A COMPARATIVE STUDY OF NEURAL NETWORK AND WAVELET DECOMPOSITION MODELS FOR PRICE FORECASTING OF TOMATO

ABSTRACT

Tomato price forecasting in Bihar marketplaces is the subject of a study. The study compares and contrasts many time series models. Three models are compared: WT-TDNN (Wavelet Transform with TDNN), TDNN (Time-Delay Neural Network), and ARIMA.

This study evaluates the efficacy of advanced and traditional time series models for forecasting tomato prices in Bihar markets, with a focus on ARIMA, Time Delay Neural Networks (TDNN), and Wavelet-TDNN (WT-TDNN). ARIMA models were optimized using the Akaike Information Criterion (AIC), identifying (0, 1, 2) and (1, 1, 1) as optimal configurations, and validated through Wald's test. TDNN models with 2:4s:1l and 1:6s:1l architectures exhibited robust performance based on RMSE, MAE, and MAPE metrics.

The WT-TDNN model, which integrates wavelet decomposition with TDNN, demonstrated superior predictive accuracy over both ARIMA and standard TDNN models. By employing the Haar filter for series decomposition into orthogonal components and applying TDNN to each component, WT-TDNN effectively captured the complex, non-linear dynamics of agricultural price series.

This innovative approach enhances the precision of price forecasting, offering significant implications for agricultural economics and machine learning applications. The findings provide valuable insights for policymakers, farmers, and market analysts, enabling improved decision-making and fostering resilience in the agricultural sector. The study highlights WT-TDNN as a transformative tool for addressing the challenges of agricultural price prediction.

Key words: TDNN, Wavelet Analysis, Performance Measures, RMSE, MAE, MAPE

1.Introduction

In the bustling markets of rural and urban landscapes, the price of agricultural commodities like tomatoes plays a crucial role in shaping the lives of farmers, consumers, and policymakers. Predicting these prices, however, is no simple task. Agricultural price series are as unpredictable as the weather that governs their production—complex, nonlinear, and often chaotic. Farmers, already grappling with challenges like market imperfections, speculative trading, and globalization, need accurate price forecasts to make informed decisions about planting, harvesting, and selling their produce.

Historically, statistical models like the Autoregressive Integrated Moving Average (ARIMA) have been the go-to tools for forecasting. These models, while effective for linear relationships, struggle to capture the intricate dynamics of agricultural price series. Recognizing these limitations, researchers began exploring nonlinear models such as Self-Exciting Threshold Autoregressive (SETAR) and Smooth Transition Autoregressive (STAR). These models brought improvements but still lacked the flexibility and adaptability required for real-world agricultural data.

As technology advanced, the focus shifted to machine learning (ML) and artificial intelligence (AI) models, which offer powerful tools for analyzing complex datasets. Among these, Time Delay Neural Networks (TDNNs), Long Short-Term Memory (LSTM) networks, and Gated Recurrent Units (GRUs) emerged as frontrunners. These models excel at capturing long-term dependencies and nonlinear patterns, making them ideal for agricultural price forecasting. Yet, even these sophisticated tools faced challenges when dealing with the nonstationary nature of agricultural price series.

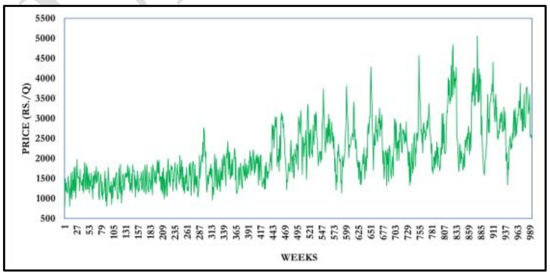
This is where decomposition techniques like Wavelet Analysis come into play. By breaking down a complex price series into simpler subseries, Wavelet Analysis provides a clearer picture of underlying patterns, enabling neural networks to make more accurate predictions. The synergy between Wavelet Transforms and AI models promises a revolutionary leap in forecasting methodologies.

Inspired by this potential, the current study embarks on a journey to address the complexities of agricultural price forecasting, with a specific focus on tomato prices. It aims to explore the capabilities of TDNNs and LSTM models, both individually and in combination with Wavelet Transforms. The objectives of this study are to investigate the effectiveness of these models and techniques in providing reliable and actionable price forecasts.

By delving into the intersection of advanced AI techniques and agricultural economics, this research aspires to empower farmers, stabilize markets, and guide policymakers, ultimately contributing to a more resilient agricultural ecosystem.

2.Materials and Methods

2.1Data Collection
This study utilized weekly tomato price data from the Bihar market, sourced from the National Horticultural Research and Development Foundation (NHRDF) website, covering January 2004 to March 2023. The dataset was divided into training (980 observations) and testing (12 weeks) sets to evaluate the effectiveness of various time series models. Figures 1 and 2 illustrate the non-stationary and non-linear behavior of the minimum and maximum price series, further supported by descriptive statistics(table 1), which highlight the distribution characteristics.



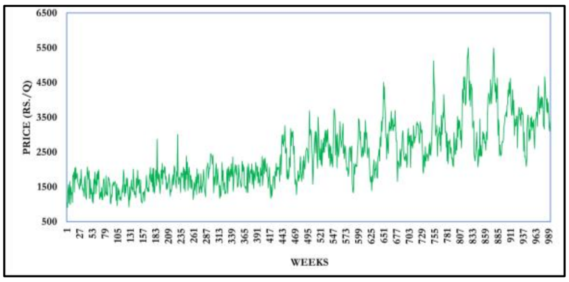


Fig 1. Time plot for daily tomato minimum price series

Fig 2. Time plot for daily tomato maximum price series

Statistics	Min Price Series	Max Price Series
No. of Observations	992	992
Max Value	5114	5069
Min Value	688	802
Mean	1884.38	2104.53
Median	1716	1963.5
Std. Deviation	846.77	740.19
Skewness	0.91	0.84
Kurtosis	0.34	0.41

Table 1 Descriptive Statistics

2.2Time Series Stationarity

Stationarity of the series was assessed using the Augmented Dickey-Fuller (ADF) and Phillips-Perron (PP) tests. Both tests confirmed that the original series were non-stationary, but stationarity was achieved after differencing. Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots provided insights into lag structures, while the Shapiro-Wilk test confirmed non-normality. The Hurst Exponent Statistic indicated anti-persistence, suggesting the absence of long memory in the series.

2.3 Model Development

The ARIMA models were developed by identifying the best-fit configurations based on Akaike Information Criteria (AIC). ARIMA (1, 1, 1) and ARIMA (0, 1, 2) were identified as optimal models for minimum and maximum price series, respectively, with significant parameter estimates.

A Time Delay Neural Network (TDNN) was applied to predict price series, leveraging logistic and identity activation functions. Optimal configurations were determined through experimentation with varying input nodes and hidden layers, trained using the Levenberg-Marquardt algorithm for fast convergence.

To enhance predictive accuracy, a Wavelet-Time Delay Neural Network (Wavelet-TDNN) model was implemented. The series were decomposed into wavelet components and modeled individually. Performance metrics such as Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) were used to evaluate model effectiveness. The Wavelet-TDNN consistently demonstrated superior accuracy in both training and validation phases, as highlighted in the forecasts for minimum and maximum price series.

This methodological approach provides a robust framework for forecasting price series, combining traditional statistical models with advanced neural network techniques for enhanced predictive performance.

3.Results And Discussions

To assess the significance of the parameter estimates in the ARIMA models, Wald's z-test was employed. The test results are further supported by Tables 3. which present the point forecasts for the ARIMA (1, 1, 1) model for the tomato minimum price series and the ARIMA (0, 1, 2) model for the tomato maximum price series. These forecasts are visually depicted in Figures 3 and 4 respectively. Table 2 details the results for the ARIMA (1, 1, 1) model, showing that the predicted values closely follow the actual test set values across several weeks, indicating the model's accuracy in forecasting the minimum price of tomatoes. Similarly it also presents the outcomes for the ARIMA (0, 1, 2) model applied to the maximum price series, with the forecasted values also demonstrating a strong alignment with the observed data. These results, illustrated in Figure 3. (a) and 3. (b), highlight the effectiveness of the selected ARIMA models in capturing the dynamics of tomato price fluctuations.

Table 2. Results of ARIMA model for tomato price series

Week	Test Set (Min Price)	ARIMA (1,1,1) (Min Price)	Test Set (Max Price)	ARIMA (0, 1, 2) (Max Price)
981	3157	3486	3836	4134
982	3250	3628	4024	4315
983	3084	3431	3693	3918
984	2801	3138	3318	3557
985	2858	3175	3409	3708
986	2535	2835	3216	3457
987	2523	2877	3211	3436
988	2818	3191	3389	3679
989	2290	2599	2982	3209
990	2031	2429	2633	2848
991	1911	2291	2454	2677
992	2058	2418	2737	2946

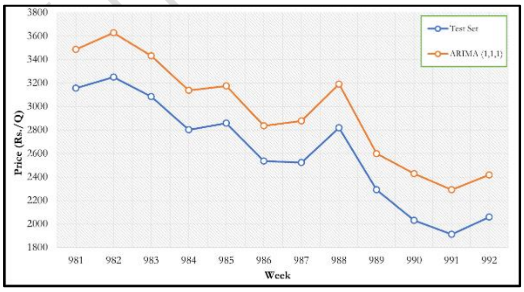


Fig 3. a) Fitted ARIMA (1, 1, 1) with tomato minimum price series

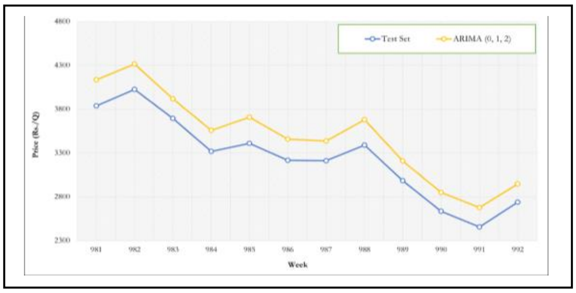


Fig 3.b)Fitted ARIMA (0, 1, 2) with tomato maximum price series

In the class of TDNN models considered in this study, it was observed that the models with structures 2:8s:1l and 1:6s:1l had the lowest performance measures for the minimum and maximum price series of tomatoes, respectively. The point forecasts up to 12 lags for both series are detailed in table 3, with the best TDNN models for each series' point forecasts illustrated graphically (figure 4 (a) and 4 (b)).

Table 3: Results of TDNN model for tomato price series

Week	Test Set (Min price)	TDNN (2:8s:1l) (Min price)	Test Set (Max price)	TDNN (1:6s:1l) (Max price)
981	3157	3288	3836	3932
982	3250	3385	4024	4101
983	3084	3208	3693	3810
984	2801	2931	3318	3414
985	2858	2978	3409	3526
986	2535	2665	3216	3293
987	2523	2647	3211	3305
988	2818	2942	3389	3504
989	2290	2410	2982	3066
990	2031	2164	2633	2737
991	1911	2039	2454	2567
992	2058	2180	2737	2819

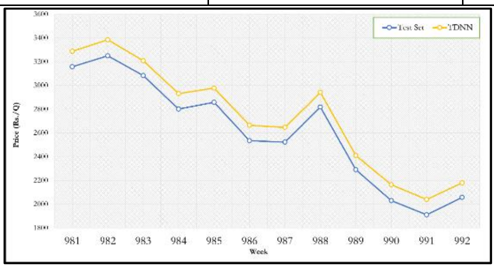


Fig. 4a) Fitted TDNN (2:8s:1l) with tomato minimum price series

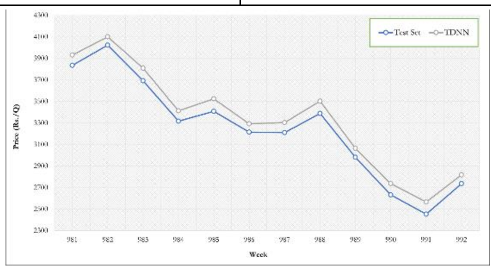


Fig.4 b) Fitted TDNN (1:6s:1l) with tomato maximum price series

To enhance predictive accuracy, both the minimum and maximum tomato price series were decomposed using a six-level wavelet transform, resulting in seven components (W_1 to W_6 and V_6). Each component was modeled using Time Delay Neural Networks (TDNN), and the model with the lowest Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) was identified as the best fit. The point forecasts are presented in Table 4 and Figure 5(a) for the minimum price series and Figure 5(b) for the maximum price series. This approach highlights the efficacy of the Wavelet-TDNN framework in capturing complex price patterns.

Table 4 Results of wavelet based TDNN model for tomato price series

Week	Test Set (Min Price)	WT-TDNN (Min Price)	Test Set (Max Price)	WT-TDNN (Max Price)
981	3157	3181	3836	3841
982	3250-‘m,	3272	4024	4034
983	3084	3122	3693	3698
984	2801	2827	3318	3325
985	2858	2888	3409	3418
986	2535	2580	3216	3221
987	2523	2572	3211	3216
988	2818	2860	3389	3399
989	2290	2327	2982	2990
990	2031	2048	2633	2640
991	1911	1927	2454	2459
992	2058	2073	2737	2746

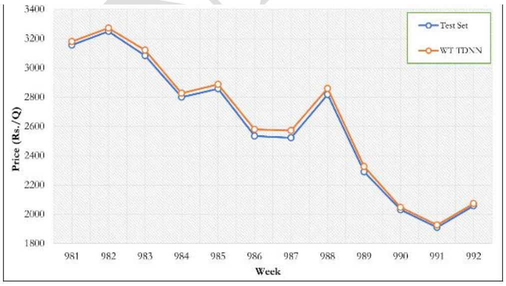


Fig.5 a) Fitted wavelet based TDNN for min price series

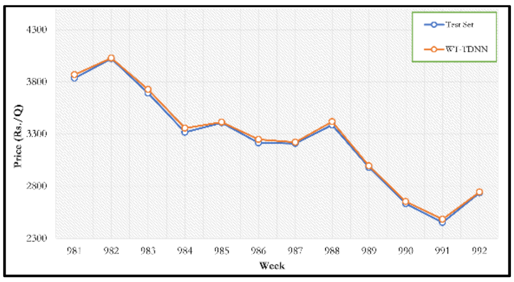


Fig 5b) Fitted wavelet based TDNN for max price series

Performance Measure in relation to Tomato Price Series

Performance measures for both series is compared with three different models viz. Autoregressive Integrated Moving Average (ARIMA), Time Delay Neural Network (TDNN) and Wavelet Based Time Delay Neural Network (WT-TDNN) considering three different measures i.e., RMSE, MAE and MAPE illustrated in Table 5 and Table 6.

Table 5: Performance measures of model for tomato minimum price series

Series	Model	Set	RMSE	MAE	MAPE (%)
Minimum Price Series	**ARIMA (2, 1, 2)**	**Training**	404.948	391.4475	12.2114
		Validation	656.734	522.6335	12.3609
	TDNN	**Training**	370.309	435.701	8.5604
		Validation	562.9005	528.697	9.7438
	WT-TDNN	**Training**	96.0445	238.051	6.1712
		Validation	149.544	224.785	7.3675

Table 6: Performance measures of model for tomato maximum price series

Series	Model	Set	RMSE	MAE	MAPE(%)
Maximum Price Series	**ARIMA (1,1,0)**	**Training**	**305.788**	**318.8195**	**11.6465**
		Validation	**373.6925**	**436.84**	**11.7908**
	TDNN	**Training**	**320.8295**	**375.937**	**7.9911**
		Validation	**278.2175**	**393.6585**	**9.1701**
	WT-TDNN	**Training**	**49.5133**	**40.4010**	**5.6013**
		Validation	**62.4449**	**147.735**	**6.7988**

Bar and line diagram for all points forecasts of different model is illustrated in Figure 6, Figure 7, Figure 8 and Figure 9. Figure 10 and Figure11 depicts various performance measures obtained from different models used in analysis of both price series respectively.

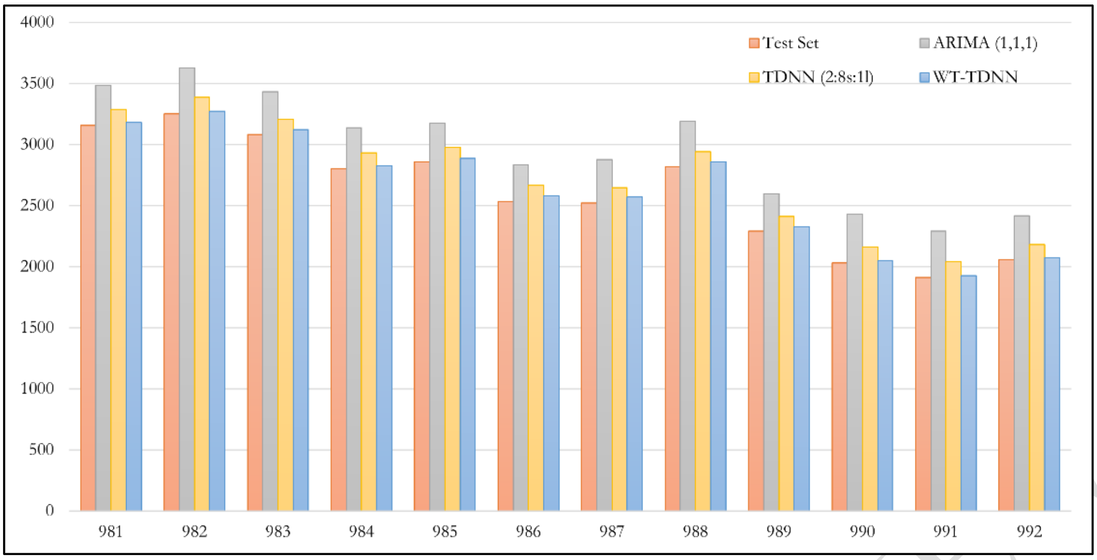


Figure 6: Bar diagram of several point forecasts for tomato minimum price series

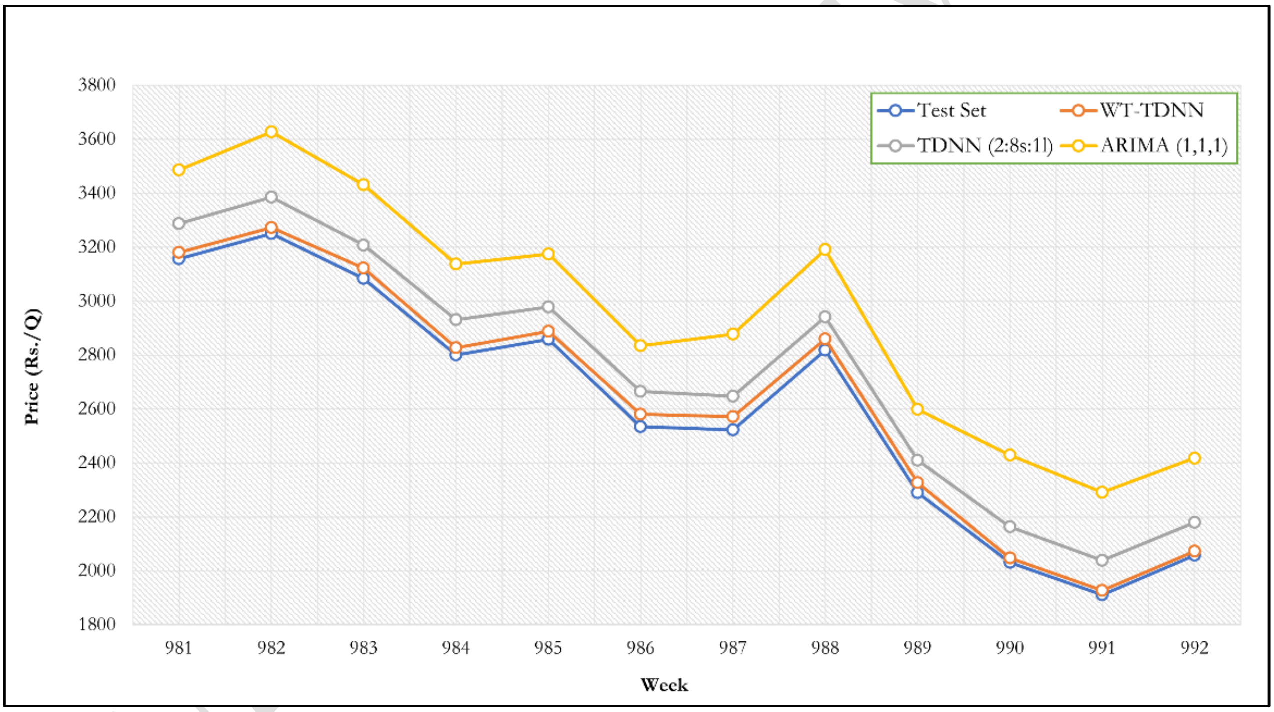


Figure 7: Line diagram of several point forecasts for tomato minimum price series

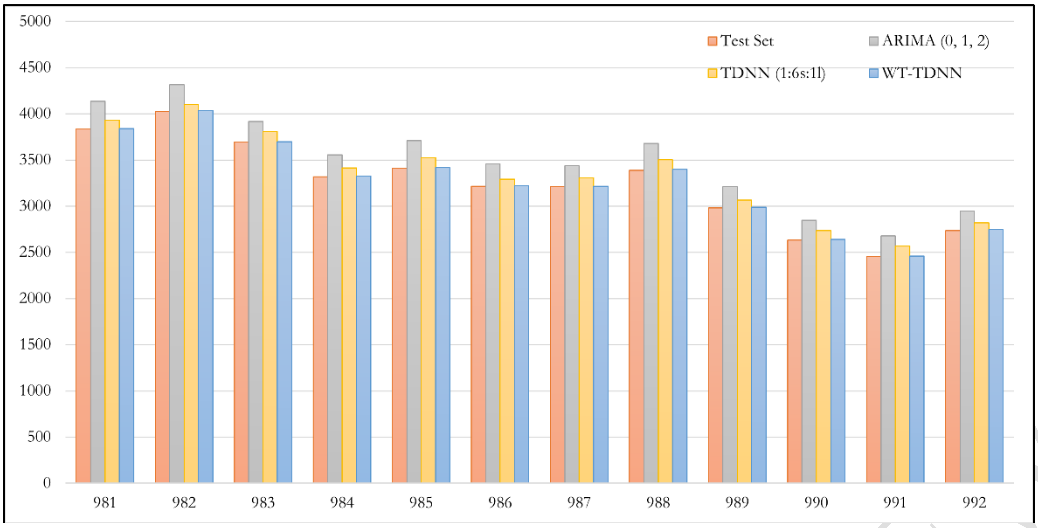


Figure 8: Bar diagram of several point forecasts for tomato (max) price series

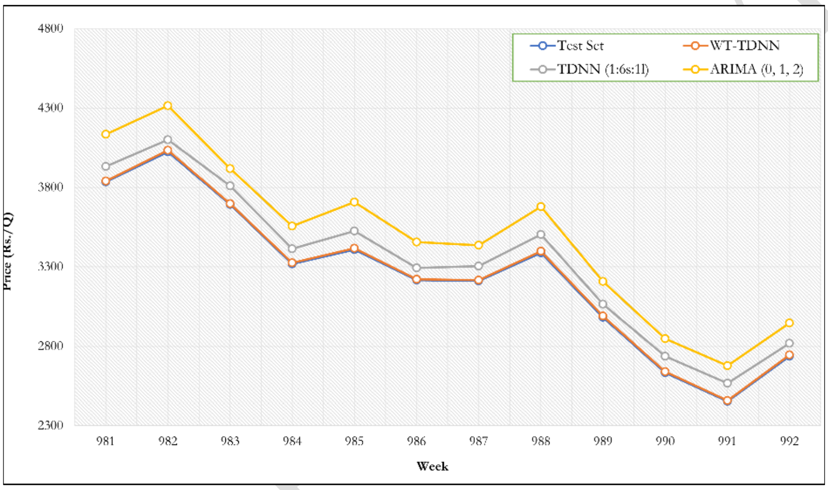


Figure 9: Line diagram of several point forecasts for tomato maximum price series

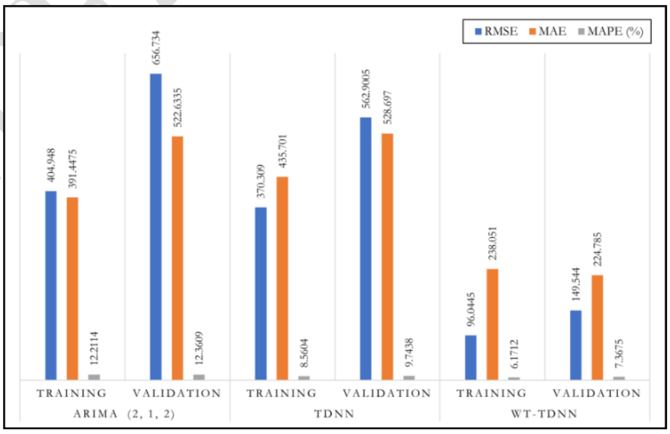


Figure 10: Bar diagram of performance measures for tomato minimum price series

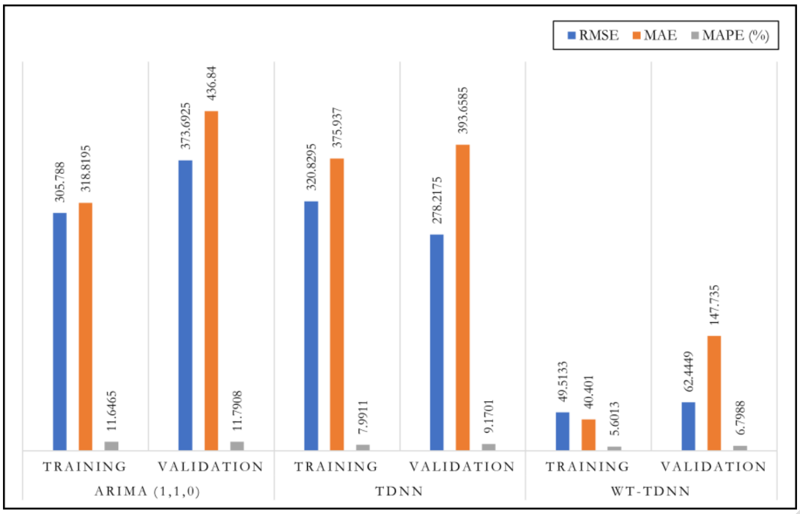


Figure 11: Bar diagram of performance measures for tomato maximum price series

Conclusion

By reducing risks and uncertainties, accurate agricultural price forecasting is essential for assisting farmers, legislators, and government organizations in making decisions. Using a comparative comparison of time series models, such as ARIMA, Time Delay Neural Networks (TDNN), and Wavelet Transformed TDNN (WT-TDNN), this study forecasts weekly tomato prices in Bihar marketplaces.

Differentiating was used to address the price series' volatility, skewness, and non-stationarity. Wald's test was used to validate the model's performance, and ARIMA (1, 1, 1) and ARIMA (0, 1, 2) were found to be the best models for minimum and maximum price series, respectively. The predictive power of the TDNN and WT-TDNN models was further investigated; the results showed that WT-TDNN was the best method for identifying the intricate trends in tomato pricing. Based on performance parameters that repeatedly show WT-TDNN's accuracy and robustness in predicting agricultural price patterns, the study comes to the conclusion that it is superior.

Disclaimer (Artificial intelligence)

Option 1:

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc.) and text-to-image generators have been used during the writing or editing of this manuscript.

Option 2:

Author(s) hereby declare that generative AI technologies such as Large Language Models, etc. have been used during the writing or editing of manuscripts. This explanation will include the name, version, model, and source of the generative AI technology and as well as all input prompts provided to the generative AI technology

Details of the AI usage are given below:

1.

2.

3.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**(6), 716-723.

Alam, W., Ray, M., Kumar, R. R., Sinha, K., Rathod, S. and Singh, K. N. (2018). Improved ARIMAX modal based on ANN and SVM approaches for forecasting rice yield using weather variables. *Indian Journal of Agricultural Sciences*, **88**(12), 1909-1913.

Anjoy, P. and Paul, R. K. (2019). Comparative performance of wavelet-based neural network approaches. *Neural Computing and Applications*, **31**(8), 3443-3453.

Bengio, Y. (2009). Learning deep architectures for AI. *Foundations and trends® in Machine Learning*, **2**(1), 1-127.

Bianchi, F. M., Maiorino, E., Kampffmeyer, M. C., Rizzi, A. and Jenssen, R. (2017). Properties and training in recurrent neural networks. *Recurrent Neural Networks for Short-Term Load Forecasting: An Overview and Comparative Analysis*, 9-21.

Burrus, C. S. (2015). Wavelets and wavelet transforms.

Chen, Q., Lin, X., Zhong, Y. and Xie, Z. (2019). Price Prediction of Agricultural Products Based on Wavelet Analysis-LSTM. In *2019 IEEE International Conference on Parallel and Distributed Processing with Applications, Big Data and Cloud Computing, Sustainable Computing and Communications, Social Computing and Networking*, IEEE, pp. 984-990.

Cho, K., Van Merriënboer, B., Bahdanau, D. and Bengio, Y. (2014). On the properties of neural machine translation: Encoder-decoder approaches. *ArXiv preprint arXiv:1409.1259*.

Daubechies, I. (1992). *Ten Lectures on Wavelets*.Society for industrial and applied mathematics. Philadelphia, USA.

Demuth, H., and Beale, M. (2002). *Neural network toolbox user's guide*. Natic, MA: Mathworks.

Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, **74**(366a), 427-431.

Estrada J. C. G., Silva C., Yanez E., Rodriguez, N. and Calvo, I. P. (2007). Monthly catch forecasting of anchovy Engraulis ringens in the north area of Chile: Non-linear univariate approach. *Fisheries Research*, **86**, 188-200.

Ghamariadyan, M. and Imteaz, M. A. (2021). Monthly rainfall forecasting using temperature and climate indices through a hybrid method in Queensland, Australia. *Journal of Hydrometeorology*, **22**(5), 1259-1273.

Goodfellow, I., Bengio, Y. and Courville, A. (2016). *Deep learning*. MIT press.

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, **9**(8), 1735-1780.

Jaiswal, R., Jha, G. K., Kumar, R. R. and Choudhary, K. (2022). Deep long short-term memory-based model for agricultural price forecasting. *Neural Computing and Applications*, **34**(6), 4661-4676.

Jozefowicz, R., Zaremba, W. and Sutskever, I. (2015). An empirical exploration of recurrent network architectures. In *International conference on machine learning*, Proceedings of Machine Learning Research, pp. 2342-2350.

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), 229-239.

Jha, G. K. and Sinha, K. (2014). Time-delay neural networks for time series prediction: an application to the monthly wholesale price of oilseeds in India. *Neural Computing and Applications*, **24**(3), 563-571.

Kajitani Yoshio, Hipel, K. W. and Mcleod A. I. (2005). Forecasting nonlinear time series with feed-forward neural networks: A case study of Canadian lynx data. *Journal of Forecasting*, **24**, 105-117.

Kumar, S., Arora, K., Singh, P., Gupta, A. K., Sharma, I. and Vatta, K. (2022). Performance comparison of ARIMA and Time Delay Neural Network for forecasting of potato prices in India. *Agricultural Economics Research Review*, **35**(2).

Lama, A., Jha, G. K., Gurung, B., Paul, R. K., Bharadwaj, A. and Parsad, R. (2016). A comparative study on time-delay neural network and GARCH models for forecasting agricultural commodity price volatility, *Journal of the Indian Society of Agricultural Statistics*, **70**(1), 7-18.

LeCun, Y. (1989). *Generalization and network design strategies*. Technical Report CRG-TR-89-4, University of Toronto.

LeCun, Y., Bengio, Y. and Hinton, G. (2015). Deep learning, *Nature*, **521**(7553), 436-444.

McLeod, A. I. and Li, W. K. (1983). Diagnostic checking ARMA time series models using squared residual autocorrelations. *Journal of Time Series Analysis*, **4**, 269-273.

Paul, R. K. and Garai, S. (2021). Performance comparison of wavelets-based machine learning technique for forecasting agricultural commodity prices. *Soft Computing*, **25**(20), 12857-12873.

Paul, R. K. and Garai, S. (2022). Wavelets Based Artificial Neural Network Technique for Forecasting Agricultural Prices. *Journal of the Indian Society for Probability and Statistics*, **23**, 47-61.

Percival, D.B. and Walden, A.T. (2000). *Wavelet methods for time series analysis*. Cambridge University Press, U.K.

Phillips, P. C. B. and Perron, P. (1988). Testing for a unit root in time series regression, *Biometrika*, **75**, 335-346.

Prakash, N. and Farzana, G. S. (2019). Short Term Price Forecasting of Horticultural Crops Using Long Short Term Memory Neural Network. In *International Conference on Computational and Bio Engineering*, Springer, Cham, pp. 111-118.

Qian, Z., Pei, Y., Zareipour, H. and Chen, N. (2019).A review and discussion of Transformation-based hybrid models for wind energy forecasting applications. *Applied Energy*, **235**, 939-953.

Rathod, S. and Mishra, G. C. (2018). Statistical models for forecasting mango and banana yield of Karnataka, India. *Journal of Agricultural Science and Technology*, **20**(4), 803-816.

Rathod, S., Singh, K. N., Patil, S. G., Naik, R. H., Ray, M. and Meena, V. S. (2018). Modeling and forecasting of oilseed production of India through artificial intelligence techniques. *Indian Journal of Agricultural Sciences*, **88**(1), 22-27.

Ray, M., Singh, K. N., Ramasubramanian, V., Paul, R. K., Mukherjee, A. and Rathod, S. (2020). Integration of wavelet transform with ANN and WNN for time series forecasting: an application to Indian monsoon rainfall. *National Academy Science Letters*, ***43***(6), 509-513.

RL, M. and Mishra, A. K. (2021). Forecasting spot prices of agricultural commodities in India: Application of deep-learning models. *Intelligent Systems in Accounting, Finance and Management*, **28**(1), 72-83.

Rumelhart, D. E., Hinton, G. E. and Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, **323**,533-536.

Said, S. E. and Dickey, D. A. (1984). Testing for unit roots in autoregressive-moving average models of unknown order, *Biometrika*, **71**, 599-607.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks*, **61**, 85-117.

Schwarz, G. E. (1978). Estimating the dimension of a model, *Annals of Statistics*, **6**(2), 461–464.

Temizel, Tugba T. and Casey, M. C. (2005). A comparative study of autoregressive neural network hybrids. *Neural Networks*. **18**, 781-789.

Yu, P. and Yan, X. (2020). Stock price prediction based on deep neural networks. *Neural Computing and Applications*, **32**, 1609-1628.