Indian Journal of Agricultural Sciences **95** (10): 1260–1266, October 2025/Article
https://doi.org/10.56093/ijas.v95i10.163270

Forecasting jute (*Corchorus* spp.) prices and arrivals in West Bengal using ARIMA and advanced machine learning techniques

SOURAV GHOSH[1], N M ALAM[1]*, DEBARATI DATTA[1], SONALI PAUL MAZUMDAR[1], BIJAN MAJUMDAR[1], SABYASACHI MITRA[1], RITESH SAHA[1], SANJOY SAHA[1] and GOURANGA KAR[1]

ICAR-Central Research Institute for Jute and Allied Fibres, Barrackpore, West Bengal 700 121, India

Received: 04 January 2025; Accepted: 02 September 2025

ABSTRACT

The study aimed to forecast and understand the price movements and volatility in jute (*Corchorus* spp.) prices and arrivals from 2007–2023 across the markets of four major jute growing districts in West Bengal namely Uttar Dinajpur, Nadia, Coochbehar, and Murshidabad. To achieve this, both traditional statistical methods (ARIMA) and advanced machine learning models, including Artificial Neural Networks (ANN), Support Vector Machine (SVM), Random Forest (RF), and Long Short-Term Memory (LSTM), were employed. The model performance was evaluated using error metrics such as Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE) and Mean Absolute Error (MAE). The LSTM model outperformed the others, with an average RMSE of 12.53 and a MAPE of 4.87%, demonstrating its superior accuracy in forecasting jute prices. In the case of jute arrivals, LSTM and SVR achieved the best performance, with LSTM recording the lowest RMSE of 14.32 and MAE of 11.87 in predicting arrivals. This study is among the first to apply the Long Short-Term Memory (LSTM) model, a specialized deep learning technique, along with hybrid statistical ML models, to jute market forecasting particularly for arrivals, an area that has received limited attention in previous literature-leveraging LSTM's capability to accurately capture complex non-linear patterns. The findings underscore the non-linearity and interdependencies between markets, providing critical insights for traders and policymakers. These insights enable more precise anticipation of price fluctuations and contribute to better-informed market strategies, ultimately benefiting the jute supply chain in West Bengal.

Keywords: ARIMA model, Jute price forecasting, Long short-term memory, Machine learning models, Random forest, Support vector regression

Jute (*Corchorus* spp.) holds significant economic and agricultural importance in India, serving as a primary source of fiber for various industries. India is the largest producer of jute globally, with approximately 0.6 million hectares under cultivation and an annual production of 1.61 million tonnes (Anonymous 2023). West Bengal is the leading state in jute cultivation, accounting for 0.5 million hectares and producing 1.3 million tonnes, which represents a substantial portion of the national output (Anonymous 2023). Jute occupies a unique position as an eco-friendly, biodegradable, and renewable natural fiber with substantial domestic market and export potential generating foreign exchange worth ₹18.2 billion (Anonymous 2021). Approximately, 4 million farmer families are involved in jute cultivation across India, covering an area of about 6.52 lakh hectares (Jute Corporation of India 2024). The jute industry's significance extends beyond fiber production; the entire plant is utilized, contributing to environmental sustainability and rural livelihoods (Rahman *et al.* 2017).

Despite the advantages, the jute industry faces challenges, including price instability and competition from synthetic fibres. Price instability is influenced by several factors, such as annual production variations, low price elasticity of demand, seasonality, market inefficiencies, weak supply chains, and market monopolies (Ghosh *et al.* 2020, Kumar *et al.* 2023). The inherent volatility of agricultural prices complicates forecasting efforts, as these time series often exhibit complex spatial, linear, and non-linear patterns (Saha *et al.* 2020). Traditional time series forecasting techniques such as the Autoregressive Integrated Moving Average (ARIMA) model, have been widely applied to capture linear relationships in price data. These models excel in capturing linear relationships within time series data but often struggle with non-linear patterns, which can lead to sub-optimal forecasts (Zhang 2003). With the emergence of advanced machine learning (ML) methodologies, including Artificial Neural Networks

[1]ICAR-Central Research Institute for Jute and Allied Fibres, Barrackpore, West Bengal. *Corresponding author email: alam.nurnabi@gmail.com

(ANN), Support Vector Regression (SVR), Random Forest (RF), and Long Short-Term Memory (LSTM) networks, researchers are now better equipped to address the non-linear and volatile nature of these series inherent in agricultural price dynamics (Jha and Sinha 2013, Nayak *et al.* 2013, Purohit *et al.* 2021, Shankar *et al.* 2024). However, it has been observed that applying linear models to non-linear data, or vice versa, can yield poor forecasting results (Khashei and Bijari 2011).

Recognizing the potential of machine learning models and deep learning techniques offers a promising solution for capturing the underlying complexities of agricultural price movements. Zhang (2003) pioneered the concept of forecasting by decomposing time series into linear and nonlinear components, applying ARIMA to capture the linear dynamics, and subsequently using ANN to model the residuals. This approach has been refined over time, with various studies demonstrating improved forecasting accuracy through both liner and machine learning models. For instance, Khashei and Bijari (2011) and Wang *et al.* (2013) presented variations that incorporate both ARIMA and ANN, enhancing predictive performance across diverse datasets. However, challenges remain, particularly in ensuring the models' suitability for non-Gaussian time series, which is common in agricultural price data. This study aims to forecast the jute prices and arrivals using ARIMA and advanced machine learning techniques such as ANN, SVR, RF and LSTM in four major markets: Uttar Dinajpur (North Dinajpur), Nadia, Coochbehar, and Murshidabad. Understanding the transmission of volatility among these markets is essential for formulating effective price policies and ensuring market stability.

MATERIALS AND METHODS

Data collection: In this study, time series data (2007–2023) of jute price (₹/q) and arrival quantity (tonnes) from major markets of four major districts of West Bengal, namely Uttar Dinajpur, Nadia, Coochbehar, and Murshidabad were collected from AGMARKNET website (https://agmarknet.gov.in). The dataset includes monthly data points, capturing both short-term fluctuations and long-term trends. Out of a total of 204 data points (January 2007 to December 2023), 198 points (January 2007–June 2023) were used for model building, and the remaining 6 points (July 2023 to December 2023) were used for validation.

Methodology: To effectively forecast jute prices, we employed a multi-stage modelling approach that integrates both linear and non-linear modelling techniques. Initially, ARIMA model were used to capture linear dependencies in the jute price series (Shankar *et al.* 2023b). To address non-linear patterns not captured by the ARIMA model, we applied a Support Vector Regression model, which excels in handling non-linear relationships by transforming input data into a higher-dimensional space using a kernel function. Additionally, Random Forest was utilized for its robustness and ability to handle complex, non-linear relationships through ensemble learning. Artificial Neural Networks were employed to model any remaining non-linear dependencies. The ANN architecture was carefully designed to optimize performance through experimentation (Shankar *et al.* 2023a). Long Short-Term Memory networks were used for capturing long-term dependencies. LSTM effectively addressed long-range dependencies and vanishing gradient issues through their sophisticated memory cell structure and gating mechanisms. The final forecast was generated by combining the outputs of these models such as ARIMA for linear components while ANN, SVR, RF and LSTM for non-linear components (Kumar *et al.* 2020). The accuracy of the combined forecast was evaluated using the reserved validation dataset (July 2023 to December 2023). Performance metrics such as Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE) and Mean Absolute Error (MAE) were employed to assess the forecasting accuracy. The detailed steps of the methodology are as follows:

ARIMA: ARIMA models are used to capture linear relationships in time series data. The ARIMA (p, d, q) model, where p is the order of autoregression, d is the degree of differencing, and q is the order of the moving average, was selected based on the Partial Autocorrelation Function (PACF) and Autocorrelation Function (ACF) plots. These plots help to determine the values of p and q, while differencing ensures data stationarity (d). The parameters were estimated using the maximum likelihood method, with the best model chosen based on the lowest Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values. The model was verified to ensure that the residuals exhibited white noise, indicating homoscedastic residuals with a mean of zero. The ARIMA model is given by:

$$\varphi(L)\Delta^{d} y_t = \theta(L)\varepsilon_t \tag{1}$$

where, φ(L) and θ(L) were the lag operators L of AR and MA polynomials with order p and q, Δ^d is difference operator.

Artificial neural network: Artificial Neural Network was used to model the non-linear residuals from the ARIMA model. ANNs consist of interconnected nodes organized in layers; input, hidden, and output layers. The hidden layers are crucial for capturing complex relationships within the data.

$$y_t + 1 = g[\sum_{j=0}^{g} \alpha_j f(\sum_{i=0}^{g} \alpha_{(i=0)}^{p} \beta_{ij}\, y_{t-i})] \tag{2}$$

Where $y_t + 1$, Observation at time t+1; f and g, Activation functions at hidden and output layer; p, Number of input nodes; q, Number of hidden nodes; β_{ij}, Weight attached to the connection between its input.

Support vector regression: SVR was employed to capture non-linear relationships within the dataset. By mapping input data into a higher-dimensional space using a kernel function, SVR can identify complex patterns that linear models might miss. The SVR model minimizes prediction errors by balancing model complexity and predictive accuracy. The model is given by:

$$f(x) = w^T \phi(x) + b \tag{3}$$

Where w, Weights vector; x, Input; b, Bias; $\phi(x)$, Non-linear mapping function i.e. kernel function.

Random forest: Random Forest is an ensemble learning technique that generates multiple decision trees using bootstrap sampling and random feature selection. Each tree is trained on different subsets of the data, ensuring diversity among the trees. For regression tasks, the predictions from all trees are averaged to enhance accuracy and reduce overfitting. This method captures complex patterns within the data by leveraging the strengths of multiple decision trees.

Long short-term memory: LSTM is a type of Recurrent Neural Network (RNN) designed to overcome the limitations of traditional RNNs, particularly in capturing long-range dependencies and mitigating the vanishing gradient problem. LSTM networks have a memory cell structure that allows them to selectively remember or forget information over time, managed by three gates: the forget gate (f_t), the input gate (i_t), and the output gate (O_t). The LSTM model can be expressed as:

$$f_t = \sigma\,[W_f * (h_{t-1}, x_t) + b_f];\ i_t = \sigma\,[W_i * (h_{t-1}, x_t) + b_i];\ O_t = \sigma\,[W_o * (h_{t-1}, x_t) + b_o]$$

Where W_f, W_i and W_o are weight matrices for the forget, input, and output gates, respectively; b_f, b_i and b_o are bias terms for the gates; h_{t-1} is the previous hidden state; x_t is the current input, σ is the activation function (usually sigmoid or tanh). LSTM networks are particularly effective in modelling time series data with complex temporal dynamics.

Model evaluation: For selecting the most effective time series model for a dataset, different error metrics are used to compare the predicted values and the actual values. Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE) and Mean Absolute Error (MAE) are used in this study. Let Y_t is the actual values, Y_t is the fitted value and n is the number of observations.

$$\text{RMSE} = \left(\sqrt{\sum_{t=1}^{n} \left\{ \frac{Y_t - \hat{Y}_t}{Y_t} \right\}^2 / _n} \right) \quad (4)$$

$$\text{MAPE} = \left(\sum_{t=1}^{n} \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right|^2 / _n \right) \times 100 \quad (5)$$

$$\text{MAE} = \frac{1}{n} \sum_{t=1}^{n} |Y_t - \hat{Y}_t| \quad (6)$$

The Diebold-Mariano (DM) test (Diebold and Mariano 1995) was applied to assess and compare the forecasting accuracy of two models by examining their forecast errors, also known as forecast residuals. This test is model-free, utilizing a loss differential $L_t = g(e_{t1}) - g(e_{t2})$, where e_{t1} and e_{t2} are forecast errors from methods 1 and 2, respectively. The null hypothesis of equal accuracy, i.e. $E(L_t) = 0$, against the alternative hypothesis of unequal accuracy, $E(L_t) = \neq 0$, Test statistic for DM test is as follows:

$$\text{DM} = \frac{\bar{L}}{\sqrt{\frac{2\pi \hat{s}_d(0)}{T}}}$$

where, $\bar{L} = \sum_{t=1}^{T} L_t$ and $s_d(0) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \infty_L(k)$, spectral density at frequency 0.

RESULTS AND DISCUSSION

In this study, conventional model (ARIMA), machine learning model (ANN, SVR, RF) and deep learning model (LSTM) have been used separately to see the trends and complex non-linear relationships inherent in the time-series data. The parameters for each model were rigorously tuned to optimize prediction accuracy and their performance was evaluated using RMSE, MAPE and MAE. The configurations for price and arrival forecasting models for each location are detailed in Table 1 and Table 2, respectively.

Descriptive analysis of jute arrivals and price variability across markets: Table 3 provides a summary of the descriptive statistics for both the arrival and price data of jute from four selected markets in West Bengal: Kaliaganj, Bethuadahri, Coochbehar, and Jangipur. The mean arrival and price values indicate variations across markets, with Jangipur having the highest mean arrival (67.93 tonnes) and Kaliaganj showing the lowest (31.81 tonnes). In terms of prices, Kaliaganj recorded the highest mean price (₹ 4157.51/q), while Jangipur had the lowest mean price (₹ 3705.28/q). Standard deviations revealed substantial variability in arrivals, particularly in Bethuadahri (64.43 tonnes) and Kaliaganj (36.84 tonnes), indicating fluctuations in the jute supply in these markets. The price data showed relatively lower variability, with standard deviations ranging between ₹1325.22 and ₹1504.86/q across markets. Kurtosis and skewness values for arrivals

Table 1 Model configuration for jute price

Models	ARIMA	ANN	SVR			RF		LSTM
			Cost	γ	ε	No of trees	Variables per split	
Bethuadahri	(1,1,0)	2-5-1	10	0.01	0	500	4	(25,20,1)
Jangipur	(0,1,1)	4-8-1	100	0.01	0	500	4	(25,20,1)
Kaliaganj	(1,1,1)	2-13-1	10	0.01	0	500	4	(25,10,1)
Coochbehar	(0,1,1)	4-15-1	10	0.01	0	500	4	(25,20,1)

ARIMA, Autoregressive Integrated Moving Average; ANN, Artificial Neural Networks; SVR, Support Vector Machine; RF, Random Forest; LSTM, Long Short-Term Memory.

Table 2 Model configuration for jute arrivals

Models	ARIMA	ANN	SVR			RF		LSTM
			Cost	γ	ε	No of trees	Variables per split	
Bethuadahri	(1,1,1)	10-6-1	10	0.01	0	500	4	(25,20,1)
Jangipur	(1,0,0)	1-2-1	1	0.01	0	500	4	(25,20,1)
Kaliaganj	(0,0,1)	2-14-1	10	0.01	0	500	4	(25,20,1)
Coochbehar	(2,1,0)	15-5-1	10	0.01	0	500	4	(25,10,1)

ARIMA, Autoregressive integrated moving average; ANN, Artificial neural networks; SVR, Support vector machine; RF, Random forest; LSTM, Long short-term memory.

Table 3 Descriptive statistics of arrival (tonnes) and price (₹) of selected jute markets (2007–2023)

	Kaliaganj		Bethuadahri		Coochbehar		Jangipur	
	Arrival	Price	Arrival	Price	Arrival	Price	Arrival	Price
Mean	31.81	4157.51	48.39	3855.64	27.26	4086.69	67.93	3705.28
Median	20.68	3898.40	17.78	3603.87	8.22	3924.32	67.92	3431.88
Standard Deviation	36.84	1504.86	64.43	1399.33	35.69	1399.60	37.40	1325.22
Kurtosis	14.55	-0.32	9.33	-0.66	3.70	-0.58	27.83	-0.19
Skewness	3.15	0.63	2.77	0.53	1.82	0.56	3.92	0.71
Minimum	1.19	1845.00	0.04	1706.35	0.27	1880.65	3.04	1761.85
Maximum	281.23	8371.43	357.61	7340.00	175.20	7300.00	364.44	7510.77
Count	204	204	204	204	204	204	204	204

suggest significant deviations from normal distribution, particularly in Jangipur and Kaliaganj, where arrivals display heavy-tailed distributions and positive skewness, indicating occasional extreme spikes in arrival quantities. The price data, on the other hand, showed near-normal distributions with slight positive skewness across all markets, reflecting more consistent price trends over time. This decoupling may be attributed to government interventions, minimum support prices, and regulated procurement, which cushion farmers against extreme price dips despite erratic arrivals

Jute price forecasting: For jute price forecasting, LSTM outperformed other models with the lowest RMSE, MAPE, and MAE across all locations, indicating superior performance in capturing both short- and long-term dependencies in the data (Table 4). In Bethuadahri market, LSTM achieved the lowest RMSE (169.40), indicating highly accurate predictions compared to other models like ARIMA (RMSE 1196.34), ANN (RMSE 1013.22), and RF (RMSE 1239.29). The MAPE of 2.12% and MAE of 114.51 further highlighted the precision of LSTM in this market. ANN and SVR also showed relatively good performance, but LSTM was superior by a significant margin. For Jangipur market, the LSTM model demonstrated the highest prediction accuracy with an RMSE of 168.77 and a MAPE of 2.65%, outperforming SVR and ARIMA, which had RMSEs of 1052.86 and 969.02, respectively. Although ANN underperformed (RMSE of 2514.68), LSTM's error metrics suggested its robustness in capturing price variations in this market. Similar results were obtained in the Kaliaganj and Coochbehar markets, where LSTM consistently outperformed other models, with RMSE values of 339.09 and 727.46, respectively. SVR, ARIMA, and RF showed higher errors, highlighting LSTM's superior capability in capturing complex temporal patterns. The results are consistent with previous studies that highlighted LSTM's efficiency in dealing with complex, long-term temporal dependencies (Dave *et al.* 2021, Purohit *et al.* 2021). For jute price forecasting, the LSTM model significantly outperformed the SVM model during the training phase (DM statistic ranging from 3.15 to 4.21, $p<0.01$). In the validation phase, the LSTM maintained superior accuracy in three markets, with DM statistics ranging from 2.37 ($p=0.04$) at Bethuadahri to 3.01 ($p=0.02$) at Jangipur. Ray *et al.* (2023) also demonstrated the robustness of LSTM in outperforming ARIMA models for various agricultural commodities.

Jute arrival forecasting: For forecasting jute arrivals, the LSTM model also generally outperformed the other models, although in some cases SVR showed comparable or slightly better performance, especially in terms of RMSE and MAPE (Table 5). For Bethuadahri market, LSTM exhibited the best performance (RMSE 2.06, MAPE 59.70%), slightly outperforming SVR (RMSE 2.52, MAPE 53.35%). In contrast, RF showed poor results with a high RMSE (9.47) and MAPE (67.37%), indicating that simpler models struggled with arrival data. In Jangipur market, SVR slightly outperformed LSTM (RMSE 4.53, MAPE 3.36%), but LSTM remained competitive with RMSE of 4.24. The close performance of SVR and LSTM here may be due to the relatively stable patterns in jute arrivals at this location, which SVR could model effectively. In Kaliaganj market, LSTM achieved the lowest RMSE (21.94)

Table 4 Model performance for predicting jute prices

Locations	Models	RMSE	MAPE	MAE
Bethuadahri	ARIMA	1196.34	20.07	1040.86
	ANN	1013.22	14.47	833.03
	SVR	950.07	13.30	859.23
	RF	1239.29	22.20	1143.23
	LSTM	169.40	2.12	114.51
Jangipur	ARIMA	969.02	16.70	822.65
	ANN	2514.68	46.78	2418.19
	SVR	1052.86	14.85	938.89
	RF	1417.98	27.27	1308.66
	LSTM	168.77	2.65	136.24
Kaliaganj	ARIMA	908.80	14.85	820.35
	ANN	1746.26	25.72	1479.42
	SVR	922.13	10.09	717.08
	RF	1331.30	19.71	1105.81
	LSTM	339.09	5.09	290.02
Coochbehar	ARIMA	1251.03	18.37	1044.31
	ANN	959.79	12.45	717.12
	SVR	1305.39	17.77	1194.54
	RF	1698.70	32.28	1583.29
	LSTM	727.46	10.65	648.28

ARIMA, Autoregressive Integrated Moving Average; ANN, Artificial Neural Networks; SVR, Support Vector Machine; RF, Random Forest; LSTM, Long Short-Term Memory; RMSE, Root Mean Square Error; MAPE, Mean Absolute Percentage Error; MAE, Mean Absolute Error.

Table 5 Model performance for predicting jute arrivals

Locations	Models	RMSE	MAPE	MAE
Bethuadahri	ARIMA	2.44	51.83	1.66
	ANN	47.06	987.03	29.32
	SVR	2.52	53.35	2.12
	RF	9.47	67.37	9.04
	LSTM	2.06	59.70	1.58
Jangipur	ARIMA	5.61	5.41	4.08
	ANN	9.38	11.22	8.29
	SVR	4.53	3.36	2.45
	RF	4.90	4.73	3.25
	LSTM	4.24	3.63	2.74
Kaliaganj	ARIMA	23.14	88.78	16.82
	ANN	27.66	52.57	18.00
	SVR	54.75	82.37	26.10
	RF	58.50	64.38	28.72
	LSTM	21.94	63.13	16.90
Coochbehar	ARIMA	2.81	289.57	2.64
	ANN	2.19	213.39	2.05
	SVR	1.03	29.85	0.81
	RF	1.38	33.64	1.06
	LSTM	1.58	48.53	0.94

ARIMA, Autoregressive Integrated Moving Average; ANN, Artificial Neural Networks; SVR, Support Vector Machine; RF, Random Forest; LSTM, Long Short-Term Memory; RMSE, Root Mean Square Error; MAPE, Mean Absolute Percentage Error; MAE, Mean Absolute Error.

compared to ARIMA (23.14) and RF (58.50). However, the high variability in arrival patterns caused relatively higher MAPE values (LSTM MAPE 63.13%), which indicates potential volatility in the data. In Coochbehar market, SVR outperformed all models with an RMSE of 1.03 and a MAPE of 29.85%, making it the most reliable model for forecasting arrivals at this location. LSTM was also competitive but showed slightly higher RMSE (1.58). This suggests that while LSTM captured the broader trends, SVR was more accurate in pinpointing specific arrival volumes in Coochbehar. Similar findings were observed by Paul *et al.* (2022), where machine learning techniques, particularly GRNN, outperformed conventional models like ARIMA for forecasting vegetable prices, further confirming the reliability of ML models for capturing short-term fluctuations in agricultural markets. For jute arrival forecasting, LSTM demonstrated significantly better forecasting accuracy than SVM in all markets during the training phase (DM statistic: 2.29 to 4.11, $p\leq0.07$), which persisted into the validation phase (DM statistic: 1.99 to 3.59, $p\leq0.06$). Our results indicated that while LSTM models are highly effective in capturing broader trends in jute arrivals, SVR may be more suitable for markets with stable or predictable arrival patterns, as highlighted by Paul *et al.* (2022). The relatively poor performance of RF, particularly in jute arrivals (RMSE of 9.47 in Bethuadahri market), reflects its limitations in handling time-series data, which is also echoed in the work of Ray *et al.* (2023), where RF models in forecasting accuracy.

Model evaluation and insights: A comparative analysis of actual and predicted values for jute prices and arrivals across four major markets in West Bengal have been presented in Fig. 1A and 1B, respectively. The LSTM model demonstrated high accuracy in forecasting jute prices (Fig. 1A), while various models, including ARIMA, RF, SVR, and LSTM, were applied for predicting jute arrivals, as shown in Fig. 1B. The results indicated that the models effectively capture both price trends and arrival patterns, albeit with some variations across markets. The LSTM model consistently outperformed standalone linear (ARIMA) and non-linear models (ANN, SVR, RF) especially in price forecasting, as demonstrated by similar results in previous studies (Dave *et al.* 2021, Ray *et al.* 2023). Accurate forecasts of jute prices and arrivals, particularly using LSTM models, can support the development of early warning systems

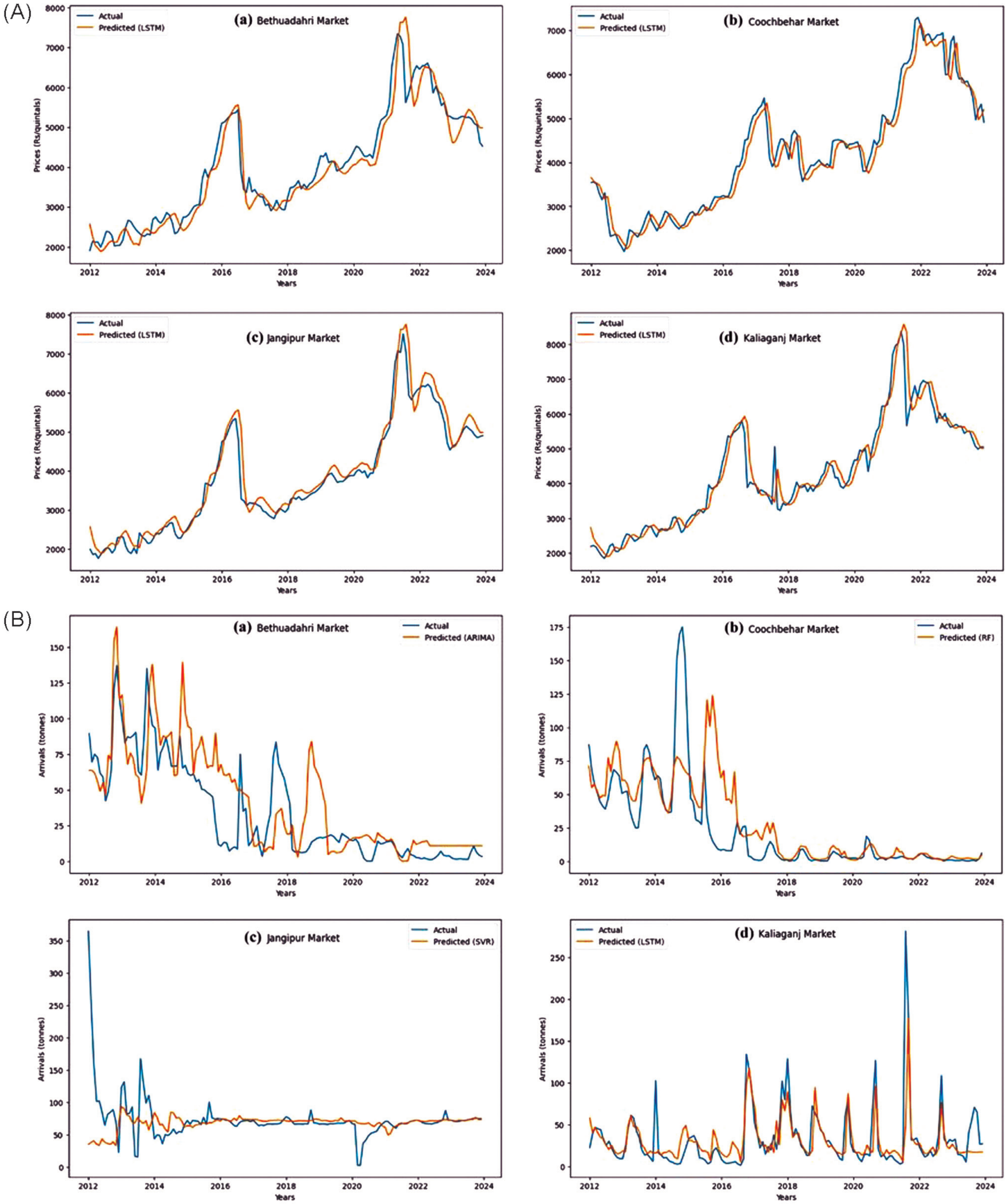

Fig. 1 Time series plot showing actual and predicted (A) jute prices (₹/q) and (B) jute arrival (tonnes) in four major markets of West Bengal: (a) Bethuadahri, (b) Coochbehar, (c) Jangipur, and (d) Kaliaganj. (The predictions are based on the LSTM model, highlighting its performance in capturing price trends across various years).

for market fluctuations. Such systems can aid farmers in making informed decisions about the timing of harvest and sale, while also assisting procurement agencies like the Jute Corporation of India in planning buffer stock and procurement operations. Moreover, these forecasts can guide policymakers in designing targeted price stabilization interventions and export strategies, thereby contributing to a more resilient and transparent jute marketing system in

West Bengal. The success of LSTM can be attributed to its ability to model long-term dependencies and its flexibility in adjusting to both price and arrival dynamics. However, in locations like Coochbehar for arrival forecasting, SVR showed that it could be more accurate in capturing short-term fluctuations, highlighting the importance of market-specific model selection (Paul *et al.* 2022). The inconsistent performance of RF and ANN suggested that simpler models are less adept at capturing the complex temporal dependencies present in our dataset (Purohit *et al.* 2021).

This study aims to forecast jute prices and arrivals using ARIMA and advanced machine learning techniques such as ANN, SVM, RF and LSTM in four major markets: Uttar Dinajpur, Nadia, Coochbehar, and Murshidabad. These models were evaluated using performance metrics such as RMSE, MAPE and MAE. Among them, LSTM emerged as the most effective model with the lowest RMSE, MAPE and MAE for predicting jute prices in the Bethuadahri, Jangipur, Kaliaganj, and Coochbehar markets. It also achieved the best results in forecasting jute arrivals in Bethuadahri and Kaliaganj with the lowest RMSE and MAE values. For Jangipur, both SVR and LSTM performed well in forecasting arrivals, though SVR slightly outperformed others in terms of RMSE and MAPE. SVR achieved the best performance in forecasting arrivals in Coochbehar with the lowest RMSE, MAPE, and MAE. These findings suggested that advanced machine learning models particularly LSTM consistently outperformed other models and highly effective for predicting jute prices. In addition, both LSTM and SVR performed well in forecasting jute arrivals. Complex time series data were forecasted using advanced machine learning and deep learning techniques. Considering the highly non-linear and non-normal characteristics of price and arrival data, the LSTM model was employed to effectively capture intricate nonlinear patterns. These patterns are typically challenging to detect using conventional statistical or other machine learning methods. The LSTM model demonstrated superior accuracy in both the training and validation phases, confirming its suitability for forecasting such complex time series data. The effectiveness of these models in forecasting jute price and arrivals highlights their potential for application to other agricultural commodities across various markets in India, providing a robust framework for anticipating market movements and supporting policy formulation.

REFERENCES

Anonymous. 2021. Annual Report 2020–2021. Ministry of Textiles, Government of India, New Delhi.

Anonymous. 2023. Agricultural Statistics at a Glance. Department of Agriculture, Cooperation and and Farmers' Welfare, Ministry of Agriculture and Farmers' Welfare, Government of India, New Delhi.

Dave E, Leonardo A, Jeanice M and Hanafiah N. 2021. Forecasting Indonesia exports using a hybrid model ARIMA-LSTM. *Procedia Computer Science* **179**: 480–87.

Diebold F X and Mariano R S. 1995. Comparing predictive accuracy. *Journal of Business and Economic Statistics* **13**(3): 253–63.

Ghosh S, Singh K N, Thangasamy A, Datta D and Lama A. 2020. Forecasting of onion (*Allium cepa*) price and volatility movements using ARIMAX-GARCH and DCC models. *The Indian Journal of Agricultural Sciences* **90**(5): 1009–13.

Jha G K and Sinha K. 2013. Agricultural price forecasting using neural network model: An innovative information delivery system. *Agricultural Economics Research Review* **26**(2): 229–39.

Jute Corporation of India. 2024. Annual Report 2023–24. Ministry of Textiles, Government of India.

Khashei M and Bijari M. 2011. A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. *Applied Soft Computing* **11**(2): 2664–75.

Kumar K S, Ilakiya T and Gowthaman T. 2023. Price instability, seasonal index and modelling for major vegetables in India. *Journal of Applied Horticulture* **25**(2): 1–6.

Kumar R R, Gupta A K and Patra C. 2020. Jute price forecasting in Murshidabad market of West Bengal using ARIMA technique. *Journal of Pharmacognosy and Phytochemistry* **9**(1): 1802–07.

Nayak D R, Mahapatra A and Mishra P. 2013. A survey on rainfall prediction using artificial neural network. *International Journal of Computer Applications* **72**(16): 32–40.

Paul R K, Yeasin M, Kumar P, Kumar P, Balasubramanian M, Roy H S, Paul A K and Gupta A. 2022. Machine learning techniques for forecasting agricultural prices: A case of brinjal in Odisha, India. *PLOS One* **17**(7): e0270553.

Purohit S K, Panigrahi S, Sethy P K and Behera S K. 2021. Time series forecasting of price of agricultural products using hybrid methods. *Applied Artificial Intelligence* **35**(15): 1388–406.

Rahman S, Kazal M, Begum I and Alam M. 2017. Exploring the future potential of jute in Bangladesh. *Agriculture* **7**(12): 96.

Ray S, Lama A, Mishra P, Biswas T, Das S S and Gurung B. 2023. An ARIMA-LSTM model for predicting volatile agricultural price series with random forest technique. *Applied Soft Computing* **149**: 110939.

Saha A, Singh K N, Ray M and Rathod S. 2020. A hybrid spatio-temporal modelling: AAN application to space-time rainfall forecasting. *Theoretical and Applied Climatology* **142**: 1271–82.

Shankar S V, Ajaykumar R, Ananthakrishnan S, Aravinthkumar A, Harishankar K, Sakthiselvi T and Navinkumar C. 2023b. Modelling and forecasting of milk production in the western zone of Tamil Nadu. *Asian Journal of Dairy and Food Research* **42**(3): 427–32.

Shankar S V, Chandel A, Gupta R K, Sharma S, Chand H, Kumar R, Mishra N, Ananthakrishnan S, Aravinthkumar A, Kumaraperumal R and Gowsar S R N. 2023a. Exploring the dynamics of arrivals and prices volatility in onion (*Allium cepa*) using advanced time series techniques. *Frontiers in Sustainable Food Systems* **7**: 1208898.

Shankar S V, Chandel A, Gupta R K, Sharma S, Chand H, Aravinthkumar A and Ananthakrishnan S. 2024. Comparative study on key time series models for exploring the agricultural price volatility in potato prices. *Potato Research* 1–19.

Wang L, Zou H, Su J, Li L and Chaudhry S. 2013. An ARIMA-ANN hybrid model for time series forecasting. *Systems Research and Behavioural Science* **30**(3): 244–59.

Zhang G P. 2003. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing* **50**: 159–75.