Received: 20 SEPT 2024 | Accepted: 01 JUNE 2026 | Published Online: 30 JUNE 2026
Recoletos Multidisciplinary Research Journal Vol. 14 no. 1 (June 2026)
DOI: https://doi.org/10.32871/rmrj2614.01.02

Original Article

Smart Crop Selection and Irrigation Control Using Stacked Ensemble Learning

Vinay Kumar Enugala[1*], Srinivas Prasad[1]

[1] Computer Science and Engineering Department, GITAM Deemed To Be University, Andhra Pradesh, India

*Correspondence email: enugalavinay@gmail.com

Abstract

Background: *Precision agriculture has continued to grow as modern farming uses data to manage resources and improve crop yields. This research aims to improve the efficiency of modern agricultural approaches by providing innovative data analysis for effective and sustainable irrigation management.*
Methods: *The ensemble of a multi-layer perceptron (MLP) and an RF is performed using logistic regression as a learner. This ensemble model is used for crop prediction. The study encompasses various databases, including crop and nutrient mapping, sustainable irrigation datasets, and IoT sensor datasets with real-time records of soil moisture, temperature, and other relevant parameters.*
Results: *The research achieved an ensemble model accuracy of 99. 34%, greater than the traditional technique in terms of precision and time efficiency.*
Conclusion: *By leveraging IoT sensor data, this research improves the accuracy and proactivity of agricultural business practices and, therefore, supports the efficient use of agricultural resources.*

Keywords

precision agriculture, stacked ensemble learning, crop classification, smart irrigation, Internet of Things (IoT), machine learning, sustainable agriculture, feature engineering, UN SDG 15, Life on Land

INTRODUCTION

Over the past decade, agriculture has undergone a significant transformation driven by advances in data analytics, the Internet of Things (IoT) technologies, and machine learning (ML). These developments have facilitated a transition from conventional experience-based farming toward precision agriculture, in which real-time environmental data and predictive computational models guide crop management decisions. Precision agriculture aims to optimize resource use, enhance crop productivity, and reduce environmental impacts through intelligent monitoring and decision-support systems (Alahmad et al., 2023). Recent studies have demonstrated that integrating IoT sensing, remote sensing, and artificial intelligence has substantially improved the accuracy of crop prediction, irrigation scheduling, and soil resource management (Akhter & Sofi, 2021; Akkem et al., 2023). Moreover, contemporary precision agriculture systems are increasingly grounded in predictive analytics and data-driven decision-support frameworks, enabling more responsive, adaptive, and efficient farm management practices.

Accurate crop classification is a fundamental component of precision agriculture because it directly influences irrigation planning, fertilizer allocation, and yield estimation. Experts used remote sensing technologies—particularly optical and radar imaging—alongside machine learning algorithms to map crops across diverse agroecological conditions. Previous research has shown that models such as artificial neural networks, decision trees, and random forest classifiers can achieve high classification accuracy using spectral, spatial, and environmental features (Elavarasan & Vincent, 2020; Farmonov et al., 2023; Ikram et al., 2022; Kiruthika & Karthika, 2023; Rani et al., 2023; Talaat, 2023). Despite these advances, most existing approaches rely on single-model architectures, which may be sensitive to noise, prone to multicollinearity, and subject to dataset-specific biases, thereby limiting robustness and generalizability across heterogeneous agricultural environments. From a computational perspective, ensemble learning theory suggests that combining heterogeneous models can reduce individual model bias and variance, thereby improving predictive stability, generalization, and robustness across complex data environments.

Beyond crop classification, efficient irrigation management represents another critical challenge in modern agriculture, especially in regions facing increasing water scarcity. IoT-based smart irrigation systems that monitor soil moisture, temperature, humidity, rainfall, and atmospheric conditions have been extensively explored to optimize water usage and minimize wastage. Machine learning–driven irrigation decision-support systems have demonstrated significant potential to improve water-use efficiency while maintaining crop yields (Ikram et al., 2022; Kayum & Rahman, 2023; Kiruthika & Karthika, 2023; Rani et al., 2023). However, many of these systems depend heavily on historical datasets and often lack adaptability to rapidly changing environmental conditions. Furthermore, the computational complexity of deep learning–based models pose challenges for real-time implementation in resource-constrained farming environments. Although several studies have reported high predictive accuracies, many existing systems remain computationally intensive, functionally fragmented, or difficult to deploy in real-time agricultural settings.

In addition to crop classification and irrigation control, nutrient management is vital in sustaining soil health and agricultural productivity. Aligning crop selection with soil nutrient availability is essential for maximizing yield while reducing excessive fertilizer application and environmental degradation (Morales & Villalobos, 2023; Priyadharshini et al., 2021; Sharma et al., 2023). Although several studies have used crop nutrient databases in conjunction with machine learning to support data-driven fertilizer recommendations, most existing approaches treat crop prediction, nutrient assessment, and irrigation management as separate analytical tasks rather than as components of an integrated decision-making framework. This fragmented approach limits the development of comprehensive precision agriculture systems capable of supporting holistic and real-time farm management decisions.

Despite substantial progress in ML- and IoT-enabled precision agriculture, several important research gaps remain. Many existing studies lack real-time data integration, focus on isolated agricultural functions, or rely on computationally intensive deep learning models that limit scalability and practical deployment (Bhimavarapu et al., 2023; Jhajharia et al., 2023; Keerthana et al., 2021; Kuradusenge et al., 2023; Nikhil et al., 2024; Oikonomidis et al., 2022; Vignesh et al., 2023). Moreover, the potential of stacked ensemble learning—which combines heterogeneous base models through a meta-learning framework—remains relatively underexplored in precision agriculture, particularly for improving predictive robustness, stability, and computational efficiency within integrated agricultural decision support systems. Despite high classification accuracies reported in previous studies, limited research has examined the integration of crop prediction, nutrient assessment, and smart irrigation control within a unified ensemble-based decision-support framework. Furthermore, few studies have adequately balanced predictive performance, computational efficiency, and real-time applicability for resource-constrained agricultural environments.

To address these limitations, this study proposes OPTIRIGA-ML, an integrated precision agriculture framework that supports smart crop selection and sustainable irrigation management through stacked ensemble learning. The proposed approach combines multi-layer perceptron and random forest classifiers as base learners, with logistic regression serving as a meta-learner to enhance predictive performance. By jointly analyzing fused optical–radar remote sensing data, crop nutrient databases, and IoT-based environmental sensor data, the framework enables comprehensive, data-driven agricultural decision-making. Multicollinearity is addressed using variance inflation factor analysis, while dimensionality reduction via principal component analysis improves model stability and computational efficiency. Unlike many existing approaches that address agricultural tasks in isolation, the proposed framework integrates crop classification, nutrient-informed crop suitability analysis, and intelligent irrigation control within a single analytical architecture. By integrating crop classification, nutrient assessment, and irrigation recommendations within a unified analytical architecture, the proposed system aims to improve prediction accuracy, optimize resource use, and advance sustainable, scalable precision agriculture. More specifically, the study seeks to contribute to the development of computationally efficient, scalable, and real-time decision support systems for modern data-driven agriculture.

METHODS

Study Design and System Overview

This study adopts an experimental, data-driven design to develop and evaluate an integrated ML framework for crop classification and irrigation control. The OPTIRIGA-ML system consists of data acquisition, preprocessing, and feature engineering, model development, ensemble learning, and performance evaluation.

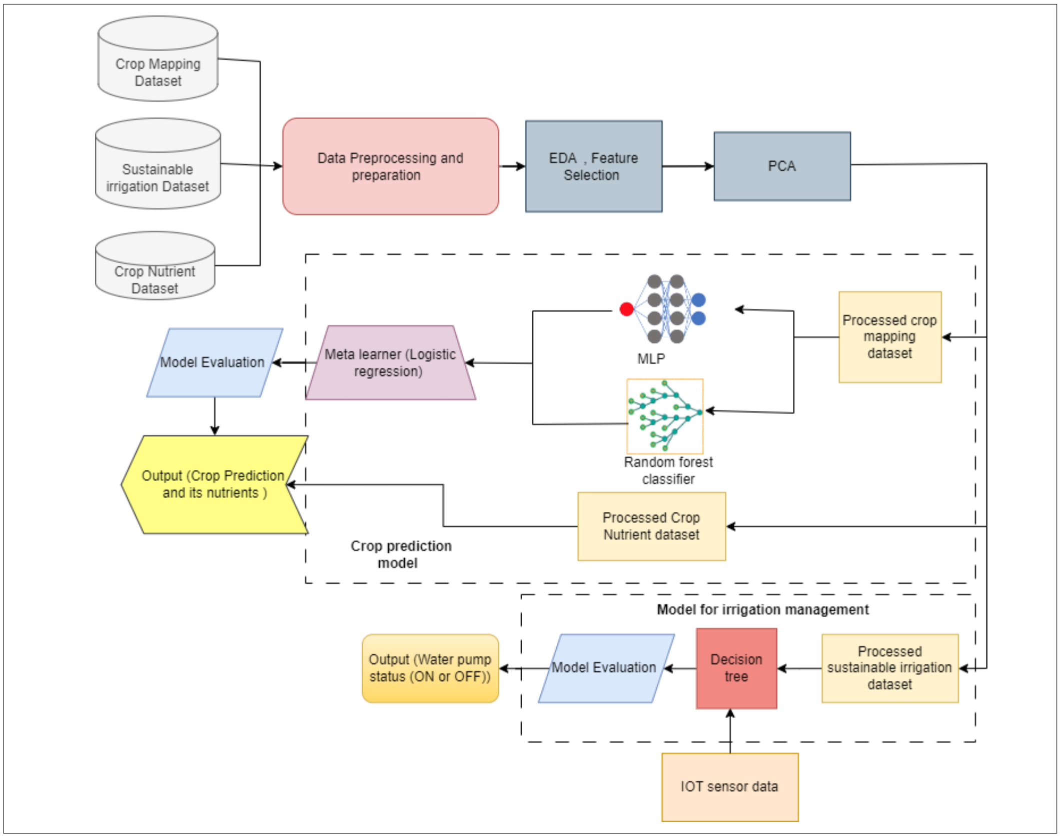

Figure 1. *System Architecture Diagram*

Data Sources

Crop Mapping Dataset: Crop mapping using a fused optical-radar dataset

This data set features combined optical and radar data for classifying croplands. The data collected by RapidEye satellites (optical and radar) European Space Agency (n.d.) over a farming area near Winnipeg, Canada, in 2012. This data set will serve as the basis for training machine learning models to predict crop types from spectral information and geographic attributes. Classifying crops such as corn, soy, wheat, and others helps this data set in agricultural planning, resource allocation, and yield estimation.

Crop nutrient database

The Crop Nutrient Dataset helps predict nutrient requirements of various crops by providing detailed nutrient profiles for each crop. It supports irrigation scheduling that matches each crop's requirements. This ensures ideal growing conditions, prevents nutrient loss, and improves crop quality. Overall, it promotes sustainable and efficient agriculture. It is an open-access database, created in collaboration with the International Fertilizer Association (IFA) and Wageningen University & Research (WUR), that provides data on nutrient concentrations, crop yields, residues, and related information for crops of nutritional and industrial significance (Crawford, 2019).

Sustainable Irrigation Dataset: Dataset for Predicting Watering the Plants

The sustainable irrigation data set includes IoT sensor data on temperature, humidity, soil moisture levels, and other environmental parameters relevant to predicting optimal irrigation schedules for agricultural fields. The researcher feeds the data into machine learning models that farmers use to automate or optimize irrigation by predicting the system's status (ON/OFF) (Harshil, 2020).

Data Preprocessing and Feature Engineering

Data pre-processing is a crucial step in preparing datasets for machine learning.

Handling Missing Values

To maintain data integrity, we identified and removed rows with missing values for the Crop Mapping Data set. This approach was feasible due to the large data set, ensuring sufficient data for analysis. In the sustainable irrigation data set, columns with high missing data rates were dropped, and rows with missing values were removed to ensure reliability.

Label Encoding

For the crop features, the labels were encoded numerically, with values from 1 to 7 corresponding to the different crops in the Crop Mapping Data set. We corrected these labels by applying bias and converting them to prevent the machine learning input format (0 to 6).

Normalization

The researcher used the StandardScaler from scikit-learn to normalize the features. This process ensures that each feature has a mean of zero and a standard deviation of one, calculated as

$$z = \frac{(x - \mu)}{\sigma} \qquad (1)$$

Where x is the original value, μ is the mean, and σ is the standard deviation. This step ensures that all features contribute equally during model training, thereby improving the model's performance and convergence during gradient descent.

Exploratory Data Analysis (EDA) and Feature Selection

In the Crop Mapping Data set, the researchers performed a wide-ranging correlation analysis as follows: using the correlation matrix and determining the degree of intercorrelation between the features. The identified features included characteristics with correlation coefficients larger than 0. Moreover, 90 is the number of predictors deemed highly correlated, chosen to eliminate multicollinearity problems since it reduces model efficiency.

The researcher then applied the Pearson correlation matrix to the sustainable irrigation data set to examine the interrelationships of the features. Based on the findings on multicollinearity, we used the variance inflation factor (VIF). They scrutinized all features with high VIF values and transformed them into other features to ensure that the dataset had the most features with low multicollinearity.

Dimensionality Reduction

Principal component analysis (PCA)

Principal component analysis (PCA) is a statistical method used for data reduction. It transforms a data set into a new coordinate system in which the variance of each variable under any data projection is maximized along the first coordinate (the first principal component), the second-largest variance is maximized along the second coordinate, and so on. PCA finds patterns in data and simplifies it, breaking it down into features that account for most of the variance in the data set. PCA transforms the original features into a set of linearly uncorrelated components that capture the maximum variance in the data.

Let X be the standardized data matrix of size $n \times p$, where n is the number of observations and p is the number of features.

Covariance matrix:
$$C = \frac{1}{n-1} X^T X \tag{2}$$

Eigenvalue Decomposition:
$$C v_i = \lambda_i v_i \tag{3}$$

Where λ_i is the eigenvalue, and v_i is the eigenvector of the covariance matrix C.

The eigenvectors corresponding to the largest eigenvalues are chosen as the principal components. The number of components retained is based on the cumulative explained variance ratio.

For the Crop Mapping Data set, we retained 15 principal components, while for the Sustainable Irrigation Data set, we retained five. We calculated the explained variance ratio to ensure the selected components captured a significant portion of the total variance. This step was crucial for maintaining information integrity while reducing dimensionality, thereby simplifying model training and improving computational efficiency.

Model Development and Training

After processing and splitting, the researchers applied the datasets to the models.

Multi-layer perceptron (MLP)

An MLP was trained using the Keras Sequential framework on the Crop Mapping Data set. The architecture included several fully connected layers, each using the 'leaky_relu' activation function to prevent vanishing gradients and allow small negative weights, enhancing learning. The final layer was a softmax layer, ideal for multi-class classification, as it normalizes outputs into a probability distribution across crop classes.

The model used the Adam optimizer for adaptive learning, helping achieve faster convergence. The loss function was sparse categorical cross-entropy, suitable for the multi-class classification task. The model was trained for 10 epochs with a batch size of 32, with a portion of the data reserved for validation to monitor performance and prevent overfitting. Accuracy and loss plots provided insight into learning progress and the potential for overfitting.

Let X_{train}^{ANN} denote the training data set for the ANN.

ANN predicts probabilities for each class using Softmax activation:

$$\mathrm{ANN_{output}} = \mathrm{Softmax}(\mathrm{W_{ANN}X + b_{ANN}}) \tag{4}$$

W_{ANN} are the weights, and b_{ANN} are the biases.

Soft max function converts logits (raw scores) into probabilities for multiple classes.

$$\mathrm{Softmax(z)_i} = \frac{\mathrm{e^{zj}}}{\sum_{\mathrm{k=1}}^{\mathrm{K}} \mathrm{e^{zk}}} \mathrm{forj} = 1\ldots,\ \ \mathrm{K} \tag{5}$$

Where z is a vector of logits and K is the number of classes.

Random forest classifier

In addition to the ANN, a random forest classifier was trained on the crop-mapping dataset. The random forest algorithm is an ensemble learning method that combines multiple decision trees to enhance classification accuracy and robustness. By averaging the predictions of multiple trees, the random forest reduces the risk of overfitting, which is typically associated with individual decision trees.

The RF model was configured with 10 estimators, each representing a separate decision tree. The model was then trained on the training data, leveraging the ensemble approach to improve the predictive performance.

RF final prediction:

$$\hat{y}_{\mathrm{rf}} = \arg\max_{k} \sum_{i=1}^{N} \hat{P}_{i,k} \tag{6}$$

Where $\hat{P}_{i,k}$ is the predicted probability from the i^{th} model for class k.

Meta learners (Logistic Regression)

The researcher employed a model stacking approach to improve the classification accuracy by combining the strengths of the ANN and the RF classifier. In model stacking, multiple base models are trained independently, and their predictions are then used as input features for a higher-level meta-learner. This meta-learner synthesizes the predictions from the base models to produce the final output.

Logistic regression was used as the meta-learner for the stacking ensemble. Logistic regression is a simple yet effective model for integrating predictions from different classifiers. By combining the outputs of the ANN and the random forest classifier, the logistic regression model can leverage their complementary strengths, potentially improving overall performance.

The prediction calculation for logistic regression:

$$\mathrm{Logit} = \mathrm{W} \bullet \mathrm{x} + \mathrm{b} \tag{7}$$

W is the weight matrix (with one column per class), and b is the bias vector.

This comprehensive training strategy, incorporating train-test splits, ANN, random forest, and model stacking, aimed to maximize the predictive accuracy and robustness of the crop-type classification model.

The decision tree classifier (DTC)

A decision tree classifier was developed with hyperparameter tuning using the Scikit-learn GridSearchCV module on the sustainable irrigation data set. GridSearchCV with 5-fold cross-validation identified the best parameters. The parameters optimized the model performance for predicting irrigation needs.

Experimental Setup and Validation Strategy

To ensure methodological rigor, the researcher adopted a standardized experimental setup and validation strategy for all machine learning models. The datasets were first randomly partitioned into training and test sets at an 80:20 split, with 80% of the data used for model training and 20% reserved for performance evaluation. This partitioning ensured that the testing data remained completely unseen during model development, thereby providing an unbiased estimate of predictive performance.

For models requiring hyperparameter optimization, such as the decision tree classifier, a five-fold cross-validation strategy was used. In this approach, the training dataset was divided into five equal subsets: four for training and one for validation. This process was repeated iteratively until each subset had served as a validation fold. The cross-validation procedure minimized overfitting, improved model generalization, and ensured robustness in parameter selection.

To prevent data leakage and maintain experimental integrity, all preprocessing steps—including normalization, feature selection, and principal component analysis—were applied exclusively to the training data and subsequently applied to the testing data using the learned transformation parameters. This process ensured that no information from the test dataset influenced model training. Additionally, class distributions were preserved during data splitting to maintain representative sampling across crop types and irrigation status categories.

All experiments were conducted using Python-based machine learning libraries, including Scikit-learn for traditional models and Keras/TensorFlow for neural network implementation. In this approach, the training dataset was divided into five equal subsets: four for training and one for validation. This standardized validation framework enabled reliable performance assessment and enhanced the transparency and reproducibility of the proposed precision agriculture system.

Model Evaluation Metrics

The input to the models is the data that has been pre-processed depending on the task to be solved. The Crop Mapping data set also contains spectral information and geographical properties for identifying crop types from optical radar sources. The sustainable irrigation data set includes the following attributes: temperature, humidity, and moisture data obtained from sensors, which are useful for predicting irrigation requirements. The Nutrients data set provides additional context on the environment for the forecasted crop types. For instance, after identifying crop types using machine learning models trained on optical radar data, the Nutrients set was used to extract more detailed nutrient descriptions of the recognized crops. The given data set associates specific crops with nutrient values and can be used for analysis and decision-making in agricultural management. By combining such information, decision-makers among farmers and researchers can choose suitable crops to cultivate, appropriate soil management methods, and the right nutrients for the plants, thereby maximizing productivity and minimizing failure due to improper soil management among other factors. The model evaluation results are then determined based on accuracy, precision, recall, and F1 score. The metric formula provides a clear, understandable definition.

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}} \tag{8}$$

It provides an overall assessment of how well a model performs across all classes.

$$\text{Precision} = \frac{\text{TruePositives}}{\text{TruePositives} + \text{FalsePositives}} \tag{9}$$

$$\text{Recall} = \frac{\text{TpruePositives}}{\text{TruePositives} + \text{FalsePositives}} \tag{10}$$

$$\text{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{11}$$

Algorithm 1: OPTIRIGA-ML Model for Crop Prediction and Sustainable Irrigation

Input: X_train_cm, y_train_cm, X_test_cm, y_test_cm (Crop Mapping dataset),
Output: final_predictions, accuracy, classification report

Initialization:
1: Import necessary libraries and modules
2: Load dataset (cm_df)
Process:
3: Preprocessing Dataset:
Remove highly correlated features using a correlation matrix
- Compute correlation matrix: *R = corr(X)*
if *|R(i, j)| > 0.90*
- Drop one feature from each pair of highly correlated features
Perform Principal Component Analysis (PCA) for feature reduction
- Standardize features: *X_std = (X - mean(X))/std(X)*
- Compute covariance matrix: *C = cov(X_std)*
- Compute the eigenvalues and eigenvectors of *C: (λ, V) = eig(C)*
- Select the top n components: *X_pca = X_std * V(:, 1:n)*
- *(X_train_cm, X_test_cm, y_train_cm, y_test_cm)* = train_test_split*(X_pca, y, test_size=0.2)*
7: Train ANN:
Initialize the ANN model with layers and activation functions
- ANN layers: [Dense (16), Dense (64), Dense (256), Dense (128), Dense (32), Dense (7)]
- Activation functions: Leaky ReLU for the hidden layers and softmax for the output layer
Comparison of the ANN model with the optimizer and loss function
Optimizer = Adam, loss = sparsity_categorical_crossentropy
The ANN model was fit to X_train_cm and y_train_cm
- θ_ANN = argmin (L *(θ; X_train_cm, y_train_cm)*)
Evaluation of the ANN model on the validation data (optional)
- Validation loss and accuracy
8: Train random forest classifier (RF):
Initialize the RF model with hyperparameters
- RF parameters: n_estimators = 10
- θ_RF = argmin (L *(θ; X_train_cm, y_train_cm)*)
9: Generate Predictions for Meta-learners:
Generate predictions from ANN on training data: preds1_train
- preds1_train = ANN (X_train_cm; θ_ANN)
Generate predictions from RF on training data: preds2_train
- preds2_train = RF (*X_train_cm; θ_RF*)
Combined predictions as meta-features for training meta-learners: train_meta_features
- train_meta_features = [preds1_train, preds2_train]
Generate predictions from ANN on testing data: preds1_test
- preds1_test = ANN (X_test_cm; θ_ANN)
generate predictions from RF on testing data: preds2_test
- preds2_test = RF (X_test_cm; θ_RF)
Combined predictions as meta-features for testing meta-learners: test_meta_features
- test_meta_features = [*preds1_test, preds2_test*]
10: Train Meta-learner (Logistic Regression):
Initialize the logistic regression model
Logistic regression model: L.R.
Fit the logistic regression model to train_meta_features and y_train_cm
- θ_LR = argmin (L (θ; train_meta_features, y_train_cm))
11: Evaluate Ensemble Model:
Generate final predictions using the meta-learner on test_meta_features
- Final predictions = LR (test_meta_features; θ_LR)
Calculate accuracy and classification report comparing final predictions with y_test_cm
Return: final_predictions, accuracy, classification report

RESULTS

This section presents a comprehensive evaluation of the proposed precision agriculture framework using three heterogeneous datasets: crop mapping, crop nutrient, and sustainable irrigation datasets. Advanced machine learning models—including Artificial Neural Networks (ANNs), Random Forest Classifiers (RFCs), and Decision Tree (DT) classifiers—were employed in conjunction with hyperparameter optimization, feature selection, dimensionality reduction, and ensemble learning strategies to enhance predictive accuracy, robustness, and generalization. Model performance was assessed using standard classification metrics, including accuracy, precision, recall, F1-score, and detailed classification reports.

Exploratory Data Analysis of Crop Mapping and Sustainable Irrigation Datasets

Figure 2 depicts a correlation matrix, visualized as a heat map, showing the relationships between variables. The color scale ranges from blue (indicating strong positive correlation) to red (indicating strong negative correlation), with yellow representing weak or no correlation. The triangular format of the matrix displays correlations only once to avoid redundancy. Variables are listed on the x- and y-axes, with each cell representing the correlation coefficient between the intersecting variables. This matrix is essential for identifying the dataset's patterns, dependencies, and multicollinearity, and for guiding further data analysis and feature selection.

Figure 3 displays a Pearson correlation matrix of the variables. The matrix is color-coded, with deeper purple hues indicating stronger correlations (both positive and negative). The diagonal elements, all equal to 1, represent the correlations among variables. The matrix helps identify relationships between variables, such as which pairs move together and the strength of those relationships, which is useful for data analysis and feature selection.

Table 1 summarizes the Variance Inflation Factors (VIF) computed to quantify multicollinearity among predictor variables in the sustainable irrigation dataset. Several features exhibit severe multicollinearity, notably pressure (217.87), air temperature (92.93), and pH (75.17), indicating strong linear dependence on other predictors. Moderate multicollinearity is also observed in soil humidity, air humidity, and wind gust speed.

High VIF values can inflate coefficient variance, degrade model stability, and compromise interpretability. To mitigate these effects, dimensionality reduction and feature selection strategies—including PCA, VIF-based elimination, and regularization-aware modelling—were adopted to improve model reliability and predictive performance.

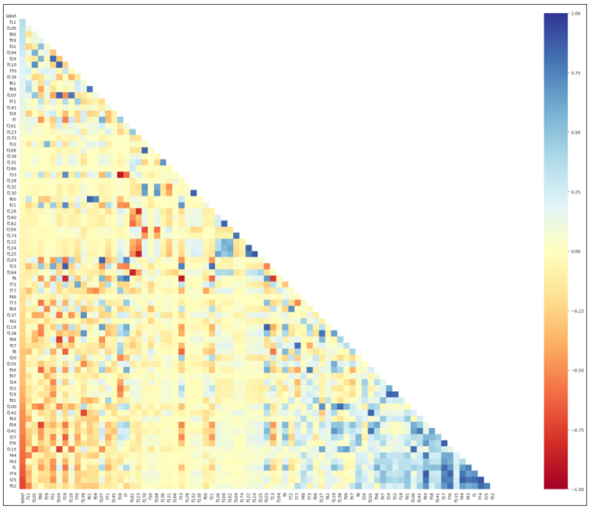

Figure 2. *Correlation Matrix of the Crop Mapping Dataset*

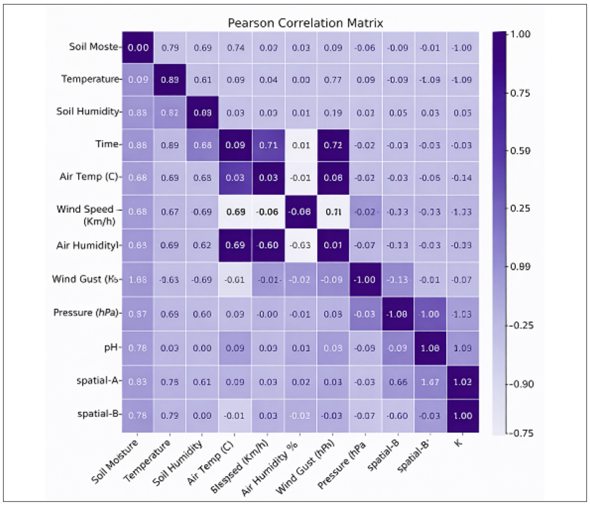

Figure 3. *Pearson Correlation Matrix for the Sustainable Irrigation Dataset*

Table 1. *Variance Inflation Factors for the Sustainable Irrigation Dataset*

	Feature	VIF
0	Soil Moisture (%)	4.112660
1	Temperature (oC)	3.985566
2	Soil Humidity (%)	10.412162
3	Time (in ms)	3.891616
4	Air temperature (oC)	92.934053
5	Wind speed (km/h)	41.029239
6	Air humidity (%)	12.869436
7	Wind gust (km/h)	20.787565
8	Pressure (kPa)	217.870085
9	pH	75.173476
10	Rainfall (in Cm)	5.106048
11	N	3.336565
12	P	9.010659
13	K	4.952706

PCA Component Analysis

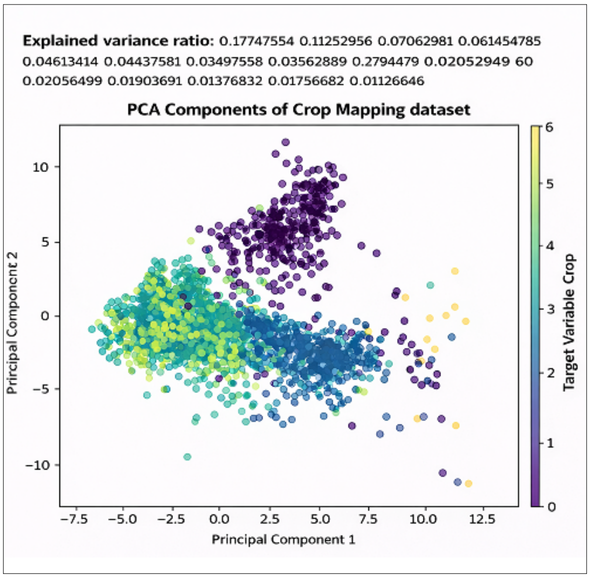

Figure 4. *PCA Components for the Crop Mapping Dataset*

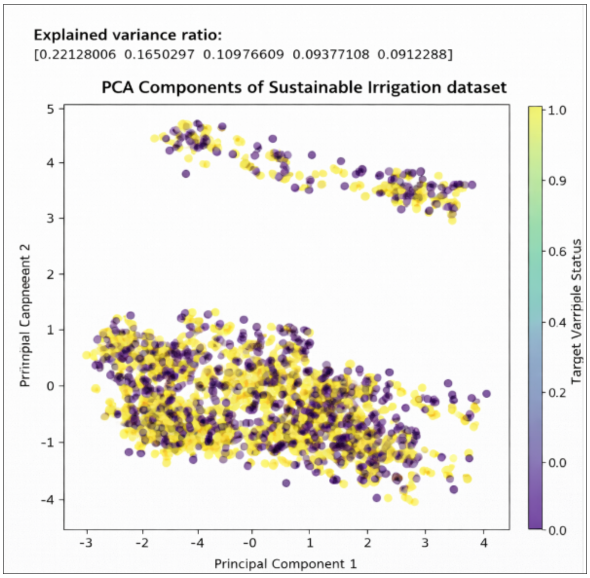

Figure 5. *PCA Component of the Sustainable Irrigation Dataset*

Figure 4 shows a PCA scatter plot applied to the crop mapping dataset, with points labeled by the "Target Variable Crop." The color spectrum from yellow to purple highlights different crop characteristics. PCA components compress the data, maximizing variance to reveal structure and patterns. The coefficient of determination indicates the variance captured by each principal component, showing the importance of the first few components. The scatter plot reveals clear groupings, indicating that PCA effectively distinguishes between crop types and helps detect patterns and relationships in the dataset.

PCA of the sustainable irrigation dataset. Every point is colored according to the 'Target Variable Status,' with a transition from yellow to purple indicating different levels of the target variable.

The variance ratio mentioned above indicates the extent to which each model component explains the total variance. Principal Component 1 captures approximately 22.12% of the variance, Principal Component 2 captures 16.50%, and the other components contribute a decreasing amount of 10. 87%, 9. 37%, and 8. 91%, respectively. In data analysis, only the first several principal components usually contain the most information about the relationships between values.

Model Evaluations Results

Table 2. *Classification Report for the Ensemble Model*

	Precision	Recall	f1-score	Support
Broadleaf	0.94	0.84	0.88	244
Canola	1.00	1.00	1.00	15130
Corn	0.99	1.00	0.99	7857
Oat	0.99	0.99	0.99	9506
Pea	0.99	0.98	0.99	690
Soy	0.99	0.99	0.99	14747
Wheat	0.99	0.99	0.99	16993
Accuracy	0.98	0.97	0.99	65167
Macro avg	0.99	0.97	0.98	65167
Weighted avg	0.99	0.99	0.99	65167

Table 3. *Classification Report for the DT Model (Water Pump Action Prediction)*

	Precision	Recall	f1-score	Support
0	0.90	0.94	0.92	213
1	0.94	0.91	0.92	227
Accuracy (macro avg)	0.92	0.92	0.92	440
Weighted avg	0.92	0.92	0.92	440

The ensemble model achieved an overall accuracy of 99.34%, demonstrating strong performance in distinguishing between crop types. Precision ranged from 0.94 to 1.00, with high recall and F1 scores across all crop classes. The model's precision in identifying each crop type and its ability to retrieve actual positives were consistently high, with most F1 scores ranging from 0.88 to 1.00. The finding validates the model's effectiveness for practical farming, enabling efficient resource management through crop categorization.

For water pump action prediction, the DT model demonstrated a solid accuracy of 92%. It achieved 0.90 precision for the "pump OFF" class (213 instances) and 0.94 for the "pump ON" class (227 instances). The general accuracy of 92% across 440 instances confirms the model's reliability in predicting pump actions based on the sustainable irrigation dataset.

Model Outputs

Figure 6 illustrates the output of the ensemble crop prediction model. Based on optical–radar data analysis, the model identified Canola (Brassica sp.) as the optimal crop under the given conditions. The predicted average moisture content (8.93%) indicates favorable harvesting conditions. Nutrient profiling revealed average dry-state concentrations of nitrogen (3.88%), phosphorus (0.62%), and potassium (0.98%), which are essential for crop growth and yield stability. Additionally, a protein content of 27.9% underscores Canola's agronomic and nutritional suitability. Furthermore, the protein content of Canola seeds was approximately 27.9% in the dry state, highlighting their nutritional richness and suitability for agricultural production under the specified conditions. Hence, the model output predicts the crop and its corresponding nutrient values.

Using data from IoT sensors, the decision tree model predicts whether to turn the water pump ON or OFF for sustainable irrigation. Based on the given input values—soil moisture at 31°C, temperature at 17°C, soil humidity at 61%, time at 17 hours, air humidity at 10.64%, wind gust at 21.82 km/h, rainfall at 206.4 mm, and nutrient levels (N: 47.0, P: 46.0, K: 52.0)—the model has recommended turning the water pump OFF. This data suggests that the current soil and environmental conditions provide sufficient moisture and nutrients, negating the need for additional irrigation.

```
For the given input from optical-radar dataset, the model has predicted the suitable crop to be:
Canola

The nutrients data related to the crop Canola is as follows

ScientificName :              Brassica sp.
Symbol :                      BRASS2
NuContAvailable :             N,P,K
PlantPartHarvested :          Seed
CropCategory :                Cereal and oil
YieldUnit :                   cwt
AvYieldUnitWeight(lb) :       100.0
AvMoisture% :                 8.93
AvN%(dry) :                   3.8887366483
AvP%(dry) :                   0.62
AvK%(dry) :                   0.98
YieldUnitWeight(lb)_set :     100
DryMatter%_M-FF :             90.5
DryMatter%_NAS :              90.5
DryMatter%_F&L :              92.2
AvDryMatter% :                91.07
N%(dry)_NAS :                 3.6
N%(dry)_F&L :                 4.464
N%(dry)_M-FF :                3.6022099448
P%(dry)_NAS :                 0.61
P%(dry)_F&L :                 0.63
K%(dry)_F&L :                 0.98
Protein%(dry)_NAS :           22.5
Protein%(dry)_F&L :           27.9
N%(wet)_M-FF :                3.26
```

Figure 6. *Output of the Ensemble Model (Crop Prediction and Nutrients Related to Crops)*

```
Now the decision tree model trained on sustainable irrigation dataset, will predict
whether to turn the water pump ON or OFF based on the information collected by the
IoT sensors. These information include:

Soil Moisture :               31
Temperature :                 17
 Soil Humidity :              61
Time :                        17
Air humidity (%) :            10.64
Wind gust (Km/h) :            21.82
rainfall :                    206.4
N :                           47.0
P :                           46.0
K :                           52.0
Volume Equivalent :           0.23

For the above given information (input), the model has predicted that, you need to turn OFF
the water pump.
```

Figure 7. *Inputs and Outputs of the DT model*

DISCUSSION

The research methodology utilized big data analysis and precision machine learning to optimize crop selection and irrigation, showcasing the potential of precision agriculture. Using datasets for crop mapping, nutrients, and sustainable irrigation, the study employed ANN, random forest classifiers (RFCs), and decision tree classifiers (DTs) with hyper-parameter tuning and ensemble techniques. The ensemble model achieved an average accuracy of 99.34%, with high precision, recall, and F1 scores across all crop types, demonstrating its effectiveness for real-world agricultural applications. The decision tree model for water pump action achieved 92% accuracy, aiding irrigation decision-making based on IoT sensor data.

Table 4. *Comparison of the Proposed Model with an Existing Model*

Reference	Model for Prediction	Accuracy
Rani et al. (2023)	LSTM RNN (Weather Prediction), Random Forest Classifier (Crop Selection)	97.235% (Crop Selection)
Kiruthika and Karthika (2023)	IDCSO-WLSTM	92.68%
Elavarasan and Vincent (2020)	Deep Recurrent Q-Network	93.7%
Jhajharia et al. (2023)	Random Forest,	96.3% (RF.)
Keerthana et al. (2021)	Decision Tree, Ada Boost Regressor, Random Forest Classifier	95.7% (DT + Ada Boost), 95.0% (DT + RF)
Proposed Model	Stacking (MLP+RF) using logistic regression as a meta-learner	99.34%

Several recent studies have investigated machine learning and deep learning techniques for crop prediction and decision support in precision agriculture, demonstrating varying levels of accuracy, scalability, and practical applicability. Rani et al. (2023) employed a hybrid framework combining an LSTM-based recurrent neural network for weather prediction with a Random Forest classifier for crop selection, achieving 97.235% accuracy; however, the sequential nature of LSTM models introduces substantial computational complexity and longer training times, which may hinder real-time deployment in resource-constrained environments. Similarly, Kiruthika and Karthika (2023) proposed an IoT-driven crop recommendation system using an IDCSO-WLSTM model, achieving 92.68% accuracy. While optimization-enhanced deep learning improved convergence, the comparatively lower performance and high computational overhead raise concerns regarding generalization and scalability for smallholder farming systems. Reinforcement learning–based approaches, such as the Deep Recurrent Q-Network proposed by Elavarasan and Vincent (2020), achieved 93.7% accuracy and offer adaptive decision-making capabilities; however, their reliance on extensive interaction data and lengthy training cycles limits immediate applicability in dynamic, real-world agricultural settings.

In contrast, traditional machine learning models have demonstrated competitive performance with reduced computational cost. Jhajharia et al. (2023) achieved 96.3% accuracy using a Random Forest model, highlighting its robustness to noise and nonlinear relationships. However, single-model architectures may exhibit bias and limited adaptability under heterogeneous conditions. Keerthana et al. (2021) further improved decision tree performance using AdaBoost and Random Forest ensembles, achieving accuracies of 95.7% and 95.0%, respectively; however, these homogeneous ensemble strategies may fail to exploit complementary feature representations fully. Building on these insights, the proposed model adopts a stacked ensemble architecture that integrates MLP and Random Forest base learners with logistic regression as the meta-learner, achieving the highest reported accuracy of 99.34%. This improvement stems from the complementary learning capabilities of neural and tree-based models, with the meta-learner effectively reducing bias and variance. Unlike deep recurrent or reinforcement learning frameworks, the proposed approach balances predictive performance and computational efficiency, enabling scalable, interpretable, and real-time deployment for IoT-enabled precision agriculture, thereby positioning it as a robust solution for high-accuracy agricultural decision support in resource-limited environments.

CONCLUSION

This study validates the effectiveness of machine learning–based precision agriculture by designing and evaluating an ensemble framework integrating MLP, Random Forest, and Decision Tree models, achieving a predictive accuracy of 99.34%. The proposed system enables reliable crop classification, nutrient-informed crop suitability assessment, and intelligent irrigation control, thereby improving resource utilization, water conservation, and evidence-based agricultural decision-making. Despite the strong performance of the ensemble approach, certain limitations persist, including dependence on historical data, sensitivity to dynamic environmental and crop health conditions, and the computational complexity of advanced models in resource-limited agricultural settings. These constraints can be addressed through continuous data acquisition, adaptive model retraining, and deployment-oriented optimization strategies. Beyond technical outcomes, the OPTIRIGA-ML framework has the potential to significantly impact agronomy and the economy by improving crop yields, reducing production costs, increasing farmers' income, and promoting sustainable practices such as water efficiency and reduced fertilizer use. Future research should emphasize integrating real-time IoT and satellite data, exploring reinforcement and deep learning techniques, and incorporating climate adaptation mechanisms to enhance scalability, robustness, and long-term sustainability in precision agriculture.

Author Contributions

V. K. Enugala: Conceptualization, Methodology, Software, Data curation, Writing-Original draft preparation, Visualization and Investigation; **S. Prasad:** Supervision, Validation and Review.

Funding

This research received no external funding.

Ethical Approval

Not applicable.

Competing interest

The authors declare no conflicts of interest.

Data Availability

Data will be made available by the corresponding author on request.

Declaration of Artificial Intelligence Use

In this work, the author's utilized ChatGPT to assist with improving language clarity and structuring the initial literature review. After using this tool/service, the authors evaluated and revised the content as necessary and takes full responsibility for the published content.

REFERENCES

Akhter, R., & Sofi, S. A. (2021). Precision agriculture using IoT data analytics and machine learning. Journal of King Saud University - Computer and Information Sciences, 34(8), 5602–5618. https://doi.org/10.1016/j.jksuci.2021.05.013

Akkem, Y., Biswas, S. K., & Varanasi, A. (2023). Smart farming using artificial intelligence: A review. Engineering Applications of Artificial Intelligence, 120, 105899. https://doi.org/10.1016/j.engappai.2023.105899

Alahmad, T., Neményi, M., & Nyéki, A. (2023). Applying IoT sensors and big data to improve precision crop production: A review. Agronomy, 13(10), 2603. https://doi.org/10.3390/agronomy13102603

Bhimavarapu, U., Battineni, G., & Chintalapudi, N. (2023). Improved optimization algorithm in LSTM to predict crop yield. Computers, 12(1), 10. https://doi.org/10.3390/computers12010010

Crawford, C. (2019). Crop nutrient database. Kaggle. https://www.kaggle.com/datasets/crawford/crop-nutrient-database
Elavarasan, D., & Vincent, P. M. D. (2020). Crop yield prediction using deep reinforcement learning model for sustainable agrarian applications. IEEE Access, 8, 86886–86901. https://doi.org/10.1109/access.2020.2992480
European Space Agency. (n.d.). RapidEye ESA archive. https://earth.esa.int/eogateway/catalog/rapideye-esa-archive
Farmonov, N., Amankulova, K., Szatmári, J., Sharifi, A., Abbasi-Moghadam, D., Mahdi, S., & Mucsi, L. (2023). Crop type classification by DESIS hyperspectral imagery and machine learning algorithms. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 16, 1576–1588. https://doi.org/10.1109/jstars.2023.3239756
Harshil. (2020). Intelligent irrigation system. Kaggle. https://www.kaggle.com/datasets/harshilpatel355/autoirrigationdata
Ikram, A., Aslam, W., Aziz, R. H. H., Noor, F., Mallah, G. A., Ikram, S., Ahmad, M. S., Abdullah, A. M., & Ullah, I. (2022). Crop yield maximization using an IoT-Based smart decision. Journal of Sensors, 2022(1). https://doi.org/10.1155/2022/2022923
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
Kayum, A., & Rahman, M. O. (2023). Appropriate crop recommended system for cultivation using IoT and ML. 2023 5th International Conference on Sustainable Technologies for Industry 5.0 (STI). https://doi.org/10.1109/sti59863.2023.10464865
Keerthana, M., Meghana, K. J. M., Pravallika, S., & Kavitha, M. (2021). An ensemble algorithm for crop yield prediction. 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), 963–970. https://doi.org/10.1109/icicv50876.2021.9388479
Kiruthika, S., & Karthika, D. (2023). IOT-BASED professional crop recommendation system using a weight-based long-term memory approach. Measurement: Sensors, 27, 100722. https://doi.org/10.1016/j.measen.2023.100722
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
Morales, A. G., & Villalobos, F. J. (2023). Using machine learning for crop yield prediction in the past or the future. Frontiers in Plant Science, 14. https://doi.org/10.3389/fpls.2023.1128388
Nikhil, U. V., Pandiyan, A. M., Raja, S. P., & Stamenkovic, Z. (2024). Machine learning-based crop yield prediction in south india: Performance analysis of various models. Computers, 13(6), 137. https://doi.org/10.3390/computers13060137
Oikonomidis, A., Catal, C., & Kassahun, A. (2022). Hybrid deep learning-based models for crop yield prediction. Applied Artificial Intelligence: An International Journal, 36(1), 1–18. https://doi.org/10.1080/08839514.2022.2031823
Priyadharshini, A., Chakraborty, S., Kumar, A., & Pooniwala, O. R. (2021). Intelligent crop recommendation system using machine learning. 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), 843–848. https://doi.org/10.1109/ICCMC51019.2021.9418375
Rani, S., Mishra, A. K., Kataria, A., Mallik, S., & Qin, H. (2023). Machine learning-based optimal crop selection system in smart agriculture. Scientific Reports, 13(1), 15997. https://doi.org/10.1038/s41598-023-42356-y
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
Talaat, F. M. (2023). Crop yield prediction algorithm (CYPA) in precision agriculture based on IoT techniques and climate changes. Neural Computing & Applications, 35(23), 17281–17292. https://doi.org/10.1007/s00521-023-08619-5
Vignesh, K., Askarunisa, A., & M. Abirami, A. (2023). Optimized deep learning methods for crop yield prediction. Computer Systems Science and Engineering, 44(2), 1051–1067. https://doi.org/10.32604/csse.2023.024475

How to cite this article:
Enugala, V. K., & Prasad, S. (2026). Smart crop selection and irrigation control using stacked ensemble learning. *Recoletos Multidisciplinary Research Journal, 14*(1), 13–27. https://doi.org/10.32871/rmrj2614.01.02