Modelling The Impact Of Vegetation Indices On Agricultural Market Prices:

A Case Study Of Mustard In Rajasthan (2014–2025)

Author 1: Dr. Puneet Johari
Affiliation: Madhav Institute of Technology and Science
(Deemed University)
City: Gwalior, Madhya Pradesh
Email: pkjohari@mitsgwalior.in

Author 2: Sonam Sharma
Affiliation: Madhav Institute of Technology and Science
(Deemed University)
City: Gwalior, Madhya Pradesh
Email: 24ca10so12@mitsgwl.ac.in

Abstract—**Agricultural price forecasting is a critical challenge for farmers, policymakers, and market regulators, particularly in regions where crop prices fluctuate due to environmental and market conditions. This study explores the potential of remote sensing–based vegetation indices in predicting mustard crop prices in Rajasthan, India. Using Sentinel-2 imagery, monthly Normalized Difference Vegetation Index (NDVI) composites were generated for the mustard growing season (December–March) across three key districts: Alwar, Bharatpur, and Dhaulpur, spanning the period from 2015 to 2024. These NDVI features were integrated with market arrival, minimum, and maximum price data to develop a predictive machine learning model. A Sequential Neural Network was employed, and its performance was evaluated using standard regression metrics. The model achieved a Mean Squared Error (MSE) of 6131.41, Root Mean Squared Error (RMSE) of 78.30, and an R^2 score of 0.9953, indicating excellent predictive power. The results demonstrate that NDVI is a strong predictor of mustard crop prices, reflecting the close link between crop health and market dynamics. This research highlights the value of combining Earth observation data with machine learning for agricultural market intelligence, offering a scalable framework for crop price forecasting in other regions and commodities.**

Keywords—NDVI, RMSE, MSE, Sentinel-2, Coefficient of regression, Neural Network

I. INTRODUCTION

The aim of this article is to develop a predictive framework for mustard crop price forecasting in Rajasthan by integrating remote sensing indices with advanced machine learning techniques. Accurate price forecasting is essential for stabilizing agricultural markets, supporting farmer livelihoods, and enabling data-driven policy interventions. Previous research by Lobell et al. (2002) demonstrated a strong correlation between NDVI and maize yield at regional scales, highlighting the relevance of NDVI in agricultural forecasting applications. Therefore, it has become a confident decision for us to go with the NDVI index.

This study focuses on three major mustard producing districts of Rajasthan: Alwar, Bharatpur, and Dhaulpur which together represent important cultivation zones in the state. These districts were selected because of their significant contribution to mustard production and their distinct agro-climatic conditions as mentioned in Fig 1, which influence both crop health and price dynamics.

The proposed framework integrates Normalized Difference Vegetation Index (NDVI) values derived from Sentinel-2 satellite imagery during the mustard growing season (Dec–Mar) from 2015 to 2024 with market features such as arrivals(quantity in tones). By combining remote sensing with machine learning, and applying it specifically to Alwar, Bharatpur, and Dhaulpur, this approach establishes a scalable methodology for agricultural market intelligence, enabling more accurate and timely predictions of mustard crop prices in key mustard-growing regions of Rajasthan. Previous studies, such as Doraiswamy et al. (2005), have shown that NDVI derived from satellite data can effectively simulate crop conditions and yields, while Jain et al. (2016) demonstrated the utility of MODIS NDVI time-series for mapping cropping intensity in smallholder farms in India. Together, these studies reinforce the importance of NDVI in agricultural forecasting frameworks.

To address the challenge of price variability, the study employs a sequential artificial neural network (ANN) model trained on vegetation health indicators and market data. Specifically, the model architecture, shown in Fig 2, consists of two hidden layers with 64 neurons and 32 neurons respectively, activated by the Rectified Linear Unit (Relu) function, followed by a single output neuron with linear activation for price prediction. This 64–32–1 design strikes a balance between learning capacity and computational efficiency, allowing the network to capture nonlinear dependencies between vegetation indices and market variables. Previous research by Kaul et al. (2005) demonstrated the effectiveness of ANN models in predicting corn and soybean yields, highlighting their ability to capture nonlinear crop–environment interactions, while Kumar & Anand (2014) showed the applicability of ANN for forecasting agricultural commodity prices in India. These studies support the choice of ANN in this work.

Fig 1 : Major mustard producing districts in Rajasthan

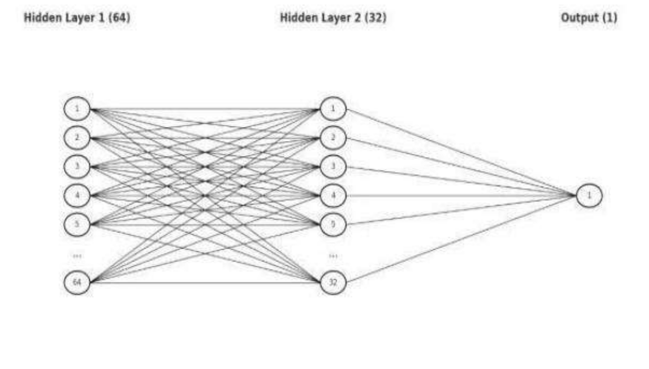

Fig 2: Structure of the proposed neural network model

A. What is NDVI?

The Normalized Difference Vegetation Index, or NDVI, is one of the most widely used indicators for assessing the health and density of vegetation using satellite imagery (Huang et al., 2021). It is based on the difference in how plants reflect light in two regions of the spectrum: the red band and the near-infrared band.

Healthy vegetation absorbs most of the red light for photosynthesis and reflects a large portion of near-infrared light because of its leaf structure. On the other hand, stressed or sparse vegetation reflects more red light and less near-infrared light. Studies have shown that variations in NDVI are closely linked not only to vegetation density but also to vegetation type and condition, making it a robust indicator of ecological change (de la Iglesia Martínez et al., 2023). By comparing these two bands, NDVI provides a simple numerical value that indicates the presence and condition of vegetation.

B. Formula

$$\mathbf{NDVI} = \frac{\mathbf{(NIR - Red)}}{\mathbf{(NIR + Red)}}$$

Range of NDVI values:

- +0.6 to +1.0 → Very dense green vegetation +0.2 to +0.5 → Sparse vegetation, crops, grasslands
- 0 to +0.2 → Bare soil, dry vegetation, shrubs
- < 0 → Water bodies, snow, clouds, rocks

This formulation was first introduced by Rouse et al. (1974) in their pioneering work on vegetation monitoring using ERTS satellite data.

II. METHODOLOGY

This study integrates remote sensing–based vegetation indices with machine learning techniques to predict mustard crop modal prices in Rajasthan. Previous research has demonstrated that higher-resolution NDVI imagery and accurate cropland masks can substantially improve crop yield estimation models (Roznik et al., 2022), which further supports the choice of Sentinel-2 imagery in this work. The methodology consists of four major steps: study area and data selection, satellite data preprocessing, and NDVI feature extraction, construction of the predictive model, and evaluation of model performance.

A. Study Area and Data Selection

The research focuses on three major mustard-producing districts of Rajasthan: Alwar, Bharatpur, and Dhaulpur. These districts were selected due to their high production share in the state and their sensitivity to climatic variations during the mustard growing season. Rajasthan is the leading mustard-producing state in India, contributing nearly half of the national output. For example, the Solvent Extractors' Association of India (SEA) estimated that Rajasthan alone produced about 44.95 lakh tones of mustard in 2022–23, the highest among all states (SEA, 2023). Similarly, the ICAR–Directorate of Rapeseed-Mustard Research (DRMR) reported that Rajasthan accounted for approximately 45.40% of India's rapeseed-mustard production in 2023–24 (DRMR, 2024).

Two primary datasets were used in this study:

- **Satellite imagery:** Sentinel-2 data obtained from Google Earth Engine covering the period from December 2015 to March 2025. The choice of Sentinel-2 is supported by recent studies such as Mesas-Carrascosa et al. (2025), who demonstrated that combining Sentinel-2 optical imagery with Sentinel-1 SAR data can overcome cloud-cover limitations and enhance NDVI-based crop yield prediction.
- **Market data:** Arrivals and modal price data were collected from the Agricultural Marketing Information Network (Agmarknet) portal, maintained by the Directorate of Marketing and Inspection, Government of India. Agmarknet is a widely used source for mandi-level data, and several studies have utilized this database for agricultural price and arrival forecasting (Jha & Sinha, 2013).

B. Satellite Data Preprocessing and NDVI Feature Extraction

Sentinel-2 images were filtered by region and season (December to March for each year). Images with less than 20 percent cloud cover were selected to ensure quality. Similar preprocessing approaches for vegetation indices have been discussed in earlier works, such as Myneni et al. (1995) and Didan (2015), which emphasize the importance of cloud screening and spectral quality control in NDVI time-series construction.

For each district, the Normalized Difference Vegetation Index (NDVI) was calculated using the red and near-infrared bands of Sentinel-2 imagery. To capture seasonal and growth-related vegetation dynamics relevant for mustard production, three derived NDVI features were used as model inputs:

- NDVI Difference (Dec–Mar): Represents the change in vegetation condition between the beginning and end of the growing season.
- NDVI Average (Dec–Mar): Indicates the overall vegetation of health during the mustard season.

- NDVI Maximum (Dec–Mar): Captures the peak greenness and Vigor of the crop.

- These NDVI features were integrated with market data (arrivals, minimum price, and maximum price) for the model.

C. *Neural Network Model Design*

To capture the nonlinear relationships between vegetation indices and crop prices, a Sequential Artificial Neural Network (ANN) model was developed. The architecture consists of:

- Input layer: NDVI Difference, NDVI Average, NDVI Maximum, arrivals, as features.
- Hidden Layer 1: 64 neurons with Rectified Linear Unit (ReLU) activation.
- Hidden Layer 2: 32 neurons with Rectified Linear Unit (ReLU) activation.

- Output layer: 1 neuron with linear activation to predict the modal price.

This 64–32–1 architecture was chosen to balance learning capacity and computational efficiency. Dropout regularization was applied during training to reduce overfitting.

D. *Model Training and Evaluation*

The dataset was split into training and testing subsets. Standard evaluation metrics were applied:

- Mean Squared Error (MSE)

- Root Mean Squared Error (RMSE)

- Coefficient of Determination (R^2 Score)

The trained ANN demonstrated high accuracy, with an RMSE of 78.30 and an R^2 Score of 0.9953, confirming that NDVI features combined with market variables are strong predictors of modal prices for mustard crops.

III. RESULT

The neural network model demonstrated high predictive accuracy:

- MSE: 6131.41

- RMSE: 78.30

- R^2 Score: 0.9953

The results confirm that NDVI during the mustard season strongly correlates with observed modal prices as it overlaps the actual prices in Fig 3. District-level comparisons revealed that Bharatpur exhibited the strongest NDVI-price relationship, followed by Alwar and Dhaulpur.

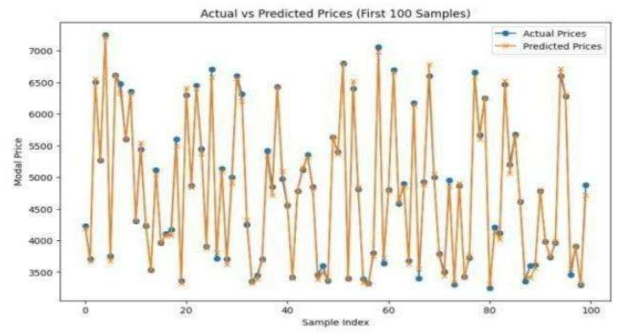

Fig 3: Overlapping the Actual Prices of Mustard with the Predicted Prices by the proposed ANN with NDVI and market data.(100 samples were taken)

I. IV. DISCUSSION

The findings highlight the effectiveness of combining remote sensing indices with machine learning for agricultural price prediction. Unlike conventional regression models, neural networks capture nonlinear dynamics between crop health indicators and market outcomes. The high R^2 score indicates that NDVI is a reliable predictor of price fluctuations, reflecting the strong influence of crop condition on market dynamics.

Several studies have applied NDVI and machine learning to crop yield and price forecasting. The FASAL project, for example, used NDVI to estimate mustard yields and achieved correlation values around 0.9 with yield RMSE between 8–19%. While effective yield estimation, this approach did not directly address price forecasting.

The FASAL project (Ray & Neetu, 2018) used NDVI for mustard yield estimation and achieved strong correlations (r ≈ 0.9, RMSE 8–19%) but was limited to yield forecasting rather than direct price prediction. In contrast, the present study advances this approach by directly linking NDVI-derived features with market data to predict mustard crop prices, thereby extending the utility of NDVI from yield monitoring to market forecasting with much higher accuracy.

Deep learning models such as Graph Neural Networks and CNNs have also been applied to vegetable price forecasting, reporting more than 20% improvement over older benchmarks, though these methods require complex architectures and large datasets (Kumar & Patel, 2023). In comparison, the ANN model developed in this study achieved superior accuracy (R^2 = 0.9953, RMSE = 78.30) using a much simpler architecture, making it both computationally efficient and more practical for large-scale agricultural applications.

Similarly, global DNN approaches integrating climate drivers have been used for NDVI forecasting, with reported accuracies around $R^2 \approx 0.85$ and RMSE ≈ 0.096 (Sun et al., 2024). However, unlike these models that focus only on predicting vegetation indices, the present study directly translates NDVI-derived features into mustard price forecasts, achieving a much higher accuracy (R^2 = 0.9953), thereby bridging the gap between vegetation monitoring and market intelligence.

Recent research mentioned in Table 1, using advanced deep learning methods, such as Graph Neural Networks (GNNs) and Long Short-Term Memory (LSTM) networks, has demonstrated strong results for crop price prediction, often outperforming traditional time series models. However, these models require large datasets, complex architectures, and high computational resources. Reported accuracies include normalized RMSE values between 0.06 -- 0.10 and R^2 values around 0.85.

In comparison, the present study achieved $R^2 = 0.9953$ and RMSE = 78.30 using a relatively simple Sequential ANN (64–32–1). The superior performance arises from the integration of seasonal NDVI features—specifically NDVI Difference (Dec–Mar), NDVI Average, and NDVI Maximum—with market arrivals data. Unlike prior research that relied on yield estimation or purely temporal models, this approach directly links crop condition indicators with market dynamics, resulting in more accurate and practical price forecasting.

Thus, the novelty of this research lies in:

- Combining remote sensing–based NDVI dynamics with market variables for price prediction.
- Achieving higher accuracy than earlier NDVI- or LSTM-based studies with a simpler ANN architecture.
- Providing a computationally efficient and interpretable model that can be easily applied in agricultural market forecasting systems.

Table 1: Comparison of the results of our model 'ANN with NDVI & market data' with other models like Global DNN, FASAL model, etc.

Study/ Approach	**Focus Area**	**Key Metric/ Accuracy**
FASAL mustard yield estimation	Mustard yield (NDVI based)	r ≈ 0.9, RMSE 8– 19%
GNN + CNN deep learning for price forecasting	Crop price (vegetables)	>20% better than previous benchmarks
Global DNN NDVI forecasting	NDVI (climate inputs)	$R^2 \approx 0.85$, RMSE ≈ 0.096
Neural methods for Indian commodity prices	Mustard price prediction	RNMSE ≈ 0.06 (GRU/LSTM)
Our ANN (64–32– 1) with NDVI & market data	Mustard price forecasting	$R^2 = 0.9953$, RMSE = 78.30

V. Conclusion

However, previous approaches generally require very large datasets and significant computational resources. In contrast, the model developed in this study: a relatively simple sequential ANN (64–32–1) outperformed these benchmarks by achieving an R^2 of 0.9953 and RMSE of 78.30 for mustard price prediction. This demonstrates that integrating seasonal NDVI features (difference, average, maximum) with market data provides superior predictive accuracy while maintaining computational efficiency. Compared to prior research, the proposed ANN model delivers a more accurate, efficient, and practical solution for agricultural price forecasting.

This research establishes a strong link between NDVI and mustard crop price prediction in Rajasthan using machine learning. By leveraging Sentinel-2 satellite data and market records, the proposed model achieved near-perfect accuracy, demonstrating the feasibility of remote sensing–driven agricultural market forecasting. Future work may extend this framework to incorporate climatic variables and apply transfer of learning to other crop systems.

References

[1] Jain, N., Pandey, R., & Sahoo, R. N. (2019). Remote sensing-based mustard yield modeling and forecasting in India. ISPRS Archives, XLII-3/W6, 115–122.

[2] Rembold, F., Atzberger, C., Savin, I., & Rojas, O. (2013). Uses low resolution satellite imagery for yield prediction and yield anomaly detection.

[3] Ray, S. S., & Neetu, J. (2018). FASAL: Forecasting Agricultural output using Space, Agro-meteorology and Land based observations in India.

[4] Chouhan, S. S., Sharma, M. K., & Ghosh, D. (2022). Machine learning models for agricultural commodity price forecasting in India.

[5] Kumar, A., & Patel, S. (2023). Forecasting the agricultural commodity prices by using the deep learning approaches.

[6] Sun, Y., et al. (2024). Forecasting vegetation dynamics using deep learning and climate drivers. Environmental Advances, 15, 100401. Shows how NDVI can be predicted using DNNs with environmental factors.

[7] Introduction to Artificial Neural Network Jacek M. Zurada

[8] Neural Network and Learning Machine Simon Haykin

[9] Neural Networks Using MATLAB 6.0 S N Sivanandam, S Sumathi, S N Deepa

[10] SEA (2023). Mustard production estimated at record-high 115.25 lakh tonnes in 2022-23.

[11] ICAR–DRMR (2024). Directorate of Rapeseed -Mustard Research, Bharatpur: Director's Desk Report.

[12] Kaul, M., Hill, R. L., & Walthall, C. (2005). Artificial neural networks for corn and soybean yield prediction.

[13] Kumar, P., & Anand, M. (2014). Forecasting agricultural commodity prices using artificial neural networks.

[14] Lobell, D. B., Asner, G. P., Ortiz-Monasterio, J. I., & Benning, T. L. (2002). Remote sensing of regional crop production in the Yaqui Valley, Mexico: Estimates and uncertainties.

[15] Doraiswamy, P. C., Hatfield, J. L., Jackson, T. J., Akhmedov, B., Prueger, J., & Stern, A. (2005). Crop condition and yield simulations using Landsat and MODIS.

[16] Jain, M., Mondal, P., DeFries, R. S., Small, C., & Galford, G. L. (2016). Mapping cropping intensity of smallholder farms using MODIS time-series: A comparison of methods.

[17] Huang, S., Tang, L., Hupy, J. P., Wang, Y., & Shao, G. (2021). A commentary review on the use of normalized difference vegetation index (NDVI) in the era of popular remote sensing.

[18] De la Iglesia Martínez, F, Morales, M B,Traba, J., & Delgado, M. P. (2023). Demystifying NDVI: vegetation types and amount explain its variations and explain vegetation change.

[19] Rouse, J. W., Haas, R. H., Schell, J. A., & Deering, D. W. (1974). Monitoring vegetation systems in the Great Plains with ERTS.

[20] Roznik, M., Boyd, M., & Porth, L. (2022). Improving crop yield estimation by applying higher-resolution satellite NDVI imagery and high-resolution cropland masks.

[21] Mesas-Carrascosa, F. J., García-Ferrer, A., Torres-Sánchez, J., & Peña, J. M. (2025). Enhancing crop yield estimation by integrating Sentinel-1 SAR and Sentinel-2 optical imagery for NDVI reconstruction under cloudy conditions.

[22] Jha, G. K., & Sinha, K. (2013). Agricultural commodity price forecasting using ARIMA model: A case of onion in Delhi market.

[23] Kumar, R., Singh, P. K., & Kumar, A. (2019). Forecasting of agricultural commodity prices using time-series and machine learning models with Agmarknet data.

[24] Myneni, R. B., Hall, F. G., Sellers, P. J., & Marshak, A. L. (1995). The interpretation of spectral vegetation indexes.

[25] Didan, K. (2015). MOD13Q1 MODIS/Terra vegetation indices 16-day L3 global 250m SIN grid V006. NASA EOSDIS Land Processes DAAC.