PLANT SCIENCE TODAY
Vol x(x): xx–xx
https://doi.org/10.14719/pst.8558

eISSN 2348-1900

RESEARCH ARTICLE

Wavelet enhanced time series forecasting for tomato price volatility in South Indian markets

V Lavanya[1], R Gangai Selvi[1*], Patil Santosh Ganapati[1], J Auxcilia[2], V Karthick[3] & S Vishnu Shankar[1]

[1]Department of Physical Sciences and Information Technology, Agricultural Engineering College and Research Institute, Tamil Nadu, Agricultural University, Coimbatore 641 003, Tamil Nadu, India

[2]Department of Fruit Science, Horticultural College & Research Institute, Tamil Nadu Agricultural University, Coimbatore 641 003, Tamil Nadu, India

[3]Department of Agricultural Economics, Tamil Nadu Agricultural University, Coimbatore 641 003, Tamil Nadu, India

*Correspondence email - gangastat@tnau.ac.in

Received: 28 March 2025; Accepted: 28 April 2025; Available online: Version 1.0: 17 May 2025

Cite this article: Lavanya V, Gangai Selvi R, Patil SG, Auxcilia J , Karthick V, Vishnu SS. Wavelet enhanced time series forecasting for tomato price volatility in South Indian markets . Plant Science Today (Early Access). https:/doi.org/10.14719/pst.8558

Abstract

Accurate price forecasting is crucial in the agricultural sector, where farmers face significant challenges due to price volatility. Price fluctuations directly influence the livelihoods of producers and the affordability for consumers, making crop management difficult for farmers. Agricultural time series data are often highly complex and nonlinear, making price prediction, a challenging task. While various forecasting approaches, including stochastic models, machine-learning techniques and hybrid models, have been explored, their effectiveness is often limited due to the inherent complexity of agricultural datasets. Recently, wavelet-enhanced models have emerged as a robust approach, effectively capturing both short-term fluctuations and long-term trends. Wavelet decomposition plays a vital role in denoising data and extracting inherent patterns, thereby improving predictive accuracy. This study investigates the application of wavelet -based models for forecasting the monthly wholesale tomato prices in key South Indian markets such as Bangalore, Chennai and Trivandrum. The findings address the forecasting challenges posed by the volatility of tomato prices, providing valuable insights for stakeholders, including farmers, traders and policymakers, to facilitate informed decision-making. Further, the study highlights the necessity of a robust price policy to stabilize market fluctuations, safeguard farmers' livelihoods and ensure fair returns. Hence, incorporating advanced forecasting techniques, such as wavelet-based models can significantly improve market stability and promote sustainable agricultural development.

Keywords: denoising; market volatility; time series forecasting; tomato; wavelet decomposition

Introduction

Price forecasting is crucial for farmers, stakeholders, consumers and policymakers, facilitating informed market planning and strategic decision-making (1). Farmers face substantial challenges due to unpredictable price fluctuations, which complicate the effective marketing of their produce (2). The lack of accurate price forecasting often leads to financial instability and increased post-harvest losses. Therefore, precise forecasting models are crucial to mitigate these challenges to enhance the market stability (3). Time series modelling plays a fundamental role in forecasting by identifying the hidden patterns and structures within the data (4). The efficiency of a forecasting method primarily depends on the characteristics and nature of the dataset (5). Advancements in statistical and machine learning techniques have significantly enhanced price forecasting accuracy across diverse domains such as finance, agriculture, economics and business (6). The wavelet decomposition method stands out as a powerful tool in analyzing complex and volatile time series data (7). Unlike traditional statistical models, wavelet-based techniques capture the patterns in both time and frequency dimensions. By decomposing the data into multiple resolution levels, wavelet-based models enhance the predictive capability of traditional forecasting techniques, making them particularly useful for agricultural price forecasting (8). Vegetables being highly perishable and seasonal exhibit significant price volatility (9). Tomato is a highly volatile crop and one of the most widely consumed staple foods globally, making their price stability crucial for economic security (10). In India, tomato production is concentrated in key states such as Andhra Pradesh, Madhya Pradesh, Karnataka, Gujarat, Odisha West Bengal (11). However, its supply chain faces challenges related to limited storage infrastructure, transportation inefficiencies and post-harvest losses. These factors lead to unstable market conditions, affecting producers and consumers (12). The fluctuating prices of tomato often create economic disruptions at the household level. Despite its significance, the tomato price fluctuation is highly susceptible to external factors such as climate variability, natural calamities and market dynamics (13). The perishable nature of tomatoes further contributes to price volatility, making it imperative to predict market trends to help farmers make strategic harvesting and selling decisions (14). Hence, accurate forecasting models are essential for optimizing market strategies, reducing uncertainty and supporting informed decision-making (15).

Recent studies have explored the diverse methodologies for price forecasting, highlighting the effectiveness of various models. A study on maize price forecasting in Ghana using various time series models such as Single Exponential Smoothing (SES), Double Exponential Smoothing (DES), Triple Exponential Smoothing (TES), Autoregressive Integrated Moving Average (ARIMA) and Seasonal ARIMA (SARIMA) found ARIMA to be the most effective model based on MAE, offering valuable insights for agricultural planning and decision-making (16). The effectiveness of ARIMA models for precise price forecasting in key potato-producing regions of India had been examined (17). The analysis of potato price volatility in India using ARIMA, SARIMA, Artificial Neural Network (ANN) and Long Short-Term Memory (LSTM) models identified LSTM as the most accurate model (18). Neural network modelling had been explored for long-term daily price forecasting of various agricultural commodities (19). A comparative analysis of ARIMA and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models for crude oil price forecasting provided valuable insights of their performance (20). An ARMA-GARCH model had been used to analyse Korean green onion price fluctuations, to estimate volatility and forecast prices, highlighting its effectiveness in modelling seasonally adjusted prices (21). A trivariate ARMA GARCH Vine Copula model for forecasting vegetable price volatility of tomato, onion and potato across three Indian markets demonstrated its superior accuracy over traditional models (22). Wavelet-based denoising techniques enhanced the accuracy of predictive models by effectively filtering out noise while preserving essential signal features (6). A hybrid Wavelet-ARIMA-LSTM model for forecasting share price index futures, demonstrated that wavelet decomposition improves prediction accuracy, with ARIMA effectively capturing stable signals and LSTM handling noisy data (23). The integration of wavelet decomposition with stochastic and machine learning models improved agricultural price prediction accuracy by effectively handling nonlinearity and non-normality in price data (4). Wavelet-based hybrid model combining wavelet decomposition with ANN outperformed ARIMA, GARCH and ANN models in forecasting tomato prices (8). Wavelet ARIMA-ANN model outperformed individual ANN and Wavelet ANN models in forecasting meteorological drought (24).

The main contribution of this study lies in the application of advanced wavelet techniques that combine traditional and machine learning models to enhance the accuracy of tomato price forecasting. While most existing studies have focused on major markets in Northern India, this study is the first to apply wavelet decomposition techniques specifically to the Southern Indian markets. The primary objective is to assess the effectiveness of wavelet-enhanced models, demonstrating their superiority over benchmark models. By utilizing wavelet techniques, this study offers valuable insights of the complex price dynamics of tomatoes in key South Indian markets, benefiting farmers, traders and policymakers. Further, the study highlights the need for well-structured price policy to stabilize market fluctuations and ensure fair returns for farmers. These findings emphasize the importance of adopting advanced forecasting methods to address the unique challenges of the agricultural sector, especially for highly volatile commodities like tomatoes. This study highlights the effectiveness of wavelet enhanced models in strengthening predictive accuracy and facilitating informed decision-making in agricultural markets.

Material and Methods

Data description

This study analyses the monthly wholesale prices of tomatoes in key South Indian markets such as Bangalore, Chennai and Trivandrum. The dataset, covering the period from January 2007 to December 2023, was collected from Indiastat. It comprises 204 monthly observations, with 184 used for model training and the remaining 20 reserved for testing purposes. This results in a 90:10 training-to-testing split, ensuring a robust assessment of model accuracy. The workflow of the study is depicted in Fig. 1.

Autoregressive Integrated Moving Average (ARIMA)

The ARIMA model, also referred to as the Box-Jenkins model, is a widely recognized tool for analysing univariate time series data (25). By incorporating both autoregressive (AR) and moving average (MA) components, ARIMA is particularly effective at identifying linear patterns within the data (26). To determine the optimal values for the model's parameters, the Auto-Correlation Function (ACF) and Partial Auto-Correlation Function (PACF) are used. To achieve stationarity, differencing is often applied, which helps stabilize the mean and variance over time (27). The ARIMA model is characterized by three parameters namely p (autoregressive lag), d (the degree of differencing) and q (moving average lag) (28). The general form of the ARMA (p, q) model is given as:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots \theta_q \varepsilon_{t-q}$$

The ARIMA (p, q, d) model is obtained by incorporating a differencing term, expressed as: $\phi(L)\Delta^d y_t = \theta(L)\varepsilon_t$

where y_t represents the actual value of the time series at time t, ϕ_i denotes the autoregressive component of order p, θ_i represents the moving average component of order q, d is the differencing term, ε_t is the random error at time t, $\phi(L)$ and $\phi(L)$ represent the AR and MA polynomials of the lag operator L with orders p and q, respectively. The ARIMA model is particularly effective for handling non-stationary data, which is common in real-world scenarios where trends and seasonality are prevalent.

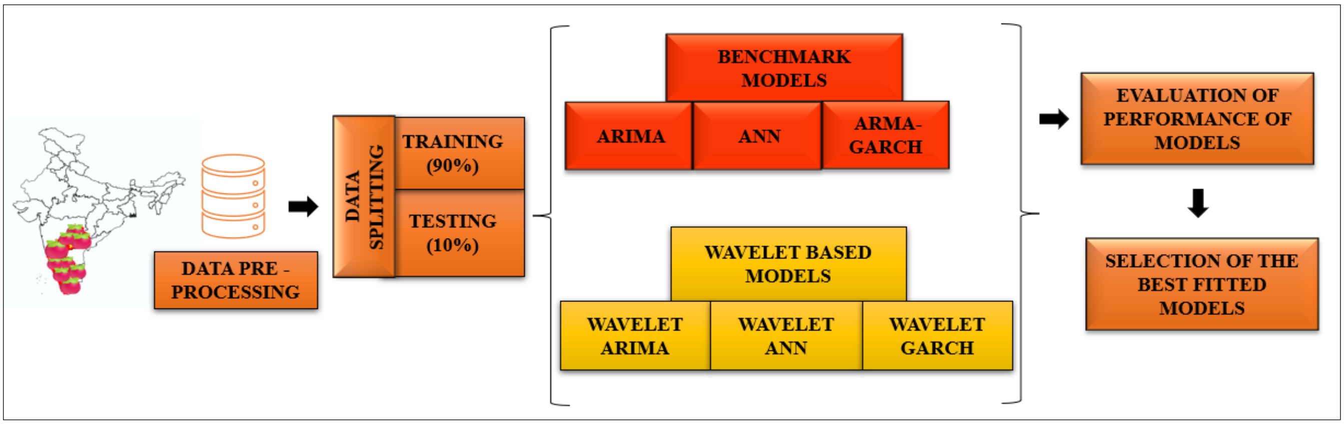

Fig. 1. Flowchart of the study.

Artificial Neural Network (ANN)

ANN is a powerful machine-learning model used to detect the complex and nonlinear patterns within data (18). ANNs are data-driven, self-learning models inspired by the architecture and functioning of biological neural networks in the human brain (29). As a non-parametric statistical method, ANNs do not require prior assumptions about the data distribution. The fundamental structure of an ANN consists of three key layers such as the input layer, which accepts external data as input features, hidden layers, where the input data is processed to identify patterns and the output layer, which generates the final predicted output (30). The structure of ANN is illustrated in Fig. 2. The learning process of an ANN involves adjusting the weights of the connections between neurons to reduce prediction errors. This optimization is achieved iteratively through methods such as backpropagation and gradient descent (31). This self-adaptive, data-driven approach makes them an ideal choice for handling complex datasets. Mathematically, an ANN model can be represented as:

$$y_t = f\left(\sum_{j=1}^{q} \omega_j g\left(\sum_{i=1}^{p} \omega_{ij} y_{t-i}\right)\right)$$

where y_t represents the observed value at time t,ω_j (j=1,2,…q) and ω_{ij} (i=1,2…p, j=1,2,…q) are the connection weights, which are the model parameters. p refers to the number of input nodes and q represents the number of hidden nodes. The functions g and f represent the activation functions applied in the hidden and output layers, respectively.

ARMA-GARCH

The hybrid time series modelling approach integrates both linear and nonlinear components to enhance forecasting accuracy (32). Traditional time series models are effective in capturing linear dependencies but often fail to capture complex nonlinear patterns. Nonlinear models can handle nonlinear dependencies but may struggle with long-term trend estimation (33). By integrating these models, hybrid approaches provide a precise forecasting. It begins with a linear model, such as ARMA, to capture linear patterns, followed by the detection of nonlinearity in residuals using the Brock-Dechert-Scheinkman (BDS) test. If nonlinear dependencies exist, an appropriate nonlinear model is applied to model the unexplained variations (34). The final forecast is obtained by combining the linear and nonlinear model predictions. For time series exhibiting volatility

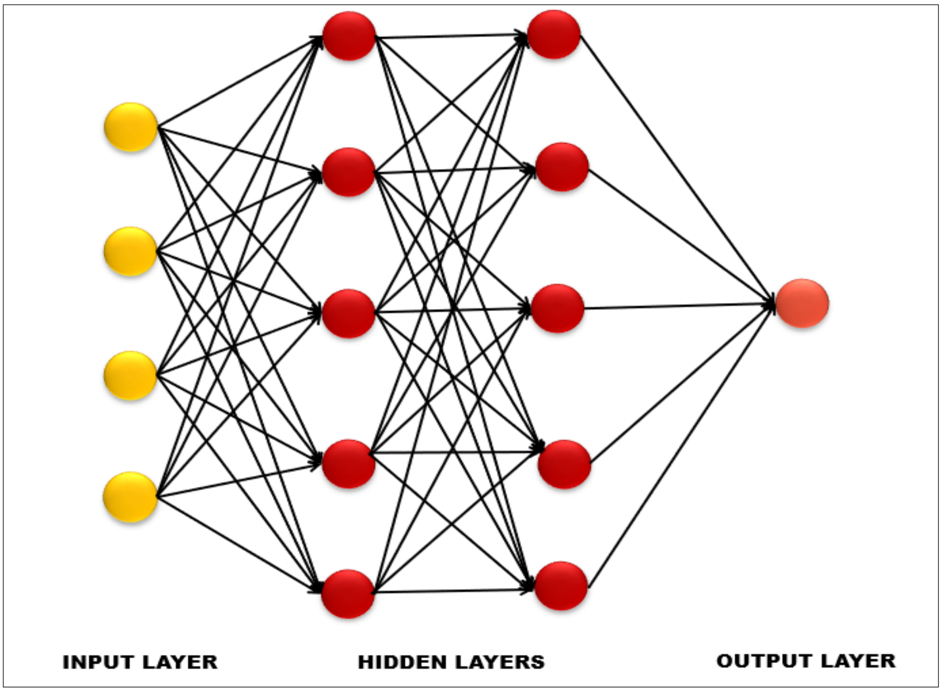

Fig. 2. Structure of ANN.

clustering, the ARCH-Lagrange Multiplier (LM test is conducted to check for conditional heteroskedasticity (35). If significant heteroskedasticity is detected, GARCH model is employed to capture time-varying volatility (36). The ARMA-GARCH hybrid approach provides a robust framework for handling complex time-dependent data, improving predictive performance and capturing intricate patterns that single-model approaches might overlook (37).

Wavelet analysis

Wavelet decomposition is an effective and powerful technique for analyzing time series data by breaking the original signal into multiple frequency components (6). This enables the identification of both short-term fluctuations and long-term trends. Wavelets act as fundamental elements, much like sine and cosine functions in trigonometry (38). However, unlike sine and cosine waves, wavelets oscillate and quickly diminish to zero, making them highly suitable for signal processing (4). In cases where standard wavelet transform methods are not applicable, the Maximal Overlap Discrete Wavelet Transform (MODWT) serves as a flexible alternative. Unlike conventional Discrete Wavelet Transform (DWT), MODWT introduces redundancy, enhancing resolution and stability in signal processing (39). DWT typically requires the data length to be a power of two for efficient processing. But MODWT does not have this restriction, making it more adaptable for real-world applications (8, 40). MODWT facilitates multi-resolution analysis by decomposing signals into multiple sub-series at different scales (41). It applies high-pass filters to extract high-frequency components and low-pass filters to capture low-frequency trends (42). Various wavelet filters, such as Haar, D4, LA8 and BI14, generate detail coefficients (from high-pass filters) and approximation coefficients (from low-pass filters), enabling effective noise reduction and trend identification (43).

The detailed coefficients are calculated as follows:

$$D_j(t) = \sum_{k=-\infty}^{\infty} w_{\psi_{j,k}} \psi_{j,k}(t)$$

The approximation coefficients are given by:

$$A_j(t) = \sum_{k=-\infty}^{\infty} V_{\phi_{j,k}} \phi_{j,k}(t)$$

The original signal is represented as the sum of its detailed and approximate components:

$$x(t) = \sum_{j=1}^{i} D_j(t) + A_j(t)$$

Where $\psi_{j,k}(t)$ is the wavelet function paired with $\phi_{j,k}(t)$, the scaling function. The wavelet coefficient is denoted as $W_{\psi j,k}$ while $V_{\phi j,k}$ is the scaling coefficient, t represents time, j and k are the scale and translation parameters, respectively and J indicates the level of decomposition, ranging from 1 to i. The number of decomposition levels is determined based on the length of the series (N). The minimum levels can be found by log N and maximum levels by $\log_2 N$ ensuring the efficient capture of the signal (6). The series is progressively decomposed until all levels are reached. This study integrates wavelet decomposition with ARIMA, ANN and GARCH models to enhance time series forecasting (44). The methodology follows these steps:

Step 1: The original time series data is divided into training and testing sets.

Step 2: The series is decomposed using the Haar wavelet filter, which provides optimal performance.

Step 3: Each decomposed component is separately modeled using ARIMA, ANN and GARCH.

Step 4: Forecasts are generated for the combined models, resulting in wavelet-based ARIMA, wavelet-based ANN and wavelet-based GARCH.

Step 5: Inverse Wavelet Transform (IWT) is applied to reconstruct the final forecast from the wavelet coefficients.

Performance metrics

The accuracy of the models is assessed by comparing the predicted values with the actual observations (45). To measure performance, error metrics such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) are employed. Models with lower values for these metrics are considered to provide a better fit, as they demonstrate greater effectiveness in capturing the patterns within the data (46). The mathematical representations of these metrics are as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}$$

$$\text{MAE} = \frac{1}{n}\sum_{t=1}^{n}|y_t - \hat{y}_t|$$

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right| \times 100$$

where y_t represents the actual values, $\hat{Y}_t$ is the forecasted values and n is the number of observations in the time series.

Results

Summary statistics

This study examines the descriptive statistics of monthly wholesale tomato prices in Bangalore, Chennai and Trivandrum to assess the price distribution and volatility across these markets (Table 1). The results indicate substantial fluctuations, as Bangalore and Chennai have minimum (₹150/qtl) and maximum (₹8600/qtl) prices. The mean price is ₹1435.82/qtl in Bangalore, ₹1690.01/qtl in Chennai and ₹2243.05/qtl in Trivandrum, with slightly lower median values, suggesting a right-skewed distribution. Price dispersion, measured by standard deviation, is highest in Bangalore (₹1186.88), followed by Chennai (₹1157.77/qtl) and Trivandrum (₹1098.04/qtl), indicating notable fluctuations. Skewness values of 2.52 (Bangalore), 1.86 (Chennai) and 1.81 (Trivandrum) confirm a positively skewed distribution, indicating occasional price spikes. Kurtosis values further

Table 1. Descriptive statistics of tomato price series (Rs./qtl)

Descriptive Statistics	Bangalore	Chennai	Trivandrum
Minimum	150	150	630
Maximum	8600	8600	6709
Mean	1435.82	1690.01	2243.05
Median	1111.50	1423.50	2014.50
Mode	727.96	885.11	850.27
Standard Deviation	1186.88	1157.77	1098.04
Skewness	2.52	1.86	1.81
Kurtosis	8.94	5.11	4.40
CV	82.66	70.35	48.95
CDVI	67.76	63.55	38.26

reveal that Bangalore (8.94) has a more peaked distribution than Chennai (5.11) and Trivandrum (4.40), indicating frequent extreme price variations. The coefficient of variation (CV) is highest for Bangalore (82.66 %), followed by Chennai (70.35 %) and Trivandrum (48.95 %), indicating that Bangalore exhibits the highest relative price volatility. Cuddy-Della Valle Index (CDVI) is also greater in Bangalore (67.76) and Chennai (63.55) than in Trivandrum (38.26), suggesting higher price instability in these markets. Bangalore exhibits the highest level of price fluctuations, making it the most volatile market among the three markets. Fig. 3 shows the tomato price distributions across different markets, confirming the above-observed findings regarding the nonlinear and non-stationary nature of prices in all the markets. These findings emphasize the significant price variations across different regions, which are essential for price forecasting and market stability.

Preliminary test

The Augmented Dickey-Fuller (ADF) test is a widely utilized statistical method in time series analysis to determine whether a series is stationary. The test is based on the null hypothesis that the series contains a unit root, indicating it is non-stationary. Rejecting this hypothesis confirms stationarity in the data. Table 2 presents the results of the preliminary test for time series analysis. ADF test results for Bangalore, Chennai and Trivandrum, indicates that test statistics for all markets are significantly lower than their respective critical values. This indicates that the data is stationary, confirming the absence of a unit root and making it suitable for further time series forecasting. The Box-Pierce test is used to check for autocorrelation in a time series dataset. A significant p-value (< 0.01) suggests that there is strong evidence of autocorrelation in the residuals of the time series model. Shapiro-Wilk test indicates that all datasets significantly deviate from normality (p-values < 0.01), highlighting the need for caution in applying parametric statistical methods. The Jarque-Bera test also confirms the non-normality across all the markets. To examine the presence of nonlinearity in the dataset, the BDS test is

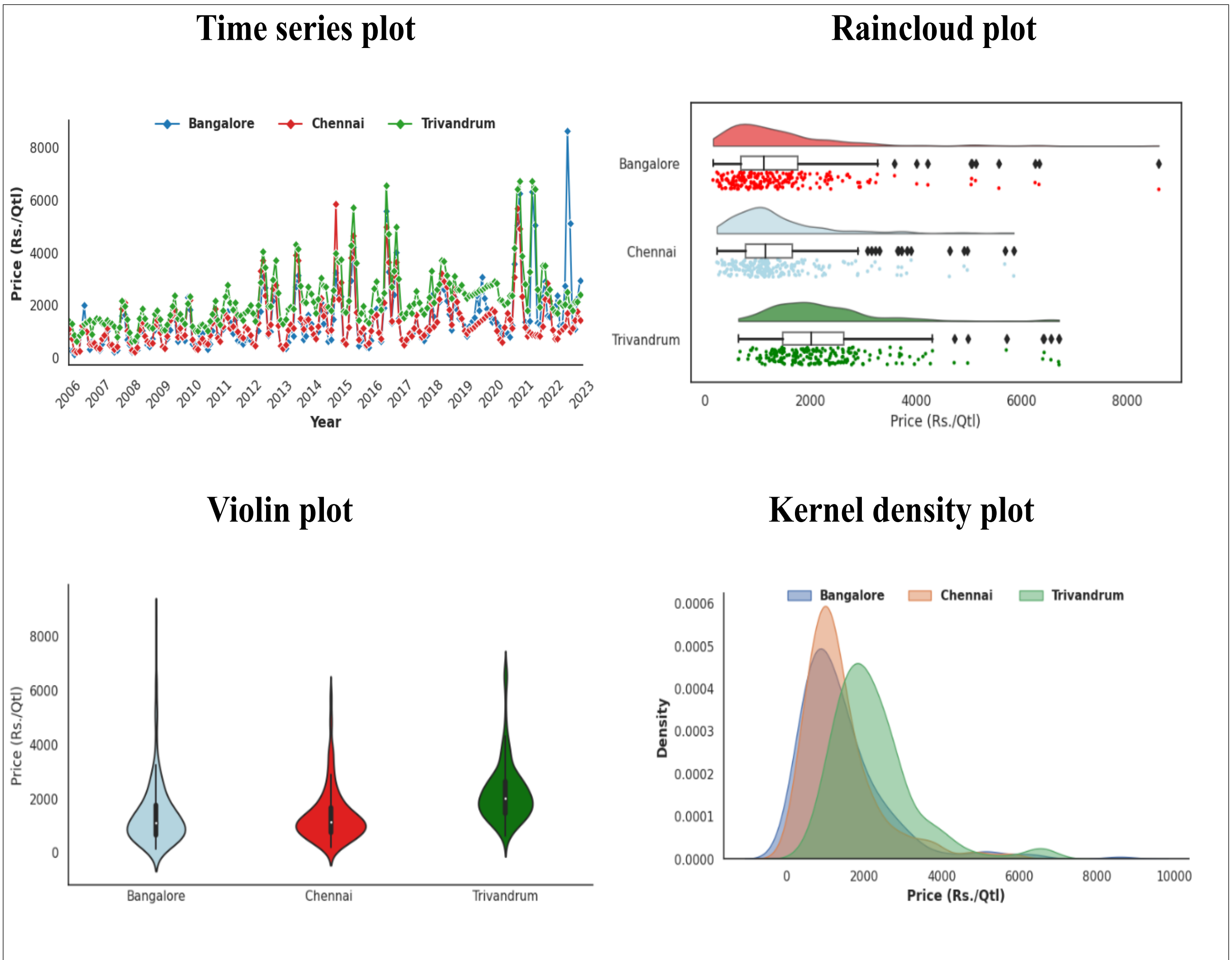

Fig. 3. Statistical charts of tomato price distribution.

Table 2. Results of the preliminary tests for time series data

Nonlinearity test

		Embedding dimension 2		Embedding dimension 3	
Bangalore					
Epsilon	Parameter	Statistic	p-value	Statistic	p-value
eps[1]	593.44	11.72	< 0.001	11.91	< 0.001
eps[2]	1186.88	7.88	< 0.01	7.03	< 0.01
eps[3]	1780.32	6.99	< 0.01	5.96	< 0.01
eps[4]	2373.76	6.16	< 0.01	5.25	< 0.01
Chennai					
Epsilon	Parameter	Statistic	p-value	Statistic	p-value
eps[1]	489.4	9.84	< 0.001	10.26	< 0.001
eps[2]	978.8	9.01	< 0.001	8.85	< 0.001
eps[3]	1468.2	8.52	< 0.001	8.55	< 0.001
eps[4]	1957.61	7	< 0.01	6.96	< 0.01
Trivandrum					
Epsilon	Parameter	Statistic	p-value	Statistic	p-value
eps[1]	549.02	22.13	< 0.001	25.36	< 0.001
eps[2]	1098.04	14.48	< 0.001	14.47	< 0.001
eps[3]	1647.06	10.41	< 0.001	9.72	< 0.001
eps[4]	2196.08	8.76	< 0.001	7.89	< 0.01

Stationary test

Market	Statistic	p-value
ADF test		
Bangalore	-5.89	0.01
Chennai	-5.46	0.01
Trivandrum	-4.21	0.01
Autocorrelation test		
Market	Box-Pierce Test Statistic	p-value
Bangalore	66.05	<0.01
Chennai	72.33	<0.01
Trivandrum	99.05	<0.01
Normality test		
Shapiro-Wilk	Statistic (W)	p-value
Bangalore	0.77	<0.01
Chennai	0.85	<0.01
Trivandrum	0.84	<0.01
Jarque Bera	Statistic (X^2)	p-value
Bangalore	916.67	<0.01
Chennai	1028.2	<0.01
Trivandrum	282.59	<0.01

conducted across different markets. As shown in Table 2, the results reveal the presence of nonlinearity in the data. The test statistics for different epsilon values show significant fluctuations, suggesting a complex underlying structure in the data. This confirms the necessity for advanced modelling techniques. Since the BDS test validates the nonlinear characteristics of the dataset, machine learning models emerge as a powerful alternative, as they can effectively capture complex dependencies without relying on rigid assumptions, thereby enhancing predictive accuracy.

Results of different models used in the study

The study employed a diverse set of stochastic, machine learning, hybrid and wavelet enhanced models, including ARIMA, ANN, ARMA-GARCH, Wavelet-ARIMA, Wavelet-ANN, Wavelet-GARCH each optimized to improve the forecasting accuracy. Following preliminary time series diagnostics, ARIMA models were initially fitted based on the lowest Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values (Table 3). This led to the selection of ARIMA (1, 0, 2) for Bangalore, ARIMA (1,0,1) for Chennai and ARIMA (2, 0, 1) for Trivandrum, effectively capturing the linear dependencies within the time series data for these markets. To further evaluate the residuals, the ARCH-LM test was conducted to detect heteroskedasticity. The results indicated a significant ARCH effect in the Chennai market, highlighting the potential volatility and the need for alternative models that can more effectively capture this pattern. In contrast, the residuals from the Bangalore and Trivandrum markets showed no signs of an ARCH effect, implying stable variance over time. Given the presence of conditional heteroskedasticity in Chennai, an ARMA-GARCH model was fitted to effectively capture the volatility (Table 4). The ARMA (1, 2) + GARCH (1, 1) configuration emerged as the best-fitting model for the Chennai market, with the

Table 3. Estimates of ARIMA model

Market						ARCH-LM test
Bangalore	**ARIMA(1,0,2)**	AIC = 2951.56		BIC = 2967.63		Statistic 16.82; p-value 0.15
	Parameter	ar1	ma1	ma2	mean	
	Estimate	-0.17	1.01	0.45	1288.64	
	SE	0.23	0.21	0.13	110.31	
Chennai	**ARIMA(1,0,1)**	AIC = 2980.47		BIC = 2993.33		Statistic 23.32; p-value 0.02
	Parameter	ar1		ma1	mean	
	Estimate	0.37		0.39	1396.42	
	SE	0.09		0.07	126.88	
Trivandrum	**ARIMA(2,0,1)**	AIC = 2939.15		BIC = 2955.23		Statistic 10.86; p-value 0.54
	Parameter	ar1	ar2	ma1	mean	
	Estimate	0.64	-0.11	0.34	2201.09	
	SE	0.14	0.11	0.12	145.57	

Table 4. Estimates of the ARMA-GARCH model

	ARMA (1,2) + GARCH (1,1)			AIC = 15.94			BIC = 16.06	
	Parameter	mu	ar1	ma1	ma2	omega	alpha1	beta1
Chennai	Estimate	1115.0	0.07	0.95	0.40	180200	0.37	0.33
	SE	240.5	0.19	0.17	0.11	44820	0.12	0.11
	p-value	<0.001	0.68	<0.001	<0.001	<0.001	0.001	0.002

Table 5. Estimates of ANN model

Market	Network architecture	Parameters (Weights)
Bangalore	5-2-1	15
Chennai	2-5-1	21
Trivandrum	4-2-1	13

lowest AIC (15.94) and BIC (16.06) values. This model captures the trend and volatility in tomato price data of Chennai market. The ANN architectures (Table 5) were optimized for each region, with Bangalore utilizing a 5-2-1 network optimized with 15 weights, Chennai employing a 2-5-1 network with 21 weights and Trivandrum adopting a 4-2-1 network with 13 weights. These configurations were determined by fine-tuning the parameters to effectively capture the intricate nonlinear relationships within the data.

While the benchmark models successfully captured the overall price trends across various markets, wavelet transformation was applied to improve forecasting accuracy by breaking down each series into distinct frequency components. Using the Haar wavelet filter, the series was decomposed up to a maximum of seven levels (with the highest J = $\log_2 N$, where N = 204). Fig. 4 represents the wavelet decomposition levels of tomato price series across different markets. The summary statistics of wavelet

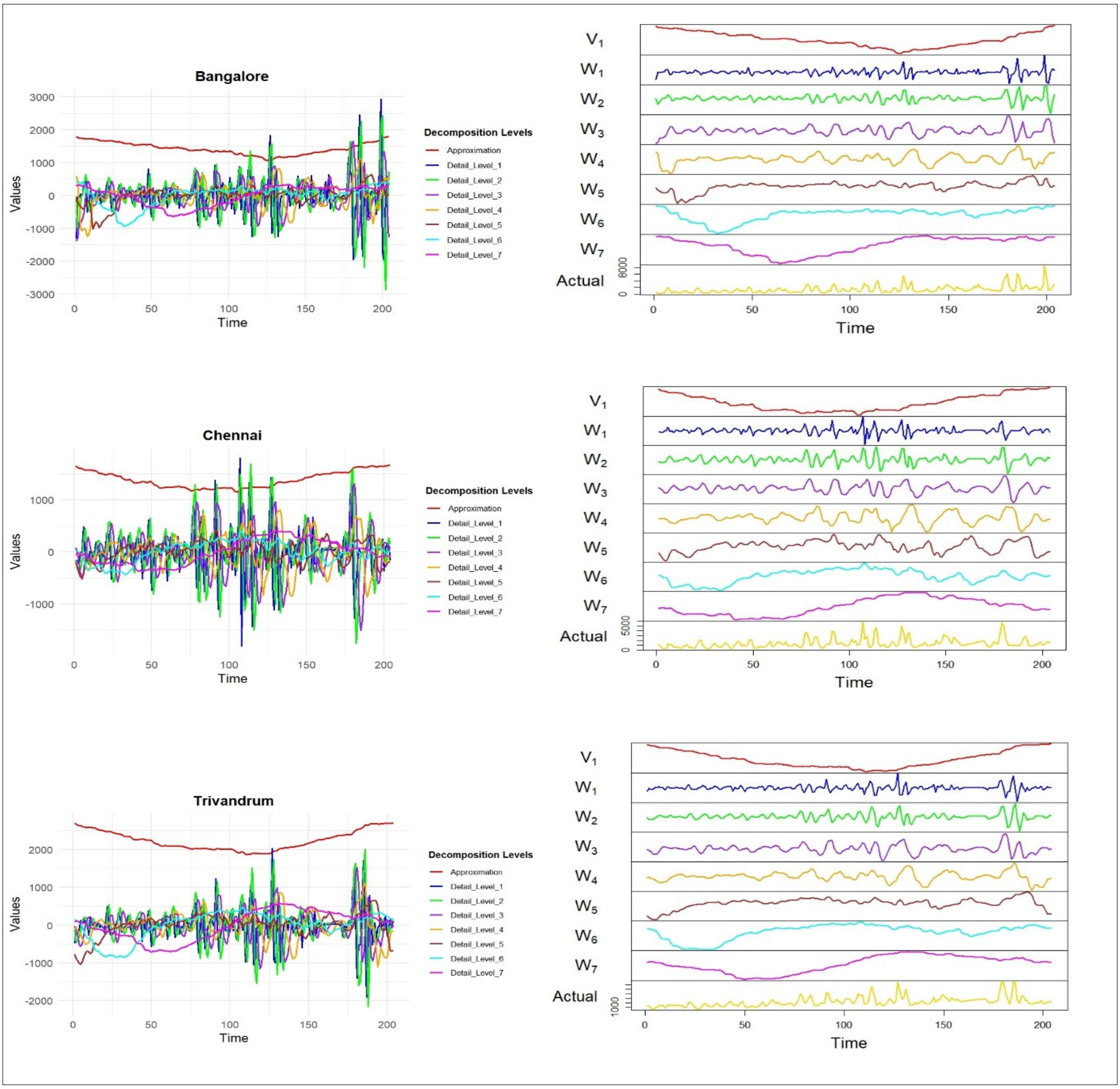

Fig. 4. Wavelet decomposition levels of tomato price data of different markets.

coefficients for tomato prices across different markets provide valuable insights of price dynamics. In wavelet analysis, V1 represents the approximation coefficients, capturing the long-term trend, while W1-W7 denotes the detail coefficients at various levels, reflecting short-term fluctuations at different frequencies. The wavelet decomposition of tomato prices of Bangalore market (Table 6, Fig. 5) reveals significant short-term volatility in W1-W3, characterized by high standard deviations and leptokurtic distributions. Lower frequency levels (W5-W7) show reduced fluctuations but exhibit negative skewness, indicating frequent downward price movements. The approximation coefficient (V1) represents the overall trend with minimal skewness, suggesting stable long-term price behavior. For Chennai (Table 7, Fig. 6), short-term fluctuations (W1-W3) exhibit higher variability, as indicated by larger standard deviations, while long-term trends (W5-W7) remain more stable. Skewness and kurtosis indicate symmetric distribution, except for slight negative skewness at finer levels. Similarly, the wavelet decomposition of Trivandrum's tomato prices (Table 8, Fig. 7) indicates substantial short-term variability in W1-W3, with high standard deviations and positive skewness. Lower wavelet levels (W5-W7) show declining volatility but exhibit negative skewness, implying more frequent downward movements. The approximation coefficient (V1) represents a stable long-term trend with slight positive skewness. These decomposed series were subsequently utilized as inputs for the ARIMA, ANN and GARCH models. This process ensured the effective utilization of the distinct frequency components extracted from the data and enhancing the forecasting performance.

Selection of the best fitting model

The forecasting performance of different models across Bangalore, Chennai and Trivandrum markets reveals varying levels of accuracy (Table 9). In Bangalore, Wavelet-ANN outperformed the benchmark models, highlighting the effectiveness of wavelet-based approaches, with RMSE (1302.42), MAPE (25.93 %) and MAE (863.48). In Chennai, Wavelet ARIMA achieved the highest predictive accuracy, achieving the lowest RMSE (548.04), MAPE (20.27 %) and MAE (406.61), followed closely by Wavelet ANN and Wavelet GARCH. Similarly, in Trivandrum, Wavelet ANN exhibited superior forecasting accuracy with RMSE (891.57), MAPE (17.87 %) and MAE (666.79). Hence, wavelet-enhanced models consistently provided more precise predictions across all the markets, highlighting their advantage in effectively capturing the price fluctuations. The radar charts

Table 6. Descriptive statistics of wavelet coefficients of tomato price data of Bangalore market

Bangalore	V1	W1	W2	W3	W4	W5	W6	W7
Minimum	1076.55	-1977.5	-2863.75	-1379	-1245.19	-1026.91	-931.77	-638.34
Maximum	1803.41	2930	2427.5	1660.5	1146.81	637	433.45	333.76
Mean	1435.82	0	0	0	0	1.26	0	1
Median	1406	-1	-34.25	-10.25	16.38	44.36	109.72	118.93
Std Deviation	186.42	555.75	643.92	479.3	400.48	280.24	311.99	294.71
Skewness	0.09	0.43	0.03	0.23	-0.32	-1.33	-1.43	-0.73
Kurtosis	-1	7.01	4.19	1.31	0.97	2.9	1.39	-0.88

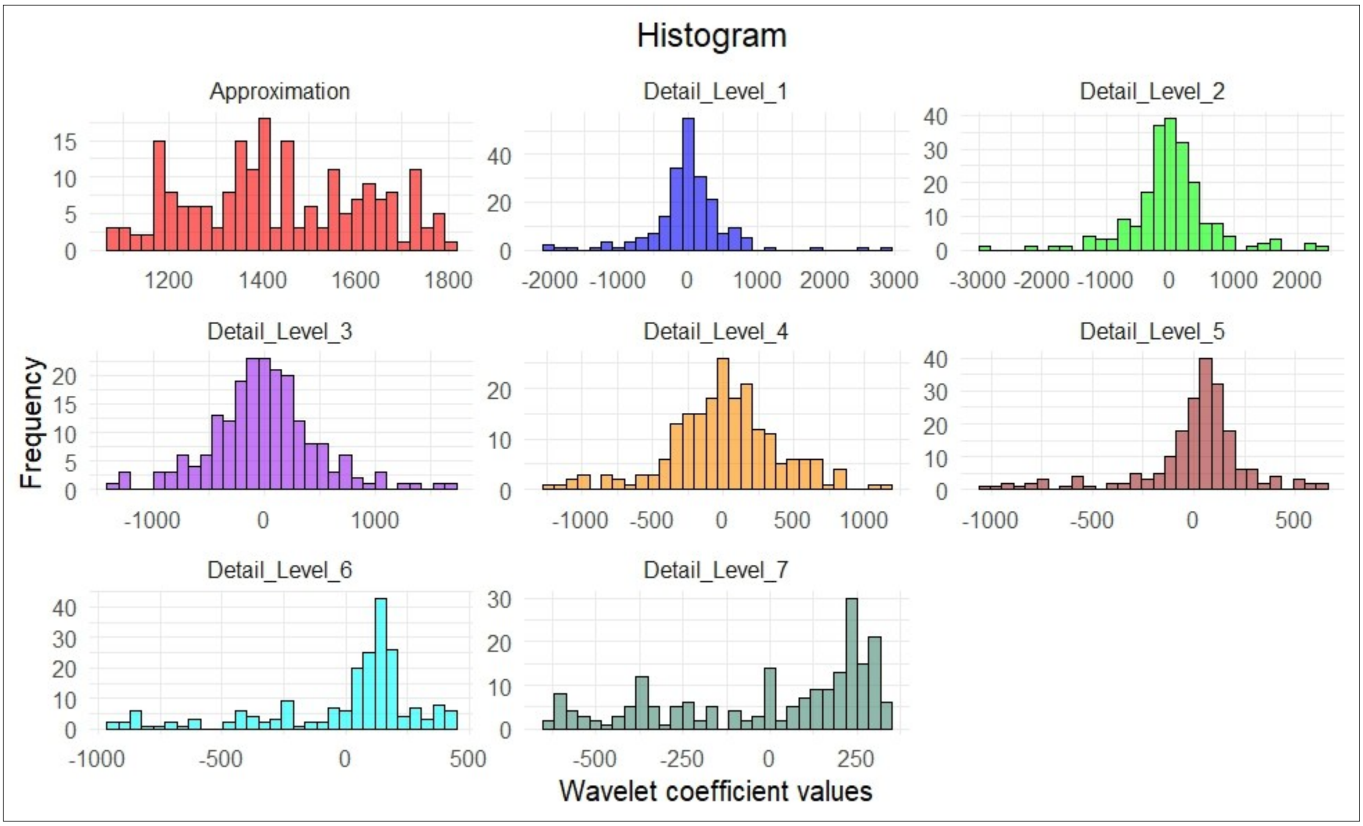

Fig. 5. Histogram of wavelet coefficients of Bangalore market.

Table 7. Descriptive statistics of wavelet coefficients of tomato price data of Chennai market

Chennai	V1	W1	W2	W3	W4	W5	W6	W7
Minimum	1153.45	-1812	-1761.75	-1514.88	-862.86	-503.33	-436.09	-365.78
Maximum	1669.97	1805.5	1690	1315.63	803.19	420.31	394.88	399.69
Mean	1391.17	0	0	0	0	29.17	25.75	-18.49
Median	1357.92	7.75	17	28.25	-27.47	29.17	25.75	-18.49
Std Deviation	155.47	440.24	526.65	455.11	347.74	194.68	199.92	238.32
Skewness	0.29	0.06	0.18	-0.11	0	-0.27	-0.51	0.08
Kurtosis	-1.33	4.12	1.96	0.76	-0.02	-0.53	-0.64	-1.29

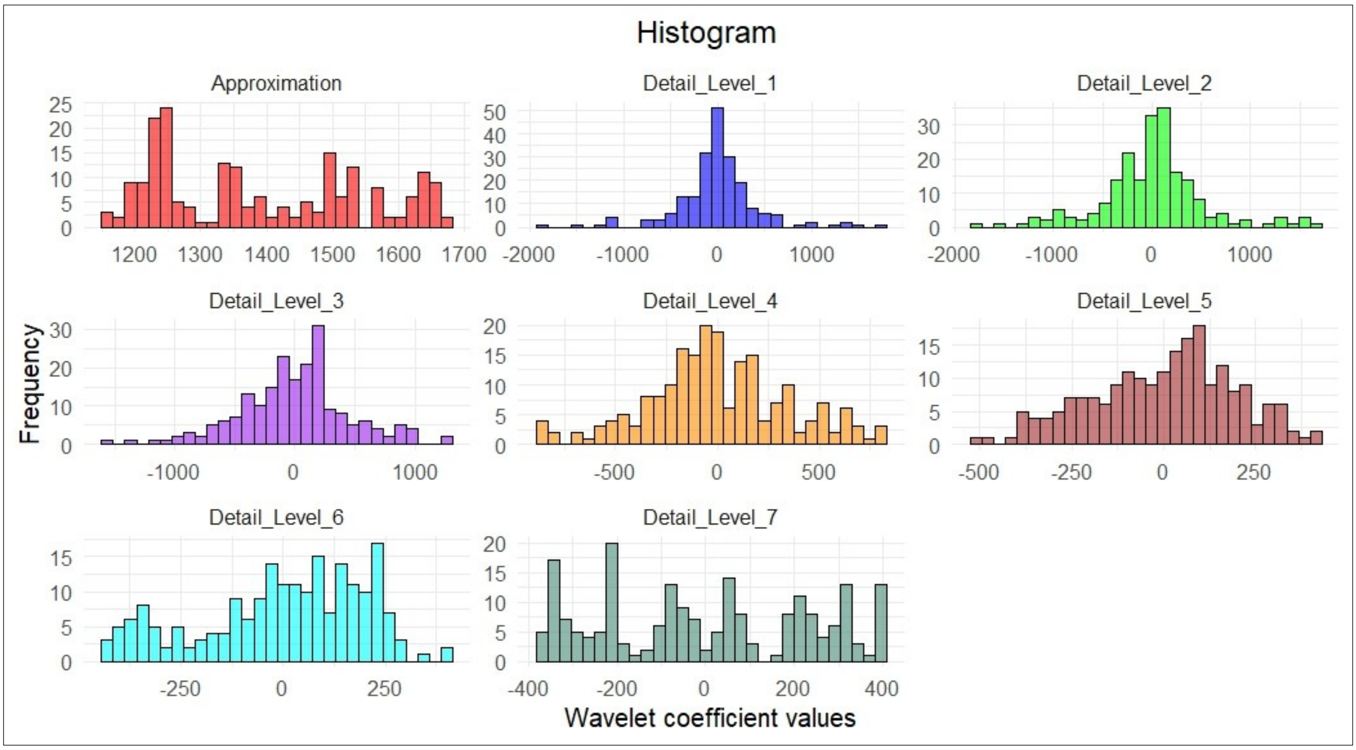

Fig. 6. Histogram of wavelet coefficients of Chennai market.

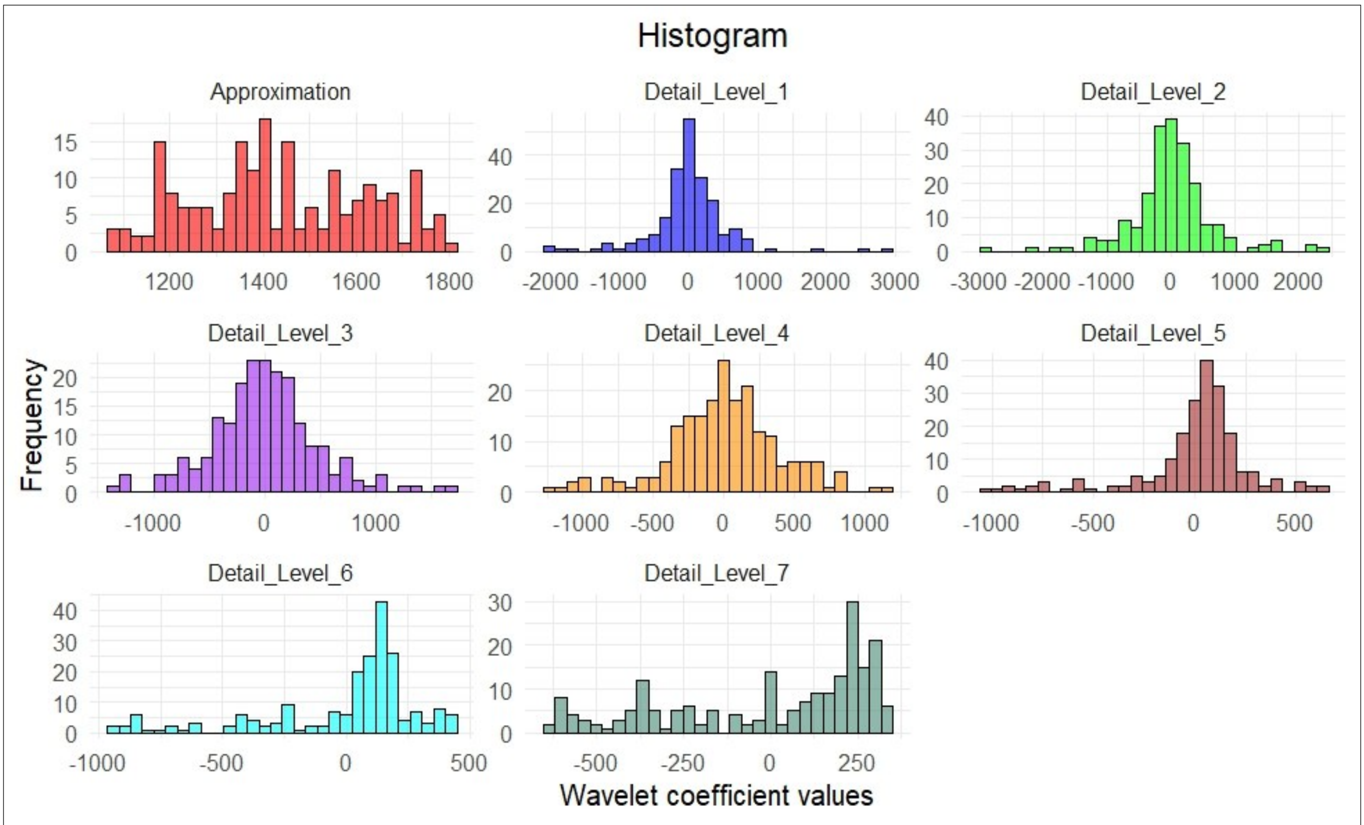

Fig. 7. Histogram of wavelet coefficients of Trivandrum market.

Table 8. Descriptive statistics of wavelet coefficients of tomato price data of Trivandrum market

Trivandrum	V1	W1	W2	W3	W4	W5	W6	W7
Minimum	1872	-1913	-2149	-1156	-1111	-1032	-847	-715
Maximum	2704	2036	2009	1533	1108	655	459	568
Mean	2243	0	0	0	0	0	0	0
Median	2200	2	15	9	-4	47	114	83
Std Deviation	254	428	528	409	367	311	356	393
Skewness	0.37	0.2	0.07	0.21	-0.11	-1.12	-1.18	-0.34
Kurtosis	-1.16	5.62	3.6	1.52	0.88	1.84	0.34	-1.11

Table 9. Performance metrics of different models

Market	Models	RMSE	MAPE	MAE
Bangalore	ARIMA	1631.84	35.90	980.63
	ANN	1450.47	31.68	906.17
	WARIMA	1256.21	28.41	886.33
	WANN	**1302.42**	**25.93**	**863.48**
Chennai	ARIMA	593.01	27.78	500.64
	ANN	609.95	29.96	530.45
	GARCH	641.25	31.47	576.86
	WARIMA	**548.04**	**20.27**	**406.61**
	WANN	551.04	22.97	448.78
	WGARCH	558.37	25.53	460.7
Trivandrum	ARIMA	1578.13	24.68	971.56
	ANN	1426.50	21.96	875.76
	WARIMA	962.7	19.47	686.94
	WANN	**891.57**	**17.87**	**666.79**

(Fig. 8) compare the performance of various forecasting models across three different markets. A model's polygon being closer to the centre indicates lower error values with superior predictive performance. Across all three markets, wavelet-enhanced models consistently exhibit lower error values, highlighting the effectiveness of integrating wavelet decomposition with statistical and machine learning models for enhanced forecasting accuracy. The Diebold Mariano (DM) test was performed to evaluate the forecasting accuracy of benchmark models against wavelet-enhanced models. As presented in Table 10, the results indicate the superior performance of wavelet-based models across different markets. Specifically, the test statistics were negative, with smaller p-values, confirming that models such as Wavelet ARIMA, Wavelet ANN and Wavelet GARCH consistently outperformed benchmark models like ARIMA, ANN and ARMA-GARCH in terms of forecasting accuracy. These findings highlight that integrating wavelet decomposition with benchmark models provides a robust approach for forecasting tomato prices, particularly in markets exhibiting high volatility and nonlinear trends. Across all the markets, the wavelet-enhanced models outperformed well than standalone individual models.

Table 10. DM test results of tomato price series of different markets

Market	Wavelet based model	Benchmark model	DM Test Statistic	p-value
Bangalore	WARIMA	ARIMA	-2.10	0.0213
	WANN	ANN	-1.78	0.0342
Chennai	WARIMA	ARIMA	-3.61	0.0041
	WANN	ANN	-3.05	0.0024
	WGARCH	ARMA-GARCH	-4.38	0.0001
Trivandrum	WARIMA	ARIMA	-3.14	0.0020
	WANN	ANN	-3.24	0.0021

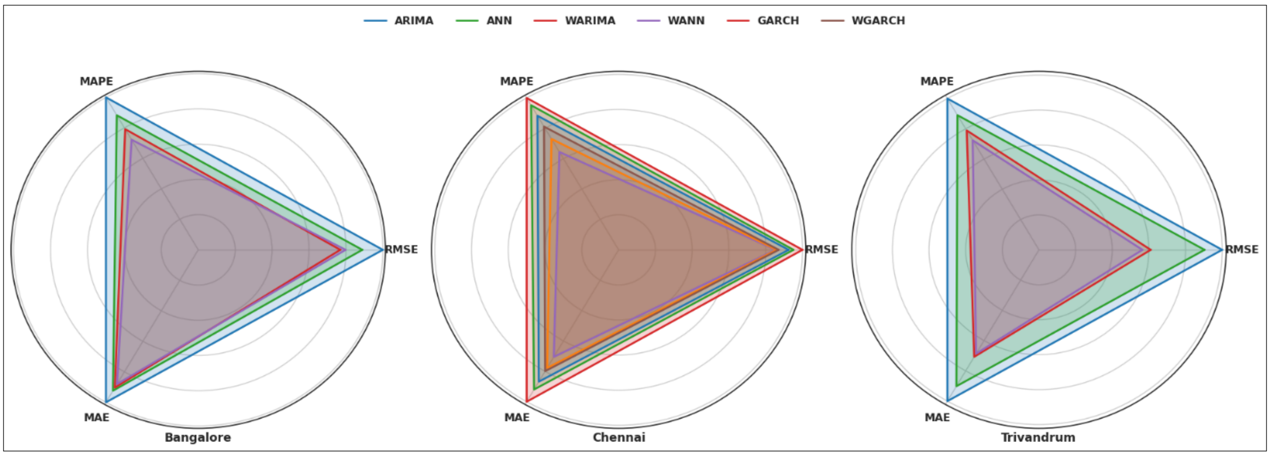

Fig. 8. Radar plot.

Discussion

This study examines the benchmark, machine learning models and wavelet-enhanced techniques in forecasting tomato prices across three markets namely Bangalore, Chennai and Trivandrum. By analysing ARIMA, ANN, ARMA-GARCH and wavelet enhanced models, the study provides insights of each model's ability to capture the patterns in the data. The forecasting performance across these markets indicates that wavelet-based models consistently outperform traditional models in terms of accuracy as observed from the previous studies (4, 6, 8, 42). In Bangalore, Wavelet ANN achieved the lowest error metrics, making it the most effective model for price forecasting. Traditional ARIMA exhibited the highest errors, highlighting its limitations in capturing complex price variations. Similarly, in Chennai, Wavelet ARIMA demonstrated the best performance, highlighting the superior predictive capability. Wavelet ANN also performed well, while ANN and ARMA-GARCH exhibited relatively higher errors, emphasizing their limitations in handling the price fluctuations. In Trivandrum, Wavelet ANN again emerged as the top-performing model, closely followed by Wavelet ARIMA, whereas ANN and ARIMA exhibited higher errors, indicating their reduced adaptability to capture the price variations. Overall, wavelet enhanced models consistently delivered predictions that are more accurate across all the markets, highlighting the advantage of wavelet-based models in capturing both short-term fluctuations and long-term price patterns effectively (39). Traditional ARIMA and ANN models performed moderately but had higher errors, making them less reliable for precise price forecasting. These findings suggest that wavelet-enhanced models offer the most accurate price forecasts for tomatoes in Bangalore, Chennai and Trivandrum markets (8). This is particularly valuable for farmers, as precise price forecasts enable them to make well-informed decisions to maximize the profits (43, 47). Forecasting price fluctuations allow farmers to mitigate financial risks, enhance supply chain efficiency and make strategic marketing decisions regarding storage and transportation (46). Since vegetables lack a Minimum Support Price (MSP), the government should provide feasible solutions to introduce price stabilization measures, particularly for highly volatile crops like tomatoes (25). This study would contribute to determining an optimal MSP and formulating effective agricultural policies, as suggested by previous research (36). Furthermore, accurate price forecasting supports policymakers and traders in market planning, ensuring price stability and minimizing the adverse effects of market volatility.

Conclusion

This study explored the wholesale tomato price data across key South Indian markets such as Bangalore, Chennai and Trivandrum by employing various predictive models, including ARIMA, ANN, ARMA-GARCH and wavelet-enhanced models such as Wavelet ARIMA, Wavelet ANN and Wavelet GARCH. The results highlight the superior performance of wavelet-based models, particularly Wavelet ANN and Wavelet ARIMA, which consistently outperformed traditional benchmark models in terms of forecasting accuracy. Lower RMSE, MAPE and MAE values demonstrated the robustness of these models in capturing both short and long-term price patterns in the data. Future research could focus on optimizing filter selection and decomposition levels to further enhance forecasting accuracy and better understand of the seasonal and irregular patterns in agricultural time series data. Tomatoes are highly volatile crops lacking an MSP, exposing farmers to unpredictable price fluctuations. To ensure financial security and market stability, the government should consider implementing a fixed price policy. Such a measure would safeguard the farmers' livelihoods and reduce the market uncertainties. For farmers and market stakeholders, this study offers the advanced decision-making approach for optimizing marketing strategies, mitigating financial risks and enhancing market efficiency. This study highlights the crucial role of advanced forecasting techniques in strengthening market stability, optimizing decision-making and creating sustainable agricultural practices.

Acknowledgements

The authors have no acknowledgements to declare.

Authors' contributions

VL carried out the research framework, methodology, review & editing, and drafted the manuscript. RGS participated in the research framework, methodology, review & editing and validation. PSG performed investigation and review & editing. JA participated in the research framework and validation of the manuscript. VK contributed to the research framework, investigation, review & editing and validation. SVS carried out the research framework, methodology, review & editing and validation of the manuscript. All authors read and approved the final manuscript.

Compliance with ethical standards

Conflict of interest: Authors do not have any conflict of interests to declare.

Ethical issues: None

References

1. Reddy AA. Price forecasting of tomatoes. Int J Veg Sci. 2019;25 (2):176-84. https://doi.org/10.1080/19315260.2018.1495674
2. Zhang Q, Yang W, Zhao A, Wang X, Wang Z, Zhang L. Short-term forecasting of vegetable prices based on lstm model-Evidence from Beijing's vegetable data. Plos one. 2024;19(7):e0304881. https://doi.org/10.1371/journal.pone.0304881
3. Purohit SK, Panigrahi S, Sethy PK, Behera SK. Time series forecasting of price of agricultural products using hybrid methods. Appl Artif Intell. 2021;35(15):1388-406. https://doi.org/10.1080/08839514.2021.1981659
4. Garai S, Paul RK, Rakshit D, Yeasin M, Emam W, Tashkandy Y, et al. Wavelets in combination with stochastic and machine learning models to predict agricultural prices. Mathematics. 2023;11 (13):2896. https://doi.org/10.3390/math11132896
5. Cho W, Kim S, Na M, Na I. Forecasting of tomato yields using attention-based LSTM network and ARMA model. Electronics.

2021;10(13):1576. https://doi.org/10.3390/electronics10131576

6. Tamilselvi C, Yeasin M, Paul RK, Paul AK. Can denoising enhance prediction accuracy of learning models? A case of wavelet decomposition Approach. Forecasting. 2024;6(1):81-99. https://doi.org/10.3390/forecast6010005
7. Iwabuchi K, Kato K, Watari D, Taniguchi I, Catthoor F, Shirazi E, et al. Flexible electricity price forecasting by switching mother wavelets based on wavelet transform and Long Short-Term Memory. Energy and AI. 2022;10:100192. https://doi.org/10.1016/j.egyai.2022.100192
8. Paul RK, Garai S. Performance comparison of wavelets-based machine learning technique for forecasting agricultural commodity prices. Soft Comput. 2021(20):12857-73. https://doi.org/10.1007/s00500-021-06087-4
9. Sun F, Meng X, Zhang Y, Wang Y, Jiang H, Liu P. Agricultural product price forecasting methods: A review. Agriculture. 2023;13(9):1671. https://doi.org/10.3390/agriculture13091671
10. Mathenge Mutwiri R. Forecasting of tomatoes wholesale prices of Nairobi in Kenya: time series analysis using Sarima model. Int JStat Distrib Appl. 2019;5(3):46. https://doi.org/10.11648/j.ijsd.20190503.11
11. Ivanisevic D, Mutavdzic B, Novkovic N, Vukelic N. Analysis and prediction of tomato price in Serbia. Econ Agric. 2015;62(4):951-62. https://doi.org/10.5937/ekoPolj1504951I
12. Liu S, Yuan H, Zhao Y, Li T, Zu L, Chang S. Research on multi-step fruit colour prediction model of tomato in solar greenhouse based on time series data. Agriculture. 2024;14(8):1211. https://doi.org/10.3390/agriculture14081211
13. Hossain MM, Abdulla F. On the production behaviours and forecasting the tomatoes production in Bangladesh. J Agric Econ Dev. 2015;4(5):66-74.
14. Chitikela G, Admala M, Ramalingareddy VK, Bandumula N, Ondrasek G, Sundaram RM, et al. Artificial-intelligence-based time-series intervention models to assess the impact of the COVID-19 pandemic on tomato supply and prices in Hyderabad, India. Agron. 2021;11(9):1878. https://doi.org/10.3390/agronomy11091878
15. Usha S, Muralibhaskaran V, Monish G, Vigneswaran G. Forecasting of tomato prices and yield based on season and weather using LSTM and Arima algorithms. International Conference on Emerging Research in Computational Science (ICERCS) 2024 (pp. 1-6). IEEE. https://doi.org/10.1109/ICERCS63125.2024.10895499
16. Osei I, Appiah B, Nyamadi M, Frimpong B, Titiati EK. Maize crop price prediction in Ghana using time series models. Open Sci J. 2025;10(1):1-12. https://doi.org/10.23954/osj.v10i1.3648
17. Badal PS, Kamalvanshi V, Goyal A, Kumar P, Mondal B. Forecasting potato prices: application of ARIMA model. Econ Affs. 2022;67(4):491-96. https://doi.org/10.46852/0424-2513.4.2022.14
18. Kumar R, Lad YA, Kumari P. Forecasting potato prices in Agra: Comparison of linear time series statistical vs. neural network models. Potato Res. 2025:1-22. https://doi.org/10.1007/s11540-024-09838-6
19. Xu X, Zhang Y. Commodity price forecasting via neural networks for coffee, corn, cotton, oats, soybeans, soybean oil, sugarand wheat. Intell Syst Account. Finance and Manag. 2022;29(3):169-81. https://doi.org/10.1002/isaf.1519
20. Yahaya AE, Etuk EH, Emeka A. Comparative performance of ARIMA and GARCH model in forecasting crude oil price data. Asian J Probab Stat. 202115(4);251–75. https://doi.org/10.9734/ajpas/2021/v15i430378
21. Qiao Y, Ahn BI. Volatility analysis and forecasting of vegetable prices using an ARMA‐GARCH model: An application of the CF filter and seasonal adjustment method to Korean green onions. Agribus. 2024. https://doi.org/10.1002/agr.21958
22. Manjunatha B, Paul RK, Ramasubramanian V, Avinash G, Paul AK, Yeasin M, et al. Trivariate-ARMA-GARCH type-Vine Copula model for time series forecasting. Commun Stat-Simul Comput. 2024:1-31. https://doi.org/10.1080/03610918.2024.2433495
23. Zhang J, Liu H, Bai W, Li X. A hybrid approach of wavelet transform, ARIMA and LSTM model for the share price index futures forecasting. NAmJ Econ Financ. 2024;69:102022. https://doi.org/10.1016/j.najef.2023.102022
24. Khan MM, Muhammad NS, El-Shafie A. Wavelet based hybrid ANN-ARIMA models for meteorological drought forecasting. J Hydrol. 2020;590:125380. https://doi.org/10.1016/j.jhydrol.2020.125380
25. Shankar SV, Chandel A, Gupta RK, Sharma S, Chand H, Aravinthkumar A, et al. Comparative study on key time series models for exploring the agricultural price volatility in potato prices. Potato Res. 2024:1-9. https://doi.org/10.1007/s11540-024-09776-3
26. Vitale J, Robinson J. In-Season price forecasting in cotton futures markets using ARIMA, neural networkand LSTM machine learning models. J Risk and Financ Manag. 2025;18(2):93. https://doi.org/10.3390/jrfm18020093
27. Reddy BD, Naik JS, Kumar SV, Kumar S, Haritha G, Reddy MR. A methodological review on time series forecasting by using ARIMA. In: Proceedings of the International Conference on Advanced Materials, Manufacturing and Sustainable Development (ICAMMSD 2024) 2025 Mar 17 pp. 709-19. Atlantis Press. https://doi.org/10.2991/978-94-6463-662-8_55
28. Kumar P, Ekka P. Demand forecasting for ensuring safety and boosting operational efficiency in hotel hospitality using ARIMA model. JHTE. 2025:1-5. https://doi.org/10.1080/10963758.2024.2436584
29. Shankar SV, Ajaykumar R, Ananthakrishnan S, Aravinthkumar A, Harishankar K, Sakthiselvi T, et al. Modeling and forecasting of milk production in the western zone of Tamil Nadu. Asian J Dairy Food Res. 2023;42(3):427-32. https://doi.org/10.18805/ajdfr.DR-2103
30. Bashir U, Singh K, Mansotra V. Examining daily closing price prediction of the NSE index using an optimized artificial neural network: A study of stock market. J Sci Res. 2025;17(1):195-209. https://doi.org/10.3329/jsr.v17i1.74640
31. Paul RK, Yeasin M, Kumar P, Kumar P, Balasubramanian M, Roy HS, et al. Machine learning techniques for forecasting agricultural prices: A case of brinjal in Odisha, India. Plos one. 2022;17(7):e0270553. https://doi.org/10.1371/journal.pone.0270553
32. Salman AA, Ansseif AAL. Modeling and forecasting volatility in the Iraq stock exchange: A survey study using ARCH and GARCH models. Int J Stat Appl Math. 2025;10(2):35–45. https://doi.org/10.22271/maths.2025.v10.i2a.1978
33. Nath B, Bhattacharya D. Forecast of agricultural commodity price in the presence of volatility. Environ Ecol. 2025;43(1):79–88. https://doi.org/10.60151/envec/TWWK2894
34. Fang Z, Han JY. Realized GARCH model in volatility forecasting and option pricing. Comput Econ. 2025:1-21. https://doi.org/10.1007/s10614-024-10826-8
35. Mehtarizadeh H, Mansouri N, Mohammad Hasani Zade B, Hosseini MM. Stock price prediction with SCA-LSTM network and statistical model ARIMA-GARCH. J Supercomput. 2025;81(2):366. https://doi.org/10.1007/s11227-024-06775-6
36. Shankar SV, Chandel A, Gupta RK, Sharma S, Chand H, Kumar R, et al. Exploring the dynamics of arrivals and prices volatility in onion (*Allium cepa*) using advanced time series techniques. Front Sustain Food Syst. 2023;7:1208898. https://doi.org/10.3389/fsufs.2023.1208898
37. Tarno, Maruddani DAI, Rahmawati R, Hoyyi A, Trimono, Munawar. Arma-garch model forvalue-at-risk (Var) prediction on stocks of pt. astra agro lestari.tbk. JMathComput Sci. 2021;11(2):2136-52. https://doi.org/10.28919/jmcs/5453

38. Nigam S, Verma S, Nagabhushan P. Wavelet RCNN: Enhancing object detection accuracy through spectral information. In 2023 IEEE 20th India Council International Conference (INDICON) 2023 Dec 14. pp. 1323-329. IEEE. https://doi.org/10.1109/INDICON59947.2023.10440842

39. Paul RK. ARIMAX-GARCH-WAVELET model for forecasting volatile data. Model Assist StatAppl. 2015;10(3):243-52. https://doi.org/10.3233/MAS-150328

40. Paul RK, Sarkar S, Yadav SK. Wavelet based long memory model for modelling wheat price in India. Indian J Agric Sci. 2021;91 (2):227-31. https://doi.org/10.56093/ijas.v91i2.111594

41. Ray M, Singh KN, Ramasubramanian V, Paul RK, Mukherjee A, Rathod S. Integration of wavelet transform with ANN and WNN for time series forecasting: an application to Indian monsoon rainfall. Natl AcadSci Lett. 2020;43(6):509-13. https://doi.org/10.1007/s40009-020-00887-2

42. Anjoy P, Paul RK. Comparative performance of wavelet-based neural network approaches. Neural Comput Appl. 2019;31:3443-53. https://doi.org/10.1007/s00521-017-3289-9

43. Rathod S, Singh KN, Paul RK, Meher SK, Mishra GC, Gurung B, et al. An improved ARFIMA model using maximum overlap discrete wavelet transform (MODWT) and ANN for forecasting agricultural commodity price. J Ind Soc Agric Stat. 2017;71:103-11.

44. Singh S, Parmar KS, Kumar J. Development of multi-forecasting model using Monte Carlo simulation coupled with wavelet denoising-ARIMA model. Math Comput Simul. 2025;230:517-40. https://doi.org/10.1016/j.matcom.2024.10.040

45. Paul RK, Gurung B, Paul AK. Modelling and forecasting of retail price of arhar dal in Karnal, Haryana. Indian J Agric Sci. 2015;85 (1):69-72. https://doi.org/10.56093/ijas.v85i1.46001

46. Paul RK, Shankar SV, Yeasin M. Forecasting area and yield of cereal crops in India: intelligent choices among stochastic, machine learning and deep learning techniques. Proc Indian Natl Sci Acad. 2024;22:1-7. https://doi.org/10.1007/s43538-024-00345-3

47. Paul RK, Garai S. Wavelets based artificial neural network technique for forecasting agricultural prices. J Indian Soc Probab Stat. 2022;23 (1):47-61. https://doi.org/10.1007/s41096-022-00128-3

Additional information

Peer review: Publisher thanks Sectional Editor and the other anonymous reviewers for their contribution to the peer review of this work.

Reprints & permissions information is available at https://horizonepublishing.com/journals/index.php/PST/open_access_policy

Publisher's Note: Horizon e-Publishing Group remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Indexing: Plant Science Today, published by Horizon e-Publishing Group, is covered by Scopus, Web of Science, BIOSIS Previews, Clarivate Analytics, NAAS, UGC Care, etc
See https://horizonepublishing.com/journals/index.php/PST/indexing_abstracting

Publisher information: Plant Science Today is published by HORIZON e-Publishing Group with support from Empirion Publishers Private Limited, Thiruvananthapuram, India.