International Journal of Advanced Biochemistry Research

ISSN Print: 2617-4693
ISSN Online: 2617-4707
NAAS Rating (2026): 5.29
IJABR 2026; SP-10(1): 181-190
www.biochemjournal.com
Received: 09-10-2025
Accepted: 12-11-2025

Study on forecasting of Indian onion export price using machine learning techniques-based hybrid models

Jhade Sunil, Abhishek Singh, Manjubala M and Abinayarajam D

DOI: https://www.doi.org/10.33545/26174693.2026.v10.i1Sc.6908

Abstract

The onion market is profoundly shaped by global trade dynamics, rapid technological advancement, and shifting consumer demand. Within this global context, India stands out as the world's largest producer and second-largest exporter of onions. Despite this prominent position, the nation's economy and export potential are threatened by significant price volatility. Consequently, accurate price predictions are crucial for farmers, policymakers, and the government to make informed decisions. Unfortunately, agricultural datasets often exhibit nonlinearity and non-stationarity, making predictions challenging. To address this, hybrid models are proposed, which combine multiple models instead of relying on individual ones. Specifically, this article compares the performance of individual models (ARIMA, TDNN, SVR) against a suite of hybrid models (EMD-ARIMA, EMD-TDNN, EMD-SVR, EEMD-ARIMA, EEMD-TDNN, EEMD-SVR). Monthly onion price data from 2008 to 2021 is used, and the dataset is decomposed into six independent intrinsic modes and one residue, revealing price volatility patterns. The decomposed components are forecasted using conventional and machine learning methods, and the forecasts are aggregated to produce a final prediction. Empirical evaluation demonstrates that the EEMD-SVR model achieves superior predictive accuracy, with the lowest RMSE (405.09) and MAPE (14.93%). This highlights its effectiveness in modeling the complex, nonlinear, and non-stationary dynamics of agricultural prices.

Keywords: Onion export, price forecasting, machine learning, hybrid model, empirical mode decomposition, ensemble empirical mode decomposition.

Introduction

The global marketplace, technological innovations, and consumer demands in the onion industry continue to evolve (Raka and Ramesh, 2017) [31]. India, the world's largest onion producer, recorded an output of about 26.7 million metric tonnes in 2023-24 (Dohlman *et al.*, 2025) [16], and exported nearly 1.6 million metric tonnes (APEDA, 2024). However, despite strong production and exports, domestic onion prices remain highly volatile, leading to instability in farmers' incomes, consumer affordability, and export competitiveness (Debopam *et al.*, 2021; Dhotre *et al.*, 2025) [14, 15].

Onion prices in the domestic market exhibited sharp fluctuations during 2019 and 2021, primarily due to inadequate information regarding future prices and severe supply shocks (Ghosh *et al.*, 2022) [18]. These shocks were driven by low yields, delayed sowing, and crop losses caused by excessive rainfalls (Satish 2018; Imran *et al.*, 2025) [33, 21]. In this situation, the government imposed an onion export ban, and resorted to imports from Afghanistan, Egypt, Turkey, and Iran to stabilize the market (Biswas 2019, and Saxena *et al.*, 2024) [6, 34]. However, such interventions often result in significant financial losses for farmers (Kumar *et al.*, 2020) [25]. Price forecasting of agricultural commodities is crucial for farmers, policymakers, and the government; however, accurate forecasting remains difficult due to unpredictable factors such as weather, market dynamics, and inadequate storage facilities (Ajmal *et al.*, 2024) [4]. Agricultural price data are often nonstationary and nonlinear in nature (Wang *et al.*, 2020) [45]. Traditional statistical approaches such as ARIMA, SARIMA, and GARCH have been widely applied. For instance, Agbo (2023) [2] examined export crop price volatility in Egypt using ARIMA/GARCH, Sahu (2024) [32] studied potato price and export

Jhade Sunil
Teaching Associate, Department of Statistics and Mathematics, BJR Agricultural College, Sircilla, PJTAU, Hyderabad, Telangana, India

Abhishek Singh
Professor, Department of Farm Engineering, Institute of Agricultural Sciences, Banaras Hindu University, Varanasi, Uttar Pradesh, India

Manjubala M
Assistant Professor in Agricultural Statistics, RPCAU, Pusa, Samastipur, Bihar, India

Abinayarajam D
Research Scholar, Department of Farm Engineering, Institute of Agricultural Sciences, Banaras Hindu University, Varanasi, Uttar Pradesh, India

Corresponding Author:
Jhade Sunil
Teaching Associate, Department of Statistics and Mathematics, BJR Agricultural College, Sircilla, PJTAU, Hyderabad, Telangana, India

behaviour with similar models, and Vinay *et al.* (2024) compared ARIMA and SARIMA for onion price forecasting, confirming their short-term effectiveness. Yadav (2024) [49] further demonstrated that SARIMA outperformed other univariate models in forecasting global wheat prices. Nevertheless, while these models are valuable for linear and stationary datasets, they often fail to capture the nonlinear and highly volatile dynamics of agricultural markets.

To address these limitations, machine learning (ML) and AI methods have gained prominence for their ability to capture complex patterns. Xu and Zhang (2021, 2022) [47-48] applied neural networks to commodities such as coffee, corn, cotton, soybeans, sugar, and wheat for price forecasting. Mohanty *et al.* (2023) [28] proposed an ML-based framework for crop price prediction, while Patil *et al.* (2023) [29] emphasized its role in ensuring food security. Brignoli *et al.* (2024) [8] applied ML to grain yield prediction futures, and Yerukala *et al.* (2024) [50] used neural networks for agricultural commodity forecasting. More recently, Theofilou *et al.* (2025) [42] demonstrated the effectiveness of ML in staple food crops. Collectively, these studies highlight the versatility of ML across diverse crops and markets, though they also caution that AI models can be prone to issues such as local minima and overfitting.

To better address the nonlinearity and nonstationarity in agricultural price series, hybrid and decomposition-based approaches have received growing attention. Among these, noise-assisted decomposition methods such as EMD and EEMD have been particularly emphasized by Zhang (2003) [52], Ince and Trafalis (2006) [22], and Chen *et al.* (2012) [9]. Guo *et al.* (2012) [19] combined EMD with neural networks to enhance wind speed forecasting, while Abadan and Shabri (2014) [1] showed that EMD-ARIMA improved rice price forecasts. Choudhary *et al.* (2019) [10] demonstrated that EEMD-TDNN outperformed TDNN for potato prices. Das *et al.* (2020) [13] reported the superiority of EMD-SVR over standard SVR. Silva *et al.* (2021) [37] highlighted the effectiveness of ensemble techniques for corn and sugar, and Purohit *et al.* (2021) [30] applied hybrid methods to tomato, onion, and potato markets. Khan *et al.* (2022) [23] employed EEMD for short-term forecasts, while Zelingher and Makowski (2023) [51] demonstrated hybrid effectiveness in maize and cocoa markets. More recent studies further confirm these advancements: Shobharani *et al.* (2024) [36] found TDNN and hybrid approaches outperforming SARIMA for tomato and capsicum, and Mao and Soonthornphisaj (2024) [27] showed that ensemble models improved maize price prediction in Thailand.

Despite these developments, onion export price forecasting remains relatively underexplored. This study addresses this gap by evaluating hybrid forecasting frameworks for Indian onion export prices with key objectives: To assess the robustness of hybrid ML models in volatile vegetable markets, and to provide actionable guidance for model selection and trade-related decision-making.

2. Methods and Materials

2.1 Data source

This study used monthly export price data for Indian onions (₹/Quintal) from January 2008 to December 2021, collected from the Agricultural and Processed Food Products Export Development Authority (APEDA). As shown in figure 3, data were split into training and testing sets, the first 156 observations were utilized for model building and last 12 data points were used for validation purpose. Pre-processing, modelling, and evaluation were performed using Python (v3.5), and R (v4.3.2) with packages such as forecast, t series, caret, e1071, EMD, and Rlibeemd.

2.2 Autoregressive Integrated Moving Average (ARIMA)

This model is one of the most widely used approaches for time series forecasting. It is denoted as ARIMA (p, d, q), where p represents the autoregressive order, d is the degree of differencing, and q is the moving average order, defined in equation 2.2.1. ARIMA is particularly useful for modelling non-stationary linear time series. For seasonal data, the model is extended to the Seasonal ARIMA (SARIMA), denoted as ARIMA (p, d, q)(P, D, Q)s, where (P, D, Q) are the seasonal parameters and *s* is the seasonal period, it is stated in equation 2.2.2 (Box *et al.*, 1976).

$$\varphi(B)\Delta^d y_t = \theta(B)u_t \quad (2.2.1)$$

$$(\Delta_s)^D \varphi(B)\Phi(B^s)y_t = \theta(B)\Theta(B^s)u_t \quad (2.2.2)$$

Where y_t is the value of the price series at time t, u_t is the disturbance term at time t which is assumed to be IID with mean zero and constant variance σ^2, the backshift operator B is defined by $(B)y_t = y_{t-1}$, $\Delta = (1-B)$ the differencing operator, $(\Delta_s)^D = (1-B^s)^D$ is seasonal differencing operator. The polynomials are defined as: $\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_p B^p$ (autoregressive), $\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q$ (moving average), $\Phi(B^S) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \cdots - \Phi_P B^{Ps}$ (seasonal autoregressive) and $\Theta(B^s) = 1 + \Theta_1 B^s + \Theta_2 B^{2s} + \cdots + \Theta_Q B^{Qs}$ (seasonal moving average). Where the degrees of the polynomials are p, q, in B, and P, Q for B^s respectively.

2.3 Time delay neural networks (TDNNs)

ANNs are human brain-inspired models that capture time series characteristics using data-driven, nonlinear, and non-parametric methods, unlike traditional forecasting methods. Neural networks consist of three interconnected neuron layers, where each layer receives input from the previous one and sends output to the next (Singh, 2021) [38]. Nonlinear activation functions are applied at hidden nodes to transform weighted inputs, introduce nonlinearity, and regulate output values. During training, the network learns optimal weights and biases stored in its nodes, which determine the mapping between inputs and outputs. The number of input neurons corresponds to the number of lagged values used as predictors, while hidden nodes process their weighted sums through nonlinear transformations (Li *et al.*, 2010) [26]. Mathematically, forward propagation is expressed as:

$$y_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j\, g\left(\beta_{0j} + \sum_{i=1}^{p} \beta_{ij}\, y_{t-i}\right) + \varepsilon_t \quad (2.3.1)$$

Where p and q are the number of input layers and hidden nodes, $y_{t-1}, y_{t-2}, y_{t-3} \ldots, y_{t-p}$ are the input patterns, β_{ij} is the synaptic weight between the i^th^ input neuron and j^th^ hidden neuron, α_j is the weight between the j^th^ hidden neuron and the output neuron, β_{0j} and α_0 is the bias, $g(.)$ are the activation functions of hidden nodes, respectively (Ahmed, 2025) [3].

2.4 Support Vector Regression (SVR)

SVR introduced by Vapnik (1998) [43], extends SVM to regression tasks using a loss function. It's working process is shown in figure 1, and the basic linear SVR model is defined as:

$$y = w^T \phi(x) + b \tag{2.4.1}$$

Where w defines the weight vector, ϕ denotes mapping function, and b is bias, these parameters are estimated by minimizing a regularized risk function:

$$R(\theta) = \frac{1}{2}\|w\|^2 + c\left[\frac{1}{N}\sum_{i=1}^{N} L_\varepsilon\big(y_i, f(x_i)\big)\right] \tag{2.4.2}$$

Here, $\frac{1}{2}\|w\|^2$ controls model complexity, $\left[\frac{1}{N}\sum_{i=1}^{N} L_\varepsilon\big(y_i, f(x_i)\big)\right]$ called 'empirical error' is estimated by Vapnik ε-insensitive loss function:

$$L_\varepsilon\big(y_i, f(x_i)\big) = \begin{cases} |y_i - f(x_i)| - \varepsilon, & |y_i - f(x_i)| \geq \varepsilon \\ 0 & otherwise \end{cases} \tag{2.4.3}$$

Here, both c and ε are user-determined hyper-parameters, ε defines a tolerance margin around the regression function, while c is a regularization constant balancing flatness and error tolerance (Singla *et al.*, 2021) [39].

To allow some training errors, slack variables (ξ_i, ξ_i^*) are introduced, leading to the primal optimization problem:

$$Minimize: R_p(\mathrm{w}, \mathrm{b}, \xi_i, \xi_i^*) = \frac{1}{2}\|w\|^2 + C\left[\sum_{i=1}^{N}(\xi_i + \xi_i^*)\right] \tag{2.4.4}$$

Subject to the constraints

$$\begin{cases} w^T\varphi(x_i) + b - y_i \leq \varepsilon + \xi_i^* \\ y_i - w^T\varphi(x_i) - b \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0;\ \mathrm{i} = 1,2, \ldots, \mathrm{N} \end{cases}$$

This concept known as soft margin regression. For nonlinear problems, SVR employs the *kernel trick*, replacing the dot product in feature space with a kernel function $\mathrm{K}(x_i, x_j) = \langle\phi(x_i), \phi(x_j)\rangle$. This enables efficient learning of nonlinear relationships without explicitly mapping to high dimensions. Among available kernels, the Radial Basis Function (RBF) is widely used for time series and price forecasting tasks (Das *et al.*, 2019) [11].

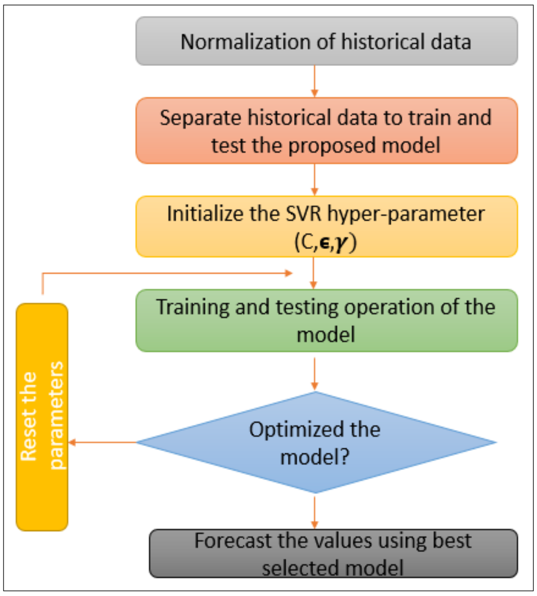

Fig 1: Working procedure of SVR Model

2.5 Empirical mode decomposition (EMD)

EMD is a type of adaptive time series decomposition approach used for nonlinear and non-stationary time series data. The method decomposes a series into a finite set of Intrinsic Mode Functions (IMFs) and a final residue (Huang *et al.*, 1998) [20]. Each IMF represents an oscillatory mode with unique amplitude and frequency modulation, and it must satisfy two conditions: (i) the number of extrema and zero-crossings should be nearly equal, and (ii) the mean of the envelopes defined by local maxima and minima should be zero (Sonam and Kumar, 2017) [40]. The original series can then be reconstructed as:

$$y(t) = \sum_{i=1}^{n} h_i(t) + r(t) \tag{2.5.1}$$

Steps of EMD

1. Identify all local maxima and minima of the series.
2. Interpolate maxima and minima using cubic splines to form the upper and lower envelopes.
3. Compute the local mean

$$m_1(t) = [y_{max}(t) + y_{min}(t)]/2 \tag{2.5.2}$$

4. Subtract the mean from the original series

$$h_1(t) = y(t) - m_1(t) \tag{2.5.3}$$

5. If $h_1(t)$ satisfies the IMF conditions, it is considered as the first IMF. If not, steps 1 to 4 are repeated by treating $h_1(t)$ as the new input, until the remainder becomes a monotonic function and no more IMF can be extracted.

The possible number of IMFs that can be extracted from a time series is approximately log_2N, Where N denotes the length of the series (Wu and Huang, 2009) [46]. While EMD effectively captures hidden oscillatory patterns in non-stationary agricultural price data, a major limitation is the mode mixing problem, where signals of different scales appear in a single IMF. This reduces the interpretability of the decomposition (Sezen, 2023) [35].

2.6 Ensemble empirical mode decomposition (EEMD)

Wu and Huang (2009) [46] introduced EEMD as an improved version of EMD to overcome the issue of mode mixing. In EEMD, white noise is repeatedly added to the original series, decomposed with EMD, and then averaged across trials to extract the true IMFs and residue. The working process is illustrated in Figure 2, and the final decomposition is represented as:

$$y(t)=\sum_{j=1}^{n} h_j\ (t)+ r(t) \qquad (2.6.1)$$

Where $h_j(t)$, j=1,2…..n are the final IMFs and r(t) is the residue.

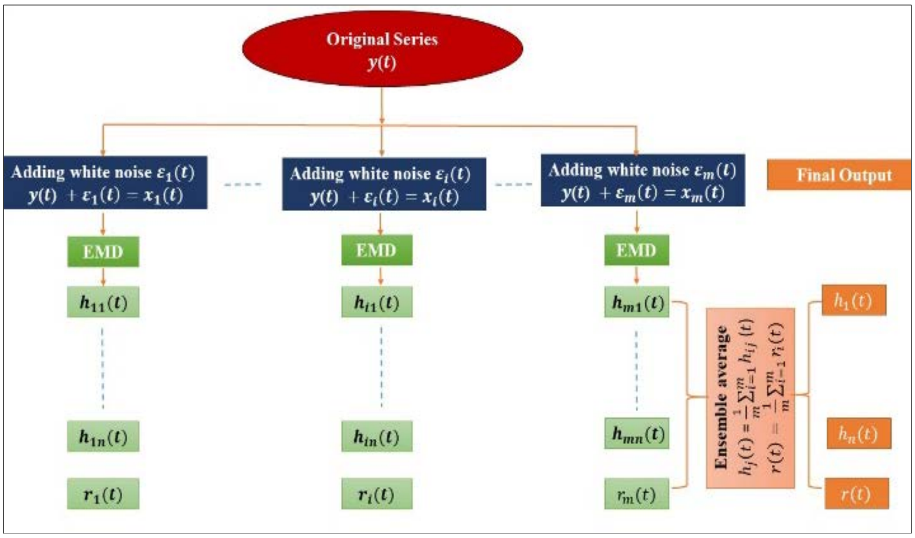

Fig 2: EEMD Algorithm Working procedure

3. Results and Discussion

In order to predict the prices of export onion, the dataset was pre-processed, a few missing values were imputed using the mean imputation method in R with the *imputeTS* package to ensure accuracy and consistency. Descriptive statistics (Table 1) revealed large fluctuations, with prices ranging from ₹271.57 to ₹11,405.45 per quintal and a high coefficient of variation (72.16%), reflecting substantial instability. The time plot (Figure 3) clearly revealed the presence of trend and volatility, while the seasonal boxplot (Figure 4) confirmed seasonality, and the ACF and PACF plots (Figure 5) indicated non-stationarity. Formal statistical tests further supported these findings: Shapiro-Wilk test (p< 0.01) Rejected normality, the Ljung-Box test (p< 0.01) confirmed significant autocorrelation, the KPSS test (Table 2) Indicated non-stationarity in original series but stationarity after first differencing, and the BDS test (Table 3) Confirmed the presence of nonlinearity (p< 0.01 across all dimensions). Collectively, these results confirm the data's complexity and justify the use of advanced forecasting methods.

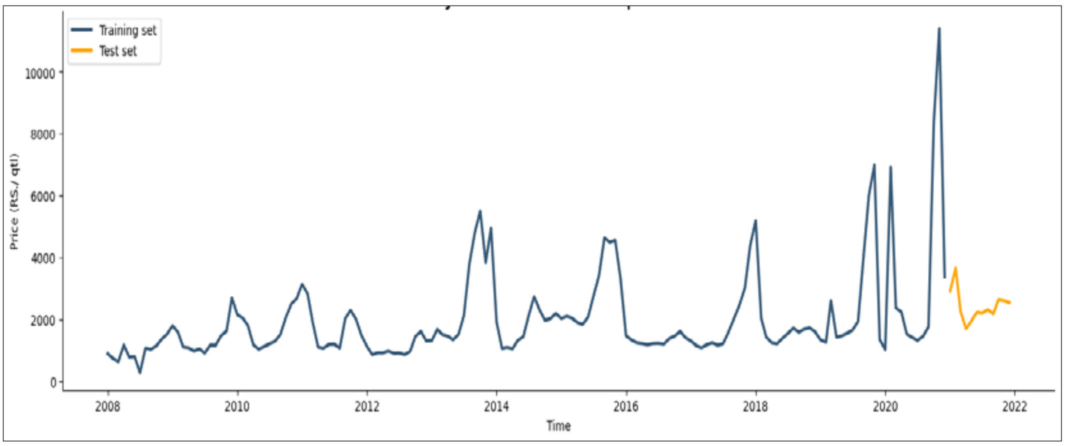

Fig 3: Monthly onion export price series

Table 1: Descriptive Statistics of Monthly Onion Export Prices (₹/Qtl)

Statistical Measures	Export
Mean	2005.583
Minimum	271.572
Maximum	11405.453
Standard Deviation	1447.274
Skewness	3.123
Kurtosis	13.331
Coefficient of variation	72.166
Shapiro-Wilk test (w)	0.682**
Ljung-Box test	63.621**

Note: ** significant at the 1% level of significance

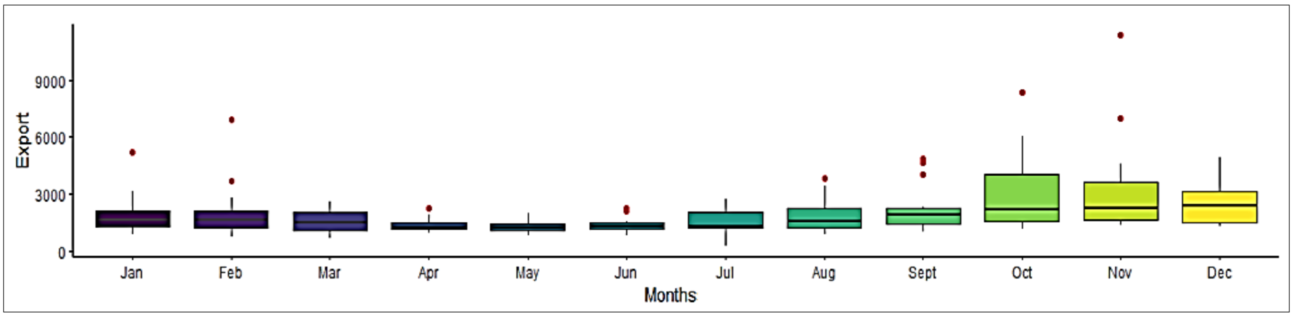

Fig 4: Seasonal box plot of onion prices in the export market

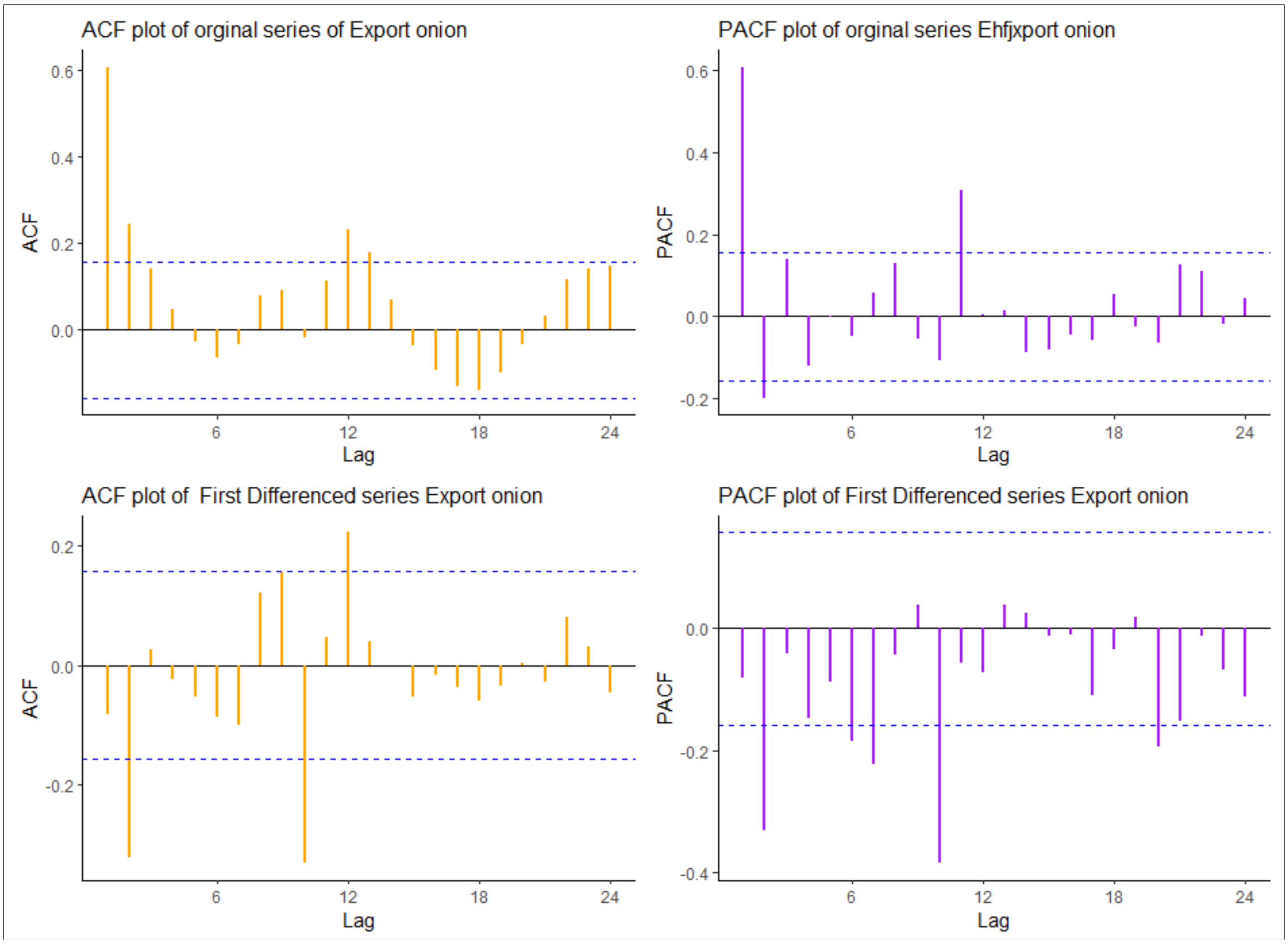

Fig 5: ACF and PACF plots of the original and transformed onion export price series

Table 2: KPSS test results for onion Price in Export Market

Nature of series	KPSS level	Truncation lag parameter	p- value	Conclusion
Original series	0.782	4	0.01	Non-stationarity
After 1st diff	0.012	4	0.1	Stationarity

Table 3: BDS test results for testing linearity of data series

Parameter	Dimension (m=2) Statistic	Dimension (m=2) Probability	Dimension (m=3) Statistic	Dimension (m=3) Probability	Conclusion
723.635	0.203	< 0.01	16.353	< 0.01	Non linear
1447.270	10.540	< 0.01	10.167	< 0.01	
2170.904	8.954	< 0.01	8.552	< 0.01	
2894.539	6.589	< 0.01	6.379	< 0.01	

Following the identification of nature of date series, several traditional forecasting models were applied, including ARIMA, TDNN and SVR. Their predictive performance on the test dataset is reported in Table 4. An ARIMA (1,1,1)(0,0,2) [12] model was selected as optimal based on its low AIC value, and its residuals passed diagnostic checks for white noise. However, consistent with the BDS test results, its linear structure failed to capture the dataset nonlinear dynamics. Subsequently, machine learning models (ANN and SVM) were used to address this limitation.

After conducting several iterations with different nodes, the optimal ANN architecture was determined to be (12: 6S:1L), consisting of 12 input nodes, 6 hidden nodes with Tanh activation function, and 1 output node, based on their holdout forecasting performance. The RMSE of the training set (213.03) indicated strong in-sample performance, whereas the substantially larger RMSE of the test set (1643.08) confirmed overfitting (Table 5). In contrast, SVR, tuned via 10-fold cross-validation including an RBF kernel with the hyper parameters (Cost = 1, Gamma = 0.1, Epsilon = 0.1), achieved better robustness and higher predictive accuracy than TDNN, as shown in Figures 9 and 10.

To enhance prediction accuracy, hybrid models using EMD and EEMD decomposition were applied. The series was decomposed into six IMFs and one residual (Figures 6-7), where amplitudes decreased from IMF1 to IMF6 and the residual captured the random trend. This decomposition revealed hidden patterns, improving the forecasting performance of ARIMA, TDNN, and SVR (Das *et al.*, 2023) [12]. Each decomposed dataset was modeled as in the single-model approach, and final fitted results are then obtained following the procedure outlined in section 2.5 and 2.6 of the methodology. The results (Table 4) indicated that the EEMD-SVR model produced forecasts closer to the test data than the other models, as illustrated in Figure 8, effectively capturing both linear and nonlinear dynamics.

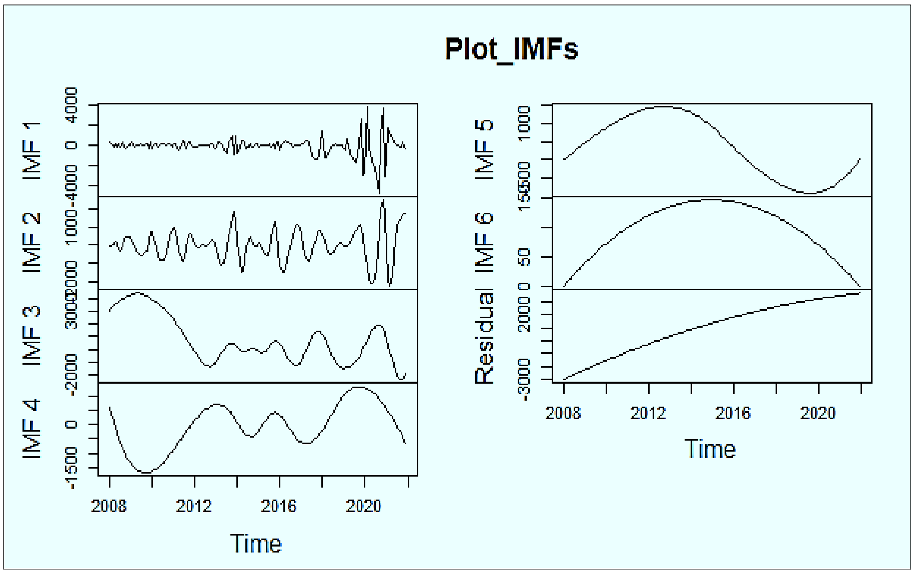

Fig 6: EMD components of monthly export onion price

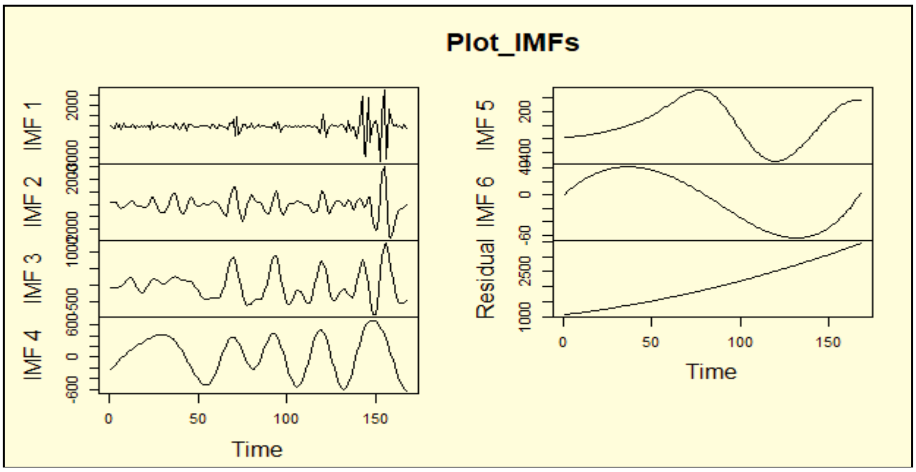

Fig 7: EEMD components of monthly export onion price

Table 4: Forecasted values of the export onion prices using different models

Time	Test	ARIMA	EMD-ARIMA	EEMD-ARIMA	TDNN	EMD-TDNN	EEMD-TDNN	SVR	EMD-SVR	EEMD-SVR
Jan-21	2912.772	2135.977	2215.123	2635.244	3108.443	2272.953	2391.306	3893.908	2034.806	3660.807
Feb-21	3650.996	6125.156	2752.232	3742.896	5223.323	2306.568	1555.354	3071.966	660.1495	3209.602
Mar-21	2248.038	3020.415	3439.036	3078.639	1598.3	2967.874	1363.932	4333.064	1295.059	2535.957
Apr-21	1694.196	2845.153	2788.226	3264.777	1380.299	3243.568	2147.729	2229.542	1652.97	2421.013
May-21	1953.881	2307.903	3008.195	2657.269	1471.921	4373.173	2788.795	1715.816	2042.65	2341.839
Jun-21	2228.614	2101.044	3729.831	3483.767	2895.819	5995.593	3739.343	1957.557	1398.867	2411.726
Jul-21	2202.52	1952.583	3949.5	3817.076	2866.374	7450.691	3851.552	2210.828	1863.424	2401.206
Aug-21	2317.634	2008.745	3787.77	3963.268	1820.009	3777.488	2929.447	2185.962	2103.546	2467.861
Sep-21	2176.981	2157.485	3673.601	3758.823	2746.535	2076.5	2383.603	2298.292	2260.75	2469.529
Oct-21	2637.548	6379.426	3610.843	3543.069	6265.09	2970.772	3287.193	2161.893	2404.513	2662.228
Nov-21	2604.104	8374.231	3471.276	3905.936	6381.878	1554.735	3265.069	2665.853	2494.247	2864.872
Dec-21	2538.223	3664.267	3290.19	3562.891	3033.048	1389.186	2619.418	2622.499	3089.186	2989.186

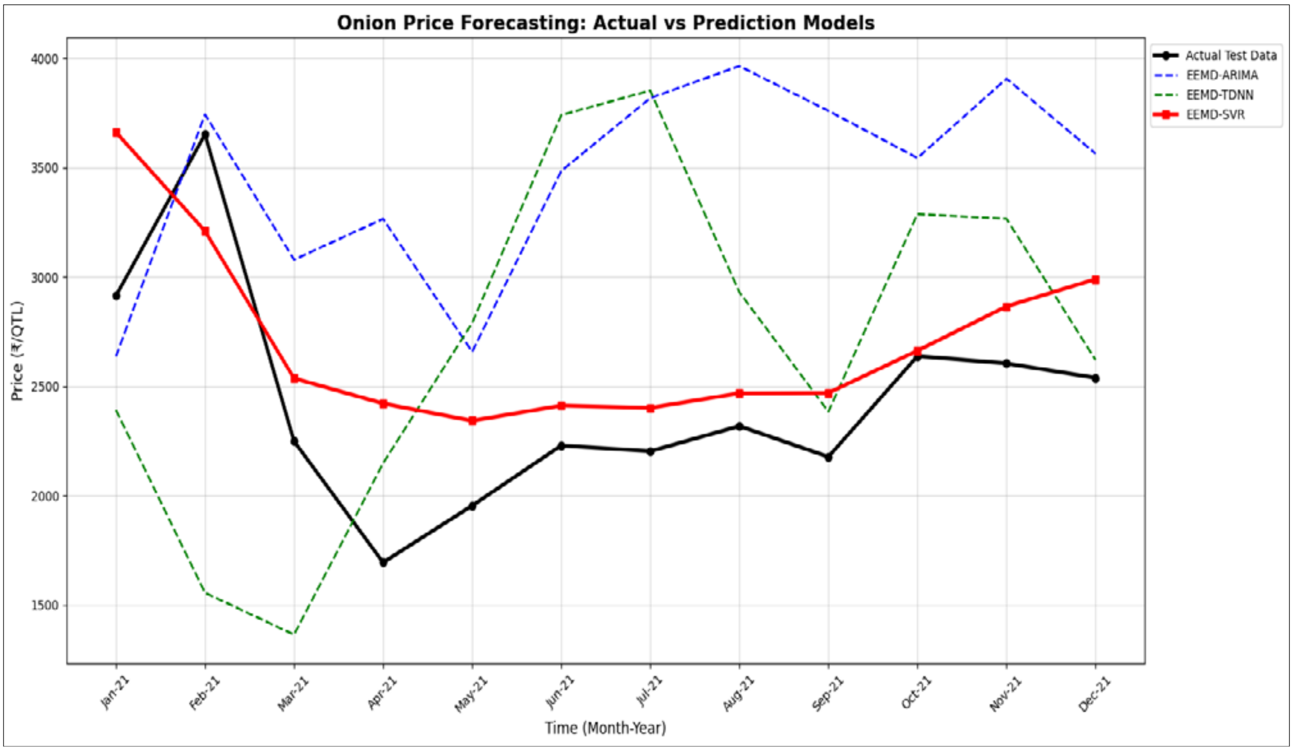

Fig 8: Forecasted values of the export onion prices using different models

The predictive performance of nine models for forecasting onion export prices was evaluated using RMSE and MAPE metrics (Table 5). The results indicate that hybrid machine learning models consistently outperformed traditional benchmarks, with the EEMD-SVR model achieving the highest accuracy. Its superiority is visually confirmed by radar plots, which show consistently lower RMSE (405.09) and MAPE (14.93%) values compared to other models (Figures 9-10). To further assess the superiority of EEMD-SVR, The Diebold-Mariano test was conducted under the alternative hypothesis that EEMD-SVR outperforms the other hybrid models in forecasting accuracy. The p-values reported in Table 6 confirm that EEMD-SVR was significantly more accurate than all other hybrids. The enhanced performance of EEMD-SVR can be attributed to the decomposition process, which isolates and identifies distinct features of price volatility, improving the forecasting process.

Table 5: Comparative assessment of prediction performance of different models

Set	TEST	ARIMA	EMD-ARIMA	EEMD-ARIMA	TDNN	EMD-TDNN	EEMD-TDNN	SVR	EMD-SVR	EEMD-SVR
Train	RMSE	1078.71	953.31	846.36	213.03	181.77	172.25	882.12	718.34	582.36
	MAPE (%)	28.54	27.23	26.96	8.32	8.44	8.36	14.85	21.90	16.25
Test	RMSE	2479.80	1363.71	1179.47	1643.08	1266.35	1026.00	726.36	994.14	405.09
	MAPE (%)	70.06	54.036	47.97	44.30	47.45	34.65	19.46	21.82	14.93

Table 6: Diebold Marino test results EEMD-SVR vs other hybrid models

Test comparison		EMD-ARIMA	EEMD-ARIMA	EMD-TDNN	EEMD-TDNN	EMD-SVR
EEMD-SVR	Test statistic	4.681	3.987	2.188	1.942	1.532
	p value	0.0003	0.001	0.039	0.041	0.053

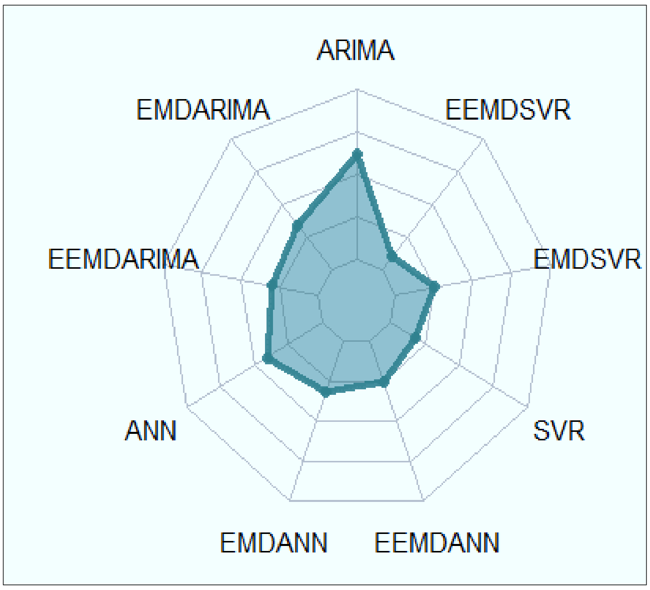

Fig 9: RMSE values for test data of export onion prices

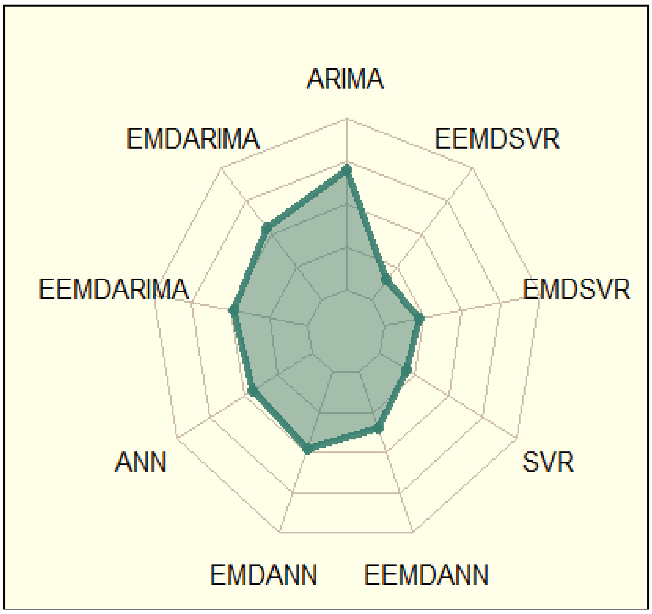

Fig 10: MAPE values for test data of export onion prices

4. Conclusion

Farmers and traders can reduce losses with advance price information, which is especially important for perishable crops like tomatoes, onions, and potatoes. In this study, we forecast onion export prices using hybrid machine learning and classical models, focusing on handling both nonstationary and nonlinear data. We used decomposition techniques, EMD and EEMD, to break the series into simpler components, predicted them with ARIMA, TDNN, and SVR, and combined the results for the final forecast. EEMD improved prediction accuracy compared to EMD. Some small errors may occur due to external factors like weather. Accurate forecasts help farmers make better decisions on production and marketing, and assist the government in planning import and export policies.

Acknowledgements

I am indeed grateful to the Editors for their guidance and counsel. I am very grateful to the reviewer for valuable comments and suggestions of generously listing many useful references.

Conflict of interest

The authors do not have any financial or non-financial conflict of interest to declare for the research work included in this article.

References

1. Abadan S, Shabri A. Hybrid empirical mode decomposition-ARIMA for forecasting price of rice. Applied Mathematical Sciences. 2014;8(63):3133-3143.
2. Agbo HMS. Forecasting agricultural price volatility of some export crops in Egypt using ARIMA/GARCH model. Review of Economics and Political Science. 2023;8(2):123-133.
3. Ahmed I. A machine learning approach to short-term price forecasting of agricultural crops in India. In: 2025 First International Conference on Advances in Computer Science, Electrical, Electronics, and Communication Technologies (CE2CT). IEEE; 2025 Feb. p. 325-331.
4. Ajmal S, Rohith S, Unniravisankar P, Nabay A. Price dynamics of tomato, onion and potato (TOP) in India. Asian Journal of Agricultural Extension, Economics & Sociology. 2024;42:134-143.
5. Agricultural and Processed Food Products Export Development Authority. Fresh onions: export volume in FY23-24. New Delhi: Agricultural and Processed Food Products Export Development Authority; 2024.
6. Biswas P. Explained: What is driving onion prices. Indian Express. 2019 Nov 18.
7. Box GE. Science and statistics. Journal of the American Statistical Association. 1976;71(356):791-799.
8. Brignoli PL, Varacca A, Gardebroek C, Sckokai P. Machine learning to predict grains futures prices. Agricultural Economics. 2024;55(3):479-497.
9. Chen CF, Lai M, Yeh CC. Forecasting tourism demand based on empirical mode decomposition. Knowledge-Based Systems. 2012;26:281-287.
10. Choudhary K, Jha GK, Kumar RR, Mishra DC. Agricultural commodity price analysis using ensemble empirical mode decomposition: A case study of daily potato price series. Indian Journal of Agricultural Sciences. 2019;89(5):882-886.
11. Das P. Study on machine learning techniques based hybrid model for forecasting in agriculture. 2019.
12. Das P, Jha GK, Lama A. Empirical mode decomposition based ensemble hybrid machine learning models for agricultural commodity price forecasting. Statistics Applications. 2023;99:1-??.
13. Das P, Jha GK, Lama A, Parsad R, Mishra D. Empirical mode decomposition based support vector regression for agricultural price forecasting. Indian Journal of Extension Education. 2020;56(2):7-12.
14. Debopam R, Paul RK, Panwar S. Asymmetric price volatility of onion in India. Indian Journal of Agricultural Economics. 2021;76(2):245-260.
15. Dhotre M, Nithin KN, Kolluru R, Desai S. Recurring onion and tomato crises in India: A critical analysis and future perspectives. In: Emerging Trends in Food and Agribusiness Marketing. Hershey (PA): IGI Global; 2025. p. 123-160.
16. Dohlman E, Hansen J, Chambers W, Interagency Agricultural Projections Committee. USDA agricultural projections to 2034. Washington (DC): United States Department of Agriculture; 2025.

17. GH HN. A comparative analysis of time series models for onion price forecasting: Insights for agricultural economics. Journal of Experimental Agriculture International. 2024;46(5):146-154.
18. Ghosh S, Datta D, Majumdar B, Mazumdar SP, Soumia PS, Thangasamy A. Onion price fluctuation in Indian market: Analysis and mitigation strategies. Vigyan Varta International E-Magazine of Science Enthusiasts. 2022;3(4):95-98.
19. Guo Z, Zhao W, Lu H, Wang J. Multi-step forecasting for wind speed using a modified EMD-based artificial neural network model. Renewable Energy. 2012;37(1):241-249.
20. Huang NE, Shen Z, Long SR, Wu ML, Shih HH, Zheng Q, et al. The empirical mode decomposition and Hilbert spectrum for nonlinear and non-stationary time series analysis. Proceedings of the Royal Society of London Series A. 1998;454:909-995.
21. Imran M, Kang H, Lee SG, Kim EH, Park HM, Oh SW. Current trends and future prospects in onion production, supply, and demand in South Korea: A comprehensive review. Sustainability. 2025;17(3):1-??.
22. Ince H, Trafalis T. A hybrid model for exchange rate prediction. Decision Support Systems. 2006;42(2):1054-1062.
23. Khan DM, Ali M, Iqbal N, Khalil U, Aljohani HM, Alharthi AS, et al. Short-term prediction of COVID-19 using novel hybrid ensemble empirical mode decomposition and error trend seasonal model. Frontiers in Public Health. 2022;10:922795.
24. Kim D, Oh HS. EMD: a package for empirical mode decomposition and Hilbert spectrum. The R Journal. 2009;1(1):40-42.
25. Kumar P, Badal PS, Paul RK, Jha GK, Venkatesh P, Kingsley IT, et al. Empowering farmers through future price information: A case of price forecasting of brinjal in eastern Uttar Pradesh. Indian Journal of Economics and Development. 2020;16(4):479-488.
26. Li G, Xu S, Li G. Short-term price forecasting for agro-products using artificial neural networks. Agriculture and Agricultural Science Procedia. 2010;3(1):278-287.
27. Mao S, Soonthornphisaj N. Thailand’s maize prices forecasting using ensemble technique. ASEAN Journal of Scientific and Technological Reports. 2024;27(4):e252279.
28. Mohanty MK, Thakurta PKG, Kar S. Agricultural commodity price prediction model: a machine learning framework. Neural Computing and Applications. 2023;35(20):15109-15128.
29. Patil A, Shah D, Shah A, Kotecha R. Forecasting prices of agricultural commodities using machine learning for global food security: towards sustainable development goal 2. International Journal of Engineering Trends and Technology. 2023;71:277-291.
30. Purohit SK, Panigrahi S, Sethy PK, Behera SK. Time series forecasting of price of agricultural products using hybrid methods. Applied Artificial Intelligence. 2021;35(15):1388-1406.
31. Raka S, Ramesh C. Understanding the recurring onion price shocks: revelations from production-trade-price linkages. Policy Paper - National Centre for Agricultural Economics and Policy Research. 2017;6(33):56-59.
32. Sahu PK, Das M, Sarkar B, VS A, Dey S, Narasimhaiah L, et al. Potato production in India: a critical appraisal on sustainability, forecasting, price and export behaviour. Potato Research. 2024;67(4):1209-1245.
33. Satish DP. Onion price hits rock bottom, sells at Rs 1 per kilo in Karnataka. News18. 2018 Nov 21.
34. Saxena R, Birthal PS, Agrawal RC, Sharma P, Paul BS, Pant DK, et al. From local to global opportunities to accelerate agricultural exports from India. New Delhi; 2024.
35. Sezen C. Pan evaporation forecasting using empirical and ensemble empirical mode decomposition (EEMD) based data-driven models in the Euphrates sub-basin, Turkey. Earth Science Informatics. 2023;16(4):3077-3095.
36. Shobharani HM, Gupta AK, Kumar S, Tripathi G. Machine learning and hybrid model building mechanism for forecasting agricultural market behaviour: a case study of vegetable crops. Indian Journal of Economics and Development. 2024;20:303-316.
37. Silva RF, Barreira BL, Cugnasca CE. Prediction of corn and sugar prices using machine learning, econometrics, and ensemble models. Engineering Proceedings. 2021;9:31.
38. Singh A. Comparison of artificial neural networks and statistical methods for forecasting prices of different edible oils in Indian markets. International Research Journal of Modernization in Engineering Technology and Science. 2021;3(10):1044-1050.
39. Singla S, Paul RK, Shekhar S. Modelling price volatility in onion using wavelet-based hybrid models. Indian Journal of Economics and Development. 2021;17(2):256-265.
40. Sonam M, Kumar A. Empirical mode decomposition: theory and applications. International Journal of Electronic and Electrical Engineering. 2017;7(8):873-878.
41. Sunil J, Sindhu D. Growth status and ARIMA model forecast of area, production of wheat crop in India. Journal of Scientific Research and Reports. 2020;26(5):117-125.
42. Theofilou A, Nastis SA, Michailidis A, Bournaris T, Mattas K. Predicting prices of staple crops using machine learning: a systematic review of studies on wheat, corn, and rice. Sustainability. 2025;17(12):5456.
43. Vapnik V. The support vector method of function estimation. In: Nonlinear modeling: advanced black-box techniques. 1998. p. 55-85.
44. Wang G, Xu S, Chen Z, Li Y. A hybrid model integrating variational mode decomposition and intelligent optimization for vegetable price prediction. Agriculture. 2025;15(9):919.
45. Wang L, Feng J, Sui X, Chu X, Mu W. Agricultural product price forecasting methods: research advances and trend. British Food Journal. 2020;118(5):1129-1145.
46. Wu Z, Huang NE. Ensemble empirical mode decomposition: a noise-assisted data analysis method. Advances in Adaptive Data Analysis. 2009;1(1):1-41.
47. Xu X, Zhang Y. Corn cash price forecasting with neural networks. Computers and Electronics in Agriculture. 2021;184:106120.

48. Xu X, Zhang Y. Commodity price forecasting via neural networks for coffee, corn, cotton, oats, soybeans, soybean oil, sugar, and wheat. Intelligent Systems in Accounting, Finance and Management. 2022;29(3):169-181.
49. Yadav A. A comparative study of time series, machine learning, and deep learning models for forecasting global price of wheat. In: Operations Research Forum. Cham: Springer International Publishing; 2024 Nov. Vol 5(4):113.
50. Yerukala S, Madala P, Battula V, Enduri MK, Tokala S, Bisiringi NP. Estimating future prices of key agricultural commodities using machine learning models. In: 2024 Beyond Technology Summit on Informatics International Conference (BTS-I2C). IEEE; 2024 Dec. p. 572-576.
51. Zelingher R, Makowski D. Investigating and forecasting the impact of crop production shocks on global commodity prices. Environmental Research Letters. 2024;19(1):014026.
52. Zhang GP. Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing. 2003;50(2):159-175.
53. Zhang N, An Q, Zhang S, Ma H. Price prediction for fresh agricultural products based on a boosting ensemble algorithm. Mathematics. 2025;13(1):71.