

Prediction of Area, Production and Productivity of Black Gram in Andhra Pradesh using Machine Learning

Pullaganti Gowri¹, Nilavathy Kutty^{1*}, Alli Pandiyathuray²

¹School of Social Sciences and Languages, Vellore Institute of Technology, Vellore, India, ²Business School, Vellore Institute of Technology, Chennai, India. *Corresponding Author's Email: nilavathy.k@vit.ac.in

Abstract

Black gram, a pulse rich in protein, is mostly cultivated during Rabi in Andhra Pradesh, helping smallholder growers and improving soil fertility. The area, production, and productivity of Black gram can be predicted based on historical data that can provide potential insights. However, historical methods often do not work because they ignore recent shifts in climate, pest status, and responses in the market. This manuscript proposes a gaussian process regression (GPR), random forest (RF) named as GPRF method for forecasting the area, production, and productivity of black gram in Andhra Pradesh using historical agricultural data. Initially, input data is collected from India stat dataset. To execute, the input data is pre-processed using GPR, which removes null values from the data in the dataset. Then the pre-processed data is provided to RF, which is employed to forecast the area, production and productivity for rainfed crop black gram. The approach achieved a lower MAE of 0.028 and an RMSE of 0.021, outperforming existing techniques such as SVR, KNN and SVM. The forecasting period spans from 2026 to 2030 and focuses on area, production, and productivity. The proposed GPRF model accurately forecasts black gram cultivation in Andhra Pradesh, with low MAPE values: 1.33% for area, 2.04% for production, and 0.73% for productivity. The results indicate a strong upward trend, and the proposed method outperforms existing methods in terms of accuracy and reliability, making it highly suitable for strategic agricultural planning and policy-making.

Keywords: Agricultural Economics, Black Gram, Climate Variability, Crop Productivity Forecasting, Precision Agriculture, Rainfed Crops.

Introduction

Food security remains a critical development challenge in India, especially as the country strives to meet the nutritional demands of over 1.4 billion people (1). Agriculture forms the backbone of the Indian economy, contributing about 17% to the gross domestic product and providing livelihoods to nearly 45% of the population (2). Despite producing sufficient food grains to be self-reliant, India continues to import large quantities of pulses and edible oils to meet internal demand (3). Among the pulses, black gram (*Vigna mungo* L.), commonly known as urad, holds significant nutritional and economic value (4). It is a key protein source for a predominantly vegetarian population and is widely cultivated across several states, with Andhra Pradesh being a major producer (5). Notably, black gram is predominantly cultivated under rainfed conditions, making it highly dependent on monsoon patterns and seasonal rainfall (6). In India, rainfed agriculture accounts for more than 60% of the net sown area, is inherently risky due

to its reliance on erratic and uncertain rainfall (7). Black gram, as a rainfed pulse crop, is particularly vulnerable to climatic fluctuations such as delayed monsoons, irregular rainfall, and prolonged dry spells, all of which significantly impact its productivity and production (8). Moreover, historical shifts in agricultural practices like those during the Green Revolution show how technological and policy interventions have shaped crop productivity (9). However, the present-day challenges including rapid urbanization, shrinking cultivable land, climatic unpredictability, and unsustainable input usage are creating new barriers to sustainable black gram cultivation under rainfed regimes (10). To address these challenges, it is essential to develop accurate, region-specific forecasting models for rainfed black gram cultivation (11). As a crop largely dependent on monsoon rains, black gram is highly vulnerable to climatic variability (12). By leveraging advanced time-series analysis and machine learning techniques, it becomes possible

This is an Open Access article distributed under the terms of the Creative Commons Attribution CC BY license (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

(Received 25th July 2025; Accepted 01st October 2025; Published 31st October 2025)

to capture complex temporal and weather-related patterns that influence its growth (13). These models can improve the accuracy of forecasting area, production, and productivity, supporting informed decisions, efficient resource use, climate risk management, and adaptive policy-making (14). Focusing on the specific needs of rainfed agriculture, particularly for a vital pulse like black gram, promotes sustainable farming and enhances India's food and nutritional security in the face of increasing environmental uncertainty. Various research works have previously existed in the literature which is based on the prediction of area, production, and productivity for pulses, including black gram. Some of them are reviewed here application of machine learning (ML) techniques for forecasting of red gram prices of Andhra Pradesh, India. The methods employed included seasonal autoregressive integrated moving average (SARIMA), generalized autoregressive conditional heteroskedasticity (GARCH), support vector regression, artificial neural network, and random forest. These approaches were evaluated utilizing standard performance metrics such as mean squared error (MSE), root mean squared error (RMSE), and the diebold-mariano test to assess predictive accuracy. While traditional models and some machine learning methods showed moderate performance, the RF method demonstrated the highest forecasting accuracy. This study highlighted the effectiveness of ensemble learning in agricultural price prediction, although it also pointed out the limitations of certain standalone models and emphasized the potential need for hybrid or ensemble strategies to enhance forecasting reliability (15). In past study presented crop recommendations using ML approach for crop recommendation to support farmers in selecting a compatible crop for their growing environment and agronomic situation, considering many factors (weather, soil type, etc.). The model offers recommendations based on data and context that can ultimately improve productivity and sustainability. The standpoint of each model recommends to the user contextually dependent data and its effectiveness will depend on the quality of the data used to build the model. The model also ignores the socio-economic factors as well as the more subjective aspects of the model, which was transparency, which can influence trust and acceptance from the user (16). Researchers

invented in the past an integrated model that works on predicting pistachio productivity and selecting important soil variables through integrating the support vector machine (SVM) model with the firefly meta-heuristic method. The model was created using a dataset of 124 pistachio orchards and also identified important soil variables on productivity such as phosphorus, potassium, salinity, and gypsum. The hybrid model has improved predictive capacity and therefore could potentially handle non-linear relationships in the data more effectively than a single, non-hybrid model. The advantages were high accuracy and efficient soil management for sustainable production. The disadvantages deal with the reliance on a dataset that features large and high-quality datasets, and the complexity of the hybrid model that could make practical implementation difficult (17). In past study suggested a hybrid crop productivity prediction model utilizing an improved feature ranking fusion process. The methodology involves data normalization followed by the application of an enhanced synthetic minority over-sampling technique (SMOTE) algorithm to improve data for feature extraction. Key features, including correlation-based, statistical, entropy, and raw data features, were selected through an optimal feature selection process combining chi-square, relief, and recursive feature elimination (RFE) methods. The model then uses a hybrid approach combining long short-term memory (LSTM) and deep belief network (DBN) for prediction. The strengths lie in its high accuracy and robustness, while limitations include computational complexity and dependency on high-quality data (18). Researchers invented in the past cultivar-accelerated crop productivity - ML and multisensory data fusion for text classification in agriculture. This paper elucidates a novel approach to text classification in the context of agriculture, which combines ML and data fusion across multiple sensors. The aim was to improve agricultural intelligence through automated extraction and classification of taxonomic relevant textual data (which could improve monitoring of crop productivity). The combination of structured and unstructured data will provide a comprehensive look at agricultural systems, but the authors acknowledged a few issues around this broad-based method of analysis including the management and integration complexity

(including spatial-temporal), data quality sensitivity, and limitations on classification-based approaches without predictive or decision-support capabilities (19). In past study introduced, a mix of an auto encoder (AE), gated recurrent unit (GRU) the AEGRU approach, for long-term traffic forecasting (LTTF) prediction. The AE extracts the most significant features from the original data prior to performing dimensionality reduction. The AE subsequently generates accurate and comprehensible entry information for the GRU to increase its operational efficiency. The GRU uses the data provided by the AE to forecast future traffic volume (20). Researchers invented in the past point-driven interrupted time series and ML methodologies for forecasting Indian food grain production. This work proposes a hybrid approach using change-point detection, interrupted time series analysis, and ML to forecast India's food grain production. This approach aims to improve the forecast accuracy, especially when the structural changes, such as policy changes or climatic events. This approach adds robustness to the hybrid approach through the conjunction of statistical and ML approaches. However, change point detection requires some level of expertise in the area of interest to identify with confidence the correct change points, change point detection could be less interpretable by non-domain experts, and may not generalize well to data sets were no clear structural changes (21). The vector error correction (VEC) approach indicates effective price adjustments, while granger causality identifies Madhya Pradesh as the primary market influencing prices. Impulse response and variance decomposition analyses confirm the propagation of price shocks among states, reinforcing their correlation (22). Genetic diversity was observed among the black gram genotypes in terms of dry matter distribution, reproductive efficiency, and yield characteristics under high-temperature stress conditions. Out of the 30 genotypes evaluated for thermo tolerance, exhibited resilience to high-temperature stress in the reproductive phase, attributable to their elevated dry matter-to-pod setting ratio and improved yield. Correlation analysis revealed a strong positive relationship between total dry matter and all reproductive efficiency traits with seed yield under heat stress conditions. Principal component analysis (PCA) results showed significant variation

among the traits, accounting for 86.6% of the total variability (23). The general survey of the recent research work indicates that while there is significant progress in the prediction of crop productivity using ML and hybrid modelling approaches, one primary drawback has been a general lack of generalizability over different crops and climates. Many of the best performing models use quality data that is domain specific to the model, which may not be obtained in many situations. In conjunction, the complexity of these models and the low interpretability can make it difficult for end users in agriculture to practically implement a ML model-design. Many authors are dealing with the same problems with variation of the different technologies presented in the literature with support vector regression (SVR), k-nearest neighbours (KNN), and support vector machine (SVM). SVR has high computational cost and hyper parameter tuning, with it being tricky in terms of noisy data and non-linear relationships. KNN also does not have the ability to deal with time dependencies and struggles with large or high-dimensional datasets. Furthermore, KNN is very sensitive to the distance metric selected and how many neighbours to include. Similar as SVR, SVM will also consume clear computational seconds for larger datasets and is sensitive to, amongst many other things, the choice of kernel. While very effective for certain datasets, SVM is simply not effective to capture sequential behaviour. To address these issues, we suggest the GPRF method. In this paper, it has been proposed the GPRF based forecasting model will adequately provide a solution to the challenges associated with existing crop productivity prediction methods, such as poor generalizability, heavy reliance on data, and low interpretability and explore a GPRF hybrid methodology that integrates GPR for robust data pre-processing and RF for powerful prediction. This provides accurate forecasting of black gram area, production, and productivity, even with limited or noisy datasets. These hybrid approaches download the obstacles of over fitting data, whilst reducing computational complexity, and improves usability in real world situations such as rainfed cropping. The intelligence of the GPRF model has demonstrated the advancement of crop forecasting is highly scalable for agricultural plans at all levels, but especially advanced planning in resource limited regions. The novelty of this study

lies in the comparative ML framework and economic implications of black gram with the comparative ML framework.

Major Contribution of this paper includes:

- Developed a novel and scalable GPRF model tailored for long-term forecasting of black gram area, production and productivity, in rainfed regions like Andhra Pradesh.
- Introduced the use of GPR to accurately impute missing values, preserving statistical integrity and temporal consistency in datasets spanning over five decades (1971-2023).
- Introduced the use of RF helps provide more precise predictions of area, production, and productivity, enabling policymakers and farmers to make informed decisions regarding resource distribution, market strategies, and crop insurance programs.
- Demonstrated the model's effectiveness on a large-scale historical dataset, enabling informed decisions in crop planning, resource

allocation, and market risk management specific to rain-dependent agro-climatic zones.

Remaining manuscripts arranged as below: section 2 displays Methodology, section 3 depicts Results Discussion, and section 4 depicts Conclusion.

Methodology

In this section, a hybrid ML model for forecasting black gram area, production and productivity in Andhra Pradesh using historical data GPRF is proposed. The process comprises three main steps: Data Acquisition, Pre-processing, and Prediction. Initially, historical agricultural data is collected from the Indiatat dataset. The input data then undergoes pre-processing to prepare it for further analysis. Following this, the pre-processed data is fed into the RF model for predicting the area, production, and productivity of the rainfed black gram crop. The block diagram of the proposed GPRF approach is displayed in Figure 1. Below is a thorough explanation of each step.

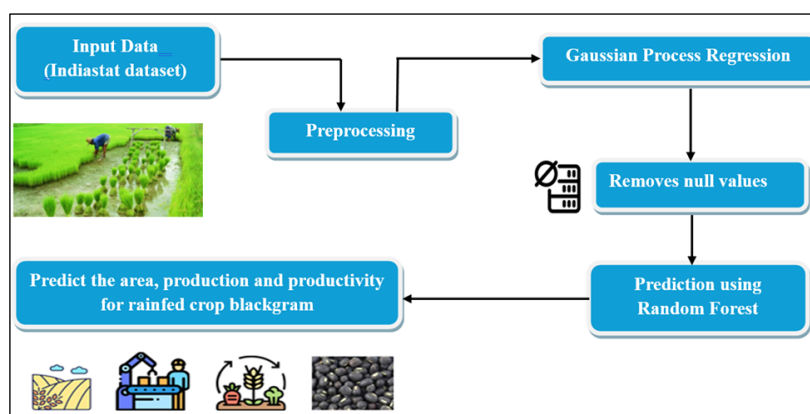


Figure 1: Block Diagram of GPRF Approach

Black Gram in Andhra Pradesh

Black gram, also called as urad dal, is a significant pulse crop cultivated extensively in Andhra Pradesh due to its adaptability to varied agro-climatic conditions. The state ranks among the top producers of black gram in India, particularly in districts like Guntur, Prakasam, Krishna, and Nellore. It is mainly cultivated in the kharif season under rain-fed conditions and in the rabi season with irrigation. Black gram is important for rotation and improving soil fertility, due to its nitrogen fixation ability. It is important in the food system as a significant protein source and holds value in the market, benefiting food security and farmers income. Government departments and services assist production with improved seeds,

pest control practices, and selling incentives through procurement to promote sustainable pulse production.

Data Acquisition

The input data is gathered from Indiatat, which provides information specifically for Andhra Pradesh black gram, offering strong secondary data that spans over five decades (1981-2025). The complete dataset is provided as annual statistics on key agricultural metrics such as area farmed (hectares), production (tonnes), and productivity (kilograms per hectare). This data is gathered from genuine government records and official documents, presented in an annual format similar to the previously mentioned Indiatat dataset, and it is dependable and consistent

enough for long-term agricultural studies (24). The dataset chronicles the growth and production of black gram across multiple jurisdictions in the Andhra Pradesh state and is useful to assess adaptation to climate variability, policy changes, technological interventions, and market changes over time. Long term temporal data allow for various trend detection, forecasting, and policy assessments and thus offer ground-breaking contributions to the body of research on food security, agricultural sustainability, and rural development.

Pre-processing using GPR

In this subsection we examine pre-processing using GPR removes missing values from the input dataset. GPR is useful for the pre-processing of agricultural forecasting, since long-term historical data are often plagued with missing values, noise and inconsistencies. GPR will allow for data

$$x = [x_1, x_2 \dots x_M]^S$$

Where, $x_1, x_2 \dots x_M$ denotes the noisy target observations corresponding to the input vector, denotes the instances, S denotes the random variable's conditional mean. Analyse the extent

$$B = \begin{bmatrix} B_x & K_{test} \\ K_{test}^S & K(Y_{test}, Y_{test}) + \frac{1}{\lambda} \end{bmatrix}$$

Where, B represents the Gaussian multivariate distribution with a covariance matrix and a mean of zero, Y_{test} represents the predicted value corresponding to the test input vector, K_{test} denotes the kernel between training inputs and

$$B_x = \begin{bmatrix} K(Y_1, Y_1) & \dots & K(Y_1, Y_M) \\ K(Y_2, Y_1) & K(Y_2, Y_2) + \frac{1}{\lambda} & \vdots \\ \vdots & \vdots & \vdots \\ K(Y_M, Y_1) & \dots & K(Y_M, Y_M) + \frac{1}{\lambda} \end{bmatrix}$$

Where, $K(Y_1, Y_1)$ represents the kernel function that evaluates the self-similarity of the input vector Y_1 , $\frac{1}{\lambda}$ represents the regularization parameter used to control model complexity and noise variance, $K(Y_2, Y_1)$ denotes the covariance between input Y_1 and Y_2 as computed using a kernel function, $K(Y_M, Y_M)$ denotes the self-covariance of M^{th} training inputs. The imputed data are validated using historical economic trends and regional agricultural benchmarks to ensure consistency with typical productivity patterns and the effects of policy changes. Finally, GPR is applied to remove

completion to permit temporal continuity of missing data estimates, to estimate missing values of variables such as area, production and productivity, for a statistically relevant and consistent forecast. GPR is non-parametric and probabilistic; thus, GPR has some advantages of perspective; it does not terminate the underlying trend and variability; it does provides a smoothed version of the variables while not terminating essential patterns present in the variables; it can provide good imputation of sparse, or irregular datasets; these are valuable and useful features in the analysis of agricultural economics, as reliable datasets are of tremendous importance for the development of policy, crop planning and resource allocation. Time-series data of black gram area, production and productivity in Andhra Pradesh are collected from official sources (25). A statistical audit will be made that will highlight missing values, as seen in Equation [1].

[1]

and distribution of missing values, considering factors like seasonal variability, droughts, or policy changes that impact agricultural economic analysis, as shown in Equation [2].

[2]

test inputs, K_{test}^S represents the Transpose of K_{test} . GPR is used to estimate missing values by modeling temporal and spatial trends, ensuring realistic and data-driven imputation aligned with agricultural dynamics, as shown in Equation [3].

$$\begin{bmatrix} K(Y_1, Y_M) \\ \vdots \\ K(Y_M, Y_M) + \frac{1}{\lambda} \end{bmatrix}$$

[3]

the null values from the input dataset. The cleaned data are then forwarded to the prediction phase.

Prediction using RF

In this section, prediction using RF is utilized to predict the area, production and productivity for rainfed crop blackgram. RF provides significant advantages for agricultural forecasting, particularly in data-rich yet heterogeneous environments such as rainfed blackgram cultivation in Andhra Pradesh. It handles non-linear relationships effectively and is resilient to overfitting, making it suitable for complex crop-

environment interactions. RF manage missing or imbalanced data better than many traditional models, providing consistent performance across varying agro-climatic zones. From an agricultural economics perspective, RF contributes to more accurate forecasts of area, production, and productivity, enabling policymakers and farmers to make data-driven decisions regarding resource allocation, market planning, and crop insurance schemes. The primary goal is to leverage the

$$Var(M) = \frac{1}{|M|} \sum_{x \in M} (x - \bar{x})^2 \quad [4]$$

Where, $Var(M)$ represents the variance of the data set, $|M|$ indicated as the Number of samples in set M, x indicated as each individual sample in the set, $\bar{x}$ indicated as the mean of the values in M. The RF

$$Reduction = Var(M) - \left[\frac{|M_l|}{|M|} Var(M_l) + \frac{|M_r|}{|M|} Var(M_r) \right] \quad [5]$$

Where, Reduction represents the gain from splitting; how much the variance is reduced, M represents the Parent node data, M_l and M_r Left and right child nodes after the split, $Var(M_l)$ and $Var(M_r)$ represents the variance of left and right split, $\frac{|M_l|}{|M|}$ and $\frac{|M_r|}{|M|}$ represents the weighted

$$\bar{x}_s = \frac{1}{E} \sum_{e=1}^E s_e x(s) \quad [6]$$

Where, $\bar{x}_s$ represents the Final predicted value for input s from the ensemble, $\frac{1}{E}$ represents the total count of models in the ensemble, $s_e x(s)$ represents the prediction of the eth model for input sample. Model predictions are evaluated using statistical metrics and are interpreted in terms of economic feasibility and regional agricultural planning. Finally, RF is used to predict the area, production and productivity for rainfed crop black gram.

Results

The actual outcomes of the proposed GPRF model are discussed in the segment. The assessments were conducted on a Windows 11 PC equipped with an Intel (R) Core (TM) i7 CPU running at 2.40 GHz and 16GB of RAM python was used to carry out the proposed strategy. Several performance criteria, like mean absolute percentage error (MAPE), root mean square error (RMSE), and mean absolute error (MAE) are evaluated. The outcomes of the proposed GPRF methodology are compared to those of current techniques such as SVR incurs significant computational and memory costs, as the training complexity increases rapidly

historical agricultural data to generate reliable and interpretable forecasts that support sustainable planning, minimize economic risk, and improve productivity in rainfed regions through informed interventions (26). Historical data on black gram area, production, and productivity in Andhra Pradesh are structured to reflect region-specific economic and climatic variations affecting rainfed agriculture, as shown in Equation [4].

model is trained using predictors like year, land use, and regional factors to capture non-linear agricultural-economic patterns, as shown in Equation [5].

proportions of each split. The trained RF model forecasts area, production, and productivity, supporting procurement planning, subsidy distribution, and regional food security assessment, as shown in Equation [6].

with dataset size, making it impractical for extremely large datasets. SVR faces difficulties with noisy data because the epsilon-insensitive margin lacks robustness against numerous outliers, which can lead to potential over fitting. Another drawback is the necessity of feature scaling, since variations in feature magnitudes can greatly impact the results (11). In KNN, the prediction step takes longer with larger datasets. Calculating precise distances is crucial for ensuring the accuracy of the algorithm. Determining the value of K in KNN can be costly when the dataset is extensive. It also requires more memory capacity than efficient classifiers or other supervised learning methods. KNN is highly sensitive to the scale of the data and to extraneous features. Irrelevant or correlated features can significantly affect the outcome and should be removed. The computational cost is considerably higher, as the distance to each training example must be calculated (12). SVM operates by identifying the optimal hyperplane that best separates the data into categories. This hyperplane is determined by the support vectors, which are the data points

closest to the boundary. SVM requires long training times, making it impractical for large datasets. Another limitation is that SVM classifiers struggle with overlapping classes. Overall, the algorithm is not well-suited for handling very large datasets. In situations where the dataset contains high levels of noise, such as overlapping target classes, SVMs tend to exhibit reduced performance (13).

Figure 2 illustrates the forecasted trend of black gram cultivation in Andhra Pradesh from 2026 to

2030. The Figure 2A area under cultivation is projected to increase from around 120,000 ha to 325,000 ha. Figure 2B production is expected to grow from 95,000 metric tonnes (mt) to nearly 290,000 mt. Figure 2C productivity also shows a steady rise from 6.5 mt/ha to 16 mt/ha (metric tonnes per hectare). These trends indicate a strong positive growth in black gram farming, highlighting the effectiveness of forecasting models for agricultural planning.

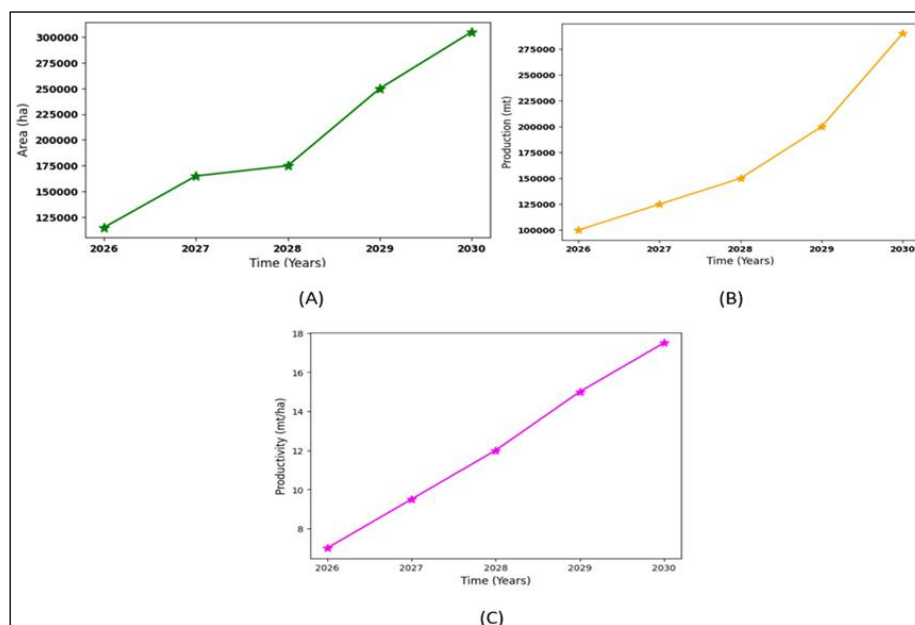


Figure 2: Forecasted Trend of Black Gram Cultivation in Andhra Pradesh: (A) Area, (B) Production, and (C) Productivity

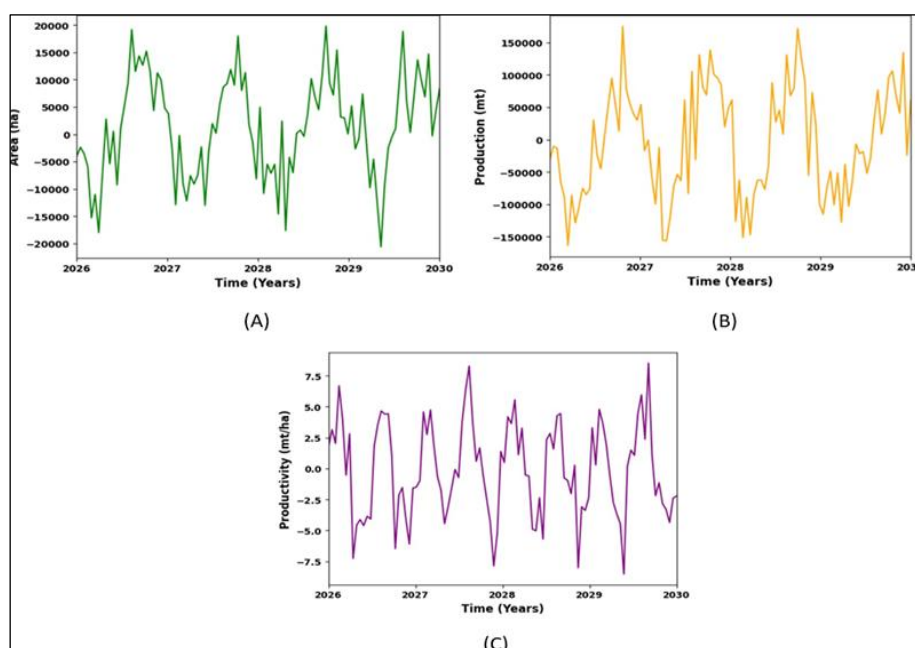


Figure 3: First Order Differencing of Black Gram Data for Stationarity: (A) Area, (B) Production, and (C) Productivity

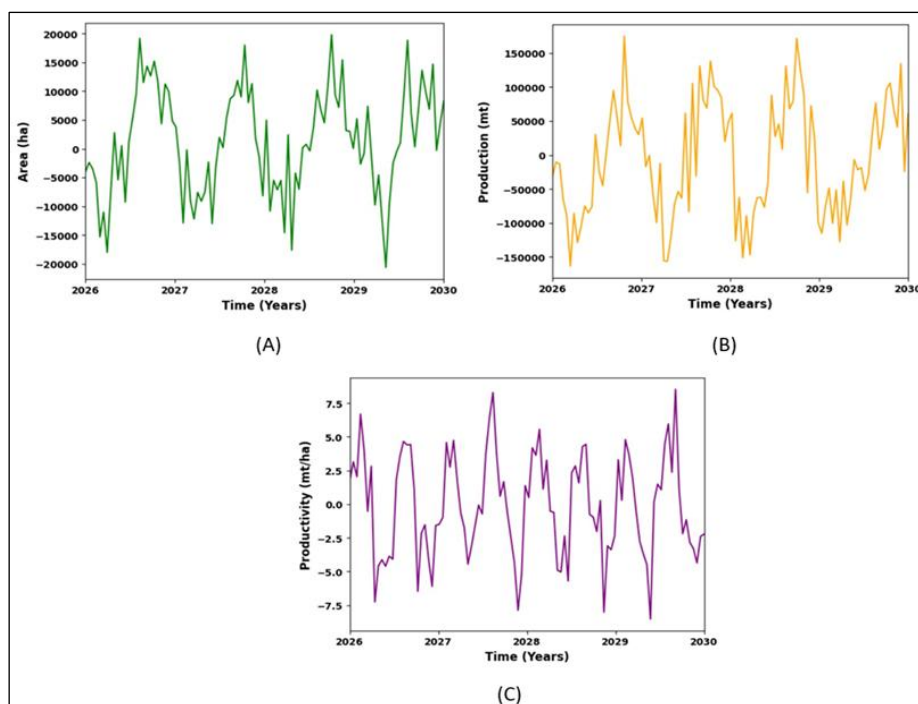


Figure 4: Second Order Differencing of Black Gram Data for Forecasting Preparation: (A) Area, (B) Production, and (C) Productivity

Figure 4 shows the second order differencing of Black gram data in Andhra Pradesh from 2026 to 2030, used for forecasting preparation. Figure 4A area fluctuates between -18,000 to 24,000 ha, Figure 4B production varies sharply from -180,000

to 320,000 mt, and Figure 4C productivity ranges from -3 to 10 mt/ha. These variations indicate volatility, and the differencing helps stabilize the data, making it suitable for accurate time series forecasting.

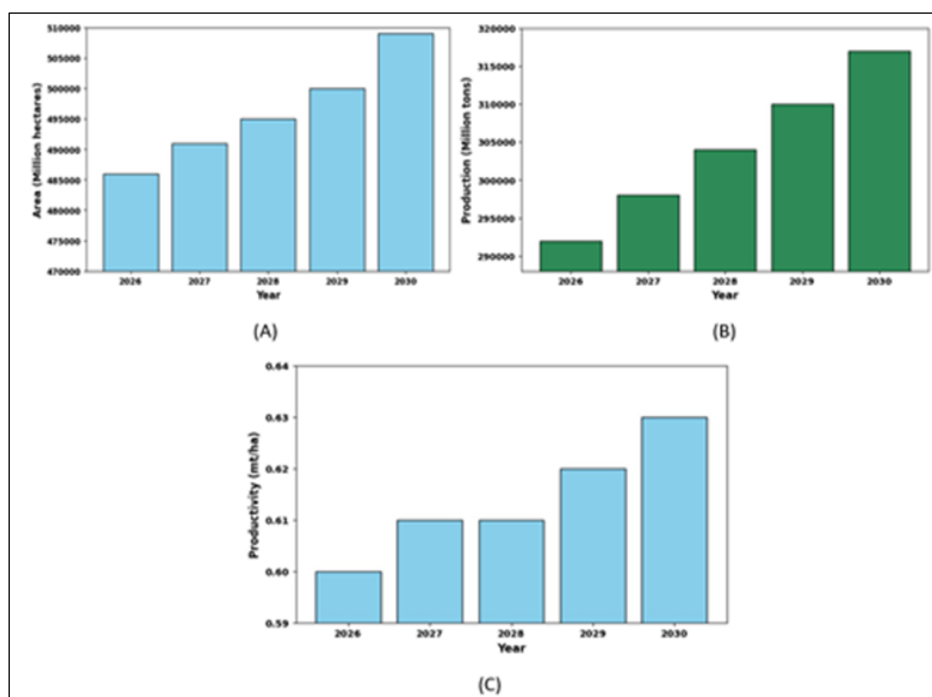


Figure 5: Forecasts of (A) Area, (B) Production, and (C) Productivity Of Black Gram

Figure 5 presents the forecasted values of area, production, and productivity of Black gram in Andhra Pradesh from 2026 to 2030 using the proposed GPRF model. Figure 5A increase in the land area cultivated is predicted from 486,200 hectares in 2026 to 504,600 hectares in 2030, indicating a slow growth of cultivation. Figure 5B production is anticipated to increase from 292,300 million tons in 2026 to 316,000 million tons in

2030, indicating an improvement in output. Figure 5C productivity is also increasing from 0.60 mt/ha in 2026 to 0.63 mt/ha in 2030. All these trends indicate a positive growth path in black gram production systems and support the use of GPRF model in predicting these trends, providing useful information for planning and resource use in the agricultural sector.

Table 1: Forecasted Area, Production, and Productivity of Black Gram

Year	Forecasted Area (ha)	Forecasted Production (MT)	Forecasted Productivity (MT/ha)
2026	486,200	292,300	0.60
2027	489,500	297,400	0.61
2028	494,300	305,000	0.61
2029	499,100	310,200	0.62
2030	504,600	316,000	0.63

Table 1 presents the forecasted area, production, and productivity of Black gram in Andhra Pradesh for 2026–2030. The area projected for cultivation shows a consistent rise from 486,200 ha in 2026 to 504,600 ha in 2030, which indicates gradual increases. Similarly, production is projected to show growth growing from 292,300 MT in 2026 to

316,000 MT in 2030. Productivity will also show gradual improvement: it increases from 0.60 MT/ha in 2026 to 0.63 MT/ha in 2030. These trends have optimistic implications for black gram cultivation in the future, as farming practices and effective use of resources will grow in the state.

Table 2: Descriptive Statistics and Trend Analysis for Black Gram

Parameter	Mean	Median	SD	Skewness	Kurtosis	MKT	Sen's Slope (Q)
Area	443000	430000	72500	0.42	2.10	6.85 (0.0001)	4500
Production (mt)	236000	220000	44800	0.58	2.45	7.24 (0.0001)	6700
Productivity (mt/ha)	0.53	0.52	0.08	0.11	1.89	5.98 (0.0001)	0.007

Table 2 shows descriptive statistics and trend analysis for black gram cultivation in Andhra Pradesh, focusing on key characteristics like area, production, and productivity. The average cultivated area of 443,000 hectares with a moderate skewness of 0.42, indicating a small right-tail distribution, and a Sens slope of 4,500 suggests a constant increasing trend. Production averages 236,000 metric tonnes, with a higher skewness of 0.58 and Sens slope of 6,700, indicating a strong upward trend. Productivity has a mean of 0.53 mt/ha and a low skewness of 0.11, indicating a statistically significant positive trend (mann-kendall test value of 5.98 and Sen's slope of 0.007). These results indicate a consistent and positive growth in all three parameters over time. The increasing patterns in area, production, and

productivity indicate that historical agricultural data can be effectively used to anticipate black gram productivity in the region, offering useful insights for planning and policy creation.

Table 3 confirms the efficacy of the GPRF model in predicting black gram area, production, and productivity in Andhra Pradesh for the period 2018 to 2022. The model's predictive ability was highly satisfactory, with low MAPE values of 1.33% (area), 2.04% (production), and 0.73% (productivity). The relative errors for area ranged from -1.82% to 1.35%, which is a reminder of how little the predicted values deviated from the actuals. The relative errors for production ranged slightly more widely than -2.64% to 2.51%. The relative estimates of productivity were reasonably stable across the time-series period of projection;

in fact, for most years the percentage error values were zero or close to that. Overall, the adjusted values confirm that the model did accurately reflect historical variations or trends in black gram growing during the chosen representative site. The data from 2018 to 2022 were selected for validation because they represented the last five years of complete and reliable historical data

existing just prior to the forecasting horizon. Validating the model on this recent dataset ensures that it reflects current agricultural patterns and conditions. Consequently, the forecasts for 2026 to 2030 are grounded in recent trends, enhancing their credibility and relevance for future agricultural planning and policy decisions.

Table 3: Validation of GPRF Model for Forecasting Black Gram

Parameter	Year	Actual Value	Predicted Value	Relative Error	MAPE (%)
Area (Million hectares)	2018	440000	432000	-1.82%	1.33%
	2019	445000	449200	0.94%	
	2020	452000	458100	1.35%	
	2021	459000	464500	1.20%	
	2022	462000	455800	-1.34%	
Production (Million tons)	2018	235000	228800	-2.64%	2.04%
	2019	239000	243000	1.46%	
	2020	245000	248900	1.59%	
	2021	251000	257300	2.51%	
	2022	258000	263100	1.98%	
Productivity (mt/ha)	2018	0.53	0.53	0.00%	0.73%
	2019	0.54	0.54	0.00%	
	2020	0.54	0.55	1.85%	
	2021	0.55	0.56	1.82%	
	2022	0.56	0.56	0.00%	

Performance Measures

Performance of proposed technique is analysed with metrics utilizing MAE, MAPE, and RMSE.

Mean Absolute Error (MAE)

Mean Absolute Error (MAE) evaluates prediction error by focusing on the median of absolute errors, making it more resistant to outliers and non-normally distributed data, shown by Equation [7].

$$MAE_{forecast} = median(|y_1 - \hat{y}_1| \dots |y_i - \hat{y}_i| \dots |y_n - \hat{y}_n|) \quad [7]$$

Where, n indicated as count of data points, y_i actual value for i^{th} data point, $\hat{y}_i$ signifies anticipated value for i^{th} data point.

Mean Absolute Percentage Error (MAPE)

MAPE, a commonly used metric in time series analysis, measures the accuracy of forecasts. The average absolute percentage difference between the expected and actual values is provided. A percentage is frequently used to express MAPE. The MAPE is given in Equation [8].

$$MAPE = mean\left(\frac{|SIC_p - SIC_g|}{SIC_g}\right) \quad [8]$$

Where SIC_p denotes the predicted value, SIC_g denotes the actual value.

Root Mean Square Error (RMSE)

The RMSE is a frequently employed statistic to evaluate a model's mistake in projecting quantitative data. It is computed as the mean of the squared discrepancies between the measured and projected values. RMSE assesses the magnitude of prediction mistakes and is easy to read since it is expressed in the same units as the target variable. The formula for RMSE is given in Equation [9].

$$RMSE_{forecast} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y})^2} \quad [9]$$

Where, N represents quantity of data points, y_i real value for $i=1$ data point, $\hat{y}$ signifies anticipated value for i^{th} data point.

Performance Analysis

The proposed GPRF model significantly outperforms the existing methods, achieving the lowest errors across all metrics with MAE of 0.028, MAPE of 0.015, and RMSE of 0.021, indicating high prediction accuracy. In comparison, SVM shows

MAE of 0.085, MAPE of 0.088, and RMSE of 0.087, while SVR records 0.090, 0.091, and 0.082, and KNN results in 0.096, 0.097, and 0.087, respectively. These findings clearly demonstrate the superior performance and reliability of the proposed GPRF method in accurately forecasting black gram agricultural parameters in the region. Table 4 presents the performance analysis of various forecasting methods SVR, KNN, SVM, and

the proposed GPRF using evaluation metrics such as MAE, MAPE, and RMSE for forecasting the area, production, and productivity of Black Gram in Andhra Pradesh. The simulation outputs of the GPRF technique are shown in Figures 6 to 8 and Table 1. The proposed GPRF method is compared with traditional machine learning methods, specifically SVR, KNN, and SVM, to evaluate its forecasting performance.

Table 4: Performance Analysis of Proposed and Existing Methods

Methods	MAE	MAPE	RMSE
SVR	0.090	0.091	0.082
KNN	0.096	0.097	0.087
SVM	0.085	0.088	0.085
GPRF (Proposed)	0.028	0.015	0.021

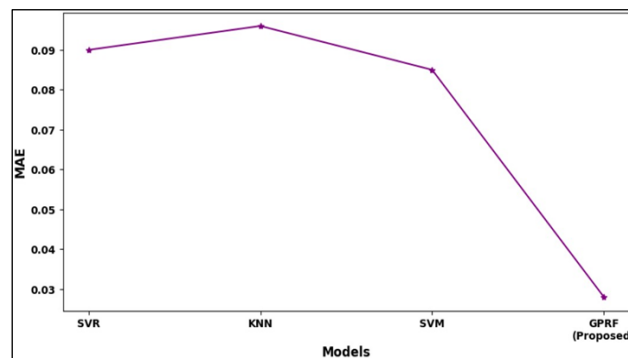


Figure 6: Performance Analysis of MAE

Figure 6 presents the performance analysis of various forecasting models based on MAE for predicting black gram productivity in Andhra Pradesh using historical agricultural data. Among the compared models, KNN exhibited the highest MAE of approximately 0.096, followed by SVR with an MAE of around 0.090. The SVM model showed improved accuracy with an MAE of about 0.085. However, the proposed GPRF model significantly

outperformed the others, achieving the lowest MAE of approximately 0.028, indicating a substantial improvement in forecasting precision. This comparison clearly demonstrates that the GPRF method is highly effective for productivity prediction tasks, showcasing superior accuracy and robustness over traditional and hybrid machine learning approaches.

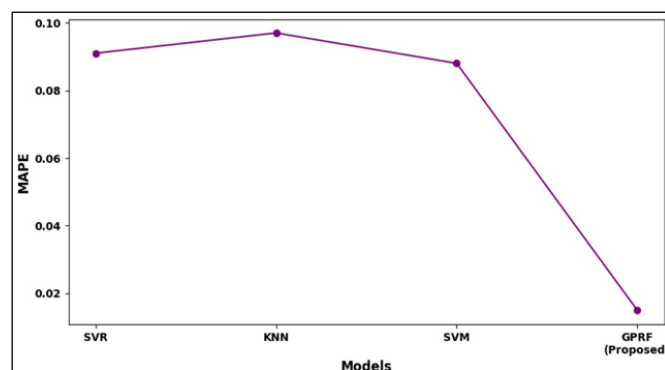


Figure 7: Performance Analysis of MAPE

Figure 7 illustrates the performance comparison of different forecasting models based on MAPE for

predicting black gram productivity in Andhra Pradesh using historical agricultural data. The

KNN model had the highest MAPE with approximately 0.097, only slightly higher than the SVR models MAPE value of approximately 0.091. The SVM model improved slightly with an MAPE of approximately 0.088. The proposed GPRF model showed the best prediction performance by achieving the lowest MAPE of approximately

0.015, representing a considerable reduction in prediction error. Overall, the results demonstrate the superior forecasting capability of the proposed GPRF model over traditional and deep learning-based methods and represent a highly effective method for precisely forecasting productivity in Precision Agriculture.

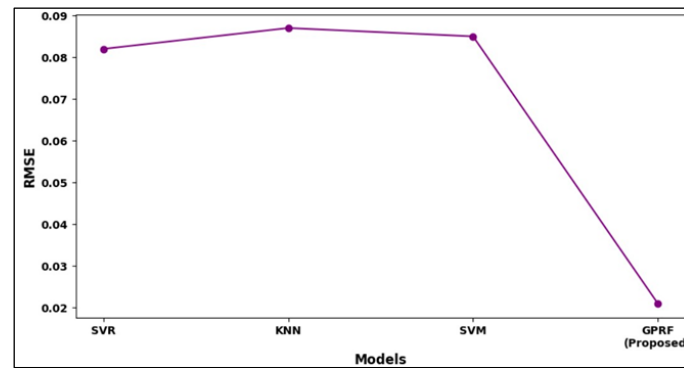


Figure 8: Performance Analysis of RMSE

Figure 8 presents the RMSE analysis of different forecasting methods used for predicting black gram productivity in Andhra Pradesh based on historical agricultural data. Among the evaluated models, SVM recorded the highest RMSE of approximately 0.099, indicating larger prediction errors, followed by KNN and SVR with RMSE values of about 0.087 and 0.082, respectively. The proposed GPRF model achieved the lowest RMSE of around 0.021, demonstrating a substantial reduction in prediction error. This clearly shows that GPRF is the most accurate and reliable model for productivity forecasting, outperforming both traditional and deep learning methods in minimizing deviation between predicted and actual values.

Discussion

The proposed GPRF forecasting model makes substantial advances in agricultural prediction by integrating GPR and RF to accurately estimate black gram area, production, and productivity under dynamic agro-climatic and market variables in Andhra Pradesh. The input data is gathered from Indiatat, which provides information specifically for Andhra Pradesh black gram, offering strong secondary data that spans over five decades (1981-2025). The complete dataset is provided as annual statistics on key agricultural metrics such as area farmed (hectares), production (tonnes), and productivity (kg/ha). This data is gathered from genuine government records and official

documents, presented in an annual format similar to the previously mentioned Indiatat dataset, and it is dependable and consistent enough for long-term agricultural studies. Unlike traditional models, which rely primarily on historical data and frequently ignore recent environmental and policy changes, the combination of GPR for pre-processing and RF for prediction enables the model to capture complicated, nonlinear interactions within the data. Autoregressive integrated moving average model (ARIMA), powered by a genetic algorithm achieved improved crop yield predictions based on evaluation metrics, namely MAE (0.80%), RMSE (3.75%), and MSE (0.07%) (27). ARIMA models are applied to both the original time series data and the first-differenced data to assess stationarity. The Box-jenkins autoregressive integrated moving average (ARIMA) model for Black gram yield shows a minimum absolute percentage error of 19.99% and a maximum of 43.29% (28). The results of the suggested GPRF model, in relation to the results of other models showed outperformed other models with lowest MAE of 0.028, RMSE of 0.021 and MAPE of 0.015. The proposed GPRF model outperformed with other models KNN (MAE 0.096, RMSE 0.087, and MAPE 0.097), SVR (MAE 0.090, RMSE 0.082, MAPE 0.091), and SVM (MAE 0.085, RMSE 0.099, MAPE 0.088), indicating higher forecasting accuracy. The GPRF model also had the shortest computation time, at 75 seconds,

compared to 190 seconds for KNN, 146 seconds for SVM, and 118 seconds for SVR, demonstrating its effectiveness in actual-time forecasting. Forecasts generated only for 2026-2030: area from 468,500 hectares to 490,100 hectares, production from 268,000 metric tonnes to 286,900 metric tonnes, and productivity from 0.57 to 0.60 metric tonnes per hectare, with narrow 95 percent prediction intervals indicating high model confidence. These findings confirm that the GPRF model enhances enabling policymakers and farmers to make data-driven decisions regarding resource allocation, market planning, crop insurance schemes, making it ideal for strategic agricultural planning and policymaking in Andhra Pradesh.

Conclusion

In this section, a hybrid ML GPRF model forecasts black gram, area, production and productivity in Andhra Pradesh using historical data was successfully implemented in Python. Across diverse evaluation metrics, the proposed method consistently demonstrates significant improvements in MAE, MAPE, and RMSE. The proposed GPRF model demonstrates superior performance in forecasting the area, production, and productivity of black gram in Andhra Pradesh using historical agricultural data. The model achieved significantly lower error values MAE of 0.028, MAPE of 0.015, and RMSE of 0.021 outperforming benchmark models such as SVR, KNN, and SVM. The GPRF model also maintained high prediction accuracy for the years 2018-2022, with MAPE values of 1.33% for area, 2.04% for production, and 0.73% for productivity. Forecasts for 2023-2027 show a consistent upward trend: the cultivated area is expected to increase from 468,500 ha to 490,100 ha, production from 268,000 mt to 286,900 mt, and productivity from 0.57 to 0.60 mt/ha, all within narrow and reliable prediction intervals. Furthermore, the model required only 75 seconds of computation time, significantly faster than other approaches. These findings highlight the GPRF models effectiveness and efficiency in providing accurate and timely forecasts, making it a valuable tool for strategic agricultural planning and policy development across the region. The proposed method has some limitations, including reliance on historical data and the exclusion of key factors like soil quality, rainfall and policy impacts. It also lacks

uncertainty estimation, which limits its practical reliability. In future, the model can be improved by integrating real-time climate, satellite data and hyperparameter tuning techniques by applying dynamic learning methods, and incorporating national-level, district-level analysis and uncertainty modelling for better decision-making.

Abbreviations

AEGRU: Autoencoder Gated Recurrent Unit, ARIMA: Autoregressive Integrated Moving Average, GARCH: Generalized Autoregressive Conditional Heteroskedasticity, GPRF: Gaussian Process Random Forest, GPR: Gaussian Process Regression, KNN: K-Nearest Neighbors, PCA: Principal Component Analysis, SARIMA: Seasonal Autoregressive Integrated Moving Average, SVR: Support Vector Regression, SVM: Support Vector Machine.

Acknowledgement

I would like to express my sincere gratitude to International Research Journal of Multidisciplinary Scope for considering my research paper titled "Prediction of Area, Production and Productivity of Black Gram using Machine Learning" for publication. I appreciate the valuable feedback and guidance provided by the editorial team and reviewers, which will contribute to enhancing the quality of my work.

Author Contributions

Each of the mentioned authors has approved the work and contributed significantly, directly, and intellectually.

Conflict of Interest

The authors declare that there are no conflicts of interest related to this research work. No financial, personal, or professional relationships have influenced the findings, analysis, or conclusions presented in this study.

Declaration of Artificial Intelligence (AI) Assistance

The authors declare no use of Artificial Intelligence AI for the write-up of this manuscript.

Ethics Approval

This study was conducted in accordance with the ethical guidelines and principles.

Funding

This research received no external funding.

References

1. Krithi MR, Varman PM. Is Preference for Horticulture Crops Growing Over the Years? An Analysis of the Trend in Consumption, Production and Area under Horticulture Cultivation in India. Responsible Production and Consumption. 2024; 1st Edition, CRC Press: 66-76.
<https://www.taylorfrancis.com/chapters/edit/10.1201/9781003542506-9/>
2. Ali M, Deo RC, Downs NJ, Maraseni T. Cotton yield prediction with Markov Chain Monte Carlo-based simulation model integrated with genetic programming algorithm: A new hybrid copula-driven approach. Agricultural and forest meteorology. 2018; 263: 428-448.
<https://doi.org/10.1016/j.agrformet.2018.09.002>
3. Feng P, Wang B, Li Liu D, Waters C, Xiao D, Shi L, Yu Q. Dynamic wheat yield forecasts are improved by a hybrid approach using a biophysical model and machine learning technique. Agricultural and Forest Meteorology. 2020; 285:107922-107934.
<https://doi.org/10.1016/j.agrformet.2020.107922>
4. Verma R, Kushwaha KP, Bijlwan A, Bisht AS. Enhancing urad bean (*Vigna mungo* L.) crop management with machine learning: Predictive analysis of pod rot severity and pod bug incidence patterns. Australasian Plant Pathology. 2024; 53(3):273-283.
<https://doi.org/10.1007/s1331C3-024-00967-7>
5. Devi M, Kumar J, Malik DP, Mishra P. Forecasting of wheat production in Haryana using hybrid time series model. Journal of Agriculture and Food Research. 2021; 5(1):100175-100180.
<https://doi.org/10.1016/j.jafr.2021.100175>
6. Subbulakshmi S. Effect of early sowing on growth and yield of blackgram varieties under rainfed condition. International Journal of Agricultural Sciences. 2022; 18(1): 489-495.
<https://doi.org/10.15740/HAS/IJAS/18.1/489-495>
7. Kumar BV, Rao PG. An effective hybrid attention model for crop yield prediction using IoT-based three-phase prediction with an improved sailfish optimizer. Expert Systems with Applications. 2024; 255(1): 124740-124752.
<https://doi.org/10.1016/j.eswa.2024.124740>
8. Basu PS, Singh UM, Kumar AN, Praharaj CS, Shivran RK. Climate change and its mitigation strategies in pulses production. Indian journal of Agronomy. 2016; 61(4):71-82.
<https://www.researchgate.net/publication/311717987>
9. Sridhara S, Manoj KN, Gopakkali P, Kashyap GR, Das B, Singh KK, Srivastava AK. Evaluation of machine learning approaches for prediction of pigeon pea yield based on weather parameters in India. International Journal of Biometeorology. 2023; 67(1):165-180.
<https://doi.org/10.1007/s00484-022-02396-x>
10. Nikhil UV, Pandiyan AM, Raja SP, Stamenkovic Z. Machine learning-based crop yield prediction in south india: performance analysis of various models. Computers. 2024; 13(6):137-165.
<https://doi.org/10.3390/computers13060137>
11. Gosai D, Raval C, Nayak R, Jayswal H, Patel A. Crop recommendation system using machine learning. International Journal of Scientific Research in Computer Science, Engineering and Information Technology. 2021; 7(3): 558-569.
<https://doi.org/10.32628/CSEIT2173129>
12. Choudhury SS, Pandharbale PB, Mohanty SN, Jagdev AK. An acquisition based optimised crop recommendation system with machine learning algorithm. EAI Endorsed Transactions on Scalable Information Systems. 2024; 11(1): 411-418.
<https://doi.org/10.4108/eetis.4003>
13. Bhattacharya B, Chowdary VM, Das A, Kumar M, Polaju S, Kumari M, Chakraborty A, Halder D, Maity S. Inventory and mapping of kharif crops using machine learning with EOS-04 time-series SAR data. Current Science. 2024; 126(9):1050-1060.
<https://doi.org/10.18520/cs/v126/i9/1050-1060>
14. Apat SK, Mishra J, Raju KS, Padhy N. An artificial intelligence-based crop recommendation system using machine learning. Journal of Scientific and Industrial Research (JSIR). 2023; 82(5): 558-567.
<https://doi.org/10.56042/jsir.v82i05.1092>
15. Swarnalatha P, Rao VS, Reddy GR, Rathod S, Ramesh D, Devi KU. Application of Machine Learning Techniques Models for Forecasting of Redgram Prices of Andhra Pradesh, India. Journal of Scientific Research and Reports. 2024; 30(7): 252-271.
<https://doi.org/10.9734/jsrr/2024/v30i72142>
16. Charishma PN, Lakshmi SR, Durga MV. Making crop recommendations using machine learning techniques. Journal of Computer Allied Intelligence. 2024; 2(2):1-12.
<https://doi.org/10.69996/jcai.2024006>
17. Seyedmohammadi J, Zeinadini A, Navidi MN, McDowell RW. A new robust hybrid model based on support vector machine and firefly meta-heuristic algorithm to predict pistachio yields and select effective soil variables. Ecological Informatics. 2023; 74: 102002-102015.
<https://doi.org/10.1016/j.ecoinf.2023.102002>
18. Boppudi S, Jayachandran S. Improved feature ranking fusion process with Hybrid model for crop yield prediction. Biomedical Signal Processing and Control. 2024; 93:106121-106133.
<https://doi.org/10.1016/j.bspc.2024.106121>
19. Reyana A, Kautish S, Karthik PS, Al-Baltah IA, Jasser MB, Mohamed AW. Accelerating crop yield: multisensor data fusion and machine learning for agriculture text classification. IEEE Access. 2023; 11:2 20795 - 20805.
<https://doi.org/10.1109/ACCESS.2023.3249205>
20. Mohandu A, Mohan K, "Long-Term Traffic Flow Prediction using Hybrid Deep Learning Technique," International Journal of Engineering Trends and Technology. 2023; 71(5):156-165.
<https://doi.org/10.14445/22315381/IJETT-V71I5P216>
21. Chitikela G, Rathod S, Vijayakumar S. Change point-driven interrupted time series and machine learning models for forecasting Indian food grain production. Discover Food. 2025; 5(1):1-7.

- <https://doi.org/10.1007/s44187-025-00350-5>
22. Singh S, Guleria A, Sheoran P, Sood P, Mahajan V, Kumar D, Singh S, Singh G, Rana RK. Price transmission and market integration analysis of black gram (urad dal) across major Indian states. *Frontiers in Sustainable Food Systems*. 2025; 9: 1589518-1589533. <https://doi.org/10.3389/fsufs.2025.1589518>
 23. Pavithra N, Jayalalitha K, Sujatha T, Harisatyanarayana N, Lakshmi NJ, Roja V. Evaluation of Blackgram (*Vigna mungo* L.) Genotypes for Dry Matter Partitioning, Reproductive Efficiency and Yield under High Temperature Stress. *Legume Research-An International Journal*. 2025; 48(1):75-85. <https://doi.org/10.18805/LR-5287>
 24. Indiastat.com [Internet]. New Delhi: Datanet India Private Limited, ITeS Company,; c2000 [updated 2025 April 14; cited 2025 Jun 10]. Available from: <https://www.indiastat.com/data/agriculture>
 25. Vinodha K, Gopi ES, Agnibhoj T. LiDAR-based estimation of bounding box coordinates using Gaussian process regression and particle swarm optimization. *Biomimetic Intelligence and Robotics*. 2024; 4(1):100140-100153. <https://doi.org/10.1016/j.birob.2023.100140>
 26. Panda SK, Mohanty SN. Time series forecasting and modelling of food demand supply chain based on regressors analysis. *IEEE Access*. 2023; 11:42679-42700. <https://doi.org/10.1109/ACCESS.2023.3266275>
 27. Guo Y. Integrating genetic algorithm with ARIMA and reinforced random forest models to improve agriculture economy and yield forecasting. *Soft Computing*. 2024; 28(2):1685-706. <https://doi.org/10.1007/s00500-023-09516-8>
 28. Mahapatra SK, Dash A. ARIMA Model for Forecasting of Black Gram Productivity in Odisha. *Journal of Asiatic Society for Social Science Research*. 2020; 2(1):131-6. <https://www.asssr.in/index.php/home/article/view/105>

How to Cite: Gowri P, Kutty N, Pandiyathuray A. Prediction of Area, Production and Productivity of Black Gram in Andhra Pradesh using Machine Learning. *Int Res J Multidiscip Scope*. 2025; 6(4):1336-1350. doi: 10.47857/irjms.2025.v06i04.07183