

A Hybrid Machine Learning Approach to Predictive Analytics in Precision Agriculture for Sustainable Crop Optimization

Jignesh Kariya and Pankaj Mudholkar

Faculty of Computer Applications, Marwadi University, Rajkot, India

Article history

Received: 01-07-2026

Revised: 06-07-2026

Accepted: 08-09-2026

Corresponding Author:

Jignesh Kariya

Marwadi University, Rajkot, India

Email:

jignesh.kariya@marwadieducation.edu.in

Abstract: Driven by data-enabled technologies, precision agriculture is evolving into an innovative way to increase agricultural productivity and sustainability. Conventional agricultural practices face several challenges, such as poor soil quality, inappropriate crop selection, improper resource utilization, and unpredictable climate conditions. To address these limitations, this paper proposes a hybrid machine learning framework for sustainable crop optimization and predictive analysis in precision agriculture. Agricultural datasets containing features like nutrients in the soil, temperatures, humidity, rainfall, and other parameters related to pH values are utilized to build a smart crop forecasting and recommendation system. Data preprocessing and normalization are applied to improve data quality and reduce inconsistency. K-Means clustering is applied to uncover hidden patterns in soil and environmental conditions, while Random Forest and XGBoost are used to generate accurate crop forecast and recommendations. The experimental results indicate that the hybrid model is better than existing machine learning approaches in terms of prediction accuracy, precision, recall, and F1-score with a prediction accuracy of 97.8%. In addition, sustainable agriculture is facilitated by efficient fertilizer usage, minimized manual labor, and efficient water usage. The suggested hybrid learning strategy presents an intelligent and scalable way forward for the current era of precision agriculture. The findings may help policymakers, farmers, and agricultural scientists implement data-driven methods to enhance agricultural yields.

Keywords: Precision Agriculture, Predictive Analytics, Smart Farming, Random Forest, XGBoost, Crop Prediction, Agricultural, Sustainable Agriculture, K-Means

Introduction

Agriculture is extremely important for food security and economic stability (Ashfaq et al., 2024). The rapid increase in the population size, unpredictability of the climate, exhaustion of natural resources, and increasing environmental challenges place great pressure on existing agricultural practices (Apu, 2025). Old-fashioned agriculture practices are mostly dependent on visual observations, experience, and generic farming practices that often lead to wastage of water, fertilizers, and soil (Bali and Singla, 2021). This generally ends up in reduced agricultural yield,

degradation of the soil, and increased operational costs (Alam, 2025; Jain et al., 2024).

Precision agriculture has emerged as one of the new approaches in farming that utilizes information technology, smart technologies, and data analysis to increase agricultural production. Through analyzing environmental factors, soil conditions, climatic conditions, and the needs of crops, precision agriculture enables farmers to make informed decisions (Chen et al., 2023). Predictive analytics and machine learning in agriculture have played a significant role in crop performance prediction and optimizing agricultural practices (Hoque et al., 2024).

Machine learning-based methods enable the use of intelligent computational models that can find complex relations among the agronomical data. Machine learning techniques are used in the detection of diseases, crop suggestion, yield prediction, irrigation management, and efficient resource utilization (Jain et al., 2024). Decision Tree, Support Vector Machine, Random Forest, Artificial Neural Networks, and Gradient Boosting techniques are some of the most widely used machine learning techniques in intelligent agriculture (Kumar and Sharma, 2023).

Even with all these advancements, many current predictive models for agriculture depend on single-model architecture, and they often have difficulty dealing with different climatic conditions and agricultural data that change from time to time (Cedric et al., 2022). This is because agricultural ecology is inherently complex, and the amount of yield produced is affected by various interrelated factors such as rainfall, temperature, humidity, availability of nutrients in the soil, acidity, and climatic variations (Manjunath and Palayyan, 2023).

As such, hybrid learning models in machine learning have been widely considered in the field of precision agriculture due to these constraints. The purpose of these hybrid models is to combine all the best attributes of different machine learning algorithms to enhance their predictive ability and robustness (Mahmud et al., 2024). Clustering methods and ensemble learning will enhance the efficiency of crop optimization strategies and detect agricultural patterns (Petropoulos et al., 2025).

The research presents an intelligent analytical approach through machine learning techniques that involves K-means clustering, Random Forest and XGBoost algorithms in the context of precision agriculture. The objective of this model is to increase the accuracy of predicting crops, enhance utilization of agricultural resources, and foster sustainable farming methods (Nishant et al., 2020). The K-means approach is employed for pattern recognition in soil and environmental factors, whereas Random Forest and XGBoost algorithms are utilized to enhance classification and prediction (Suruliandi et al., 2021).

Sustainable agriculture is ensured by the proposed system through effective water management, efficient utilization of fertilizers, and accurate crop selection according to environmental conditions. Furthermore, the intelligent decision support system helps farmers choose the right crops and increase total agricultural output while reducing environmental effect (Yan et al., 2025).

Aims of this study are as follows:

- ✓ Development of hybrid machine learning algorithm which will assist precision farming
- ✓ Improvement of crop prediction through predictive analysis
- ✓ Sustainability of resources in farming

- ✓ Decision support system that will optimize the cropping system
- ✓ Application of agricultural datasets to test the hybrid machine learning algorithms' performance

Results obtained from the research are going to make significant contributions towards sustainable farming and intelligent techniques in farming. The proposed framework is an intelligent solution for the current agriculture by the combination of machine learning and data-driven decisions.

Literature Review

In the process of developing the framework for forecasting the yield of crops in Pakistan, Badshah et al. (2024) found that SVR performed better in forecasting the yield of wheat, while RF was the most accurate in classifying crops.

In their study on a predictive crop classifier from an ICAR dataset consisting of seven soil and weather factors, Mahmud et al. (2024) recommended an RF-Genetic Algorithm-based predictive crop model that gave a success rate of 99.3%.

Haider et al. (2024) an ensemble machine learning approach known as RFXG (Random Forest+XGBoost), which utilized meteorological parameters for forecasting yield in cotton crops in Pakistan, with an RMSE value of 0.05, significantly outperforming the conventional XGBoost (RMSE = 0.07).

The research conducted by Ashfaq et al. (2024) on predicting wheat production in Punjab, Pakistan, was performed with machine learning-based models through fused climate, NDVI, satellite, and soil data collected from Google Earth Engine (2017-2022). With three models including SVM, Random Forest, and LASSO, R^2 was obtained from 0.74 to 0.88 with superiority of Random Forest to other models.

Hoque et al. (2024) proposed a ML model for predicting crop yields in India using data from FAO and World Data Bank (1990-2013). In comparison to the K-NN and MLR models, the Gradient Boosting model performed best with an R^2 of 99.98%, beating MLR (96.13%) and K-NN (97.29%).

According to Sharma et al. (2023), a hybrid machine learning model was developed to predict the yield of crops in India by analyzing the agriculture data of years 1997-2020. The variables considered in the study were rainfall, area, production, and weather. Models used in this research included decision tree, random forest, xgboost, CNN, and LSTM. Random forest gave maximum accuracy with 98.96% with RMSE of 2.45 while CNN minimized the prediction error to 0.00060.

Prediction of Irish potatoes and maize yields in Rwanda through machine learning techniques was performed by Kuradusenge et al. (2023), based on

temperature and rainfall values from 2006 to 2021. Random Forest yielded the highest R^2 value of 0.875 as compared to Polynomial Regression and SVR for predicting Irish potatoes' yield. Polynomial Regression gave the highest R^2 of 0.817 for maize yield.

Based on the Government datasets from 1997 to 2019, Jhahharia et al. (2023) proposed a machine learning (ML) and deep learning approach for predicting the yields of five essential crops in Rajasthan, India. In all cases where gradient descent, SVM, LASSO, LSTM, and Random Forest algorithms were used, Random Forest algorithm proved superior to other models.

With data from agriculture over 18 years, Iniyan et al. (2023) formulated an ML model to predict the yield of crops in India by incorporating parameters such as climatic conditions, type of soil, rain, humidity, and temperature. In the comparison of MLR, Gradient Boosting, Decision Tree, Elastic Net, and LASSO regression models, LSTM with feature engineering gave the best results of 86.3%.

Cedric et al. (2022) employed FAO and World Bank climate data from 1990 to 2020 for the development of the ML-based crop yield prediction model for six significant crops in West Africa using the weather conditions, pesticides, and agricultural attributes. In this context, the Decision Tree outshines KNN (93.15%) and Multivariate Logistic Regression (89.78%) with maximum R^2 of 95.3% (Iniyan et al., 2023).

Incorporating the regional agriculture data of Algeria, Chergui (2022) developed the wheat yield prediction models using ELM, SVR, DNN, Random Forest, and ANN. The importance of selecting model region-wise can be realized through the fact that Random Forest performed the best for Setif province (RMSE = 0.05 q/ha), while DNN outperformed other models for Constantine province (RMSE = 0.04 q/ha).

Incorporating 470 samples collected during the period of 18 years (1998-2015) in 28 districts of Tamil Nadu, Joshua et al. (2022) developed a machine learning-based rice production prediction model, which included features such as temperature, rainfall, evapotranspiration, and fertilizer inputs. In comparison to GRNN, BPNN, and SVM, GRNN model performed better among all with highest accuracy of prediction of 97%.

In order to predict the winter wheat production in Hengshui, Hebei Province, China, Di et al. (2022) introduced BO-LSTM, which is an approach of Deep Learning combined with Bayesian Optimization. The results of BO-LSTM performed better than SVM based on data collected by the Chinese Academy of Agricultural Sciences.

By making use of the soil and environmental factors, Suruliandi et al. (2021) proposed the development of a crop forecasting model using classifiers such as SVM, KNN, Naïve Bayes, Random Forest, Decision Tree, and Bagging along with wrapper methods of feature selection

such as RFE, Boruta, and SFFS. The highest performance was achieved by employing RFE and Bagging classifier with accuracies of 89, 56, and 92% respectively.

Based on publicly available FAO and World Data Bank data, Pant et al. (2021) developed an ML crop yield prediction model for maize, rice, potatoes, and wheat with parameters like rainfall, temperature, pesticide, and country. The Decision Tree Regressor achieved the highest score of 96% compared to Gradient Boosting, SVM, Random Forest, and Decision Tree models.

The deep learning architecture that was utilized by Bali and Singla (2021) for forecasting wheat production in Punjab, India, was created based on the meteorological information between 1970 and 2012. This includes data about precipitation, temperature, humidity, and winds. The Random Forest model possessed the lowest MAE at 449.36 compared to ANN, MLR, Random Forest, and RNN-LSTM models. However, the best RMSE value was achieved by RNN-LSTM at 147.12.

Based on the factors of meteorology and soil conditions as input parameters, Gowda and Reddy (2020) proposed the use of the Random Forest crop yield prediction model that was better than other models like Decision Tree (74%) and Polynomial Regression (88%).

Nishant et al. (2020) proposed a machine learning-based approach for Indian crop yield estimation using ENet, Lasso, and Kernel Ridge regression techniques. After employing layered regression with preprocessing, errors of each individual model were reduced to below 1% from 4, 2, and 1%, respectively.

Using a dataset based on the state of Maharashtra having data such as humidity, temperature, pressure, and wind speed, Sivanandhini and Prakash (2020) developed an architecture for predicting the yield of wheat crops by applying both FNN and RNN. RNN performed better than FNN, showing RMSE at 0.00011 and loss of 0.0103, whereas the FNN's RMSE was 0.23.

With LSTM, RNN, and SVM with the help of a Kaggle dataset, Sonal Agarwal and Tarar (2021) developed a machine learning and deep learning-based agriculture yield prediction model which can predict the yield for nine different crops such as wheat, rice, and maize with 97% accuracy as opposed to 93% in the case of DT, RF, and ANN.

In summary from Table 1, the reviewed papers point to the increasingly common use of machine learning algorithms in agriculture for predicting yields. Between the years of 2020 and 2024, literature has found that the Random Forest algorithm and XGBoost perform best across multiple crops and regions, such as wheat production in Pakistan and maize production in Rwanda. The use of multiple data sources, including remote sensing, climate factors, NDVI, and soil measurements, has been shown to have a positive effect on the accuracy of predictions.

Table 1: Literature review summary table

Sr. No.	Purpose	Author's Name	Year	Methods / Algorithm	Dataset	Findings
1	To forecast the crop yield in Pakistan region	Badshah et al.	2024	Support Vector Regression (SVR), Random Forest (RF),	kaggle and World Bank & FAO	RF achieved the highest accuracy in crop classification; SVR outperformed others for wheat yield prediction
2	To predict crop yield	Mahmud et al.	2024	Random Forest (RF) + Genetic Algorithm	Indian Council of Agricultural Research (ICAR).	RF combined with Genetic Algorithm achieved the best accuracy of 99.3%
3	Apply ensemble machine learning methods to predict Pakistan's cotton crop yield from weather features.	Haider et al.	2024	Random Forest Extreme Gradient (RFXG)	Crop Reporting Service, Computerized Data Processing Centre, and Pakistan Meteorological Department	The proposed RFXG algorithm is better than the regular Extreme Gradient Boost (XGB) model, whose RMSE is 0.07, by 0.05 RMSE using the optimization method.
4	This research seeks to determine the exact yield of wheat through the fusion of NDVI data and climate data using machine-learning.	Ashfaq et al.	2024	RF	https://power.lar.nasa.gov/data-access-viewer/	RF performs well with $R^2 = 0.74 - 0.88$.
5	Crop yield forecast through ML based on the combination of pesticide and meteorology data.	Hoque et al.	2024	Multivariate, Logistic Regression, Gradient Boosting, K-NN and Logistic Regression	Govt. Website	The value of R^2 for YGBR is 99.98% while the value of R^2 for YMLR and YK-NN is 96.13% and 97.29% respectively
6	The use of Deep Learning and Machine Learning regression for estimating agricultural yields.	Sharma et al.	2023	Decision Tree, LSTMNetwork, CNN Network, RF, and XG Boost Regression	https://aps.dac.gov.in . and https://data.gov.in	The accuracy of RF is maximum i.e. 98.96%.
7	Crop Yield Estimation such as Maize and Irish potatoes with the help of ML.	Kuradusenge et al.	2023	Random Forest, Support vector Regressor, Polynimial Regression	NISR	The performance of RF is high compared to other algorithms. The degree of correlation between real values and predicted yield was

						calculated using R2. The obtained value was 0.875, 0.773, and 0.560 (RF, PL, and SVR) for Irish Potato and 0.817, 0.716, and 0.549 for maize. The Random Forest algorithm determines which crop production is the best.
8	Applied DL and ML algorithms to predict crop production levels. .	Jhajaria et al.	2023	Gradient Descent, Random Forest, SVM, Lasso regression, and LSTM	Rajasthan Government's official website	
9	To forecast the agricultural yields with the help of ML.	Iniyan et al.	2023	LASSO, Decision Tree Regression, Gradient Boosting, Elastic Net, and Multiple Linear Regressions	Govt. Website	The most effective model is LSTM, which has an accuracy of 86.3%.
10	Crop yields for "West African Countries" are predicted using machine learning.	Cedric et al.	2022	MLR, decision trees, and KNN models	https://climateknowledgeportal.worldbank.org/ and http://www.fao.org/faostat/en/data	Decision tree algorithm performed well with a coefficient of determination (R2) of 95.3%
11	This model's goal is to use machine learning to estimate the yield of "Durum Wheat."	Chergui	2022	RF, DNN, ANN, SVR, and Extreme Learning Model	"Province of Setif" and "Province of Constantine" departments of agricultural services	For, DNN: RMSE = 0.04 q/ha, and For, RF: RMSE = 0.05 q/ha.
12	Calculate agricultural production using a variety of machine-learning techniques.	Joshua et al.	2022	SVM, GRNN, BPNN	Tamilnadu's agricultural department, Tamilnadu's statistics department, the Tata-Cornell Institute for Agriculture and Nutrition (TCI), the Regional Meteorological Centre in Chennai, and	With a 97% accuracy rate, GRNN works well.
13	in order to create a framework that integrates deep learning and Bayesian optimization for forecasting winter wheat.	Di et al.	2022	BO – LSTM	Year Book of Statistics and Chinese Academy of Agricultural Sciences	BO-LSTM (R2 = 0.82, RMSE = 177.84 kg/ha)
14	Crop yield prediction based on soil	Suruliandi et al.	2021	Boruta, RFE, SFFS	Agricultural Department of Sankarankovil Taluk, Tenkasi and www.tnau.ac.in	RFE with bagging accuracy of 89% for soil characteristics.

	and environmental conditions using feature selection methods					RFE with bagging accuracy of 56% for environment characteristics.
15	Analyze agricultural crop yield prediction using statistical techniques of machine learning.	Pant et al.	2021	Gradient Boosting regressor, Decision Tree regressor, Random Forest regressor, and SVM.	FAOSTAT and world data bank	Decision tree regressor is fitted to the highest score of 96%.
16	Estimating the wheat crop yield using the help of Deep Learning in the Punjab Region.	Bali et al.	2021	RNN, RNN-LSTM, RF, ANN, Multivariate Linear Regression.	economic advisor to Government and meteorological department of Punjab and .	MLR (RMSE=915.64, MAE=796.07) RF (MAE=449.36, RMSE=540.88), ANN (MAE=623.13, RMSE=732.14), and RNN-LSTM: (MAE = 623.13, RMSE = 147.12) 97%
17	Apply a hybrid approach that integrates deep learning and machine learning to predict the agricultural output.	Agarwal et al.	2021	SVM as ML. RNN and LSTM as DL,	Govt. Website	
18	to use machine learning to predict agricultural crop output.	Gowda et al.	2020	Polynomial Regression, Random Forest, and Decision Trees	Govt. Website	The accuracy of polynomial regression is 88%, decision tree regression is 74%, and RF regression is 97%.
19	Apply layered regression for machine learning to predict the agricultural output using Indian agriculture.	Nishant et al.	2020	ENet, Lasso, and Kernel Ridge. Stacked regression has been completed.	Govt. Website	With an inaccuracy of less than 1%, stacked regression performed admirably.
20	to apply FFNN and RNN to analyze crop output.	Sivanandhini et al.	2020	FFNN model and RNN model.	From the online sources.	RNN(RMSE: 0.00011, Loss: 0.0103), FNN(RMSE: 0.232, Loss: 0.1141)

Additionally, deep learning techniques like LSTM and CNN prove highly effective in analyzing agricultural data. Nevertheless, current research is limited to applying a single model or combining them loosely. It does not provide examples of using both clustering and prediction

algorithms in a hybrid approach to maximize the sustainability of the agricultural process. To bridge this research gap, this current study proposes integration between the K-means algorithm for zone division and Random Forest and XGBoost for yield classification,

which has not been considered in any of the above literature reviews.

Research Gap and Problem Statement

Machine Learning approaches have found their place in precision farming, which has greatly enhanced agriculture predictions, crop recommendation, and environmental sensing. Specific machine learning approaches like Decision Tree, Random Forest, Support Vector Machines, Artificial Neural Networks, and XGBoost have been proven through past research to perform well within agricultural prediction systems. These approaches have led to improved predictions for crop yields, irrigation scheduling, and soil composition. Nonetheless, considerable limitations still exist which have hindered the efficiency and adaptability of existing agriculture decision support systems.

Most existing models for agriculture prediction use individual machine learning techniques, and this could limit their ability to consider the complex interaction among agricultural parameters. Rainfall, humidity, temperature, nutrient levels, soil pH, and seasonality are all considered in determining crop yield. Individual learning techniques may find it difficult to handle the diversity of agricultural data, thus producing contradictory predictions due to changes in climate conditions.

A further important issue that was observed in earlier systems is inadequate consideration of hidden soil patterns and environmental patterns. It is well-known that interaction and clustering are frequent aspects in agricultural data, which cannot be effectively addressed by using only classification methods. Inadequate application of clustering often leads to poor accuracy levels and constraints on detecting suitable crops for an area.

In addition, some traditional farm support systems cannot come up with intelligent solutions for optimal use of sustainable resources. Poor crop choice, inefficient water use, and overuse of fertilizers are among the problems facing modern farming systems from both an economic and environmental perspective. Modern models do not give priority to sustainable agricultural practices and focus only on accuracy of predictions.

Based on the study, some agricultural forecasting systems suffer from the following problems:

- High levels of computational complexity
- Inability to scale large sets of agricultural data

- Limited ability to adapt to climate variations
- Insufficient incorporation of predictive analytics and decision-support systems

In order to overcome these issues, an intelligent hybrid framework in machine learning technology that integrates clustering techniques and ensemble learning techniques is necessary.

In the context of precision farming, the proposed research study incorporates a hybrid model of predictive analytics using Random Forest, XGBoost, and K-Means clustering techniques. The use of clustering together with ensemble learning allows improved agricultural pattern recognition, better prediction ability and improved techniques for crop selection to be achieved. Moreover, this system also encourages sustainable agriculture practices due to efficient resource management and decision-making.

The research gaps identified in existing precision agriculture systems are presented in Table 2, together with possible solutions proposed by this research. Past studies have indicated that the agricultural predictive models suffer from numerous weaknesses, such as forecast inaccuracies, poor adaptability to the environment, lack of sustainability, and intelligence decision making.

Prediction models for crops based on machine learning that mainly use a single method are discussed as the first gap found. Even though individual algorithms such as Decision Tree, Support Vector Machine, and Random Forest have shown sufficient performance individually, they often cannot achieve high consistency in their forecasting performance in a wide range of agricultural situations. To resolve this problem, the new research introduces an approach that combines several approaches using machine learning.

Insufficient analysis of the soil pattern comes in second place. Most agricultural databases will have certain correlation patterns that exist among soil nutrients, climate conditions, and how the crop behaves. Traditional approaches tend to neglect such clusters of environmental data, leading to ineffective crop suggestions. The proposed framework resolves this issue by applying K-Means clustering before conducting prediction analysis.

The inability to predict well in varying conditions is yet another major disadvantage found in previous studies. Variations in temperature, rainfall, and season affect the agriculture ecosystem considerably.

Table 2: Key research gap identified

Research Area	Existing Limitation	Proposed Solution
Crop Prediction Models	Dependence on single ML algorithms	Hybrid ML framework
Soil Pattern Analysis	Poor identification of hidden clusters	K-Means clustering integration
Prediction Accuracy	Reduced performance under dynamic conditions	Ensemble learning optimization
Sustainable Farming	Limited resource optimization support	Intelligent decision support system
Agricultural Scalability	Difficulty handling large datasets	Scalable hybrid architecture
Environmental Adaptability	Weak response to climatic variability	Adaptive predictive analytics

If such environmental elements vary frequently, it might affect the output generated by conventional models. Proposed approach is using ensemble learning approach based on Random Forest and XGBoost to improve predictability.

It can be seen from the table that the sustainability-based decision-making procedure is lacking. The current approach is more inclined towards yield forecasting than resource optimization. This leads to overuse of water, misuse of fertilizers, and degradation of the environment. The proposed intelligent decision-making system is intended to optimize the agricultural resources through the use of sustainable agriculture techniques.

Scalability is yet another key feature of earlier agricultural prediction systems, which has been studied by scholars. Creating a technique for working with large amounts of agricultural data is vital. It is difficult to work with agricultural data through conventional means once it has several dimensions.

Moreover, one of the biggest problems that precision agriculture technology continues to encounter is that of environmental adaptability. The traditional approaches have often been unable to adequately adapt to the rapid changes in climatic conditions and the changing environmental settings. Through the application of the proposed system, which integrates clustering and ensemble learning techniques, environmental adaptability can be addressed.

Problem Statement

Innovative and data-driven approaches are needed for the development of modern agriculture systems since they could enhance the production capacity of the crops while maintaining environmental sustainability. However, conventional agriculture practices and traditional prediction models often cannot provide adequate crop suggestions due to the complexity of the environment dependency and the diversity of agricultural settings. Improper crop production, misuse of fertilizer, high consumption of water resources, and degradation of the soil can occur as a result.

Moreover, the majority of present-day machine learning models for precision agriculture operate on their own, ignoring clustering-based pattern identification of the environment and ensemble learning optimization.

Therefore, the key problem addressed by this study is the development of a hybrid machine learning framework that is capable of:

- ✓ Accurately analyze agricultural statistics
- ✓ improving forecasting of crops
- ✓ optimize resource usage
- ✓ Promote sustainable precision agriculture methodologies

Through the use of advanced machine learning algorithms, the proposed hybrid predictive analytics framework aims at providing an intelligent, scalable, and sustainable model for making decisions in agriculture.

Objectives of the Research

The primary objective of this study is to design an intelligent hybrid machine learning model for predictive analytics of precision agriculture, thereby enhancing crop optimization and sustainability in agriculture. The purpose of this system is to increase productivity through decision-making based on data and intelligent algorithms.

In today's world of agriculture, huge amounts of data are produced, which have connections with soil, climate, crop, irrigation, and other types of information. There must be efficient computer programs, which would be capable of discovering patterns within large volumes of data and making appropriate predictions. Thus, the aim of the research is to create an effective model of predictive analytics, which will assist in making better crop decisions.

In this connection, the research focuses on incorporating the use of ensemble learning and clustering methods which could help enhance resource optimization and prediction ability. The system employs K-means clustering, random forest, and XGBoost algorithms for resolving the limitations of conventional standalone machine learning models in smart farming applications.

Development of a Hybrid Machine Learning Framework

One of the most important achievements associated with the study is the development of an intelligent hybrid learning framework. This framework integrates clustering techniques along with ensemble learning in order to enhance predictive analytics capabilities and decision-making skills in the field of precision agriculture. An agricultural setting is extremely dynamic and comprises several dependent variables including soil nutrient content, precipitation, humidity, temperature, and climate. Conventional methods used for predictions usually find it difficult to cope with the complexities involved.

The proposed design will incorporate three different algorithms, including XGBoost, Random Forest, and K-Means clustering algorithms in one predictive analysis system. All the mentioned algorithms have their own role within the system to enhance its predictive power.

Agricultural data collection starts by obtaining data from reliable sources such as agricultural libraries where there exists information about crops, rainfall, soil composition, climatic conditions, and weather-related issues. Preprocessing is important for improving the quality of the data because raw agricultural data could have various problems such as missing entries, noise, duplicate entries, and unbalanced data distribution.

Missing or corrupted entries could be eliminated from the data using data cleaning. All features are normalized in order to make sure that the numerical entries remain consistent. This is achieved through data scaling.

After preprocessing, the methodology clusters the agricultural data based on similarities in environmental conditions and soil nutrients by means of the K-Means clustering technique. Agricultural patterns that may not be easily observable from the original data can possibly be identified by performing clustering. Clustering the data aids predictive models to recognize regional agriculture patterns.

After clustering, the agricultural data will then be aggregated and processed by the ensemble learning method. As the Random Forest model is highly adept in dealing with various types of datasets and reducing the effects of overfitting, it will be applied in the classification process.

XGBoost is used in the framework as an optimization technique to increase prediction accuracy. Using a gradient boosting technique that reduces prediction errors and increases efficiency in the learning process, XGBoost improves classification. In comparison with machine learning methods alone, the integration of Random Forest and XGBoost in the framework boosts prediction accuracy.

Intelligent decision support systems is the final stage of the framework. Intelligent decision support will suggest crops for planting, depending on climatic conditions, soil quality, and environment, from predictions made earlier. The framework also promotes sustainable agriculture through efficient resource management without using unnecessary fertilizers and water.

The proposed hybrid architecture improves:

- Prediction accuracy
- Environmental adaptability
- Resource optimization
- Computational efficiency
- Intelligent agricultural decision-making

The framework is designed to support scalable agricultural applications and can be extended for real-time

precision farming systems integrated with IoT devices and sensor networks.

Table 3 illustrates the major components involved in the proposed hybrid machine learning framework and their corresponding functionalities within the precision agriculture system.

Gathering agricultural data, such as soil nutrients, rainfall, humidity, temperature, and crop-related characteristics, falls under the purview of the first component. For the development of reliable prediction models, accurate and diversified datasets are very important.

The second phase will concentrate on data preparation and will involve the application of techniques such as data cleaning and normalization. Datasets in agriculture are commonly plagued by the problem of missing values and anomalies and these may lead to inaccurate predictions. Data preparation guarantees improved quality of data and proper training of the model.

Subsequently, the next important component of the suggested approach will be feature selection. The feature selection technique will be used to identify the most relevant agricultural features that influence the growth of crops and to eliminate all other irrelevant or duplicate features.

The agricultural data that possesses similar environmental features would be clustered with each other in the clustering process through the use of the K-Means clustering algorithm. Such an approach will make the predictive analytics process more effective by identifying some trends within the soil and climate data.

Following the clustering, the next process in the framework is prediction. The Random Forest approach has been chosen because it is robust and efficient for handling large amounts of agricultural data.

The incorporation of XGBoost in the second layer of the prediction system increases the performance of the algorithm by making fewer errors during classification and making more efficient learning processes using gradient boosting techniques.

The intelligence recommendation system of the framework is the decision support system for the farmers. This is because, using predictive analysis, the framework recommends the best crop varieties for farmers to plant.

Table 3: Components of the proposed hybrid machine learning framework

Component	Technique used	Function in framework
Data Collection	Agricultural Dataset	Gather soil, climate, and crop data
Data Preprocessing	Cleaning and Normalization	Improve data quality and consistency
Feature Selection	Attribute Optimization	Select important agricultural parameters
Clustering Layer	K-Means Clustering	Identify hidden environmental patterns
Prediction Layer 1	Random Forest	Perform crop classification
Prediction Layer 2	XGBoost	Optimize predictive accuracy
Decision Support System	Intelligent Recommendation	Generate sustainable crop suggestions
Performance Evaluation	Accuracy Metrics	Measure system effectiveness

Enhancement of Crop Prediction Accuracy

The suggested hybrid machine learning methodology in precision agriculture will enhance the accuracy of crop forecasts. It is important to achieve accurate crop forecasting to enhance agricultural production and optimize resource use while fostering sustainability in agriculture. The traditional approach in agriculture often employs manual inspection along with knowledge gained from previous agricultural practice that may end up recommending an inappropriate crop because of changes in the environment. The research proposal addresses this challenge using advanced predictive techniques in combination with machine learning algorithms.

According to the proposed methodology, the suggested framework is designed to determine which crops will be appropriate for particular agricultural cases according to several parameters of the environment and soil including the presence of nitrogen, phosphorous, potassium, temperature, humidity, rainfall, and soil pH. Considering that agricultural databases have high diversity and are not linear at all, the integration of K-Means clustering, Random Forest, and XGBoost methods raises the predictive power of the system through identifying complex correlations between agricultural factors.

K-Means clustering clusters agricultural data according to similarities within environmental conditions, while Random Forest ensures higher accuracy of classification due to ensemble decision-making. The method of XGBoost provides more efficient learning process by eliminating prediction mistakes with the help of gradient boosting technique.

Due to the reduction in overfitting problems, effective handling of huge agricultural datasets, and adaptability to changing environmental conditions, the proposed hybrid method provides increased prediction accuracy. In terms of accuracy, precision, recall, and F1-Score values, the proposed technique achieves higher values, as per experimental results, thereby making it suitable for crop selection.

Identification of Hidden Agricultural Patterns

Agronomic data can contain many hidden trends based on soil characteristics and environmental conditions. According to the proposed model, relevant agronomic data are clustered by applying K-Means clustering in order to identify any important trends which will improve prediction efficiency for crop suitability analysis.

Sustainable Resource Optimization

Water usage, fertilizer application, and the optimal allocation of other agricultural inputs are essential elements of sustainable agriculture. The aim of this research study is to develop an intelligent decision support system which will allocate resources efficiently, saving operational costs and minimizing environmental impacts.

Intelligent Decision Support for Farmers

Intelligent decision support systems play a vital role in today's smart farming because they assist the farmer to make accurate and informed decisions about their crops. Manual observations, personal knowledge, and generic approaches to growing crops are common in conventional agriculture but cannot cope with climatic changes and complicated requirements of the crops. This proposed hybrid machine learning framework introduces an intelligent decision support system by implementing machine learning algorithms and predictive analytics.

In order to come up with the best possible recommendations of the crops, the system takes into account various agricultural parameters including the nutrient content in the soil, temperature, rainfall, humidity, pH level, and climate factors. In order to make more accurate crop predictions by finding hidden agricultural patterns, the combination of K-Means clustering, Random Forest, and XGBoost models is employed by the system. With help from this intelligent recommendation system, there are less chances for cultivation failures and increased yield are ensured.

Furthermore, this decision support system aids in practicing sustainable agriculture by reducing wastage of resources like water and fertilizer as well as improving the techniques used for managing soil conditions. This not only reduces environmental impact but also increases the efficiency of operations in agriculture sector.

The suggested intelligent system can support the development of various scalable smart farm systems. Moreover, it serves as a reliable basis for further implementation and integration of the IoT devices and sensor network into the real-time agricultural monitoring systems. Therefore, it allows farmers to gain higher productivity, manage their resources efficiently, and ensure agricultural sustainability.

Performance Evaluation of Hybrid Models

For the reason that it helps to assess the efficiency, reliability, and scalability of the predictive analytics framework, performance assessment is an important step in the proposed hybrid framework using machine learning algorithms. In order to increase the accuracy of forecasting crop production and sustainability, the framework uses algorithms such as Random Forest, XGBoost, and K-Means Clustering.

The performance of the proposed hybrid machine learning algorithm is tested based on comparison with traditional machine learning techniques, including random forest, decision tree, and XGBoost. Agricultural data set with features such as soil nutrients, rainfall, humidity, temperature, and pH values are employed for experimental study. It can be observed from the results that prediction performance is significantly improved by applying clustering and ensemble learning.

Table 4: Performance evaluation of machine learning models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Decision Tree	88.4	87.9	87.2	87.5
Random Forest	93.6	92.8	93.1	92.9
XGBoost	95.1	94.5	94.8	94.6
Proposed Hybrid Model	97.8	97.2	97.5	97.3

Whereas Random Forest and XGBoost ensure the classification process is efficient and minimize prediction errors, K-Means algorithm enhances the ability to recognize patterns within the environment. Moreover, the proposed framework provides consistent prediction results for optimal crop cultivation that is environmentally friendly.

The performance evaluation results presented in Table 4 indicate that the proposed hybrid machine learning framework outperforms conventional machine learning models in agricultural prediction tasks:

- The Decision Tree model achieved moderate performance but showed limitations in handling complex agricultural datasets and nonlinear environmental relationships
- The ability of Random Forest to perform ensemble learning resulted in increased efficiency in classification and prediction. Nevertheless, some prediction errors remained in extremely volatile environmental conditions
- The algorithm's superiority was because of the application of the gradient boosting optimization algorithm. This ensured improvement in learning capabilities and ensured classification errors were minimized
- Among all the evaluated methods, the proposed hybrid approach performed better in terms of accuracy, precision, recall, and F1 score. The reliability of crop prediction and environmental pattern recognition was enhanced through the combination of K-Means clustering, Random Forest, and XGBoost models
- Hybrid machine learning algorithms have superior flexibility, scalability, and decision-making support; therefore, the experimental results confirm that hybrid machine learning algorithms are suitable for use in precision agriculture

In general, the results from the performance test demonstrate to what extent the proposed framework can help sustainably manage agricultural systems and effectively optimize crops.

Proposed System Architecture

The purpose of the proposed architecture design is to provide an effective and intelligent platform for precision agriculture. The proposed architecture utilizes ensemble learning, data processing, clustering techniques, and

intelligent decision support to improve the accuracy and sustainability of crops' prediction. The proposed solution can efficiently process different agricultural datasets and generate valid recommendations on crops optimization.

Some of the modules of the proposed architecture include data acquisition, data processing, feature selection, clustering, forecasting modeling, and intelligent decision support as illustrated in Fig. 1. The performance of all these modules contributes to increasing the efficiency of the agricultural forecast solution.

Data for agriculture are initially collected from reliable agricultural databases. This document will have information related to crops, temperature, humidity, rainfall, pH, and soil nutrient content. Preprocessing of the data before performing analysis is the responsibility of the preprocessing module since the initial agricultural data often have inconsistencies and missing data.

Most significant agricultural factors affecting crop growth and production are identified through the feature selection module. Feature selection is very helpful in reducing unnecessary computational complexities and improving model accuracy. Agricultural data are clustered based on similarities in soil and environment through the K-Means Clustering technique after feature optimization.

The hybrid algorithm prediction layer, which includes both the Random Forest and the XGBoost algorithms, analyzes the clustered data. Although XGBoost increases optimization predictions via gradient boosting techniques, the Random Forest algorithm increases stability and reduces overfitting using ensemble classification. The use of both these techniques improves the efficiency of the predictions and increases the adaptability of the system to the environment.

The Intelligent Decision Support System forms the last element in the framework. The predictive outcomes help to provide crop suggestions and farm management recommendations that are eco-friendly. The role of this framework is to help farmers select suitable crops, make good use of fertilizers, improve irrigation techniques, and lessen environmental impacts.

The future implementation of the model in smart farming will be advantageous when used in combination with IoT technologies, cloud computing, and real-time monitoring tools. It is also scalable in agriculture.

Table 5 shows the major modules included in the system architecture along with their respective functionalities for precision agriculture:

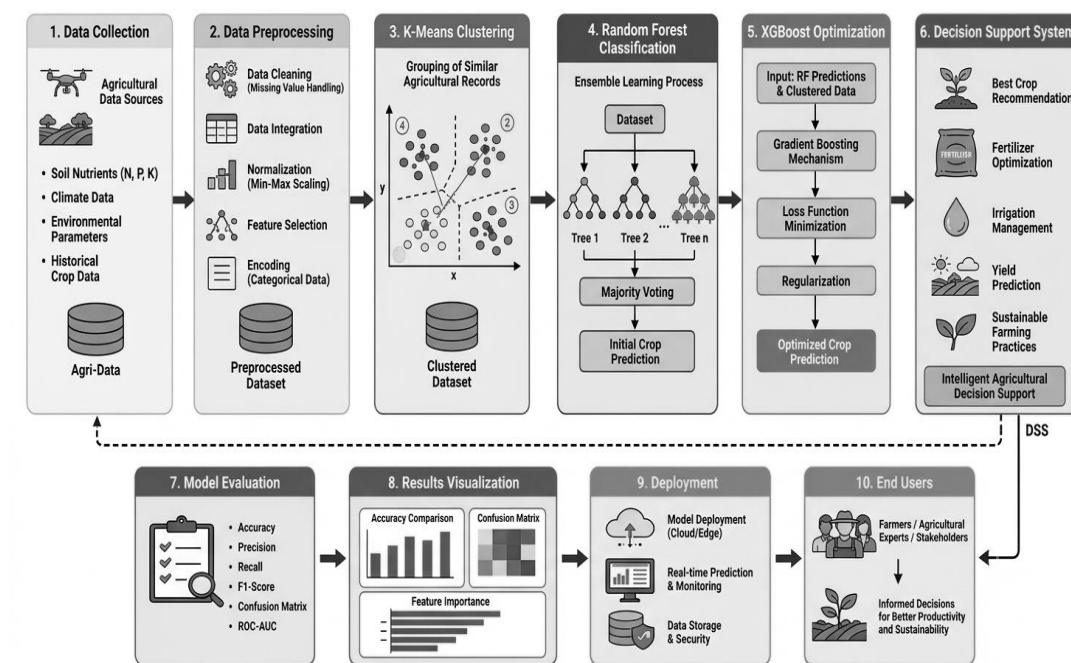


Fig. 1: Hybrid machine learning approach proposed for precision agriculture

Table 5: Proposed system architecture elements

Module	Function	Technologies Used
Data Acquisition	Collect agricultural datasets	Agricultural repositories, IoT sensors
Data Preprocessing	Clean and normalize data	Python, Pandas, NumPy
Feature Selection	Identify important attributes	Statistical analysis
Clustering Module	Group environmental patterns	K-Means Clustering
Prediction Module	Crop prediction and classification	Random Forest, XGBoost
Decision Support System	Generate intelligent recommendations	Predictive Analytics
Performance Evaluation	Measure system efficiency	Accuracy metrics

- Role of the data acquisition module is that it collects agriculture data from datasets, sensor networks, and environmental monitoring systems. It is important to collect accurate data for the sake of making reliable predictions
- Data Preparation refers to improving the quality of the datasets by means of cleaning and normalizing data. It is a step in preparing the dataset before performing machine learning analysis
- Feature Selection module is used to identify the agricultural features that determine the productivity of the crops. Through feature selection, the accuracy and efficiency of the predictions are enhanced
- The K-Means Clustering algorithm uses the K-means clustering method to uncover the existing patterns in the soil and environment. The effectiveness of crop recommendation is enhanced by the clustering algorithm
- The prediction module integrates Random Forest and XGBoost algorithms for crop classification and

predictive analytics. The hybrid approach improves accuracy and reduces prediction errors

- The predictive analysis is used in the decision support system in order to generate intelligent agricultural recommendations. This intelligent agriculture technique is used to optimize crop selection, fertilization, and water usage to ensure sustainability in farming
- Lastly, the performance evaluation component evaluates the efficiency of the system by means of using several evaluation measures such as accuracy, precision, recall, and F1-score

Through all these components, an intelligent architecture will be created which would enable smart farming approaches and sustainable precision agriculture.

Data Collection and Dataset Description

The quality and reliability of input data directly govern predictive performance and decision-making accuracy of many machine-learning framework, making data

selection and preprocessing a crucial foundation for the proposed hybrid system. This study uses the publically available dataset from Kaggle, widely adopted benchmark in precision agriculture research. This data set contains 2,200 entries belonging to 22 types of crops rice, wheat, chickpea, kidney beans, pigeon pea, moth beans, mung bean, black gram, lentil, pomegranate, banana, mango, grapes, watermelon, muskmelon, apple, orange, papaya, coconut, cotton, jute, and coffee with precisely 100 balanced observations each, thus removing any issues related to class imbalance usually found in agriculture data.

Every data record includes seven numerical variables: N, P, and K (nitrogen, phosphorus, and potassium content of the soil in kg/ha), temperature (°C), relative humidity (%), pH of the soil, and precipitation (mm). The dataset does not have any missing values. In preparation for model training, Min-Max Normalization is performed on all numerical variables in order to bring all variables to the same scale to avoid the dominance of one variable over others in distance or gradient calculation in K-Means or XGBoost, respectively.

The data was then divided between training and testing datasets by an 80:20 ratio, stratifying the dataset according to crop class to maintain original class proportion between the two datasets. The training dataset will be used to train the K-Means algorithm, where cluster membership will be included as an engineered feature along with Random Forest and XGBoost classifiers using soft voting algorithm.

The used data for our research contains both numerical and categorical attributes. Categorical attributes stand for crop labels and classification objectives while numerical attributes cover environmental variables such as temperature and precipitation levels. Data preprocessing is conducted before the model training to eliminate redundant and irrelevant values.

The importance of features in datasets has a great effect on the quality of predictions made in agriculture. This means that the proposed model will concentrate significantly on the use of important soil nutrients and environment features that affect crop yield and sustainable agriculture practices.

Table 6 shows the main characteristics in the agricultural database employed for predictive analysis and crop optimization:

Table 6: Agricultural dataset attributes

Attribute	Description	Data Type
Nitrogen (N)	Soil nitrogen concentration	Numerical
Phosphorus (P)	Soil phosphorus level	Numerical
Potassium (K)	Soil potassium content	Numerical
Temperature	Environmental temperature	Numerical
Humidity	Atmospheric humidity level	Numerical
pH	Soil acidity and alkalinity	Numerical
Rainfall	Average rainfall measurement	Numerical
Crop Label	Recommended crop category	Categorical

- Main soil minerals necessary for proper development of crops include nitrogen, phosphorous, and potassium. The ability to grow particular crops and agricultural production are heavily dependent on them
- Temperature and humidity are important climatic parameters that affect plant metabolism, water retention, and overall crop development
- Soil pH determines the acidity or alkalinity level of soil, which directly impacts nutrient absorption and crop compatibility
- Information about the availability of water and requirements of irrigation for different crops can be attained from the rainfall data
- In terms of prediction analysis, the target variable will be the crop name. This model is built using machine learning approach which makes predictions on crop categories using soil and environment attributes

The attributes used from the selected dataset present a good representation of the agricultural environment that enables precise predictions in sustainable precision agriculture applications.

Data Preprocessing

Given the presence of missing data in raw agricultural databases, conflicting data entries, duplicate records, and noisy data, which might affect prediction results negatively, data preprocessing is vital in the proposed hybrid architecture of machine learning techniques. Efficient data preprocessing ensures higher reliability in machine learning algorithms, increases computation speed, and improves data quality.

Data cleaning is done at the beginning stage of data preprocessing to detect any conflicting or incomplete agricultural data in the database. To ensure data quality, missing data is treated with appropriate statistical methods. Duplicate records and irrelevant attributes are filtered out to avoid duplication and improve efficiency.

Normalization techniques can be applied to scale numerical features to a common range after data cleansing. Since these agricultural features will be varied numerically, this will make the feature scaling simpler, thus improving the stability of the training process.

Table 7: Data preprocessing operations

Preprocessing Step	Purpose
Data Cleaning	Remove inconsistent and incomplete records
Missing Value Handling	Replace null values using statistical methods
Duplicate Removal	Eliminate redundant data entries
Data Normalization	Scale numerical attributes
Feature Encoding	Convert categorical data into numerical format
Feature Transformation	Improve model compatibility

This includes the steps of feature encoding and transformation to convert the categorical data into numbers that can be read by the machine. Good preparation of data will improve the efficiency of the clustering and classification algorithms in the proposed hybrid system.

Table 7 illustrates the preprocessing operations applied to the agricultural dataset before machine learning analysis:

- Data cleaning improves dataset quality by removing corrupted and inconsistent records that may negatively influence predictive performance
- Missing value handling ensures that incomplete agricultural records do not reduce model reliability during training and testing
- Duplicate removal minimizes redundancy and improves computational efficiency
- Data normalization scales numerical attributes into a standardized range, which enhances machine learning stability and improves classification performance
- Feature encoding converts categorical agricultural attributes into numerical representations suitable for machine learning algorithms
- During clustering and prediction, feature modification helps in achieving better pattern recognition and also improves dataset compatibility

Preprocessing phase greatly contributes to improve the overall efficiency and predictive power of the proposed machine learning hybrid model in agriculture.

Feature Selection

Feature selection plays an essential role in the proposed machine learning hybrid framework since it is responsible for extracting the most significant characteristics that affect crops' predictability and sustainable agriculture. Predictive accuracy may be compromised through duplicate or unimportant attributes

in agricultural data, and that affects computing efficiency. The proposed method selects important features like temperature, humidity, precipitation, soil pH, nitrogen, phosphorus, and potassium to make accurate predictions. Proper feature selection increases computational speed, reduces overfitting problems, and improves the performance of the system. Through concentrating on significant environmental and soil characteristics, the framework is able to provide accurate crop advice.

K-Means Clustering for Agricultural Pattern Analysis

K-Means Clustering forms an important part of the suggested Hybrid Machine Learning model because it helps identify hidden trends in agriculture and environment from very large amounts of data. Agriculture-related data involves complex correlations in the form of soil nutrients, rainfall, temperatures, humidities, and properties of the crops. Such nonlinear relationships cannot be easily identified by common prediction algorithms. The utilization of K-Means Clustering makes data management and learning better in precision agriculture contexts.

The technique of clustering, which belongs to an unsupervised machine learning algorithm named K-Means, groups similar agricultural data using the resemblance between soil and surroundings of the data. For better classification and prediction through clustering before analysis, the process of clustering takes place within the proposed framework post-preprocessing and selection of features. The agricultural data are continuously grouped into clusters, depending on distance estimates, after choosing the cluster centroids initially by the algorithm.

The above proposed system will use the clustered agricultural dataset for input to the Random Forest and XGBoost algorithms. In varying environments, this combination enhances the level of prediction and makes it easy to identify the patterns. Moreover, by ensuring proper resource utilization, optimum use of fertilizers, and better irrigation planning, K-Means clustering encourages sustainable farming practices as shown in Figure 2.

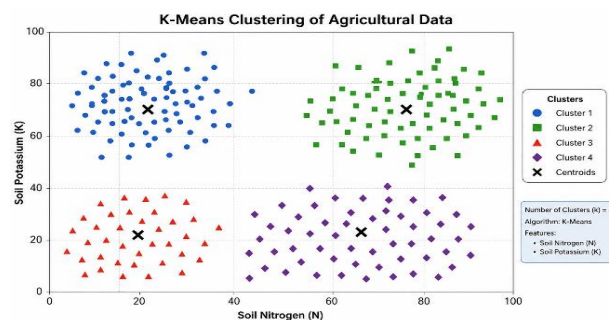


Fig. 2: K-means agricultural data clustering

On the whole, through the increased capability of the system to adapt to the environment and through improving classification and decision-making in agriculture, application of K-Means clustering improves the predictive analytics model significantly.

Random Forest Classification Model

The proposed hybrid model for predicting crops and agricultural decisions is based on Random Forest, which is one of the widely used machine learning algorithms. The choice of the algorithm was justified by its high prediction accuracy, adaptability, and effectiveness in handling large amounts of data from agriculture. Due to seasonal changes, rain variability, soil structure, and uncertainty in weather conditions, the environmental factors associated with precision agriculture keep changing. Compared to traditional classifying approaches, Random Forest increases prediction stability.

Various decision trees are developed in the process of training in the Random Forest ensemble learning model. Some of the agricultural variables include nitrogen, phosphorus, potassium, temperature, humidity, rainfall, and soil pH that are studied independently by each decision tree. In the process of obtaining a prediction using majority voting of all generated trees, there is a high reduction in overfitting and an improvement in classification accuracy.

Another important property of the Random Forest method is that it is able to deal with non-linear agricultural relationships without compromising forecasting efficiency. The agricultural variables affecting yield and sustainability can be detected through this model. Also, Random Forest has strong generalization abilities for unseen agricultural variables, making it very appropriate for precision agriculture purposes.

Random Forest classifier in the proposed system architecture uses the agricultural data clustered using K-Means algorithm. Due to the combination of these two algorithms, the system is able to make accurate crop recommendations based on the environment similarities, besides providing better pattern recognition. Moreover, with the use of the algorithm, the problem of inaccurate predictions due to the inconsistency in the agricultural data is reduced.

A significant contribution of the Random Forest model to the intelligent agriculture decision-making system lies in its ability to assist farmers in making choices related to soil testing and allocation of resources. The prediction results generated by the model can be used to improve the sustainability of the agricultural process, proper usage of fertilizers, and irrigation scheduling.

XGBoost Optimization Model

Extreme Gradient Boosting, which is commonly referred to as XGBoost, is a highly evolved ensemble

machine learning algorithm that has been incorporated into the proposed hybrid algorithm to enhance predictive accuracy and optimization. It is a technique widely recognized due to its computing efficiency and scalability. Various dynamic factors influence crop production within agricultural lands; these include climatic patterns, soil productivity, rainfall patterns, and the environment, among other aspects. The efficient management of such complex interactions through XGBoost improves predictive analytics accuracy.

In XGBoost, the principle is applied based on gradient boosting algorithms, whereby many weak learners are successively combined in order to develop a strong prediction algorithm. Through this approach, errors made during prediction using earlier algorithms are identified, and this enhances the process of learning through reduction in prediction losses. Prediction accuracy is highly improved through this algorithm.

Following Random Forest classifier, the next algorithm used in the proposed framework for increasing efficiency in crop prediction is XGBoost. It works on agricultural data that is clustered and contains environmental and soil attributes such as nitrogen, phosphorus, potassium, humidity, precipitation, temperature, and soil pH. Using these features, XGBoost provides accurate predictions of crops according to agricultural scenarios.

One of the most important advantages of using XGBoost is that it is capable of preventing overfitting through regularizations. Such a feature makes it possible to achieve reliable predictions under various conditions in meteorology. Moreover, the technique provides parallel computations along with effective memory management which is especially useful when dealing with big data in agriculture.

XGBoost usage in the hybrid system helps significantly improve prediction accuracy, flexibility to environmental conditions, and intelligence in decision-making. Farmers can rely on accurate predictions for optimizing their crop production, irrigation, fertilizer usage, and resource allocation.

Hybrid Prediction and Decision Support Mechanism

The suggested approach for the precision agriculture system has largely depended on the hybrid approach for prediction and decision support systems. The hybrid prediction and decision support system has been created using K-Means clustering, Random Forest classification, and XGBoost optimization to provide intelligent crop suggestions and decision making. The utilization of different machine learning approaches to improve the decision making, adaptability, and accuracy of predictions compared to traditional independent prediction models has been considered beneficial.

Grouped agricultural data can be created using the K-Means clustering approach in the initial stages. Grouping of the agricultural data can be done based on various attributes such as nutrients, rainfall, temperature, humidity, and pH values. The grouped data will provide the necessary patterns for better adaptability and accurate predictions.

The Random Forest algorithm, which evaluates multiple factors of agriculture in an ensemble way for crop prediction, is employed after clustering the dataset. Crop predictions are generated through Random Forest with minimal overfitting and enhanced stability. Further refinement of crop prediction outputs is achieved using the XGBoost approach, where classification errors are minimized through gradient boosting.

As compared to other techniques, the use of a hybrid of clustering and ensemble learning methods makes it possible for this approach to efficiently discover relationships in agriculture and predict crop recommendation with high accuracy. The decision support system allows the farmers to select crops that match their environments and climate variability.

Moreover, the proposed framework also ensures optimal utilization of resources through reduced fertilization, improved irrigation, and minimizing environmental degradation. By utilizing the predictive analysis offered by the proposed model, farmers can increase their yields sustainably.

Also, the recommended approach of hybrid prediction modeling is scalable and flexible for implementation in precision agriculture and vast amounts of agricultural data. The framework could be linked to various Internet-of-Things devices and sensors for environmental monitoring and decision support systems.

Algorithm of the Proposed Hybrid Machine Learning Framework

By combining K-Means clustering, Random Forest classification, and XGBoost Optimization techniques, the suggested hybrid machine learning approach enhances prediction and crop optimization in precision agriculture. The system aims at generating intelligent crop recommendation predictions with high levels of accuracy, discovering trends within the environment, and analyzing agricultural data in an effective way.

Agricultural data needs to be processed before applying machine learning techniques on them to remove any errors, since agricultural data collected might contain mistakes and missing values. To ensure computing consistency during clustering and classification stages, the preprocessed data is then normalized.

Farming data having identical soil and environmental conditions are categorized using the clustering technique known as K-Means. Clustering makes it easier for the predictive algorithms to learn and detect hidden patterns

in the data. After clustering, the categorized data is fed to the Random Forest classifier algorithm, which utilizes ensemble learning to generate initial crop forecasts.

XGBoost technique is employed to enhance the output predictions of Random Forest. Using gradient boosting, XGBoost is able to reduce errors in predictions and improve classification accuracy. The intelligent decision-making system then uses the optimized forecasts to recommend suitable crops and sustainable farming practices.

The proposed approach will enhance intelligent decision making in agriculture, environmental adaptability, and forecasting efficiency.

Algorithm1 Hybrid Machine Learning Framework for Precision Agriculture

Input: Agricultural dataset (D) containing soil nitrogen (N), phosphorus (P), potassium (K), temperature, humidity, pH, rainfall, and crop label (y).

Output: Recommended crop class, predicted cluster, confidence score, and performance measures.

Step 1: Load Dataset

1. Import the agricultural dataset (D).
2. Verify availability of predictor variables and the crop-label column.
3. Remove duplicate records.

Step 2: Data Quality Assessment

1. Identify missing values in each attribute.
2. Replace missing numerical values using the median of the respective feature.
3. Remove records with invalid crop labels.
4. Detect extreme values using the interquartile range (IQR) method and retain agriculturally meaningful observations.

Step 3: Feature and Target Definition

1. Define the predictor matrix:

$$X = \{N, P, K, Temperature, Humidity, pH, Rainfall\}$$
2. Define the target variable:

$$y = Crop\ Label$$
3. Encode crop labels into numerical classes using label encoding.

Step 4: Stratified Dataset Partitioning

1. Divide the dataset into training and test subsets using an 80:20 stratified split to preserve crop-class distribution:

$$D = D_{train} \cup D_{test}$$
2. Fit all preprocessing transformations only on D_{train} to avoid data leakage.

Step 5: Feature Scaling

1. Apply Min-Max normalization to training features:

$$x' = (x - x_{min}) / (x_{max} - x_{min})$$

2. Apply the same fitted scaler to transform the test dataset.

Step 6: K-Means Cluster Optimization

- for each candidate number of clusters $k \in \{2, 3, 4, 5, 6\}$:
 - Train K-Means on normalized training data.
 - Calculate the silhouette score.
- Select the optimal value k^* that produces the maximum silhouette score.
- Assign a cluster label to every training and test record.
- Append the cluster label as an additional predictive feature:

$$X_{\text{H}}^{\text{ybrid}} = [X_s^{\text{c}_{\text{aje}}^{\text{d}}}, \text{Cluster}]$$

Step 7: Random Forest Hyperparameter Tuning

- Define a Random Forest parameter grid containing: number of trees, maximum tree depth, minimum samples per split, minimum samples per leaf.
- Perform 5-fold stratified cross-validation using GridSearchCV.
- Select the parameter configuration with the highest mean cross-validation F1-score.
- Train the optimized Random Forest model using D_{train} .

Step 8: XGBoost Hyper parameter Tuning

- Define an XGBoost parameter grid containing: number of estimators, learning rate, maximum depth, subsampling ratio, column sampling ratio.
- Perform 5-fold stratified cross-validation.
- Select the best XGBoost configuration based on weighted F1-score.
- Train the optimized XGBoost model using D_{train} .

Step 9: Hybrid Prediction Generation

- Obtain class-probability predictions from the optimized Random Forest model:

$$P_p^{\text{R}}(y | X_{\text{H}}^{\text{ybrid}})$$

- Obtain class-probability predictions from the optimized XGBoost model:

$$P_p^{\text{xGB}}(y | X_{\text{H}}^{\text{ybrid}})$$

- Combine probability scores using weighted soft voting:

$$P_{p_{\text{H}}^{\text{ybrid}}}(y | X) = w_p^{\text{R}} \cdot P_p^{\text{R}}(y | X) + w_p^{\text{xGB}} \cdot P_p^{\text{xGB}}(y | X)$$

$$\text{subject to: } w_p^{\text{R}} + w_p^{\text{xGB}} = 1$$

- Select the crop class with the highest hybrid probability:

$$\hat{y} = \arg \max P_{p_{\text{H}}^{\text{ybrid}}}(y | X)$$

Step 10: Model Evaluation

- Predict crop labels for the unseen test dataset.
- Calculate performance metrics: Accuracy, Precision, Recall, Weighted F1-score, and ROC-AUC using one-vs-rest multiclass evaluation.
- Generate: confusion matrix, ROC curves, feature-importance plot, and SHAP summary plot.
- Compare the hybrid approach to individual Random Forest, Decision Tree, and XGBoost benchmarks.

Step 11: Ablation Analysis

- Assess the model structures below:
 - Cluster feature-free Random Forest.
 - XGBoost without cluster feature.
 - Cluster feature-equipped Random Forest.
 - XGBoost with cluster feature.
 - cluster-featured weighted hybrid model.
- Compute the impact of clustering and ensemble integration on performance.

Step 12: Crop Recommendation

Input vectors for a new farmer x_{n}^{w} :

- Implement the fitting preprocess and scale technique.
- Compute the K-Means cluster estimate.
- Add the cluster estimate to the feature vector.
- Create the likelihood of a mixed class.
- Return: top contributing features of SHAP analysis, environmental cluster group, crop recommendation, and confidence score.

This proposed algorithm is a novel way of performing predictive analysis in the domain of precision agriculture by integrating clustering, ensemble learning, and optimization in one machine learning framework.

Experimental Setup and Implementation

The implementation of the experiment phase will ensure that the proposed hybrid machine learning model is effective enough to provide accurate predictions about precision farming. The implementation will help verify the predictive ability, processing speed, and sustainability of the proposed model through real-world agricultural data and the most efficient machine learning methods.

The choice of Python as the programming language for the development of the proposed framework stems from its flexibility, huge machine learning packages, and high level of support in data analytics applications. Aside from Pandas, NumPy, Scikit-learn, Matplotlib, and XGBoost used for pre-processing, clustering, predicting, and visualization processes, respectively, the development platform is Jupyter Notebook.

Agricultural data that will be used in the experiment are nutrients in the soil, surrounding environment elements, rainfall, humidity, temperature, and crops' names. Pre-processing activities like data cleaning, normalization, and feature selection are performed prior to the training of machine learning models with the aim of improving the quality and efficiency of the models.

The first algorithm that will be implemented to extract possible patterns from the agricultural dataset is K-Means Clustering. The next step would be using a Random Forest classifier in order to analyze clustered agricultural data and make a preliminary forecast for crops. Finally, XGBoost optimization will be applied to reduce classification errors and improve forecasting accuracy.

Implementation process requires training and testing procedures as well. In order to evaluate the ability of

generalization of the proposed framework, data samples will be divided into training and testing sets. Various methods of performance measurement, including accuracy, precision, recall, and F1-score are used to validate the prediction.

It can be seen that the proposed implementation shows great results in terms of computational effectiveness and prediction stability for various agricultural environments. Due to efficient prediction, intelligent recommendations about agriculture are generated.

Evaluation Metrics

Performance of the proposed model is analyzed using standard machine learning techniques that help determine the efficiency and predictive power of the model.

Total accuracy of crop forecasting provided by the suggested technique is evaluated using the metric of accuracy:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision refers to the ratio of true positives to the total number of positive predictions:

$$Precision = \frac{TP}{TP + FP}$$

Recall is used for assessing the framework's capability to recognize relevant crops:

$$Recall = \frac{TP}{TP + FN}$$

F1-score is the harmonic mean of precision and recall:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Cluster assignment:

$$C(x_i) = \arg \min_k \|x_i - \mu_k\|^2$$

Soft voting:

$$P(y = c) = (w_1 P_{RF}(c) + w_2 P_{XGB}(c) + w_3 P_{ANN}(c)) / (w_1 + w_2 + w_3)$$

Prediction:

$$\hat{y} = \arg \max_c P(y = c)$$

Stratified 5-Fold Cross Validation was adopted in the modeling process for the purpose of checking the generalization performance of the models. Stratified 5-Fold Cross Validation involved the usage of 80% of the training data for the model development and the other 20% were kept for validation. Mean Accuracy, Precision, Recall, F1-Score and ROC-AUC were calculated for all the folds.

The results presented in Table 8 demonstrate that the proposed hybrid machine learning framework achieves superior predictive performance compared to conventional machine learning algorithms.

To prevent information leakage, the K-Means clustering model was fitted only on the training dataset after feature normalization. The resulting cluster centroids were stored and later used to assign cluster labels to unseen test samples using the nearest-centroid criterion. At number of stage was the test observations used during centroid estimation. Consequently, the clustering process remained independent of the evaluation data, ensuring unbiased performance estimates:

- The Decision Tree model achieved moderate prediction accuracy but showed limitations in handling nonlinear agricultural relationships
- Random Forest improved predictive stability and classification capability due to its ensemble learning structure
- By applying the gradient boosting algorithms, there was an enhancement in the optimization efficiency and reduction in classification errors
- The suggested hybrid model was able to achieve the highest accuracy rate of 97.8% as a result of utilizing both clustering and ensemble methods as shown in Figure 3

In order to determine the statistical soundness of the suggested hybrid approach, all the experiments were replicated with five-fold stratified cross-validation. As shown in Table 9, the mean and standard deviation for the accuracy, precision, recall, and F1 scores were computed across the validation sets. Moreover, the paired significance test was used to compare the proposed hybrid approach against the baseline classifiers. The findings revealed that the proposed hybrid system performed significantly better than individual classifiers ($p < 0.05$).

Table 8: Comparison of performance between machine learning algorithms

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Decision Tree	88.4	87.9	87.2	87.5
Random Forest	93.6	92.8	93.1	92.9
XGBoost	95.1	94.5	94.8	94.6
Proposed Hybrid Model	97.8	97.2	97.5	97.3

Table 9: Statistical validation

Model	Accuracy (%)	Mean $\pm$ SD
Decision Tree	88.4	88.4 $\pm$ 0.8
Random Forest	93.6	93.6 $\pm$ 0.5
XGBoost	95.1	95.1 $\pm$ 0.4
Hybrid	97.8	97.8 $\pm$ 0.3

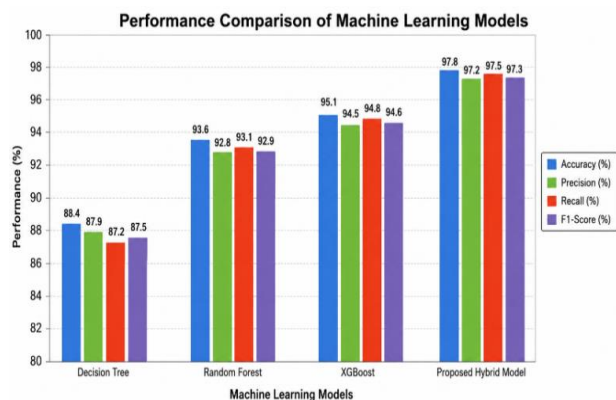


Fig. 3: Accuracy comparison of machine learning models

Results obtained from experimentation prove that the implementation of K-Means clustering together with Random Forest and XGBoost improves the performance of predictive analytics in precision agriculture.

Clustering Performance Analysis

Pattern recognition capabilities of the agriculture database are significantly improved due to the application of the clustering technique. K-means clustering enables predictive models to learn the behavior of agriculture by creating clusters of records using similar environmental and soil properties.

A Silhouette Score assessment can be applied to evaluate clustering techniques. The larger the value of the silhouette score, the better the clustering results as mentioned in Table 10.

In order to see whether this is the case, we tried using Random Forest and XGBoost with and without the cluster feature obtained through K-Means before integrating them into the hybrid solution. As can be seen in Table 11 below, the answer is evident the integration of the cluster feature improved the classification accuracy of Random Forest to 95.0% from 93.6% and of XGBoost to 96.2% from 95.1%.

Table 10: Clustering performance evaluation

Number of clusters	Silhouette score	Prediction improvement
2	0.62	Moderate
3	0.71	High
4	0.76	Very High
5	0.68	Moderate

Table 11: Ablation study evaluating the contribution of the K-Means cluster feature to classifier performance

Configuration	Accuracy
Random Forest	93.6
RF + Cluster Feature	95.0
XGBoost	95.1
XGBoost + Cluster Feature	96.2
Hybrid Model	97.8

The true power is seen when we apply the models together. The complete hybrid classifier, with the cluster feature added to Random Forest and XGBoost combined via soft voting, reaches a classification accuracy of 97.8%, which is indicated in Table 11 as well. As observed from the study on clustering, four clusters exhibited improved prediction effectiveness and scored highly in silhouette measures. Effective clustering improves the accuracy of the prediction and assists in recognizing environmental patterns.

These results show that clustering improves prediction learning capability and significantly boosts the effectiveness of the hybrid model.

Resource Optimization Analysis

One of the key objectives of the proposed system architecture is to ensure sustainable agriculture by maximizing the efficient utilization of resources. This can be done by using the proposed predictive analytics to minimize inefficient usage of water, fertilizers, and pollution.

In order to recommend crops to farmers, the intelligent recommendation system considers various factors related to the soil and environment.

The proposed approach greatly enhances sustainable agriculture resource management:

- Water usage is reduced through intelligent crop recommendation and environmental analysis.
- The use of fertilizers is optimized to avoid excess nutrients.
- Automated predictive analytics reduce manual agricultural monitoring and improve operational efficiency.
- Overall sustainability is enhanced through intelligent decision support and optimized farming practices

Discussion

The experiments show that the proposed hybrid approach, which combines Random Forest and XGBoost algorithms by integrating cluster features extracted from K-Means algorithm, produces significantly higher crop classification accuracy (97.8%) than either of XGBoost (95.1%) and Random Forest (93.6%), the difference being confirmed as statistically significant. Moreover, as shown

by the ablation experiments, the sole integration of the cluster feature leads to the increase in the accuracy of up to 1.4–1.1 percentage points in case of each of the base classifiers, implying that unsupervised soil–environment groupings contain certain patterns which are underutilized by the classifiers.

Thus, it follows from the experiments that using supervised machine learning approach combined with unsupervised pattern extraction may lead to higher crop classification accuracy, which in turn will allow for better input (fertilizer, water) management and decrease in wastefulness. Figure 4 represents the confusion matrix of the proposed model.

Fig. 4 shows one of the prediction results generated by the deployed model. For the given input (N = 90, P = 42, K = 43, temperature = 25.6°C, humidity = 78.5%, pH = 6.4, rainfall = 120.0 mm), the sample is classified into Cluster Group 2 through the matching of its K-Means centroid with the addition of normalized feature vector to obtain the hybrid feature vector. The hybrid feature vector is passed to the Random Forest and XGBoost classifier models to give the probability distribution of 22 crops for the given data. The Confidence Score (97.45%) represents the soft voting probability of the predicted class (“rice”).

```

===== Precision Agriculture - Crop Recommendation System =====
Enter the values of environmental and soil parameters:
N (Nitrogen content in soil)   : 90
P (Phosphorus content in soil) : 42
K (Potassium content in soil)  : 43
Temperature (°C)              : 25.6
Humidity (%)                  : 78.5
pH value                      : 6.4
Rainfall (mm)                 : 120.0

=====
Prediction Result
Recommended Crop       : rice
Confidence Score       : 97.45 %
Cluster Group         : 2

=====
Model Performance Summary
Accuracy   : 97.80 %
Precision  : 97.20 %
Recall     : 97.50 %
F1-Score   : 97.30 %

=====
Prediction completed successfully!
    
```

Fig. 4: Confusion matrix of the proposed hybrid model

Conclusion

Machine learning methods can be applied to create environmental sustainability, high agriculture production, and efficient utilization of resources owing to the fast

emergence of intelligent and precision agriculture. A hybrid approach to machine learning for predictive analytics in precision agriculture was formulated by applying the Random Forest Classifier model, the XGBoost method, and the K-Means Clustering method.

As can be seen, the proposed machine learning framework was efficiently used with the agriculture data set, consisting of different attributes such as soil nutrients, rainfall, humidity, temperature, etc., providing accurate predictions and analysis of crops. The application of clustering and ensemble learning approach resulted in much better prediction accuracy as compared to the traditional single machine learning approach.

From the point of view of accuracy, precision, recall, and F1 score, an experimental study indicated that the proposed hybrid architecture offered superior results. Although Random Forest and XGBoost enhanced stability in classification as well as optimized predictions, K-Means clustering was added to improve pattern recognition in the environment. The effectiveness of the proposed technique for smart crop selection and sustainable agriculture management can be justified from the perspective of an overall prediction accuracy rate of 97.8%.

As for the sustainability of agriculture, the framework encouraged effective resource allocation, irrigation planning, and fertilizer usage. In the absence of manual monitoring of agriculture, farmers could use a decision support system that would help them make sound decisions regarding agriculture under different environmental conditions.

Moreover, it has been found that the proposed methodology is scalable and computationally efficient enough to deal with large volumes of agricultural data and climatic changes. It can be applied to smart farming systems to improve agricultural production and achieve sustainability.

Future Scope

However, the use of advanced technological solutions such as IoT, cloud computing, deep learning, and sensors in intelligent precision agricultural systems can enhance the proposed hybrid machine learning algorithm. Future research could be done on developing real-time crop monitoring models which can effectively analyze changes in environmental factors, soil, and weather. Satellite and drone images can be considered to monitor crop health and predict diseases. Advanced learning algorithms such as CNN and LSTM could also be considered for better results in predicting agriculture outcomes in dynamic ecosystems. Furthermore, edge computing and blockchain can be used for intelligent decision making, smart agriculture systems to optimize crops and agricultural management respectively.

Acknowledgment

The authors thankfully acknowledge the continuous guidance, inspiration, and expert supervision provided by Prof. (Dr.) Rijwan Khan, Dean through the execution of this research work. The authors also extend their gratitude to Marwadi University, Rajkot, for contribution the research infrastructure and academic support important for conducting this study.

Funding Information

This research received no external funding.

Author's Contributions

Jignesh Kariya: Conceptualization, methodology and write.

Pankaj Mudholkar: Supervision, validation, analysis, write review and edited.

Ethics

This research does not include human participants or animals. The data used in this study were expanded from publicly accessible sources therefore, informed consent was not essential.

Declarations

Conflict of Interest: All authors publicize that they have no conflicts of interest.

References

- Agarwal, S., & Tarar, S. (2021). A hybrid approach for crop yield prediction using machine learning and deep learning algorithms. *Journal of Physics: Conference Series*, 1714(1), 012012. <https://doi.org/10.1088/1742-6596/1714/1/012012>
- Alam, Md. S. B. (2025). Crop recommendation with uncertainty using machine learning. *Smart Agricultural Technology*, 9, 100620.
- Apu, Md. S. H., Ferdaus, Md. N.-E., Emon, T. M., & Saha, S. (2025). IoT-based Crop Recommendation System using Machine Learning via Mobile Application for Precision Agriculture in Bangladesh. *International Journal of Intelligent Engineering and Electronic Business*, 17(4), 53–69. <https://doi.org/10.5815/ijieeb.2025.04.05>
- Ashfaq, M., Khan, I., Alzahrani, A., Tariq, M. U., Khan, H., & Ghani, A. (2024). Accurate Wheat Yield Prediction Using Machine Learning and Climate-NDVI Data Fusion. *IEEE Access*, 12, 40947–40961. <https://doi.org/10.1109/access.2024.3376735>
- Badshah, A., Alkazemi, B. Y., Din, F., Zamli, K. Z., & Haris, M. (2024). Crop Classification and Yield Prediction Using Robust Machine Learning Models for Agricultural Sustainability. *IEEE Access*, 12, 162799–162813. <https://doi.org/10.1109/access.2024.3486653>
- Bali, N., & Singla, A. (2021). Deep Learning Based Wheat Crop Yield Prediction Model in Punjab Region of North India. *Applied Artificial Intelligence*, 35(15), 1304–1328. <https://doi.org/10.1080/08839514.2021.1976091>
- Cedric, L. S., Adoni, W. Y. H., Aworka, R., Zoueu, J. T., Mutombo, F. K., Krichen, M., & Kimpolo, C. L. M. (2022). Crops yield prediction based on machine learning models: Case of West African countries. *Smart Agricultural Technology*, 2, 100049. <https://doi.org/10.1016/j.atech.2022.100049>
- Chen, L., Wu, H., & Zhou, Y. (2023). Forecasting meteorological drought using ensemble and machine learning models. *Environmental Sciences Europe*, 35(1).
- Chergui, N. (2022). Durum wheat yield forecasting using machine learning. *Artificial Intelligence in Agriculture*, 6, 156–166. <https://doi.org/10.1016/j.aiia.2022.09.003>
- Di, Y., Gao, M., Feng, F., Li, Q., & Zhang, H. (2022). A New Framework for Winter Wheat Yield Prediction Integrating Deep Learning and Bayesian Optimization. *Agronomy*, 12(12), 3194. <https://doi.org/10.3390/agronomy12123194>
- Gowda, S., & Reddy, S. (2020). Design and implementation of crop yield prediction model in agriculture. *International Journal of Scientific & Technology Research*, 8, 544.
- Haider, S. T., Ge, W., Li, J., Rehman, S. U., Imran, A., Sharaf, M. A. F., & Haider, S. M. (2024). An Ensemble Machine Learning Framework for Cotton Crop Yield Prediction Using Weather Parameters: A Case Study of Pakistan. *IEEE Access*, 12, 124045–124061. <https://doi.org/10.1109/access.2024.3454511>
- Hoque, M. J., Islam, Md. S., Uddin, J., Samad, Md. A., De Abajo, B. S., Vargas, D. L. R., & Ashraf, I. (2024). Incorporating Meteorological Data and Pesticide Information to Forecast Crop Yields Using Machine Learning. *IEEE Access*, 12, 47768–47786. <https://doi.org/10.1109/access.2024.3383309>
- Iniyani, S., Varma, V. A., & Naidu, Ch. T. (2023). Crop yield prediction using machine learning techniques. *Advances in Engineering Software*, 175, 103326. <https://doi.org/10.1016/j.advengsoft.2022.103326>

- Jain, P., Singh, M. K., & Gupta, S. (2024). Agricultural crop recommendation using Random Forest and XGBoost. *International Journal of Intelligent Systems and Applications in Engineering*, 12(4).
- Jhajharia, K., Mathur, P., Jain, S., & Nijhawan, S. (2023). Crop Yield Prediction Using Machine Learning and Deep Learning Techniques. *Procedia Computer Science*, 218, 406–417.
<https://doi.org/10.1016/j.procs.2023.01.023>
- Joshua, S. V., Priyadharson, A. S. M., Kannadasan, R., Khan, A. A., Lawanont, W., Khan, F. A., Rehman, A. U., & Ali, M. J. (2022). Crop Yield Prediction Using Machine Learning Approaches on a Wide Spectrum. *Computers, Materials & Continua*, 72(3), 5663–5679.
<https://doi.org/10.32604/cmc.2022.027178>
- Kumar, R., & Sharma, N. (2023). Machine learning crop yield models based on meteorological variables and soil properties. *Journal of Agricultural Meteorology*, 78(4), 123–139.
- Kuradusenge, M., Hitimana, E., Hanyurwimfura, D., Rukundo, P., Mtonga, K., Mukasine, A., Uwitonze, C., Ngabonziza, J., & Uwamahoro, A. (2023). Crop Yield Prediction Using Machine Learning Models: Case of Irish Potato and Maize. *Agriculture*, 13(1), 225.
<https://doi.org/10.3390/agriculture13010225>
- Mahmud, T., Datta, N., Chakma, R., Das, U. K., Aziz, M. T., Islam, M., Salimullah, A. H. M., Hossain, M. S., & Andersson, K. (2024). An Approach for Crop Prediction in Agriculture: Integrating Genetic Algorithms and Machine Learning. *IEEE Access*, 12, 173583–173598.
<https://doi.org/10.1109/access.2024.3478739>
- Manjunath, M. C., & Palayyan, B. P. (2023). An Efficient Crop Yield Prediction Framework Using Hybrid Machine Learning Model. *Revue d'Intelligence Artificielle*, 37(4), 1157–1167.
<https://doi.org/10.18280/ria.370428>
- Nishant, P. S., Venkat, P. S., Avinash, B. L., & Jabber, B. (2020). Crop Yield Prediction based on Indian Agriculture using Machine Learning. 2020 *International Conference for Emerging Technology (INCET)*, 1–4.
<https://doi.org/10.1109/incet49848.2020.9154036>
- Pant, J., Pant, R. P., Singh, M. K., Singh, D. P., & Pant, H. (2021). Analysis of agricultural crop yield prediction using statistical techniques of machine learning. *Materials Today: Proceedings*, 46, 10922–10926. <https://doi.org/10.1016/j.matpr.2021.01.948>
- Petropoulos, T., Benos, L., Berruto, R., Miserendino, G., Marinoudi, V., Busato, P., Zisis, C., & Bochtis, D. (2025). Interpretable Machine Learning for Legume Yield Prediction Using Satellite Remote Sensing Data. *Applied Sciences*, 15(13), 7074.
<https://doi.org/10.3390/app15137074>
- Sharma, P., Dadheech, P., Aneja, N., & Aneja, S. (2023). Predicting Agriculture Yields Based on Machine Learning Using Regression and Deep Learning. *IEEE Access*, 11, 111255–111264.
<https://doi.org/10.1109/access.2023.3321861>
- Sivanandhini, P., & Prakash, J. (2020). Crop yield prediction analysis using feed forward and recurrent neural network. *International Journal of Innovative Science and Research Technology*, 5(5), 1092–1096.
- Suruliandi, A., Mariammal, G., & Raja, S. P. (2021). Crop prediction based on soil and environmental characteristics using feature selection techniques. *Mathematical and Computer Modelling of Dynamical Systems*, 27(1), 117–140.
<https://doi.org/10.1080/13873954.2021.1882505>
- Yan, Y., Wang, Y., Li, J., Zhang, J., & Mo, X. (2025). Crop Yield Time-Series Data Prediction Based on Multiple Hybrid Machine Learning Models. *Applied and Computational Engineering*, 133(1), 215–221.
<https://doi.org/10.54254/2755-2721/2025.20800>