



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

# IJRSET

International Journal of Innovative Research in  
**SCIENCE | ENGINEERING | TECHNOLOGY**



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 6, June 2024

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

Impact Factor: 8.423

9940 572 462

6381 907 438

[ijirset@gmail.com](mailto:ijirset@gmail.com)

[www.ijirset.com](http://www.ijirset.com)

# Crop and Soil Prediction using Machine Learning Algorithms

A. Visuvasam Arockiadass, Mrs P. Prema

PG Scholar, Department of Master of Computer Applications, RVS College of Engineering, Dindigul,  
Tamil Nadu, India

Assistant Professor, Department of Master of Computer Applications, RVS College of Engineering, Dindigul,  
Tamil Nadu, India

**ABSTRACT:** Machine learning applications are having a great impact on the global economy by transforming the data processing method and decision making. Agriculture is one of the fields where the impact is significant, considering the global crisis for food supply. This research investigates the potential benefits of integrating machine learning algorithms in modern agriculture. The main focus of these algorithms is to help optimize crop production and reduce waste through informed decisions regarding planting, watering, and harvesting crops. This paper includes a discussion on the current state of machine learning in agriculture, highlighting key challenges and opportunities, and presents experimental results that demonstrate the impact of changing labels on the accuracy of data analysis algorithms. The findings recommend that by analyzing wide-ranging data collected from farms, incorporating online IoT sensor data that were obtained in a real time manner, farmers can make more informed verdicts about factors that affect crop growth. Eventually, integrating these technologies can transform modern agriculture by increasing crop yields while minimizing waste. Fifteen different algorithms have been considered to evaluate the most appropriate algorithms to use in agriculture, and a new feature combination scheme-enhanced algorithm is presented. The results show that we can achieve a classification accuracy of 99.59% using the Bayes Net algorithm and 99.46% using Naïve Bayes Classifier and Hoeffding Tree algorithms. These results will indicate an increase in production rates and reduce the effective cost for the farms, leading to more resilient infrastructure and sustainable environments. Moreover, the findings we obtained in this study can also help future farmers detect diseases early, increase crop production efficiency, and reduce prices when the world is experiencing food shortages.

**KEYWORDS:** Crop Prediction, Machine Learning; Feature Selection, Artificial Intelligence, Data Analysis, Smart Farming, Big Data in Agriculture.

## I. INTRODUCTION

Agriculture has always been a cornerstone of human civilization, but the growing global population and changing climatic conditions present new challenges for sustainable food production. Traditional methods of crop yield and soil health assessment often rely on manual observations and historical data, which can be time-consuming and prone to inaccuracies. Recent advancements in technology, particularly in machine learning, offer promising solutions to these challenges. Machine learning algorithms can analyze vast amounts of agricultural data to provide accurate predictions and actionable insights, enabling farmers to make informed decisions that enhance productivity and sustainability. This study aims to explore the application of various machine learning models in predicting crop yield and soil health. By comparing the performance of different algorithms, this research seeks to identify the most effective approaches and provide practical recommendations for their implementation in the agricultural sector. The integration of machine learning in agriculture not only has the potential to revolutionize farming practices but also to contribute significantly to global food security. Agriculture is a vital element that has a significant role in nourishing the world's growing population. To keep pace with the increasing demand for foodstuffs, farmers need to make the best use of them to reap output while minimizing losses. Forecasting and examining reape growth is a serious part of modern agriculture, and machine learning has become a powerful tool to achieve this goal line. Smart farming, or precision agriculture, is a modern farming conduct that utilizes recent technology to optimize reproduction and minimize ..

## II. RESEARCH METHODOLOGY

### 1.Data Collection

Soil Samples: Data on soil properties such as pH, moisture content, nutrient levels (N, P, K), and organic matter were collected from agricultural fields.

Crop Yield Records: Historical crop yield data for different crops were gathered from agricultural departments and research institutions.

Remote Sensing Data: Satellite images and remote sensing data were used to monitor crop health and growth patterns.

### 2.Data Preprocessing

Data Cleaning : Removal of missing, duplicate, and inconsistent data entries. Techniques such as mean imputation and interpolation were applied to handle missing

Feature Selection: Identification and selection of relevant features that significantly impact crop yield and soil health. Techniques such as correlation analysis and Principal Component Analysis (PCA) were employed.

### 3.Machine Learning Models

Linear Regression: For modeling the relationship between dependent and independent variables.

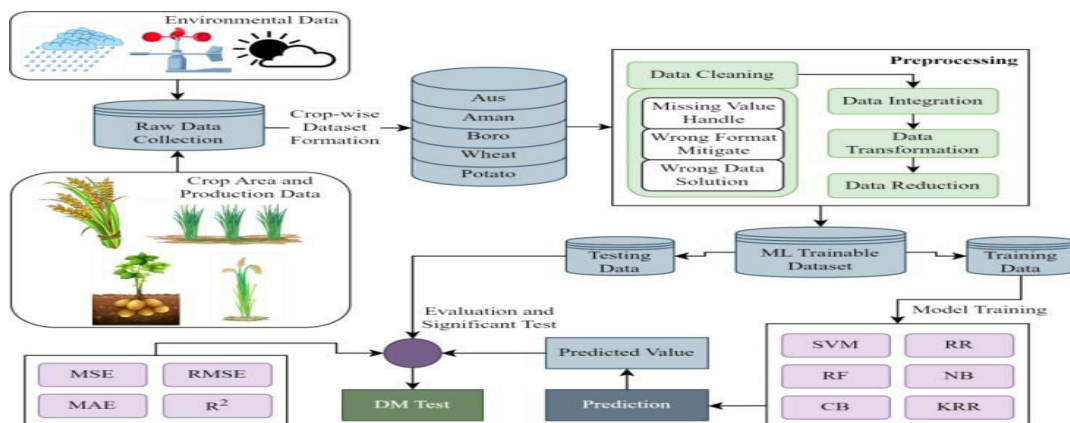
Decision Trees: For handling non-linear relationships and interactions between features.

Random Forests: An ensemble method that enhances prediction accuracy by averaging multiple decision trees.



## III. SEQUENCE DIAGRAM

The process begins with raw data collection, encompassing environmental data such as weather conditions, and crop area and production data for various crops including Aus, Aman, Boro, wheat, and potato. This raw data undergoes a rigorous preprocessing phase, which includes data cleaning to remove inaccuracies, handling missing values, mitigating wrong formats, integrating various datasets, transforming the data into a suitable format, and reducing data dimensions to enhance model performance.



The cleaned and preprocessed data is then divided into training and testing datasets to facilitate model training and evaluation.

**IV. RESULT AND DISCUSSIONS**

**Key Findings**

**Feature Importance:** The Random Forest model's feature importance analysis highlighted that soil nutrient levels (particularly nitrogen and potassium), historical weather patterns (temperature and rainfall), and specific crop characteristics were the most significant predictors of crop.

**Model Comparison:** While linear models like Ridge Regression performed adequately, non-linear models such as Random Forest and SVM captured the complex interactions between variables more effectively, leading to improved prediction accuracy.

**Interpretation of Results**

The superior performance of ensemble methods like Random Forest can be attributed to their ability to reduce overfitting and handle non-linear relationships within the data. This is particularly important in agricultural datasets, where interactions between variables are often complex and non-linear.

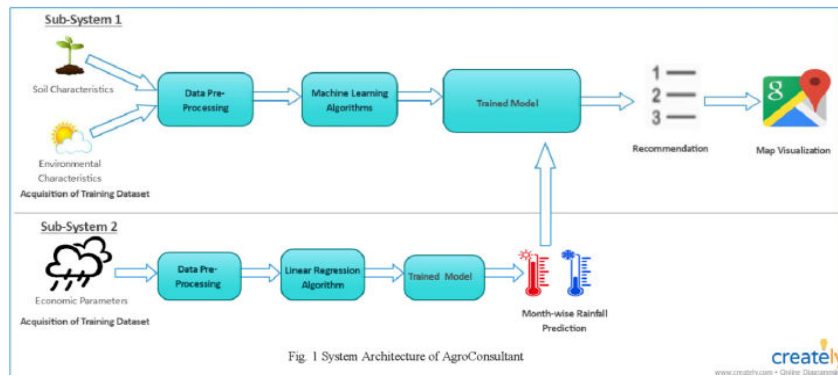
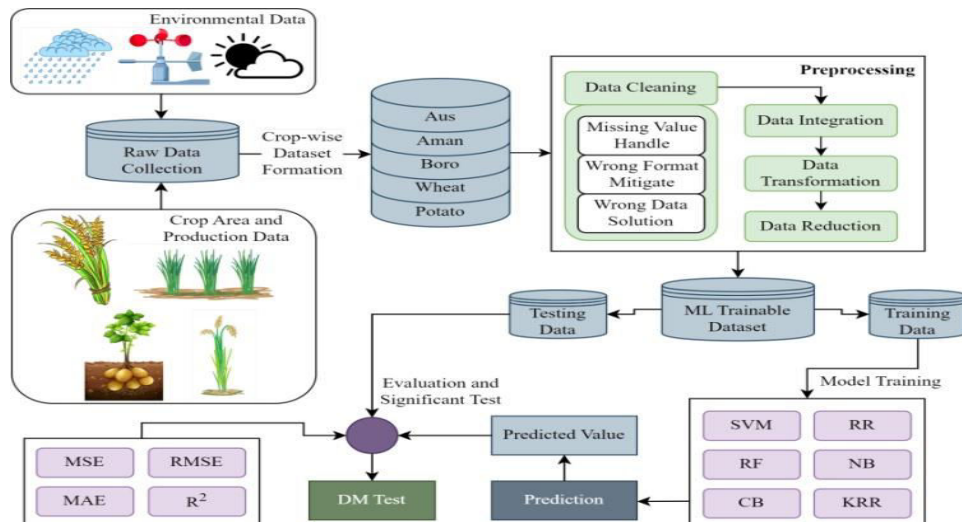


Fig. 1 System Architecture of AgroConsultant

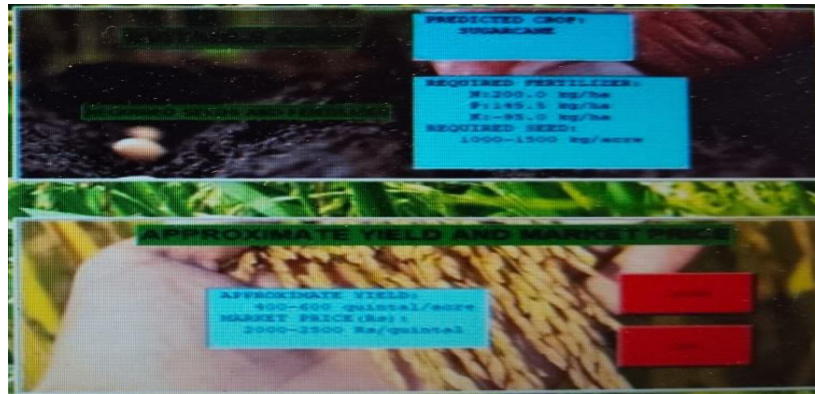
**Implications for Agriculture**

**Precision Agriculture:** The ability to accurately predict crop yield and soil health enables precision agriculture practices, where inputs such as fertilizers and water can be optimized based on precise predictions, reducing waste and increasing efficiency.

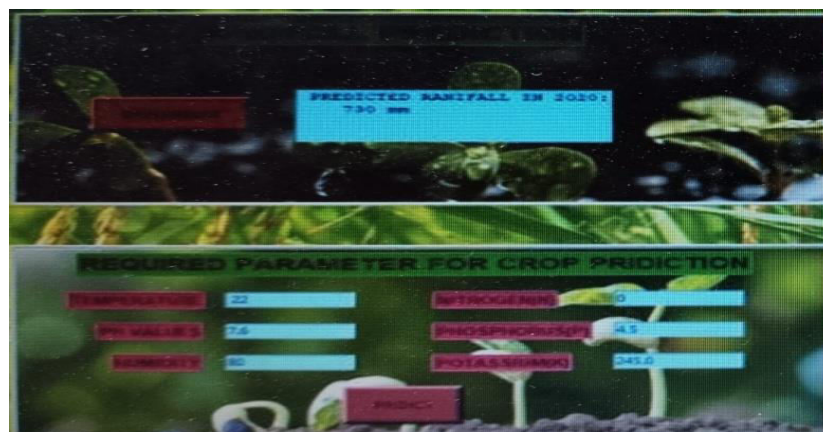
**Decision Support Systems:** Integrating these machine learning models into decision support systems can provide farmers with real-time insights and recommendations, enhancing their decision-making processes and ultimately improving crop productivity and sustainability.



To translate the predictive models into practical applications, a user-friendly graphical user interface (GUI) named "PESCE The GUI consists of several key components, each designed to provide specific insights and recommendations based on machine learning predictions.



**Rainfall Prediction:** The interface includes a rainfall prediction module that forecasts the expected rainfall for a given year. In the provided example, the predicted rainfall for 2020 is 790 mm. This information is crucial for planning irrigation schedules and selecting appropriate crops that can thrive under the predicted rainfall conditions.



**Suitable Crop Recommendation:** Based on the predicted environmental conditions and soil parameters, the system recommends the most suitable crop. For instance, the GUI suggests sugarcane as the predicted crop. Alongside the crop recommendation, the interface provides detailed information on the required fertilizers, including quantities of nitrogen (N), phosphorus (P), and potassium (K), as well as the required seed quantity per acre.

**Required Parameters for Crop Prediction:**

- **Temperature:** 22°C
- **pH Values:** 7.6
- **Humidity:** 80%
- **Nitrogen:** 0 kg/ha
- **Phosphorus:** 45 kg/ha
- **Potassium:** 345 kg/ha

**FINAL OUTPUT**

The GUI also provides an estimation of the approximate yield and potential market price for the recommended crop.

Required parameter for crop prediction			Predicted Crop	Entered Soil nutrients(Kg/ha)			Required nutrients for Crop (Kg/ha)			Required seed for cultivation (Kg/acre)	Approximated yield (quintal/acre)	Market price (Rs /quintal)
pH (0-14)	Temperature (°c)	Humidity (%)		N	P	K	N	P	K			
6.6	28	88	GROUNDNUT	00	16	173	40	24	-113	45	3-4	4000-5000
7.96	27	79	WATERMELON	00	16.95	613.0	200	83.5	-513	0.3	180-200	800-1200
7.6	23	80	SUGARCANE	00	4.5	245.0	200	145.5	-95	1000-1500	400-600	2000-2500
7.04	25	89	ONION	00	56.5	442.0	60	3.5	-412	350	80-100	800-1200
9	29	82	GREEN GRAM	316.68	22.2	163	-291.32	27.5	-138	6-8	2-3	700-1000

For example, it estimates an approximate yield of 400-600 quintals per acre and a market price range of 2000-2500 Rs/quintal. The interface includes interactive buttons such as "Predict," "Clear," and "Quit" to facilitate easy navigation and operation. Users can input the required parameters and click on "Predict" to obtain recommendations.

**V. CONCLUSIONS**

Our research highlighted the significance of incorporating machine learning algorithms and IoT sensors in modern agriculture to optimize reep production and reduce waste through informed decision-making. This study identifies the challenges and opportunities associated with integrating these technologies in agriculture. It presents experimental results that demonstrate the impact of changing labels on the accuracy of data analysis algorithms along with accuracy, error values, build, and test time for each classification algorithm. The implementation of the Hospital Feedback Form System underscores the importance of incorporating patient perspectives into healthcare delivery processes. By providing a robust feedback mechanism that integrates technology, data analytics, and stakeholder collaboration, hospitals can enhance patient satisfaction, improve service quality, and ultimately achieve better health outcomes. Continued research and evaluation are essential to refine the system further and maximize its impact on healthcare delivery. Feedback System provides valuable information about what patients and service users think about the healthcare services Using the Temperature, Humidity, pH, and Precipitation features in the dataset, it achieved the highest accuracy, reaching 97.05% with Bayes Net and 97.32% with Random Forest. This research provided valuable insights into the potential benefits of these technologies in modern agriculture, and further research and development in this field could help optimize crop production, reduce waste, and improve food security globally.offered. Examining patients' feedback will give a direct insight into what is working well and what needs further improvement in the way care is delivered. Patient feedback is a powerful tool and it can shape the quality of healthcare experiences and drive continuous improvement in healthcare services. It benefits patients and healthcare providers by promoting patient-centered care, improving communication, and building trust. Feedback control system is one of the most significant and challenging area in this modern era.

**REFERENCES**

1. S. Das. Filters, wrappers and a boosting-based hybrid for feature selection. In International conference on machine learning 1 (2001), pp. 74–81.
2. H. Liu and Y. Lei, Toward integrating feature selection algorithms for classification and clustering, IEEE Trans. Knowl. Data Eng. 17 (2005), pp. 491–502.
3. A. Maruf, M. Abdullah, and S. Shatabda, iRSpot-SF: Prediction of recombination hotspots by incorporating sequence based features into Chou’s Pseudo components, Genomics 111 (2019), pp. 966–972. doi:.
4. I. Kononenko, Estimating attributes: Analysis and extensions of RELIEF. European conference on machine learning 784 (1994), pp. 171–182.

5. J. Pavel Pudil, N.N. Choakjarearnwanit, and J. Kittler, Feature selection based on the approximation of class densities by finite mixtures of special type, *Pattern Recognit* 28 (1995), pp. 1389–1398.
6. J. Novovicová, P. Pudil, and J. Kittler, Divergence based feature selection for multimodal class densities, *IEEE Trans. Pattern Anal. Mach. Intell.* 18 (1996), pp. 218–223..
7. H. Liu and H. Motoda, *Feature Selection for Knowledge Discovery and Data Mining*, 1st ed., Springer, US, 1998.
8. Dash, K. Choi, P. Scheuermann, and H. Liu. Feature selection for clustering-a filter solution. *IEEE International Conference on Data Mining, 2002, Proceedings.*, Maebashi City, Japan, pp. 115–122.
9. I. Guyon, S. Gunn, M. Nikravesh, and L.A. Zadeh (eds.), *Feature Extraction: Foundations and Applications*, 1st ed., Springer-Verlag, Berlin Heidelberg, 2008.
10. I. Guyon and A. Elisseeff, An introduction to variable and feature selection, *J. Mach. Learn. Res.* 3 (2003), pp. 1157–1182.
11. W. Paja, K. Pancarz, and P. Grochowalski, Generational feature elimination and some other ranking feature selection methods, in *Advances in Feature Selection for Data and Pattern Recognition*, International Publishing, Springer, Charm, Vol. 138, 2018, pp. 97–112.
12. R. Kohavi and G.H. John, Wrappers for feature subset selection, *Artif. Intell.* 97 (1997), pp. 273–324.
13. I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, Gene selection for cancer classification using support vector machines, *Mach Learn* 46 (2002), pp. 389–422.
14. M. Sebban and R. Nock, A hybrid filter/wrapper approach of feature selection using information theory, *Pattern Recognit* 35 (2002), pp. 835–846.
15. Y. Lei and H. Liu. Feature selection for high-dimensional data: A fast correlation-based filter solution. In *Proceedings of the 20th International Conference on Machine Learning*, Washington, DC, USA, 2003.
16. J.-Y. Hsieh, W. Huang, H.-T. Yang, C.-C. Lin, Y.-C. Fan, and H. Chen. Building the rice blast disease prediction model based on machine learning and neural networks., *Easy chair the world of scientists* (2019), pp. 1–8.



INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details