The Journal of Agricultural Sciences - Sri Lanka
Vol. 20 No 3, September 2025. pp393–409
http://doi.org/10.4038/jas.v20i3.10534

Original Research

Predicting Agriculture Commodity Price Trends Using Ensemble Regression Approach

R Ragunath[1] and R Rathipriya[2*]

Received: 08th September 2024 / Accepted: 07th July 2025

ABSTRACT

Purpose: *This study aims to improve the forecasting of agricultural commodity price trends by developing an ensemble approach that integrates multiple regression models. Addressing this research problem, it aims to provide more accurate predictions for farmers and dealers, leveraging Wholesale Pricing Index data and Indian rainfall patterns.*

Research Method: *The research uses various regression models and competitive ensemble techniques utilizing Wholesale Pricing Index data and Indian rainfall information to forecast Agricultural Commodity Price trends.*

Findings and Values: *Empirical evidence supports the efficacy of the proposed ensemble approach over traditional regression models in accurately capturing agricultural price fluctuations. The competitive ensemble method proves beneficial for the food and financial industries in making informed decisions. This research offers practical utility by presenting a more accurate ensemble approach for forecasting Agricultural Commodity Price trends, supporting strategic planning in the food and economic industries.*

Keywords: *Agricultural commodity price, Competitive ensemble approach, Price forecasting, Regression models.*

INTRODUCTION

Nowadays, many small-scale farmers experience substantial losses due to limited access to or understanding of market information, often resulting in the sale of their produce at less price. Agricultural product prices are highly volatile and influenced by various factors, such as weather condition, pests, production risks, global demand and supply, governmental policies, and economic factors. (Yao *et al.* 2010, Yang & Liu 2012, Chen *et al.* 2013). Accurate and reliable price forecasting techniques are increasingly vital for mitigating pricing risks and enabling farmers and stakeholders to make well-informed decisions regarding the sale of agricultural commodities. Traditional agricultural commodity price (ACP) forecasting approaches have relied on a range of linear and nonlinear statistical models; however, recent developments suggest that machine learning (ML) offers more robust predictive capabilities (Flach 2012).

In recent years, ML techniques have become central to the data science landscape, with several empirical studies confirming their superiority over classical time series models, particularly in forecasting financial and commodity markets (Kumar & Rath 2020).

[1]Department of Computer Science, Periyar University, Salem-636011.
[2]Department of Computer Science, Periyar University, Salem-636011.
*rathipriyar@gmail.com
https://orcid.org/0000-0002-3970-262X

Among the most commonly utilized ML algorithms are artificial neural networks (ANN), generalized regression neural networks (GRNN), support vector regression (SVR), random forests (RF), and gradient boosting machines (GBM). These techniques are inherently data-driven and nonparametric, allowing them to effectively capture complex, stochastic relationships within datasets (Milunovich 2020). Traditional statistical methods, such as linear regression and Box-Jenkins models, often underperform compared to neural networks in prediction tasks (Werbos 1988). Studies by (Zhang & Qi 2005, Zhang & Kline 2007) and (Weng *et al.* 2019) also affirm the superiority of ML and deep learning methods over conventional models for forecasting agricultural prices.

Furthermore, various regression models—including multiple linear regression (MLR) (Sheehy *et al.* 2006), random forest (RF), Lasso, K-nearest neighbors (KNN), Gaussian process regression (GPR), gradient boosting decision trees (GBDT), and SVR—have demonstrated improved accuracy in agricultural product price prediction. However, the performance of these individual models is often constrained by their inherent assumptions and limitations (Zhang & Na 2018, Paul *et al.* 2022). To address the limitations of relying on a single predictive model, ensemble learning has gained prominence in recent years, particularly in agricultural forecasting. Advances in machine learning and computational power have enabled the adoption of ensemble methods across various research domains (Ustuner & Balik Sanli 2019, Chatzimparmpas *et al.* 2021, Razavi-Termeh *et al.* 2021, Li *et al.* 2022).

Ensemble learning involves integrating multiple base models to form a more powerful and generalized predictor. The fundamental principle is that even if one base learner performs suboptimally, the collective model can compensate for errors and enhance predictive performance. Common ensemble strategies include bagging, boosting, and stacking. Building on this, the current research introduces a novel approach to ACP prediction using competitive ensemble learning models.

This study will look into the following research questions: Can a trustworthy forecasting model be developed to forecast future ACP patterns based on historical trends in agricultural prices and monthly rainfall? These future ACP trends can be used to help farmers make wise judgments about selling their crops.

Several factors need to be considered in order to create an effective predictive model to answer this issue.

- First, the real-time datasets contain a significant amount of missing values.
- Second, there is a long-term temporal relationship between forecasted crop prices and historical pricing trends.
- Third, crop prices and monthly rainfall have a relationship.

The development of a forecasting model that can overcome these difficulties is crucial. In light of these factors, a novel competitive ensemble regression model is proposed, which uses historical ACP data and monthly rainfalls to forecast future ACP trends.

MATERIALS AND METHODS

Preliminaries

This section discussed the machine learning based prediction models for agriculture commodities and rainfall data in the literature. The various existing machine learning and ANN techniques used for agricultural commodity prediction in different models are regression, ensemble, classification, deep learning models, etc., and different statistical measures such as MAE, RMSE, MAPE, F1, R-squared, SMAPE, MAD, etc., used in the prediction models. It explores significant research works that have already been done in detail under several topics. Explored the prediction of crop yield using remote sensing data. They employed various models such as Bagging, Boosting, Stacking, RF, ERT, XGBoost, LightGBM, and CatBoost, all falling under the regression model category. The authors discussed how the ensemble of deep learning algorithms improved prediction accuracy (Zhang *et al.* 2022). Focused on predicting the prices of corn, cattle, and hogs. They compared the performance of ANN (Artificial Neural Network) and Multivariate Relevance Vector Machine (MVRVM) models in regression. The evaluation measures used were RMSE (Root Mean Square Error) and RMSPE (Root Mean Square Percentage Error) (Ticlavilca *et al.* 2010). Studied the price forecasting of onions and potatoes. They explored models such as ARIMA (AutoRegressive Integrated Moving Average), SARIMA (Seasonal ARIMA), and LSTM (Long Short-Term Memory) in the classification model category. The evaluation measure used was RMSE, and the LSTM model showed promising results for improving farmer income (Madaan *et al.* 2019).

Focused on predicting the price of corn. They compared multiple regression models including LR (Linear Regression), RF (Random Forest), XGBoost, LightGBM, LSTM, AttLSTM, and AttLSTM-ARIMA-BP. Evaluation measures such as MAPE (Mean Absolute Percentage Error), RMSE, MAE (Mean Absolute Error), and R^2 (Coefficient of Determination) were used. The AttLSTM-ARIMA-BP model was found to be the best in accurately predicting agricultural prices (Guo *et al.* 2022). Examined the prediction of cabbage and radish prices. They compared benchmark models including LSTM, GCN-LSTM, STL-AttLSTM, DA-RNN, and DIA-LSTM. Evaluation measures used were RMSE and MAPE. The DIA-LSTM model demonstrated the highest accuracy in predicting agricultural commodity prices (Gu *et al.* 2022). Focused on the relationship between weather data and crop productivity using the IMD (Indian Meteorological Department) dataset. They explored models such as DCRN (Deep Convolutional Recurrent Network), DT (Decision Tree), SOM (Self-Organizing Map), SVR (Support Vector Regression), and ESN (Echo State Network) in the classification model category. Evaluation measures used were RMSE, MAE, Accuracy, F-measure, Precision, and Recall. The DCRN model achieved an accuracy of 97.89% in accurately predicting rainfall and crop productivity, particularly in high rainfall regions (Talasila *et al.* 2020). Focused on predicting the prices of cabbage and carrot using big data and machine learning techniques. They compared models such as ARIMA (AutoRegressive Integrated Moving Average) and NAIVE. Evaluation measures used included MSE (Mean Squared Error), RMSE, MAE (Mean Absolute Error), and MAPE (Mean Absolute Percentage Error).

The ARIMA model was utilized to forecast vegetable prices, highlighting the application of big data and ML techniques in agricultural price prediction (Mao *et al.* 2022). Conducted research on forecasting the prices of soybean and sunflower seeds. They compared models such as ARIMA, MA (Moving Average), ARMA (AutoRegressive Moving Average), ANN (Artificial Neural Network), and SVM (Support Vector Machine) in the regression model category. Evaluation measures used were MAPE and RMSPE (Root Mean Square Percentage Error). The ANN model outperformed the ARIMA model in accurately forecasting the price of agricultural commodities (Mahto *et al.* 2021). Focused on predicting the price of hogs. They explored models such as LightGBM, XGBoost, GBDT (Gradient Boosting Decision Tree), WOA (Whale Optimization Algorithm), CEEMDAN (Complete Ensemble Empirical Mode Decomposition with Adaptive Noise), SVR, and EEMD (Ensemble Empirical Mode Decomposition). These models were evaluated using measures such as MSE, MAE, RMSPE, MAPE, R^2, and DA (Directional Accuracy). The combined model WOA-LightGBM-CEEMDAN achieved the highest prediction accuracy compared to other models (Wang *et al.* 2022).

These studies highlight the diverse range of models, measures, and algorithms used for agricultural commodity price prediction in various contexts. From ensemble algorithms to deep learning techniques, researchers are continuously exploring and refining methods to improve the accuracy of predicting agricultural prices, thereby benefiting farmers and stakeholders in the agricultural industry. Conducted research on forecasting the price of brinjal. They compared models such as ARIMA, SVR (Support Vector Regression), RF (Random Forest), GRNN (Generalized Regression Neural Network), and GBM (Gradient Boosting Machine) in the neural network model category. Evaluation measures used included ME (Mean Error), RMSE, MAE, and MAPE. The GRNN model demonstrated better performance compared to other techniques for forecasting agricultural commodity prices (Paul *et al.* 2022). Focused on predicting the wholesale prices of crops and their relationship with rainfall. They utilized RF (Random Forest) and DT (Decision Tree) models in the regression model category. The evaluation measure used was RMSE, and the DT model showed good prediction accuracy (Soni & Raut 2022). Investigated the forecast performance of various models for predicting piglet, pork, wheat bran, corn, and other agricultural commodity prices. They compared models such as RF (Random Forest), SVM (Support Vector Machine), ANN (Artificial Neural Network), ELM (Extreme Learning Machine), and MRMR (Minimum Redundancy Maximum Relevance). Evaluation measures used were MAPE, ACC (Accuracy), and IR (Information Rate).

The RF model demonstrated better forecast performance compared to the other models for agricultural commodity prices (Zhang *et al.* 2020). Focused on stock market data prediction for diversified financials, petroleum, non-metallic minerals, and basic metals. They explored deep learning models such as DT, Bagging, RF, Adaboost, GB (Gradient Boosting), XGBoost, ANN, RNN (Recurrent Neural Network), and LSTM (Long Short-Term Memory). Evaluation measures used were MAPE, MAE, RMSE, and MSE. The LSTM model provided more accurate results for stock market data with the highest model fitting ability (Nabipour *et al.* 2020).

Conducted research on predicting the prices of chicken, chili, and tomato using data from the FAMA (Federal Agricultural Marketing Authority) official government website. They compared models such as ARIMA, SVR, Prophet, XGBoost, and LSTM in the regression model category. The evaluation measure used was MSE, and the LSTM model with an MSE of 0.304 was selected as the best model for predicting agricultural commodity price data (Chen *et al.* 2021). Conducted a comparative study on forecasting wholesale prices of tomato, onion, and potato. They compared models such as ARIMA, RNN (Recurrent Neural Network), LSTM, VAR (Vector Autoregression), RFR (Random Forest Regression), and XGBoost in the regression model category.

The evaluation measure used was RMSE, and the LSTM model showed promising results in the comparative analysis (Suresha *et al.* 2021). Focused on forecasting permeability in carbonate reservoirs using well log data. They applied an ensemble approach using ANN, SVM, KNN (K-Nearest Neighbors), and NCC (Nearest Competitor Classifier). Evaluation measures used were RMSE and CC (Correlation Coefficient). The competitive ensemble modeling method demonstrated effective forecasting of permeability in heterogeneous oil and gas reservoirs (Adeniran *et al.* 2019). Successfully implemented the forecasting of agricultural product prices and provided information on the nearest market details based on the farmers' location using big data concepts. They utilized the PNN (Probabilistic Neural Network) model in the context of big data for accurate price forecasting. The PNN model achieved an impressive accuracy of 98.3% in predicting the prices of agricultural products, showcasing the successful application of big data techniques in agricultural price forecasting (Utomo *et al.* 2022).

Conducted research on the prediction of agricultural commodity prices using various neural network-based algorithms such as BPNN (Backpropagation Neural Network), GA (Genetic Algorithm), ANN-GA (Artificial Neural Network-Genetic Algorithm), RBFNN (Radial Basis Function Neural Network), SVM, and ARIMA. Evaluation measures used included MAE, MAPE, MSE, RMSE, MAD (Mean Absolute Deviation), R^2, SMAPE (Symmetric Mean Absolute Percentage Error), MIV (Mean Impact Value), SPA (Sum of Product of Accuracy), among others.

These algorithms were employed to improve

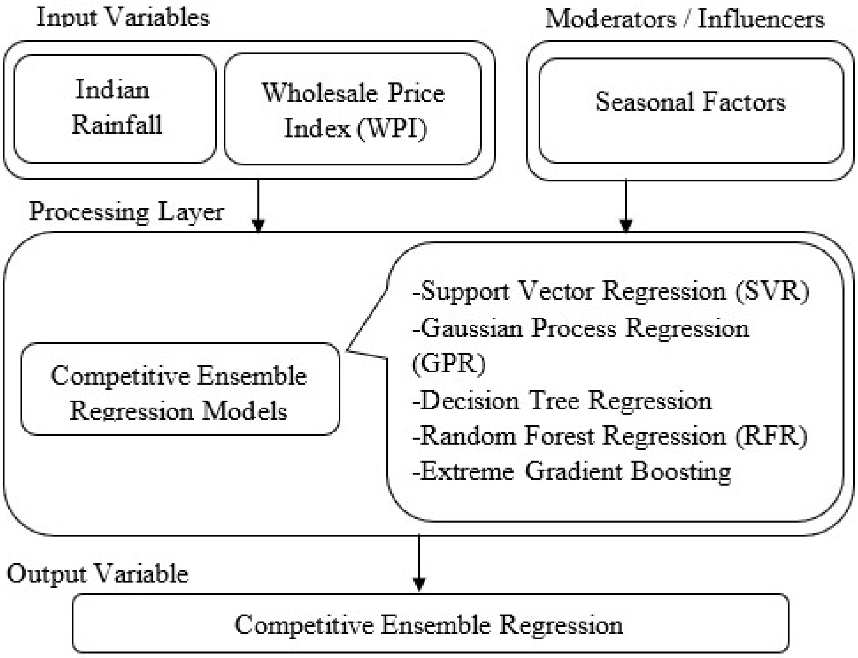

Figure 1. Conceptual framework diagram

the accuracy of price prediction for agricultural products (Bayona-Ore *et al.* 2021). Conducted numerical experiments on multiple datasets from the UCI Machine Learning Repository. They explored ensemble methods such as bagging, boosting, and stacking/blending for regression tasks. The focus was on identifying optimal hyperparameter values for these ensemble models. The authors emphasized the superiority of the ensemble approach, which demonstrated higher accuracy in predictive tasks compared to individual models (Shahhosseini *et al.* 2022). Focused on predicting temperature, rainfall, and soil conditions for agricultural purposes. They employed popular deep learning algorithms such as CNN (Convolutional Neural Network), DNN (Deep Neural Network), and LSTM in the neural network model category. While specific evaluation measures were not mentioned, these algorithms were widely recognized as effective for prediction tasks in the agricultural domain (Van Klompenburg *et al.* 2020). (Zhang *et al.* 2023) planned a hybrid ensemble model to combine the Variational Mode Decomposition (VMD) with Extreme Learning Machine (ELM) to forecast agricultural futures prices, attaining lower RMSE compared to ARIMA and machine learning baselines, and effectively capturing nonlinear trends. Similarly, (Zhang & Tang 2024) introduced a hybrid quadratic technique with LSTM to forecast commodity stocks such as soybean and corn, achieving the lowest RMSE of 0.297.

(Sari *et al.* 2024) applied the GA-ELM technique to forecast 11 agricultural commodities, achieving an MSE of 0.312 over ARIMA and GA-LSTM. (Guo *et al.* 2022) developed a spatio-temporal ensemble with ARIMA, BP neural networks, and LSTM, showing a 12% improvement in prediction accuracy using RMSE and MAPE metrics when factoring in seasonal and regional data. (Choudhary *et al.* 2025) presented a GA-optimized hybrid VMD-LSTM model for predicting vegetable prices such as onion and tomato, and reported the lowest MSE of 0.288, highlighting the effectiveness of decomposition and genetic optimization in agricultural price forecasting.

These studies collectively highlight the diverse range of models, algorithms, and evaluation measures used in predicting agricultural commodity prices. From traditional methods such as ARIMA and SVR to advanced techniques like LSTM and ensemble models, researchers are continuously exploring and refining approaches to enhance accuracy. These efforts ultimately aim to support farmers, stakeholders, and decision-makers in making informed decisions and managing risks in the

Table 1. Regression models used for agricultural commodity price prediction

Base Regressor Model (BRM)	Definition	Why
Support Vector Regression (SVR)	A regression technique based on Support Vector Machines that uses a margin of tolerance and kernel functions to fit complex, non-linear relationships in the data.	Offers probabilistic predictions and confidence intervals.
Gaussian Process Regression (GPR)	A non-parametric, Bayesian approach to regression that assumes a Gaussian distribution over functions and provides a measure of uncertainty in predictions.	Often achieves state-of-the-art results in Kaggle competitions.
Decision Tree Regression (DTR)	A tree-based model that splits data based on feature values to predict outcomes. Each internal node represents a decision rule, and each leaf node gives a prediction.	Transparent and easy to visualize.
Random Forest Regression (RFR)	An ensemble method that builds multiple decision trees and averages their predictions to improve accuracy and reduce overfitting.	Robust, handles missing data and overfitting well.
Extreme Gradient Boosting (XGBoost)	A boosting-based ensemble method that builds models sequentially, each correcting errors made by the previous one, optimized for speed and performance.	Effective with proper kernel choice and small datasets.

BRM = Base Regressor Model.

agricultural industry. However, existing models often lack integrated approaches that combine rainfall data with market variables specific to the Indian agromarket. This study addresses this research gap by proposing a competitive ensemble regression model using WPI and rainfall data. The main research questions are: How does monthly rainfall impact Indian commodity prices? Which ensemble methods best improve forecasting accuracy? A conceptual framework diagram is presented in Figure 1 to illustrate the relationships between variables and decision making.

Predicting techniques

A summary of the various regression models is presented in Table 1. It is important to note that several other machine learning techniques exist for this task. However, the models included in this study were selected based on their frequent application in the existing literature for predicting permeability, highlighting their relevance and effectiveness in similar predictive contexts.

Performance measures

Mean Absolute Error (MAE): Average of the absolute differences between actual and predicted values. It is shown in equation 1.

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |actualY - predictedY| \quad (1)$$

Mean Squared Error (MSE): Average of the squared differences between actual and predicted values. It is estimated by using equation 2.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(actualY - predictedY\right)^2 \quad (2)$$

Root Mean Squared Error (RMSE). Square root of the MSE. It is shown in equation 3.

$$\begin{aligned} RMSE &= \sqrt{MSE} \\ &= \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_{\text{actual}} - Y_{\text{predicted}}\right)^2} \end{aligned} \quad (3)$$

R-squared (R^2) : Proportion of the variance in the dependent variable explained by the regression model. It is shown in equation 4.

$$R^2 = 1 - \frac{\sum\left(actualY - predictedY\right)^2}{\sum\left(actualY - predictedY\right)^2} \quad (4)$$

Mean of the absolute percentage errors (MAPE) between actual and predicted values. Equation 5 explains it.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{abs(|actualY - predictedY|)}{actualY} \quad (5)$$

Equation 6 explains percentage ratio of the total absolute error to the total actual values.

$$PE = \frac{\sum_{i=1}^{n}|actualY - predictedY|}{\sum_{i=1}^{n} actualY} \times 100 \quad (6)$$

Proposed work

The proposed ensemble regression framework consists of several stages, including data collection, preprocessing, and prediction of wholesale prices for crops such as *paddy, maize, ragi, barley, and wheat.*

The summary of the main contribution of this ensemble learning approach using Roulette Wheel Selection (RWS) is as follows.

The proposed work competitive ensemble learning-based forecasting strategy is used to predict the price forecasting trend for *paddy, ragi, barley, maize and wheat* in India, which consists of four basic forecasting models, and algorithms to achieve more accurate and robust performance under various change types.

Stage 1: Data preprocessing

Data were obtained from the official portal (data.gov.in) and prepared for regression modelling. Real-world datasets often contain missing values, which can negatively impact machine learning performance and reduce model accuracy. Since many algorithms cannot handle missing data directly, this study uses suitable imputation techniques to replace missing entries before modelling. In this study, missing values were addressed using imputation techniques supported by Python packages such as Pandas and Scikit-learn. These tools offer efficient methods to detect and replace missing entries with statistically meaningful values (e.g., mean, median, or mode) or model-based imputations.

Stage 2: Base regressor models

The predictive framework, outlined in Algorithm 1 and Figure 1, consists of training and testing phases. First, the dataset is split—70% for training and 30% for testing. Five regression models are trained: XGBoost Regressor (XGB), Random Forest (RF), Support Vector Regressor (SVR), Decision Tree Regressor (DT), and Gaussian Process Regressor (GPR).

During training, each model is validated (using 10-fold cross-validation) to monitor performance and mitigate overfitting. Once trained, the models are evaluated on the test set, which shares characteristics with the training data, using performance metrics like RMSE, R^2, and Percentage Error (PE).

To select the optimal regressor, this study applies Roulette Wheel Selection (RWS). Each model i has a computed fitness f_i, and its selection probability is:

$$p_i = \frac{f_i}{\sum_j f_j} \quad (7)$$

Where p_i is the probability that individual regressor model i will be selected, f_i is the fitness of individual regressor model i and $\sum_j f_j$ represents the sum of all the fitness of the

Algorithm 1: Crop Price Prediction Model using Penalty Reward Fitness Function

Input: Dataset D (Rainfall, Crop_WPI); - Base year price (P_base) per 100 kg
Output: Forecasted Crop_WPI; Forecasted crop price for 100 kg
procedure PREDICTION ()

1. Split dataset D into training set D_train and test set D_test
2. Initialize:
   - modelList ← list of regression models
   - fitnessList ← empty list to store g(Model_i)
   - performanceDict ← dictionary to store RMSE, R^2, PEs
3. For each model M_i in modelList:
   a. Train M_i on D_train
   b. Predict: WPI_train_pred ← M_i(D_train)
   c. Compute:
      - RMSE_train ← RMSE (WPI_train, WPI_train_pred) using equation 3
      - R^2 ← R^2 (WPI_train, WPI_train_pred) using equation 4
      - PE ← Percent Error using equation 6
   d. Define Fitness Flags:
      - f1(M_i) = calculate f1(M_i) using equation 9
      - f2(M_i) = calculate f2(M_i) using equation 10
   e. Predict on test set: WPI_test_pred ← M_i(D_test)
   f. Compute RMSE_test ← RMSE (WPI_test, WPI_test_pred)
   g. g(M_i) = Compute fitness of M_i using equation?
   h. Append g(M_i) to fitnessList
   i. Store metrics in performanceDict[M_i] ← (RMSE_test, R^2, PE)
4. Perform Roulette Wheel Selection:
   a. For each g_i in fitnessList:
      w_i = 1 / (g_i + ε), where ε = small constant (e.g., 10^{-6})
   b. Normalize weights:
      total_weight = Σ w_i
      P_i = w_i / total_weight
   c. Generate random number r ∈ [0,1]
   d. Traverse cumulative probabilities C_i:
      Select model M_opt where C_i-1 < r ≤ C_i
5. Forecast Crop_WPI using selected model:
   - Forecasted_WPI ← M_opt(D_test_latest)
6. Calculate Crop Price per 100 kg using WPI:
   - Price_per_100kg = (Forecasted_WPI / 100) × Base year price (P_base) per 100 kg
7. Return Forecasted_WPI, Forecasted Price_per_100kg

End

regressors in the population.

The function $g(x)$ penalizes models less by subtracting 2 from RMSE if both performance indicators (R^2 and PE) are satisfactory. This is known as a valid penalty-reward-based fitness design, shown in equation (8). $F_1(\text{Model}_i)$ and $F_2(\text{Model}_i)$ are shown in equations (9) and (10).

$$\text{Fitness function } f = \arg\min(g(x)) \tag{8}$$

where

$$g(x) = \begin{cases} \text{RMSE}(\text{Model}_i, \text{TestData}) - 2, \\ \text{if } f_1(\text{Model}_i) = 1 \text{ and } f_2(\text{Model}_i) = 1, \\ \text{RMSE}(\text{Model}_i, \text{TestData}), \\ \text{if } f_1(\text{Model}_i) = 1, \\ \text{RMSE}(\text{Model}_i, \text{TestData}), \\ \text{if } f_1(\text{Model}_i) = 0 \text{ or } f_2(\text{Model}_i) = 0. \end{cases} \tag{9}$$

$$f_1(\text{Model}_i) = \begin{cases} 1, & R^2(\text{Model}_i) \geq 0.9 \\ 0.7, & 0.7 < R^2(\text{Model}_i) < 0.9 \\ -1, & R^2(\text{Model}_i) < 0 \\ 0, & \text{Otherwise} \end{cases} \tag{10}$$

Table 2. Performance metric values of BRMs for paddy dataset

Model	Train RMSE	Test	MAE	Train R^2	Test ^{2}	MAPE	Train Percent Error	Test Percent Error
XGB Regressor	0.002	1.86	1.66	0.99	0.98	0.01	1.40	0.01
Random Forest Regressor	0.56	1.74	1.38	0.99	0.98	0.009	0.002	0.009
Decision Tree Regressor	0	1.50	1.23	1	0.98	0.008	0	0.008
Gaussian Process Regressor	1.3	134.74	130.76	1	-85.81	0.88	8.53	0.88
SVR	16.91	14.55	13.14	-0.01	-0.01	0.09	0.09	0.08

$$f_2(\text{Model}_i) = \begin{cases} 1, & PE(\text{Model}_i) \geq 10 \\ 0.7, & 10 < PE(\text{Model}_i) < 25 \\ 0, & \text{Otherwise} \end{cases} \tag{11}$$

Competitive ensemble learning using penalty-reward-based fitness function

A competitive ensemble, also known as a competitive learning ensemble, refers to a machine learning technique that combines multiple models in order to improve performance and achieve better predictive accuracy. It is based on the concept of ensemble learning, which leverages the diversity and collective intelligence of multiple models to make more accurate predictions compared to individual models. Five regressor algorithms were used to provide the forecasted crop's WPI in stage 3 competitive ensemble learning. Based on the hyperparameters and model behaviour, each method has a little variation in their predicted values. For optimal performance, all hyperparameters are chosen experimentally by trial-and-error.

Data preparation

This section discusses the experiments carried out to quantify the performance of the proposed model. Initially, the commodities Wholesale price dataset was collected from the official government website, office of the economic adviser department for promotion of industry and internal trade (Ministry of Commerce and Industry, 2025). Monthly rainfall in India data was collected from website The Hindu business line (The Hindu Business Line, 2023). The pre-processed data were then cleaned and filtered to fit the regression models. Then, the performance of the aforementioned methodology was evaluated in two setups (i.e., classical ensemble learning and competitive ensemble learning) according to the input data. On the basis of the RMSE, the best and worst regressor models were examined.

RESULTS AND DISCUSSION

Tables 2, 3, 4, 5 and 6 display the performance results of five BRMs for the crops of paddy, maize, ragi, barley, and wheat. The metric includes RMSE value, R2 Value, and PE. The test datasets MAE and MAPE results are also included. When comparing the RMSE values of the baseline regression models for the training and test datasets, we can determine whether a model is overfitting or not.

In this study, a comprehensive analysis of model fitness for a dataset was conducted using a two-step process. Initially, fitness criteria were established based on the performance metrics 'Test_R2value' and 'Test_PercentError.' Specifically, three fitness levels were defined for 'Test_R2value': a fitness value of 1 if the value equalled or exceeded 0.9, a fitness value of 0.7 if the value fell between 0.9 and 0.7, and a fitness value of 0 for values below 0.7. Similarly, for 'Test_PercentError,' a fitness value of 1 was assigned if the value was 2% or less, 0.7 for values between 2% and 5%, and 0 for values exceeding 5%. These

Table 3. Performance metric values of BRMs for maize dataset

Model	Train $_{R}MSE$	Test _RMSE	MAE	Train _R2Value	Test _R2Value	MAPE	Train _Percent Error	Test _Percent Error
XGB Regressor	0.007	11.39	6.24	0.99	0.72	0.04	3.68	0.04
Random Forest Regressor	2.22	11.96	6.39	0.98	0.69	0.04	0.01	0.04
Decision Tree Regressor	0	14.95	8.44	1	0.53	0.06	0	0.06
Gaussian Process Regressor	1.26	126.12	120.98	1	-32.36	0.88	8.63	0.89
SVR	22.74	21.92	17.16	-0.04	-0.008	0.12	0.12	0.12

Table 4. Performance metric values of BRMs for barley dataset

Model	Train RMSE	Test RMSE	MAE	Train R2Value	Test R2Value	MAPE	Train Percent Error	Test Percent Error
XGB Regressor	0.01	5.99	5.002	0.99	0.96	0.03	9.07	0.03
Random Forest Regressor	2.91	7.47	4.73	0.99	0.94	0.03	0.01	0.03
Decision Tree Regressor	0	8.82	5.69	1	0.91	0.03	0	0.03
Gaussian Process Regressor	1.36	137.6	131.28	1	-18.95	0.89	8.58	0.89
SVR	30.78	30.85	22.33	-0.02	-0.003	0.14	0.14	0.15

Table 5. Performance metric values of BRMs for wheat dataset

Model	Train RMSE	Test RMSE	MAE	Train R2Value	Test R2Value	MAPE	Train Percent Error	Test Percent Error
XGB Regressor	0.004	6.39	4.36	0.99	0.85	0.02	2.39	0.03
Random Forest Regressor	1.69	5.67	3.65	0.99	0.88	0.02	0.008	0.02
Decision Tree Regressor	0	7.78	4.63	1	0.79	0.03	0	0.03
Gaussian Process Regressor	1.27	128.35	124.39	1	-55.62	0.89	8.56	0.89
SVR	19.07	17.13	14.88	-0.008	-0.009	0.10	0.10	0.10

Table 6. Performance metric values of BRMs for ragi dataset

Model	Train RMSE	Test RMSE	MAE	Train R2Value	Test R2Value	MAPE	Train Percent Error	Test Percent Error
XGB Regressor	0.01	5.10	4.43	0.99	0.97	0.02	6.52	0.02
Random Forest Regressor	2.96	4.67	3.72	0.99	0.97	0.02	0.01	0.01
Decision Tree Regressor	0	4.96	4.26	1	0.97	0.02	0	0.02
Gaussian Process Regressor	1.76	174.39	164.20	1	-33.4	0.85	8.43	0.86
SVR	38.97	36.95	30.76	-0.15	-0.54	0.18	0.15	0.16

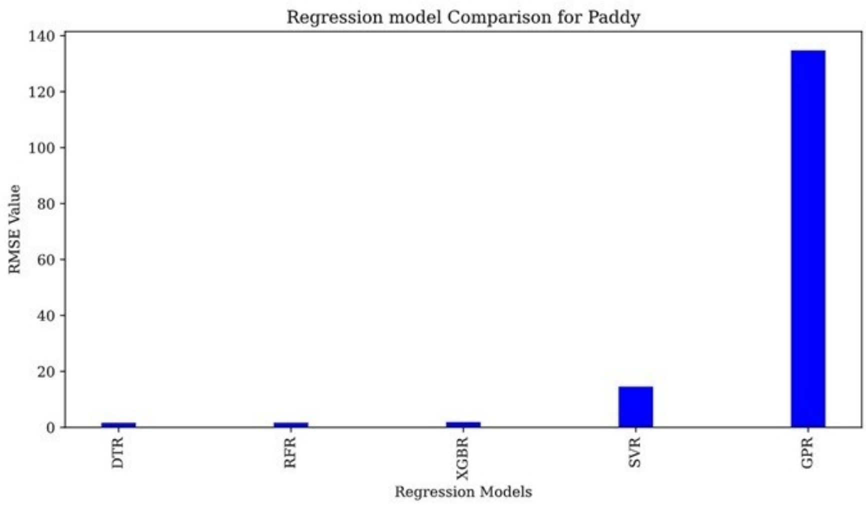

Figure 2 (A). RMSE Comparison with different models on paddy wholesale price data.

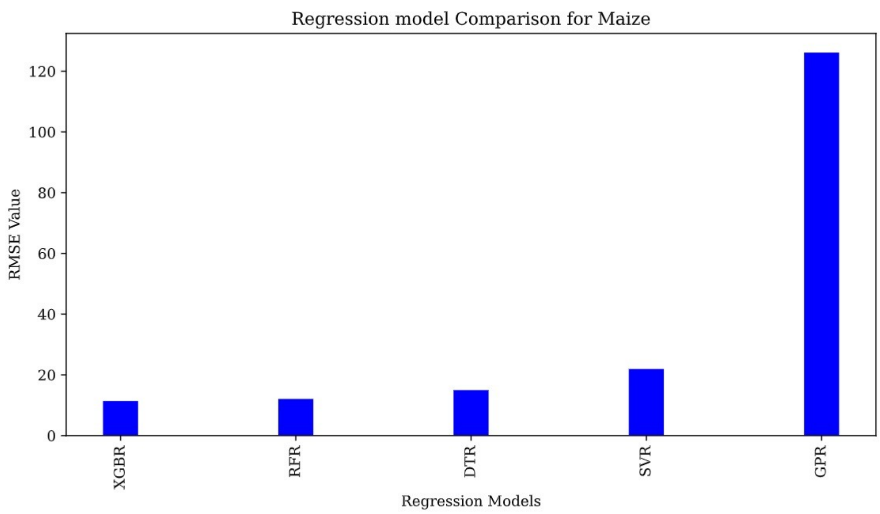

Figure 2 (B). RMSE Comparison with different models on maize wholesale price data.

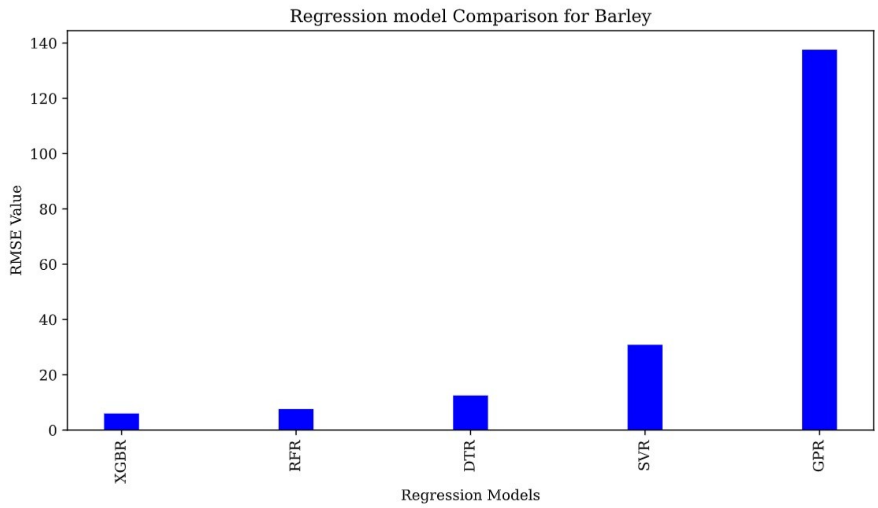

Figure 2 (C). RMSE Comparison with different models on barely wholesale price data.

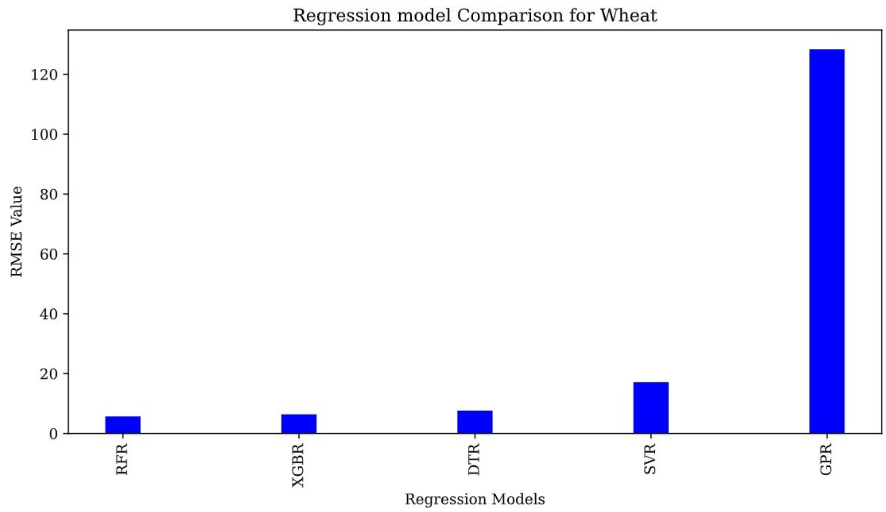

Figure 2 (D). RMSE Comparison with different models on wheat wholesale price data.

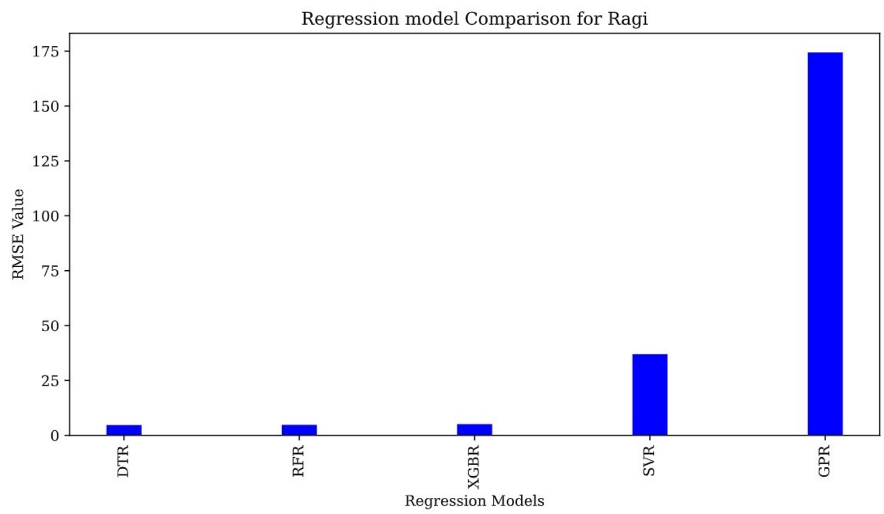

Figure 2 (E). RMSE Comparison with different models on ragi wholesale price data.

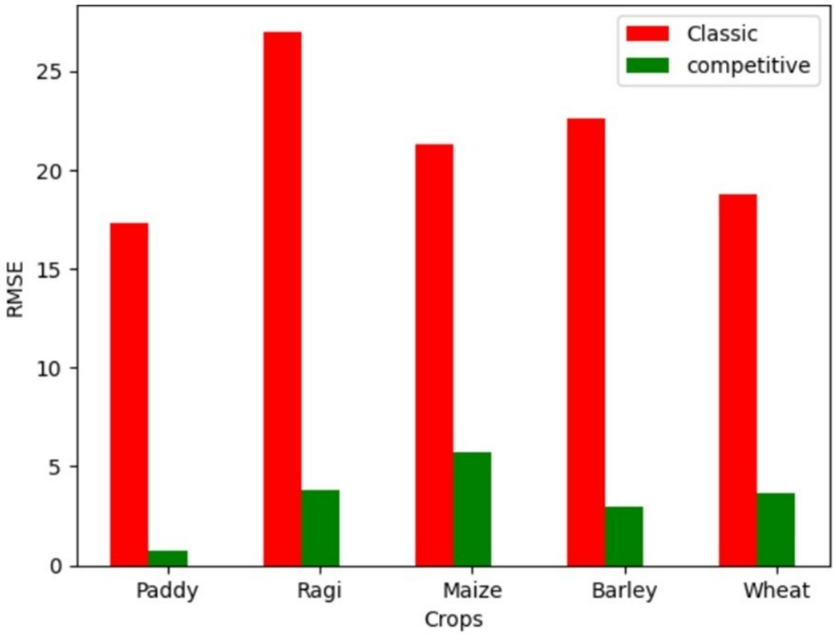

Figure 3. Comparative performance of ensemble methods.

fitness criteria were thoughtfully designed to capture the predictive quality of model performance.

Subsequently, these criteria were seamlessly integrated into the analysis framework. The computed fitness values for each model were aggregated into a unified fitness metric denoted as 'FitnessValue' by summing the 'F1' and 'F2' components, facilitating an assessment of each model's overall fitness within our analysis. Additionally, to select the best model, a roulette wheel selection algorithm was implemented, allowing for the probabilistic identification of the most suitable model based on its fitness score, thereby enhancing the model selection process in the context of this study.

Fig. 2 (A), (B), (C), (D), and (E) illustrate the RMSE-based comparison of regression models across all crop datasets. The X-axis represents the base regression models, while the Y-axis shows their corresponding RMSE values. The bar charts clearly indicate that the XGBoost Regression model consistently achieves the lowest RMSE across all crops. Table 7 shows the relative performance of ensemble methods across all crop datasets.

It includes RMSE value for training datasets, test datasets and average RMSE value. It is clearly noted that competitive ensemble method has lower RMSE for all datasets. The competitive ensemble method showed the improved performance by 53.9%, 93.1%, 101.9%, 121.5% and 16.6% for barley, Maize, Paddy, Wheat and Ragi datasets respectively. Figure 3 shows the Comparative Performance of Ensemble Methods. Table 08 shows the crops of paddy, maize, ragi, barley, and wheat forecast prices for next one year. Discussion: The observations drawn from these results are summarized below:

- For the paddy dataset, SVR, RFR, and XGB Regressor models did not suffer from overfitting, whereas DTR and GPR exhibited overfitting issues.
- For the maize dataset, SVR, RFR, and XGB Regressor models also avoided overfitting, while DTR and GPR showed clear signs of overfitting.
- In the case of barley and wheat datasets, only SVR did not suffer from overfitting, and the remaining models showed overfitting tendencies.
- For the ragi dataset, SVR and XGB Regressor did not suffer from overfitting, whereas RFR, DTR, and GPR exhibited overfitting.
- Interestingly, even though SVR avoided overfitting, the XGBoost Regressor consistently achieved lower RMSE values on the test datasets than SVR, across all datasets.

A comprehensive evaluation reveals that XGBoost is the most robust and generalizable model across all five crop datasets. As seen in Tables 2 to 6, XGBoost

Table 7. Comparative performance of ensemble methods

Crop Dataset	Classical Ensemble Model			Competitive Ensemble Model		
	Train RMSE	Test RMSE	Average RMSE	Train RMSE	Test RMSE	Average RMSE
Barley	7.016	38.14	22.58	0.01	5.99	3.004
Maize	5.248	37.27	21.26	0.007	11.39	5.69
Paddy	3.755	30.88	17.319	0	1.50	0.75
Wheat	4.409	33.06	18.73	1.69	5.67	3.68
Ragi	8.740	45.218	26.98	2.96	4.67	3.82

Table 8. Forecast price for commodities

Month	Paddy	Maize	Barley	Wheat	Ragi
January	3556.10	3293.96	3196.97	3358.92	8871.76
February	3529.91	3395.51	3277.87	3369.30	8937.25
March	3529.91	3559.69	3251.74	3445.32	8948.06
April	3588.85	3786.97	3819.42	3471.90	8943.29
May	3604.13	3805.14	3985.04	3488.87	8881.76
June	3612.86	3764.48	4039.20	3479.74	8704.95
July	3612.86	3925.73	4012.34	3532.35	8585.77
August	3684.90	3992.28	4033.86	3634.67	8347.28
September	3748.21	4040.94	4071.56	3650.75	8055.37
October	3789.68	3885.89	3988.64	3725.49	8137.13
November	3783.13	3814.75	4126.93	3842.84	8174.28
December	3783.13	3876.68	4101.48	3883.30	8155.78

consistently produced the lowest test RMSE values (e.g., 1.86 for paddy, 11.39 for maize, 5.99 for barley, 6.39 for wheat, and 5.1 for ragi) and high test R^2 scores (ranging from 0.85 to 0.98), indicating strong predictive performance. In contrast, although SVR avoided overfitting, it yielded significantly higher RMSE values and lower R^2 scores, such as 14.55 for paddy and 36.95 for ragi, indicating poor model fitness.

Moreover, models like DTR and GPR showed clear signs of overfitting, with nearly perfect training accuracy but large test RMSEs and drastically negative test R^2 values (e.g., -85.81 for paddy using GPR), indicating poor generalization. Notably, the competitive ensemble models outperformed classical ensemble models such as RFR. XGBR delivered significantly lower RMSE values across all datasets. This validates the effectiveness and practical utility of the proposed framework in enhancing crop price prediction.

CONCLUSION

This study examined the forecasting of agricultural price trends using competitive ensemble learning with a RWS operator. The proposed framework applies competitive ensemble learning to the outputs of multiple BRMs instead of relying on a single BRM. This approach significantly improved performance and increased model robustness. The results demonstrated that the competitive ensemble learning method outperformed both individual regression models and classical ensemble approaches in terms of predictive performance. However, a key limitation of this study is the relatively small number of features in the datasets. To address this limitation and further improve accuracy, future research should incorporate additional relevant features into the dataset. Overall, this work establishes competitive ensemble learning with RWS operators as a promising and effective strategy for agricultural price prediction.

REFERENCES

Adeniran AA, Adebayo AR, Salami HO, Yahaya MO, Abdulraheem A (2019) A competitive ensemble model for permeability prediction in heterogeneous oil and gas reservoirs. *Applied Computing and Geosciences*, 1, 100004. https://doi.org/10.1016/j.acags.2019.100004.

Bayona-Ore S, Cerna R, Tirado Hinojoza E (2021) Machine learning for price prediction for agricultural products. https://doi.org/10.37394/23207.2021.18.92.

Chatzimparmpas A, Martins RM, Kucher K, Kerren A (2021) StackGenVis: Alignment of data algorithms and models for stacking ensemble learning using performance metrics. *IEEE Transactions on Visualization and Computer Graphics*, 27(2), 1547–1557. https://doi.org/10.1109/TVCG.2020.3030352.

Chen LJ, Ye C, Hu SW, Wang V, Wen J (2013) The effect of a target zone on the stabilization of agricultural prices and farmers' nominal income. *Journal of Agricultural and Resource Economics*, 34–47. https://www.jstor.org/stable/23496737.

Chen Z, Goh HS, Sin KL, Lim K, Chung NKH, Liew XY (2021) Automated agriculture commodity price prediction system with machine learning techniques. *Advances in Science, Technology and Engineering Systems*, 6(4), 376–384. https://doi.org/10.25046/aj060442.

Choudhary K, Jha GK, Jaiswal R, Kumar RR (2025) A genetic algorithm optimized hybrid model for agricultural price forecasting based on VMD and LSTM network. *Scientific Reports*, 15(1), 9932. https://doi.org/10.1038/s41598-025-94173-0.

Flach P (2012) *Machine learning: the art and science of algorithms that make sense of data*. Cambridge University Press. https://doi.org/10.1017/CBO9780511973000.

Gu YH, Jin D, Yin H, Zheng R, Piao X, Yoo SJ (2022) Forecasting agricultural commodity prices using dual input attention LSTM. *Agriculture*, 12(2), 256. https://doi.org/10.3390/agriculture12020256.

Guo Y, Tang D, Tang W, Yang S, Tang Q, Feng Y, Zhang F (2022) Agricultural price prediction based on combined forecasting model under spatial-temporal influencing factors. *Sustainability*, 14(17), 10483. https://doi.org/10.3390/su141710483.

Kumar D, Rath SK (2020) Predicting the trends of price for ethereum using deep learning techniques. In: *Artificial intelligence and evolutionary computations in engineering systems*, 103–114.

Li Z, Chen Z, Cheng Q, Duan F, Sui R, Huang X, Xu H (2022) UAV-based hyperspectral and ensemble machine learning for predicting yield in winter wheat. *Agronomy*, 12(1), 202. https://doi.org/10.1007/978-981-15-0199-9_9.

Madaan L, Sharma A, Khandelwal P, Goel S, Singla P, Seth A (2019) Price forecasting & anomaly detection for agricultural commodities in India. In: *Proceedings of the 2nd ACM SIGCAS Conference on Computing and Sustainable Societies*, 52–64. https://doi.org/10.1145/3314344.3332488.

Mahto AK, Alam MA, Biswas R, Ahmed J, Alam SI (2021) Short-Term Forecasting of Agriculture Commodities in Context of Indian Market for Sustainable Agriculture by Using the Artificial Neural Network. *Journal of Food Quality*, 2021(1), 9939906. https://doi.org/10.1155/2021/9939906.

Mao L, Huang Y, Zhang X, Li S, Huang X (2022) ARIMA model forecasting analysis of the prices of multiple vegetables under the impact of the COVID-19. *PLoS ONE*, 17(7), e0271594. https://doi.org/10.1371/journal.pone.0271594.

Milunovich G (2020) Forecasting Australia's real house price index: A comparison of time series and machine learning methods. *Journal of Forecasting*, 39(7), 1098–1118. https://doi.org/10.1002/for.2678.

Nabipour M, Nayyeri P, Jabani H, Mosavi A, Salwana E, Shahab S (2020) Deep learning for stock market prediction. *Entropy*, 22(8), 840. https://doi.org/10.3390/e22080840.

Paul RK, Yeasin M, Kumar P, Balasubramanian M, Roy HS, Paul AK, Gupta A (2022) Machine learning techniques for forecasting agricultural prices: A case of brinjal in Odisha, India. *PLoS ONE*, 17(7), e0270553. https://doi.org/10.1371/journal.pone.0270553.

Razavi-Termeh SV, Sadeghi-Niaraki A, Choi SM (2021) Spatial modeling of asthma-prone areas using remote sensing and ensemble machine learning algorithms. *Remote Sensing*, 13(16), 3222. https://doi.org/10.3390/rs13163222.

Sari M, Duran S, Kutlu H, Guloglu B, Atik Z (2024) Various optimized machine learning techniques to predict agricultural commodity prices. *Neural Computing and Applications*, 36(19), 11439–11459. https://doi.org/10.1007/s00521-024-09679-x.

Shahhosseini M, Hu G, Pham H (2022) Optimizing ensemble weights and hyperparameters of machine learning models for regression problems. *Machine Learning with Applications*, 7, 100251. https://doi.org/10.1016/j.mlwa.2022.100251.

Sheehy JE, Mitchell PL, Ferrer AB (2006) Decline in rice grain yields with temperature: Models and correlations can give different estimates. *Field Crops Research*, 98(2–3), 151–156. https://doi.org/10.1016/j.fcr.2006.01.001.

Soni N, Raut J (2022) Crop price prediction using machine learning techniques. *International Advanced Research Journal in Science, Engineering and Technology*, 9(5), 757–762. https://doi.org/10.17148/IARJSET.2022.95107.

Suresha HP, Mutturaj U, Krishna KT (2021) Indian commodity market price comparative study of forecasting methods: A case study on onion, potato and tomato. *Asian Journal of Research in Computer Science*, 12, 147–159. https://doi.org/10.9734/ajrcos/2021/v12i430300.

Talasila V, Prasad C, Sagar Reddy GT, Aparna A (2020) Analysis and prediction of crop production in Andhra region using deep convolutional regression network. *International Journal of Intelligent Engineering & Systems*, 13(5). https://doi.org/10.22266/ijies2020.1031.01.

Ticlavilca AM, Feuz DM, McKee M (2010) Forecasting agricultural commodity prices using multivariate Bayesian machine learning regression. https://doi.org/10.22004/ag.econ.285320.

Ustuner M, Balik Sanli F (2019) Polarimetric target decompositions and light gradient boosting machine for crop classification: A comparative evaluation. *ISPRS International Journal of Geo-Information*, 8(2), 97. https://doi.org/10.3390/ijgi8020097.

Utomo AH, Gumilang MA, Ahmad A (2022) Agricultural commodity sales recommendation system for farmers based on geographic information systems and price forecasting using probabilistic neural network algorithm. In: *IOP Conference Series: Earth and Environmental Science*, 980(1), 012061. https://doi.org/10.1088/1755-1315/980/1/012061.

Van Klompenburg T, Kassahun A, Catal C (2020) Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture*, 177, 105709. https://doi.org/10.1016/j.compag.2020.105709.

Wang X, Gao S, Guo Y, Zhou S, Duan Y, Wu D (2022) A combined prediction model for hog futures prices based on WOA–LightGBM–CEEMDAN. *Complexity*, 2022(1), 3216036. https://doi.org/10.1155/2022/3216036.

Weng Y, Wang X, Hua J, Wang H, Kang M, Wang FY (2019) Forecasting horticultural products price using ARIMA model and neural network based on a large-scale data set collected by web crawler. *IEEE Transactions on Computational Social Systems*, 6(3), 547–553. https://doi.org/10.1109/TCSS.2019.2914499.

Werbos PJ (1988) Generalization of backpropagation with application to a recurrent gas market model. *Neural Networks*, 1(4), 339–356. https://doi.org/10.1016/0893-6080(88)90007-X.

Yang D, Liu Z (2012) Does farmer economic organization and agricultural specialization improve rural income? Evidence from China. *Economic Modelling*, 29(3), 990–993. https://doi.org/10.1016/j.econmod.2012.02.007.

Yao S, Guo Y, Huo X (2010) An empirical analysis of the effects of China's land conversion program on farmers' income growth and labor transfer. *Environmental Management*, 45, 502–512. https://doi.org/10.1007/s00267-009-9376-7.

Zhang DZL, Ling L (2023) Decomposition and integrated forecasting model of agricultural futures prices based on VMD-ELM. *Operations Research and Management*, 32(1), 127. https://doi.org/10.12005/orms.2023.0021.

Zhang D, Chen S, Liwen L, Xia Q (2020) Forecasting agricultural commodity prices using model selection framework with time series features and forecast horizons. *IEEE Access*, 8, 28197–28209. https://doi.org/10.1109/ACCESS.2020.2971591.

Zhang GP, Kline DM (2007) Quarterly time-series forecasting with neural networks. *IEEE Transactions on Neural Networks*, 18(6), 1800–1814. https://doi.org/10.1109/TNN.2007.896859.

Zhang GP, Qi M (2005) Neural network forecasting for seasonal and trend time series. *European Journal of Operational Research*, 160(2), 501–514. https://doi.org/10.1016/j.ejor.2003.08.037.

Zhang T, Tang Z (2024) Agricultural commodity futures prices prediction based on a new hybrid forecasting model combining quadratic decomposition technology and LSTM model. *Frontiers in Sustainable Food Systems*, 8, 1334098. https://doi.org/10.3389/fsufs.2024.1334098.

Zhang Y, Na S (2018) A novel agricultural commodity price forecasting model based on fuzzy information granulation and MEA-SVM model. *Mathematical Problems in Engineering*, 2018(1), 2540681. https://doi.org/10.1155/2018/2540681.

Zhang Y, Liu J, Shen W (2022) A review of ensemble learning algorithms used in remote sensing applications. *Applied Sciences*, 12(17), 8654. https://doi.org/10.3390/app12178654.