International Journal of Engineering Trends and Technology
ISSN: 2231–5381 / https://doi.org/10.14445/22315381/IJETT-V71I12P226
Volume 71 Issue 12, 277-291, December 2023

Original Article

Forecasting Prices of Agricultural Commodities using Machine Learning for Global Food Security: Towards Sustainable Development Goal 2

Anket Patil[1], Dhairya Shah[2], Abhishek Shah[3], Radhika Kotecha[4]

[1,2,3,4]*K. J. Somaiya Institute of Technology, University of Mumbai Mumbai, India*.

[1]*Corresponding Author : anket.patil@somaiya.edu*

Received: 25 August 2023 Revised: 22 November 2023 Accepted: 01 December 2023 Published: 06 December 2023

Abstract - *Global food security is vital for promoting human health, upholding social well-being, and ultimately achieving the United Nations' Sustainable Development Goal (SDG) 2: Zero Hunger. Conversely, it is influenced by a multitude of factors, with the dynamics of agricultural commodity prices playing a significant role. Recognizing the potential of Machine Learning in agricultural applications, this work delves into exploring the price dynamics of key agricultural commodities across various global producers. Through rigorous experimentation and performance comparison, this study analyses suitable Machine Learning methods and proposes a Hybrid SARIMA-LSTM (HySALS) to forecast global prices of agricultural commodities. The effectiveness of the proposed approach is evaluated using historical price data for five important commodities: Wheat, Millet, Sorghum, Maize, and Rice, both on a global average scale and with specific emphasis on developing nations that are either global leaders in the production of these crops or hold a significant production share within their own borders. The training data encompasses the years 2005 to 2017, while testing is conducted for the period from 2018 to 2022, followed by forecasting global prices for these commodities from 2023 to 2030. The insights derived from these forecasts are aimed to assist the decision-making processes of various stakeholders, from farmers to policymakers, thereby contributing to the efforts towards achieving global food security.*

Keywords - *Sustainable Development Goals, Global food security, Machine learning, Agricultural research, Price dynamics, Price forecasting.*

1. Introduction

Agriculture is an imperative sector of the global economy with a significant impact on employment and rural development. According to the Food and Agriculture Organization (FAO), nearly 33% of the global population relies on agriculture for their livelihoods [1]. In fact, agriculture accounts for up to 60% of employment and over 25% of the GDP in many emerging countries [2]. Beyond its economic dimensions, agriculture intertwines with health and environmental aspects, making it a cornerstone of sustainable development. Yet, amidst its paramount importance worldwide, ensuring food security remains a formidable challenge.

Global food security entails assuring that all people, regardless of their geographical location or socio-economic status, have consistent access to sufficient, safe, and nutritious food. The United Nations' Sustainable Development Goals (SDG) 2: Zero Hunger focuses on the widespread commitment to changing our agricultural systems to fulfil the world's rising food demand and achieve global food security [3].

While various factors such as population growth, climate change, agricultural methods, distribution networks, etc. impact global food security, the fluctuation in the prices of agricultural commodities plays a pivotal role. Both consumers and producers can have a negative impact due to price hikes. Consumers, particularly those in low-income households, may even face malnutrition as food costs rise. The World Bank estimates that food price hikes between 2008 and 2011 drove 100 million people into poverty [4]. Furthermore, recent figures show a jump in food price inflation, with a 40% global increase in 2020 alone [5].

Balancing the agricultural supply chain is crucial in this situation. To ensure that supply meets demand, prices are steady, and all chain participants make a profit, it necessitates controlling the production, distribution, and consumption of agricultural commodities. According to the World Food Programme, up to 40% of the entire cost of producing food in underdeveloped nations comes from inefficient supply chains [6]. Furthermore, recent research has emphasized the considerable impact of supply chain interruptions brought on by the COVID-19 pandemic, which led to market instability and price volatility, and the effect of this shall last for a longer duration [7]. The management of these difficulties depends heavily on forecasting. Accurate forecasts of world food prices can support strategic planning and assist decision-makers at all levels, from farmers to policymakers, in making informed choices. According to the International Food Policy Research Institute, 20% less price volatility can be achieved with precise forecasting [8]. The accuracy of forecasting models has recently increased due to the developments in Artificial Intelligence (AI) and Machine Learning (ML) [9]. According to the World Economic Forum, digital technology, such as AI and ML, might open a $2.3 trillion market for the world's agriculture sector by 2030 [10]. The proposed work employs Machine Learning to create a reliable and accurate model for predicting global food prices.

The contributions of this work are:

1. This research offers a fundamental analysis of the global price dynamics of five agricultural commodities, which are staple foods for many populations worldwide, thus providing a foundation for identifying trends and developing accurate price forecasting models.
2. The work analyses suitable Machine Learning methods for forecasting global agricultural prices and proposes the Hybrid SARIMA-LSTM (HySALS) approach for improved forecasts.
3. The work harnesses the potential of Machine Learning driven forecasting of agricultural commodity prices to assist policy-making on global food security for sustainable development, specifically focusing on developing countries.

The paper is organized as follows: Section 2 provides a comprehensive literature review, summarizing prior work in the field of agriculture using Machine Learning, as well as a brief discussion of the promising forecasting methods. Section 3 presents the proposed approach, which focuses on the forecasting of prices for different agricultural commodities and the analysis of price dynamics. Section 4 discusses the implementation results, demonstrating the effectiveness of the proposed approach as a step towards achieving global food security through global price analysis and forecasting. Section 5 concludes the work and suggests directions for future research.

2. Related Work

The existing work on applications of Machine Learning algorithms for agriculture and the prominent forecasting methods are presented in this section, which has been the base for this study.

2.1. Machine Learning for Agriculture

The applications of Machine Learning in agriculture are depicted in Fig. 1, and the recent research corresponding to these applications is presented in this section for a deeper understanding of the subject matter.

Fig. 1 Applications of machine learning in agriculture

Machine Learning techniques like Convolutional Neural Networks (CNN) and Support Vector Machines (SVM) have been widely applied to detect pests early and precisely. Deepa et al. [11] cover using the Alexnet model to identify tomato leaf diseases in their work. The authors stress the need to quickly and precisely assess the severity of infections to let the farmers take further intervention measures and stop the damage from worsening. Their work provides a better classification model, which reduces the number of training sessions while improving computation precision and gradient flow to recognize and classify tomato leaf disease.

Intelligent fertilizer spraying employs AI/ML technology to provide efficient and targeted application of fertilizers and pesticides, lowering chemical use and adverse environmental effects. ML systems can pinpoint the specific locations that need treatment by analyzing data like crop health, pest presence, and environmental variables. This method minimizes the use of chemicals, lessens environmental impact, and increases resource effectiveness. Researchers in their work describe using Deep Learning methods to eradicate harmful insect infestations in different plants [12].

Another application, yield prediction, utilizes AI/ML models and Data Analytics, which helps farmers plan better and make smarter choices to predict crop yields. These systems can predict crop yields precisely by looking at historical data, weather patterns, soil conditions, and other pertinent aspects. Farmers may use this information to make well-informed choices about managing their crops, allocating resources, and planning their markets. Klompenburg et al. [13] in their work present the synthesis of 50 papers using Machine Learning and 30 Deep Learning-based papers on crop yield prediction.

Automated Irrigation Systems optimize water usage based on several factors such as soil composition, weather data, and plant water consumption and ensure that plants get the right amount of water at the right time. Researchers have also conducted a two-year study investigating the performance of the Irrigation Scheduling Supervisory Control and Data Acquisition (ISSCADA) system as a tool to manage deficit irrigation scheduling for cotton [14]. The system uses sensor feedback for decision support. The authors highlight how the ISSCADA system, automated by sensor feedback, can maintain seed cotton yield while saving water. This study provides valuable insights into the practical application of AI in optimizing irrigation systems, demonstrating the potential for significant improvements in water efficiency and sustainability in agriculture.

Weather forecasting with the assistance of AI/ML algorithms enables farmers to make decisions based on the weather, as it significantly impacts agriculture, especially extreme events. Researchers analyse vast volumes of weather data and utilize Machine Learning algorithms to produce precise and localized forecasts. Farmers can plan their planting, irrigation, disease management, and harvesting operations according to these projections. The literature discusses an ensemble prediction system using a Deep Learning weather prediction model that iteratively forecasts six important meteorological variables, including temperature, precipitation, humidity, wind speed and direction, atmospheric pressure, and solar radiation with a six-hour temporal precision [15]. Convolutional Neural Networks (CNNs) on a cubed sphere grid are used to provide global predictions in this computationally effective approach. The trained model can provide a 320-member set of six-week predictions at 1.4° precision in under three minutes on a single GPU. A collection of 32 DLWP models with slightly varying learnt weights is created by randomizing the CNN training process and the main method used to construct an ensemble spread.

Price forecasting suggests the use of combined forecasting models using AI/ML to anticipate agricultural prices. Guo et al. [16] focus on forecasting Maize prices in the Sichuan Province. To determine the spatial-temporal influencing variables of price fluctuations, they use the Apriori algorithm. They integrate the Long Short-Term Memory (LSTM), Autoregressive Integrated Moving Average (ARIMA), Back Propagation (BP), and Attention Mechanism Algorithm models to create their LSTM-ARIMA-BP model. Even though their study is limited to Maize in a single province, it offers insightful information about how ML might be used to estimate agricultural prices.

To sum up, the corpus of research that has already been done emphasizes the revolutionary potential of Machine Learning in many facets of agriculture, from precision farming and pest detection to ethical issues. The use of AI/ML in agricultural price predictions stands out as one area, though. As shown by Linanza et al. [17], there is still a sizable gap in the knowledge of and use of ML in projecting global food prices. The rising volatility of food prices and its effects on global food security and economic stability need urgent attention. The work presented in this paper attempts to address this gap by creating a reliable ML model for predicting the price of agricultural commodities worldwide. By forecasting the possible price rises, the research aims to assist policymakers, farmers, and society in designing and implementing sustainable agricultural practices, eventually promoting the accomplishment of global food security and SDG 2.

2.2. Machine Learning-based Forecasting Methods

The promising methods for forecasting have been studied and presented in this section for analysis.

2.2.1. Autoregressive Integrated Moving Average(ARIMA)

ARIMA [18, 19] combines autoregressive, differencing, and moving average features in the data. The ‘AR’ component considers how the variable relates to its own past values. The ‘I’ component is responsible for changing the data to make it more stable over time. The ‘MA’ component investigates the patterns and relationships among the errors or discrepancies in the data, both in the present and in the past. ARIMA is mathematically modelled, as shown in equation (1).

$$\boldsymbol{Y_t = C + \emptyset_1 Y_{t-1} + \emptyset_2 Y_{t-2} + \ldots + \emptyset_p Y_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} + \varepsilon_t} \quad (1)$$

Where,

Y_t represents the time series

C is a constant

$\emptyset_1, \emptyset_2, \ldots, \emptyset_p$ are the autoregressive coefficients

$\theta_1, \theta_2, \ldots, \theta_p$ are the moving average coefficients

$\boldsymbol{\varepsilon_t, \varepsilon_{t-1}, \ldots, \varepsilon_{t-q}}$ are error terms

In the context of global food price forecasting, ARIMA can be used to capture and forecast trends and patterns in food prices over time.

2.2.2. Seasonal Autoregressive Integrated Moving Average (SARIMA):

SARIMA [19] is a popular time series forecasting method that expands the ARIMA model to recognise seasonal trends in data. It mixes seasonal terms with moving average, autoregressive, and differencing components. The model is immensely helpful when examining time series data with obvious seasonal patterns or trends. SARIMA is mathematically modelled, as shown in equation (2).

$$\begin{aligned}\boldsymbol{Y_t} = &+\emptyset_1(\boldsymbol{Y_{t-1}} - \boldsymbol{Y_{t-s-1}}) + \emptyset_2(\boldsymbol{Y_{t-2}} - \boldsymbol{Y_{t-s-2}}) + \\ &\ldots + \emptyset_p(\boldsymbol{Y_{t-p}} - \boldsymbol{Y_{t-s-p}}) \\ &+ \boldsymbol{\Phi_1}(\boldsymbol{Y_{t-s}} - \boldsymbol{Y_{t-s-s}}) + \boldsymbol{\Phi_2}(\boldsymbol{Y_{t-2s}} \\ &- \boldsymbol{Y_{t-s-s}}) + \cdots . + \boldsymbol{\Phi_p}(\boldsymbol{Y_{t-Ps}} + \boldsymbol{Y_{t-s-Ps}}) \\ &+ (\boldsymbol{1} - \boldsymbol{B})^{\boldsymbol{d}}\boldsymbol{\varepsilon_t} \\ &- \boldsymbol{\theta_1\varepsilon_{t-1}} - \boldsymbol{\theta_2\varepsilon_{t-2}} - \ldots . - \boldsymbol{\theta_q\varepsilon_{t-q}} \\ &- \boldsymbol{\Theta_1\varepsilon_{t-s}} - \boldsymbol{\Theta_2\varepsilon_{t-2m}} - \ldots . - \boldsymbol{\Theta_Q\varepsilon_{t-Qs}}\end{aligned} \quad (2)$$

Where,

Y_t represents the time series at time t

C is the constant term or intercept

$\emptyset_1, \emptyset_2, \ldots, \emptyset_p$ are the non-seasonal autoregressive (AR) coefficients

$\Phi_1, \Phi_2, \ldots, \Phi_p$ are the seasonal autoregressive (SAR) coefficients

s represents the seasonal period or the number of time steps in a complete seasonal cycle

Y_{t-q} refers to the lagged values of the time series

d represents the order of non-seasonal differencing.

B is the backshift operator.

ε_t represents the error term or residual at time t

$\theta_1, \theta_2, \ldots, \theta_p$ are the moving average (MA) coefficients

$\Theta_1, \Theta_2, \ldots, \Theta_p$ are the seasonal moving average (SMA) coefficients

p, d, q are the orders of the AR, differencing (I), and MA components, respectively

P and Q are the orders of the seasonal AR and seasonal MA components, respectively.

2.2.3. Support Vector Regression (SVR)

SVR [20] is a type of Support Vector Machine (SVM) that supports linear and nonlinear regression. SVM aims to identify a function with a maximum tolerance deviation from the actual target values for all the training data while remaining as flat as possible. In the context of global food price forecasting, SVR can be used to model complex, nonlinear relationships utilizing historical price data.

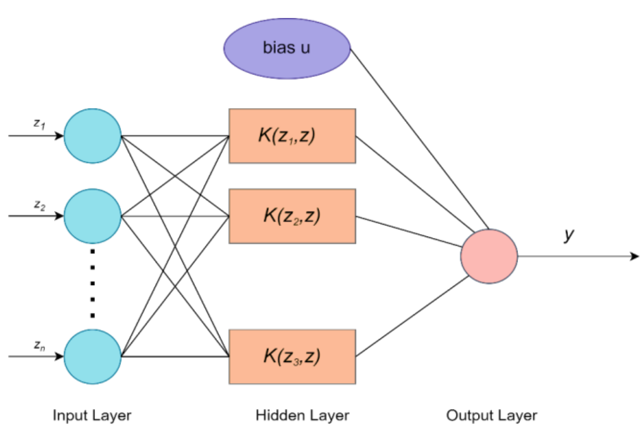


Fig. 2 General architecture of SVR

The basic principle of SVR is to map the features of sample data from low dimension to high dimension and perform regression analysis on them in high dimension by the usage of the kernel function as shown in Fig. 2. The kernel function K of SVR is mathematically modelled in equation (3).

$$K = min(w, u, z_1 \ldots z_n, z) : \frac{||w||^2}{2} + \sum_{k=1}^{n}(\xi_k + \xi^*{}_k) \quad (3)$$

Where,

w is the weight vector

u is the bias

$z_1 \ldots z_n$ and z are slack variables

C is the parameter of the penalty

$\xi_k + \xi^*{}_k$ captures the extent of the deviation or error

2.2.4. Extreme Gradient Boosting (XGBoost)

XGBoost [21] is a Machine Learning approach that builds a powerful predictive model by combining the predictions of multiple smaller models using the gradient boosting framework, often with decision trees as the base learners. In the context of forecasting global food prices, where there is access to historical price data as the feature, it effectively captures the complex, nonlinear patterns in the price dynamics. By iteratively optimizing and combining the predictions from multiple decision trees and creating a robust and accurate predictive model for global food price forecasting, leveraging the strengths of decision trees to capture intricate relationships and patterns within the price data. Considering $f_k(x)$ as the prediction of the k[th] tree, the output $\hat{y}$ is a combination of all K trees, mathematically modelled as in equation (4).

$$\hat{y} = \sum_{k=1}^{K} f_k(x) \quad (4)$$

2.2.5. Long Short-Term Memory (LSTM)

LSTM [22] is a special kind of Recurrent Neural Network (RNN) architecture designed to recall sequence data dependencies, something regular RNNs are unable to do due to its capability to capture long-term dependencies and model complicated nonlinear interactions. The general architecture of LSTM is depicted in Fig. 3 and is mathematically modelled as shown in equations (5) - (10).

$$\gamma_t = \sigma(W_f[z_{t-1}, x_t] + b^M{}_f) \quad (5)$$

$$i_t = \sigma(W_i[z_{t-1}, x_t] + b^M{}_i) \quad (6)$$

$$u_t = \tanh(W_u[z_{t-1}, x_t] + b^M{}_u) \quad (7)$$

$$C_t = \gamma_t C_{t-1} + i_t u_t \quad (8)$$

$$O_t = \sigma(W_o[z_{t-1}, x_t] + b^M{}_O) \quad (9)$$

$$z_t = O_t \tanh(C_t)| \quad (10)$$

Where,

γ_t represents forget gate

i_t represents input gate

u_t represents cell update

c_t represents the final cell state

o_t represents the output gate

z_t represents the hidden state

t represents the current state, whereas t - 1 represents the previous state

Z_t represents the hidden state

X_t denotes the current state

X_t represents the current input

b symbolizes the bias vector

W symbolizes the weight matrix

The sigmoid (σ) and tanh functions are nonlinear activation functions that introduce nonlinearity into the LSTM model. They are used to control the flow of information and regulate the output values of the gates and cell state in the LSTM architecture. The forget gate in an LSTM network decides what parts of the information in the cell state should be discarded. The input gate determines how much new information should be added to the cell state. The cell update then generates new potential values for the cell state. The final cell state is updated by combining the old cell state with these new potential values. The output gate then controls how much of this cell state should be revealed as the hidden state, which is the output of the LSTM cell. All these components work together, enabling the LSTM network to hold onto valuable information over long sequences. In the context of the proposed work, the temporal relationships in the price data may be modelled using LSTM and utilized for predicting the global price of commodities. In the context of global food price forecasting, LSTM can be used to capture temporal patterns such as trends and seasonality and forecast them over time.

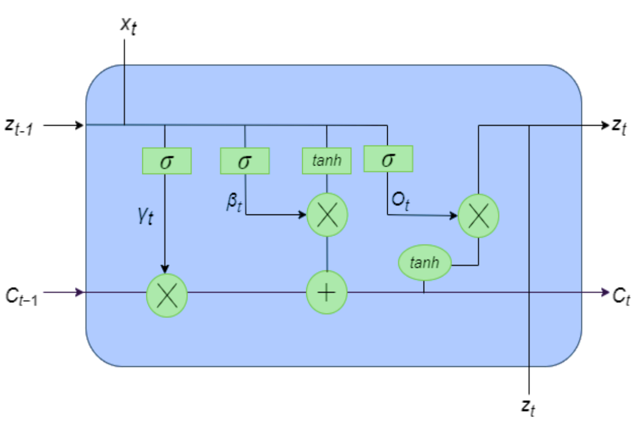


Fig. 3 General architecture of LSTM model

2.3. Price Forecasting of Agricultural Commodities

As the work focuses on forecasting global prices of agriculture commodities, the subject-specific literature review is presented in this section. Researchers [23] investigate the forecasting accuracies of individual food price models and consider their cross-dependence. The authors focus on three commodities' prices: Corn, Soybeans, and Wheat. They used an equilibrium correction model (EqCM) for each food price and used performance parameters indicated by the Mean Absolute Percentage Error (MAPE). On average, the forecasting results had 10% MAPE, therefore indicating the need for further optimization.

Wu et al. [24], in their work, forecast prices for fisheries products based on Variational Modal Decomposition (VMD) and Improved Bald Eagle Search (IBES) algorithm optimized Long Short-Term Memory Network (VMD-IBES-LSTM). They conduct empirical research utilizing data on fish prices from the Chinese Ministry of Agriculture and Rural Affairs' Department of Marketing and Informatics. The work analyses only one type of commodity, but it provides the motivation for a similar study for more crops.

Authors [25] of related work present simple approaches based on open data that make use of various parametric and non-parametric models to create a strong and approachable model that can help decision-makers optimize their harvesting operations. They employ regression models to deliver precise and trustworthy insights for thoughtful decision-making. The research does not go into detail about how the model works with various agricultural products or in other regions or climates and does not capture seasonal trends.

However, this research highlights the value of forecasting in agriculture decision-making processes, which is pertinent to the proposed work. Further, research [26] has also been employed a time series forecasting method for the future prices of agricultural products, and it is suggested to use a combination of the methods for improvement in forecasting. It is also observed that regional division and external factors significantly influence the pricing dynamics and overall profitability of agricultural commodities, which needs attention.

To sum up, the existing literature offers insightful information on agricultural product price forecasting research and identifies a few areas that call for additional study. As one of the implemented aspects, rigorous evaluation of how well the proposed forecasting models function across different geographic locations or for various agricultural products is lacking. By utilizing the Machine Learning models for global food price forecasting for multiple countries and products, our proposed work seeks to close these gaps, thereby contributing to global food security.

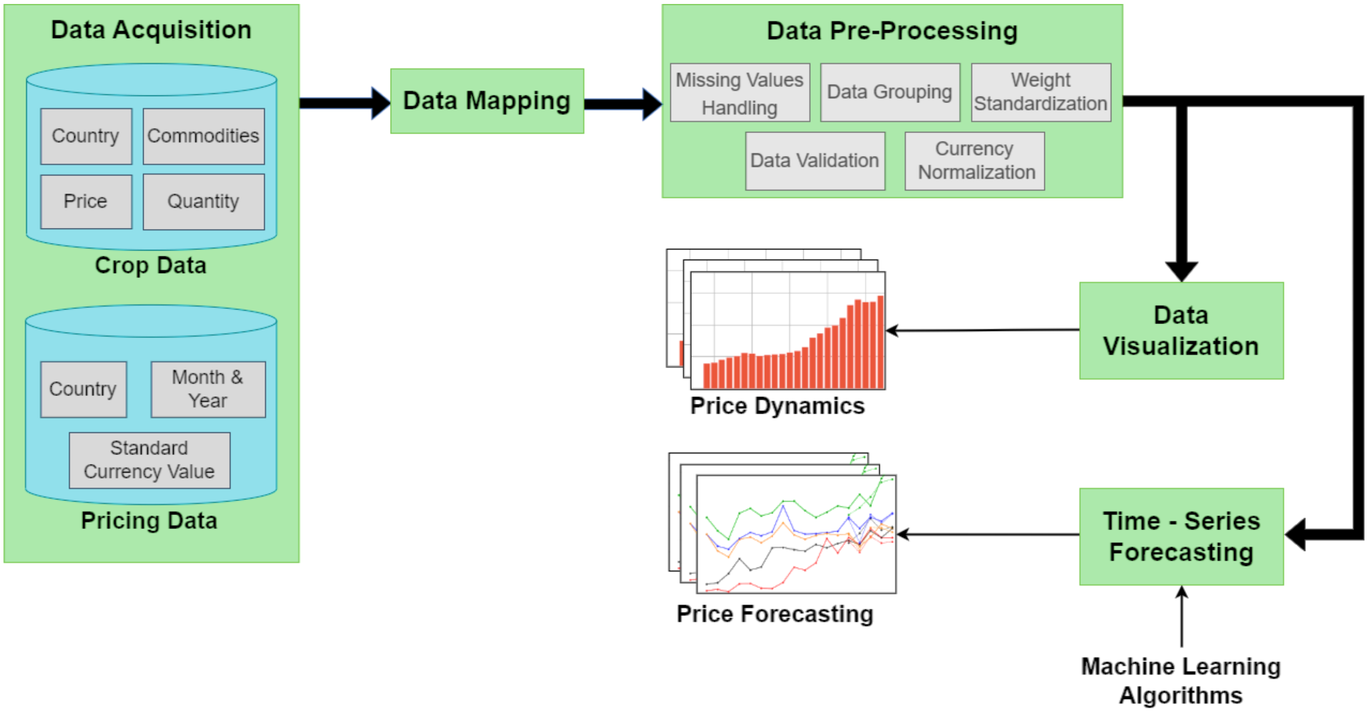


Fig. 4 Proposed approach for analysis of price dynamics and price forecasting

3. Proposed Approach

The proposed approach for Price Forecasting and Analysis of Price Dynamics is presented in Fig. 4, followed by its phase-wise description.

3.1. Data Acquisition

To address the global food security issue targeted in this work, we consider two kinds of data: One is the Crop data, which contains in-depth information on a variety of aspects affecting the cost of the crop globally, including commodity name, country, price (in country-specific currency), quantity (corresponding to different months and years). Another is the pricing data, which contains country, month, year, and standard currency value, which are used in converting currency units per U.S dollar.

3.2. Data Mapping

The two data are mapped based on common attributes such as country, year and price. This ensures comprehensive, integrated data creation, facilitating more accurate and insightful analysis.

3.3. Data Pre-Processing

The data is pre-processed using the methods discussed in the following subsection.

3.3.1. Missing Values Handling

The average price for each unique combination of country and year is calculated and used to fill in any missing values in the dataset.

3.3.2. Data Grouping

To group pertinent data points from the dataset, selecting suitable countries and commodities within a range are filtered and grouped.

3.3.3. Weight Standardization

The process of weight standardization is used to convert weights into a common unit and quantity, specifically to 1 KG. This ensures that measurements are uniform and consistent throughout the system.

3.3.4. Data Validation

Data consistency tests were performed by crosschecking the precision of crop information from the available data, thus ensuring that the values across various attributes are logically consistent and confirming the accuracy of the recorded information.

3.3.5. Currency Normalization

As the work focuses on predicting global food prices, normalizing all rates into a uniform currency is a prerequisite. The prices are normalized to USD based on exchange rates respective to the other countries in the data at respective timestamps.

3.4. Data Visualization

Data visualization is performed to examine the yearly price dynamics of different commodities. The pricing data is grouped by the year for equal weight (quantity), providing data on the peaks and troughs of crop prices.

Table 1. Algorithmic parameters

Model	Parameter	Value
ARIMA	Non-Seasonal Order	(1, 1, 1)
SARIMA	Non-Seasonal Order	(1, 1, 1)
	Seasonal Order	(1, 1, 1, 12)
SVR	Regularization	2.0
	Regression Epsilon	0.15
	Kernel Type	Radial Basis Function
XG Boost	Learning Rate (Eta)	0.1
	Maximum Depth of Tree	4
	Random Number Seed	42
LSTM	Number of LSTM Units	128
	Shape of Input Data	(1, 1)
	Number of Units in Dense Layer	1
	Number of Epochs	200
	Batch Size	8

3.5. Time-Series Forecasting

For time-series forecasting, choosing the right algorithm is essential since various algorithms may function better in certain situations or capture particular patterns more effectively. Hence, the proposed work considers several well-known Machine Learning algorithms, including ARIMA, SARIMA, SVR, XGBoost, and LSTM, and recommends a hybrid approach using the weighted average predictions made by the best two methods emerging from extensive experimentation.

4. Implementation Details

In order to address the targeted issue of global food security using Machine Learning, we use historical price data for various commodities on a global scale spanning from 2005 to 2022. Particularly, we consider the five commodities, including the cereals: Wheat, Millet, Sorghum Maize, and Rice. Furthermore, we consider the average prices of these commodities on a global scale and specifically focus on five developing countries for each commodity. The selection of these countries is based on their status as global leaders in the corresponding crop production or their being the highest producers of that crop within their national boundaries.

4.1. Datasets

The current state of research in food price prediction often faces limitations regarding dataset scope and size due to challenges in obtaining comprehensive and precise global food cost information. The details of the dataset utilized in our study are as follows:

4.1.1. Global Food Prices

The dataset is procured from the Global Food Prices Dataset (WFP) [27] and comprises over 815,000 instances with diverse parameters, including country and location, commodity, price with currency, traded quantities, and transaction timestamps.

4.1.2. World Bank Official Exchange Rates

To ensure the robustness of our work, we incorporate data from the World Bank's Official Exchange Rates dataset [28]. This dataset provides historical and current exchange rates for 266 countries from 1960 to 2022. Attributes like country name, indicator name, and separate columns for each year from 1960 through 2022 make up the data's structure.

4.2. Algorithmic Parameters

Implementing Machine algorithms in this work involved careful selection and tuning of algorithmic hyperparameters. The algorithmic parameters used are depicted in Table 1.

4.3. Evaluation Criteria

To analyze the performance of the Machine Learning models, we use Mean Absolute Percentage Error (MAPE) [29], a metric used to evaluate the error made by the forecasting models. It measures the average absolute percentage difference between the predicted and actual values, as depicted in equation (11).

$$MAPE = \frac{100}{N}\sum_{i=1}^{N}\frac{|y_i - \hat{y}_i|}{y_i} \tag{11}$$

Where,

y_i is the actual value of the dependent variable

$\hat{y}_i$ is the predicted value by the model

N is the size of the instances

4.4. Results and Discussion

The findings derived from the conducted research are presented in this section.

4.4.1. Identification of Artificial Intelligence Algorithms

As mentioned in section 3.5, the performance comparison of the Machine Learning algorithms ARIMA, SARIMA, SVR, XGBoost, and LSTM is used to identify the two most suitable classifiers for employing as the final forecasting methods as this can be seen from the results of Fig. 5, when compared on the mentioned agricultural commodities, SARIMA and LSTM have MAPE between 4.31% - 7.83% and demonstrate promising results as compared to ARIMA, SVR, and XGBoost.

The SARIMA and LSTM models can capture seasonal trends, the effect of external parameters, and previous data patterns, leading to accurate price predictions. The MAPE is an average of results derived for 05 different countries on data from 05 years.

Consequently, we propose the Hybrid SARIMA-LSTM (HySALS) as the optimal approach for forecasting global agricultural prices, aiming for better accuracy, low error rates, and consistent performance across different crop kinds. The HySALS approach implies using the weighted average of the forecast by SARIMA and LSTM to predict the prices of agricultural products, as shown in equation (12).

$$\hat{y} = w_1\, f_1(x) + w_2 f_2(x) \quad (12)$$

Where,

$\hat{y}$ is the predicted value or weighted average of prices

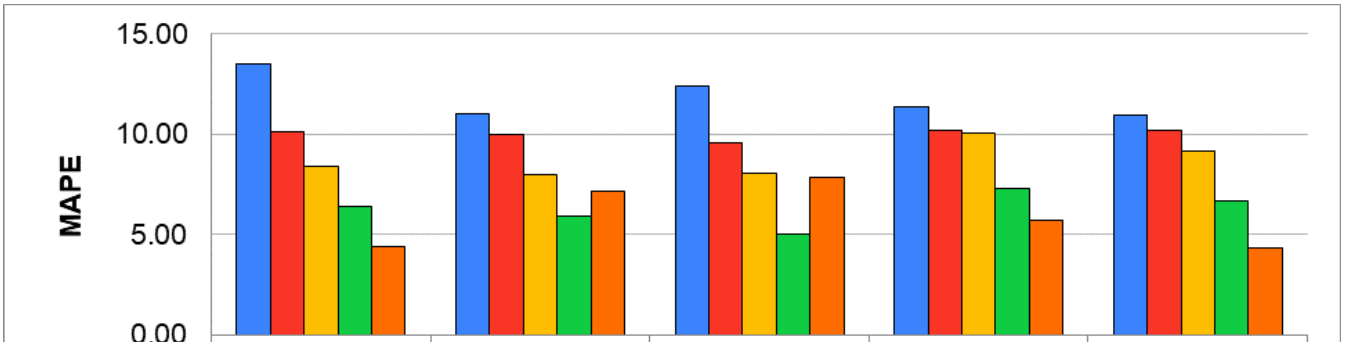
f_1 represents the predicted price using the SARIMA model

f_2 represents the predicted price using the LSTM model

w_1 represents the weight assigned to the SARIMA model's prediction, which is equal to 0.55

w_2 represents the weight assigned to the LSTM model's prediction, which is equal to 0.45

The weights are proportional to their commodity-wise performance, summing to the unit through a process of trial and error, allowing us to strike the most accurate balance between the two models when predicting the prices of agricultural products.



	Wheat	Millet	Sorghum	Maize	Rice
SVR	13.48	11.01	12.40	11.37	10.94
XGBoost	10.08	9.95	9.59	10.20	10.15
ARIMA	8.41	7.97	8.04	10.06	9.15
LSTM	6.41	5.93	5.04	7.28	6.64
SARIMA	4.38	7.14	7.83	5.73	4.31

Fig. 5 Performance comparison of machine learning algorithms on different crops

4.4.2. Price Dynamics

The results in Fig. 6 - 10 present a comprehensive analysis of the price dynamics of key agricultural commodities, including Wheat, Millet, Sorghum Maize, and Rice, across various global producers. As stated at the beginning of this section, we focus on developing countries, which either are the foremost global producers or cultivate their respective crops to the greatest extent within their own nations. Alongside this, the average price of the crops across the globe is also presented to enhance comprehension of price dynamics. The analysis reflects the price trends affected by various events during the specified periods. A highlight of these events and the corresponding dynamics serve as a suggestion to the policy makers for mitigation of food insecurity in case of similar events in the future.

Fig. 6 illustrates the price trend of Wheat in India, Ethiopia, Nepal, Afghanistan, and Tajikistan, as well as the global average from 2005 to 2022. The analysis reveals a consistent increase in Wheat prices over the specified time, except for Afghanistan, where a severe drought in 2008 led to a significant price hike [30].

Fig. 7 depicts Millet prices of global producers on average and specifically including the developing countries Niger, Mali, Senegal, Burkina Faso, and Nigeria. It is observed that there are major fluctuations in price taking place throughout the period of analysis. Particularly, there was a major price dip between 2005 and 2007 in all the nations, ranging from 0.20 USD to 0.35 USD. After analysis, it was found that one of the significant factors contributing to the price dip was favourable weather conditions and improved agricultural productivity. After 2018, there was a gradual increase in prices until the year 2011, while there was a huge increase in prices between 2012 and 2013. Particularly in Nigeria, the prices hiked up to 0.55 USD due to factors such as the insurgency attack in the northeast region, which resulted in farmers' displacement, farmlands' destruction, and disruption of agricultural activities [31].

Fig. 8 presents the Sorghum prices on average for various global producers, namely Mali, Niger, Senegal, Burkina Faso, and Gambia. The graph shows intriguing patterns in the historical price movements. Particularly, crop prices in Gambia show consistent growth, showing a constant increase in Millet's price. However, a considerable increase of 0.45 USD was seen in Mali in 2012. A military coup that resulted in political upheaval and a rise in Sorghum prices is a possible reason for this increase [32]. Another food and nutrition crisis also occurred that year in Burkina Faso, costing 0.35 USD affecting a sizable population across the Sahel Region of Western Africa. Drought, rising grain prices, a drop in remittance, environmental degradation, population relocation, persistent poverty, and vulnerability all contributed to this disaster. As a result, more than 16 million people experienced food insecurity, and more than 1 million children under the age of five were at risk for severe acute malnutrition [33].

According to Fig. 9, which illustrates the Maize prices for developing countries including Malawi, Niger, Mali, Burkina Faso and Tajikistan, there is an overall gradual upward trend in Maize prices across these regions. However, it is noteworthy that Mali experienced a sudden spike from 0.28 USD to 0.4 USD in Maize prices in 2012. This surge can be attributed to a severe security and political crisis resulting from attacks by armed groups in the northern part of the country. In response to this crisis and to prevent a decline in GDP, measures were implemented to increase agricultural production and adjust prices to support Mali during this challenging period [34]. Correspondingly, the average price of Maize has also seen a rise globally.

Fig. 10 presents the Rice prices for global producers such as India, Nepal, El Salvador, Indonesia, and the United Republic of Tanzania. The graph clearly indicates that the average prices in India and Nepal have consistently remained lower, ranging from 0.15 USD to 0.4 USD, compared to other countries from 2005 to 2022. Unlike the other nations, where prices exhibit fluctuations and occasional spikes, India and Nepal have shown minimal gradual increases in Rice prices.

This is because India and Nepal hold prominent global rice-producing positions. Their high domestic production contributes to the lower prices observed due to abundant supply. The Food and Agriculture Organization of the United Nations reports that India ranked as the world's second-largest Rice producer in 2019 [35]. While Nepal may not be among the largest global producers, it achieves substantial self-sufficiency in rice production in most years. In both countries, government policies and subsidies play a vital role in maintaining lower Rice prices [36].

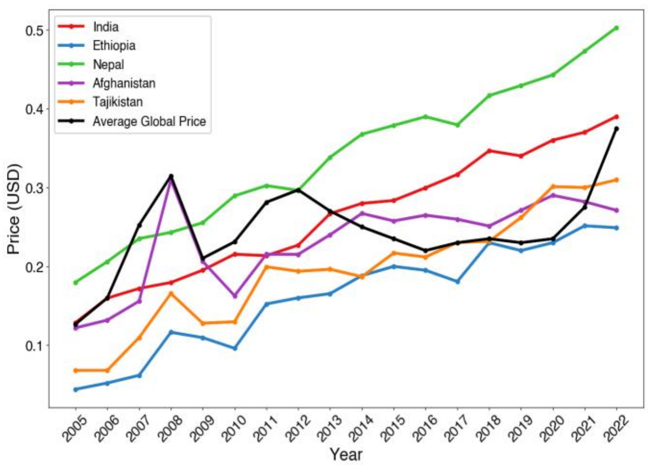


Fig. 6 Global price dynamics for agricultural commodity – Wheat

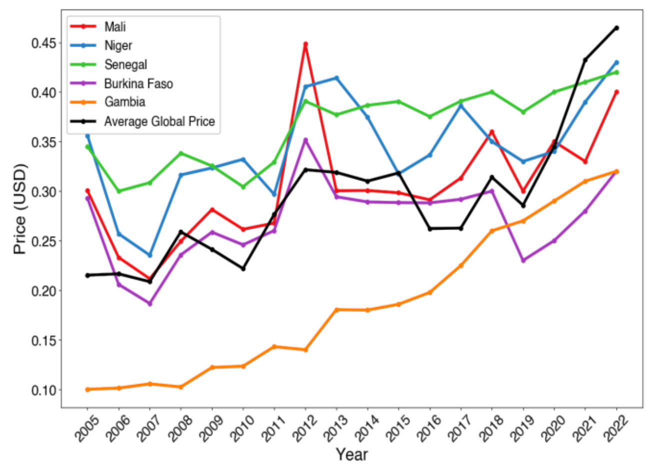


Fig. 8 Global price dynamics for agricultural commodity – Sorghum

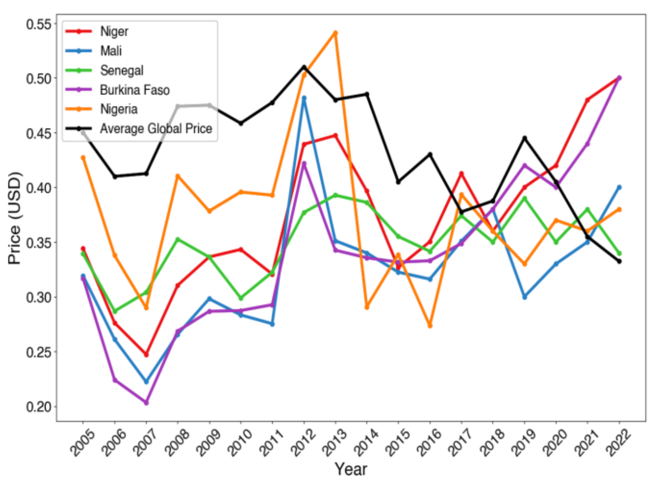


Fig. 7 Global price dynamics for agricultural commodity – Millet

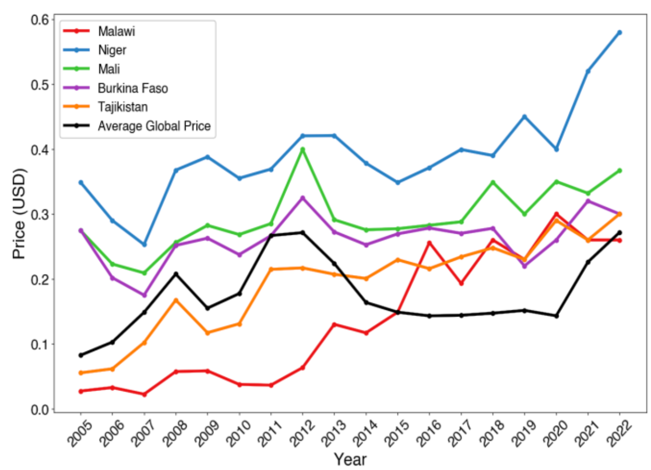


Fig. 9 Global price dynamics for agricultural commodity – Maize

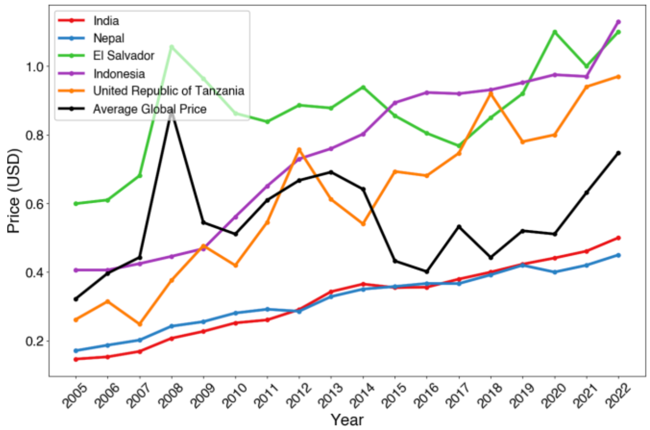
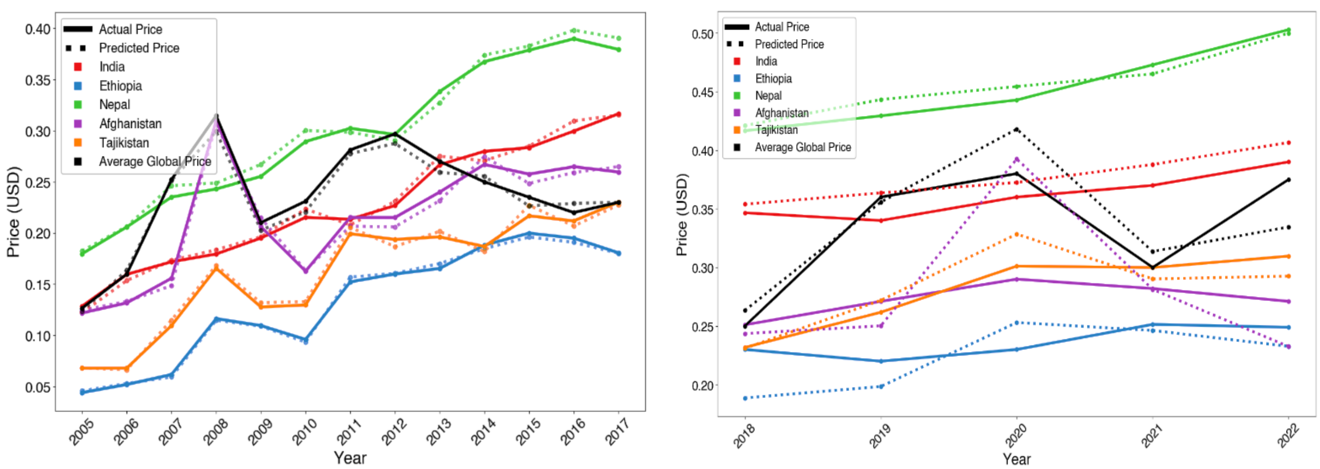


Fig. 10 Global price dynamics for agricultural commodity – Rice

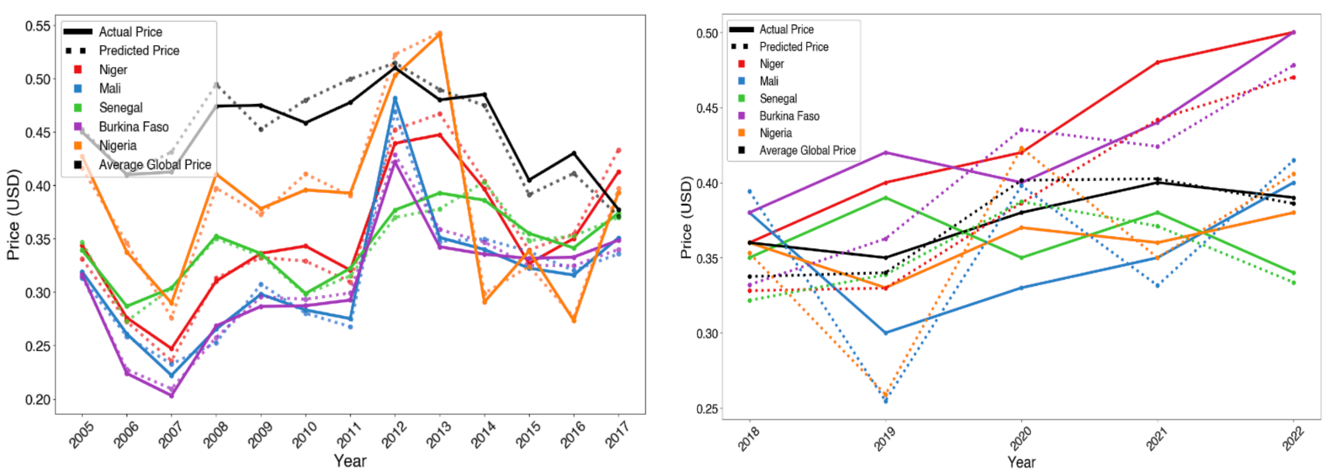
To sum up the collective analysis, the analysis of global price dynamics for Wheat, Millet, Sorghum Maize, and Rice shows variations, sudden increases, and steady rises over different periods and countries. These trends are shaped by factors such as natural disasters, political unrest, armed conflicts, climate change, economic hurdles, increasing population, etc., affecting global food security. Grasping these price trends and their root causes empowers policymakers, farmers, researchers, and all stakeholders to make knowledgeable decisions and devise strategies to tackle the issues in the agricultural sector.

By acknowledging the effects of influencing factors, steps can be taken to lessen the adverse impacts, such as enhancing disaster readiness, advocating for sustainable farming methods, and enforcing effective policies to ensure global food security and stability in the global food market. Furthermore, this analysis aids in pinpointing commodities and producer countries that are especially susceptible to price swings, allowing for focused interventions and support to ensure the supply and affordability of vital food items for communities around the globe.



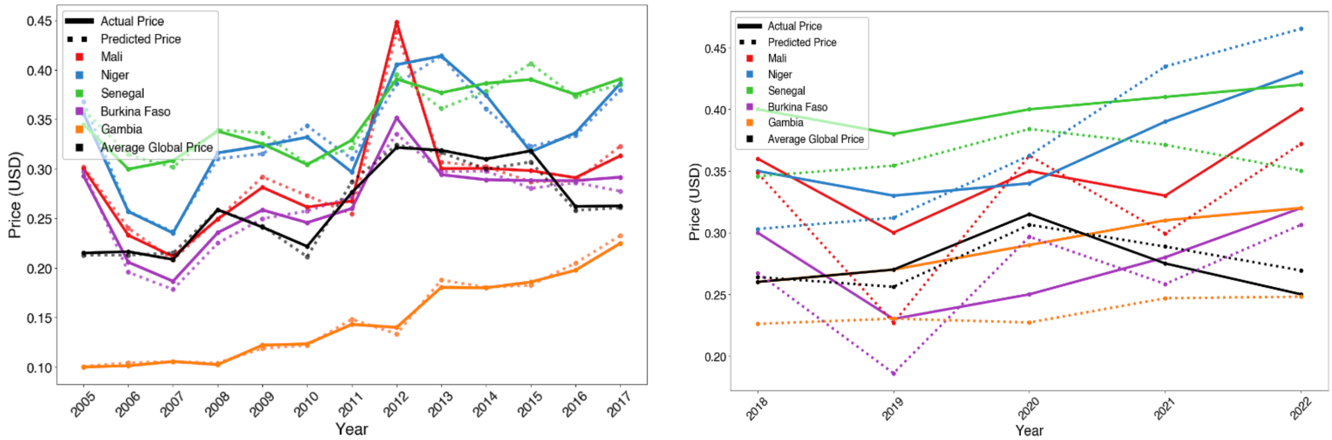
(a) Price forecasting on training data **(b) Price forecasting on testing data**

Fig. 11 Country-wise performance comparison of actual price versus predicted price using HySALS for wheat



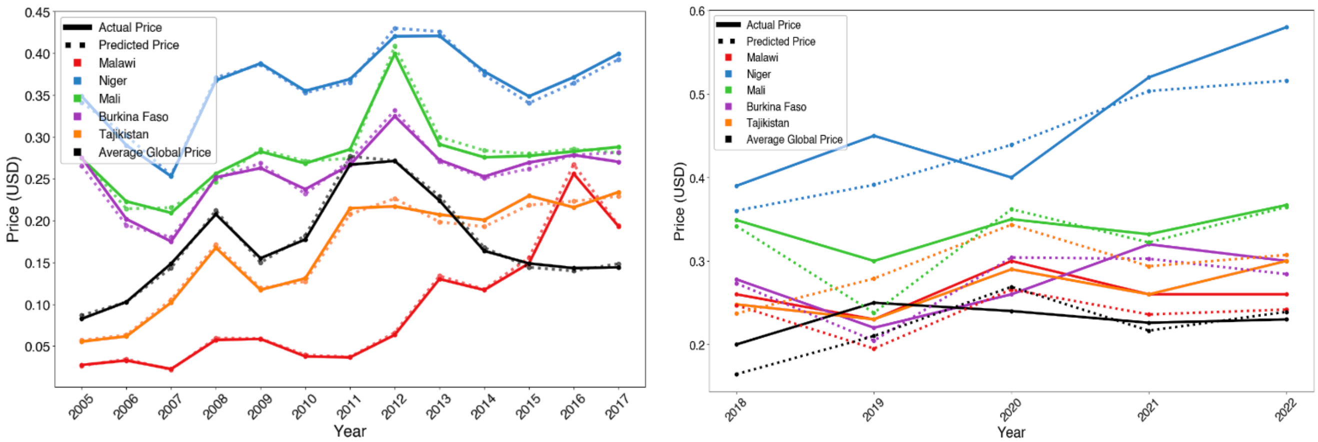
(a) Price forecasting on training data **(b) Price forecasting on testing data**

Fig. 12 Country-wise performance comparison of actual price versus predicted price using HySALS for millet



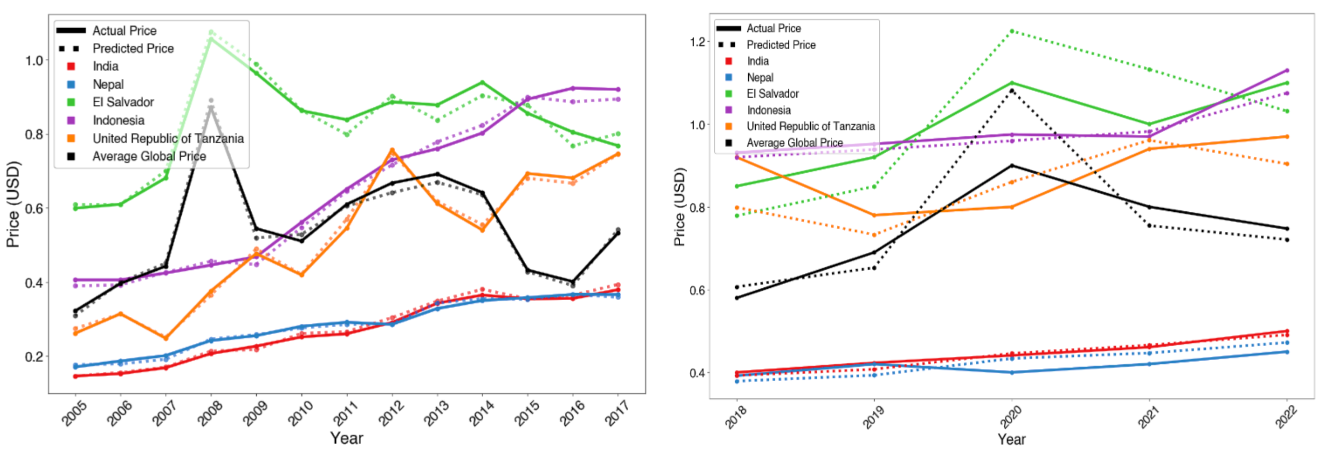
(a) Price forecasting on training data **(b) Price forecasting on testing data**

Fig. 13 Country-wise performance comparison of actual price versus predicted price using HySALS for sorghum



(a) Price forecasting on training data **(b) Price forecasting on testing data**

Fig. 14 Country-wise performance comparison of actual price versus predicted price using HySALS for maize

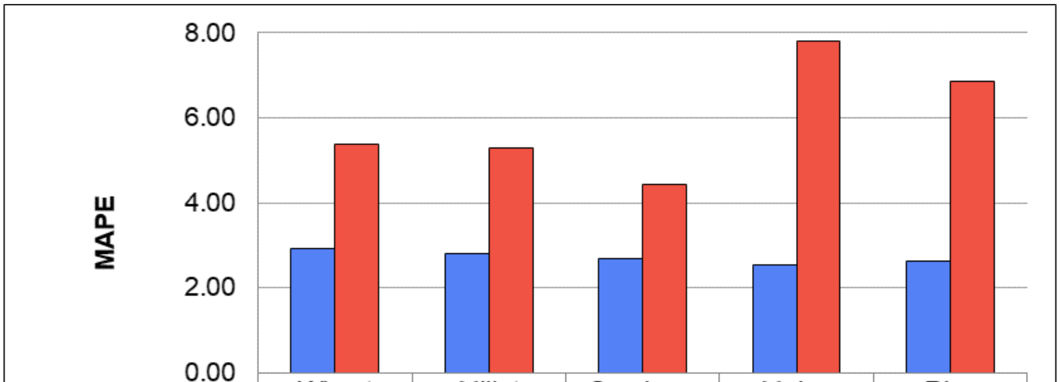


(a) Price forecasting on training data **(b) Price forecasting on testing data**

Fig. 15 Country-wise performance comparison of actual price versus predicted price using HySALS for rice

4.4.3. Performance Evaluation on Training and Test Data

This section presents the results of performance evaluation and prices forecasted using the proposed HySALS approach. The training is conducted using the data for the years 2005 to 2017, the results of HySALS are validated on the data for the years 2018 to 2022, and further, it is used for forecasting global prices of the chosen agriculture commodities from the year 2023 up to 2030.



	Wheat	Millet	Sorghum	Maize	Rice
Training MAPE	2.92	2.80	2.67	2.53	2.63
Testing MAPE	5.37	5.28	4.43	7.80	6.87

Fig. 16 Performance analysis (based on MAPE) of HySALS on training and testing data of average global prices

The performance comparison results of the prices forecasted using the HySALS approach in contrast to the actual prices for different countries during the training and testing phases are presented in Fig. 11 – 15 for Wheat, Millet, Sorghum, Maize, and Rice, respectively. The countries are chosen for crop-focused study based on their global leadership in production or being top producers within their own borders.

The Mean Absolute Percentage Error (MAPE) in predicting the average global prices on Training Data and Testing Data are presented in Fig. 16. As can be observed, our approach, HySALS, which incorporates seasonal trends in addition to examining previous data patterns, long as well as short term trend capturing shows a high degree of accuracy in its predictions.

The training MAPE in learning the average global prices has been less than 3% for all the crops, whereas the testing MAPE has been in the range of 4.43% – 7.80%, which is promising enough. As the actual and predicted values are close, the findings demonstrate the effectiveness of HySALS in making accurate forecasts for the agricultural crops under consideration and as a step towards SDG 02.

4.4.4. Price Forecasting Using Proposed HySALS Approach

The HySALS approach is further employed to predict the prices of the chosen agricultural commodities for the years 2023 to 2030. Fig. 17 – 21 demonstrate the forecasted average global prices and the forecasted prices for the developing countries that either are leading global producers of these crops or possess a substantial production share in their countries.

The findings, as observed in Fig. 17 – 21, demonstrate a general upward trend in prices for all the crops considered, on an average globally, as well as for the stated developing nations. However, the forecasts also indicate sporadic price spikes, highlighting the potential challenges to food security, such as the projected increases in Millet, Sorghum, and Maize prices for Mali in 2029. On the other hand, Rice prices in India and Nepal are expected to experience gradual growth over time without significant fluctuations.

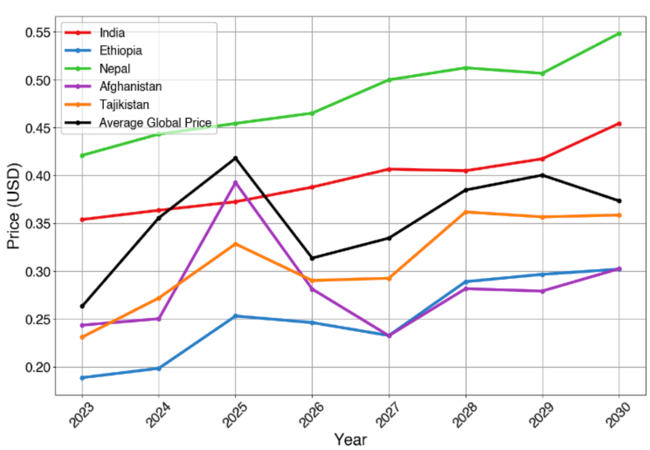


Fig. 17 Global average and country-wise price forecasting for wheat

Fig. 18 Global average and country-wise price forecasting for millet

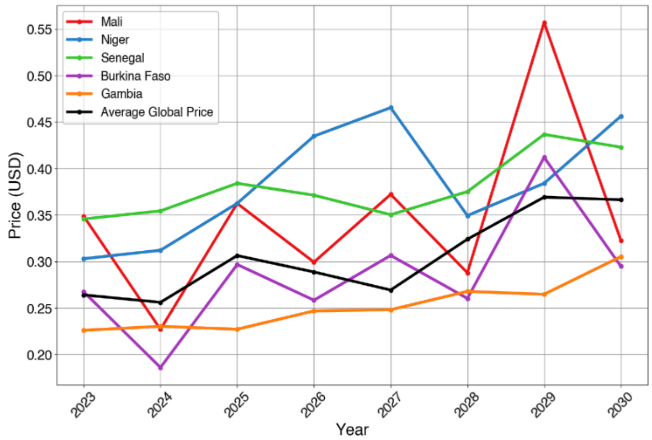


Fig. 19 Global average and country-wise price forecasting for sorghum

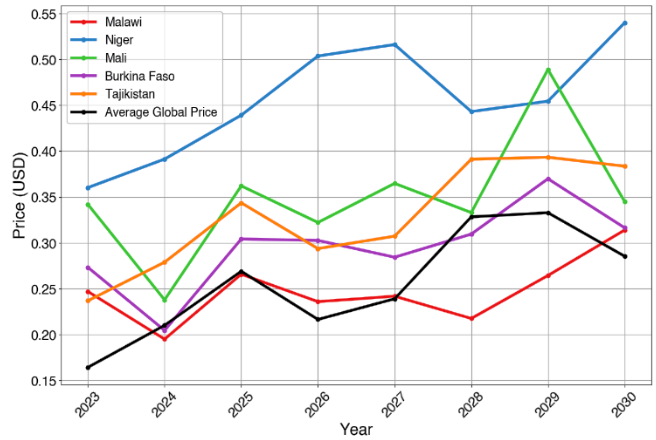


Fig. 20 Global average and country-wise price forecasting for maize

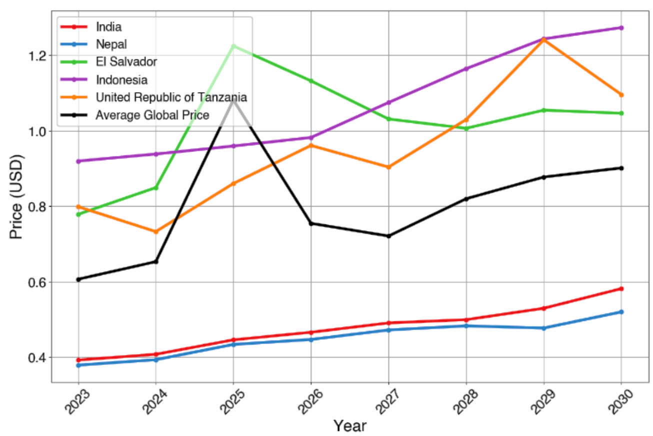


Fig. 21 Global average and country-wise price forecasting for rice

5. Conclusion

The research leverages the potential of Machine Learning (ML) to address the critical issue of global food security, which is directly impacted by the prices of agricultural commodities. The work concludes with two major outcomes: Firstly, the price dynamics of key agricultural commodities are analyzed, particularly for Wheat, Millet, Sorghum, Maize, and Rice. The analysis provides highlights the trends and fluctuations in the prices on a global scale. It is specially focused on the developing nations that are the leading producers of these crops, or they achieve the highest production of this crop among the others in their country. Secondly, a Hybrid SARIMA-LSTM (HySALS) is proposed to capture seasonal trends and dynamic patterns in these crop prices' historical data to forecast future prices accurately. The work derives motivation from SDG 02: Zero Hunger, and as a contribution towards achieving the same, the HySALS approach is trained (with <3% MAPE), tested (with <8% MAPE), and further employed to forecast prices up to the year 2030. The price dynamics analysis and the forecasted prices provide valuable insights to policymakers, farmers, researchers, and all the stakeholders to mitigate the effects of the influencing factors, foster sustainable agriculture, and make informed decisions to ensure global food security.

Future research can focus on enhancing price forecasting by considering the population dynamics, supply-demand ratio for each country, warehouse availability, and the effect of climate change as inputs to the Machine Learning models, as these factors have a crucial impact on the price dynamics and, ultimately, the food security.

Acknowledgments

The authors extend their gratitude to Dr. Suresh Ukarande, Principal - K. J. Somaiya College of Engineering, for the motivation towards this research and for engaging in highly insightful scientific discussions that significantly enriched the depth of the study.

References

[1] Food and Agriculture Organization, The State of Food Security and Nutrition in the World 2021. [Online]. Available: https://www.fao.org/state-of-food-security-nutrition/en

[2] World Bank, World Development Indicators. [Online]. Available: https://databank.worldbank.org/source/world-development-indicators

[3] United Nations, Sustainable Development Goals, Goal 2: Zero Hunger. [Online]. Available: https://www.un.org/sustainabledevelopment/hunger/

[4] World Bank, High and Volatile Food Prices Continue to Threaten the World's Poor, 2011. [Online]. Available: https://www.worldbank.org/en/news/press-release/2011/04/14/high-volatile-food-prices-continue-threaten-worlds-poor

[5] *The State of Food Security and Nutrition in the World 2021, Transforming Food Systems for Food Security, Improved Nutrition and Affordable Healthy Diets for All*, Food and Agriculture Organization of the United Nations, pp. 1-240, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[6] FAO, *In Brief to The State of Food Security and Nutrition in the World 2021, Transforming Food Systems for Food Security, Improved Nutrition and Affordable Healthy Diets for all*, pp. 1-40, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[7] Stella Nordhagen et al., "COVID-19 and Small Enterprises in the Food Supply Chain: Early Impacts and Implications for Longer-Term Food System Resilience in Low- and Middle-Income Countries," *World Development*, vol. 141, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[8] Derek Headey, and Shenggen Fan, *Reflections on the Global Food Crisis: How did it Happen? How has it Hurt? And how can we Prevent the Next One?*, International Food Policy Research Institute (IFPRI), pp. 1-122, 2010. [CrossRef] [Google Scholar] [Publisher Link]

[9] Mohd Javaid., "Understanding the Potential Applications of Artificial Intelligence in Agriculture Sector," *Advanced Agrochem*, vol. 2, no. 1, pp. 15-30, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[10] World Economic Forum, Digital Transformation of Industries, 2016. [Online]. Available: https://www.weforum.org/reports/digital-transformation-of-industries

[11] D. Deepa, R. Yaswanth, and K. Vasantha Kumar, "Tomato Leaf Diseases Classification using Alexnet," *Applied and Computational Engineering, EWA*, vol. 2, pp. 635-642, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[12] Waleed Albattah et al., "A Novel Deep Learning Method for Detection and Classification of Plant Diseases," *Complex & Intelligent Systems*, vol. 8, pp. 507-524, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[13] Thomas van Klompenburg, Ayalew Kassahun, and Cagatay Catal, "Crop Yield Prediction using Machine Learning: A Systematic Literature Review," *Computers and Electronics in Agriculture*, vol. 177, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[14] Susan A. O'Shaughnessy, Paul D. Colaizzi, and Craig W. Bednarz, "Sensor Feedback System Enables Automated Deficit Irrigation Scheduling for Cotton," *Frontiers in Plant Science*, vol. 14, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[15] Jonathan A. Weyn et al., "Sub-Seasonal Forecasting with a Large Ensemble of Deep-Learning Weather Prediction Models," *Journal of Advances in Modeling Earth Systems*, vol. 13, no. 7, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[16] Yan Guo et al., "Agricultural Price Prediction Based on Combined Forecasting Model under Spatial-Temporal Influencing Factors," *Sustainability*, vol. 14, no. 17, pp. 1-18, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[17] Maria Teresa Linaza et al., "Data-Driven Artificial Intelligence Applications for Sustainable Precision Agriculture," *Agronomy*, vol. 11, no. 6, pp. 1-14, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[18] Sima Siami-Namini, Neda Tavakoli, and Akbar Siami Namin, "A Comparison of ARIMA and LSTM in Forecasting Time Series," *17th IEEE International Conference on Machine Learning and Applications, IEEE*, pp. 1394-1401, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[19] Uppala Meena Sirisha, Manjula C. Belavagi, and Girija Attigeri, "Profit Prediction using ARIMA, SARIMA and LSTM Models in Time Series Forecasting: A Comparison," *IEEE Access*, vol. 10, pp. 124715-124727, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[20] Yu Zhang et al., "The Prediction of Spark-Ignition Engine Performance and Emissions Based on the SVR Algorithm," *Processes*, vol. 10, no. 2, pp. 1-15, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[21] Iliana Paliari, Aikaterini Karanikola, and Sotiris Kotsiantis, "A Comparison of the Optimized LSTM, XGBOOST and ARIMA in Time Series Forecasting," *12th International Conference on Information, Intelligence, Systems & Applications, IEEE*, pp. 1-7, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[22] Yupeng Wang, Shibing Zhu, and Changqing Li, "Research on Multistep Time Series Prediction Based on LSTM," *3rd International Conference on Electronic Information Technology and Computer Engineering, IEEE*, pp. 1155-1159, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[23] H. Ahumada, and M. Cornejo, "Forecasting Food Prices: The Case of Corn, Soybeans and Wheat," *International Journal of Forecasting*, vol. 32, no. 3, pp. 838-848, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[24] Junhao Wu et al., "An Aquatic Product Price Forecast Model Using VMD-IBES-LSTM Hybrid Approach," *Agriculture*, vol. 12, no. 2, pp. 1-26, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[25] Ivan Herranz-Matey, and Luis Ruiz-Garcia, "Agricultural Combine Remaining Value Forecasting Methodology and Model (and Derived Tool)," *Agriculture*, vol. 13, no. 4, pp. 1-15, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[26] Koichi Kurumatani, "Time Series Forecasting of Agricultural Product Prices Based on Recurrent Neural Networks and its Evaluation Method," *SN Applied Sciences*, vol. 2, no. 8, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[27] Global Food Prices Database (WFP), Humanitarian Data Exchange. [Online]. Available: https://data.humdata.org/dataset/wfp-food-prices

[28] World Bank, World Bank Open Data, 2020. [Google Scholar] [Publisher Link]

[29] Gareth James et al., *An Introduction to Statistical Learning with Applications in R*, Springer, pp. 1-426, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[30] Food and Agriculture Organization of the United Nations, Regional Overview of Food Insecurity, 2007. [Online]. Available: http://www.fao.org/3/a-j2442e.pdf
[31] S.A. Adebisi, O.O. Azeez, and R. Oyedeji, "Appraising the Effect of Boko Haram Insurgency on the Agricultural Sector of Nigerian Business Environment," *Journal of Law and Governance*, vol. 11, no. 1, pp. 15-26, 2016. [CrossRef] [Google Scholar] [Publisher Link]
[32] Human Rights Watch, Rights Trends in World Report 2013: Mali, 2012. [Online]. Available: https://www.hrw.org/world-report/2013/country-chapters/mali
[33] European Commission, 2012 Sahel Food & Nutrition Crisis: ECHO's Response at a Glance. [Online]. Available: https://ec.europa.eu/echo/files/aid/countries/ECHO_2012_Response_Sahel_Crisis_en.pdf
[34] World Bank, The Malian Economy Holds Steady in the Face of Crisis, 2013. [Online]. Available: https://www.worldbank.org/en/news/feature/2013/03/14/the-malian-economy-holds-steady-in-the-face-of-crisis
[35] Food and Agriculture Organization, FAOSTAT Data, 2023. [Online]. Available: https://www.fao.org/faostat/en/#data/QCL
[36] Department of Food & Public Distribution, Government of India, 2023. [Online]. Available: https://dfpd.gov.in/