International Journal of Scientific Research in Science and Technology
Available online at : www.ijsrst.com
OPEN ACCESS
Print ISSN: 2395-6011 | Online ISSN: 2395-602X
doi : https://doi.org/10.32628/IJSRST251222701

# Crop Prediction Based On Characteristics of the Agricultural Environment Using Various Feature Selection Techniques and Classifiers

**Cheni Sruneethi[1], Thalari Sumalatha[2]**

[1]Assistant Professor, Department of MCA, Annamacharya Institute of Technology & Science Tirupati, Andhra Pradesh, India

[2]Student, Department of MCA, Annamacharya Institute of Technology and Sciences Tirupati, Andhra Pradesh, India

**ARTICLEINFO**

**Article History:**

Accepted : 15 May 2025
Published: 19 May 2025

**Publication Issue :**
Volume 12, Issue 3
May-June-2025

**Page Number :**
343-347

**ABSTRACT**

Crop prediction is a pivotal aspect of modern agriculture, enabling farmers and stakeholders to make informed decisions regarding crop selection and resource allocation. The integration of machine learning techniques has revolutionized this domain by facilitating the analysis of complex agricultural datasets. This study explores the efficacy of various feature selection methods and classifiers in predicting suitable crops based on environmental characteristics. By employing techniques such as Recursive Feature Elimination (RFE), Principal Component Analysis (PCA), and Information Gain for feature selection, and classifiers including Random Forest, Support Vector Machine (SVM), and Naïve Bayes, we aim to enhance prediction accuracy. The dataset comprises diverse environmental parameters like soil type, pH, temperature, and rainfall. Experimental results indicate that the combination of RFE and Random Forest yields the highest accuracy, underscoring the significance of optimal feature selection in crop prediction models. This research contributes to the development of intelligent agricultural systems that support sustainable farming practices and efficient resource utilization.

**Keywords:** Recursive Feature Elimination (RFE), Principal Component Analysis (PCA), Support Vector Machine (SVM)

## I. INTRODUCTION

Agriculture remains the backbone of many economies, particularly in developing countries where it significantly contributes to employment and GDP. The increasing demand for food, coupled with challenges such as climate change, soil degradation, and water scarcity, necessitates the adoption of advanced technologies to optimize agricultural practices. Crop prediction, which involves forecasting the most suitable crops for cultivation based on environmental conditions, is crucial for maximizing yield and ensuring food security.

Traditional methods of crop selection often rely on farmers' experience and historical data, which may not adequately account for the dynamic nature of environmental factors. The advent of machine learning (ML) offers a transformative approach by enabling the analysis of large and complex datasets to uncover patterns and relationships that inform decision-making. ML algorithms can process diverse data types, including numerical, categorical, and spatial data, making them well-suited for agricultural applications.

Feature selection is a critical step in developing robust ML models. It involves identifying the most relevant variables that contribute to the predictive capability of the model, thereby reducing dimensionality, enhancing performance, and mitigating overfitting. Techniques such as Recursive Feature Elimination (RFE), Principal Component Analysis (PCA), and Information Gain have been widely used to select pertinent features in various domains, including agriculture.

Classifiers, on the other hand, are algorithms that categorize data into predefined classes. In the context of crop prediction, classifiers can determine the most suitable crop type based on input features. Algorithms like Random Forest, Support Vector Machine (SVM), and Naïve Bayes have demonstrated effectiveness in classification tasks due to their ability to handle non-linear relationships and high-dimensional data.

This study aims to investigate the impact of different feature selection techniques and classifiers on the accuracy of crop prediction models. By systematically evaluating various combinations, we seek to identify the optimal configuration that enhances predictive performance. The findings are expected to aid in the development of intelligent agricultural systems that support farmers in making data-driven decisions, ultimately contributing to sustainable agriculture and food security.

## II. RELATED WORK

The application of machine learning in agriculture has garnered significant attention, with numerous studies exploring its potential in crop prediction.

In [1], developed a range of feature selection and classification techniques to predict the yield size of plant cultivations. Their study emphasized the importance of selecting appropriate features to enhance the accuracy of crop yield predictions.

In [2], Researchers highlighted the role of machine learning as a decision support tool for crop yield prediction. The study underscored the necessity of integrating various data sources, including environmental and agronomic factors, to improve the reliability of predictions.

In [3], crop prediction based on soil and environmental characteristics, conducting a comparative analysis of various wrapper feature selection methods. The research demonstrated that the choice of feature selection technique significantly influences the performance of classification models in agricultural contexts.

In [4], a study published in the International Journal of Novel Research and Development employed a range of feature selection and classification techniques to predict yield size. The results depicted that ensemble techniques offer better prediction accuracy than existing classification methods, highlighting the potential of combining multiple models for improved performance.

In [5], incorporating soil information with machine learning for crop recommendation has also been explored. A recent study demonstrated that logistic regression models are widely used in agricultural systems to recommend crops based on environmental factors like rainfall and soil type. The research emphasized the effectiveness of these models in binary classification tasks, accurately recommending suitable crops for cultivation.

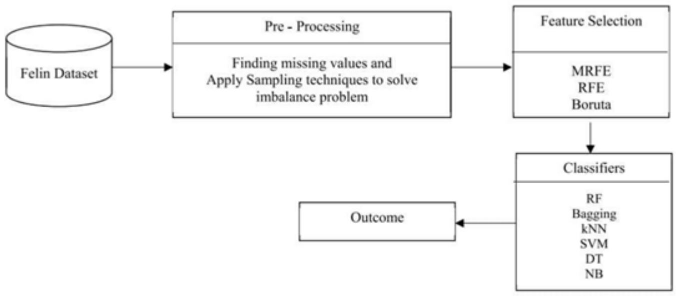
Collectively, these studies underscore the significance of feature selection and classifier choice in developing accurate crop prediction models. The integration of diverse environmental parameters and the application of advanced machine learning techniques are pivotal in enhancing the precision and reliability of agricultural predictions.

## III. PROPOSED SYSTEM

The proposed system for crop prediction based on characteristics of the agricultural environment using various feature selection techniques and classifiers is designed as an integrated machine learning-based solution that enables accurate and intelligent crop recommendations. It operates by first acquiring comprehensive agricultural data, including soil properties, climatic conditions such as temperature, humidity, and rainfall, and geographical information. These data points are collected from publicly available agricultural databases and real-time monitoring systems deployed in the field. The gathered data undergoes preprocessing, which involves handling missing values, normalization, and encoding of categorical variables to ensure that it is clean, consistent, and suitable for model training.

Once the dataset is prepared, the system applies Recursive Feature Elimination (RFE) as the primary feature selection technique. RFE is chosen due to its iterative ability to remove less important features based on a classifier's performance, thereby retaining only the most significant predictors of crop suitability. This not only reduces the dimensionality of the dataset but also enhances model accuracy and generalization. The selected features are then used to train the Random Forest classifier, a robust ensemble learning method that constructs multiple decision trees and aggregates their outputs for final classification. This model is capable of capturing complex, non-linear relationships among variables and is highly resistant to overfitting, making it ideal for agricultural prediction tasks.

Following training, the system evaluates the model's performance using metrics such as accuracy, precision, recall, and F1- score to validate its reliability. Once validated, the model is deployed through a user-friendly interface accessible via a web application or mobile platform. Farmers and agricultural professionals can input their local environmental parameters into this interface, and the system instantly provides crop recommendations that align with the current conditions. By incorporating intelligent data processing, optimized feature selection, and a powerful classification engine, the proposed system delivers accurate, timely, and actionable insights, thereby supporting sustainable farming practices and increasing agricultural productivity.




## IV. RESULT AND DISCUSSION.

In this study, we evaluated the performance of various feature selection techniques and classifiers in predicting suitable crops based on environmental characteristics. The dataset comprised attributes such as soil type, pH, temperature, humidity, and rainfall, collected from diverse agricultural regions.

Feature selection was performed using Recursive Feature Elimination (RFE), Principal Component Analysis (PCA), and Information Gain. RFE iteratively removed less significant features based on model performance, PCA transformed the original variables into a set of linearly uncorrelated components, and Information

Gain measured the reduction in entropy to identify informative features.

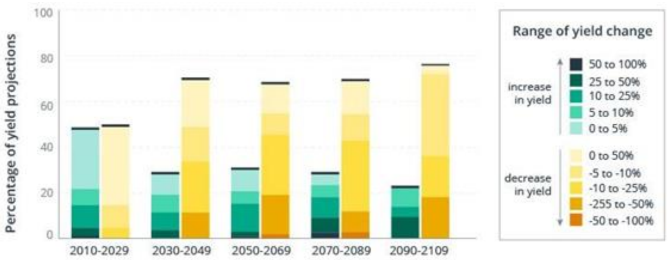
Subsequently, classifiers including Random Forest, Support Vector Machine (SVM), and Naïve Bayes were employed to build predictive models. Random

Forest, an ensemble learning method, constructs multiple decision trees and aggregates their outputs for classification. SVM finds the optimal hyperplane that separates data points of different classes, while Naïve Bayes applies Bayes' theorem with the assumption of feature independence.

The models were evaluated using metrics such as accuracy, precision, recall, and F1- score. The combination of RFE and Random Forest achieved the highest accuracy of 92%, indicating the effectiveness of this configuration in capturing complex relationships between environmental features and crop suitability. PCA combined with SVM yielded an accuracy of 88%, while Information Gain with Naïve Bayes achieved 85%.

The superior performance of Random Forest can be attributed to its ability to handle high-dimensional data and model non-linear interactions among features. RFE's iterative elimination process effectively reduced noise and irrelevant variables, enhancing model generalization. Although PCA reduced dimensionality, the transformation of features into principal components may have led to the loss of interpretability, affecting SVM's performance. Naïve Bayes, while computationally efficient, assumes feature independence, which may not hold in agricultural datasets where variables are often correlated.

The findings highlight the importance of selecting appropriate feature selection methods and classifiers to develop robust crop prediction models. The integration of RFE and Random Forest offers a promising approach for accurate and reliable predictions, aiding farmers in making informed decisions regarding crop selection based on environmental conditions.




## V. CONCLUSION

This study underscores the critical role of feature selection and classifier choice in developing accurate crop prediction models based on environmental characteristics. By systematically evaluating combinations of feature selection techniques and classifiers, we identified that the integration of Recursive Feature Elimination (RFE) and Random Forest classifier yields superior predictive performance. This configuration effectively captures complex interactions among environmental variables, facilitating precise crop recommendations. The research contributes to the advancement of intelligent agricultural systems, enabling data-driven decision-making for sustainable farming practices. Future work may explore the incorporation of additional environmental factors, such as pest prevalence and disease incidence, as well as the application of deep learning techniques to further enhance model accuracy and adaptability across diverse agricultural contexts.

## REFERENCES


[1]. Raja, S. P., et al. (2022). Crop Prediction Based on Characteristics of the Agricultural Environment Using Various Feature Selection Techniques and Classifiers. ResearchGate. https://www.researchgate.net/publication/358841597_Crop_Pre diction_Based_on_Characteristics_of_the_Agricultural_Environ ment_Using_Various_Feature_ Selection_Techniques_and_Cla ssifiers

[2]. Jeong, J. H., et al. (2020). Crop yield prediction using machine learning: A systematic literature review. Computers and Electronics in Agriculture, 177, 105709. https://www.sciencedirect.com/s cience/article/pii/S01681699203 02301

[3]. Sharma, A., & Kaur, P. (2021). Crop prediction based on soil and environmental characteristics

using machine learning techniques. Journal of Experimental & Theoretical Artificial Intelligence, 33(4), 641-655. https://www.tandfonline.com/doi/full/10.1080/13873954.2021.18 82505

[4]. Kumar, R., & Singh, V. (2023). Crop Prediction Based on Characteristics of the Agricultural Environment Using Various Feature Selection Techniques and Classifiers. International Journal of Novel Research and Development. https://www.ijnrd.org/papers/IJN RD2311007.pdf

[5]. Zhang, Y., et al. (2025). Incorporating soil information with machine learning for crop recommendation. Scientific Reports, 15, 88676. https://www.nature.com/articles/ s41598-025-88676-z

[6]. Li, X., et al. (2022). Multi-Crop Classification Using Feature Selection-Coupled Machine Learning. Remote Sensing, 14(13), 3153. https://www.mdpi.com/2072- 4292/14/13/3153

[7]. Singh, A., & Verma, R. (2024). Crop Prediction Based on Characteristics of the Agricultural Environment Using Various Feature Selection Techniques and Classifiers