Vol. 24, No. 2, pp. 177-189 (2024)
Journal of Agricultural Physics
ISSN 0973-032X
http://www.agrophysics.in

Research Article

Evaluating MLR and ANN for Mustard Yield Forecasting in the Tarai Region of Uttarakhand

KARAN KUMAR[1,*], RAVI KIRAN[2] , SUBASH NATARAJA PILLAI[1], ANURAG SATPATHI[3] AND SUMAN[1]

[1]Division of Agricultural Physics, ICAR-Indian Agricultural Research Institute, New Delhi-110012
[2]Department of Agrometeorology, G.B Pant University of Agriculture and Technology, Pantnagar, U.S Nagar-263145, Uttarakhand
[3]Division of Agrometeorology, Sher-e-Kashmir University of Agricultural Sciences and Technology of Kashmir, Srinagar-190025, Jammu and Kashmir

ABSTRACT

The present study focuses on developing and comparing mustard yield prediction models using statistical and machine learning approaches in the Udham Singh Nagar district of Uttarakhand, India. A statistical model based on Stepwise Multiple Linear Regression (SMLR) and Artificial Neural Network (ANN) was constructed using 21 years of mustard yield and weather data (2001–2021). Both the models incorporated weighted and unweighted weather indices, exploring the relationship between meteorological factors—such as temperature, rainfall, humidity, sunshine hours, and wind speed—and mustard yield. The SMLR model performed well during calibration (R^2 = 0.72, nRMSE < 30%) but showed reduced accuracy during validation (R^2 = 0.32). The ANN model used crop yield and weather parameters across key phenological stages of mustard growth. The ANN model achieved a high R^2 value of 0.84, demonstrating a strong correlation between predicted and observed yields during testing. The validation phase also showed promising results, with an R^2 value of 0.71 and low RMSE and nRMSE values, reflecting its reliability in forecasting yield. Key independent variables influencing yield were identified, including sunshine hours, wind velocity, and their interactions, highlighting the model's ability to capture complex, non-linear relationships. While the ANN model slightly overestimated yields during calibration and underestimated them during validation, these deviations were minimal, indicating the model's robustness and suitability for practical applications. The findings underscore the ANN model's superiority in predicting mustard yield, offering a reliable tool for real-time estimation, improved agricultural planning, and informed decision-making. This highlights the potential of machine learning approaches to advance sustainable agriculture in the region and future research should delve into advanced deep learning methods, including Convolutional Neural network (CNN), Deep Neural Networks (DNN), and Recurrent Neural Networks (RNN), individually or in combination, to enhance crop yield prediction.

Key words: Mustard, Yield forecasting, Machine learning, ANN

Introduction

Rapeseed-mustard, an important oilseed crop, holds a significant position in global agriculture and the Indian economy. With an annual global production of approximately 40 million tonnes, it is considered as a crucial crop for edible oil production for mankind. In India, rapeseed mustard plays a prominent role in the agricultural sector, contributing substantially to the country's oilseed production.

*Corresponding author,
Email: kk5299.kk@gmail.com

According to the Directorate of Oilseeds Development, total oilseeds production increased from 280.50 lakh tonnes in 2011-12 to 361.01 lakh tonnes in 2020-21, with a peak of 365.65 lakh tonnes in 2019-20. The yield of total oilseeds increased from 1066 kg/ha in 2011-12 to 1154 kg/ha in 2020-21, with a peak of 1284 kg/ha in 2017-18.

In the past, assessing crop productivity required labour-intensive crop-cutting trials that required a significant amount of human resources. Predicting crop yields via crop yield models, which may be formulated using a variety of statistical and machine learning approaches, is a modern alternative to this traditional approach. Setiya *et al.* (2022) claimed that the SMLR strategy is the simplest method to develop a yield forecast model using a dataset of yield and meteorological characteristics. This technique enables the selection of the top predictors from a vast pool of predictors**.** Statistical models have been widely used in agriculture for growth and yield prediction. The method of stepwise linear regression allows for the simultaneous deletion of unimportant variables while performing many regressions. In essence, stepwise regression repeats multiple regression many times while eliminating the variable with the least association each time. What is left at the end is the set of elements that best explains the distribution. The data must be regularly distributed, and the independent variables must not be correlated, as the only requirements. Stepwise regression combines the forward and backward selection methods, and variable selection is done automatically. Before adding them to the final model, all predictor variables in the approach are examined to see if their significance has decreased below the defined tolerance threshold. A variable is eliminated from the model if it is determined to be non-significant. The simplest method for creating a yield forecast model based on a dataset of yield and weather characteristics is the stepwise multiple linear regression (SMLR) technique. This strategy enables the selection of the top predictors from a vast pool of predictors (Das *et al.*, 2018).

An artificial neural network, a type of non-linear machine learning technique, has three layers: input, hidden, and output. In this method, information travels from the input layer through the hidden layer to the output layer (Kaul *et al.*, 2005). The number of independent predictors affects how many nodes there are in the input layer. Artificial neural networks employ learning algorithms that may, in a sense, learn when they are exposed to fresh data. They thus develop into a very potent tool for non-linear statistical modelling. The use of artificial neural networks to forecast crop yield utilizing multiple crop performance indicators is reviewed in this research. The application of ANN and the foundations of neural network design are also covered (Khairunniza-Bejo *et al.*, 2014). In comparison to previous methodologies, ANN has been shown to offer a superior understanding of crop variability.

The researchers have made numerous efforts to construct pre-harvest yield forecast models based on yield and meteorological data (Das *et al.*, 2020; Aravind *et al.*, 2022) . Using weather data, a model constructed using multiple linear regression equations, artificial neural networks, randomized and ridge regression approaches, and other techniques has the potential to give a reliable, timely, and cost-effective prediction of rice yield. Pre-harvest agricultural yield prediction utilizing Artificial Neural Network (ANN), Least Absolute Shrinkage and Selection Operator (LASSO), and Elastic Net (ELNET) are receiving a lot of interest at the present time.

The use of Artificial Neural Networks (ANNs), Fuzzy Systems, and Genetic Algorithms, among other Artificial Intelligence (AI) applications, has demonstrated greater effectiveness in handling the issue. Several studies have highlighted the poor performance of multi-linear regression (MLR) based PTFs, as these models fail to account for the non-linear relationships between input and output variables (Bhattacharya *et al.*, 2021; Sarkar *et al.*, 2023). Since real-world data predominantly exhibit non-linear characteristics, artificial intelligence (AI)-based machine learning techniques such as artificial neural network (ANN) and support vector machine (SVM) etc, which utilize pattern recognition methods, can serve as effective tools. Studies have shown that approximately 97% of PTFs are based on empirical equations, while only 3% employ pattern recognition techniques (Amanabadi *et al.*, 2019). AI-based models leveraging pattern recognition have demonstrated higher efficiency compared to empirical PTFs (Bhattacharya *et al.*, 2018). Moreover, in AI-based machine learning models, prior knowledge of the relationship between input and output variables is not a prerequisite (Gocic *et al.*, 2015). The use of ANN can simplify and improve the accuracy of models derived from intricate natural systems with numerous inputs. Using ANNs, multiple crop yield prediction models have been developed. If we create an artificial neural network (ANN) that correctly learns the relationships between the effective climatic parameters and crop output, we can use it to estimate crop production in the short- and long term, and we can also create an ANN model for each region. Additionally, the most beneficial crop yield parameters can be found by utilizing ANNs. Therefore, some parameters whose measurements are challenging and economical can be ignored. An incredibly adaptable method of artificial neural networks is quickly emerging to manage such a problem. A feed-forward back propagation artificial neural network is the most often used ANN. As an example, the method has been used to model and predict different crop yields based on a variety of predictor variables, including soil type, pH, nitrogen, phosphate, potassium, organic carbon, calcium, magnesium, sulphur, manganese, copper, iron, depth, temperature, rainfall, and humidity (Dai *et al.,* 2011). Consideration has been given to ANN with zero, one, and two hidden layers. Computing MSEs has allowed researchers to determine the ideal number of hidden layers and units within each hidden layer. To address the multifaceted challenges associated with crop yield prediction, this study employs Stepwise Multiple Linear Regression (SMLR) and Artificial Neural Network (ANN) methodologies. These advanced modeling approaches were utilized to enhance the precision and robustness of mustard yield estimation.

Materials and Methods

Study area

This study centers on a prominent mustard-producing Udam Singh Nagar district located within the Tarai belt of Uttarakhand (Fig.1). The selection of this district is informed by its critical role in the

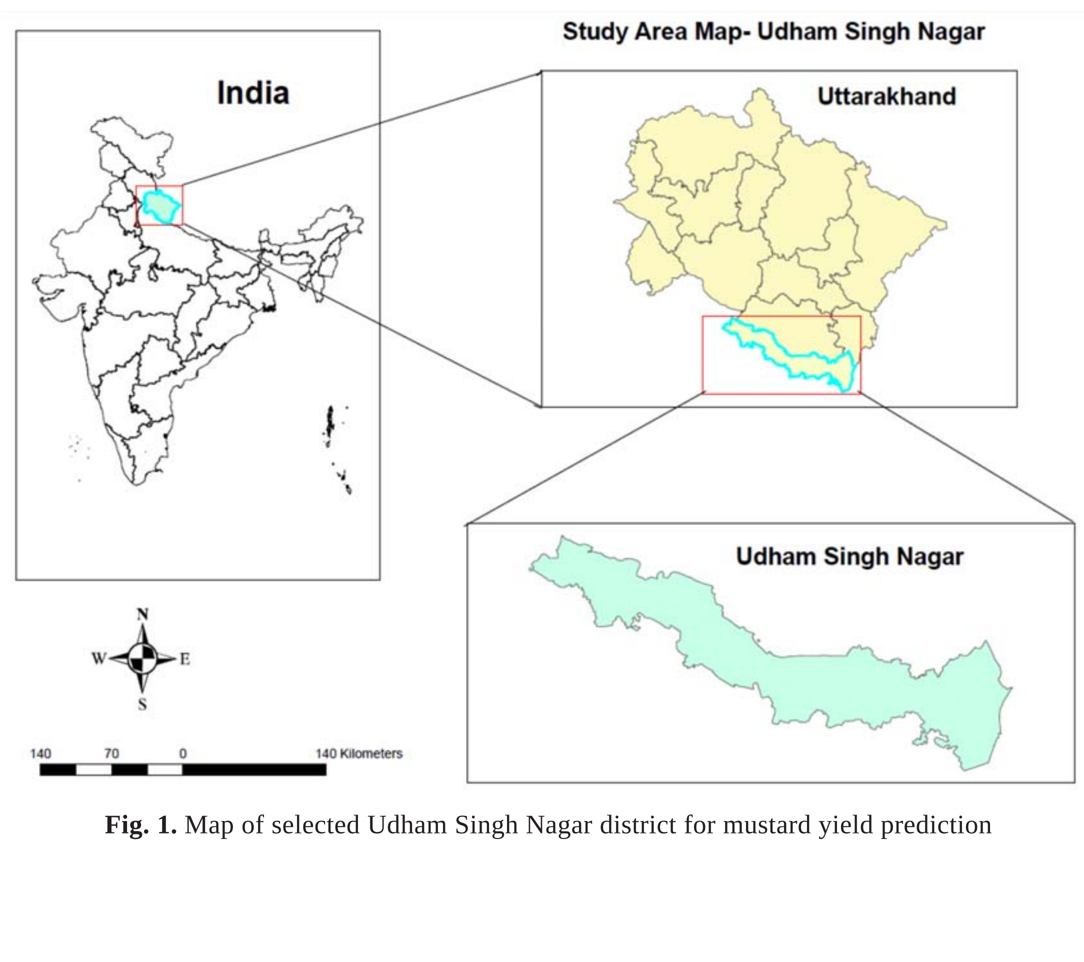


Fig. 1. Map of selected Udham Singh Nagar district for mustard yield prediction

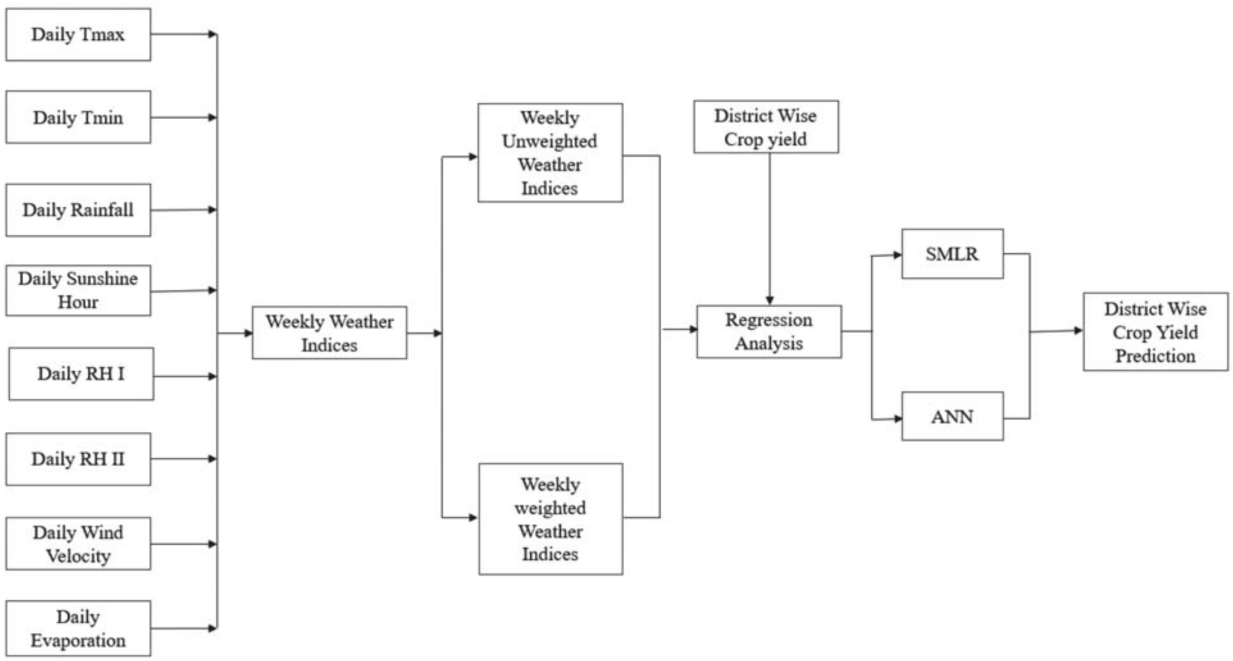


Fig. 2. Step involved in model development

mustard production landscape of Uttarakhand, as documented by the Directorate of Economics and Statistics, Ministry of Agriculture, Government of India (DACNET, 2021–2022).

Software used

1. Statistical Package for Social Sciences (SPSS)

For the management and statistical analysis of social science data, the SPSS Version 27 software package was used. Figure 2 illustrated the procedures required in development of statistical models.

1.1 Simple or unweighted and weighted weather indices

Simple or unweighted weather indices are numerical values that represent the sum of the different weather variables affecting the crop during a certain period, without assigning any specific weights to them and Weighted weather indices are numerical values that represent the sum of the different weather variables affecting the crop during a certain period, with reference to their respective influences on the variable to be predicted during each week (Aditya *et al.*, 2012; Ghosh *et al.*, 2014). Simple and weighted weather indices have been formulated for the Udham Singh Nagar district of Uttarakhand. The computation of simple and weighted weather indices was based on the following formulas provided by Das *et al.* (2020).

Unweighted weather indices

$$Zij = \sum_{w=1}^{n} X_{iw'}$$

$$Z_{ii'j} = \sum_{w=1}^{n} X_{iw} X_{i'} w'$$

Weighted weather indices

$$Zij = \sum_{w=1}^{n} r_{iw}^{j} X_{iw'}$$

$$Z_{ii'j} = \sum_{w=1}^{n} r_{ii'w}^{j} X_{iw} X_{i'} w'$$

Where, Here, Z represents the weather index, n is the week of the forecast, $X_{iw}/X_{i'w}$ is the value of the i^{th}/i^{2th} weather variable, the value of j is 0 for all unweighted indices and 1 for all weighted indices, and $r^{j}_{iw}/r^{j}ii'w$ is the value of the correlation coefficient of the detrended yield with the i^{th} weather variable/ product of the i^{th} and i'^{th} weather variables in the w^{th} week.

1.2 Stepwise Multiple Linear Regression (SMLR)

Stepwise multiple linear regression (SMLR) is the simplest method for generating a yield forecast based on a dataset of yield and weather parameters. Through a series of automated procedures, this technique assists in choosing the best predictors from a large pool of predictors. The t-statistics and p-value are commonly used to evaluate the importance of the new variable included in subsequent steps at each stage (Singh *et al.*, 2014; Das *et al.*, 2018).

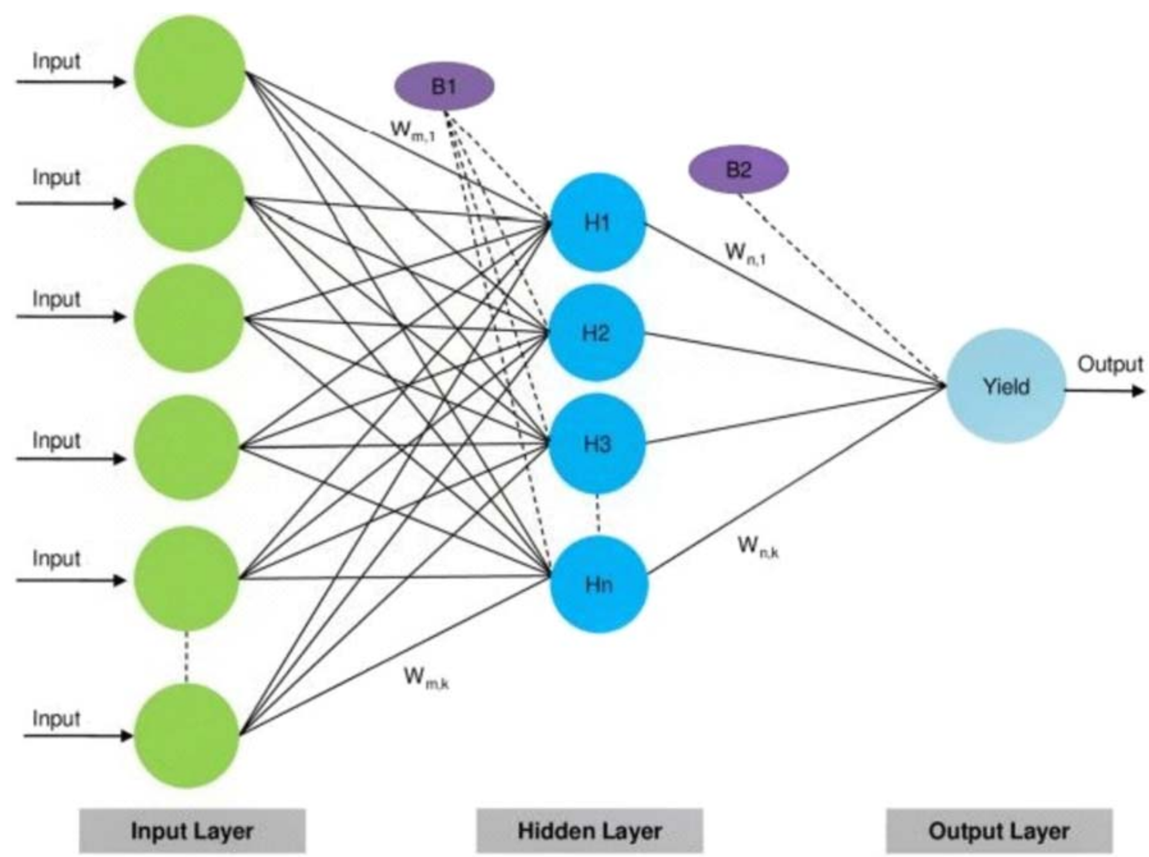
In the current study, the addition and removal of the variables were both taken into consideration with p-values of 0.10. For the US Nagar district of Uttarakhand, meteorological data from the previous 21 years were gathered along with historical mustard yield data from the Directorate of Economics and Statistics, Ministry of Agriculture and Farmers Welfare (https://aps.dac.gov.in/), in order to create a model using SMLR.

2. R Studio

R Studio (Version 2023.06.1 Build 524) is a free, open-source IDE (integrated development environment) for R. The end user's view of graphs, data tables, R code, and output can all be readily seen simultaneously thanks to the interface's organization.

2.1 Artificial neural network (ANN)

An Artificial Neural Network (ANN) is an information-processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information (Kaur and Sharma, 2019) (Fig. 3). The key element of this paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements (neurons) working in unison to solve specific problems. In a neural network, the neurons in the hidden layer play a crucial role. They use an activation function to transform the neuron's



Where, B1 and B2 are bias corrections ; Wm,(1…k) and Wn, (1…k) are connecting weights. Analysis using ANN was carried out by fixing 70% of the data for calibration and the remaining dataset for validation.

Fig. 3. Basic three-layer ANN structure used in the study

activation level into an output signal. This is where most of the important computations happen. The processed outputs from the hidden layer are then passed to the output layer. The number of neurons in the input layer is determined by the number of independent predictors in the dataset. The output (h_i) of a neuron in the hidden layer can be mathematically described, as explained by Wang and Wang (2003).

$$h_i = \sigma\left(\Sigma_{j=1}^{N} V_{ij} x_j + T_i^{hid}\right)$$

Here, σ is the activation function, N is the quantity of input neurons, V_{ij} is the weights, x_j is the neuron inputs and T_i^{hid} is the hidden neurons' threshold. In this neural network model, the hidden layer consists of 3 neurons, and the training process is configured to run for a maximum of 100 iterations. ANNs, like people, learn by example. An ANN is configured for a specific application, such as pattern recognition or data classification, through a learning process.

3. Evaluation of Models Performance

R^2, root mean square error (RMSE), normalized root mean square error (nRMSE), and mean biased error (MBE) were used to assess the performance of the models. The value of R^2 close to 1 and the value of RMSE and MBE near to 0 indicate better model performance. In addition to this, the model performance is scored excellent, good, fair, and poor based on the value of nRMSE lies between 0-10%, 10-20%, 20-30% and >30% respectively. Formulas of the following measures are mentioned below.

$$R^2 = \left(\frac{\frac{1}{n}\Sigma_{i=1}^{n}(y_i - \overline{y})(\hat{y}_i - \overline{\hat{y}})}{\sigma_y \sigma_{\hat{y}}}\right)^2$$

$$\text{RMSE} = \sqrt{\frac{\Sigma_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}}$$

$$nRMSE = \sqrt{\frac{\Sigma_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \times \frac{100}{\overline{A}}$$

$$MBE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)$$

Where,

y_i is the observed value and

$\hat{y}_i$ is the predicted value for i=1, 2……, n.

σ_y and $\sigma_{\hat{y}}$ are the standard deviation of actual and predicted observations respective.

Results and Discussion

Development of mustard yield prediction model using statistical approach

In the context of mustard crop production, weather conditions play a significant role in determining the yield. A model based on linear regression utilizing weather data has been developed to accurately predict mustard yield in a reliable, timely, and cost-effective manner. To create this model, data on mustard yield and weather parameters were gathered over a period of 21 years (2001-2021) in the Udham Singh Nagar district of Uttarakhand. The stepwise multiple linear regression (SMLR) technique was employed for model development.

This part of the study finds the correlations between meteorological indices and mustard yield. Through this analysis, it seeks to uncover how particular weather conditions influence the growth and productivity of mustard crops.

This study seeks to enhance comprehension of the relationship between weekly weather indices and crop production by examining the correlations between mustard yield and various meteorological factors. The analysis focused on investigating the connections between mustard output and maximum temperature (T_{max}), minimum temperature (T_{min}), total rainfall (RF), morning (RH I) and evening (RH II) relative humidity, sunshine hours (SSH), wind speed (WS), and evaporation (Evap.) at different phases of the mustard growth cycle. To accomplish this, the Statistical Package for Social Sciences (SPSS) software was utilized to develop a prediction model for mustard yield, employing long-term weather data as inputs. The objective of these models is to provide a reliable means of forecasting mustard crop yield based on historical weather patterns and conditions. For the creation of a multivariate model, weighted and unweighted weather indices were chosen (Table 1). To create weighted and unweighted weather indices, a total of 72 parameters were used.

Table 1. Interaction table for Weighted and Unweighted Weather indices for development of multivariate model

Variables	Unweighted Weather indices								Weighted Weather indices							
	T_{max}	T_{min}	RH I	RH II	RF	WS	SSH	Evap	T_{max}	T_{min}	RH I	RH II	RF	WS	SSH	Evap
T_{max}	Z10								Z11							
T_{min}	Z120	Z20							Z121	Z21						
RH I	Z130	Z230	Z30						Z131	Z231	Z31					
RH II	Z140	Z240	Z340	Z40					Z141	Z241	Z341	Z41				
RF	Z150	Z250	Z350	Z450	Z50				Z151	Z251	Z351	Z451	Z51			
WS	Z160	Z260	Z360	Z460	Z560	Z60			Z161	Z261	Z361	Z461	Z561	Z61		
SSH	Z170	Z270	Z370	Z470	Z570	Z670	Z70		Z171	Z271	Z371	Z471	Z571	Z671	Z71	
Evap.	Z180	Z280	Z380	Z480	Z580	Z680	Z780	Z80	Z181	Z281	Z381	Z481	Z581	Z681	Z781	Z81

Meteorological models were constructed for the mustard crop, utilizing weighted weather indices of crop growth. To create these weighted weather indices, various approaches were employed. Firstly, simple weather indices were generated by individually summing up each weather variable. Secondly, combinations of two weather variables were considered, and their product or interaction was calculated. The correlation between these weighted weather indices and the adjusted crop yield was then analyzed. The incorporation of weighted weather indices in the model allows for a more refined understanding of the relationship between meteorological factors and mustard crop yield, facilitating more accurate predictions and providing deeper insights into the effects of specific weather conditions on crop production.

The regression equation developed during the statistical model development approach is as follows;

$$Y=2.999+0.37*X+0.008*Z41$$

Where,

Y= Yield (t/ha)

X= Number of years

Z41= Relative Humidity II

The scatter plots 4 and 5 depicting the relationship between observed and predicted yield in mustard cultivation for calibration and validation for 21 years showed a positive correlation, indicating satisfactory agreement between the model predictions and actual observations. During the calibration stage, the SMLR model demonstrated

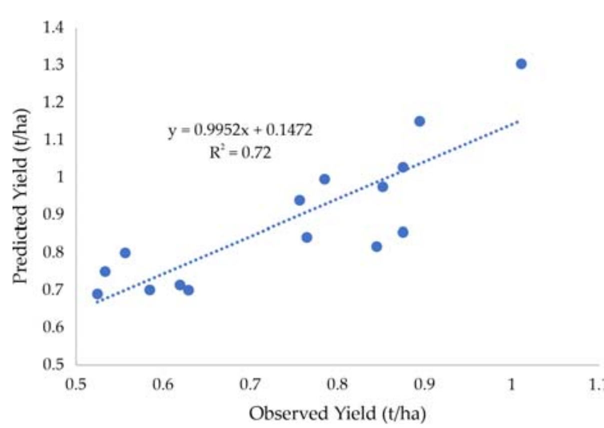


Fig. 4. Scatter plot showing the comparison between predicted and observed yield for calibration

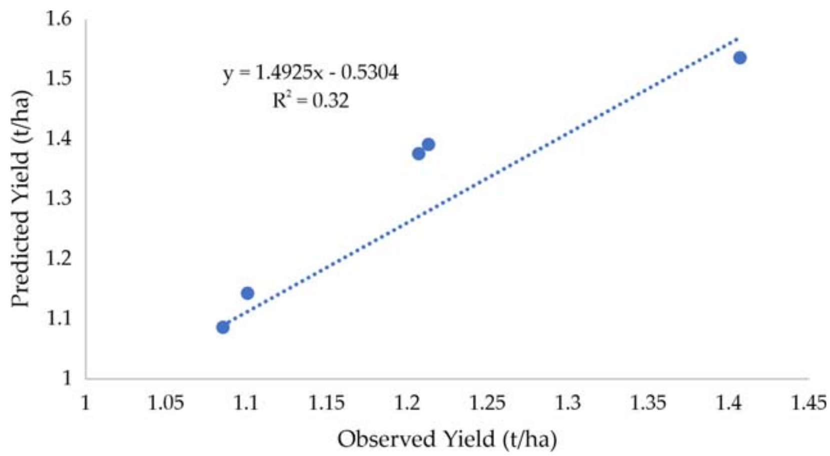


Fig. 5. Scatter plot showing the comparison between predicted and observed yield for validation

good performance with an R^2 value of 0.72 and nRMSE values below 30%, predicted within the range of 0.75 t ha^{-1} to 1.26 t ha^{-1}. However, at the validation stage, the model's performance declined, with an R^2 value of 0.32 and an nRMSE value of 0.32, indicating limited explanatory capability and diminished accuracy.

The findings reveal that the SMLR model exhibited promising predictive ability during the calibration phase, encompassing a significant proportion of yield variability and yielding accurate results. Nevertheless, its performance faltered during validation, raising concerns about its generalizability beyond the calibration dataset. To enhance the model's applicability for yield prediction at district levels, further investigations and refinement are necessary, considering the similarities in statistical approaches used by other studies for yield prediction.

Crop yield estimation using machine learning approach

In the context of mustard yield estimation in the Udham Singh Nagar district of Uttarakhand, artificial neural network (ANN) methodologies were employed for real-time predictions. The prediction model was constructed using crop yield data (t/ha) and meteorological data gathered at different phenological stages of mustard over a period of 21 years. Specifically, the model aimed to predict the yield of mustard.

To develop and validate the ANN-based approach, the entire dataset spanning 21 years (2001-2021) was divided into a calibration set comprising 70% of the data and a validation set comprising the remaining 30%. The calibration set, containing 16 years of yield data from 2001 to 2016, was utilized for training the ANN model. On the other hand, the validation set, encompassing four years of yield data from 2017 to 2021, was used to assess the performance and accuracy of the trained ANN model. This approach facilitated the real-time estimation of mustard yield, providing valuable insights into the factor influencing crop productivity in the study area. The use of ANN as a machine learning technique allowed for robust yield predictions, contributing to improved agricultural planning and decision-making processes.

In the training set, the relationship between prediction, predictors, and the dependent variable is established. The testing data was used to assess the models' accuracy. The size of the model represents the number of neurons in the hidden layer, and decay displays the decay rate of the gradient descent.

Figures 6 and 7 compare predicted and observed yields during calibration and validation.The coefficient of determination (R^2) values was calculated to assess the goodness of fit between the observed and predicted yield data. Notably, a high R^2 value of approximately 0.83 was achieved during the calibration, indicating a strong correlation between the observed and predicted yields. Similarly, in the validation phase, the R^2 value reached 0.72, further demonstrating a significant relationship between the observed and predicted yield data. These

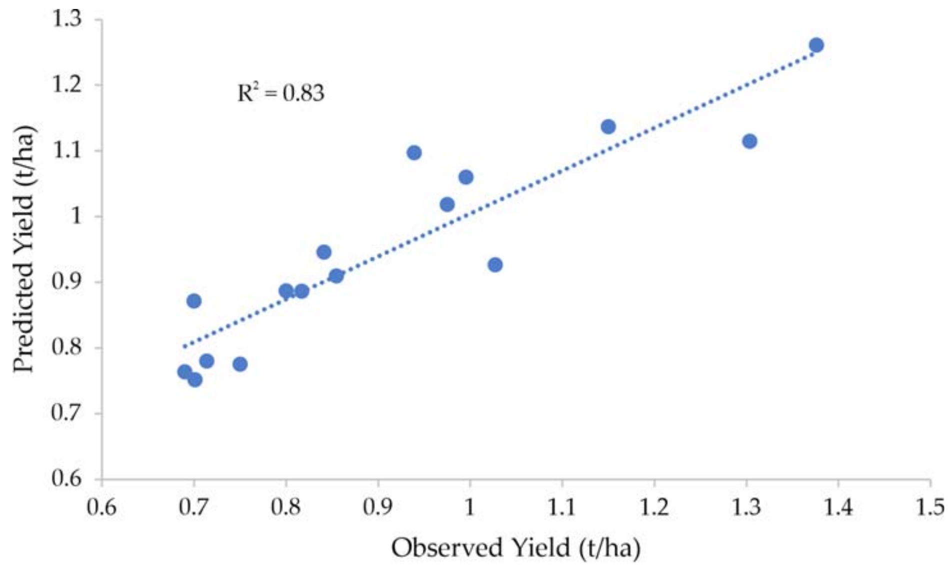


Fig. 6. Scatter plot showing the comparison between predicted and observed yield for calibration

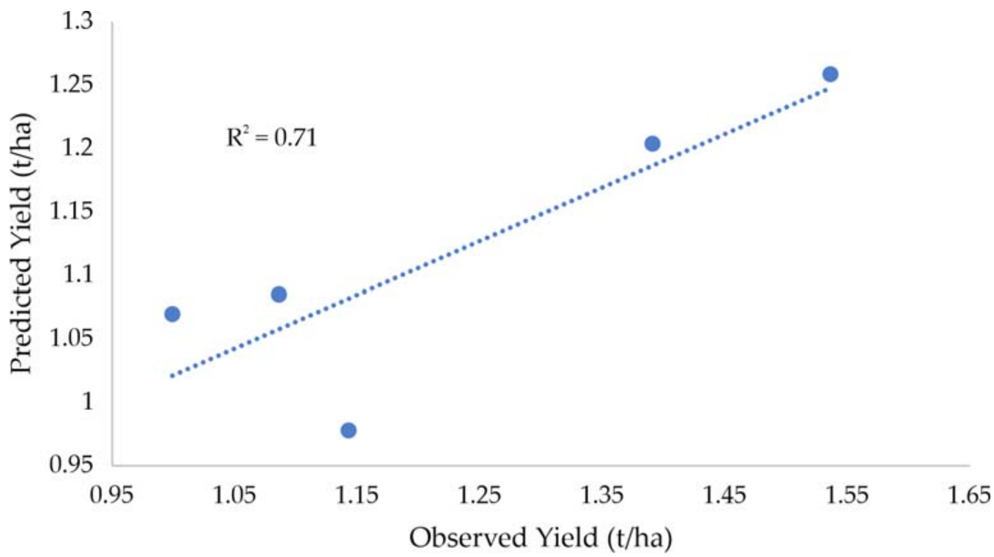


Fig. 7. Scatter plot showing the comparison between predicted and observed yield for validation

results validate the accuracy and reliability of the prediction model, highlighting its capability to effectively estimate mustard yield, both during the calibration and validation periods.

Importance of independent variables

This study used Karl Pearson's correlation coefficient to assess the impact of various indices on mustard yield, selecting the top 10 indices for analysis.Figure 8 highlights the findings, showing time as the most critical independent variable for the ANN model.The following time, the second most significant factor affecting mustard yield was the weighted index Z671, representing the product of sunshine hours and wind velocity. Among all the 10 indices considered, Z61 exhibited the least impact on mustard yield. These results provide valuable insights into the relative importance of different indices in relation to mustard crop productivity and offer useful information for agricultural decision-making and management practice.

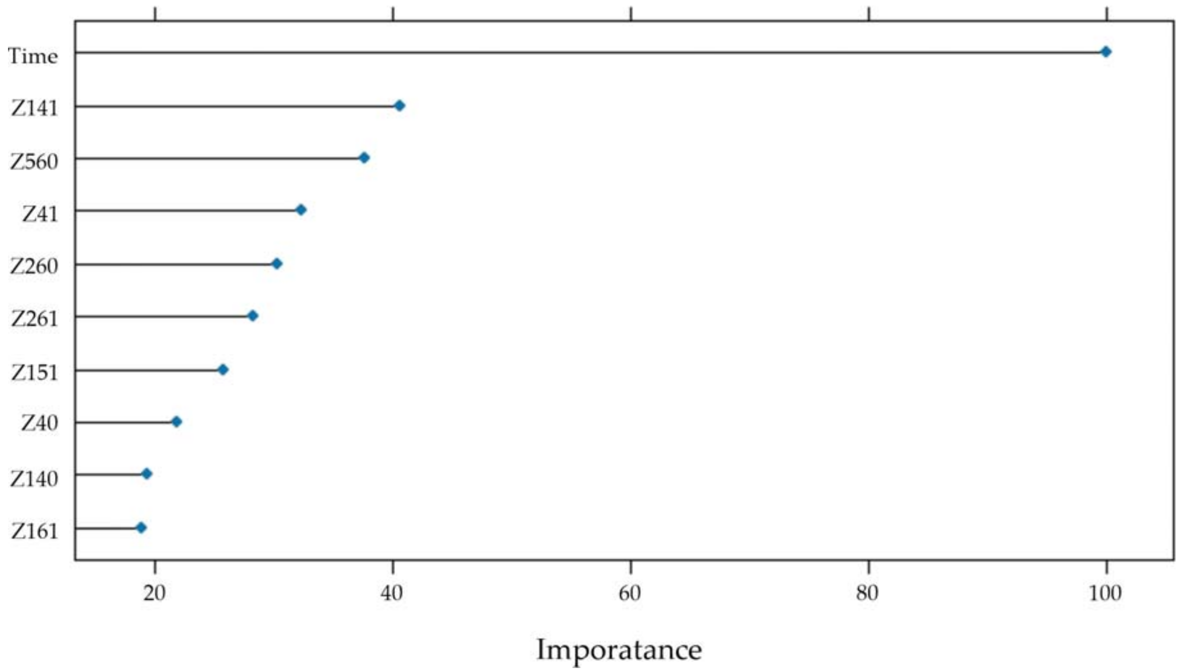


Fig. 8. List of some important independent variables used in ANN

To compare the performance of statistical and machine learning approaches

Mustard yield prediction models were developed using the ANN approach based on 21 years of data (2001–2021). These models were developed by utilizing both crop yield data and weather data from various phenological stages of crop growth. To evaluate the performance of the statistical models, key metrics such as the coefficient of determination (R^2), root mean square error (RMSE), and mean bias error (MBE) were employed. These metrics were computed for calibration and validation datasets to assess the models' accuracy and reliability in predicting mustard yield over the specified time frame.

The performance evaluation of two predictive models for mustard yield prediction in the US Nagar region is presented in Tables 2 and 3. The Artificial Neural Network (ANN) model achieved an R^2 value of 0.84 during calibration and 0.71 during validation, indicating a moderate correlation between predicted and observed yields. The RMSE and nRMSE values for the ANN model were relatively low, suggesting reasonable predictive accuracy, with the model

Table 2. Performance of the model developed using ANN technique for mustard yield prediction of US Nagar

During calibration				During validation			
R^2	RMSE (t ha^{-1})	nRMSE (t ha^{-1})	MBE	R^2	RMSE (t ha^{-1})	nRMSE (t ha^{-1})	MBE
0.84	0.011	0.0014	0.035	0.71	0.140	0.316	-0.11

Table 3. Performance of the model developed using multiple regression technique

During calibration				During validation			
R^2	RMSE (t ha^{-1})	nRMSE (t ha^{-1})	MBE	R^2	RMSE (t ha^{-1})	nRMSE (t ha^{-1})	MBE
0.72	0.170	0.277	0.150	0.30	0.120	0.32	-0.11

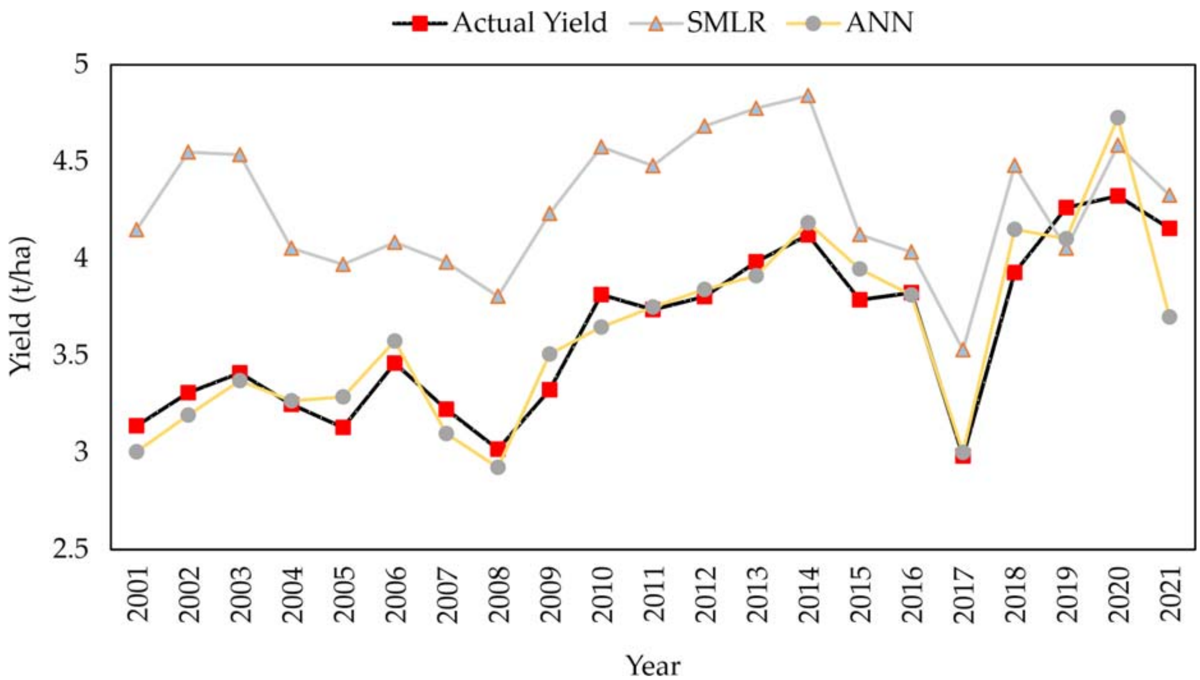


Fig. 9. Single-line graph showing the actual and predicted yield by MLR and ANN

exhibiting a slight overestimation during calibration (MBE = 0.035 t ha^{-1}) and a slight underestimation during validation (MBE = -0.11 t ha^{-1}). On the other hand, the multiple regression model showed an R^2 value of 0.72 during calibration and 0.30 during validation, suggesting a weaker correlation. The RMSE and nRMSE values were higher for the multiple regression model compared to the ANN model, indicating larger prediction errors, while the MBE values were close to zero during both calibration and validation, indicating a negligible bias in the predictions.

The results from Table 2 and Table 3 indicate that the ANN model outperformed the multiple regression technique in mustard yield prediction for the US Nagar region. The ANN model demonstrated better predictive accuracy with higher R^2 values, indicating a stronger correlation between predicted and observed yields Fig 9. Additionally, the ANN model exhibited lower RMSE and nRMSE values, implying more precise predictions with smaller errors compared to the multiple regression model (Satpathi *et al*., 2025). The slight overestimation during calibration and slight underestimation during validation by the ANN model can be further improved through model optimization. Conversely, the multiple regression model showed a weaker correlation and higher prediction errors, suggesting limitations in capturing the complex relationships influencing mustard yield. Overall, the effectiveness of the ANN technique in accurately forecasting mustard yield has significant implications for informed decision-making and improved agricultural practices in the *Tarai* region of US Nagar. Similarly, Aravind *et al*. (2022) highlighted the effectiveness of artificial neural networks (ANNs) in forecasting wheat yield in Patiala district, demonstrating their superior predictive accuracy when compared to traditional approaches such as multiple linear regression (MLR), LASSO, and ELNET. And Uno *et al*. (2005) observed that ANN-based yield models outperformed conventional models, particularly during the validation phase, showcasing their enhanced reliability.

Conclusion

The comparison of statistical and machine learning (ANN) approaches for yield prediction revealed that the machine learning approach achieved a coefficient of determination (R^2) of 84% between observed and predicted yield. During calibration, the root mean square error (RMSE), normalized root mean square error (nRMSE), and mean bias error (MBE) were 0.011, 0.0014, and 0.035, respectively. For validation, the corresponding values were 0.71, 0.140, 0.316, and -0.112, respectively.

For the statistical approach (SMLR), the coefficient of determination (R^2) between observed and predicted yield was 0.72. During calibration, the root mean square error (RMSE), normalized root mean square error (nRMSE), and mean bias error (MBE) were 0.170, 0.277, and 0.150, respectively. For validation, the corresponding values were 0.30,0.120, 0.32, and -0.11, respectively. Two distinct methodologies, namely SMLR and ANN, were applied to investigate the correlation between yield and weather parameters in Udham Singh Nagar district, Uttarakhand, India. The comparison of results obtained from both models revealed that the ANN model exhibited the highest correlation in prediction. Furthermore, the artificial neural network demonstrated the advantageous ability to work independently and estimate errors, enhancing its utility as a reliable prediction tool. In conclusion, the ANN method outperformed SMLR for predicting mustard yield in Udham Singh Nagar district, Uttarakhand, India.

Future research should delve into deep learning techniques such as CNN, DNN, and RNN, either individually or synergistically, to harness their proven potential in revolutionizing crop yield prediction. By incorporating advanced datasets like high-resolution satellite imagery and remote sensing data, researchers can gain a more nuanced understanding of the spatial and temporal dynamics that shape crop productivity. These innovative approaches promise to refine prediction accuracy, optimize resource allocation, and address the complex challenges posed by climate variability in agriculture.

References

Aditya, K. and Das, S. 2012. Crop yield forecasting using discriminant function analysis. *LAP Lambert Academic Publishing*.

Amanabadi, S., Vazirinia, M., Vereecken, H., Vakilian, K.A. and Mohammadi, M.H. 2019. Comparative study of statistical, numerical and machine learning-based pedotransfer functions of water retention curve with particle size distribution data. *Eurasian Soil Science* **52**(12): 1555-1571.

Aravind, K.S., Vashisth, A., Krishanan, P. and Das, B. 2022. Wheat yield prediction based on weather parameters using multiple linear, neural network and penalised regression models. *Journal of Agrometeorology* **24**: 18–25. https://doi.org/10.54386/jam.v24i1. 1002.

Bhattacharya, Priya, Maity, P.P., Ray, Mrinmoy and Krishnan, P. 2018. Comparison of artificial neural network and multi-linear regression for prediction of field capacity soil moisture content. *Journal of Agricultural Physics* **18**(2): 173-180.

Bhattacharya, P., Maity, P.P., Ray, M. and Mridha, N. 2021. Prediction of mean weight diameter of soil using machine learning approaches. *Agronomy Journal* **113**(2): 1303-1316.

Dai, X., Huo, Z. and Wang, H. 2011. Simulation for response of crop yield to soil moisture and salinity with artificial neural network. *Field Crops Research* **121**(3): 441-449.

Das, B., Nair, B., Arunachalam, V., Reddy, K.V., Venkatesh, P., Chakraborty, D. and Desai, S. 2020. Comparative evaluation of linear and nonlinear weather-based models for coconut yield prediction in the west coast of India. *International Journal of Biometeorology* **64**(7): 1-13.

Das, B., Nair, B., Reddy, V.K. and Venkatesh, P. 2018. Evaluation of multiple linear, neural network and penalised regression models for prediction of rice yield based on weather parameters for west coast of India. *International Journal of Biometeorology* **62**(10): 1809-1822.

Das, B., Nair, B., Arunachalam, V., Reddy, K.V., Venkatesh, P., Chakraborty, D. and Desai, S. 2020. Comparative evaluation of linear and nonlinear weather-based models for coconut yield prediction in the West Coast of India. *International Journal of Biometeorology* **64**: 1111–1123.

Ghosh, K., Balasubramanian, R., Bandopadhyay, S., Chattopadhyay, N., Singh, K.K. and Rathore, L.S. 2014. Development of crop yield forecast models under FASAL-a case study of kharif rice in West Bengal. *Journal of Agrometeorology* **16**: 1–8.

Gocic, M., Motamedi, S., Shamshirband, S., Petkoviæ, D., Ch, S., Hashim, R. and Arif, M. 2015. Soft computing approaches for forecasting reference evapotranspiration. *Computers and Electronics in Agriculture* **113**: 164-173.

Kaul, M., Hill, R.L. and Walthall, C. 2005. Artificial neural networks for corn and soybean yield prediction. *Agricultural Systems* **85**(1): 1-18.

Kaur, R. and Sharma, S. 2019. An ANN Based Approach for Software Fault Prediction Using Object Oriented Metrics. In: Luhach, A.K., Singh, D., Hsiung, P.-.-A., Hawari, K.B.G.,Lingras, P., Singh, P.K. (Eds.), Advanced Informatics for Computing Research. ICAICR 2018. Communications in Computer and Information Science, Vol 955. Springer Singapore, pp. 341–354.

Khairunniza-Bejo, S., Mustaffha, S. and Ismail, W.I.W. 2014. Application of artificial neural network in predicting crop yield: A review. *Journal of Food Science and Engineering* **4**(1): 1.

Sarkar, A., Maity, P.P., Ray, M., Chakraborty, D., Das, B. and Bhatia, A. 2023. Inclusion of fractal dimension in four machine learning algorithms improves the prediction accuracy of mean weight diameter of soil. *Ecological Informatics* **74**: 101959.

Satpathi, A., Chand, N., Setiya, P., Ranjan, R., Nain, A. S., Vishwakarma, D.K. and Kisi, O. 2025. Evaluating statistical and machine learning techniques for sugarcane yield forecasting in the tarai region of North India. *Computers and Electronics in Agriculture* **229**: 109667.

Setiya, P., Satpathi, A., Nain, A.S. and Das, B. 2022. Comparison of weather-based wheat yield forecasting models for different districts of Uttarakhand using statistical and machine learning techniques. *Journal of Agrometeorology* **24**(3): 255-261.

Singh, R.S., Patel, C., Yadav, M.K. and Singh, K.K. 2014. Yield forecasting of rice and wheat crops for eastern Uttar Pradesh. *Journal of Agrometeorology* **16**: 199–202.

Uno, Y., Prasher, S.O., Lacroix, R., Goel, P.K., Karimi, Y., Viau, A. and Patel, R.M. 2005. Artificial neural networks to predict corn yield from Compact Airborne Spectrographic Imager data. *Computers and Electronics in Agriculture* **47**: 149–161.

Wang, S.-C. and Wang, S.-C. 2003. Artificial neural network. *Interdisciplinary Computing in Java Programming* 81–100.

Received: 4 September 2024; Accepted: 30 November 2024