INTERNATIONAL JOURNAL FOR INNOVATIVE RESEARCH IN MULTIDISCIPLINARY FIELD
ISSN(O): 2455-0620 [Impact Factor: 9.47]
Monthly, Peer-Reviewed, Refereed, Indexed Journal with IC Value : 86.87
Volume - 12, Issue - 7, July - 2026

DOIs:10.2015/IJIRMF/202607008 --:-- Research Paper / Article / Review

An Enhancing Crop Yield and Resource Efficiency with Machine Learning

[1]Ms. Bhumikaben L Chaudhari, [2]Ms. Jalpa Bhatt
[1]Post Graduate Student Dept. of Computer Science & Engineering, Dr. Kiran & Pallavi Global University, Vadodara Gujarat, India.
[2]Assistant Professor Dept. of Computer Science & Engineering, Dr. Kiran & Pallavi Global University, Vadodara Gujarat, India.
Email id: [1]bhumikachaudhari1396@gmail.com
Email id: [2]jalpabhattpj@gmail.com

Abstract: *Agriculture faces increasing pressure to meet global food demands while coping with limited resources and environmental challenges. Machine Learning (ML) and data-driven techniques to enable smarter, more sustainable farming practices. By integrating data on soil conditions, environmental factors, weather, and historical crop performance, the study develops predictive models to forecast crop yield and optimize resource usage such as water, fertilizers, and nutrients. The research aims at integrating these*

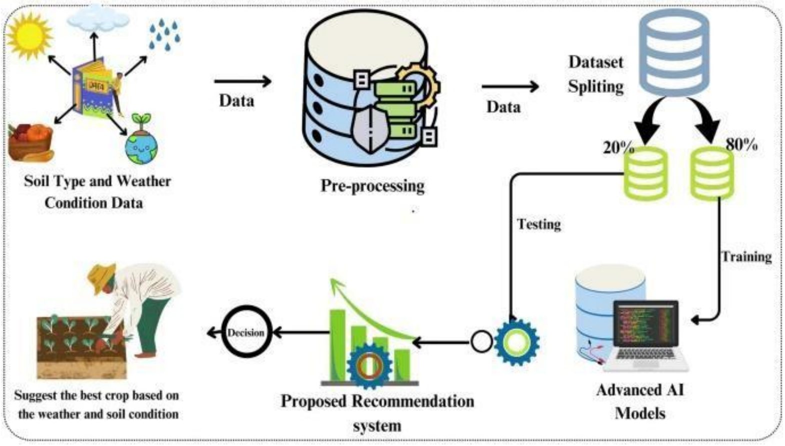
Agriculture must dramatically increase productivity to feed a growing global population with minimal land expansion. This challenge is compounded by climate change, soil degradation, and water scarcity. Traditional forecasting methods (e.g. statistical models, time-series) struggle to model complex interactions among weather, soil, genetics, and management. In contrast, ML techniques like Random Forest, SVM, and neural networks excel at handling nonlinearity and high-dimensional data. Recent advances in sensor networks, satellite and drone imagery, and IoT enable the collection of rich, real-time data. climate change, weather variability, soil quality, genetics of seeds, and crop management practices. Crop yield variation is brought by intricate GxExM interactions of genetics, environment and farm management practices involving soil type, weather conditions, seed genetics and farmer farm management decision [1]. Knowledge of technologies with the Random Forest classifier to maximize crop productivity, optimize resource allocation, enhance crop health monitoring, automate farming operations, predict weather effects, ensure sustainable practices, and predict crop pricing. The research tackles issues such as climate change, resource shortages, and food insecurity by utilizing sophisticated machine learning, specifically random forests, to contribute towards enhanced global farming practices through intelligent data-driven approaches.

Keywords: *Artificial Intelligence, Internet of Things ,Support Vector Machine, Random Forest, Machine Learning, Extreme Gradient Boosting.*

1.INTRODUCTION

These properties is vital to crafting efficacious farm strategy. Soil type variability, for instance, can either soften or intensify the effects of climate and weather on crop yields. A reduction in variability of crop yields has been suggested as an efficient means of enhancing food Availability without area expansion [2]. Various machine learning models have been applied in agriculture, such as Support Vector Machines (SVM), Decision Trees, and Neural Networks. Among these, the Random Forest algorithm has gained significant attention due to its high accuracy, robustness, and ability to handle large datasets with minimal overfitting. Studies demonstrate that Random Forest performs effectively in crop yield prediction, soil classification, and disease detection.[3].

Fig 1: Architecture diagram of crop yield prediction



Classical statistical methods are generally unable to capture complex interdependencies between parts in agriculture systems and consist of time series analysis and regression analysis. Using ML algorithms like SVM, decision trees, and random forests produces enhanced accuracy and flexibility when it comes to predicting crop yields [14]. These techniques can handle high-dimensional data and nonlinear relations and are hence perfect for complex agricultural systems. Recent advances in technology, such as remote sensing, satellite imagery, and IoT sensors, have also enhanced crop forecasting models by offering real-time farm data so that it is possible to have more timely and precise forecasts that help farmers and policymakers make better decisions [15].

2. LITERATURE REVIEW

Proposed a smart agriculture system using Machine Learning techniques to enhance crop yield and optimize resource utilization. The study demonstrated improved efficiency by integrating environmental and soil data, but lacked large-scale real-time validation [1]. Presented a systematic literature review on crop yield prediction models. The study compared various algorithms and concluded that ensemble methods perform better, though issues such as data inconsistency and scalability remain unresolved [2]. Focused on soil moisture prediction using multimodal remote sensing data fusion and machine learning. Their approach improved prediction accuracy in semi-arid regions, but required high computational resources and complex data integration [3]. Provided a comprehensive review of machine learning and deep learning techniques for crop prediction. The study highlighted improved accuracy with deep learning but noted challenges in interpretability and high computational cost [5]. Developed a smart IoT-based precision agriculture system integrating land mapping, crop prediction, and automated irrigation. The system uses IoT sensors to collect real-time environmental data such as soil moisture, temperature, and humidity. These data are then processed using machine learning techniques to enable crop prediction and optimize irrigation scheduling. The study emphasizes real-time decision-making and automation, allowing efficient water management and improved crop productivity. The integration of IoT devices enhances the system's responsiveness and adaptability to changing environmental conditions [7]. Presented a comprehensive review of IoT-based smart drip irrigation systems, focusing on architectures, machine learning techniques, and emerging trends. The study highlights the importance of integrating sensors, cloud computing, and ML algorithms to optimize water usage in agriculture. Various machine learning models such as Decision Trees, Artificial Neural Networks, and Support Vector Machines are discussed for irrigation prediction and control. The review also emphasizes automation, sustainability, and efficient resource utilization as key benefits of smart irrigation systems [8]. ML models such as Random Forest, Support Vector Machines, and deep learning approaches applied to high-resolution UAV imagery. The use of UAV data enables precise monitoring of crop health, vegetation indices (e.g., NDVI), and field variability, leading to improved prediction accuracy. However, the study highlights key limitations including high operational costs, data processing complexity, and challenges in handling large-scale datasets. It also emphasizes the need for efficient data pre-processing pipelines and real-time analytics for practical agricultural deployment [9]. Application of Artificial Intelligence (AI) and Internet of Things in enhancing security for crop yield prediction systems. The authors discuss how IoT-enabled agricultural environments rely on continuous data collection from sensors such as soil moisture, temperature, and humidity, which are then processed using AI models for prediction. A key contribution of the work is the emphasis on securing data transmission and storage through techniques like encryption and intrusion detection systems, ensuring reliability and integrity of agricultural data [11].

INTERNATIONAL JOURNAL FOR INNOVATIVE RESEARCH IN MULTIDISCIPLINARY FIELD
ISSN(O): 2455-0620 [Impact Factor: 9.47]
Monthly, Peer-Reviewed, Refereed, Indexed Journal with IC Value : 86.87
Volume - 12, Issue - 7, July - 2026

3. METHODOLOGY

The study proposes a machine learning-based multi-task learning framework for intelligent agricultural decision-making. The framework is designed to improve crop yield prediction, soil fertility assessment, and resource optimization by leveraging structured agricultural data. The overall workflow consists of data collection, preprocessing, feature engineering, model development, and evaluation.

A. Data Collection

The dataset used in this study is collected from agricultural databases and consists of both soil and environmental attributes. The input features include soil parameters such as pH, nitrogen (N), phosphorus (P), and potassium (K), along with environmental factors including temperature, humidity, and rainfall. Additional attributes such as crop type and historical yield data may also be incorporated to enhance prediction performance. These features provide a comprehensive representation of agricultural conditions.

B. Data Pre-Processing

Data preprocessing is performed to ensure data quality and consistency. Missing values in the dataset are using statistical imputation techniques such as mean or median substitution. Noise and outliers are identified using methods such as Z-score and Interquartile Range (IQR) and are appropriately treated. Furthermore, feature scaling techniques such as Min-Max normalization or standardization are applied to maintain uniformity across all input variables. If categorical variables are present, encoding techniques such as label encoding or one-hot encoding are utilized.

C. Feature Engineering

Feature engineering is carried out to enhance the predictive capability of the model. Derived features such as soil fertility index are computed using soil nutrient values (NPK) and pH levels. Rainfall trends are analyzed by aggregating temporal data to capture seasonal patterns. Additionally, interaction features are generated by combining soil and environmental parameters to model complex relationships. Other derived indicators, such as nutrient ratios and temperature-humidity index, are also included to improve model learning.

Fig 3.1 : Random Forest Accuracy

Random Forest Accuracy: 0.9931818181818182

	precision	recall	f1-score	support
0	1.00	1.00	1.00	23
1	1.00	1.00	1.00	21
2	1.00	1.00	1.00	20
3	1.00	1.00	1.00	26
4	1.00	1.00	1.00	27
5	1.00	1.00	1.00	17
6	1.00	1.00	1.00	17
7	1.00	1.00	1.00	14
8	0.92	1.00	0.96	23
9	1.00	1.00	1.00	20
10	0.92	1.00	0.96	11
11	1.00	1.00	1.00	21
12	1.00	1.00	1.00	19
13	1.00	0.96	0.98	24
14	1.00	1.00	1.00	19
15	1.00	1.00	1.00	17
16	1.00	1.00	1.00	14
17	1.00	1.00	1.00	23
18	1.00	1.00	1.00	23
19	1.00	1.00	1.00	23
20	1.00	0.89	0.94	19
21	1.00	1.00	1.00	19
...				
accuracy			0.99	440
macro avg	0.99	0.99	0.99	440
weighted avg	0.99	0.99	0.99	440

D. Model Development

The proposed framework employs the XGBoost algorithm as the core learning model due to its high efficiency and accuracy with structured data. XGBoost is an ensemble learning technique based on gradient boosting of decision trees, incorporating regularization to prevent overfitting. The model is trained using the processed dataset with appropriate objective functions depending on the task. For crop yield prediction, a regression objective is used, whereas classification objectives are applied for tasks such as crop recommendation or soil fertility classification. Hyperparameters such as learning rate, maximum tree depth, number of estimators, and subsampling ratio are optimized using techniques like grid search and cross-validation to improve model performance.

E. Multi-Task Learning Framework

The proposed system adopts a multi-task learning approach to handle multiple agricultural prediction tasks simultaneously. These tasks include crop yield prediction, soil fertility classification, and resource requirement estimation. This is achieved either by training separate models for each task or by sharing a common feature representation with task-specific outputs. The multi-task approach improves computational efficiency and enhances overall system performance.

F. Model Evaluation

The performance of the proposed model is evaluated using standard evaluation metrics. For regression tasks, metrics such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and coefficient of determination (R^2) are used. For classification tasks, accuracy, precision, recall, and F1-score are considered. The results are compared with baseline models such as Decision Trees and Random Forests to validate the effectiveness of the proposed approach

.

G. Deployment and Decision Support

The trained model is integrated into a decision support system that assists farmers in making informed decisions. The system provides crop recommendations based on soil and climatic conditions, predicts crop yield, and suggests optimal resource utilization strategies, including water and fertilizer usage. The system can be deployed as a web-based or mobile-based application for real-time accessibility.

4. IMPLEMENTATION AND RESULTS

The proposed system for crop yield prediction and resource optimization is implemented using Python, leveraging its extensive ecosystem for data analysis and machine learning. The implementation integrates data preprocessing, model development, evaluation, and visualization into a unified pipeline to ensure efficient and accurate prediction. Initially, the agricultural dataset, which includes parameters such as soil characteristics, weather conditions, and historical crop yield data, is preprocessed using Pandas and NumPy. Data cleaning techniques are applied to handle missing and inconsistent values, followed by feature selection to retain relevant attributes. Furthermore, normalization and scaling of features are performed using Scikit-learn to improve model stability and performance. Subsequently, multiple machine learning models, including Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), and XGBoost, are implemented for comparative analysis. The dataset is divided into training and testing sets in an 80:20 ratio. Each model is trained using the training dataset, and hyperparameter tuning is conducted to optimize performance, particularly for the XGBoost model, which employs a gradient boosting approach to enhance predictive accuracy. The performance of the models is evaluated using accuracy as the primary metric. As illustrated in Fig. 1, Random Forest achieved the highest accuracy of 0.99, followed by Decision Tree (0.98), SVM (0.97), and Logistic Regression (0.96). These results indicate that ensemble-based models outperform linear models due to their ability to capture complex, non-linear relationships in agricultural data.

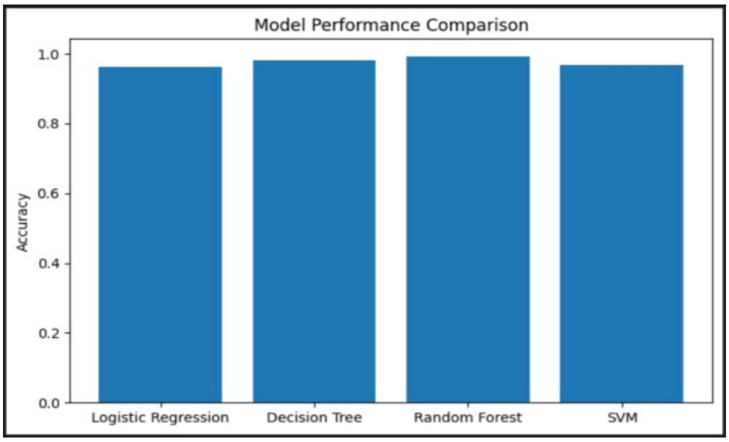


Fig 4: Model Performance Comparison

5. Discussion

The superior performance of Random Forest can be attributed to its ensemble learning approach, where multiple decision trees are combined to reduce overfitting and improve generalization. This makes it particularly effective for handling complex and non-linear relationships in agricultural datasets, such as those involving soil properties, weather patterns, and crop characteristics. The Decision Tree model also performed well due to its ability to capture non-linear decision boundaries. However, its slightly lower accuracy compared to Random Forest suggests a tendency toward overfitting when used as a standalone model. SVM demonstrated strong performance, indicating that the dataset is reasonably separable in a high-dimensional space. However, its performance is slightly lower than ensemble methods, possibly due to sensitivity to parameter tuning and kernel selection. Logistic Regression, being a linear model, achieved comparatively lower accuracy. This suggests that the underlying relationships in the dataset are not purely linear and require more complex odelling techniques.

Soil Condition Evaluation: Utilizing machine learning models, soil parameters including Ph level, moisture content, and nutrients, are analyzed to suggest the best crop and fertilizer combinations. This approach maximizes crop production as plants receive the right nutrients. Pest and Disease Detection: Modern computer vision models deployed on drones and satellites employ image analysis and recognize damage by pests or diseases, allowing for narrow window of intervention which significantly reduces damages and use of chemicals. Crop Growth Monitoring: Plant growth stages are monitored, predicted, and analyzed through time-series data and imageries, enabling yield prediction and necessary changes, for instance, water or nutrient supply adjustment to achieve best performance.

6. CONCLUSION

The system demonstrates strong predictive capabilities on agricultural datasets. The experimental results indicate that ensemble-based models significantly outperform traditional approaches. In particular, Random Forest achieved the highest accuracy of 0.99, followed by Decision Tree (0.98), Support Vector Machine (0.97), and Logistic Regression (0.96). These findings highlight the importance of using non-linear and ensemble techniques to capture complex relationships among agricultural parameters such as soil, weather, and crop conditions. Additionally, the adoption of XGBoost further strengthens the framework by providing a robust and scalable boosting mechanism for improved prediction performance. the proposed approach provides a reliable and accurate solution for crop yield prediction, supporting data-driven decision-making in agriculture. Future work may focus on incorporating real-time IoT data, expanding dataset diversity, and further optimizing model performance to enhance practical applicability in precision agriculture systems.

REFERENCES

1. R. Jaykumar et al., “Enhancing crop yield and resource efficiency with machine learning: Smart agriculture solution,” in Proc. Parul Univ. Int. Conf. Eng. Technol. (PiCET), 2025, doi: 10.1049/icp.2025.1274.
2. S. M. Shawon et al., “Crop yield prediction using machine learning: An extensive and systematic literature review,” Smart Agricultural Technol., vol. 10, 100718, 2025.
3. M. Lamichhane, S. Mehan, and K. R. Mankin, “Surface soil moisture prediction using multimodal remote sensing data fusion and machine learning algorithms in semi-arid agricultural region,” Science of Remote Sensing, vol. 12, 100255, 2025.
4. Ben Ghorbal et al., “Sustainable soil organic carbon prediction using machine learning and the ninja optimization algorithm,” Frontiers in Environmental Science, vol. 13, 1630762, 2025.
5. M. A. Jabed and M. A. A. Murad, “Crop yield prediction in agriculture: A comprehensive review of machine learning and deep learning approaches, with insights for future research and sustainability,” Heliyon, vol. 10, e40836, 2024.
6. T. Petropoulos et al., “Interpretable machine learning for legume yield prediction using satellite remote sensing data,” Appl. Sci., vol. 15, no. 13, p. 7074, 2025.
7. G. Saha et al., “Smart IoT-driven precision agriculture: Land mapping, crop prediction, and irrigation system,” PLoS One, vol. 20, no. 3, e0319268, 2025.
8. Jaiswal, T. V. Kumar, and C. Shukla, “Smart drip irrigation systems using IoT: a review of architectures, machine learning models, and emerging trends,” Discover Agriculture, vol. 3, 253, 2025.
9. J. Yuan et al., “Grain crop yield prediction using machine learning based on UAV remote sensing: A systematic literature review,” Drones, vol. 8, no. 10, p. 559, 2024.
10. J. Botero-Valencia et al., “Machine learning in sustainable agriculture: Systematic review and research

perspectives," Agriculture (MDPI), vol. 15, no. 4, 377, 2025.

11. Malhotra, K. & Firdaus, M., 2022. Application of Artificial Intelligence in IoT Security for Crop Yield Prediction, s.l.: Birsa Agricultural University ss(BAU).
12. Nagesh, O. S. et al., 2024. Boosting enabled efficient machine learning technique for accurate prediction of crop yield towards precision agriculture. Discover Sustainability, pp. 1 9.
13. Pukrongta, N., Taparugssanagorn, A. & Sangpradit, K., 2024. Enhancing Crop Yield Predictions with PEnsemble 4: IoT and ML-Driven for Precision Agriculture. Applied Science, pp. 1-29.