International Journal of Environment and Climate Change

Volume 13, Issue 10, Page 2425-2435, 2023; Article no.IJECC.104851
ISSN: 2581-8627
(Past name: British Journal of Environment & Climate Change, Past ISSN: 2231–4784)

Weather-Based Rice Crop Yield Forecasting using Different Regression Techniques & Neural Network Approach for Prayagraj Region

Nilesh Kumar Singh [a*], Shraddha Rawat [a] and Shweta Gautam [a]

[a] *Department of Environmental Science and Natural Resource Management, Sam Higginbottom University of Agriculture Technology and Sciences, Allahabad, Uttar Pradesh, India.*

Authors' contributions

This work was carried out in collaboration among all authors. All authors read and approved the final manuscript.

Article Information

DOI:10.9734/IJECC/2023/v13i102908

Open Peer Review History:
This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: https://www.sdiarticle5.com/review-history/104851

Original Research Article

Received: 20/06/2023
Accepted: 26/08/2023
Published: 06/09/2023

ABSTRACT

Rice crop yield data and weather data were considered in this study, covering the past twenty-nine years (1991-2019) in Prayagraj District, Uttar Pradesh. The data was sourced from DACNET and the College of Forestry, SHUATS Prayagraj. The analysis comprised a calibration period of 26 years (90% of the dataset) and a validation period using the remaining data (10%). In this study, 75.9% of the data were utilized for training the Artificial Neural Network (ANN) model, while the remaining 24.1% were allocated for testing and validation, ensuring comprehensive model assessment. The primary evaluation metric employed for model efficiency was the Normalized Root Mean Squared Error (nRMSE), with a focus on achieving the lowest values. Both a Stepwise Linear Regression technique and a Neural Network were employed for rice yield prediction. Notably, the regression-

---

**Corresponding author: E-mail: singhnilesh7007@gmail.com;*

based model exhibited superior performance compared to the ANN model, as indicated by the nRMSE values. This conclusion was drawn from the observation that the regression-based model yielded the best-fitting results. The study's findings highlight the significance of Bright Sunshine Hours in relation to nRMSE and the coefficient of determination, which were recorded at 0.00025 and 0.94, respectively. This underlines the importance of this meteorological factor in accurately predicting rice crop yield.

Keywords: Regression; yield; model; parameter; artificial neural networks; coefficient of determination.

1. INTRODUCTION

Agriculture has consistently stood as one of humanity's crucial pursuits, contributing significantly to livelihoods and employment. However, the escalating population has led to worsening nutritional conditions among the impoverished, necessitating improvements. The ramifications of population growth are starkly evident in the environment, where the accelerating damage poses a hindrance to agricultural production. The prediction of crop yield holds paramount importance in this scenario. Each farmer strives to estimate the potential yield from their fields, a task that has historically relied on analyzing past crop results. Crop yield hinges on various factors, including weather conditions, pest impact, and harvest planning. Precise historical crop yield information becomes a pivotal factor for informed decisions pertaining to agricultural risk management.

The proposed approach employs Regression and Neural Network techniques for rice yield prediction. Accurate crop yield prediction plays a pivotal role in addressing contemporary challenges in food security, especially amidst the backdrop of global climate change. Beyond aiding farmers in informed economic and managerial choices, accurate yield predictions contribute to efforts in averting famines.

Numerous yield prediction models have been developed using a variety of weather variables and crop data, aiming to establish regression and artificial neural network models. Crop yield prediction in agriculture heavily relies on regression-based models. Multiple linear regression (MLR) endeavors to establish a model that captures the connection between the regressed variable and multiple regressors [1]. These models have gained widespread usage in assessing the correlation between weather parameters and crop yield across different regions" [3,4,5,6] Bankara *et al.,*[2] compared "the technique of Multiple Linear Regression (MLR)". [3] also developed "crop yield forecast model by employing stepwise linear regression technique and found that temperature (maximum & minimum) and relative humidity were significant predictors in crop yield forecast". Parallel endeavors were conducted by (Kalubarme and Ahuja, [7]. Chauhan *et al.,* [8] to "develop agrometeorological data based rice yield prediction model for Karnal, central Punjab and Bulsar district of Gujarat respectively".[9] ventured into the creation of rice forecast models for various districts in West Bengal, extending the scope of such predictive endeavors. The simplest technique to generate a yield forecast based on a dataset of yield and weather parameters is stepwise multiple linear regression (SMLR) [10].

This research paper introduces the utilization of regression models and the Artificial Neural Network (ANN) technique for predicting rice crop yields based on weather and crop parameters. The primary objective is to propose an optimal yield prediction method, determined by evaluating the normalized Root Mean Square Error (nRMSE).

1.1 Study Area

The study was carried out in the Prayagraj district of Uttar Pradesh, situated in the south-eastern part of the state. Prayagraj is located between the parallels of 24º 47' north latitude and 81 º 19' east longitudes. It shares its eastern border with Sant Ravi Das Nagar district (Varanasi).

2. MATERIALS AND METHODS

Data Collected for the period 1991 – 2019 from the Directorate of Economics and Statistics, Department of Agriculture, Cooperation, and Farmers Welfare, Ministry of Agriculture and Farmers Welfare, were yearly production (in kg) and area (in ha) data under the Rice yield in Prayagraj District. Weekly and monthly data, encompassing maximum and minimum temperatures (ºC), relative humidity (%), rainfall (mm), and the number of rainy days, were gathered from the agrometeorological observatory located at Sam Higginbottom University of Agriculture, Technology, and Sciences. This observatory is positioned within the Department of Environmental Science and

Natural Resource Management at the College of Forestry in Prayagraj.

The present study employed both artificial neural network (ANN) and linear regression techniques for rice yield forecasting. Utilizing the SPSS software, the research conducted statistical analyses and developed a multiple regression model for predicting rice yield.

2.1 Development of Statistical Regression Model using Weekly Weather Parameter

The overall analysis involved utilizing a dataset spanning 26 years, allocated with 90% for calibration and the remaining portion for validation. Within this methodology, the rice yield prediction model was formulated using weekly weather parameters corresponding to the crop's growth season, specifically from the 22nd to the 37th week. Within the study, weekly weather parameters were treated as independent variables, while rice yield served as the dependent variable in the SPSS software. To develop the multivariate statistical models, the stepwise regression technique within SPSS was employed.

2.2 Development of Statistical Regression Model using Weighted and Un-weighted Parameter

In this context, the dataset for rice crop yield comprises a total of 29 years. Out of these, 26 years' worth of records have been earmarked for both training and testing purposes. Weighted and un-weighted parameters are regarded as independent variables, while rice yield is treated as the dependent variable within the SPSS software. The development of multivariate statistical models utilized the stepwise regression technique in SPSS.

Simple and weighted weather indices have been formulated specifically for the Prayagraj District. For generating simple weather indices, the summation of individual weather variables or the interaction of two weather variables at a time was employed. Conversely, weighted weather indices were derived from the sum product of individual weather variables or interactions of weather variables, all while considering their correlation with the adjusted crop yield.

The computation of both simple and weighted weather indices was carried out based on the formula [11] mentioned below (Table 1).

Table 1. Simple and weighted weather indices

Simple weather indices	Weighted weather indices
$Zij = \sum_{w=1}^{m} Xiw$	$Zij = \sum_{w=1}^{m} r^j iwXiw$
$Zij = \sum_{w=1}^{m} Xiw$	$Zii'j = \sum_{w=1}^{m} r^j ii'wXiwXi'w$
$Zii'j = \sum_{w=1}^{m} Xiw\,Xi'w$	

Where,

$Xiw/Xii'w$= value of i'^{th}/i'^{th} weather variable under study in weather week,

$r^j iw/r^j\, ii'w$ = correlation coefficient of yield with i^{th} weather variable or product of i'^{th}/i'^{th} weather variable in the week.

2.3 Development of Statistical Regression Model using Bright Sunshine Hour with Weighted & Un-weighted Parameter

In this study, the rice yield prediction model was formulated utilizing the Bright Sunshine Hour during the crop growing season, specifically from the 22nd to the 37th week. Both weighted and un-weighted indices were incorporated as independent variables, while rice yield served as the dependent variable within the SPSS software. The statistical modeling process involved utilizing the stepwise regression method available in SPSS.

2.4 Development of Statistical Regression Model using Actual Evapotranspiration with Weighted & Un-weighted Parameter

In this technique the regression equation for rice yield model was developed using Actual Evapotranspiration and weighted and un-weighted parameters were considered of SMW 22nd to 37th week. A multivariate model was then formed for the rice yield prediction using statistical regression technique in SPSS.

2.5 Rice Yield forecasting using Machine Learning Approach (ANN)

The Multilayer Perceptron (MLP) technique stands out as one of the prominent types among neural networks and being used in the study. This network is conceptualized as an input-output model, featuring weights and thresholds (biases) as the model's adjustable parameters. The configuration of hidden layers and the number of neurons within each hidden layer are often fine-tuned to optimize the final model's performance [12]. The determination of the number of nodes in each layer typically involves a trial-and-error process. Training the MLP involves utilizing a training set comprised of input data and corresponding known output data. The MLP facilitates the transformation of m inputs into n outputs through nonlinear functions. With the activation of units in the output layer, the MLP network's output is calculated as follows:

$$X_o = f(\sum X_h w_{ho} + b_j)$$

In this equation, (*f*) symbolizes the activation function, X_h represents the activation of the hidden layer node; "w_{ho}" designates the interconnection between hidden and output layer nodes, and bj indicates the bias. The study employs the back propagation algorithm utilized by artificial neural networks (ANN) [13], which is widely acknowledged as a prominent approach for training the MLP network. This recognition is substantiated by citations such as Wasserman, Fausett, and Haykin [14,15,16]. Different ANN model structures were analyzed, including variations in the number of neurons in the hidden layer [17].

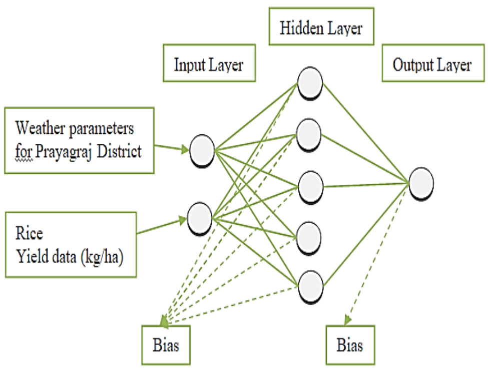

Fig.1. Diverse layers and varying network structural depiction of the ANN

The network includes an input layer of non-linear elements, known as neurons (also called nodes), which receive predictor values. Subsequent layers of neurons receive input from previous layers. Outputs from nodes in each layer serve as inputs for nodes in the following layer. The final layer is the output layer, while layers between input and output are hidden layers. Each hidden layer's nodes process input through an activation or transfer function, generating transformed output for the subsequent layer. MLP architecture allows variable hidden layers and neurons. A feed forward network is fully connected, with layers linking every neuron to those in the next. It facilitates one-way flow without cycles. Fig. 1 illustrates a fully connected MLP featuring a single hidden layer [18].

2.5.1 Training ANN model

In this investigation, 76% of the data were allocated for training the ANN model, while the remaining 24% were utilized for testing. The entire dataset (100%) was employed for validation. As shown in (Table 3), the ANN model employing Hyperbolic Tangent as the transfer function yielded the most favorable outcomes during both the training and testing phases. The optimization of the artificial neural network structure was performed by obtaining the minimum validation error. The selection of the neural network model was also guided by the size of the training and test set errors and other important quality parameters [17]. This was evident from the attainment of the lowest Root Mean Square Error (RMSE) and normalized Root Mean Square Error (nRMSE) values.

2.5.2 Model performance

For testing the performance of developed statistical forecasting models, R^2, Root Mean Square Error (RMSE) and Normalized Root Mean Square Error (nRMSE) were calculated using the following formula:

$$R^2 = \left(\frac{\frac{1}{n}\sum_{i=1}^{n}(M_i - \bar{M})\,(O_i - \bar{O})}{\sigma M \sigma O} \right)$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(O_i - M_i)^2}$$

$$\text{nRMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(O_i - M_i)} \times \frac{100}{\bar{O}}$$

Where,

RMSE stands for Root Mean Square Error, while nRMSE represents Normalized Root Mean Square Error. In these formulas, P_i denotes the predicted value, O_i signifies the observed value, n represents the number of observations, and M stands for the mean of the observed value.

Table 2. Simple and weighted weather indices used for developing model

Weather Parameter	Simple weather indices					Weighted Weather indices				
	Tmax	**Tmin**	**R/F**	**RH(I)**	**RH(II)**	**Tmax**	**Tmin**	**R/F**	**RH(I)**	**RH(II)**
Tmax	Z10					Z11				
Tmin	Z120	Z20				Z121	Z21			
R/F	Z130	Z230	Z30			Z131	Z231	Z31		
RH (I)	Z140	Z240	Z340	Z40		Z141	Z241	Z341	Z41	
RH (II)	Z150	Z250	Z350	Z450	Z50	Z151	Z251	Z351	Z451	Z51

**Simple and weighted weather indices employed for models development are presented in Table 4.*

Table 3. The network Information

Input Layer	Number of Units[a]	30
	Rescaling Method for Covariates	Standardized
Hidden Layer(s)	Number of Hidden Layer	1
	Number of Units in Hidden Layer 1[a]	5
	Activation Function	Hyperbolic Tangent
Output Layer	Dependent Variables	Yield
	Number of Units	1
	Rescaling Method for Scale Dependents	Standardized
	Activation Function	Identity
	Error Function	Sum of Squares

a. Excluding the bias unit

The model exhibits exceptional performance when the nRMSE value exceeds 10%, demonstrates effectiveness with an nRMSE value falling between 10-20%, and shows satisfactory results when the nRMSE value falls within the 20-30% range.

3. RESULTS

3.1 Rice Yield Prediction Using Statistical Modeling Based on Different Weather Parameters

This approach involved creating the rice yield prediction model by utilizing diverse weather parameters during the crop's growth season, specifically from the 22nd to the 37th week (SMW). Within this study, a range of weather parameters were regarded as independent variables, while rice yield was treated as the dependent variable in the SPSS software. The development of multivariate statistical models occurred through the utilization of the stepwise regression technique within SPSS.

3.2 Development of Statistical Regression Model using Weekly Weather Parameter

In this study, a single model, namely Model 1, was developed. This model incorporated time as an independent factor to estimate rice yield for the Prayagraj region, achieving an R^2 value of approximately 0.55. This R^2 value signifies that the model can account for about 55% of the variation in yield. Consequently, Model 1 was utilized to predict the rice yield for Prayagraj, as evidenced in (Table 4). However, it's noteworthy that for the years 2018 and 2019, the model underestimated the yield of the rice crop, as depicted in (Table 5). Conversely, for the year 2017, the model's predicted rice yield closely aligned with the actual yield, indicating its favorable performance for that year. The results presented in (Table 5) also reveal a percent deviation ranging from 7.05% to 15.99%, further affirming the model's strong fit.

The aforementioned findings can be substantiated by a study on yield forecasting for Rice and Wheat in Central Uttar Pradesh using a statistical analysis [19]. In this research, a dataset spanning twenty-three years (1992-2015) was utilized, encompassing weather variables like rainfall (mm), maximum and minimum temperatures (°C), as well as maximum and minimum relative humidity (%). Yield data for rice and wheat crops across twelve districts were employed to predict yields using statistical methods as part of the FASAL Project, led by the Department of Agronomy at Chandra Shekhar Azad University of Agriculture & Technology in Kanpur, Uttar Pradesh.

For this endeavor, a regression equation was established through statistical analysis facilitated by the SPSS software package. The generated models underwent validation using data from two years, specifically 2016 and 2017. The results indicated that the models managed to account for a variation in rice crop yield ranging from 45% to 73% and a variation in wheat crop yield from 49% to 74% across different districts.

3.3 Development of Statistical Regression Model using Weighted and Un-weighted Parameter

In this scenario, only two models were developed. Model 1 utilized time as an independent variable to estimate rice yield for the Prayagraj region, achieving an R^2 value of approximately 0.55. Conversely, Model 2 incorporated both time and Z241 as independent variables, collectively explaining a 75% variation in the yield. Consequently, Model 2 was employed to project the forecasted rice yield for Prayagraj, as demonstrated in (Table 4). Notably, for the years 2018 and 2019, the models underestimated the rice crop yield, as evidenced in (Table 5). On the other hand, the model's prediction closely aligned with the actual yield for the year 2017, suggesting a favorable performance during that year.

In support of this work, a study conducted [20] for yield forecasting models based on both simple weather variables and composite weather variables (weighted and un-weighted) for the period of 1992-2018. The study derived various multiple linear regression equations to forecast the yields of rapeseed and mustard crops.

3.4 Development of Statistical Regression Model Using Bright Sunshine Hour with Weighted & Un-weighted Parameter

A total of four multivariate statistical models for rice yield were constructed using the stepwise regression technique in SPSS. The most optimal model among these four was employed to predict the forecasted rice yield for Prayagraj. In this particular instance, Model Four utilized Z251, Z131, Z250, and time as independent factors to estimate the rice yield for the Prayagraj region, yielding an impressive R^2 value of approximately 0.94. As observed in (Table 4), Model Four emerged as the most effective choice, accounting for 94% of the variability through the parameters incorporated in the equation.

Notably, for the year 2019 (as presented in Table 5), Model 4 overestimated the rice yield, while for the year 2017, it underestimated the yield of the rice crop. However, in the case of the year 2018, Model 4 closely approximated the actual rice yield, implying its strong performance for that year.

A similar line of research was undertaken [21] focusing on forecasting district-level rice yields in Chhattisgarh during the mid-season. In their study, multivariate models were developed based on BSS for the years 2014 and 2015. The resulting forecasted yields were then compared with the actual yields of the corresponding years to assess the accuracy of the developed models.

3.5 Development of Statistical Regression Model Using Actual Evapotranspira-tion with Weighted & Un-weighted Parameter

This study developed four distinct yield prediction models for the rice crop. Model 4, encompassing variables such as minimum temperature and maximum relative humidity, maximum temperature and minimum temperature, as well as minimum temperature and maximum relative humidity, alongside time, achieved an R2 value of 0.87. All four models were utilized to predict the forecasted rice yield for Prayagraj. As evident in (Table 4), Model 4 emerged as the most appropriate choice, explaining 87% of the variability. Notably, for the year 2019, Model 4 overestimated the rice yield, while for the year 2017, it underestimated the yield of the rice crop. However, for the year 2018, Model 4 closely approximated the actual rice yield, indicating its favorable performance, as depicted in (Table 5). This model was subsequently employed to project the forecasted rice yield for Prayagraj.

Building upon the results obtained by [22], a predictive model was generated for finger millet yield based on time in the Banglore area. A multi-linear statistical equation was developed using Growing Degree Days (GDD), actual Evapotranspiration, and BSS for the period spanning 1988 to 2005. This identical model was then employed to forecast the yield of finger millet for the Kharif season of 2005.

3.6 Rice Yield Forecasting using Machine Learning Approach (ANN)

The model summary encompasses both the Training and Testing phases, with the sum of squared errors being 3.462 and 1.641, respectively, as detailed in Table 6. From Fig. 2, it becomes evident that the coefficient of determination stood at 0.66, representing the relationship between the yield predicted by the

Table 4. Rice yield forecast equation using stepwise regression method based on different parameters

Parameter	Model	Yield forecast equation	R^2	Std. Error
Bright Sunshine Hour	1	Y=3050.840+1.680(Z251)	0.406	260.872
	2	Y=2952.795+2.193(Z251)+51.442(TIME)	0.651	212.258
	3	Y=5520.653+2.084(Z251)+90.284(TIME)+2.942(Z131)	0.866	140.41
	4	Y=4218.547+2.711(Z251)+49.740(TIME)+3.797(Z131)+0.079(Z250)	0.946	96.51
Actual Evapotranspiration	1	Y= 1510.268+43.649*(TIME)	0.673	263.577
	2	Y=1944.048+51.035*(TIME)+1.40*(Z251)	0.782	219.598
	3	Y=3446.649+60.969*(TIME)+1.327*(Z251)+1.811*(Z131)	0.844	189.546
	4	Y=2260.633+55.417*(TIME)+1.796*(Z251)+1.783*(Z131)+0.043*(Z250)	0.875	173.237
Weighted and Unweighted	1	Y= 1588.077+35.641*(TIME)	0.556	248.822
	2	Y=2049.774+43.823(TIME)+1.499(Z241)	0.756	188.501
Weekly Weather	1	Y= 1588.077+35.641*(Time)	0.556	248.82

Table5. Validation of rice yield forecast models & error analysis

Methods		2017		2018		2019	
		Predicted Yield	Error %	Predicted Yield	Error %	Predicted Yield	Error %
Bright Sunshine Hour	Model 1	2315	15.6	2113	13.75	2438	13.74
	Model 2	2608.60	4.93	2397.43	23.15	2873.21	7.939
	Model 3	2521.54	8.10	2951.41	5.403	3307.75	-5.98
	Model 4	2396.01	12.6	3050.16	2.238	3347.20	-7.24
Actual Evapotranspiration	Model 1	2823.79	-2.9	2872.44	7.934	2921.08	6.405
	Model 2	2708.40	1.29	2591.85	16.92	2913.67	6.642
	Model 3	2560.40	6.69	2825.12	9.451	3058.41	2.005
	Model 4	2605.20	5.05	2894.44	7.229	3159.65	-1.23
Weighted and Unweighted	Model 1	2550.38	7.05	2586.02	17.11	2621.66	15.99
	Model 2	2576.01	6.12	2440.40	21.78	2774.16	11.11
Weekly Weather	Model 1	2550.40	7.05	2586.04	17.11	2621.68	15.99
Artificial Neural Network	Model 1	2533.87	7.65	2861.5	8.285	2954.19	5.344

Table 6. Model summary of neural network model for training and testing for single network

Training	Sum of Squares Error	3.462
	Relative Error	0.346
	Stopping Rule Used	1 consecutive step with no decrease in error
	Training Time	0:00:00.02
Testing	Sum of Square Error	1.641
	Relative Error	1.279

ANN and the observed yield. The rice yield model, developed through the ANN approach, elucidates that 66% of the variability in rice yield can be attributed to weather parameters. Notably, for the year 2019, the model underestimated the rice yield, while for the year 2017, the prediction closely aligned with the actual yield. This suggests that the ANN model formulated using Multilayer Perceptron (MLP) techniques demonstrated robust performance for the year 2017.

During the testing phase, the percent deviation ranged from 5.34% to 8.28% (as shown in Table 5), falling within an acceptable range. This signifies that the model accurately predicted the rice crop yield, with the highest deviation being around 8%.

In a research investigation titled 'Forecasting Sesame Seed Yield through Artificial Neural Networks,'[23] two techniques, namely artificial neural networks (ANN) and multiple regression models (MLR), were utilized to gauge the sesame seed yield (SYS) based on easily quantifiable plant characteristics. Field data were employed to evaluate the effectiveness of both ANN and MLR. The findings indicated that the ANN performed exceptionally well, providing precise predictions with a root-mean-square error (RMSE) of 0.339 t/ha and a coefficient of determination (R2) of 0.901.

Fig 3 illustrates the relative influence of predictors, which includes both weighted and un-weighted weather variables, on the rice yield prediction obtained through ANN. The vertical axis depicts the top ten input variables (weighted and un-weighted indices), with the variable making the most substantial contribution (largest sum of square derivative, SSD value) positioned at the top. Other variables are placed below in descending order of their contribution to predicting a given variable. The horizontal axis's scale varies according to the different mean importance levels. The variable with the highest SSD value, signifying the greatest contribution, is assigned a value of 100%. Remaining variables are assigned values (in %) corresponding to their SSD values relative to the SSD value of the most influential variable.

The predictors significantly contribute to the prediction of rice yield (as depicted in Fig. 3). The findings highlight that the predictor (indices), followed by the morning relative humidity (Z41), ranks as the most vital determinants for yield prediction.

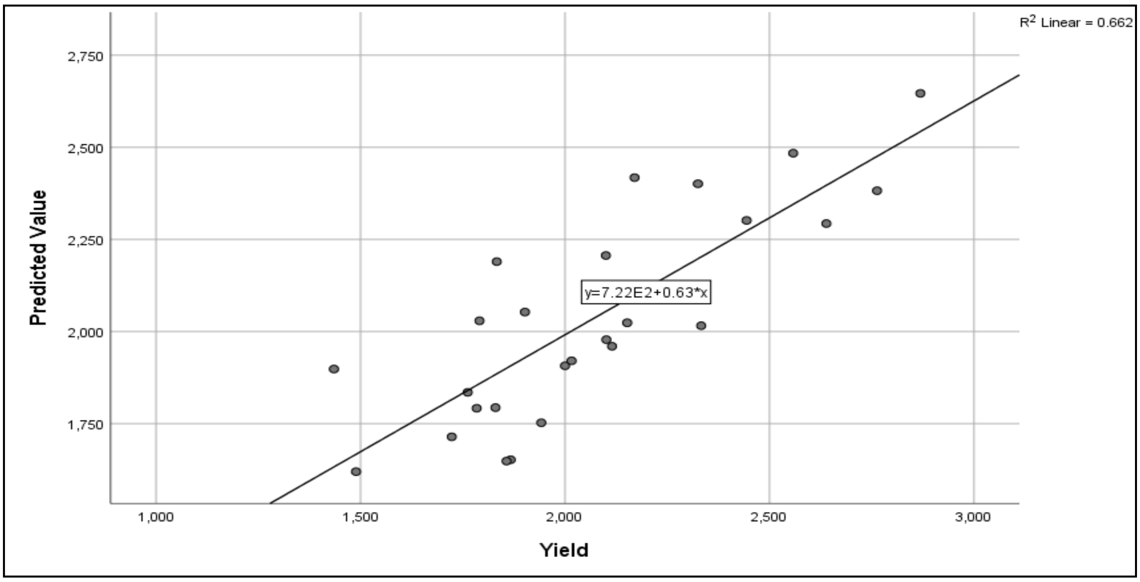

Fig.2. Comparison between actual and predicted yield

Table 7. Inter comparative study of regression and ANN based fitted best models with accuracy measure of RMSE & nRMSE

Weekly Weather		Bright Sunshine Hour		Weighted and Un-weighted		Actual Evapo-transpiration		Artificial Neural Network	
RMSE	nRMSE	RMSE	nRMSE	RMSE	nRMSE	RMSE	nRMSE	RMSE	nRMSE
0.436	0.0061	0.242	0.00025	0.451	0.00135	0.154	0.00027	0.215	0.00051

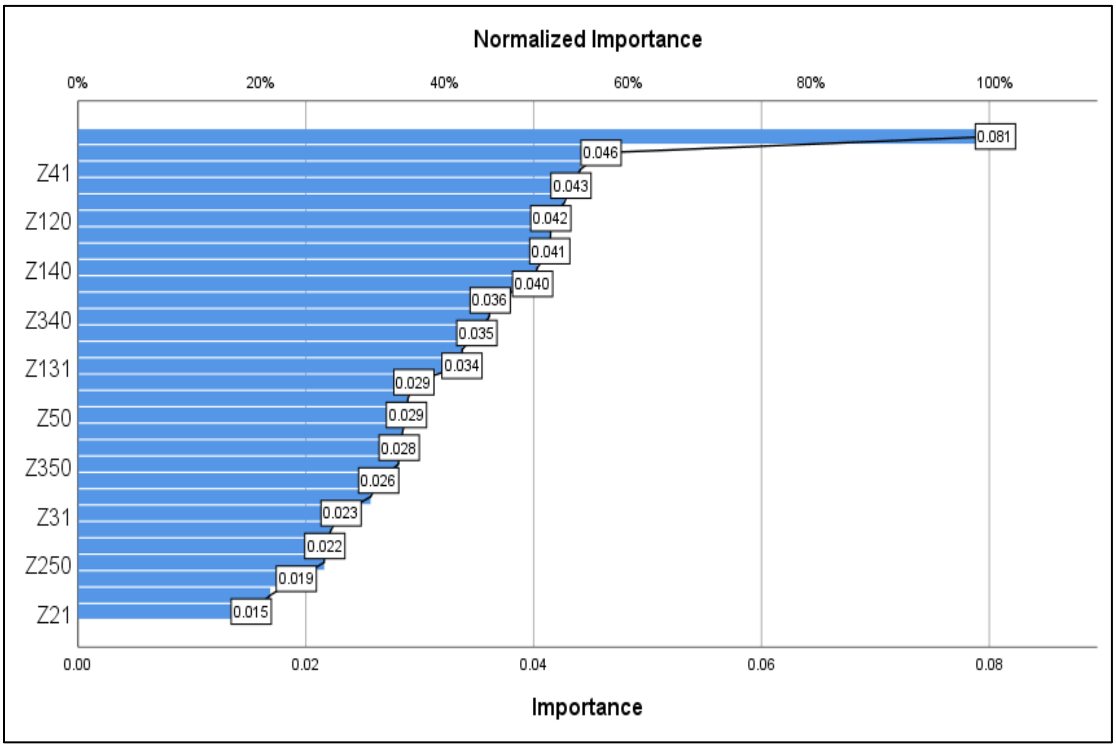

Fig.3. Relative contribution of predictors (weighted and un-weighted weather variables)

3.7 Performance Analysis of Fitted Models for Different Methods

The precision of the projected model's performance is assessed through the utilization of Root Mean Square Error (RMSE) and Normalized Root Mean Square Error (nRMSE). From the pool of regression models devised using the stepwise regression technique, the most favorable models are selected for comparison with the ANN model.

As can be discerned from (Table 7), the model developed through the regression technique, incorporating BSS and AET as additional inputs alongside weighted and un-weighted weather parameters, has demonstrated the most favorable outcomes. This model yields the lowest nRMSE value. It is followed in sequence by the ANN approach, the weighted and un-weighted parameter model, and lastly by the model crafted using weekly weather parameters. On the other hand, if we focus on the RMSE metric, then the model developed utilizing AET has showcased the finest performance among all alternatives.

4. CONCLUSION

The study demonstrated that regression techniques can be effectively employed for yield prediction in specific areas, yielding satisfactory results. In order to forecast yield, both Regression models and Artificial Neural Networks (ANN) are harnessed as prediction tools. In these models, the yield serves as the dependent variable, while weather parameters serve as the independent variables. Each model undergoes multiple runs to account for potential variations in normalized root mean square and R^2 statistical values. Employing the most suitable method for analysis, the prediction of rice crop production for selected years is achieved.

The outcomes underscore that the proposed regression model serves as a fitting approach for yield prediction. A comparative assessment of diverse models relies on normalized root mean square, R^2 statistics, and percentage prediction error. The model exhibiting the lower normalized root mean square, lower percentage prediction error, and higher R^2 statistics values is identified as the optimal choice for crop yield prediction.

Drawing from the recent study's findings, the yield prediction model for rice crops using various statistical methods and a machine learning approaches showcases superior performance. Notably, the well-fitted model, Model-4 based on Bright Sunshine Hour with weighted and un-weighted weather indices, boasts the lowest nRMSE value of 0.00025 and the highest variability of 94% among all other models.

Given that the Bright Sunshine Hour-based model exhibited the lowest nRMSE value among all selected models using different techniques in the study, it can be concluded that Bright Sunshine Hour, in conjunction with weighted and un-weighted weather indices, emerges as the most precise and dependable weather variable for predicting rice crop yield.

5. FUTURE RESEARCH ON CROP YIELD PREDICTION

1. **Integration of Advanced Machine Learning Techniques**: Future research can explore the integration of advanced machine learning techniques, such as deep learning and ensemble methods, to improve the accuracy of crop yield prediction models. These techniques have shown promising results in various domains and could potentially enhance the predictive capabilities of existing models.

2. **Incorporation of Remote Sensing Data**: Incorporating remote sensing data, such as satellite imagery and vegetation indices, can provide valuable insights into crop growth and health. Future research can focus on integrating these data sources into crop yield prediction models to capture the spatial and temporal variability of agricultural fields.

3. **Inclusion of Socioeconomic Factors**: Crop yield is influenced not only by weather conditions but also by socioeconomic factors, such as market prices, government policies, and farmer practices. Future research can explore the inclusion of these factors in yield prediction models to provide a more comprehensive understanding of the factors affecting crop production.

4. **Validation and Comparison of Different Models**: There is a need for further validation and comparison of different crop yield prediction models using independent datasets. This can help identify the strengths and weaknesses of different approaches and provide insights into the most effective models for specific crops and regions.

5. **Development of Decision Support Systems**: Future research can focus on developing decision support systems that integrate crop yield prediction models with other agricultural management tools. These systems can provide real-time recommendations to farmers, helping them make informed decisions regarding crop planning, resource allocation, and risk management.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Bharti, Das P, Banerjee R, Ahmad T, Devi S, Verma G. Artificial Neural Network Based Apple Yield Prediction Using Morphological Characters. *Horticulturae*. 2023;9(4):436. https://doi.org/10.3390/horticulturae9040436
2. Banakara KB, Popat RC, Amaresh, Pandya HR. Pre-harvest forecast of Kharif Rice using weather parameters in Bharuch district of Gujarat state. Int. J. Chem. Stud., 2018;6(6):1111- 1116.
3. Latwal A, Nain AS, Jha A, Ranjan R. Wheat yield prediction using satellite data and meteorological parameters for Udham Singh Nagar district of Uttarakhand. Int. J. Basic Appl. 2017;15(3):130-132.
4. Diwan UK, Puranik HV, Das GK, Chaudhary JL. Yield prediction of wheat at preharvest stage using regression based statistical model for 8 district of Chhattisgarh, India. Int. J. Curr.Microbiol. 2018;7(1):2180-2183,
5. Diwan UK, Puranik HV, Das GK, Chaudhary JL. Yield prediction of wheat at preharvest stage using regression based statistical model for 8 district of Chhattisgarh, India. Int. J. Curr.Microbiol. 2018;7(1):2180-2183,
6. Kumar N, Pisal R, Shukla SP, Pandey KK. Crop yield forecasting of paddy, sugarcane and wheat through linear regression technique for south Gujarat. Mausam. 2014;65:361-364.
7. Kalubarme MH, Ahuja SC. Agrometeorological rice yield modelling in Karnal. J. Indian Soc. Remote Sens. 1996;24:125–132.
8. Chauhan VS, Shekh AM, Dixit SK, Mishra AP, Kumar S. Yield prediction model of rice in Bulsar district of Gujarat. J. Agrometeorol. 2009;11(2):162-168.

9. Ghosh K, Balasubramanian R, Bandopadhyay S, Chattopadhyay N, Rathore SK, LS, Development of crop yield forecast models under FASAL- a case study of kharif rice in West Bengal. J Agrometeorol. 2014;16(1):1–8.
10. Satpathi A, Setiya P, Das B, Nain AS, Jha PK, Singh S, Singh S. Comparative Analysis of Statistical and Machine Learning Techniques for Rice Yield Forecasting for Chhattisgarh, India. Sustainability. 2023; 15(3):2786. Available:https://doi.org/10.3390/su15032786
11. Das, B.; Nair, B.; Arunachalam, V.; Reddy, K.V.; Venkatesh, P.; Chakraborty, D.; Desai, S. Comparative Evaluation of Linear and Nonlinear Weather-Based Models for Coconut Yield Prediction in the West Coast of India. Int. J. Biometeorol. 2020, 64, 1111–1123.
12. Tufail M, Ormsbee L, Teegavarapu R, Artificial intelligence-based inductive models for prediction and classification of fecal coliform in surface waters. J. Environ. Eng. 2008;134(9) :789e799.
13. Rumelhart D, Hinton G, Williams R. Learning representations by back-propagating errors. Nature. 1986;323:533–536.
14. Wasserman PD. Neural Computing: Theory and Practice. Van Nostrand Reinhold Co., New York; 1989.
15. Fausett LV. Fundamentals of Neural Networks: Architectures, Algorithms, and Applications. Prentice-Hall, Upper Saddle River, NJ, USA; 1994.
16. Haykin S. Neural Networks: A Comprehensive Foundation, second ed. Prentice Hall, Upper Saddle River, NJ, USA; 1994.
17. Hara, P.; Piekutowska, M.; Niedbała, G. Prediction of Pea (*Pisum sativum* L.) Seeds Yield Using Artificial Neural Networks. Agriculture 2023;*13*:661.
18. A Gupta, K. Sarkar, D. Dhakre, and D. Bhattacharya, “Weather based crop yield prediction using artificial neural networks : A comparative study with other approaches”, MAUSAM. 2023;74(3):825–832.
19. Naushad Khan, Ajay Kumar, Vijay Dubey, CB. Singh, Sanjeev Kumar and Shubham Singh. Estimated Yield Forecasting of Rice and Wheat for Central Uttar Pradesh using Statistical Modal. Int. J. Curr. Microbiol. App. Sci. 2020;9(8):2490-2495.
20. Kakati N, Deka RL, Das P, Goswami J, Khanikar PG, Saikia H. Forecasting yield of rapeseed and mustard using multiple linear regression and ANN techniques in the Brahmaputra valley of Assam, North East India. Theor. Appl. Climatol. 2022;150: 1201–1215.
21. Rajavel M, Khare P, Prasad JR, Singh KK, Puranik HV, Das GK. Development of Rice Yield Forecast in Midseason Using Weather Indices based Agrometeorological Model in Chhattisgarh Vayu Mandal. 2018;44(1):38-45.
22. Patil CS, Puttanna B. 56Influence of actual evapotranspiration, growing degree days and bright sunshine hours on yield of finger millet", Mausam. 2009;60(3):343-348
23. Samad Emamgholizadeh, M. Parsaeian, Mehdi Baradaran,. Seed yield prediction of sesame using artificial neural network, European Journal of Agronomy. 2015; 68:89-96.
ISSN 1161-0301,

Peer-review history:
The peer review history for this paper can be accessed here:
https://www.sdiarticle5.com/review-history/104851