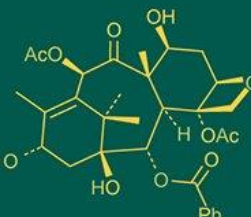
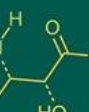
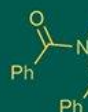
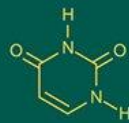


International Journal of Advanced Biochemistry Research



ISSN Print: 2617-4693
ISSN Online: 2617-4707
NAAS Rating (2026): 5.29
IJABR 2026; SP-10(7): 452-457
www.biochemjournal.com
Received: 04-05-2026
Accepted: 07-06-2026

Sohilali R Saiyad

Ph.D. Scholar, Department of
Agricultural Statistics, B. A.
College of Agriculture, Anand
Agricultural University
(AAU), Anand, Gujarat, India

AD Kalola

Professor, Department of
Agricultural Statistics, B. A.
College of Agriculture, Anand
Agricultural University
(AAU), Anand, Gujarat, India

Uttamkumar S Baladaniya

Ph.D. Scholar, Department of
Agricultural Statistics, B. A.
College of Agriculture, Anand
Agricultural University
(AAU), Anand, Gujarat, India

Corresponding Author:**Sohilali R Saiyad**

Ph.D. Scholar, Department of
Agricultural Statistics, B. A.
College of Agriculture, Anand
Agricultural University
(AAU), Anand, Gujarat, India

Comparative evaluation of ARIMA and machine learning models for paddy price forecasting in Gujarat

Sohilali R Saiyad, AD Kalola and Uttamkumar S Baladaniya

DOI: <https://www.doi.org/10.33545/26174693.2026.v10.i7Sf.9018>

Abstract

Paddy (*Oryza sativa* L.) is the most critical staple food crop in India, with production of 150,180 thousand tonnes from 51,270 thousand hectares at a yield of 2,929 kg/ha during 2024-25. Price volatility of paddy is driven by seasonal supply cycles, weather-induced yield shocks, government procurement policies and macroeconomic forces, making reliable price forecasting essential for farmers, traders and policymakers. This study applied statistical and machine learning techniques to forecast weekly paddy prices in two major Gujarat markets Khambhat (2009-2023) and Dahod (2002-2023) sourced from the Agmarknet portal. Four models were developed and compared: ARIMA, ANN, RNN and a Hybrid (ARIMA-ANN) model. Stationarity was confirmed using the Augmented Dickey-Fuller (ADF) test after first-order differencing, and nonlinearity was verified using the Brock-Dechert-Scheinkman (BDS) test. ARIMA (1,1,2) with drift was selected for Khambhat and ARIMA (0,1,1) without drift for Dahod. ANN achieved MAPE of 5.09% (Khambhat) and 2.74% (Dahod), outperforming ARIMA significantly. The RNN model recorded the best performance with MAPE of 3.20% and 2.07% for Khambhat and Dahod respectively, surpassing ANN, Hybrid and ARIMA models. The Hybrid ARIMA-ANN improved over standalone ARIMA but could not match ANN or RNN. Overall model ranking was RNN > ANN > Hybrid > ARIMA across both markets. The RNN model was used to generate 52-week forecasts for 2024, with PFE within $\pm 10\%$ for most weeks in both markets, confirming its practical utility for paddy price forecasting in Gujarat.

Keywords: Forecasting, ARIMA, ANN, RNN, Hybrid model, Agmarknet, Gujarat, Time series, BDS test, MAPE

Introduction

Cereals play an important role in Indian agriculture, constituting the principal component of the human diet for thousands of years. Rice, wheat, maize and to a lesser extent jowar and millets constitute the major cereals, with more than 50 per cent of the world's daily caloric intake derived directly from cereal grain consumption. India is the second-largest producer of rice in the world with paddy production reaching approximately 150,180 thousand tonnes from an area of 51,270 thousand hectares at a yield of 2,929 kg per hectare during 2024-25 (Anon., 2025a). In Gujarat, paddy occupies an area of 9,100 thousand hectares with a production of 22,320 thousand tonnes and a yield of 2,453 kg per hectare during 2024-25 (Anon., 2025b). Agricultural price forecasts are meant to assist farmers, governments and agribusiness industries in making decisions that influence producers, consumers and policymakers. As a result, policymakers, traders and farmers are concerned about the ability to reliably estimate the prices of cereal crops. Price volatility of paddy crop is high owing to various factors including seasonal supply cycles, weather-induced yield shocks, government procurement policies and macroeconomic forces. To overcome these issues, the present study was undertaken to develop appropriate models for forecasting weekly paddy price series in major markets of Gujarat. The data-driven approach of machine learning models, including ANN and RNN, offers a feasible price forecasting technique when the underlying relationship is nonlinear and complex (Harshith and Adhikari, 2023) ^[4].

2. Materials and Methodology

2.1 Data: The data on weekly prices (₹/quintal) of paddy crop of Gujarat were collected from the official Agmarknet portal (<https://agmarknet.gov.in/>) of the Ministry of Agriculture and

Farmers Welfare, Government of India (Anon., 2023). Daily prices were converted into weekly averages. Markets were selected based on highest arrivals and data availability as presented in Table 1. The data were split into 80 per cent for training and 20 per cent for testing.

Table 1: Data availability for the paddy crop in selected Market

Crop	Market	Data Availability
Paddy	Khambhat*	2009 to 2023
	Dahod	2002 to 2023

*excluding 2011, 2012, 2013 and 2022

2.2 Data Preprocessing

The daily price data were cleaned and converted into weekly average prices for analysis. Missing values were imputed using a rolling average of the previous five days. Input variables were scaled to the range (0, 1) using Min Max Scaler from the Scikit-learn library before training neural network models, and forecasted outputs were inverse-transformed to the original scale after prediction.

2.3 Stationarity Test

The time series was said to be stationary if its mean and variance remain constant over time. The Augmented Dickey-Fuller (ADF) test given by Dickey and Fuller (1981) [5] was used to assess stationarity. The regression equation estimated is:

$$\Delta Y_t = \alpha_1 + \alpha_2 t + \delta Y_{t-1} + \sum_{i=1}^h \beta_i Y_{t-i} + \varepsilon_t \quad (1)$$

Where, $\Delta Y_t = Y_t - Y_{t-1}$, α_1 , α_2 and β_i are parameters of the regression model, h is lag length and $\delta = \rho - 1$ and $-1 \leq \rho \leq 1$

- **H0:** Time series is nonstationary
- **H1:** Time series is stationary

If the p-value from the ADF test falls below the significance level (0.05), H_0 is rejected and the series is declared stationary. Differencing was applied to achieve stationarity where required.

2.4 Nonlinearity Test

Before applying machine learning models, the presence of nonlinearity in the residuals was verified using the Brock-Dechert-Scheinkman (BDS) test (Brock *et al.*, 1996) [6]. The BDS statistic is based on the correlation integral:

$$C_{m,T}(\epsilon) = \sum_{t < s} I_{\epsilon} (X_t^m, X_s^m) \left\{ \frac{2}{T_m(T_m-1)} \right\} \dots \quad (2)$$

Where, $T_m = T - (m - 1)$ and $I_{\epsilon} X_t^m, X_s^m$ is an indicator function which equals 1 if the sup norm $\|X_t^m - X_s^m\| < \epsilon$ and equals 0 otherwise. The BDS statistic is:

$$V_{m,\epsilon} = \sqrt{T} \frac{C_{m,T}(\epsilon) - C_{1,T}(\epsilon)}{S_{m,T}} \dots \quad (3)$$

Where, $S_{m,T}$ is the standard deviation. Under fairly moderate regularity conditions, the BDS statistic converges in distribution to $N(0, 1)$ (Bisaglia and Gerolimetto, 2014) [7].

- **H0:** Time series is linear
- **H1:** Time series is nonlinear
- **H0:** Time series is linear
- **H1:** Time series is nonlinear

2.5 ARIMA Model

The future value of a variable is assumed to be a linear function of past observations and random errors. The ARIMA (p, d, q) model is expressed as:

$$y_t = \alpha + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$$

Where, y_t and ε_t are the actual value and random error at time t respectively; ϕ_i ($i = 1, \dots, p$) and θ_j ($j = 1, \dots, q$) are model parameters; p is the AR order; d is the order of differencing; and q is the MA order. The Box-Jenkins methodology was followed in five steps:

1. Stationarity testing using the ADF test.
2. Model identification using ACF and PACF plots.
3. Parameter estimation by Maximum Likelihood Estimation (MLE).
4. Diagnostic checking using AIC, BIC and the Ljung-Box test.
5. Forecasting using the best-fit model.

The most suitable model was selected using the smallest Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC):

$$AIC = -2 \log L + 2m \dots \quad (5)$$

$$BIC = \log \sigma^2 + (m \log n) / n \dots \quad (6)$$

Where, $m = p + q$, L is the likelihood function, σ^2 is error variance and n is the number of observations. Adequacy of the fitted model was confirmed through the Ljung-Box Q^* statistic:

$$Q^* n (n + 2) \sum_{k=1}^h (n - k) - 1 r_k^2 \dots \quad (7)$$

which follows a chi-square distribution with $(h - m)$ degrees of freedom. Non-significance of Q^* confirms that residuals are white noise.

2.6 Artificial Neural Network (ANN) Model

ANN is a data-driven nonlinear model that learns complex patterns directly from data without requiring prior specification of functional form. The ANN output is computed as:

$$y_t = a_0 + \sum_{j=1}^q a_j g (\beta_{0j} + \sum_{i=1}^k \beta_{ij} y_{t-i}) + \varepsilon_t \dots \quad (8)$$

Where, y_t is the output; y_{t-i} ($i = 1, \dots, k$) are the k lagged inputs; α_j and β_{ij} are the connection weights; g is the activation function; and ε_t is the random error. In this study, 13 lagged weekly prices were used as inputs. The Rectified Linear Unit (ReLU) activation function was employed in the hidden layers:

$$f(x) = 0 \text{ when } x < 0; f(x) = x \text{ when } x \geq 0 \dots \quad (9)$$

A linear activation function was used in the output layer. Models were optimised using the Adaptive Moment Estimation (Adam) algorithm with Mean Squared Error (MSE) as the loss function. The optimal architecture (number of layers, neurons, learning rate, batch size and epochs) was determined by systematic hyperparameter tuning.

2.7 Recurrent Neural Network (RNN) Model

Unlike feedforward networks, RNNs include recurrent connections that allow information to be retained across time steps, making them suited to sequential price data. The hidden state at each time step is updated as:

$$h_t = \sigma_h (W_{xh} x_t + W_{hh} h_{t-1} + b_h) \dots \quad (10)$$

Where, x_t is the input vector, h_t is the hidden state, W_{xh} and W_{hh} are weight matrices, b_h is the bias vector and σ_h is the activation function. The output is given by:

$$y_t = \sigma_y (W_{hy} h_t + b_y) \dots \quad (11)$$

In this study, two SimpleRNN layers with 64 units each were used, followed by a dense output layer with linear activation. Data were reshaped into the format (batch size, time steps, features) required by the RNN architecture. The Adam optimiser and MSE loss function were applied.

2.8 Hybrid Model

Real-world price series typically contain both linear and nonlinear patterns. Since neither ARIMA nor ANN alone models both adequately, a hybrid ARIMA-ANN model was used. The time series is decomposed as:

$$y_t = L_t + N_t \dots \quad (12)$$

Where, L_t is the linear component modelled by ARIMA and N_t is the nonlinear component. The residuals from the ARIMA model are:

$$e_t = y_t - \hat{L}_t \dots \quad (13)$$

These residuals, confirmed to be nonlinear by the BDS test, are then modelled using an ANN:

$$e_t = f(e_{t-1} e_{t-2} \dots, e_{t-n}) + \varepsilon_t \dots \quad (14)$$

The combined forecast is obtained as:

$$\hat{y}_t = \hat{L}_t + \hat{N}_t \dots \quad (15)$$

This two-step approach exploits the linear modelling strength of ARIMA and the nonlinear modelling capability of ANN simultaneously.

2.9 Forecast Evaluation Methods

The forecasting performance of all models was compared using four error criteria:

2.9.1 Mean Square Error

$$MSE = \frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2 \dots \quad (16)$$

2.9.2 Root Mean Square Error

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2} \dots \quad (17)$$

2.9.3 Mean Absolute Error

$$MAE = \frac{1}{n} \sum_{t=1}^n |Y_t - \hat{Y}_t| \dots \quad (18)$$

2.9.4 Mean Absolute Percentage Error

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \times 100 \dots \quad (19)$$

Where, Y_t is the actual value, $\hat{Y}_t$ is the forecasted value and n is the number of observations in the test set. Lower values of all four criteria indicate better forecasting performance.

3. Results and discussion

3.1 Data

The data of weekly wholesale price (₹/quintal) for paddy crop of Gujarat were collected from the official Agmarknet portal of the Ministry of Agriculture and Farmers Welfare, Government of India (Anon., 2023). Paddy prices were collected for two markets, Khambhat (2009 to 2023, excluding 2011, 2012, 2013 and 2022) and Dahod (2002 to 2023). Daily prices were converted to weekly averages and missing values were filled using a five-day rolling average. The dataset was divided into 80 per cent for training and 20 per cent for testing.

3.2 Discussion

Visualisation of the time series data was the foremost process in the analysis. Time plots of weekly paddy prices for both markets showed a consistent upward trend over the study period, reflecting the influence of rising input costs, government procurement prices and market-level demand pressures. Descriptive statistics of the price series used in the experiments are presented in Table 2.

Table 2: Descriptive statistics of weekly paddy price series

Statistic	Khambhat	Dahod
Observations	572	1144
Mean (₹/q)	1567.23	1275.15
Median (₹/q)	1564.17	1363.75
Maximum (₹/q)	2336.50	2300.00
Minimum (₹/q)	940.00	640.00
Std. Deviation (₹/q)	236.45	402.93
Skewness	0.42	-0.03
Kurtosis	0.99	-1.10

The Augmented Dickey-Fuller (ADF) test was used to test stationarity of the price series. Results presented in Table 3 revealed that original paddy price series of both markets were non-stationary; after first-order differencing each series became stationary.

Table 3: ADF stationarity test results for paddy price series

Market	Series	t-statistic	p-value	Conclusion
Khambhat	Original	-1.67	0.45	Non-stationary
	1st Difference	-26.79	<0.01	Stationary
Dahod	Original	-1.05	0.73	Non-stationary
	1st Difference	-10.62	<0.01	Stationary

The BDS nonlinearity test was applied to each series. Results presented in Table 4 showed that for Khambhat, the null hypothesis of linearity could not be rejected at 0.5σ but was rejected at 1.0σ , 1.5σ and 2.0σ ($p < 0.001$) for both embedding dimensions. For Dahod, the null hypothesis was rejected at all epsilon levels (0.5σ to 2.0σ) with p-values below 0.001. This confirmed significant nonlinear patterns in both markets, justifying the use of machine learning models.

Table 4: BDS nonlinearity test results for paddy price series

Market	Epsilon	Dim. 2 Stat.	p-value	Dim. 3 Stat.	p-value	Conclusion
Khambhat	0.5σ	1.72	0.09	1.82	0.07	Linear
	1.0σ	2.01	<0.001	2.34	<0.001	
	1.5σ	2.35	<0.001	2.64	<0.001	
	2.0σ	2.89	<0.001	3.47	<0.001	
Dahod	0.5σ	10.49	<0.001	12.72	<0.001	Non-linear
	1.0σ	9.28	<0.001	10.29	<0.001	
	1.5σ	6.80	<0.001	7.15	<0.001	
	2.0σ	5.40	<0.001	5.89	<0.001	

3.3 ARIMA Model Results

After visualizing the ACF-PACF plots and confirming stationarity through the ADF test, various ARIMA (p, d, q) structures with and without drift were fitted and compared using AIC and BIC values. The possible ARIMA structures

for both paddy markets are presented in Tables 5 and 6. Based on the lowest AIC and BIC values, the best ARIMA models selected were ARIMA (1, 1, 2) with drift for Khambhat and ARIMA (0, 1, 1) without drift for Dahod.

Table 5: ARIMA model selection for paddy price in Khambhat market

Sr. No.	Model	AIC	BIC	RMSE	MAPE
With Drift					
1	ARIMA (0, 1, 1)	4823.07	4835.44	369.85	13.99
2	ARIMA (1, 1, 0)	4823.32	4835.69	370.47	14.01
3	ARIMA (2, 1, 0)	4824.95	4841.44	370.49	14.01
4	ARIMA (0, 1, 2)	4825.06	4841.55	369.81	13.99
5	ARIMA (1, 1, 1)	4825.07	4841.56	369.76	13.99
6	ARIMA (1, 1, 2)*	4827.02	4847.63	369.26	13.97
7	ARIMA (2, 1, 1)	4827.25	4847.87	370.13	14.00

* Selected model (RMSE 369.26; MAE 276.42; MAPE 13.97; Ljung-Box Q*: stat = 1.22, p = 0.99)

Table 6: ARIMA model selection for paddy price in Dahod market

Sr. No.	Model	AIC	BIC	RMSE	MAPE
Without Drift					
1	ARIMA (0, 1, 1)*	8838.06	8847.70	232.02	11.38
2	ARIMA (1, 1, 0)	8838.50	8848.14	232.12	11.39
3	ARIMA (1, 1, 1)	8840.04	8854.49	231.99	11.38
4	ARIMA (0, 1, 2)	8840.04	8854.49	232.00	11.38
5	ARIMA (2, 1, 0)	8840.14	8854.59	231.98	11.38
6	ARIMA (2, 1, 1)	8841.97	8861.24	232.03	11.39
7	ARIMA (1, 1, 2)	8842.05	8861.32	232.00	11.38

* Selected model (RMSE 232.02; MAE 187.74; MAPE 11.38; Ljung-Box Q*: stat = 16.93, p = 0.07)

The BDS test applied to ARIMA residuals of both markets confirmed significant nonlinearity ($p < 0.001$) at all embedding dimensions, indicated that the ARIMA residuals contained nonlinear patterns not captured by the linear model. This provided the justification for proceeding with the neural network-based models.

3.4 ANN Model Results: The ANN models for both paddy markets used 13 lagged weekly prices as inputs. The ReLU activation function was used in the hidden layers and linear activation in the output layer. The Adam optimiser with MSE loss function was employed. The optimal architectures determined by hyperparameter tuning are presented in Table 7.

Table 7: ANN model architecture and performance for paddy markets

Market	Hidden Layer (neurons)	Learning Rate	Batch / Epochs	R ²	RMSE	MAE	MAPE	MSE
Khambhat	2 layers × 64	0.001	8 / 100	51.60	135.00	124.63	5.09	18225.00
Dahod	1 layer × 16	0.001	8 / 100	91.30	67.42	49.31	2.74	4545.45

The ANN model substantially outperformed the ARIMA model in both paddy markets, with MAPE values of 5.09 per cent (Khambhat) and 2.74 per cent (Dahod) compared to ARIMA MAPE values of 13.97 and 11.38 per cent respectively. The lower explanatory power of the Khambhat ANN model relative to its MAPE performance. This confirmed the structural advantage of neural network models in capturing complex nonlinear price dynamics, consistent with findings of Jha and Sinha (2013) and Suthar *et al.* (2023) [8, 9, 11].

3.5 RNN Model Results

The RNN models were developed using two SimpleRNN layers with 64 units each, followed by a dense output layer with linear activation. The input tensor was structured as (batch size, 4 time steps, 1 feature). The Adam optimiser with MSE loss function was applied. The architecture and performance of the RNN models are summarised in Table 8.

Table 8: RNN model architecture and performance for paddy markets

Market	Architecture	Total Params	Batch / Epochs	R ²	RMSE	MAE	MAPE	MSE
Khambhat	2×SimpleRNN(64)+Dense(1)	25092	8 / 150	94.21	91.95	62.56	3.20	8454.80
Dahod	2×SimpleRNN(64)+Dense(1)	25092	8 / 200	93.69	57.31	37.78	2.07	3284.44

The RNN model achieved the lowest error values across both paddy markets, recording MAPE values of 3.20 and 2.07 per cent for Khambhat and Dahod respectively. The internal memory mechanism of the RNN, which retains information from previous time steps through the hidden state, enabled superior capture of temporal dependencies in the sequential weekly price data. Similar findings were reported by Kumari *et al.* (2023) for banana prices in Gujarat and Ghetiya *et al.* (2024)^[10] for rice prices in West Bengal, who demonstrated that^[12]

RNNs with proper preprocessing and hyperparameter tuning consistently outperform conventional statistical models.

3.6 Hybrid ARIMA-ANN Model Results: After confirming nonlinearity in the ARIMA residuals through the BDS test, the hybrid ARIMA-ANN models were developed. The ARIMA residuals were used as inputs to an ANN to capture the remaining nonlinear patterns. The final hybrid forecast was obtained by combining the ARIMA and ANN predictions. Performance of the hybrid models is presented in Table 9.

Table 9: Hybrid ARIMA-ANN model architecture and performance for paddy markets

Market	ANN on Residuals	Epochs	R ²	RMSE	MAE	MAPE	MSE	ARIMA Model
Khambhat	1 layer × 8 neurons (ReLU)	50	87.50	265.69	236.05	9.58	70591.18	ARIMA (1,1,2)+drift
Dahod	1 layer × 32 neurons (ReLU)	50	84.65	89.52	70.95	4.00	8013.83	ARIMA (0,1,1)+without

The hybrid ARIMA-ANN model outperformed standalone ARIMA across both paddy markets, with RMSE values reduced to nearly half those of ARIMA. However, the hybrid model did not surpass the standalone ANN or RNN models. Similar findings were reported by Sinha (2011)^[11], he observed that the hybrid ARIMA-ANN model improved upon ARIMA but did not outperform the standalone ANN model,

particularly when the linear and nonlinear structures interact multiplicatively rather than additively.

3.7 Comparison of ARIMA, ANN, RNN and Hybrid Models: A comprehensive comparison of the forecasting performance of all four models for each paddy market is presented in Table 10. The best model for each market is identified on the basis of the lowest RMSE, MAE, MAPE and MSE values.

Table 10: Comparison of ARIMA, ANN, RNN and Hybrid models for paddy price forecasting

Market	Model	RMSE	MAE	MAPE (%)	MSE
Khambhat	ARIMA (1,1,2) with drift	369.26	276.42	13.97	136352.75
	ANN	135.00	124.63	5.09	18225.00
	RNN*	91.95	62.56	3.20	8454.80
	Hybrid	265.69	236.05	9.58	70591.18
Dahod	ARIMA (0,1,1) without drift	232.02	187.74	11.38	52909.20
	ANN	67.42	49.32	2.74	4545.45
	RNN*	57.31	37.79	2.07	3284.44
	Hybrid	89.52	70.95	4.00	8013.83

* Best performing model

Results revealed that the RNN model had the least RMSE, MAE, MAPE and MSE values for both paddy markets, indicating superior forecasting efficiency. The ANN model ranked second, the Hybrid ARIMA-ANN third and the ARIMA model recorded the highest error values across both markets. These findings are consistent with those of Ghetiya *et al.* (2024) and Kumari *et al.* (2023)^[10, 12], who reported that RNN outperformed both ANN and classical models for agricultural price forecasting.

3.8 Weekly Paddy Price Forecasting for 2024
Following model evaluation, the best-performing RNN model was employed in recursive multi-step forecasting mode to generate 52-week price projections for 2024 for both paddy markets. A representative set of forecasted and actual prices along with the Percentage Forecast Error (PFE) for both markets of paddy crop is presented in Table 11.

Table 11: Weekly paddy price forecast (representative weeks), Khambhat and Dahod market in 2024

Week	Khambhat			Dahod		
	Actual Price (₹/q)	Forecasted Price (₹/q)	PFE (%)	Actual Price (₹/q)	Forecasted Price (₹/q)	PFE (%)
1	2975.00	3058.38	2.80	1875.00	1851.87	-1.23
5	3000.00	2803.13	-6.56	1955.00	1943.54	-0.59
10	3000.00	2759.30	-8.02	2093.33	2198.69	5.03
20	2500.00	2286.01	-8.56	2081.25	2258.44	8.51

30	3250.00	3133.01	-3.60	1942.50	1853.36	-4.59
40	3250.00	2854.57	-12.17	1774.17	1640.99	-7.51
50	3000.00	2718.85	-9.37	1937.50	1739.03	-10.24
52	3000.00	3097.28	3.24	1935.83	1803.94	-6.81

The RNN model provided satisfactory forecasts of paddy prices across both markets during 2024. In Khambhat, PFE values ranged from -13.97 to +13.80 per cent, with 15 of 52 weeks exceeding ± 10 per cent. Dahod recorded PFE values between -10.24 and +10.06 per cent, with only 3 weeks exceeding ± 10 per cent. Despite some week-to-week deviations associated with recursive multi-step forecasting, the model effectively captured the seasonal fluctuations and overall price trend in both markets. These results suggest that the RNN model is most effective for forecasting weekly paddy prices in Gujarat markets with satisfactory practical accuracy for use by farmers, traders and policy makers.

4. Conclusion

This study evaluated ARIMA, ANN, RNN and Hybrid ARIMA-ANN models for weekly paddy price forecasting in the Khambhat and Dahod markets of Gujarat. ADF test confirmed non-stationarity in the original series, which was resolved by first-order differencing. The BDS test confirmed nonlinear structure in both series, justifying machine learning approaches. Among ARIMA models, ARIMA (1,1,2) with drift (Khambhat) and ARIMA (0,1,1) without drift (Dahod) were selected based on AIC and BIC. The ANN model outperformed ARIMA with MAPE values below 6% for both markets. The RNN model was the best performer overall, achieving the lowest RMSE, MAE, MAPE and MSE in both markets, due to its ability to capture temporal dependencies through internal memory. The Hybrid ARIMA-ANN outperformed standalone ARIMA but ranked below ANN and RNN. The 52-week RNN forecasts for 2024 showed satisfactory accuracy, with minimal deviations for Dahod and acceptable errors for Khambhat. These results confirm that RNN is the most effective technique for short-term paddy price forecasting, providing actionable and reliable price intelligence for farmers, agribusiness traders and government policymakers in Gujarat.

References

1. Anonymous. Agriculture Marketing Website. 2023. Agmarknet. <http://agmarknet.gov.in>.
2. Anonymous. Ministry of Agriculture of India. 2025. IndiaStatAgri. <https://www.indiastatagri.com>.
3. Anonymous. Directorate of Agriculture, Government of Gujarat. 2025. <https://dag.gujarat.gov.in>.
4. Harshith N, Adhikari V. Cracking the code of insect communication: a machine learning approach in acoustic entomology. *Krishi Science*. 2023;4(4):81-85.
5. Dickey DA, Fuller WA. Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica*. 1981;49(4):1057-1072.
6. Brock W, Dechert W, Scheinkman J. A test for independence based on the correlation dimension. *Econom Rev*. 1996;15:197-235.
7. Bisaglia L, Gerolimetto M. Testing for (non) linearity in economic time series: a Monte Carlo comparison. Working Paper Series, Department of Statistical Sciences, University of Padua. 2014;3.
8. Jha GK, Sinha K. Agricultural price forecasting using neural network model: an innovative information delivery system. *Agric Econ Res Rev*. 2013;26(2):229-239.
9. Suthar B, Pundir RS, Popat R. Groundnut price forecasting in Gondal market of Gujarat: a comparison of ARIMA and ANN models. *Indian J Agric Mark*. 2023;37(1):51-60.
10. Kumari P, Goswami V, Harshith N, Pundir RS. Recurrent neural network architecture for forecasting banana prices in Gujarat, India. *PLoS One*. 2023;18(6):e0275702.
11. Sinha K. A study on combining ARIMA and artificial neural networks for time series forecasting [MSc (Agriculture) thesis]. New Delhi: Indian Agricultural Statistics Research Institute; 2011.
12. Ghetiya RL, Chandan GY, Kumari P, Nihala ACK, Pansuriya A. Price forecasting models for rice in West Bengal, India. *Int J Res Agron*. 2024;7(11):183-188.