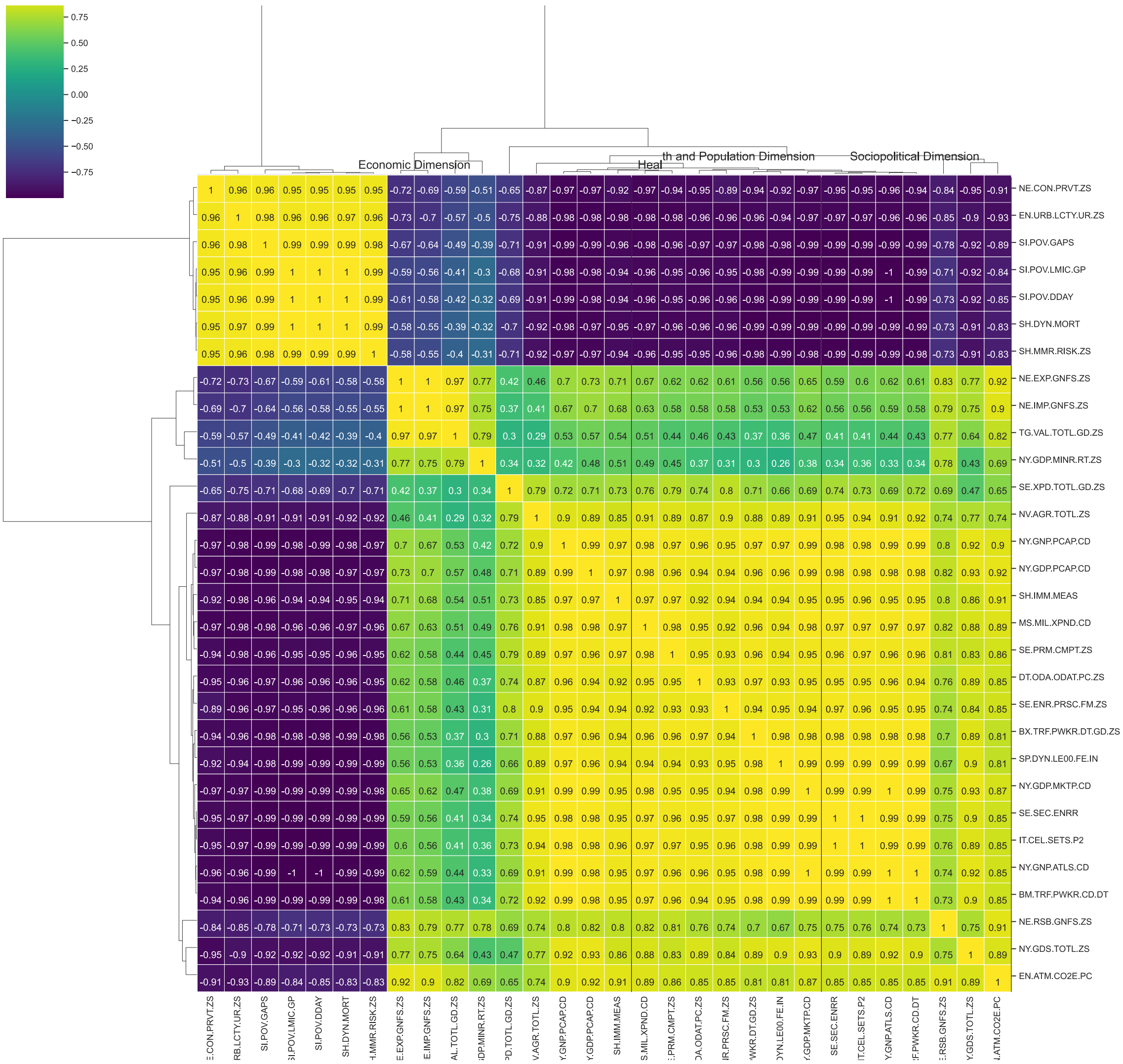
Continued on next page

Table 2 continued

Variable	Mean	Median	SD	IQR	CI Lower Bound	CI Upper Bound
SP.DYN.LE00.FE.IN Life expectancy at birth, female (years)	73.23	73.36	1.54	2.51	72.48	73.99
SH.MMR.RISK.ZS Lifetime risk of maternal death (%)	0.62	0.60	0.10	0.17	0.57	0.67
EN.URB.LCTY.UR.ZS Urban population living in areas where elevation is below 5 m (% of total population)	16.15	16.07	0.18	0.26	16.06	16.24
SI.POV.GAPS Poverty gap at national poverty lines (%)	4.89	4.35	2.07	3.30	3.87	5.90
SI.POV.LMIC.GP Poverty gap at USD 1.90 a day (2011 PPP) (%)	13.83	13.25	4.42	7.03	11.66	15.99
SI.POV.DDAY Poverty headcount ratio at USD 1.90 a day (2011 PPP) (% of population)	16.23	15.25	6.24	10.07	13.17	19.28
SH.IMM.MEAS Immunization, measles (% of children ages 12–23 months)	81.44	83.91	4.58	6.17	79.19	83.68
SE.PRM.CMPT.ZS Primary completion rate, total (% of relevant age group)	87.74	88.98	2.28	3.48	86.63	88.86
SE.SEC.ENRR School enrollment, secondary (% gross)	70.21	71.67	5.59	10.44	67.47	72.95
SE.ENR.PRSC.FM.ZS School enrollment, primary and secondary (gender parity index)	0.97	0.97	0.02	0.03	0.96	0.98
EN.ATM.CO2E.PC CO_2 emissions (metric tons per capita)	4.47	4.54	0.22	0.27	4.36	4.58
IT.CEL.SETS.P2 Mobile cellular subscriptions (per 100 people)	69.35	79.59	32.16	56.54	53.59	85.11
SE.XPD.TOTL.GD.ZS Government expenditure on education, total (% of GDP)	4.17	4.20	0.18	0.25	4.09	4.26

Notes: This table presents a detailed summary of exploratory data analysis, including measures of central tendency (mean, median); variability (standard deviation [SD], interquartile range [IQR]), and 95 percent confidence interval (CI) bounds (CI Lower Bound, CI Upper Bound) for each variable within the subset of the database. Here, "×10^n" denotes the magnitude for large numbers and the precision of each numerical value to maintain clarity and readability.

Figure 4. Heatmap of clustering results depicting an interconnected variable matrix



income (GNI), both at the state and per capita levels ("NY.GNP.ATLS.CD" and "NY.GNP.PCAP.CD, respectively), serves as an important measure of economic performance. Although a higher GNI often enables countries to invest more in agricultural development and food imports, it does not guarantee equitable food security, particularly in the absence of equal wealth distribution (Brinkman et al. 2010). Furthermore, a higher GNI per capita can increase individual purchasing power but may also increase fuel demand, potentially leading to price inflation. The GDP and its per capita variant ("NY.GDP.MKTP.CD" and "NY.GDP.PCAP.CD" respectively) are also critical indicators. A high GDP typically signals enhanced food security and facilitates investment in agricultural production (Barrett 2021; Wheeler and von Braun 2013; Nelson et al. 2010). However, interpreting GDP in isolation can be misleading because the percentage of GDP originating from agriculture, forestry, and fishing ("NV.AGR.TOTL.ZS") offers deeper insights into the stability of a country's food supply. A higher percentage of agriculture often signals

Table 3. Economic indicators impacting global food prices

Indicator	Interpretation
NY.GNP.ATLS.CD	GNI calculated using the Atlas method in current USD
NY.GNP.PCAP.CD	GNI per capita, computed using the Atlas method in current USD
NY.GDP.MKTP.CD	GDP in current USD
NV.AGR.TOTL.ZS	Percentage of GDP generated from agriculture, forestry, and fishing
NE.EXP.GNFS.ZS	Exports of goods and services as a percentage of GDP
NE.IMP.GNFS.ZS	Imports of goods and services as a percentage of GDP
TG.VAL.TOTL.GD.ZS	Merchandise trade as a percentage of GDP
BM.TRF.PWKR.CD.DT	Personal remittances paid in current USD
DT.ODA.ODAT.PC.ZS	Net ODA received per capita in current USD
NE.RSB.GNFS.ZS	Net balance of goods and services as a percentage of GDP.
NY.GDS.TOTL.ZS	Gross domestic savings as a percentage of GDP
NE.CON.PRVT.ZS	Private consumption as a percentage of GDP
NY.GDP.MINR.RT.ZS	Value added by the mining sector as a percentage of GDP
BX.TRF.PWKR.DT.GD.ZS	Personal remittances received as a percentage of GDP
MS.MIL.XPND.CD	Military expenditure in current USD
NY.GDP.PCAP.CD	GDP per capita in current USD

stability but may expose the country to global price shocks, particularly if it is excessively reliant on a single sector.

Trade dynamics, including export and import rates ("NE.EXP.GNFS.ZS" and "NE.IMP.GNFS.ZS", respectively) and the total value of goods traded ("TG.VAL.TOTL.GD.ZS"), directly influence food prices. For instance, countries with high export rates, particularly those in the agriculture sector, may deplete their local food reserves and drive domestic prices. Conversely, reliance on food imports exposes a country to fluctuations in international prices (Barrett 2010; 2013). Financial inflows and outflows, such as personal remittances both paid and received ("BM.TRF.PWKR.CD.DT" and "BX.TRF.PWKR.DT.GD.ZS" respectively), can impact household income, thereby influencing food security. Nevertheless, higher income levels may escalate demand, subsequently increasing food prices. Official development assistance (ODA) ("DT.ODA.ODAT.PC.ZS"), often directed at agricultural and food security programs, can stabilize food prices in recipient countries. However, it may also contribute to a negative balance between goods and services ("NE.RSB.GNFS.ZS"), rendering them susceptible to international price fluctuations (Prakash 2011).

Economic factors, such as domestic savings ("NY.GDS.TOTL.ZS") and private consumption rates ("NE.CON.PRVT.ZS") can also influence food prices. Increased domestic savings can foster better investment opportunities in agriculture, potentially stabilizing local food prices. By contrast, increased private consumption may drive up demand, consequently impacting food prices.

Military expenditure ("MS.MIL.XPND.CD") and the value added by the mining sector ("NY.GDP.MINR.RT.ZS") can indirectly influence food security. Multiple studies have shown that conflict negatively impacts food security, which is defined as "having, at all times, both physical and economic access to sufficient food to meet dietary needs for a productive and healthy life" (USAID 2024; Shemyakina 2022). For instance, countries with high military spending may allocate fewer resources to agriculture, thereby creating an imbalance that affects food security (Brinkman et al. 2010). Conflict leads to the destruction and deterioration of key production factors such as land, labor, and entrepreneurship. It also impacts total factor productivity and economic efficiency (Collier et al. 2003). Armed conflict affects access to markets and market efficiency as well. Furthermore, conflict negatively impacts the accumulation of human capital. In terms of food security, declines in food availability, access, and utilization have been shown to negatively impact long-term health outcomes.

Table 4. Demographic indicators affecting global food prices

Indicator	Interpretation
SH.DYN.MORT	Mortality rate for children under 5 years, expressed per 1,000 live births
SP.DYN.LE00.FE.IN	Life expectancy at birth for females
SH.MMR.RISK.ZS	Maternal mortality risk
EN.URB.LCTY.UR.ZS	Percentage of the population living in urban areas
SI.POV.GAPS	Poverty gap at USD 1.90/d (2011 international prices) as a percentage
SI.POV.LMIC.GP	Poverty headcount ratio at low- to middle-income levels
SI.POV.DDAY	Poverty headcount ratio at USD 1.90/d (2011 international prices) as a percentage of the population

Understanding these diverse economic indicators is crucial for formulating a comprehensive strategy to manage global food price volatility. An effective policy response requires an integrated approach that incorporates these economic indicators, as well as social, political, and demographic variables.

Demographic Indicators and Their Impacts on Global Food Prices

Demographic indicators offer another invaluable perspective to understand the complexities of global food prices (Table 4). Mortality rates for children under five years of age ("SH.DYN.MORT"), life expectancy for women ("SP.DYN.LE00.FE.IN"), and maternal mortality risks ("SH.MMR.RISK.ZS") serve as significant measures of societal well-being, indirectly reflecting the status of food and nutritional security (Barrett 2010). Elevated child and maternal mortality rates often signal underlying issues pertaining to food insecurity and malnutrition, which can influence the food price dynamics. Further, the demographic composition of the urban and rural populations ("EN.URB.LCTY.UR.ZS") has a profound influence on food distribution systems. Urban areas typically rely on complicated logistics and often depend on imported food, rendering them susceptible to global price fluctuations and exacerbating food security concerns (Brinkman et al. 2010). Indicators related to poverty, such as the poverty gap at USD 1.90 a day ("SI.POV.GAPS") and poverty headcount ratios at low- to middle-income levels ("SI.POV.LMIC.GP") and at USD 1.90/d ("SI.POV.DDAY"), elucidate socioeconomic disparities within populations. These indicators are particularly relevant because individuals living near or below the poverty line are more susceptible to even minor fluctuations in food prices. Their reduced purchasing power affects their ability to access nutritious food, thereby adversely affecting the stability and security of food prices (Brinkman et al. 2010; Barrett 2010). The complexity of these demographic indicators highlights the necessity for an integrated policymaking approach. Considering demographic factors along with economic, social, and political variables offers a comprehensive strategy for effectively managing the global food price ecosystem.

Sociopolitical Features and Their Impacts on Global Food Prices

Among the sociopolitical features (Table 5), the percentage of children vaccinated against measles ("SH.IMM.MEAS") serves as an indicator of public health infrastructure, which in turn affects agricultural labor productivity (Barrett 2010). Similarly, primary school attainment ("SE.PRM.CMPT.ZS") and gross secondary school enrollment rates ("SE.SEC.ENRR") reflect a country's investment in human capital, influencing the availability of skills for agricultural productivity and innovation, and thereby impacting food prices in the long term (Brinkman et al. 2010). The gender parity index for school enrollment in primary and secondary education ("SE.ENR.PRSC.FM.ZS") can also indicate the level of gender equality within a society, which has been shown to affect food security by fostering more equitable distribution and decision-making within

Table 5. Sociopolitical factors affecting global food prices

Indicator	Interpretation
SH.IMM.MEAS	Percentage of children aged 12–23 months who have been immunized against measles
SE.PRM.CMPT.ZS	Primary school completion rate, measured as a percentage of the relevant age group
SE.SEC.ENRR	Gross enrollment rate in secondary education
SE.ENR.PRSC.FM.ZS	Gender parity index for school enrollment in primary and secondary education
EN.ATM.CO2E.PC	CO_2 emissions per capita in metric tons
IT.CEL.SETS.P2	Mobile cellular subscriptions per 100 people
SE.XPD.TOTL.GD.ZS	Total expenditure on education as a percentage of GDP

Table 6. Model performance

Model	MAE	RMSE	R^2
SVR	0.471065	0.315638	0.993144
Ridge regression	0.768931	1.123438	0.975596
Linear regression	1.046580	2.525464	0.945140
Gradient boosting	2.320526	9.801146	0.787093
Random forest	2.523905	11.825548	0.743118
Decision tree	3.721429	22.082024	0.520320

Note: RMSE - root MSE

households and communities (Barrett 2010). CO_2 emissions per capita ("EN.ATM.CO2E.PC") are commonly used as indicators of a country's level of industrialization, and can potentially affect the sustainability of agriculture, thus influencing food prices. Mobile subscriptions per 100 people ("IT.CEL.SETS.P2") serve as a measure of the rate of technological dissemination, impacting food prices through market efficiency and information symmetry (Brinkman et al. 2010). Finally, total expenditure on education as a percentage of GDP ("SE.XPD.TOTL.GD.ZS") represents a commitment to skill development that could improve agricultural processes and promote food price stability (Barrett 2010). These data suggest that investments in education, gender equity, technology adoption, and environmental sustainability collectively influence global food prices and food security.

The multidimensional impact of these sociopolitical indicators on food price security underscores the need for an integrated policy approach. They also highlight the intricate connections among education, gender equity, technological reach, environmental sustainability, and their collective impact on food prices and security. Although these categories serve distinct roles, they are also interconnected. Economic indicators primarily encompass economic and trade dimensions that directly impact food security. Demographic indicators offer insights into health and population distributions that are crucial for targeted food security interventions. Finally, sociopolitical indicators mirror society's level of development, infrastructure, and values, indirectly influencing food security through policies and awareness initiatives.

Model Efficacy Based on SVR

The model's validation process includes normalization, grid search with cross-validation, and performance metrics and evaluation. Features are normalized to ensure equal weighting in the model calculations; a grid search with five-fold cross-validation is conducted on the SVR model to determine optimal parameters. All models then undergo five-fold cross-validation to assess their generalizability. Models are evaluated using mean absolute error (MAE), mean squared error (MSE), and the coefficient of determination (R^2). These metrics help identify the best-performing model based on prediction accuracy and ability to explain variability. The ridge regression and SVR demonstrated excellent performance in predicting changes in food prices influenced by various socioeconomic factors, as outlined in Table 6. These machine learning algorithms accurately forecast food prices and security as well as decipher the intricate connections among diverse social factors.

However, ridge regression encounters challenges in managing high-dimensional spaces and handling numerous training examples. Determining the optimal penalty value for ridge regression requires a comprehensive understanding of the algorithm, as well as considerable experimentation and validation. Therefore, it is imperative to develop a model that is both accurate and generalizable. In contrast, SVR distinguishes itself through its resilience to overfitting and proficiency in handling nonlinear relationships.

Algorithm 1: SVR for food price impact

Input

- Feature matrix X containing economic, demographic, and environmental indicators
- Target vector y representing global food prices.

Output

- Optimized SVR model parameters
- Evaluation metrics: MAE = 0.471065, MSE = 0.315638, R^2 = 0.993144

SVR function

- $f(\mathrm{x}) = \mathrm{w}^{\mathrm{T}} \boldsymbol{\phi}(\mathrm{x}) + b$
- Subject to the optimization of:
  - $\min_{\mathbf{w},b,\xi,\xi^*} \frac{1}{2} \parallel \mathbf{w} \parallel^2 + C \sum_{i=1}^{n}(\xi_i + \xi_i^*)$
  - Constraints for ξ_i, ξ_i^* given ϵ insensitivity zone:
    - $y_i - \mathbf{w}^\top \phi(\mathbf{x}_i) - b \leq \epsilon + \xi_i$
    - $\mathbf{w}^\top \phi(\mathbf{x}_i) + b - y_i \leq \epsilon + \xi_i^*$
    - $\xi_i, \xi_i^* \geq 0$

Procedure

1. Preprocess X by scaling features to a mean of zero and a variance of one.
2. Initialize the SVR model with a hyperparameter space comprising:
   - Regularization parameter C
   - Epsilon ϵ
   - Kernel coefficient γ
   - Kernel type kernel
3. Apply a five-fold cross-validation grid search to determine the optimal set of parameters for the SVR model.
4. Train the SVR model using the optimal parameters on the scaled dataset.
5. Predict food prices using the trained SVR model, employing the kernel trick $K(\mathrm{x}_i, \mathrm{x}_j) = \phi(\mathrm{x}_i)^{\mathrm{T}} \phi(\mathrm{x}_j)$ to project data into a higher-dimensional space.[8]
6. Compute performance metrics: MAE, MSE, and R^2 using the SVR model's predictions.

End procedure

Machine Learning Reveals the Importance of Interdisciplinary Approaches in Addressing Food Price Security

Advanced machine learning techniques have been employed to investigate food price security, including ridge regression, decision tree regressors, random forest regression, and SVR, and have demonstrated superior predictive performance. Through data mining, feature extraction, and automated database interactions, the study identified a refined list of 30 key multidimensional variables spanning economic, social, and political factors through a comprehensive review of existing research. These variables serve as a foundation for understanding the complex dynamics that influence food prices and, consequently, security. The high effectiveness of SVR along with the commendable performance of the linear and ridge

8 In this algorithm, $\phi(\mathrm{x})$ represents the feature mapping to a higher-dimensional space, which is implicit in the kernel trick, with K being the kernel function. The regularization parameter C governs the trade-off between the model's complexity and the extent to which deviations larger than ϵ are tolerated. ξ and ξ^* are slack variables that permit violations of the ϵ insensitivity zone, which is essential for capturing errors in the model. This optimization is typically solved using a quadratic programming solver in the dual space, which is reflected in the grid search step of the algorithm.

regression models underscores the potential of machine learning as a tool in this context.

Using machine learning techniques not only enhances predictive power but also highlights the crucial role of interdisciplinary approaches. These approaches facilitate identifying and integrating critical socioeconomic factors, such as health expenditures and economic contributions, which are essential for formulating robust strategies to address diverse challenges. These results highlight the need for an interdisciplinary approach that incorporates expertise from relevant fields, including economics, political science, environmental science, and public health to address the complexities of food price security on a global scale. For instance, understanding the political stability index in the context of health care spending can offer nuanced perspectives that may be overlooked with a single-discipline focus. Similarly, incorporating environmental variables along with economic indicators such as GDP or GNI allows for a more informed analysis of food production and distribution networks.

Such an interdisciplinary perspective is invaluable for policymakers who often operate at the intersection of these distinct yet interconnected domains. The insights gained from machine learning models provide a quantifiable basis for targeted policy interventions, offering empirical evidence for strategies aimed at stabilizing food prices, thereby enhancing food security and the overall well-being of the population.

STUDY LIMITATIONS AND PROSPECTS

This study utilized macro-level data rather than individual factors for analysis. A consequence of the approach and methodology employed to address macro-level issues is that inherent limitations arise when addressing specific factor issues and will not address local and regional scale dynamic issues. A limitation lies in the temporal scope of the study, spanning from 2000 to 2022, which may not adequately capture the long-term trends affecting global food price security. Further, reliance on data sourced from the WDI and FAO restricts the scope of the study and may overlook critical factors not included in these datasets. Additionally, the predominantly quantitative focus of this study might not fully account for certain qualitative aspects such as political events and cultural shifts. Although this study effectively presents significant factors and their associations with food prices, it is important to note that these do not necessarily imply causal relationships. Establishing causal links within multilevel data remains a challenge, particularly given current methodological constraints.

The constantly evolving landscape of global food prices offers several avenues for future research that warrant close attention. One such avenue is the customization of models to specific regions, considering geographical differences in socioeconomic, political, and environmental conditions. Moreover, the paradox of overfitting still exists. Exploring advanced computational methods, including machine learning and artificial intelligence, is promising but requires careful consideration of the substantial data and computational resources they demand, as well as their potential for overfitting, especially with large datasets and high dimensions.

In addition to methodological richness, longitudinal studies that examine changes in food prices and security over longer periods can offer invaluable insights into evolving trends and challenges; hence, enabling more accurate forecasts for future scenarios. While this study represents progress in identifying the factors influencing food prices and security, subsequent studies could benefit from incorporating additional variables such as the impacts of climate change or international trade relations in order to provide comprehensive policy recommendations.

From a governance standpoint, integrating insights from policy studies into future research endeavors can aid in developing robust strategies for managing food price volatility and security issues. This policy-oriented approach provides a structured framework for decision-makers and stakeholders concerned with food security. Furthermore, assessing the generalizability of the results across different local and global settings will refine the models for broader applicability and real-world impact. Finally, given the intricate nature of food security challenges, future research should advocate interdisciplinary collaboration among fields such as economics, other social sciences, environmental science, and computer science. Such interdisciplinary engagement is crucial for fostering innovative solutions for complex and pressing global issues. Future research in these areas can build upon the methodological framework presented herein to contribute not only to academic discourse but also to pragmatic, data-driven strategies aimed at addressing this important problem.

ACTIONABLE RECOMMENDATIONS

Based on the above findings, the research proposes a framework for food security through six strategic actions, focusing on policy revision, resource allocation, and regulatory enforcement to manage external shocks and market volatility (Table 7). It emphasizes using advanced analytical tools and international cooperation to predict and mitigate factors affecting food security, while also enhancing public awareness through educational campaigns. The framework advocates for reallocating military budgets and increasing investments in health and environmental sustainability to ensure robust and localized responses to food security challenges.

CONCLUSION

This study highlights the importance of economic and demographic indicators for global food price security. The findings reveal that factors such as GNI and GDP are indicators of more than just a country's national income or output; they are essential for understanding and ensuring food availability for the population. These economic metrics, along with demographic indicators such as child mortality rates and poverty levels, offer a multifaceted perspective on food security that is closely linked to household income and urbanization. Governments could redirect some military spending toward agricultural development to enhance food security and tie international military aid to food security benchmarks. Integrating health and agricultural policies could mitigate the impact of rising healthcare costs on food affordability, offering subsidies for essential foods and investments that benefit both health and agriculture.

Table 7. Framework for addressing food security

Category	Action
Governance	Revise economic policies to cushion impacts from external shocks by creating reserves, diversifying food sources, and enhancing local production.
Health and environment	Allocate more resources to health systems and sustainable practices to minimize impacts from health crises and climate change on food security.
Military spending review	Reassess military budgets to support critical sectors such as agriculture, ensuring allocations do not undermine food security initiatives.
Comprehensive analytical tools	Deploy machine learning and advanced analytics to predict and mitigate factors affecting food security; establish training programs for analysts.
Collaborative international efforts	Foster international cooperation to strengthen global and regional food security responses, sharing data, resources, and technology.
Resilience and sustainability	Launch campaigns to raise awareness about food price dynamics and promote sustainable behaviors like reducing food waste and supporting local producers.
	Develop and enforce regulations to prevent speculative and manipulative practices that cause food price volatility.

The devised approach, which integrates data mining and machine learning techniques, identifies 30 key characteristics that influence food security across multiple socioeconomic dimensions. The use of diverse models, including SVR, ridge regression, and decision tree analysis, provides new insights into the factors that influence food price and security. The effectiveness of these models highlights the need for an interdisciplinary strategy to address the food price security complexities.

In summary, addressing global food price security represents a multifaceted challenge that necessitates a comprehensive multidimensional approach. Beyond economic considerations, a comprehensive understanding of demographic characteristics is required. Both sets of indicators are indispensable for developing effective policy responses to address the complexities of global food price security. The findings revealed significant shifts in the FPI, particularly in relation to military expenditure, healthcare expenditure, and economic contributions. This study advances our understanding of the complex relationship between socioeconomic variables and food price security. By harnessing machine learning techniques and considering multiple dimensions, policymakers can make informed decisions to improve food prices and security on a global scale. Furthermore, with further studies and local adaptations, the results hold promise for generalizing to other contexts.

ACKNOWLEDGMENT

The author expresses gratitude for the comments and feedback received during the panel discussion at the 4th International Conference on Agriculture, Food Security and Safety held on 14–15 July 2023 via online platform.

DATA AVAILABILITY

The data that support the findings of this study are available at https://github.com/shanshanfy/GlobalFoodPrices.git. Accessibility is upon request.

CODE AVAILABILITY

The analysis code can be accessed at https://github.com/shanshanfy/GlobalFoodPrices.git. Accessibility is upon request.

DEVELOPER ENVIRONMENT AVAILABILITY

The GlobalFoodPricesPackages.yaml developer environment file for the Conda open-source package management system is provided through: https://github.com/shanshanfy/GlobalFoodPrices.git. This file allows for isolated environments to manage packages without interference. The file contains the configuration of the project's Python environment, including channels, dependencies, and library versions. Accessibility is upon request.

REFERENCES

Ali, J., R. Khan, N. Ahmad, and I. Maqsood. 2012. "Random Forests and Decision Trees." *International Journal of Computer Science Issues* 9(5): 272–78.

Awan, S.E., M. Bennamoun, F. Sohel, F.M. Sanfilippo, B.J. Chow, and G. Dwivedi. 2019. "Feature Selection and Transformation by Machine Learning Reduce Variable Numbers and Improve Prediction for Heart Failure Readmission or Death." *PloS One* 14(6): e0218760. DOI: 10.1371/journal.pone.0218760

Bar-Yam, Y. 2004. "Multiscale Variety in Complex Systems." *Complexity* 9(4): 37–45. https://doi.org/10.1002/cplx.20014

Barrett, C.B. 2010. "Measuring Food Insecurity." *Science* 327(5967): 825–28. DOI: 10.1126/science.118276

——— (ed.). 2013. *Food Security and Sociopolitical Stability*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199679362.001.0001

———. 2021. "Overcoming Global Food Security Challenges through Science and Solidarity." 2021. *American Journal of Agricultural Economics* 103(2): 422–47. https://doi.org/10.1111/ajae.12160

Bisong, E. 2019. *Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide to Beginners*. Berkeley, California: Apress. https://doi.org/10.1007/978-1-4842-4470-8_24

Brinkman, H.J., S. de Pee, I. Sanogo, L. Subran, and M.W. Bloem. 2010. "High Food Prices and the Global Financial Crisis Have Reduced Access to Nutritious Food and Worsened Nutritional Status and Health." *The Journal of Nutrition* 140(1): 153S-161S. https://doi.org/10.3945/jn.109.110767

Collier, P., V.L. Elliott, H. Hegre, A. Hoeffler, M. Reynal-Querol, and N. Sambanis. 2003. *Breaking the Conflict Trap: Civil War and Development Policy*. Washington, DC: World Bank and Oxford University Press.

Collier, P. Breaking the Conflict Trap: Civil War and Development Policy, vol. 41181, no. 4 (World Bank Publications, 2003).

Costa, V.G., and C.E. Pedreira. 2023. "Recent Advances in Decision Trees: An Updated Survey." *Artificial Intelligence Review* 56: 4765–4800. https://doi.org/10.1007/s10462-022-10275-5

Ding, Y., L. Cheng, W. Pedrycz, and K. Hao. 2015. "Global Nonlinear Kernel Prediction for Large Data Set with a Particle Swarm-Optimized Interval Support Vector Regression." *IEEE Transactions on Neural Networks and Learning Systems* 26(10): 2521–34. DOI: 10.1109/TNNLS.2015.2426182

Dinga, R., L. Schmaal, B.W.J.H. Penninx, D.J. Veltman, and A.F. Marquand. 2020. "Controlling for Effects of Confounding Variables on Machine Learning Predictions." *BioRxiv* 2020.08.17.255034. https://doi.org/10.1101/2020.08.17.255034

FAO (Food and Agriculture Organization). 2006. "Food Security." *Policy Brief* 2. https://www.fao.org/fileadmin/templates/faoitaly/documents/pdf/pdf_Food_Security_Cocept_Note.pdf

Fratello, M., and R. Tagliaferri. 2019. "Decision Trees and Random Forests." In *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics* 1: 374–83.

Freeman, E.A., G.G. Moisen, J.W. Coulston, and B.T. Wilson. 2016. "Random Forests and Stochastic Gradient Boosting for Predicting Tree Canopy Cover: Comparing Tuning Processes and Model Performance." *Canadian Journal of Forest Research* 46(3): 323–39. https://doi.org/10.1139/cjfr-2014-0562

Grimmer, J., M.E. Roberts, and B.M. Stewart. 2021. "Machine Learning for Social Science: An Agnostic Approach." *Annual Review of Political Science* 24: 395–419. https://doi.org/10.1146/annurev-polisci-053119-015921

Hendrix, C.S., and S. Haggard. 2015. "Global Food Prices, Regime Type, and Urban Unrest in the Developing World." *Journal of Peace Research* 52(2): 143–57. https://doi.org/10.1177/0022343314561599

Huber, F., A. Yushchenko, B. Stratmann, and V. Steinhage. 2022. "Extreme Gradient Boosting for yield estimation compared with Deep Learning Approaches." *Computers and Electronics in Agriculture* 202: 107346. https://doi.org/10.1016/j.compag.2022.107346

Jain, A.K., R.P.W. Duin, and J. Mao. 2000. "Statistical Pattern Recognition: A Review." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22(1): 4–37. https://doi.org/10.1109/34.824819

James, G., D. Witten, T. Hastie, and R. Tibshirani. 2013. *An Introduction to Statistical Learning with Applications in R*. New York: Springer.

Jordan, M.I., and T.M. Mitchell. 2015. "Machine Learning: Trends, Perspectives, and Prospects." *Science* 349(6245): 255–60. DOI: 10.1126/science.aaa8415

Jun, M.J. 2021. "A Comparison of a Gradient Boosting Decision Tree, Random Forests, and Artificial Neural Networks to Model Urban Land Use Changes: The Case of the Seoul Metropolitan Area." *International Journal of Geographical Information Science* 35(11): 2149–67. https://doi.org/10.1080/13658816.2021.1887490

Kigo, S.N., E.O. Omondi, and B.O. Omolo. 2023. "Assessing Predictive Performance of Supervised Machine Learning Algorithms for a Diamond Pricing Model." *Scientific Reports* 13: 17315. https://doi.org/10.1038/s41598-023-44326-w

Kirasich, K., T. Smith, and B. Sadler. 2018. "Random Forest vs Logistic Regression: Binary Classification for Heterogeneous Datasets." *SMU Data Science Review* 1(3): 9. https://scholar.smu.edu/datasciencereview/vol1/iss3/9

Liao, Z., S. Dai, and T. Kuosmanen. 2024. "Convex Support Vector Regression." *European Journal of Operational Research* 313(3): 858–70. https://doi.org/10.1016/j.ejor.2023.05.009

Lou, Y., R. Caruana, and J. Gehrke. 2012. "Intelligible Models for Classification and Regression." In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 150–58. https://doi.org/10.1145/2339530.2339556

Moradi, R., R. Berangi, and B. Minaei. 2020. "A Survey of Regularization Strategies for Deep Models." *Artificial Intelligence Review* 53: 3947–986. https://doi.org/10.1007/s10462-019-09784-7

Nelson, G.C., M.W. Rosegrant, A. Palazzo, et al. 2010. *Food Security, Farming, and Climate Change to 2050: Scenarios, Results, Policy Options.* Washington DC, USA: International Food Policy Research Institute (IFPRI). https://doi.org/10.2499/9780896291867

Peña, E.A., and E.H. Slate. 2006. "Global Validation of Linear Model Assumptions." *Journal of the American Statistical Association* 101(473): 341–54. https://doi.org/10.1198/016214505000000637

Prakash, A. (ed.). 2011. *Safeguarding Food Security in Volatile Global Markets*. Rome: Food and Agriculture Organization of the United Nations. http://www.fao.org/docrep/013/i2107e/i2107e.pdf

Rousseeuw, P.J., and A.M. Leroy. 2005. *Robust Regression and Outlier Detection*. John Wiley & Sons.

Roy, A., and S. Chakraborty. 2023. "Support Vector Machine in Structural Reliability Analysis: A Review." *Reliability Engineering & System Safety* 233: 109126. https://doi.org/10.1016/j.ress.2023.109126

Sahin, E.K. 2020. "Assessing the Predictive Capability of Ensemble Tree Methods for Landslide Susceptibility Mapping Using XGBoost, Gradient Boosting Machine, and Random Forest." *SN Applied Sciences* 2: 1308. https://doi.org/10.1007/s42452-020-3060-1

Shan, S. 2023. "Machine Learning Insights into Multifaceted Social Impacts on Global Urban Slum Population." *Research Square.* https://doi.org/10.21203/rs.3.rs-3137402/v1

———. 2024. "Identification of Socioeconomic Factors Influencing Global Food Price Security Using Machine Learning." *arXiv*: 2403.04231. https://doi.org/10.48550/arXiv.2403.04231

Shemyakina, O. 2022. "War, Conflict, and Food Insecurity." *Annual Review of Resource Economics* 14: 313–32. https://doi.org/10.1146/annurev-resource-111920-021918

Sun, Z., G. Wang, P. Li, H. Wang, M. Zhang, and X. Liang. 2024. "An Improved Random Forest Based on the Classification Accuracy and Correlation Measurement of Decision Trees." *Expert Systems with Applications* 237 (Part B): 121549. https://doi.org/10.1016/j.eswa.2023.121549

USAID (United States Agency for International Development). 2024. "Agriculture and Food Security." *USAID: What We Do*. Accessed on 15 November 2024. https://www.usaid.gov/agriculture-and-food-security#:~:text=Food%20security%20means%20having%2C%20at,hunger%20or%20fear%20of%20hunger

Wheeler, T., and J. von Braun. 2013. "Climate Change Impacts on Global Food Security." *Science* 341(6145): 508–13. https://doi.org/10.1126/science.1239402

Zmami, M., and O. Ben-Salha. 2023. "What Factors Contribute to the Volatility of Food Prices? New Global Evidence." *Agricultural Economics - Czech* 69(5): 171–84. https://doi.org/10.17221/99/2023-AGRICECON

APPENDICES

Appendix Table 1. The target variable and 104 features used in the analysis

Abbreviation	Full Name
NY.GDP.MKTP.CD	GDP at market prices (current USD)
NY.GDP.MKTP.KD.ZG	GDP growth (annual %)
NY.GDP.DEFL.KD.ZG	GDP deflator (annual %)
NV.AGR.TOTL.ZS	Agriculture, value added (% of GDP)
NV.IND.TOTL.ZS	Industry, value added (% of GDP)
NE.EXP.GNFS.ZS	Exports of goods and services (% of GDP)
NE.IMP.GNFS.ZS	Imports of goods and services (% of GDP)
NE.GDI.TOTL.ZS	Gross domestic investment (% of GDP)
CM.MKT.LCAP.GD.ZS	Market capitalization of listed companies (% of GDP)
MS.MIL.XPND.GD.ZS	Military expenditure (% of GDP)
IT.CEL.SETS.P2	Mobile cellular subscriptions (per 100 people)
TG.VAL.TOTL.GD.ZS	Trade (% of GDP)
BM.TRF.PWKR.CD.DT	Personal remittances, received (current USD)
BX.KLT.DINV.CD.WD	Foreign direct investment, net inflows (BoP, current USD)
DT.ODA.ODAT.PC.ZS	ODA and official aid received (per capita)
FP.CPI.TOTL.ZG	Inflation, consumer prices (annual %)
BX.KLT.DINV.WD.GD.ZS	FDI, net inflows (% of GDP)
BM.KLT.DINV.WD.GD.ZS	FDI, net outflows (% of GDP)
BM.KLT.DINV.CD.WD	Foreign direct investment, net outflows (BoP, current USD)
FM.LBL.BMNY.GD.ZS	Broad money (% of GDP)
FS.AST.CGOV.GD.ZS	Claims on Central Government (% of GDP)
EN.ATM.CO2E.KD.GD	CO_2 emissions (kg per 2010 USD of GDP)
EN.ATM.CO2E.PP.GD.KD	CO_2 emissions (kg per PPP USD of GDP)
EN.ATM.CO2E.PP.GD	CO_2 emissions (metric tons per capita and PPP of GDP)
NY.GDP.COAL.RT.ZS	Coal rent (% of GDP)
SH.XPD.CHEX.GD.ZS	Current health expenditure (% of GDP)
FS.AST.PRVT.GD.ZS	Domestic credit to private sector (% of GDP)
FD.AST.PRVT.GD.ZS	Domestic credit to private sector by banks (% of GDP)
SH.XPD.GHED.GD.ZS	Government health expenditure (% of GDP)
GC.XPN.TOTL.GD.ZS	Government expenditure (% of GDP)
NE.RSB.GNFS.ZS	Reserves of foreign exchange and gold (% of GDP)
NY.GDP.FRST.RT.ZS	Forest rent (% of GDP)
NY.GDP.MKTP.KD	GDP at market prices (constant 2010 USD)
NY.GDP.PCAP.KD	GDP per capita (constant 2010 USD)
NY.GDP.PCAP.KD.ZG	GDP per capita growth (annual %)
NY.GDP.PCAP.PP.KD	GDP per capita, PPP (constant 2011 international USD)
NY.GDP.PCAP.PP.CD	GDP per capita, PPP (current international USD)
SL.GDP.PCAP.EM.KD	GDP per capita, employed (constant 2010 USD)
NY.GDP.MKTP.PP.KD	GDP, PPP (constant 2011 international USD)
NY.GDP.MKTP.PP.CD	GDP, PPP (current international USD)
NE.CON.GOVT.ZS	Government consumption (% of GDP)
SE.XPD.TOTL.GD.ZS	Total expenditure on education (% of GDP)
NY.GDS.TOTL.ZS	Gross domestic savings (% of GDP)

Continued on next page

Appendix Table 1 continued

Abbreviation	Full Name
NY.GNS.ICTR.ZS	Gross national savings (% of GDP)
NE.GDI.FTOT.ZS	Gross fixed capital formation (% of GDP)
NE.CON.PRVT.ZS	Household final consumption expenditure (% of GDP)
NV.IND.MANF.ZS	Manufacturing, value added (% of GDP)
NY.GDP.MINR.RT.ZS	Mineral rent (% of GDP)
FM.AST.PRVT.GD.ZS	Private sector credit (% of GDP)
NY.GDP.NGAS.RT.ZS	Natural gas rent (% of GDP)
GC.NLD.TOTL.GD.ZS	Net lending (+) / Net borrowing (–) (% of GDP)
NY.GDP.PETR.RT.ZS	Oil rent (% of GDP)
BX.TRF.PWKR.DT.GD.ZS	Personal remittances, paid (% of GDP)
NV.SRV.TOTL.ZS	Services, value added (% of GDP)
NY.GDP.TOTL.RT.ZS	Total natural resources rent (% of GDP)
NE.TRD.GNFS.ZS	Trade in goods and services (% of GDP)
BG.GSR.NFSV.GD.ZS	Net service exports (% of GDP)
MS.MIL.XPND.ZS	Military expenditure (% of GNI)
MS.MIL.XPND.CD	Military expenditure (current USD)
NV.IND.MANF.KD.ZG	Manufacturing output growth (% annual)
SP.DYN.LE00.FE.IN	Life expectancy at birth, female (years)
SP.DYN.LE00.MA.IN	Life expectancy at birth, male (years)
SH.MMR.RISK.ZS	Maternal mortality ratio (modeled estimate, per 100,000 live births)
EG.CFT.ACCS.UR.ZS	Access to clean fuels and technologies for cooking (% of urban population)
EG.ELC.ACCS.UR.ZS	Access to electricity (% of urban population)
SH.STA.ODFC.UR.ZS	Open defecation (% of urban population)
SH.H2O.BASW.UR.ZS	Basic water services (% of urban population)
SH.STA.BASS.UR.ZS	Basic sanitation services (% of urban population)
SH.H2O.SMDW.UR.ZS	Safely managed drinking water services (% of urban population)
SH.STA.SMSS.UR.ZS	Safely managed sanitation services (% of urban population)
EN.URB.LCTY.UR.ZS	Urban population living in slums (% of urban population)
EN.URB.MCTY.TL.ZS	Urban population living in megacities (% of total)
SP.URB.TOTL	Total urban population
SP.URB.TOTL.IN.ZS	Urban population (% of total population)
SP.URB.GROW	Urban population growth (annual %)
NE.EXP.GNFS.KD.ZG	Export of goods and services (real growth %)
NE.CON.TOTL.KD.ZG	Consumption, total (real growth %)
NY.GDP.PCAP.CD	GDP per capita (current USD)
SI.POV.GAPS	Poverty gap at national poverty lines (%)
SI.POV.LMIC.GP	Poverty gap at USD 3.20/d (2011 PPP) (%)
SI.POV.UMIC.GP	Poverty gap at USD 5.50/d (2011 PPP) (%)
SI.POV.DDAY	Poverty headcount ratio at USD 1.90/d (2011 PPP) (%)
SI.POV.LMIC	Poverty headcount ratio at USD 3.20/d (2011 PPP) (%)
SI.POV.UMIC	Poverty headcount ratio at USD 5.50/d (2011 PPP) (%)

Note: Each feature is detailed with its corresponding World Bank abbreviation and full descriptive name.

Appendix Figure 1. Variables that passed the Anderson–Darling (AD) test

Appendix Figure 2. Variables that passed the Anderson–Darling (AD) test