



ISSN Print: 2618-060X
 ISSN Online: 2618-0618
 NAAS Rating (2026): 5.20
 IJRA 2026; SP-9(10): 196-205
www.agronomyjournals.com
 Received: 01-08-2026
 Accepted: 24-09-2026
 Published: 08-10-2026

PV Gaikwad

PG Scholar, Department of Agril. Meteorology, College of Agriculture, Pune, Mahatma Phule Krishi Vidyapeeth, Rahuri, Maharashtra, India

Dr. SK Upadhye

Senior Scientist (SWCE), AICRPDA, ZARS, Solapur, Mahatma Phule Krishi Vidyapeeth, Rahuri, Maharashtra, India

Dr. AA Bhagat

Assistant Professor of Statistics, Division of Economics, College of Agriculture, Pune, Mahatma Phule Krishi Vidyapeeth, Rahuri, Maharashtra, India

Dr. SV Bagade

Assistant Professor of Agril. Meteorology, Department of Agril. Meteorology, College of Agriculture, Pune, Mahatma Phule Krishi Vidyapeeth, Rahuri, Maharashtra, India

Dr. SV Khadtare

Agrometeorologist, AICRP on Agrometeorology, ZARS, Solapur, Mahatma Phule Krishi Vidyapeeth, Rahuri, Maharashtra, India

AC Aiwale

PG Scholar, Department of Agril. Meteorology, College of Agriculture, Pune, Mahatma Phule Krishi Vidyapeeth, Rahuri, Maharashtra, India

Corresponding Author:

PV Gaikwad

PG Scholar, Department of Agril. Meteorology, College of Agriculture, Pune, Mahatma Phule Krishi Vidyapeeth, Rahuri, Maharashtra, India

Machine learning-based prediction of pigeon pea area and production in relation to weather parameters in Solapur district, Maharashtra

PV Gaikwad¹, SK Upadhye², AA Bhagat³, SV Bagade⁴, SV Khadtare⁵, AC Aiwale⁶

DOI: <https://www.doi.org/10.33545/2618060X.2026.v9.i10Sc.6826>

ABSTRACT

Pigeon pea (*Cajanus cajan* L.) is an important pulse crop in India and its area and production are influenced by prevailing weather conditions, particularly in semi-arid and rainfed regions. The present study was undertaken to identify weather-based prediction models for area and production of pigeon pea in Solapur district of Maharashtra using machine learning algorithms. Secondary time series data on pigeon pea area and production along with annual weather parameters, namely maximum temperature, minimum temperature, rainfall, relative humidity I relative humidity II, wind speed, evaporation and bright sunshine hours were collected for the period 1968–69 to 2022–23. Descriptive statistics, time series trend analysis and Pearson correlation analysis were employed to examine the variability, trends and association between weather parameters and pigeon pea area and production. Machine learning algorithms including Linear Regression, Ridge Regression, Least Absolute Shrinkage and Selection Operator Regression, Elastic Net Regression, Artificial Neural Network, K-Nearest Neighbour, Decision Tree, Random Forest, Support Vector Regression and Gradient Boosting were evaluated using an 80:20 training-testing data split. Model performance was assessed using the coefficient of determination (R^2) and Root Mean Square Error (RMSE). The time series analysis indicated a decreasing trend in pigeon pea area, whereas production showed a slight increasing trend with considerable year-to-year variability. Correlation analysis revealed that rainfall had a positive association with pigeon pea production, while maximum temperature showed a negative association with pigeon pea cultivation. Wind speed, rainfall and bright sunshine hours were positively associated with pigeon pea area. Among the evaluated models, Gradient Boosting showed the highest R^2 for pigeon pea area ($R^2 = 0.99$) with RMSE of 100.98, while Linear Regression performed best for pigeon pea production ($R^2 = 0.83$) with RMSE of 48.13. Validation using training and testing datasets showed good agreement between observed and predicted values. Five-year forecasting for 2023–2027 indicated a gradual decline in both pigeon pea area and production during the forecast period. The study demonstrates the usefulness of weather-based machine learning models for forecasting pigeon pea area and production and provides information that can support crop planning, resource allocation and weather-based agricultural decision-making in Solapur district.

Keywords: Pigeon pea, Area, Production, Machine learning algorithms, Weather parameters, Forecasting

Introduction

Pigeon pea (*Cajanus cajan* L.) is an important kharif pulse crop cultivated widely in India, particularly in semi-arid and rainfed regions. It is well adapted to relatively dry conditions and can be grown on marginal soils. Pigeon pea is commonly known as Arhar in Hindi and Tur in Marathi and belongs to the family Fabaceae. Its seeds contain about 20–22 per cent protein and provide important nutrients such as carbohydrates, minerals and vitamins, making it an important component of the human diet. In addition to its nutritional value, pigeon pea contributes to soil fertility through biological nitrogen fixation and supports sustainable agricultural production. Maharashtra is one of the major pigeon pea producing states in India, along with Madhya Pradesh, Karnataka and Telangana. As the crop is predominantly cultivated under rainfed conditions, its growth and production are closely influenced by seasonal weather variability.

Solapur district of Maharashtra has a semi-arid climate characterized by low and erratic rainfall, high temperatures and frequent drought conditions. Pigeon pea is an important pulse crop cultivated in the district along with chickpea, green gram and black gram. Since its cultivation largely depends on rainfall, variations in the amount and distribution of rainfall may result in considerable fluctuations in crop area and production. Although pigeon pea has comparatively good drought tolerance, moisture stress and high temperature during critical growth stages, particularly flowering and pod filling, can adversely affect crop growth and yield. Consequently, variations in prevailing weather conditions may lead to year-to-year fluctuations in pigeon pea area, production and productivity.

Weather parameters such as temperature, rainfall, relative humidity, wind speed, evaporation and bright sunshine hours have an important influence on crop growth and development. Rainfall determines the availability of soil moisture, while temperature influences physiological processes and different stages of crop development. Relative humidity, wind speed, evaporation and bright sunshine hours also affect the crop microclimate and water requirement. Weather-based crop models are therefore used as scientific tools to quantify the relationship between weather variables and crop performance and to provide useful information for agricultural planning and decision-making under variable climatic conditions. Kumar and Gautam (2014) [1], reported that weather parameters play an important role in crop growth and productivity, whereas Singh *et al.*, (2016) [2], reported that variations in weather conditions can influence crop area, productivity and total production of major crops.

Traditional statistical techniques such as Linear Regression and Multiple Regression are commonly used for crop prediction because of their simple structure and ease of interpretation. However, these approaches may have limitations in capturing complex and nonlinear relationships between weather parameters and crop performance (Challinor *et al.*, 2009) [3]. With the increasing availability of long-term agricultural and weather datasets, machine learning techniques have gained importance in crop prediction because of their ability to identify complex patterns and relationships within datasets. Various machine learning algorithms, including Random Forest, Artificial Neural Network, Support Vector Regression, Decision Tree, K-Nearest Neighbour and Gradient Boosting, can be employed to develop weather-based prediction models. These approaches can be used to estimate crop area and production based on historical agricultural and weather observations.

In rainfed and semi-arid regions such as Solapur, the development of weather-based prediction models for pigeon pea area and production can provide valuable information for agricultural planning and resource management. Forecast information on expected crop area and production can assist farmers and planners in crop planning, input management, storage and marketing. Therefore, the present study was undertaken to examine the relationship between pigeon pea area and production and weather parameters and to develop and validate suitable machine learning models for predicting pigeon pea area and production in Solapur district of Maharashtra.

Materials and methods

Data collection

Secondary time series data on pigeon pea area and production for the period from 1968–69 to 2022–23, covering 55 years, were collected from the Directorate of Economics and Statistics, ICRISAT district-level data and official publications of the Government of Maharashtra. The data were obtained from the (<http://data.icrisat.org/dld/src/crops>) and (<https://data.desagri.gov.in/website/crops-apy-report-web>). Weather data were collected from the India Meteorological Department, Pune, Research Station, State Department of Agriculture and Department of Agricultural Meteorology. The weather parameters considered in the study included maximum temperature (°C), minimum temperature (°C), rainfall (mm), relative humidity I (%), relative humidity II (%), wind speed (km h⁻¹), evaporation (mm) and bright sunshine hours (h).

Software used

The entire statistical analysis and machine learning modelling were carried out using Python software. It is an open-source, high-level programming language extensively used for data analysis, statistical computing, machine learning and predictive modelling. It provides a flexible and efficient environment for handling large datasets, applying machine learning algorithms, evaluating their performance and generating graphical outputs. Python is freely available and can be

downloaded from the official Python Software Foundation website. Data pre-processing is an essential step that prepares the dataset before model development. Pre-processing improves model performance and ensures reliable prediction results. Pre-processing techniques included data cleaning, Missing value analysis, Feature selection, Feature scaling was used. The dataset was divided into training and testing subsets using 80:20 ratio which is commonly used in machine learning studies for model training and testing.

Statistical analysis and modelling approach

Descriptive statistics were computed to summarize the basic characteristics of the dataset. These measures provided information on the central tendency, variability and distribution of the crop and weather variables over the study period (1968-69 to 2022-23). The descriptive statistics included Mean, Median, Standard deviation, Minimum and Maximum values. The mean represents the average value of a variable over the study period and provides an estimate of its central tendency. The standard deviation measures the dispersion of observations around the mean, indicating the degree of variability present in the data. The minimum and maximum values describe the range of observations, while the quartiles summarize the distribution of the dataset and help identify the spread of the data.

Correlation analysis: Correlation analysis was performed to examine the relationship between weather variables and the response variables (crop area and production). It measures the strength and direction of the linear association between two variables.

The Pearson correlation coefficient (*r*) was used in this study. The value of the correlation coefficient ranges from -1 to +1.

The Pearson correlation coefficient is calculated as:

$$r = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2}}$$

Where,

- X_i = Observed value of the independent variable
- Y_i = Observed value of the dependent variable
- $\bar{X}$ = Mean of the independent variable
- $\bar{Y}$ = Mean of the dependent variable

The following machine learning algorithms were used

1. Linear Regression
2. Ridge Regression
3. LASSO Regression
4. Elastic Net Regression
5. Artificial Neural Network (ANN)
6. K-Nearest Neighbour (KNN)
7. Decision tree
8. Random Forest
9. Support Vector Regression (SVR)
10. Gradient Boosting Regression

Each model was trained using the training dataset and evaluated using the testing dataset.

1. Linear Regression: Linear Regression was applied to examine the relationship between area, production and weather parameters. It estimates the effect of independent variables on the dependent variable assuming a straight-line relationship.

$$\hat{Y}_i = \beta_0 + \beta_1 X_i$$

Where,

- Y_i = dependent variable (area/production)

- X_i = independent (predictor) variable (Tmax, Tmin, Rainfall, RHI, RHII, BSS, WS, ET)
- β_0 = intercept (value of Y when $X = 0$)
- β_1 = slope coefficient (change in Y for a one-unit change in X)

2. Ridge Regression: Ridge regression is a regularized version of linear regression that adds a penalty term to shrink coefficient values. This helps to reduce overfitting and improves prediction accuracy, especially when independent variables are highly correlated.

$$\text{Ridge: } \min[\sum (y - \hat{y})^2 + \lambda \sum \beta^2]$$

Where,

- y = actual value
- $\hat{y}$ = predicted value from the model
- $\sum (y - \hat{y})^2$ = Sum of Squared Errors (prediction error)
- β = regression coefficients (slopes)
- $\sum \beta^2$ = sum of squared coefficients
- λ = regularization parameter (controls penalty strength)
- min = minimize

3. LASSO Regression: LASSO (Least Absolute Shrinkage Selector Operator) is a method used in linear regression models for shrinkage

$$\min_{\beta} \left\{ \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 + \lambda \sum_{j=1}^p |\beta_j| + (1 - \alpha) \lambda \sum_{j=1}^p \beta_j^2 \right\}$$

Where,

- Y_i = Observed value
- $\hat{Y}_i$ = Predicted value
- β_j = Regression coefficient
- λ = Regularization parameter
- α = Mixing parameter ($0 \leq \alpha \leq 1$)
- p = Number of predictors

5. Artificial Neural Networks (ANNs): The Artificial Neural Networks (ANNs) are a class of machine learning models inspired by the structure and function of the human brain. ANN consists of input, hidden, and output layers that process information through interconnected neurons.

$$\hat{Y}(x) = \frac{1}{K} \sum_{i \in N_K(x)} y_i$$

Where,

- $\hat{Y}(x)$ = Predicted value
- K = Number of nearest neighbours
- $N_K(x)$ = Set of K nearest neighbours
- Y_i = Observed value of the neighbouring sample

6. K-Nearest Neighbour (KNN): It is a non-parametric method that predicts values based on the similarity between observations. It assigns predictions based on average of the nearest neighbouring data points.

and variable selection method. Its goal is to find out the best set of predictors that minimizes the error and gives most accurate predictions. LASSO regression shrinks some of the coefficients (the numerical values attached to variables) in a model towards zero. If variable's coefficient becomes zero, it means that variable is not important and is removed from the model. Only variables with non-zero coefficients are kept.

$$\text{LASSO: } \min[\sum (y - \hat{y})^2 + \lambda \sum |\beta|]$$

Where,

- y = actual value
- $\hat{y}$ = predicted value
- $\sum (y - \hat{y})^2$ = Sum of Squared Errors
- β = regression coefficients
- $|\beta|$ = absolute value of coefficients
- $\sum |\beta|$ = sum of absolute coefficients
- λ = regularization parameter
- min = minimize the total error

4. Elastic Net Regression: The Elastic Net Regression combines the properties of both Ridge and LASSO regression. It is particularly useful when the predictor variables are highly correlated and when variable selection is required.

$$\hat{Y}(x) = \frac{1}{K} \sum_{i \in N_K(x)} y_i$$

Where,

- $\hat{Y}(x)$ = Predicted value
- K = Number of nearest neighbours
- $N_K(x)$ = Set of K nearest neighbours

Y_i = Observed value of the neighbouring sample

7. Decision tree: Decision Tree is a supervised machine learning algorithm used for classification and regression. In this study, it was used to predict crop area and production using weather parameters and other selected variables. The algorithm repeatedly splits the data based on the variable and threshold that minimize prediction error, forming a tree-like structure with decision nodes and leaf nodes. It can capture non-linear relationships and is easy to interpret; however, deep trees may lead to overfitting.

$$\hat{y} = f(x)$$

Where,

- $\hat{y}$ = Predicted value
- X = Input feature vector
- $f(\cdot)$ = Decision tree mapping function

8. Random Forest: The random Forest is a machine learning algorithm that uses many decision trees to make predictions. Each tree gives a result, and the forest takes the majority vote or average to decide the final answer. It helps make predictions more accurate and stable than a single tree.

$$\hat{y} = \frac{1}{B} \sum_{b=1}^B T_b(x)$$

Where,

- B = number of trees
- $T_b(x)$ = prediction from each tree
- $\hat{y}$ = final prediction

9.Support Vector Regression (SVR)

Support Vector Regression (SVR) is a machine learning algorithm used to predict continuous values. It tries to find the best line or curve that represents the relationship between the input variables and the target variable. Unlike ordinary regression, SVR allows a small amount of prediction error. This error is called the ε (ϵ) margin. Errors within this margin are ignored, while larger errors are reduced during training. This helps the model make better predictions.

$$f(X) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(X_i, X) + b$$

Where,

- $f(X)$ = Predicted output
- X = Input feature vector
- X_i = Training sample
- α_i, α_i^* = Lagrange multipliers
- $K(X_i, X)$ = Kernel function
- b = Bias term
- n = Number of training samples

10.Gradient Boosting: Gradient Boosting Regression (GBR) is a machine learning algorithm used to predict continuous values such as

$$R^2 = 1 - \frac{\text{error sum of squares}}{\text{total sum of squares}} = 1 - \frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{\sum_{t=1}^n (y_t - \bar{y})^2}$$

Where,

- y_t = actual observed value
- $\hat{y}_t$ = predicted value
- $\bar{y}_t$ = mean of observed values
- $\sum_{t=1}^n (y_t - \hat{y}_t)^2$ = residual sum of squares
- $\sum_{t=1}^n (y_t - \bar{y}_t)^2$ = total sum of squares

Root Mean Square Error (RMSE): It is the standard deviation of residuals (prediction errors). The RMSE tells us how concentrated the data is around the line of best fit. It has also been used for model selection. The minimum value of RMSE indicated the better model fit. Mathematically,

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{n}}$$

crop area or production. It improves prediction accuracy by combining many small decision trees instead of using only one tree. The model builds the trees one after another. The first tree makes an initial prediction. The next tree learns from the mistakes made by the first tree and tries to correct them. Each new tree continues to reduce the prediction errors. Finally, the predictions from all the trees are combined to produce the final result. Because each tree learns from the errors of the previous trees, Gradient Boosting can capture complex relationships between weather factors and crop production.

$$FM(X) = F_0(X) + \sum_{m=1}^M \gamma_m h_m(X)$$

Where,

- $F_M(X)$ = Final prediction after M boosting iterations
- $F_0(X)$ = Initial prediction
- $h_m(X)$ = Prediction from the m^{th} weak learner (decision tree)
- γ_m = Learning rate (step size) for the m^{th} tree
- M = Total number of boosting iterations
- X = Input feature vector

Goodness of fit: Model performance was assessed using the coefficient of determination (R (Singh *et al.*, 2016) [2]) and Root Mean Square Error (RMSE) (Elavarasan *et al.*, 2018) [4].

Coefficient of determination (R (Singh *et al.*, 2016) [2]): R (Singh *et al.*, 2016) [2] gives overall variation explained by the model due to independent variables. It represents the proportion of total variance in dependent variable that is explained by independent variables. The value of R (Singh *et al.*, 2016) [2] ranges from 0 to 1. A higher R (Singh *et al.*, 2016) [2] indicates better model performance.

Where,

- y_t = actual value
- $\hat{y}_t$ = predicted value
- n = number of observations

Model performance was evaluated using the coefficient of determination (R (Singh *et al.*, 2016) [2]), root mean square error (RMSE). The best-performing model was selected based on these evaluation metrics and forecasted the area and production of Pigeon pea for next five years.

Results and discussion

Descriptive statistical analysis

The results of descriptive statistics analysis are presented in Table 1. The results of descriptive statistics for area and production of pigeon pea are presented in Table 1. The average cultivated area showed greater variability as indicated by its higher S.D. compared to production.

Table 1. Results of descriptive statistics for area and Production of Pigeon pea in Solapur district (1968-69 to 2022-23)

Pigeon pea	Mean	Median	S.D.	Max.	Min.	Skewness	Kurtosis
Area ('00 ha)	378.01	353.15	184.47	772	53	0.15	-0.89

(continued on next page)

Pigeon pea	Mean	Median	S.D.	Max.	Min.	Skewness	Kurtosis
Production ('00 t)	121.58	106	80.31	340	7	0.99	0.83

The cultivated area varied from 53 to 772 hundred hectares, whereas production ranged from 7 to 340 hundred tonnes. These results indicated the noticeable fluctuations in pigeon pea area and production from year to year. The skewness values were positive for both area and production of pigeon pea showed that a few years recorded higher values than the average, particularly for production. The kurtosis values of area under pigeon pea indicated a flatter distribution, while the production of pigeon pea kurtosis value of 0.83 indicated a slightly more peaked distribution with some extreme production values.

Table 2. showed that the mean maximum and minimum temperatures were 34.05 °C and 20.02 °C, respectively, with low variation. The standard deviations were 0.56 °C and 0.47 °C respectively, indicated that both maximum and minimum temperature remained relatively stable during the study period.

Mean morning and evening relative humidity were 71.39% and 40.40%, respectively. The mean wind speed was 7.36 km/hr, positive skewness (1.72) indicated that a few years experienced relatively

higher wind speed than the average. Annual rainfall averaged 712.17 mm and showed considerable variation (270.30–1292.40 mm). Mean evaporation was 7.27 mm, while bright sunshine hours averaged 7.47 hours. Overall, rainfall showed the highest variability, whereas temperature and evaporation remained relatively stable during the study period.

Correlation analysis for area and production of pigeon pea with weather parameters

From Fig. 1, the correlation analysis results revealed that the cultivated area was negatively correlated with maximum temperature ($r = -0.63^{**}$), indicated that an increase in maximum temperature was associated with a reduction in the area under pigeon pea cultivation. Minimum temperature ($r = -0.29^*$) and RHII ($r = -0.10$) showed weak negative relationship with crop area. The wind speed (WS) ($r = 0.41^{**}$) had the positive significant correlation with the cultivated area, followed by rainfall (RF) ($r = 0.29^*$).

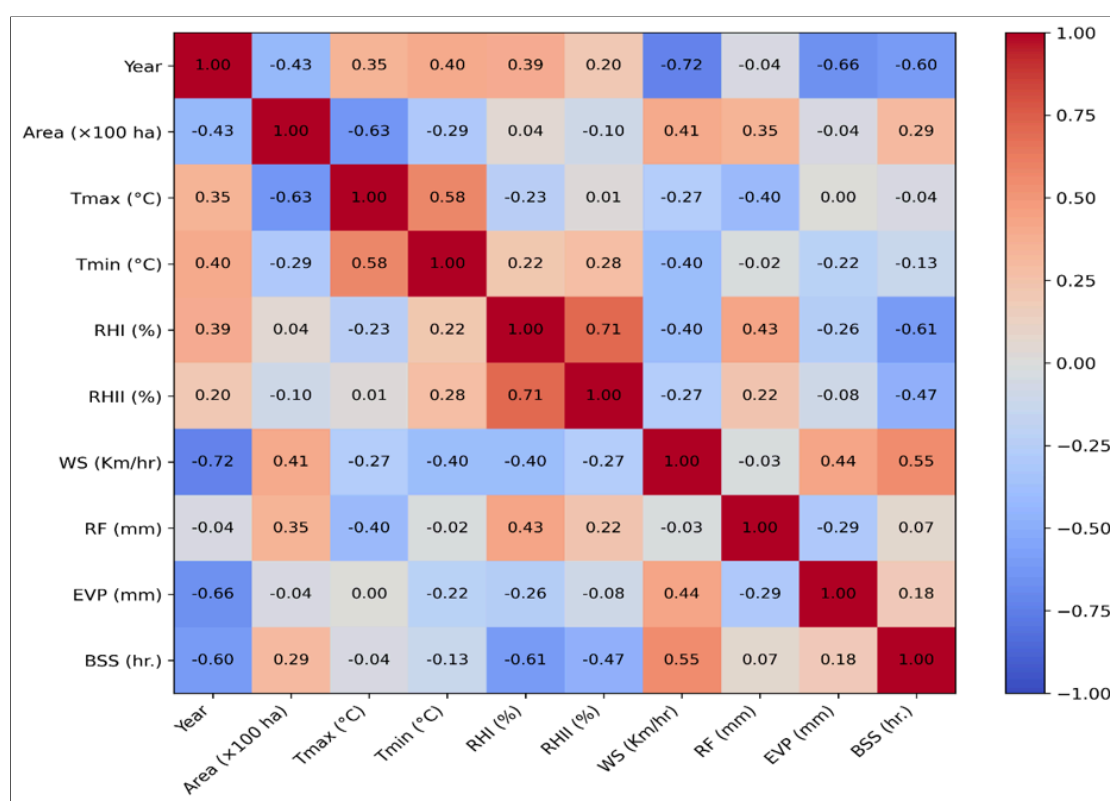


Figure 1. Correlation heatmap for area under and weather parameters

These positive correlations indicated that favourable wind conditions, adequate rainfall and more sunshine hours were associated with an increase in the area under pigeon pea cultivation. Bishnoi *et al.*, (1994) [5], reported that weather parameters influenced biomass production in pigeon pea, indicating that weather conditions play an important role in pigeon pea performance. The morning relative humidity (RHI) ($r = 0.04$) and evaporation (EVP) ($r = -0.04$) showed non-significant correlation with the area of pigeon pea in Solapur district.

Fig. 2 showed and revealed the results of correlation between pigeon pea production with weather parameters. that pigeon pea production was positive significantly correlated with rainfall ($r = 0.37^{**}$), indicated that higher rainfall was associated with increased production. RHI (0.28^*) and RHII ($r = 0.11$) showed positive non-significant correlation with production of pigeon pea.

Table 2. Descriptive statistics of weather parameters (1968-69 to 2022-23)

Particular	Mean	Median	S.D.	Max.	Min.	Skewness	Kurtosis
Tmax(°C)	34.05	34.06	0.56	35.42	32.81	0.01	-0.04
Tmin(°C)	20.02	20.13	0.47	21.15	18.86	-0.28	0.20
RH I (%)	71.39	71.00	3.27	78.32	63.90	0.06	-0.31

(continued on next page)

Particular	Mean	Median	S.D.	Max.	Min.	Skewness	Kurtosis
RH II (%)	40.40	40.26	3.63	48.54	32.10	-0.17	-0.36
WS (km/hr)	7.36	6.45	2.27	14.79	5.17	1.72	2.75
RF (mm)	712.17	692.4	209.37	1292.4	270.3	0.77	0.95
Evp. (mm)	7.27	7.28	0.94	9.40	4.96	-0.05	0.58
BSS (hr.)	7.47	7.61	0.75	8.76	4.77	-1.07	2.23

Evaporation had highly negative significant correlation ($r = -0.43^{**}$) with production of pigeon pea, suggested that increased evaporation reduced crop production. Maximum temperature ($r = -0.31^{*}$) had a significant negative correlation, indicated that higher maximum temperature was associated with minimum pigeon pea production. Wind speed ($r = -0.04$) and bright sunshine hours ($r = -0.01$) showed non-significant correlation with production, while minimum temperature ($r = 0.02$) had a negligible or non-significant positive correlation with

production of pigeon pea. Overall rainfall was positively correlated with both pigeon pea area and production, while higher maximum temperature adversely affected both the cultivated area and production. Patel *et al.*, (2026) [6], also reported that adequate rainfall promoted pigeon pea growth and yield, whereas elevated temperatures adversely affected crop performance, particularly during the flowering and pod development stages.

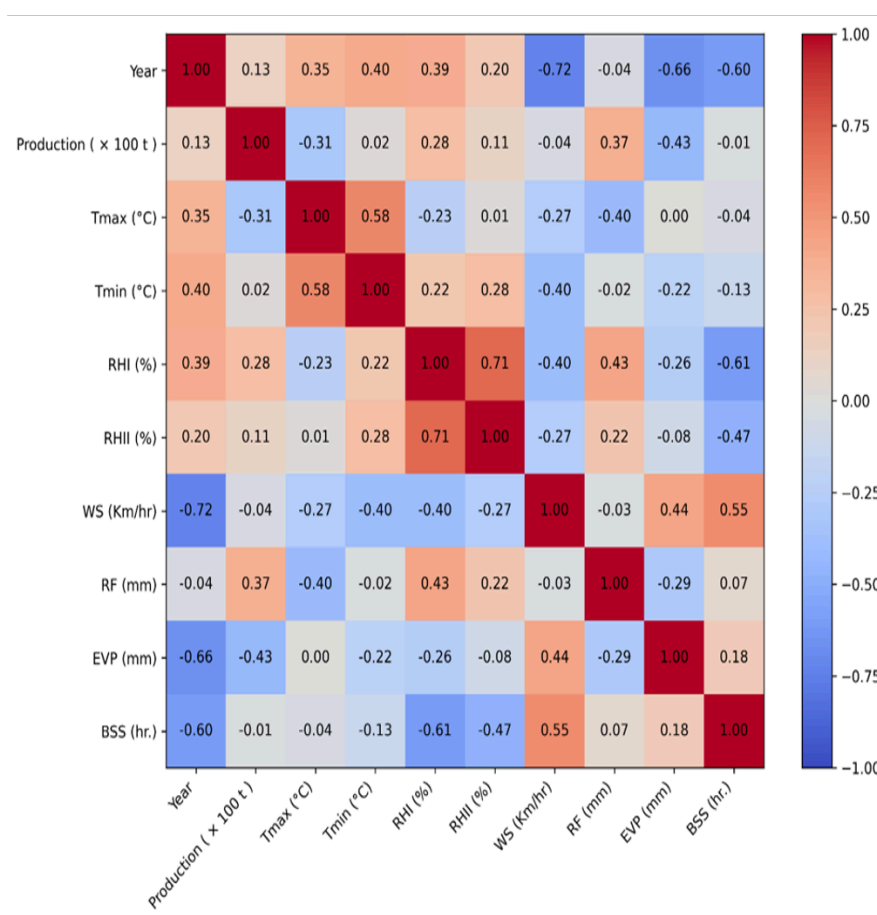


Figure 2. Correlation for production of Pigeon pea and weather parameters

Model performance

Different machine learning models were applied to analyse and predicted the area and production of Pigeon pea crop in Solapur district. The area and production of Pigeon pea were treated as dependent variables, while weather parameters as independent variables. Model performance was evaluated using R^2 and RMSE. R^2 values close to 1 and RMSE values close to 0 indicate better model performance.

Area prediction models

Gradient Boosting showed an R (Singh *et al.*, 2016) [2] of 0.99 and the lowest RMSE (100.98) among the evaluated models and was selected for pigeon pea area forecasting. Decision tree and ANN showed R (Singh *et al.*, 2016) [2] values (0.97 and 0.95). Ridge regression, elastic net, random forest, LASSO, linear regression, SVR

and KNN exhibited lower R (Singh *et al.*, 2016) [2] values and higher RMSE indicated comparatively lower prediction accuracy. The use of R (Singh *et al.*, 2016) [2] and RMSE for comparing the predictive performance of different machine-learning models in agricultural prediction studies has also been reported by Klompenburg *et al.*, (2020) [7]. Overall, Gradient boosting was the best fit model for pigeon pea area prediction.

Production prediction models

From the Table 3. Linear regression performed the best fit model, achieving the highest R (Singh *et al.*, 2016) [2] (0.83) and the lowest RMSE (48.13). Ridge regression, elastic net and LASSO showed good performance with R (Singh *et al.*, 2016) [2] values of 0.82 and RMSE values below 50.

Table 2. Performance comparison of machine learning algorithms for pigeon pea area prediction

MODELS	R ²	RMSE
Gradient boosting	0.99	100.98
Decision tree	0.97	102.47
ANN	0.95	104.69
Ridge regression	0.66	113.80
Elastic net	0.62	117.29
Random forest	0.91	130.87
LASSO	0.68	132.12
Linear regression	0.64	152.68
SVR	0.22	156.59
KNN	1.00	159.02

In contrast, decision tree, Gradient boosting, extra trees, random forest, SVR, KNN produced lower R (Singh *et al.*, 2016) [2] values and higher RMSE values, indicated comparatively lower prediction accuracy. Based on the evaluation metrics, Linear regression achieved the highest prediction accuracy among the evaluated models.

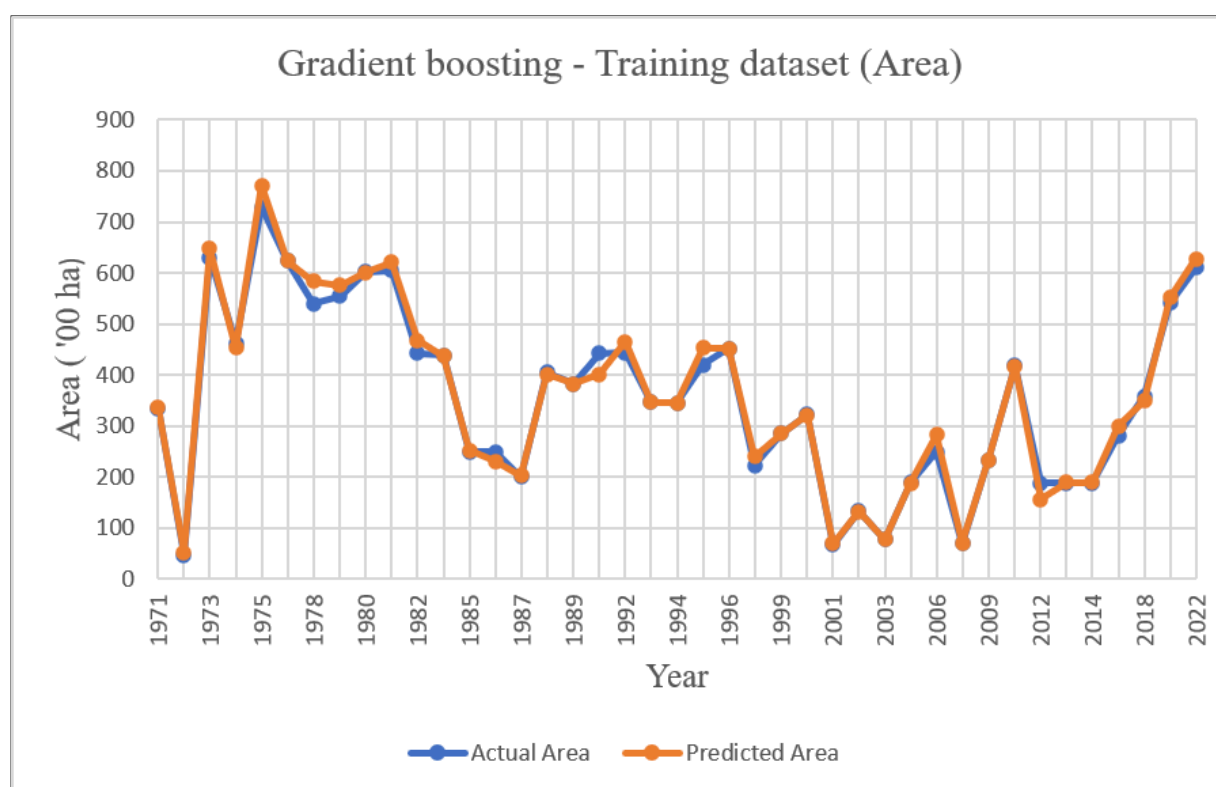
Table 3. Performance comparison of machine learning algorithms for pigeon pea production prediction

MODELS	R ²	RMSE
Linear regression	0.83	48.13

MODELS	R ²	RMSE
Ridge regression	0.82	48.92
Elastic net	0.82	49.06
LASSO	0.82	49.25
Decision tree	0.65	71.83
Gradient boosting	0.62	71.83
Extra trees	0.58	75.52
Random forest	0.54	79.58
SVR	0.49	83.85
KNN	0.27	99.84

Validation and forecasting of pigeon pea area by best fit model

The Gradient boosting model was applied using a random train-test split. Since the model used information from previous years (lag features), the first three years of data could not be included because no data from previous years were available to generate the lag features. Therefore, 52 years of data were used for model development, of which 41 years were used for training and the remaining 11 years for testing. Fig. 3 showed the comparison between the actual and predicted pigeon pea area using the gradient boosting model for the training dataset. The predicted values closely followed the actual cultivated area throughout the study period. The cultivated area increased sharply from about 47 in 1972 to nearly 729 in 1975, after which it gradually declined and remained around 540–623 hundred hectares during 1976–1981.

**Figure 3.** Actual and predicted values of pigeon pea area using the Gradient boosting model (Training dataset)

A further decline was observed during 1982–1986, with the area decreasing to about 200 hundred hectares. The cultivated area increased again during 1988–1996. After some year-to-year fluctuations between 2004 and 2017, the cultivated area increased rapidly from 2018 onwards, reaching about 610 hundred hectares in 2022. Only minor differences between the actual and predicted values were observed, indicating that the gradient boosting model represented the historical pattern of pigeon pea area satisfactorily.

Fig. 4 presented the comparison between the actual and predicted pigeon pea area for the testing dataset. The predicted values closely matched the actual values in all the selected test years. The cultivated area declined from about 550 in 1977 to around 160 in 2004. A slight increase was observed in 2007, followed by another decline during 2011 and 2015.

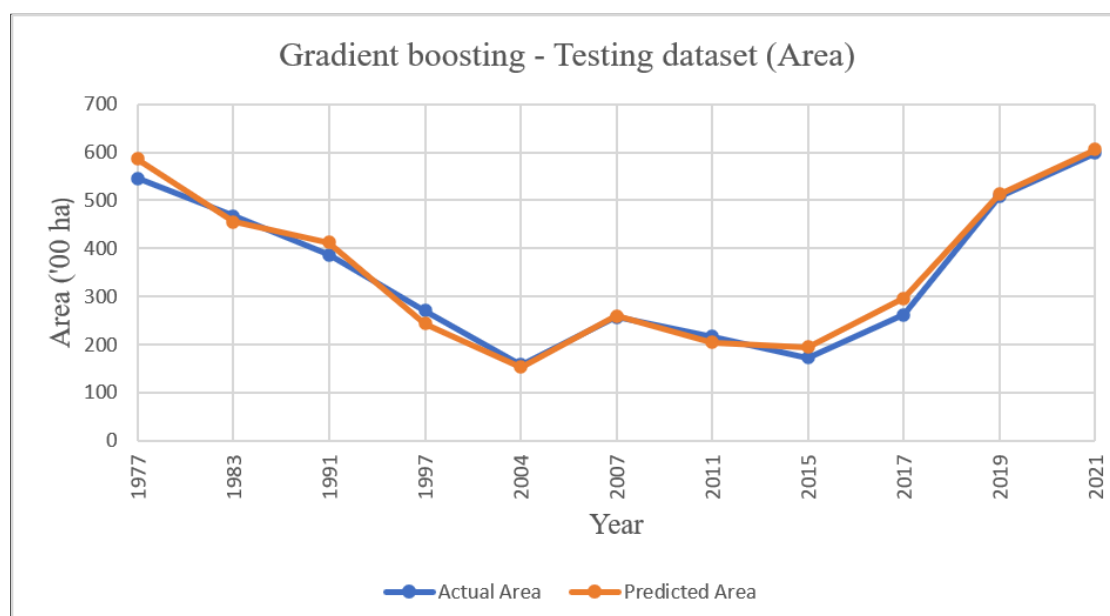


Figure 4. Actual and predicted values of pigeon pea area using the Gradient boosting model (Testing dataset)

From 2017 onwards, the cultivated area increased steadily and reached nearly 598 in 2021. The predicted values closely followed these changes, indicating that the gradient boosting model provided satisfactory predictions for the testing dataset.

Table 4. Forecasted values of pigeon pea area for the next five years in Solapur district.

Year	Forecasted values of Pigeon pea area ('00 ha)
2023	255.62
2024	251.56
2025	247.51

Year	Forecasted values of Pigeon pea area ('00 ha)
2026	243.45
2027	239.40

Figure 5. demonstrated the historical trend and five-year forecast of pigeon pea area using the Gradient Boosting model. The historical predicted values closely followed the actual cultivated area throughout the study period. The forecast indicated a gradual decline in pigeon pea area during 2023–2027. The forecast showed a gradual downward trend in pigeon pea area during 2023–2027. The cultivated area was projected to decrease from 255.62 to 239.40 in 2023–2027, indicating a slow decline in area over the forecast period

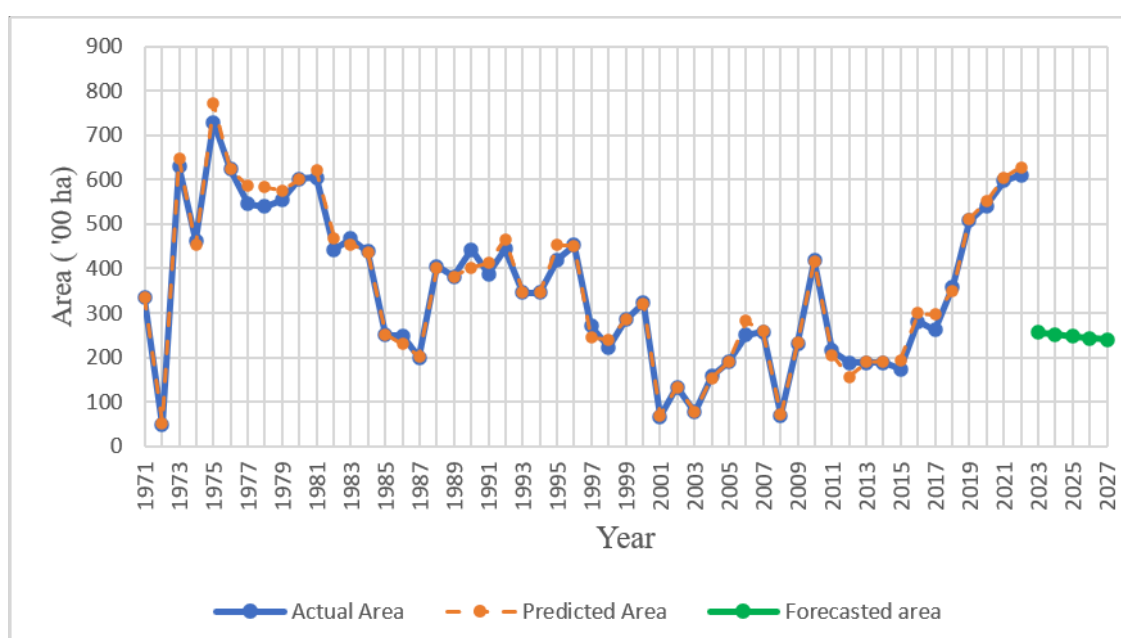


Figure 5. Actual, predicted and forecasted values of area under Pigeon pea using best fit Gradient boosting model

Validation and forecasting of pigeon pea production by best fit model

Fig. 6 showed the comparison between the actual and predicted pigeon pea production using the linear regression model for the training dataset.

The predicted values closely matched the actual production values throughout most of the study period, indicating that the model esti-

mated production with good accuracy. The model successfully captured the overall production pattern, with only slight deviations observed in 1979, 1980, 2008, and 2019. The figure shows that there is no difference between the actual and predicted values indicated that the linear regression model effectively represented the relationship between weather parameters and pigeon pea production in the training dataset.

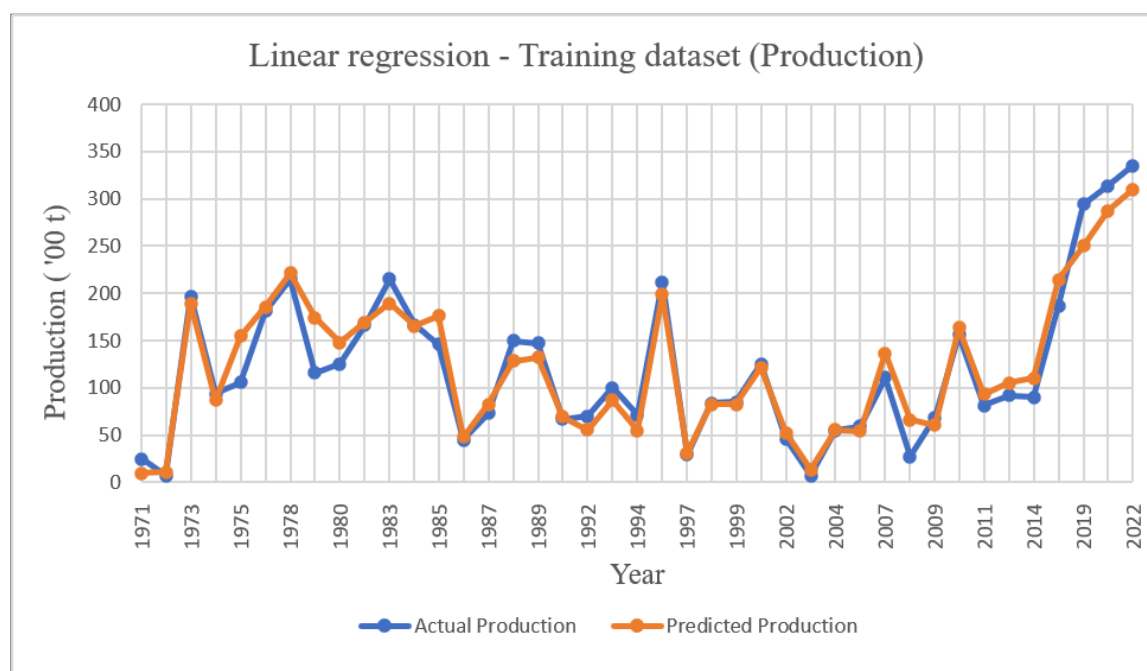


Figure 6. Actual and predicted values of pigeon pea production using the linear regression model (Training dataset)

Fig. 7 showed the comparison between the actual and predicted pigeon pea production for the testing dataset. The predicted values

closely matched the actual production values in most of the test years, indicating satisfactory predictive performance on unseen data.

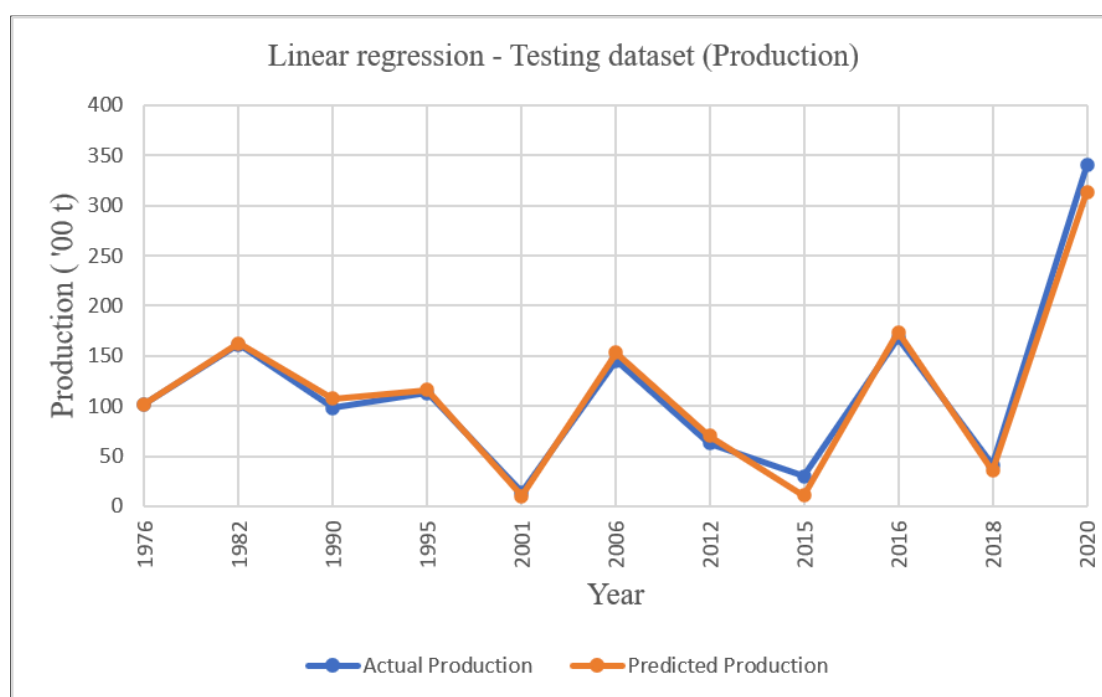


Figure 7. Actual and predicted values of pigeon pea production using the linear regression model (Testing dataset)

Slight overestimation and underestimation were observed in particularly 2015 and 2020; however, these differences were small. The overall production trend was captured well, indicating that the linear regression model provided reliable predictions for the testing dataset.

Table 5. Forecasted values of pigeon pea production for the next five year in Solapur district.

Year	Forecasted values of Pigeon pea production ('00 t)
2023	295.16
2024	284.74
2025	277.74
2026	272.19
2027	268.53

Fig. 8 illustrated the historical production trend along with the five-year forecast obtained using the linear regression model. The historical predicted values closely followed the actual production throughout the study period, indicating that the model represented the long-term production pattern satisfactorily. The forecast indicated a gradual decline in pigeon pea production during the period 2023–2027. The forecasted production decreased from 295.16 in 2023 to 268.53 in 2027, showing a small year to year reduction. These forecasted values suggested that pigeon pea production is expected to remain relatively stable but may decrease slightly in coming years. Overall, the linear regression model provided a satisfactory representation of the historical production pattern and generated reasonable forecasts for the next five years.

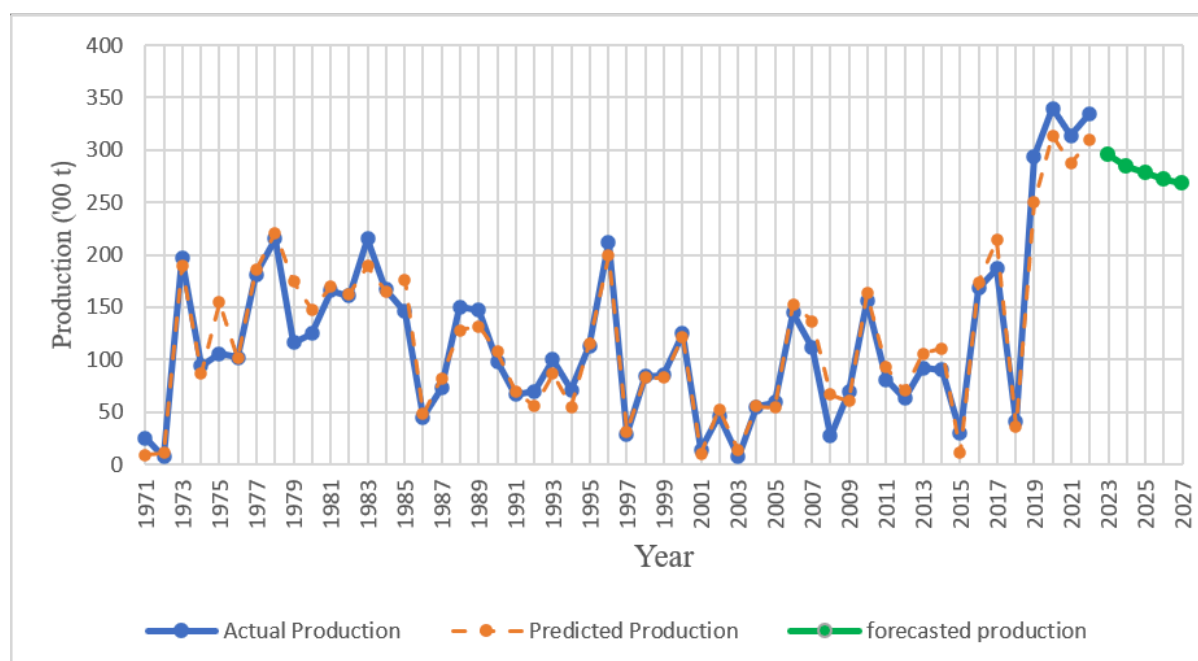


Figure 8. Actual, predicted and forecasted values of Pigeon pea production using best fit linear regression model

Conclusion

This study evaluated the predictive models of pigeon pea area and production. Pigeon pea area in Solapur district showed a decreasing trend over the years, whereas production showed a slight increasing trend with considerable year-to-year variability. Weather parameters such as rainfall and wind speed showed positive relationships with pigeon pea area, while maximum temperature showed a negative relationship with both area and production. Rainfall showed a positive relationship with production, whereas evaporation showed a negative relationship with production. Among all machine learning algorithms, the Gradient Boosting model performed best for pigeon pea area prediction, recording the highest R^2 (0.99) and lowest RMSE (100.98). For production prediction, Linear Regression performed best, with the highest R^2 (0.83) and lowest RMSE (48.13). Forecast results indicated a gradual decline in pigeon pea area from 255.62 to 239.40 hundred hectares and production from 295.16 to 268.53 hundred tonnes during 2023–2027. Overall, the developed weather-based models provided useful predictions of pigeon pea area and production and can support agricultural planning, resource allocation and weather-based decision-making in Solapur district.

Acknowledgements

The Department of Agricultural Meteorology, College of Agriculture, Pune and Department of Agricultural Economics and Statistics, College of Agriculture, Pune, Mahatma Phule Krishi Vidyapeeth, Rahuri provided research facilities to perform this experiment and statistical analysis, so I am thankful to the departments.

References

1. Kumar R, Gautam HR. Climate change and its impact on agricultural productivity in India. *J of Clim. & Weather Forecasting*. 2014;2(1):1-3.
2. Singh KK, Kalra N. Simulating impact of climatic variability and extreme climatic events on crop production. *Mausam*. 2016;67(1):113-130.
3. Challinor AJ, Ewert F, Arnold S, Simelton E, Fraser E. Crops and climate change: progress, trends, and challenges in simulating impacts and informing adaptation. *J of exper. Bot*. 2009;60(10):2775-2789.
4. Elvarasan D, Vincent DR, Sharma V, Zomaya AY, Srinivasan K. Forecasting yield by integrating agrarian factors and machine learning models: A survey. *Computers and Electr. in Agril*. 2018;155:257-282.
5. Bishnoi OP, Niwas R, Taneja KD, Rao VU. Effect of weather parameters on biomass production in pigeon pea. *Ann. of Arid Zone*. 1994;33(2):117-119.
6. Patel J. Climate change impact on pigeon pea (*Cajanus cajan*) yield in Maharashtra and Karnataka: A panel regression approach. *J of Agronomet*. 2026;28(1):92-95.
7. Klompenburg V, Kassahun A, Catal C. Crop yield prediction using machine learning: A systematic literature review. *Computers and Electr. in Agril*. 2020;177:105709.

How to cite this article: Gaikwad P, Upadhye S, Bhagat A, Bagade S, Khadtare S, Aiwale A. Machine learning-based prediction of pigeon pea area and production in relation to weather parameters in Solapur district, Maharashtra. *International Journal of Research in Agronomy*. 2026;SP-9(10):196-205. doi: <https://www.doi.org/10.33545/2618060X.2026.v9.i10Sc.6826>