Journal of Innovative Agriculture

Research Article

Machine learning-based predictive modeling of banana crop yield: a comparative analysis

Anandhi Santhosh[1*], Sujatha Prabhakaran[2]

[1]Department of Agricultural Economics, Anbil Dharmalingam Agricultural College and Research Institute, Tamil Nadu Agricultural University, Tiruchirappalli 620027, Tamil Nadu, India; anandhi.s@tnau.ac.in.
[2]Dr. M.S. Swaminathan Agricultural College and Research Institute, Tamil Nadu Agricultural University, Thanjavur 614 902, Tamil Nadu, India; sujathap@tnau.ac.in.

[*]**Correspondence**
Anandhi Santhosh
anandhi.s@tnau.ac.in

Volume: 12, Issue: 3, Pages: 1-9
DOI: https://doi.org/10.37446/jinagri/rsa/12.3.2025.1-9
Received: 4 April 2025 / Accepted: 1 August 2025 / Published online: 01 September 2025

Background: Agriculture is a primary livelihood provider in India, sustaining over 58% of rural households, with banana ranking as the country's second most significant fruit crop after mango. Banana cultivation spans 3.8 million hectares across 122 countries, with India contributing approximately 25.7% to global production. Crop yield prediction using machine learning techniques can optimize field operations and support pre-harvest planning decisions for farmers.

Methods: The study evaluated machine learning models for predicting banana crop yields across 31 districts of Tamil Nadu, India. Historical yield data from 2011-2018 were collected from governmental sources, with rainfall data from 2016-2018. After data preparation and pre-processing, three regression techniques, Multiple Linear Regression, Random Forest Regression, and Polynomial Regression, were implemented and compared. Multiple Linear Regression was selected to establish baseline linear relationships between cultivation parameters and yield outcomes, providing interpretable coefficients for agricultural decision-making. Random Forest Regression was chosen for its superior ability to capture complex non-linear interactions between multiple agricultural variables and handle real-world data inconsistencies in datasets. Polynomial Regression was utilized to examine non-linear relationships in the data, specifically curved patterns between cultivation area size and yield performance. The models were trained on key agricultural parameters, including cultivation area, productivity metrics, and rainfall patterns.

Results: Analysis revealed a weak negative correlation between cultivation area and productivity, with smaller areas (under 6000 Ha) achieving some of the highest productivity levels (70-90 tonnes / ha). Rainfall showed minimal impact on productivity, suggesting effective irrigation systems and water management practices in the region. The Random Forest model demonstrated superior performance with a 36% lower Root Mean Square value compared to other models. Polynomial Regression proved less effective due to data nonlinearity, while Multiple Linear Regression provided straightforward predictions but with lower accuracy.

Conclusion: The study confirms that Random Forest Regression is the most effective machine learning technique for banana yield prediction in Tamil Nadu's agricultural context. The findings suggest that successful banana cultivation in the region relies more on intensive farming practices in smaller areas rather than extensive cultivation.

Keywords: *crop yield, machine learning, regression, banana, model*

Introduction

Agriculture serves as the cornerstone of India's economic framework, sustaining over 58% of rural households as their primary source of livelihood. Among global fruit commodities, banana (*Musa* sp.) stands as a paramount crop, ranking as India's second most significant fruit crop after mango (Balraj et al., 2021). The crop's year-round availability, economic accessibility, diverse varieties, nutritional profile, and therapeutic properties have established its universal appeal across socioeconomic strata while maintaining substantial export potential (Singh et al., 2016). Banana cultivation, dating back to ancient civilizations in Southeast Asia and Indonesia, has evolved into a globally significant agricultural commodity. With a cultivated expanse of 3.8 million hectares across 122 countries, the banana ranks as the fifth most valuable agricultural food crop worldwide (FAO, 2020). India's prominence in global banana production is noteworthy, contributing approximately 25.7% to the global output, while the Asia-Pacific region commands 61% of global consumption (Tripathi et al., 2019). The crop's nutritional significance is underscored by its rich composition of essential nutrients, providing 67 calories per 100g and meeting 23% of daily potassium requirements, alongside substantial calcium, phosphorus, and nitrogen content (Sharma et al., 2020). Beyond its conventional use as a fruit, banana plants offer versatile applications from unripe fruit to inflorescence, leaves, stem, and rhizome serving both culinary and animal feed purposes. Recent research has highlighted banana's pharmaceutical potential, attributing it to abundant bioactive compounds including carotenoids, flavonoids, phenolics, amines, phytosterols, and vitamins, which exhibit significant antioxidant properties (Kumari et al., 2023). In contemporary agriculture, modeling techniques have become instrumental in crop growth and yield prediction, facilitating precise decision-making for current and future cultivation strategies (Liakos et al., 2018). Crop yield, a complex trait influenced by genotype, environmental factors, and their interactions, requires sophisticated prediction models. Understanding these intricate relationships demands comprehensive datasets and robust algorithmic approaches (van Klompenburg et al., 2020). The emergence of machine learning, a branch of Artificial Intelligence (AI), offers promising solutions for yield prediction by analyzing multiple variables simultaneously. This approach becomes particularly valuable with the increasing availability of agricultural data (Chlingaryan et al., 2018). By integrating agronomic principles with machine learning algorithms, researchers have developed sophisticated models for large-scale crop yield forecasting. These models, trained on historical datasets, have demonstrated significant potential in predicting yields based on various environmental and managerial parameters (Crane-Droesch, 2018). The present study focuses on developing and implementing machine learning algorithms for banana yield prediction, with particular emphasis on weather parameters and their correlation with crop productivity. This research aims to bridge the existing gap in tropical perennial crop modeling, specifically addressing the unique challenges in banana cultivation. Machine learning, with its predictive power, has shown promise in providing accurate yield forecasts by modeling complex relationships among various environmental and management factors. The study (Venugopal et al., 2020) uses Machine Learning Classification Models to forecast crop output based on temperature, rainfall, and area. The research study demonstrated the effectiveness of classification models, comparing Logistic Regression (87.8% accuracy), Naïve Bayes (91.58%), and Random Forest Classifier (92.81%). Their research highlighted Random Forest's superior performance due to its bagging technique, achieving 91.34% accuracy on test data. A comprehensive evaluation was conducted incorporating contradictory rainfall and temperature data spanning 100 years (Nigam et al., 2019). Their research compared time series models, Long Short-Term Memory (LSTM) and Recurrent Neural Network (RNN), ensemble learning models, Random Forest and XGBoost, and traditional machine learning algorithms, K Nearest Neighbour and Logistic Regression. Their findings revealed that LSTM performed better for temperature prediction, while Simple RNN excelled in rainfall predictions. The Random Forest Classifier achieved 67.80% accuracy in crop identification. Reddy & Kumar (2021) comparative analysis of various prediction models highlighted some limitations of existing approaches. They noted that machine learning algorithms can be complex and expensive, while Linear Regression's inability to handle non-linear data affects its reliability. They recommended reducing input components for Artificial Neural Network (ANN) training and emphasized the importance of comprehensive soil datasets for improved accuracy (Reddy & Kumar, 2021). The study which focused on automating agricultural activities through various regression techniques, found that Support Vector Machine Regression outperformed other models with an R-squared value of 0.968 and RMSE of 5.48, surpassing both Multivariate Polynomial Regression and Random Forest Regression (Shah et al., 2018). A detailed investigation was done on various environmental factors' impact on crop yield using Linear Regression, focusing specifically on rice cultivation. Their analysis, spanning 1990-2000, achieved an R^2 value of 0.72 and established clear correlations between Annual Rainfall, Area Under Cultivation, and Food Price Index with crop yield (Sellam & Poovammal, 2016). Several studies (Kumar et al., 2020, Sharma et al., 2020) have identified Random Forest as a particularly effective algorithm for crop yield prediction. Mishra et al. (2016) conducted a comprehensive review of machine learning applications in agricultural crop production, highlighting the need for objective statistical approaches over traditional qualitative assessments. Their study evaluated emerging techniques including artificial neural networks, decision trees, regression analysis, and various clustering methods to extract meaningful patterns from large agricultural datasets. The authors emphasized that these techniques enable significant relationships to be identified among variables, ultimately supporting more accurate and timely crop production forecasting essential for policy decisions. These studies highlight the utility of machine learning for crop yield prediction, with Random Forest, Linear Regression, and Support Vector Machines proving particularly effective in handling complex environmental datasets. The study suggests that while Random Forest generally performs well for crop yield prediction, the optimal choice of algorithm depends on specific use cases and data characteristics. Recent advancements in agricultural technology have focused on integrating IoT systems with machine learning for real-time soil analysis and crop recommendations. An Agricultural IoT system achieving 98% accuracy using a Decision Tree-AdaBoost combination for soil nutrient analysis and crop selection guidance (Cheema & Pires, 2025). Similarly, an ML-enabled IoT device utilizing various sensors to collect real-time soil and environmental data, enhancing crop productivity through customized recommendations was created (Islam et al., 2023). Extending this approach, a hybrid system combining static data with real-time IoT measurements to suggest optimal crops for maximizing yield, providing farmers with an efficient and accurate low-cost solution was proposed (Ramzan et al., 2024). Machine learning models for crop yield prediction have grown increasingly sophisticated, with several researchers developing advanced frameworks to address complex agricultural challenges. The deep neural networks for predicting Australian crop yields while accounting for climate change impacts, demonstrating 23-40% lower mean absolute error compared to benchmark methods was introduced by (Demirhan, 2025). For large-scale applications, the research study developed a national soybean yield forecasting system utilizing multiple ML approaches, finding that ANN and ensemble models achieved the best performance with approximately 6% rRMSE three months before harvest (von Bloh et al., 2023). Patrick et al. (2023) specifically addressed banana crop forecasting in Tanzania through time series and ensemble models, emphasizing the importance of temporal dynamics in agricultural predictions (Patrick et al., 2023). Motamedi & Villányi (2024) introduced a Bayesian optimized ensemble decision tree model for crop recommendation, achieving exceptional accuracy rates of 99.5%. Machine learning techniques have shown significant potential in revolutionizing agricultural crop prediction and management. The comparative analyses conducted by various researchers demonstrate that algorithms such as Random Forest, Support Vector Machine Regression, and LSTM neural networks can achieve impressive prediction accuracies when properly implemented.

This paper explores supervised learning approaches for predicting crop yields, focusing on the development of functions that effectively map input variables to desired outputs. This study makes three key contributions to agricultural machine learning methods by comparing three regression techniques for banana yield prediction across Tamil Nadu districts. The research demonstrates that smaller, intensively managed farms achieve higher productivity than larger extensive operations. Additionally, it validates Random Forest as an effective tool for practical agricultural decision-making using real data from 31 districts over multiple years. The authors identify promising avenues for future research, including the integration of additional economic parameters such as Minimum Support Price and development of more sophisticated techniques for addressing non-linear relationships within agricultural datasets. These advancements could significantly enhance prediction capabilities and provide more comprehensive support for agricultural decision-making processes.

Materials and Methods

Tamil Nadu, ranking as India's eleventh largest state by geographical area, presents a diverse agricultural landscape strategically divided into seven distinct agro-climatic zones.

Data sources and parameters

This research utilizes comprehensive agricultural datasets sourced from the official Tamil Nadu government data portaltn.data.gov.in. The analysis examines banana cultivation across all 31 districts of Tamil Nadu using a comprehensive dataset. It incorporates seven years of agricultural metrics, including cultivation area distribution, production volumes at district and aggregate levels, and productivity measurements from 2011-2018. The study also integrates three years of rainfall data from 2016-2018 to assess potential relationships between precipitation patterns and cultivation outcomes throughout the state's diverse agricultural regions. This methodical data collection framework enables a robust examination of banana cultivation trends, regional performance variations, and the potential correlation between climatic conditions and agricultural outcomes across Tamil Nadu's diverse agro-climatic regions.

Data preparation

Data preparation establishes the foundation for effective machine learning implementation. This initial phase involves systematically collecting and organizing data in appropriate storage structures while conducting a comprehensive assessment of its characteristics. Analysts thoroughly examine data sources, formats (structured, unstructured, semi-structured), distribution patterns, and inherent quality issues. This evaluation provides critical insights into data completeness, consistency, and relevance for the specific machine learning application. Data preparation also involves initial exploration through statistical summaries and visualizations to understand underlying trends, outliers, and potential relationships between variables.

Data Pre-processing

Pre-processing transforms raw data into a refined format optimized for model training through multiple specialized techniques:

- Handling missing values through imputation methods based on mean, median, mode or more advanced algorithms
- Detecting and removing duplicate records to prevent training bias and overfitting
- Normalizing or standardizing numerical features to ensure equal consideration by the algorithm
- Encoding categorical variables using one-hot encoding, label encoding, or target encoding
- Feature selection and extraction to identify the most predictive variables
- Addressing class imbalance through resampling techniques when necessary
- Executing comprehensive data wrangling to create derived features and transform variables into formats compatible with the chosen algorithm.

Training phase

With properly processed data available, model development proceeds by selecting appropriate machine learning techniques based on the specific problem domain. Common approaches include linear regression for continuous predictions, classification algorithms for categorical outcomes, and clustering for unsupervised pattern discovery. During training, the selected algorithm iteratively processes the prepared dataset, adjusting internal parameters to minimize error metrics. This process involves hyperparameter tuning through techniques like grid search or random search to optimize model configuration. Cross-validation methodologies are implemented to ensure robust performance across different data subsets and prevent overfitting.

Testing and validation

The final evaluation phase rigorously assesses model performance using separate test data withheld during training. Multiple evaluation metrics appropriate to the problem type are calculated, including accuracy, precision, recall, and F1-score for classification problems; and RMSE, MAE, and R-squared for regression problems. Performance is benchmarked against predetermined business requirements and baseline models. This comprehensive validation confirms whether the model generalizes effectively to new data points or requires additional refinement through feature engineering, algorithm selection, or parameter adjustments. The testing phase may also include stress testing with edge cases and analysis of prediction confidence levels to fully understand model capabilities and limitations.

Prediction models

In this study, the regression models such as Multiple linear regression and Random Forest regression, and Polynomial regression are used for the prediction of Banana crop yield. The systematic sequence of steps involved in ML, from data collection to prediction and application, is clearly shown below.

Standard machine learning process flow

The overall workflow (Figure 1) and machine learning process (Figure 2) are explained below,

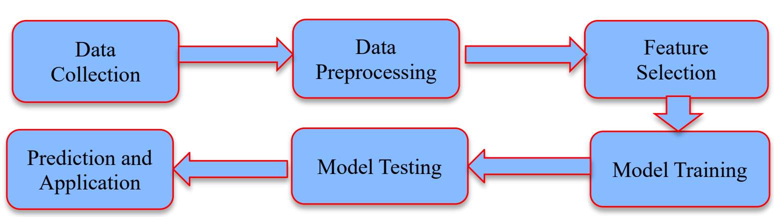

Figure 1. The flow chart shows the systematic sequence of steps involved in ML from data collection to prediction and application

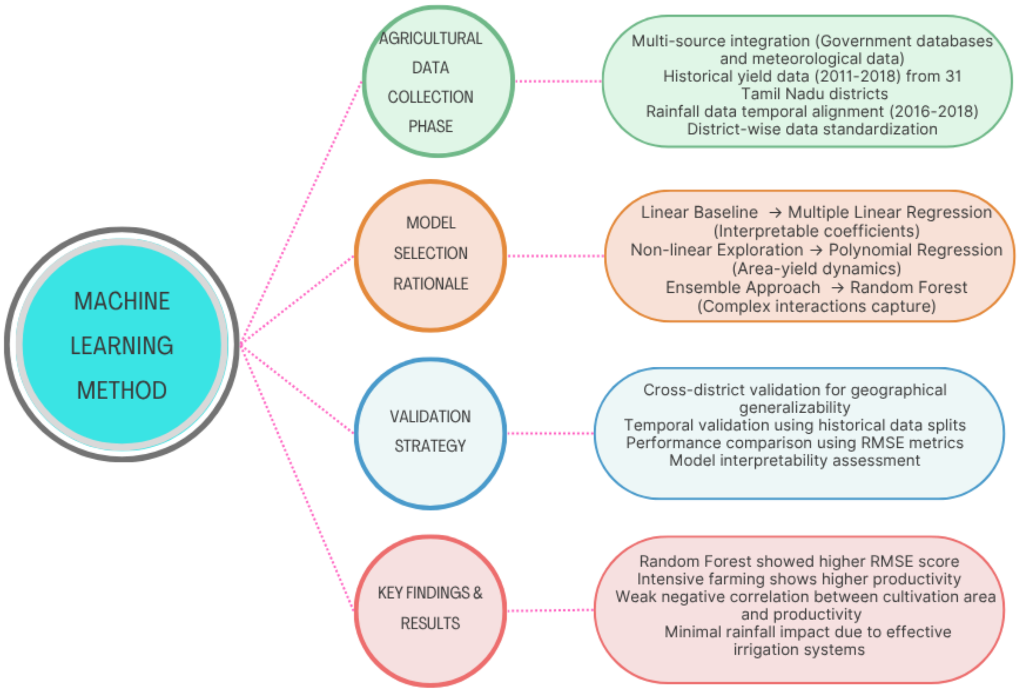

Figure 2. Machine learning process flow for the banana yield prediction

Multiple linear regression

Multiple Linear Regression (MLR) represents one of the fundamental techniques in statistical analysis and predictive modeling. While Simple Linear Regression utilizes a single independent variable to predict an outcome, MLR extends this principle by incorporating two or more independent variables to predict the value of a dependent variable with greater precision.

In quantitative research and data science applications, MLR serves as a powerful tool for analyzing relationships between multiple predictors and a continuous outcome variable. The technique works by fitting a linear equation to observed data, thereby modeling the relationship between the dependent variable and a set of independent variables.
The standard form of the Multiple Linear Regression model can be expressed as:

$y = b_0 + b_1x_1 + b_2x_2 + \ldots + b_nx_n + \varepsilon$, where y represents the dependent variable, $x_1, x_2, \ldots, x_n$ denote the independent variables, b_0 is the y-intercept, $b_1, b_2, \ldots, b_n$ are the regression coefficients that quantify the relationship between each independent variable and the dependent variable and ε represents the error term.

This mathematical framework allows researchers and analysts to assess how multiple factors simultaneously influence an outcome of interest, providing valuable insights for prediction, explanation, and decision-making across various domains.

Random forest regression

Random Forest is an ensemble learning technique that performs regression and classification tasks using multiple decision trees combined with bootstrap aggregation. This method leverages the collective output of numerous trees to determine final predictions, rather than relying on single decision trees. The basic idea behind this is often to mix multiple decision trees in determining the ultimate output instead of counting on individual decision trees. Generally, it has multiple decision trees as base learning models. Randomly perform row sampling and have sampling from the dataset, forming sample datasets for each model. The algorithm involved the following process:

1. Select random samples from the training dataset
2. Construct a decision tree for each sample
3. Obtain individual predictions from each tree
4. Select the most frequently predicted value as the final result

For implementation, the scikit-learn Python library provides the Ensemble Random Forest Regressor module specifically designed for regression applications.

Polynomial regression

Polynomial Regression models relationships between dependent and independent variables as nth degree polynomials. This approach is appropriate for non-linear datasets. The method transforms original features into polynomial features of specified degree before applying standard linear modeling techniques.

Technical requirements

Software: Jupyter Notebook
Operating System: Windows 10
Development Tools: Google ChromeWeb Browser
Python Libraries: NumPy, pandas, scikit-learn, matplotlib, seaborn

Results

Crop yield prediction plays a significant role in the Agricultural sector which can be performed using Machine learning algorithms. The aim of this study is to analyse how Machine learning techniques can be employed to develop an accurate crop yield prediction model for Banana in 31 Districts of Tamil Nadu. The study initially compared the area, productivity, and rainfall to know the relationship between the variables. Then the research study focused on the comparison of different prediction machine learning models

Comparison of model features

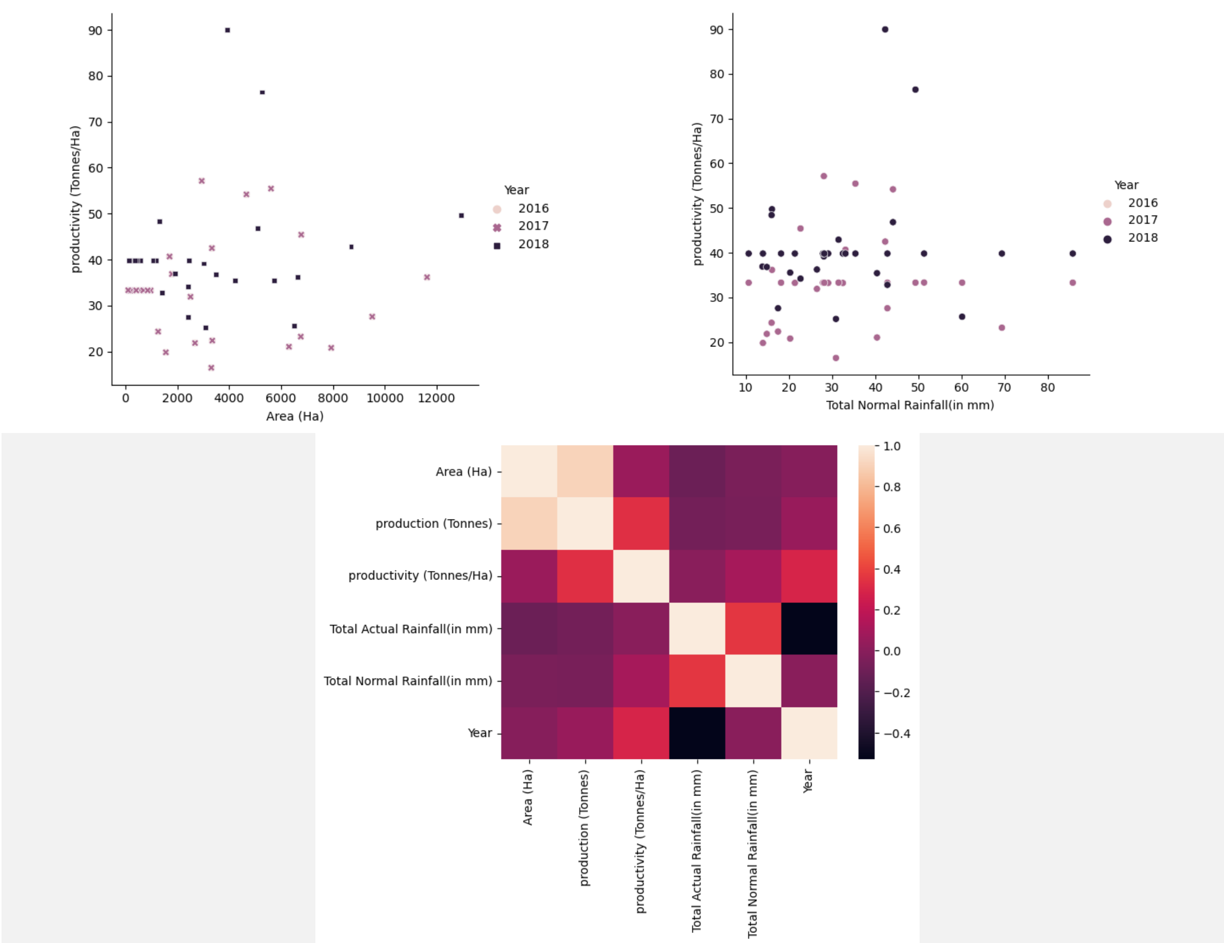

Figure 3. The figures show the comparison result between the productivity and the area, and rainfall

The analysis of banana crop productivity in Tamil Nadu districts from 2016-2018 reveals several interesting patterns in agricultural practices and outcomes. The data shows a weak negative correlation between cultivation area and productivity, with smaller areas under 6000 ha achieving some of the highest productivity levels of 70-90 tonnes/ha. Notably, rainfall appears to have minimal impact on productivity, with districts achieving high yields even under moderate rainfall conditions (30-50mm), indicating the effectiveness of existing irrigation infrastructure and water management systems across Tamil Nadu**.** The correlation matrix in Figure 3 indicates a strong relationship between area and production, while temporal analysis shows remarkable stability in productivity patterns across the three-year period. When comparing different prediction machine learning models, the Random Forest model outperformed other regression techniques in predicting banana yield. The model successfully utilized variables such as area, rainfall, and production to establish robust yield predictions and it showed superior performance with 36% lower Root Mean Square value compared to other models. Polynomial Regression was found less effective due to the model's sensitivity to data nonlinearity. Multiple Linear Regression model provided a straightforward predictive framework but lacked the accuracy of Random Forest.

Discussion

The findings from this study provide valuable insights into banana cultivation practices and the application of machine learning for crop yield prediction in Tamil Nadu. The weak negative correlation between cultivation area and productivity suggests that intensive farming in smaller areas may be more effective than extensive cultivation for banana crops in Tamil Nadu. This pattern indicates that factors beyond simple land allocation play crucial roles in determining crop productivity (Liakos et al., 2018). The minimal impact of rainfall on productivity suggests effective irrigation systems and water management practices are in place across the districts. This demonstrates the adaptation of farming techniques to local conditions, reducing dependence on natural rainfall patterns (Patrick et al., 2023). The stability in productivity patterns across the three-year period indicates consistent agricultural practices and resilience to yearly variations in environmental conditions. These findings collectively indicate that advanced machine learning techniques, particularly Random Forest, offer promising tools for developing accurate crop yield prediction models for banana cultivation in Tamil Nadu, potentially extending to other crops and regions with similar agricultural characteristics (Sharma et al., 2020). The model successfully utilized variables such as area, rainfall, and production to establish robust yield predictions. Polynomial Regression was found less effective due to the model's sensitivity to data nonlinearity, while MLR provided a straightforward predictive framework but lacked the accuracy of Random Forest (Reddy & Kumar, 2021). Random Forest method is considered better than MLR. This is a technique where algorithm create decision trees on different samples of the dataset and returns the average of the results. The salient finding of this study reveal that the Random Forest showed superior performance with 36% lower Root Mean Square value compared to other models. The Polynomial Regression proved less effective due to data nonlinearity and the Multiple Linear Regression provided straightforward predictions but with lower accuracy than Random Forest.

Conclusion

We presented a machine learning approach for banana crop yield prediction based on key agricultural parameters of 31 districts of Tamil Nadu. The approach used various Regression algorithms to make yield predictions including crop production based on area and total actual rainfall. Among the Regression Models Random Forest Regression is considered as the best model under this study as the Root mean square value obtained is lower (36%) when compared to other models. Its ability to successfully utilize multiple variables area, rainfall, and production, simultaneously offers a comprehensive prediction framework that mirrors real-world agricultural complexity**.** This multi-variable approach provides more robust predictions than single-factor models, reducing prediction errors and improving farm management decisions**.** The model's ensemble nature makes it less susceptible to data outliers and missing information, which is particularly valuable given the variability inherent in agricultural datasets**.** The polynomial model is not considered to be the fit for this data as the root mean square value is very low. Multiple Linear Regression provided a straightforward predictive framework and lacked the accuracy of Random Forest, demonstrating that agricultural yield prediction requires advanced algorithms capable of capturing complex variable interactions**.** Limitations included the need for more extensive data to improve prediction reliability. Future research should incorporate larger datasets and explore additional machine learning techniques to further improve prediction accuracy and reliability.

Acknowledgment

NIL.

Author contributions

S. Anandhi: Conceptualized the research methodology, designed the machine learning framework, performed data collection and analysis, developed and optimized the predictive models, and prepared the manuscript.
P. Sujatha: Edited the manuscript and provided formatting support of the research paper.

Funding

NIL.

Conflict of interest

The author declares no conflict of interest. The manuscript has not been submitted for publication in any other journal.

Ethics approval

Not applicable.

AI tool usage declaration

The authors declare that no AI and associated tools are used for writing scientific content in the article.

References

Balraj, G., Palpandian Preethi, M. G., & Selvaraj, R. (2021). Banana (*Musa* sp.) In book: Tropical fruit crop: Theory to Practical (pp. 43-114).

Cheema, S. M., & Pires, I. M. (2025). AIoT based soil nutrient analysis and recommendation system for crops using machine learning. *Smart Agricultural Technology, 11*, 100924.

Chlingaryan, A., Sukkarieh, S., & Whelan, B. (2018). Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. *Computers and Electronics in Agriculture*, *151*, 61-69. https://doi.org/10.1016/j.compag.2018.05.012

Crane-Droesch, A. (2018). Machine learning methods for crop yield prediction and climate change impact assessment in agriculture. *Environmental Research Letters, 13*(11), 114003. https://doi.org/10.1088/1748-9326/aae159

Demirhan, H. (2025). A deep learning framework for prediction of crop yield in Australia under the impact of climate change. *Information Processing in Agriculture*, *12*(1), 125-138.

Food and Agriculture Organization. (2020). Banana market review: Preliminary results *2019*. Rome.

Islam, M. R., Oliullah, K., Kabir, M. M., Alom, M., & Mridha, M. F. (2023). Machine learning enabled IoT system for soil nutrients monitoring and crop recommendation. *Journal of Agriculture and Food Research, 14*, 100880.

Kumar, Y. J. N., Spandana, V., Vaishnavi, V. S., Neha, K., & Devi, V. G. R. R. (2020). Supervised machine learning approach for crop yield prediction in agriculture sector. In 5th *International Conference on Communication and Electronics Systems (ICCES)* (pp. 736-741). IEEE.

Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors, 18*(8), 2674. https://doi.org/10.3390/s18082674

Mishra, S., Mishra, D., & Santra, G. H. (2016). Applications of machine learning techniques in agricultural crop production: A review paper. *Indian Journal of Science and Technology, 9*(38), 1-14. https://doi.org/10.17485/ijst/2016/v9i38/95032

Motamedi, B., & Villányi, B. (2024). A predictive analytics model with bayesian-optimized ensemble decision trees for enhanced crop recommendation. *Decision Analytics Journal*, *12*, 100516.

Nigam, A., Garg, S., Agrawal, A., & Agrawal, P. (2019). Crop yield prediction using machine learning algorithms. *In Fifth International Conference on Image Information Processing (ICIIP)* (pp. 125-130). https://doi.org/10.1109/ICIIP47207.2019.8985951

Patrick, S., Mirau, S., Mbalawata, I., & Leo, J. (2023). Time series and ensemble models to forecast banana crop yield in Tanzania, considering the effects of climate change. *Resources, Environment and Sustainability*, *14*, 100138.

Ramzan, S., Ghadi, Y. Y., Aljuaid, H., Mahmood, A., & Ali, B. (2024). An ingenious IoT based crop prediction system using ML and EL. *Computers, Materials and Continua, 79*(1), 183-199.

Reddy, D. J., & Kumar, M. R. (2021). Crop yield prediction using machine learning algorithm. *In 5th International Conference on Intelligent Computing and Control Systems (ICICCS) (*pp. 1466-1470). IEEE.

Kumari, P., Gaur, S. S., & Tiwari, R. K. (2023). Banana and its by-products: A comprehensive review on its nutritional composition and pharmacological benefits. *EFood*, *4*(5), e110.

Sellam, V., & Poovammal, E. (2016). Prediction of crop yield using regression analysis. *Indian Journal of Science and Technology*, 9(38), 1-5. https://doi.org/10.17485/ijst/2016/v9i38/91714

Shah, A., Dubey, A., Hemnani, V., Gala, D., & Kalbande, D. R. (2018, April). Smart farming system: Crop yield prediction using regression techniques. In Proceedings of International Conference on Wireless Communication: ICWiCom 2017 (pp. 49-56). Singapore: Springer Singapore. https://doi.org/10.1007/978-981-10-8339-6_6

Sharma, A., Jain, A., Gupta, P., & Chowdary, V. (2020). Machine learning applications for precision agriculture: A comprehensive review. *IEEe Access*, *9*, 4843-4873.

Singh, B., Singh, J. P., Kaur, A., & Singh, N. (2016). Bioactive compounds in banana and their associated health benefits - A review. *Food Chemistry*, *206*, 1-11. https://doi.org/10.1016/j.foodchem.2016.03.033

Tripathi, L., Ntui, V. O., & Tripathi, J. N. (2019). Application of genetic modification and genome editing for developing climate-smart banana. *Food and Energy Security*, *8*(4), e00168. https://doi.org/10.1002/fes3.168

van Klompenburg, T., Kassahun, A., & Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture*, *177*, 105709. https://doi.org/10.1016/j.compag.2020.105709

Venugopal, A., Aparna, S., Mani, J., Mathew, R., & Williams, V. (2021). Crop yield prediction using machine learning algorithms. I*nternational Journal of Engineering Research & Technology (IJERT) NCREIS*, *9*(13).

von Bloh, M., Nóia Júnior, R. d. S., Wangerpohl, X., Saltık, A. O., Haller, V., Kaiser, L., & Asseng, S. (2023). Machine learning for soybean yield forecasting in Brazil. *Agricultural and Forest Meteorology, 341*, 109670.