International Journal of Statistics and Applied Mathematics 2023; SP-8(5): 980-984

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2023; SP-8(5): 980-984

https://www.mathsjournal.com
Received: 03-08-2023
Accepted: 07-09-2023

Navigating soybean price volatility: A deep learning perspective

Moumita Baishya, G Avinash, Kamal Sharma, Veershetty and Harish Nayak GH

Moumita Baishya
The Graduate School, ICAR-Indian Agricultural Research Institute, New Delhi, India

G Avinash
The Graduate School, ICAR-Indian Agricultural Research Institute, New Delhi, India

Kamal Sharma
The Graduate School, ICAR-Indian Agricultural Research Institute, New Delhi, India

Veershetty
The Graduate School, ICAR-Indian Agricultural Research Institute, New Delhi, India

Harish Nayak GH
The Graduate School, ICAR-Indian Agricultural Research Institute, New Delhi, India

Corresponding Author:
Kamal Sharma
The Graduate School, ICAR-Indian Agricultural Research Institute, New Delhi, India

DOI: https://doi.org/10.22271/maths.2023.v8.i5Sn.1316

Abstract
Soybean, a significant oilseed crop, has become increasingly vital in India over the past decade, serving as an essential protein source for both human consumption and livestock feed. With soaring production and demand, especially in regions such as Madhya Pradesh, Maharashtra, Rajasthan, Karnataka, and Gujarat, there's an amplified need for reliable soybean futures price predictions. Forecasting in the futures market is not only of immense value but also technically challenging. This study delves into a comparative evaluation of soybean futures prices using various deep learning models, including Time Delay Neural Network (TDNN), Recurrent Neural Network (RNN), Gated Recurrent Unit (GRU) and Long Short-Term Memory (LSTM). Our findings reveal that the LSTM and GRU models substantially outperform the TDNN and RNN in terms of forecasting accuracy. Specifically, the LSTM model emerges as the pinnacle, delivering unparalleled directional forecasting results. The efficacy of the models was further assessed using Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE), wherein LSTM was identified as the most representative model for soybean price predictions. This research provides pivotal insights for futures price forecasting applications, presenting a robust model that could serve as a crucial policy tool for farmers, processors, and traders.

Keywords: Price forecasting, soybean market, TDNN, RNN, LSTM, GRU

1. Introduction
Soybean (*Glycine max*) serves as a pivotal crop, renowned for its high-quality vegetable oil and protein-rich meal, which find applications in animal feed and processed foods. Its cultivation presents farmers with a lucrative alternative to conventional staples, fostering crop diversification. Furthermore, soybean enhances both economic viability through increased income and Agro-ecological health via nitrogen fixation. As of recent metrics, India ranks fifth globally in soybean production, boasting an impressive yield of 13.27 million tonnes. Pioneering this agricultural movement are states like Madhya Pradesh, Maharashtra, and Rajasthan.

Price forecasting of soybean remains a cornerstone concern for all agricultural stakeholders, encompassing both producers and consumers. Accurate forecasting provides invaluable insights into the intricacies of soybean production, supply dynamics, and price trends. Mastery in price prediction facilitates the anticipation of market shifts, allowing for judicious planting and harvesting choices and streamlined supply chain operations. This ensures a dual benefit: food security and economic equilibrium. By equipping producers with the tools to maximize yields and adapt to market fluctuations, forecasting also ensures that consumers maintain a consistent supply of essential soybean derivatives.

The past decade has witnessed a pronounced escalation in both the production and demand for soybean in India, leading to its broader adoption among farmers across states like Madhya Pradesh, Maharashtra, Rajasthan, Karnataka, and Gujarat. This growth underscores the pressing need for trustworthy forecasts of soybean's future prices. The Indore soybean market, a prominent trading nexus in India, holds significant sway over both national and international soybean pricing metrics.

As per the AGMARKNET 2022 report, Madhya Pradesh emerges as the frontrunner with an expansive cultivation area of 55.84 lakh hectares, closely trailed by Maharashtra's 46.01 lakh hectares. Of particular note is Madhya Pradesh's marked surge in the Wholesale Prices Monthly, registering an impressive increase of 12.93%. Given this backdrop, our study embarked on an analytical journey, harnessing a decade of weekly soybean price data from Indore, Madhya Pradesh, spanning from 1st July 2006 to 15th October 2016. Utilizing advanced deep learning models, including TDNN, RNN, GRU, and LSTM, our endeavour aims to astutely predict the future volatility of soybean prices.

2. Review of Literature

Time series analysis remains a foundational pillar in various domains. Historically, the ARIMA model has been the touchstone in this field (Box *et al.* 1995) [2]. However, while ARIMA shines in deciphering linear relationships, it stumbles when faced with non-linear data patterns. Researchers have developed models like Bilinear (Granger and Anderson 1978) [7], ARCH (Engle 1982) [5], GARCH (Bollerslev 1986) [1], and TAR (Tong and Lim 1980) [16] to address this non-linearity. Despite their capabilities, these statistical models often buckle under the weight of real-world data, constrained by inherent assumptions that may not always mirror reality.

Enter the domain of machine learning (ML), a paradigm shift that brought flexibility and adaptive learning to the fore (Salcedo-Sanz *et al.* 2016) [19]. In specific areas like rainfall prediction, ML has shown a marked superiority over traditional statistical models (Ojo and Ogunjo 2022) [15]. For instance, Cramer *et al.* (2017) [3] leveraged algorithms like K-Nearest Neighbor, SVR, and RBF to analyze precipitation data. Yet, the Achilles' heel of these ML approaches is their dependency on precise feature extraction, a labour-intensive process requiring domain-specific expertise (Janiesch *et al.* 2021) [10].

Deep Learning (DL), a subset of ML, heralded a new era by sidestepping the manual feature extraction conundrum. Through intricate architectures, DL models can autonomously discern and harness pertinent features, making them adept at managing vast and chaotic datasets. Their proficiency has earned them accolades across diverse applications, positioning DL as a forerunner in the big data challenge (Torres *et al.* 2021) [17]. Many studies have shown that the stock price is predictable and many classic algorithms such as Recurrent Neural Network (RNN) of LSTM and GRU and its extensions are used in time-series prediction nowadays (Gao, *et al.*, 2022) [6].

In summation, while the journey of financial forecasting has been enriched by various methodologies, LSTM stands out, heralding a new era of precision and accessibility in predicting financial time series.

3. Theoretical background

3.1 Time Delay Neural Network (TDNN)

The artificial neural network, inspired by the functioning of the human brain, consists of abstractions of processing elements in the form of mathematical functions called artificial neurons or nodes. The group of neurons operating together forms a layer of neurons and in general, three distinct layers are formed in a standard ANN model. These three layers namely the input layer, hidden layer and output layer are so interconnected with their nodes that each layer receives input from its preceding layer and passes the output to the subsequent layer. This neural network model builds a short-term memory, in particular, heteroassociative memory, in its network through the use of time delays of a univariate time series to capture the temporal dimension of the series by Haykins (2009) [8].

3.2 Recurrent Neural Network (RNN)

It is a type of neural network where the previous outputs are fed as the input to the current step. The advantage of RNN is the hidden state (internal memory) that captures information calculated so far in a sequence has been depicted in Fig. 1.

Though the RNNs work effectively in many application domains, they may suffer from a problem called vanishing gradients Li *et al.* (2018) [13]. To cope with this problem, two variants of RNN have been developed: Long Short-Term Memory (LSTM) by Hochreiter and Schmidhuber (1997) [9] and Gated Recurrent Units (GRU) networks by Kyunghym *et al.* (2014) [12]. LSTM is capable of learning long-term dependencies with a special memory unit.

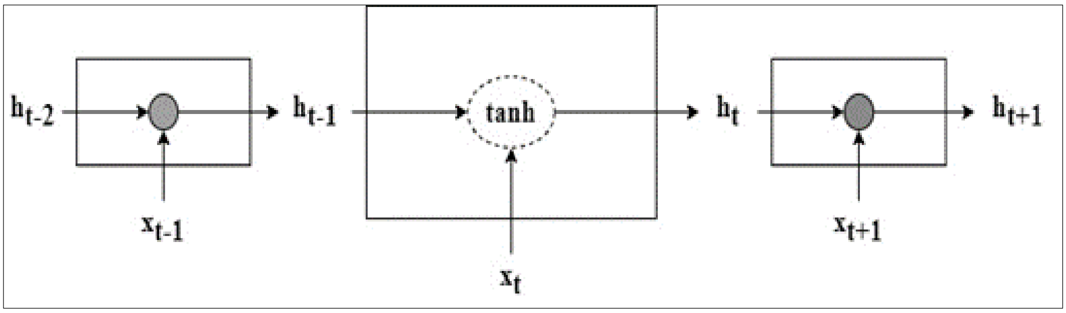

Fig 1: Architecture of the RNN

3.3 Gated recurrent unit (GRU)

Gated recurrent unit (GRU) is a kind of RNN that uses gating mechanisms to control the flow of information between cells in the neural network derived from LSTM and was introduced in 2014 by Kyunghyun Cho *et al.* (1997) [12]. GRU is composed of two gates, an update gate and a reset gate. These gates are used to filter out what information should remain and what should be disposed of. Different from traditional RNN, GRUs solve the vanishing and exploding gradient problems. Unlike LSTM, GRU has fewer parameters than LSTM due to the lack of one gate. Another difference is that GRUs also lack the cell state from LSTM so that GRU can only store both long and short-term memory in the hidden state. Recently, GRUs have been shown to perform better than LSTM on certain smaller and less frequent datasets. Fig. 2 shows the internal architecture of a GRU unit cell.

3.4 Long-Short term memory (LSTM)

Long short-term memory (LSTM) is a specific recurrent neural network (RNN) architecture. It was proposed

(Hochreiter *et al*., 1997) [9]. Unlike a traditional feed-forward neural network, it includes feedback connections. Furthermore, it can be utilized on single-point data and the sequence of data as well. The essential components of LSTM are an input gate, an output gate and a forget gate, and the LSTM network was developed to resolve the vanishing gradient problem while training the traditional RNNs. LSTM is a cell memory unit that means that LSTM can remove or add information to the cell state. LSTM has overcome the vanishing gradients and the exploding gradients problem that appeared in RNN through the units' specific internal structure built in the model. Nowadays, LSTM has been known as a powerful method capable of processing, classifying, and making predictions based on time series data.

There are three gate controls: input gate (i_t), output gate (o_t), and forget gate (f_t) in LSTM cell as shown in Fig. 3.

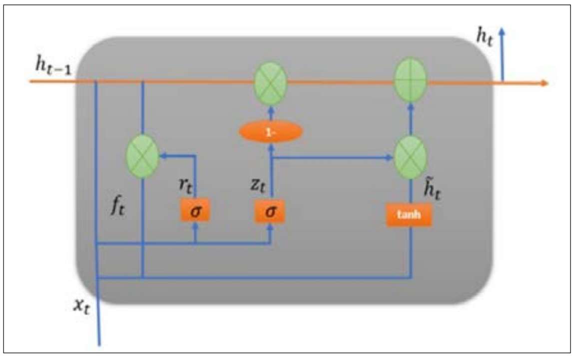

Fig 2: Architecture of the GRU

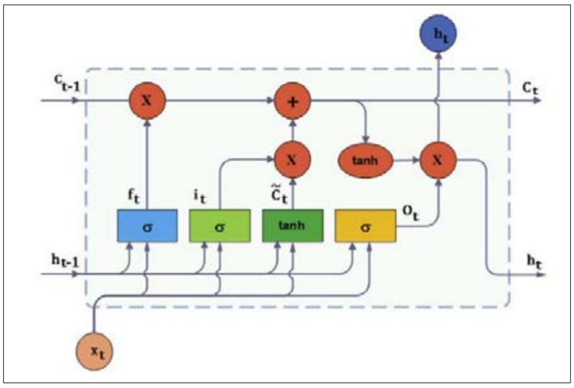

Fig 3: Architecture of the LSTM

4. Result and Discussion

4.1 Data description

In this study, the daily price of soybean price (in Rs. /quintal) is used as an experimental dataset. The weekly price series were acquired from the Agmarknet website https://agmarknet.gov.in/ (from 01 Jan 2013, through 16 Oct 2023).

The descriptive statistics of the Soybean price series, as presented in Table 1, reveal a dataset with 538 observations spanning a wide range from Rs. 2103.00 to Rs. 9788.33 per quintal. With an average price around Rs. 4063.09 and a substantial standard deviation of Rs. 1267.21, the data underscores the dynamic nature of the Soybean market. The relative variability, as captured by the coefficient of variation at 31.88%, further accentuates this volatility as shown in Fig.4. The positive skewness value of 1.72 indicates periods with notably higher prices, possibly due to factors like supply shortages or surging demand. Though the kurtosis value of 2.93 suggests a resemblance to a normal distribution in terms of its tails, this is contradicted by the Jarque-Bera and Shapiro-Wilk's tests, both of which strongly reject the hypothesis of normality with p-values less than 0.001. This departure from normality implies that standard analytical methods, which assume normally distributed residuals, might not be optimal for modeling this data. Therefore, alternative strategies catering to non-normal distributions could be more suitable for understanding and forecasting the Soybean price series. ARCH-LM test, ADF and BDS test has been done and found volatility, non-stationary and non-linearity in the data which inspires to go for deep learning models.

Table 1: Descriptive statistics of Soybean price (in Rs. /quintal)

Descriptive Statistics	Price series
Count	538.00
Minimum	2103.00
Mean	4063.09
Maximum	9788.33
Standard Deviation	1267.21
Coefficient of Variation (%)	31.88
Skewness	1.72
Kurtosis	2.93
Jarque-Bera test	459.48 (<0.001)
Shapiro-Wilk's test	0.80(<0.001)

Indication of p-values in parentheses

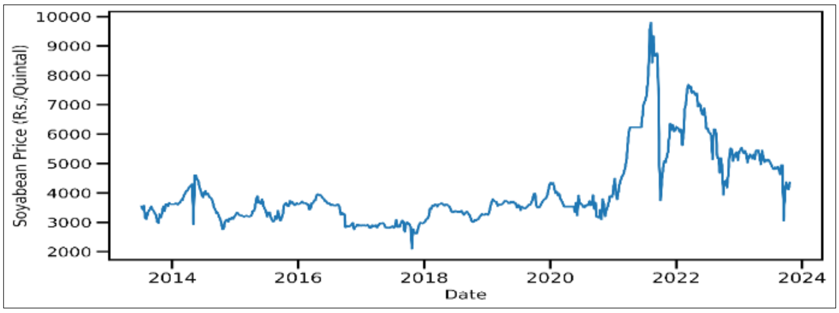

Fig 4: Time series plot of soybean price series of Indore Market, M.P

Table 2: Performance comparison of various models on test data

Models	RMSE (Rs. /quintal)	MAE (Rs. /quintal)	MAPE (%)
TDNN	419.96	279.48	5.49
RNN	428.69	274.43	5.67
GRU	413.04	259.99	5.57
LSTM	400.13	250.89	5.02

This study is done in Python software using tensor flow as the platform for training the DL models. After training these DL models, namely Time Delay Neural Network (TDNN), Recurrent Neural Network (RNN), Gated Recurrent Unit (GRU) and Long Short-Term Memory (LSTM) model with their respective loss function "Relu" and "Adam" optimizer is depicted in Fig. 5(a), 5(b), 5(c) and 5(d).

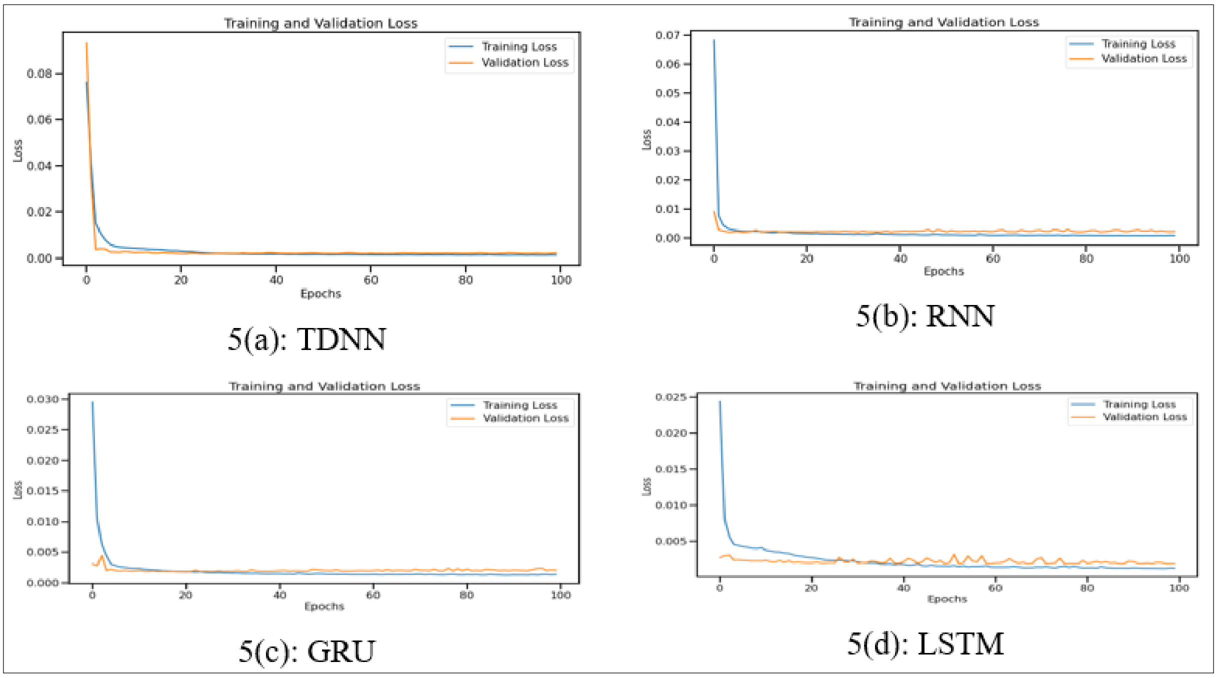

Fig 5: Loss function plots

Upon evaluating various deep learning models for their efficacy in forecasting Soybean prices, as delineated in Table 2, it was observed that the LSTM model outperformed TDNN, RNN, and GRU models in terms of precision and accuracy. The LSTM model registered the lowest RMSE and MAE values at Rs. 400.13/quintal and Rs. 250.89/quintal, respectively, indicating its superior ability to predict price fluctuations with minimal error. Furthermore, its MAPE score of 5.02% was the lowest among the models tested, showcasing its proficiency in maintaining consistency between the forecasted and actual prices. The GRU model also demonstrated commendable performance, particularly in reducing the MAE, suggesting its effectiveness in capturing the temporal dependencies characteristic of the price series. However, conventional RNN lagged slightly behind its more advanced counterparts, indicating possible limitations in handling the complex patterns within the data as shown in Fig. 6. These insights underscore the importance of selecting an appropriate modeling technique, with a preference for LSTM in this scenario, to enhance the reliability and accuracy of future market predictions in the Soybean market.

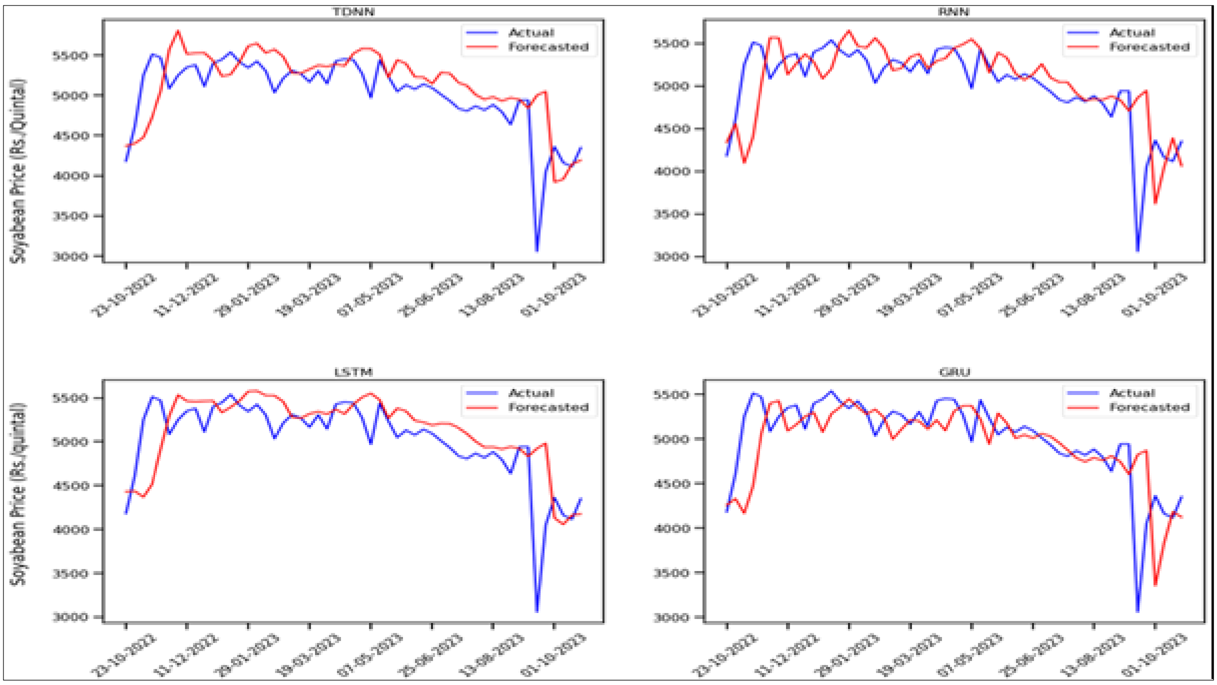

Fig 6: Optimized results obtained by different models for the Soybean price on test data

5. Conclusions

While initial assessments indicated that the assumptions required for traditional statistical modeling were not satisfied by the Soybean price series, our exploration into deep learning offered promising avenues. Among the array of models evaluated, the LSTM model distinctly stood out. It adeptly navigated the inherent volatility and intricate temporal intricacies of the data, outpacing other conventional deep learning models in accuracy and reliability. Thus, in contexts where data assumptions challenge traditional approaches, LSTM presents itself as a superior and robust alternative for predicting trends in the Soybean market.

6. References

1. Bollerslev T. Generalized autoregressive conditional heteroskedasti city. J Econom. 1986;31:307-327. https://doi.org/10.1016/0304-4076(86)90063-1
2. Box G, Jenkins G, Reinsel G, Ljung G. Time Series Analysis: Forecasting and Control. John Wiley & Sons.; c1995.
3. Cramer S, Kampouridis M, Freitas AA, Alexandridis AK. An extensive evaluation of seven machine learning methods for rainfall prediction in weather derivatives. Expert Syst Appl. 2017;85:169-181. https://doi.org/10.1016/J.ESWA.2017.05.029
4. Devi G, Parmar P. Trend of market price and seasonality of soybean in Gujarat. Gujarat Journal of Extension Education. 2022;33(2):97-101.
5. Engle RF. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. Econometrica. 1982;50:987. https://doi.org/10.2307/1912773
6. Gao N, Xue H, Shao W, Zhao S, Qin KK, Prabowo A, *et al*. Generative adversarial networks for spatio-temporal data: A survey. ACM Transactions on Intelligent Systems and Technology (TIST). 2022;13(2):1-25.
7. Granger CWJ, Anderson AP. Introduction to bilinear time series models. Vandenhoeck and Ruprecht, Gottingen; c1978.
8. Haykin S. Neural networks and learning machines, 3/E. Pearson Education India; c2009.
9. Hochreiter S, Schmidhuber J. Long Short-Term Memory. Neural Comput. 1997;9:1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735
10. Janiesch C, Zschech P, Heinrich K. Machine learning and deep learning. Electron Mark. 2021;31:685-695. https://doi.org/10.1007/s12525-021-00475-2
11. Jeon HI, Cho WR, Park SH, Kim U, Cho WJ, Kim MW, *et al*. An, Sunghyuck/and Rhee, Bum Ku Quantum the Journal of the Optical Society of Korea. 1997;1(2):121-122.
12. Junyoung C, Caglar G, KyungHyun C, Yoshua B. Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv 2014. arXiv preprint arXiv:1412.3555; c2014.
13. Li S, Jin X, Xuan Y, *et al*. Enhancing the Locality and Breaking the Memory Bottleneck of Transformer on Time Series Forecasting. In: Wallach H, Larochelle H, Beygelzimer A, *et al*. (eds) Advances in Neural Information Processing Systems. Curran Associates, Inc.; c2019.
14. Mishra S, Nahatkar SB, Patel S. Growth and instability of soybean in Central India: A district-level analysis; c2023.
15. Ojo OS, Ogunjo ST. Machine learning models for prediction of rainfall over Nigeria. Sci. African. 2022;16:e01246. https://doi.org/10.1016/J.SCIAF.2022.E01246
16. Tong H, Lim KS. Threshold autoregression, limit cycles and cyclical data with discussion. J R Stat Soc Ser B Methodol. 1980;42:245-292.
17. Torres JF, Hadjout D, Sebaa A, *et al*. Deep Learning for Time Series Forecasting: A Survey. Big Data. 2021;9:3-21. https://doi.org/10.1089/big.2020.0159
18. Vashist J. Agricultural market information through technology. Indian Journal of Agricultural Marketing. 2022;36(2):50-61.
19. Salcedo-Sanz S. Modern meta-heuristics based on nonlinear physics processes: A review of models and design procedures. Physics Reports. 2016 Oct 19;655:1-70.