Ensuring Robustness in IoT-Based Precision Agriculture: A Stacked Ensemble Model Resilient to Sensor Noise and Data Failures

Iftikhar Hussain, Muhammad Afzal, Sadam Hussain, Benish, Ali Raza Sheikh.
Department of Computer Science & Information Technology, Ghazi University, Dera Ghazi Khan
*Correspondence: iftikharh014@gmail.com

Citation | Hussain. I, Afzal. M, Hussain. S, Benish, Sheikh. A. R, "Ensuring Robustness in IoT-Based Precision Agriculture: A Stacked Ensemble Model Resilient to Sensor Noise and Data Failures", IJIST, Vol. 8 Issue. 3 pp 1139-1165, June 2026
Received| April 17, 2026 **Revised|** May 20, 2026 **Accepted|** May 27, 2026 **Published|** June 09, 2026.

The increasing need for sustainability in agriculture has driven the use of data-based techniques in crop management and fertilizer recommendations. Conventional methods, which depend on either a static set of rules or basic machine learning classification algorithms, are sometimes inadequate in addressing the issues presented by a dynamic agricultural environment. In this paper, a resilient and hierarchical stacked ensemble method that can provide precise, reliable, and uncertainty-aware fertilizer recommendations for precision agriculture is proposed. The methodology involves combining three gradient boosting learners: XGBoost; LightGBM; and CatBoost under a Ridge Regression model, using stratified five-fold cross-validation. Hyperparameters are optimized through Bayesian optimization with the Optuna library, and the Genetic Algorithm is employed as an optimization algorithm after prediction to reduce the fertilizer amount used while maintaining the nutrient content of the soil. Newly engineered features include environment-based indices such as Heat Index, Climate Stress Index, Soil Fertility Index, and different ratios of nutrients to enrich the existing feature set. Dimensionality reduction in terms of IoT deployment feasibility is ensured by Recursive Feature Elimination. Explainability is achieved by interpreting SHAP values, while uncertainty is quantified using confidence intervals. Using the public Crop Recommendation dataset consisting of 2,200 samples in 22 crop classes, the classification results for the proposed framework show an accuracy of 93.44%, F1 score of 93.47%, Cohen's Kappa of 0.9313, and Matthews Correlation Coefficient of 0.9315. The confidence of this model stands at 92.02%, whereas its accuracy against 5% Gaussian noise is 91.57%. The proposed stacked ensemble results in higher accuracy (93.44%) as compared to a baseline Random Forest model (93.21%) with a smaller number of features (50%). Bayesian optimization using the Optuna algorithm achieved the best cross-validated F1 score of 0.9520. Optimization using the Genetic Algorithm decreased the nitrogen dosage by 2.4% (29.00 → 28.31) and kept the target levels of phosphorus and potassium fixed. The average prediction uncertainty was 7.98% with the average confidence of 92.02% for 22 crop classes.

Keywords: Precision Agriculture; Stacked Ensemble Learning; Genetic Algorithm; SHAP; Explainable AI; Fertilizer Recommendation; Bayesian Optimization

Introduction:

Agriculture, however, is still among the most important sectors that contribute to human life on this planet, meeting needs of billions of people on the planet, and contributing greatly to the global economy. Taking into account that the world population will reach ten billion by 2050, the challenges facing agriculture become even more serious [1]. In addition to this, the excessive use of chemical fertilizers has also been responsible for degrading soil, polluting water resources, emission of greenhouse gases, and depletion of the productive capacity of farmland [2]. It is therefore clear that a paradigm shift needs to be brought about in decision making related to agriculture. The advent of precision agriculture represents one solution paradigm to these problems due to its reliance on data-based technology in order to tailor management decisions concerning crops and soil on a highly spatially and temporally detailed level [3]. The recent development of IoT sensor networks has further advanced the use of data-driven precision agriculture, facilitating the acquisition of soil and environment parameters in real time at field scale. The recent development of IoT sensor networks has further advanced the use of data-driven precision agriculture, facilitating the acquisition of soil and environment parameters in real time at field scale [4]. Instead of applying blanket solutions in a field, precision agriculture aims to apply different fertilizers, water, and chemicals according to local circumstances, thus maximizing efficiency and reducing waste. This is entirely dependent upon the sophistication of the recommendation systems that inform management decision-making [5].

Machine learning has been established as the prevailing tool in developing intelligent recommendation systems for agriculture, showing high efficacy with both supervised classification and regression approaches [6]. Numerous methods ranging from simple decision trees to more complex deep learning systems have been explored to solve a variety of agronomic issues like crop prediction, soil nutrient classification, yield estimation, and crop irrigation [7]. There is a considerable difference between the efficiency of such models during experiments conducted under laboratory conditions and when implemented in practice [8]. The approach of ensemble learning has gained widespread recognition since it is one of the most efficient approaches to improving the classification prediction performance [9]. Since ensemble learning methods decrease both variance and bias from the prediction results of several base classifiers, they make it possible to achieve greater accuracy and generalization than a single predictor alone [10]. Stacking allows developing a meta-classifier that will define the best way to integrate the prediction results from several base classifiers, making use of their complementary nature [11]. Algorithms of gradient boosting, including such implementations as XGBoost [12], LightGBM [13], and CatBoost [14], became state-of-the-art methods of tabular classification.

Some issues still arise with regard to the application of machine learning algorithms in precision farming in spite of all the literature present. The systems that have been built have been tested under perfect circumstances while disregarding such aspects as absence of data cleansing, loss of data, or incorrect data coming from the sensors [15]. What is more, there are still examples of researchers who did not try to use the techniques of optimizing hyperparameters with the help of Bayesian methods [16]. The main difficulty stems from the fact that the majority of the systems that have been developed generate deterministic results without measuring uncertainties, and thus, it is quite risky since making decisions relies heavily on the information provided by the system [17]. The need for explainability in AI models employed in agriculture has gained recognition. The SHAP model (Shapley Additive explanations) is a mathematical framework derived from cooperative game theory, and it is a feature attribution method that is not tied down to any specific machine learning model [17]. Through the use of ensemble pipelines along with SHAP, it becomes possible for experts in agriculture to determine which soil and environmental factors are more responsible for generating particular fertilizer recommendations.

Another aspect that has not yet received much attention in precision agriculture is the optimal amount of fertilizer needed. Classification algorithms allow predictions regarding whether to use any fertilizer and which one; however, determining the amount that results in the highest yield while minimizing harm to the environment is a complicated multi-objective problem [18]. Genetic Algorithms have proven useful in solving various optimization tasks in agriculture, including those involving non-linear, non-convex, multi-dimensional parameter spaces [19]. Adding a genetic optimization step downstream of a classification pipeline is therefore a logical step in closing the prediction gap.

The rest of this paper is structured as follows. In Section 2, a detailed literature review is carried out on machine learning for precision agriculture, ensembles, explainability, and optimization. In Section 3, the proposed approach is described in detail. In Section 4, the results of the experiments conducted are reported and analyzed. In Section 5, conclusions are drawn.

Research Objectives:

The primary objective of this research is to develop an intelligent and resilient precision agriculture framework capable of generating accurate crop and fertilizer recommendations under realistic IoT operating conditions. Specifically, the study aims to improve prediction accuracy through stacked ensemble learning, enhance robustness against sensor failures and noisy data, provide transparent decision-making through explainable artificial intelligence, quantify prediction uncertainty, and optimize fertilizer dosage using evolutionary optimization techniques. The novelty of the proposed approach is that all these elements have been integrated in an integrated framework including robust data pre-processing, domain-specific feature engineering, hyperparameter optimization using Bayesian optimization, stacking ensembles learning, model explainability via SHAP technique, uncertainty quantification, and fertilizer optimization based on Genetic Algorithm. In contrast to the previous research efforts that were mainly concerned with the predictive performance, the proposed framework tackles multiple challenges at the same time.

Literature Review:

Use of computational intelligence and machine learning in agriculture-related problems has attracted many researchers' attention in the last ten years. In this paper, we review related studies in five thematic topics that include machine learning for crop and fertilizer recommendation, ensemble learning for agricultural applications, data preprocessing, hyperparameter tuning, and explainable artificial intelligence in agriculture.

Machine Learning for Crop and Fertilizer Recommendation:

Data-driven crop recommendation techniques have traditionally used rule-based expert systems that encapsulate agronomic information obtained from field experiments. Although easily comprehensible, such systems failed to possess sufficient flexibility to cope with the non-linear correlation among various interrelated factors [1]. Supervised machine learning methods started challenging rule-based models in the early 2010s with decision trees, k-nearest neighbors, and naive Bayes algorithms achieving decent accuracy on structured agronomic data [6]. Recent research has greatly improved the complexity of recommendation systems. Random forest type classifiers perform better than single decision tree classifiers in terms of classification accuracy of crop recommendation, reaching high accuracy rates over 90% on standard test datasets [7]. Support vector machines utilizing radial basis kernel functions have proven their ability to successfully generalize in multi-class crop classification tasks [20]. Deep learning models including multi-layer perceptron networks and one-dimensional CNNs have also been implemented successfully, though they require large-sized datasets and are prone to overfitting on small agricultural examples [21].

A two-stage recommendation system based on classification for the first step and regression for the second step was suggested by [22]. A study conducted by Rusool et al. explored the use of AI-based intelligent farming for efficient resource utilization, using a combination of Random Forest and CatBoost classifiers for crop and fertilizer recommendations [23]. Although this research proved that a hybrid model could be used to achieve precision farming, it did not account for data impurities, did not carry out uncertainty estimation, and utilized regular label encoding without complex feature engineering techniques. This paper aims to build on these findings.

Ensemble Learning in Agricultural Applications:

In agriculture, ensemble approaches are superior to those using only one model in most applications. Gradient boosting algorithms have shown supremacy in competitions involving tabular datasets and practical applications. The proposed XGBoost framework by Chen & Guestrin uses regularization to avoid overfitting by limiting tree complexity [12]. The development of leaf-wise growing decision tree with gradient-based one-side sampling and exclusive feature bundling by Ke et al. made LightGBM more efficient than XGBoost in terms of computational cost, while retaining competitive accuracy [13]. The CatBoost framework uses ordered boosting and symmetric trees, avoiding prediction shift. Stacked generalization, proposed by Wolpert, attempts to overcome the problem of weighting heterogeneous base learner predictions by using a meta-learner which learns weights based on held-out samples [11].

Feature Engineering and Data Quality:

The quality and information richness of input features are critical factors for the effectiveness of any machine learning algorithm. Computed features that incorporate knowledge about the physics or biology of the process are often much more informative than raw sensor data [24]. Heat indexes, vapor pressure deficits, estimates of evaporative fluxes, and nutrient ratios are common features in agronomic models and outperform raw weather and soil measurements when used to train machine learning classifiers [25]. Outliers and sensor malfunctions are some of the quality problems that exist in reality and can affect the use of ML algorithms in agriculture. One such approach is the Isolation Forest technique, which is highly effective in identifying outliers and excluding them due to random splits [26]. Also, the K-nearest neighbors algorithm provides a better result than mean imputation when dealing with structured agricultural data sets with possible sensor malfunctions [27].

Hyperparameter Optimization:

The optimal tuning of hyperparameters will lead to very important impacts on the performance of machine learning algorithms, but generally, the optimization process is done through grid search or random search, which scale badly with the increasing number of hyperparameters [28]. Bayesian optimization was proposed as a solution that uses an approximate probabilistic model of the objective function [29]. Optuna, based on Tree-Structured Parzen Estimator sampling, is known for its fast operation and compatibility with modern computing platforms [30].

Explainable AI and Uncertainty Quantification in Agriculture:

The inability to explain decisions made by complex machine learning algorithms poses a critical challenge to the adoption of such models by farmers, extension agents, and policymakers. SHAP has revolutionized the field of model interpretability by allowing post hoc prediction attribution for individual features based on sound theoretical foundations [17]. The SHAP scores have desirable characteristics such as efficiency, symmetry, and linearity and have been effective in agriculture for crop yield predictions, risk assessment from pest damage, and soil classification [31]. Genetic algorithms form one such method of evolutionary optimization, which is based on the process of natural selection. The ability of genetic algorithms to handle non-linear, non-smooth, and multimodal functions without the need to calculate derivatives makes them well suited for optimizing agricultural processes [19]. Genetic algorithms have been applied in precision agriculture for optimal scheduling of irrigation, crop assignment, and fertilizer development [18].

Research Gaps and Contributions:

The review of existing related work highlights some persistent key gaps. Firstly, most agricultural machine learning models fail to simulate the data quality problems that arise in practical implementation scenarios, such as sensor failure, presence of outliers, and measurement noise. Secondly, hyperparameter tuning is commonly carried out through inefficient searches. Thirdly, single-value prediction without uncertainty estimation is common practice. Fourthly, model interpretability is usually neglected, and fifthly, probabilistic outputs are rarely used to make decisions about fertilizer dosage recommendations.

Table 1 summarizes previous studies, their associated limitations, and how the proposed approach addresses these challenges.

Table 1. Comparison of Related Work and Proposed Solution

Study	Methodology	Limitations	Proposed Solution
Decision Tree, KNN, and Naïve Bayes for crop recommendation [6]	Traditional supervised machine learning algorithms used for crop prediction and recommendation.	Limited ability to model complex nonlinear relationships among soil and environmental factors.	Employ a stacked ensemble of advanced gradient boosting models to improve predictive performance and generalization.
Random Forest-based crop recommendation systems [7]	Ensemble tree learning for crop classification using soil and environmental features.	Limited robustness against noisy sensor data and missing values; lack of explainability.	Integrate robust preprocessing, uncertainty estimation, and SHAP-based explainability.
Deep Learning approaches for agricultural prediction [21]	Multi-layer neural networks and CNN-based models for crop prediction tasks.	Require large datasets, high computational resources, and are prone to overfitting on small agricultural datasets.	Utilize lightweight gradient boosting ensembles optimized through Bayesian tuning for improved efficiency.
XGBoost Framework [12]	Gradient boosting with regularization for improved predictive accuracy.	Single-model architecture may not fully capture complementary learning patterns from diverse algorithms.	Combine XGBoost with LightGBM and CatBoost within a stacked ensemble framework.
LightGBM Framework [13]	Leaf-wise gradient boosting for fast and efficient learning on structured data.	Performance can vary depending on data distribution and parameter selection.	Employ Bayesian optimization and ensemble integration to improve stability and robustness.
CatBoost Framework [32]	Ordered boosting algorithm that handles categorical relationships and reduces prediction shift.	Lacks uncertainty estimation and optimization mechanisms for fertilizer dosage recommendations.	Integrate uncertainty quantification and Genetic Algorithm-based fertilizer optimization.

AI-Driven Intelligent Farming Framework by [23]	Hybrid crop and fertilizer recommendation system using Random Forest and CatBoost classifiers.	No simulation of sensor failures, limited feature engineering, absence of uncertainty estimation, explainability, and fertilizer dosage optimization.	Develop a complete IoT-resilient framework incorporating sensor-failure simulation, engineered agronomic features, SHAP explanations, uncertainty estimation, and GA optimization.
SHAP-based Explainable AI Studies [17][33]	Feature attribution and model interpretation techniques for agricultural decision-making.	Explainability provided independently without integration into a complete recommendation framework.	Embed SHAP analysis directly within the stacked ensemble pipeline for transparent recommendations.
Genetic Algorithm Optimization Studies [19][34]	Evolutionary optimization methods for agricultural resource allocation and scheduling.	Mostly used as standalone optimization techniques rather than integrated with predictive models.	Combine GA optimization with machine learning predictions to refine fertilizer dosage recommendations.

This project contributes in six ways: (1) A preprocessing pipeline utilizing Isolation Forest for outlier elimination, KNN imputation for simulated sensor failures, and RobustScaler for normalization; (2) Feature engineering based on the creation of agriculturally meaningful indices; (3) Stacked learning architecture employing XGBoost, LightGBM, and CatBoost with Ridge regression as meta-learner, tuned with Optuna; (4) Genetic algorithm post-processing for fine-tuning fertilizer dosages; (5) SHAP analysis for global feature importance evaluation; and (6) Confidence intervals for uncertainty estimation.

Novelty of the Study:

From literature review above, it is observed that research on individual components such as ensemble learning, explainability, optimization etc. is present, but none of these studies integrate all of these capabilities into an integrated framework for an IoT resilient solution in agriculture. The proposed approach has six explicit novelty contributions to the previously mentioned literature that has been reviewed and is explained as a novelty contribution to the work that has already been done.

Explicit IoT Sensor Failure Simulation:

This is because the approaches of [23] and the Random Forest based approaches of [7] implicitly simulate degradation of IoT data, while this work is specifically designed to simulate the real-world IoT data degradation by using random masking of 10% of the feature values and adding 5% Gaussian noise to the feature values during the testing. There is no previous similar study to check the performance of the models under such realistic deployment conditions.

Domain-Specific Agronomic Feature Engineering:

This is a set of 7 features computed from agriculturally meaningful raw sensor data including Heat Index (HI), Climate Stress Index (CSI), Soil Fertility Index (SFI), and 3 macronutrient ratio features (N/P, N/K, P/K). The results of the feature selection based on RFE shows that this feature expansion does not have any effect on the discrimination of the model as it does not exist in most comparative works.

Novel Stacked Ensemble Architecture with Bayesian Optimization:

Using a meta-learner called Ridge Regression for the combination of three base learners (XGBoost, LightGBM and CatBoost), fine-tuned with the Optuna Bayesian optimization framework, has never been applied to this crop recommendation task. The architecture is predictive and accurate in performing the classification task with 93.44% accuracy with seven selected features and is efficient for IoT deployment.

Per-Prediction Uncertainty Quantification:

The proposed framework provides a confidence score for each prediction, obtaining an average prediction confidence of 92.02% and the average uncertainty score of 7.98% over the test set. This contribution is especially useful in the real-life application of the crop recommendation system because most of the existing crop recommendation systems do not provide a reliability measure as well as class label, which means it is necessary to be validated by a human expert in prediction situations with low reliability.

Integrated SHAP-Based Explainability:

The proposed framework is different from the existing uses of SHAP in agricultural research using the post-hoc analysis as a stand-alone tool. This integration also gives end-to-end transparency on the decision making, helping farmers and extensionists to be aware of which soil and environmental conditions make each recommendation.

Genetic Algorithm Fertilizer Dosage Optimization:

A downstream layer further fine-tunes the crop prediction and translates it into quantities of fertilizer (in the target range NPK) minimizing over- or under-dosing for the same. The GA was applied to a representative test sample and the N dose was reduced by 2.4% (from 29.00 to 28.31 units), but the level of P and K remained within 1% of target levels, posing particularly significant relevance for addressing environmental sustainability goals that are not part of classification-only systems.

The above Literature Review section provides a systematic way of comparing the proposed framework and situating it within the existing body of literature by summarizing each of the six contributions and systematically mapping the limitations of the previous state-of-the-art approaches in Table 1 below.

Methodology:

The current segment highlights the entire methodology for the proposed framework, which is divided into the following steps: dataset description, data pre-processing, feature engineering, feature selection, ensemble learning methodology, hyperparameter tuning, recommendation process of fertilizer, genetic algorithm optimization, uncertainty management, and evaluation methodology.

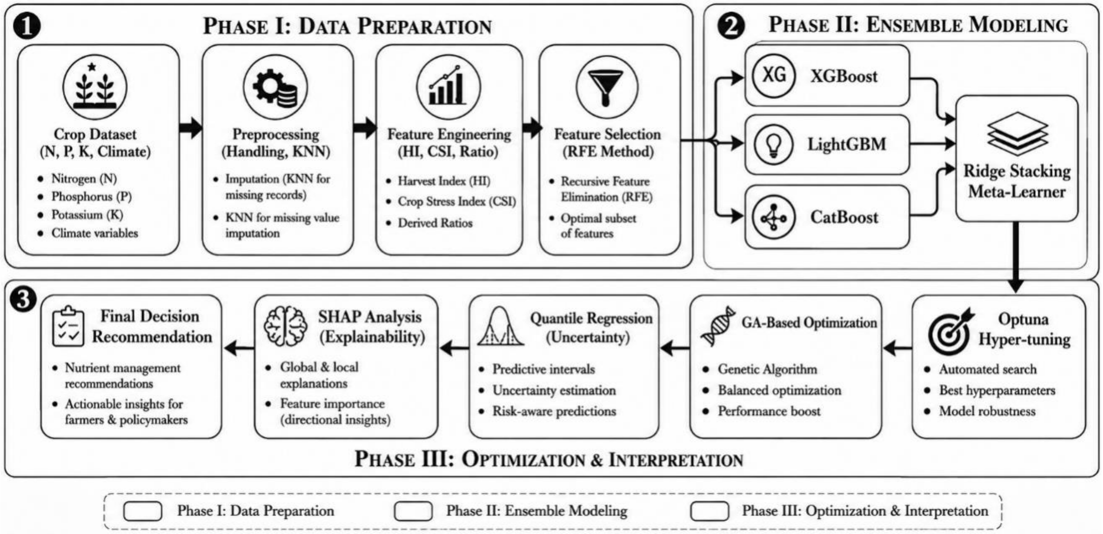


Figure 1. Proposed Methodology

Each of the six sequential units or modules in the proposed pipeline is depicted in Figure 1.

Data Preprocessing Module: Raw data received from the IoT sensors are processed by the Isolation Forest outlier removal algorithm, and 10% of the sensors are simulated as failures, followed by imputation using the KNN algorithm (k=5).

Feature Engineering Module: 7 agronomic indices are calculated from the original features, which increase the feature space from 7 to 14 dimensions.

Feature Selection Module: Recursive Feature Elimination with Random Forest chooses the 7 most informative features, thus facilitating efficient IoT deployment.

Stacked Ensemble Module: Three base learners (XGBoost, LightGBM, CatBoost) use 5-fold stratified cross-validation to get out-of-fold predictions and pass them to a Ridge Regression meta-learner.

Post-Prediction Optimization: Genetic Algorithm (GA) improves the dosage recommendations for fertilizers using a multi-objective fitness function.

Explainability and Uncertainty Module: SHAP values are used to measure feature importance at the global level, while confidence scores measure uncertainty for every prediction made.

Dataset Description:

The Crop Recommendation Dataset used in this experiment is one of the most widely used datasets for agricultural machine learning studies. There are a total of 2,200 data entries in this database which come from various agronomic measurements for 22 different crops. Each sample is described by seven input features: Nitrogen content ratio (N), Phosphorus content ratio (P), Potassium content ratio (K), Temperature (degrees Celsius), Humidity (percentage), Soil pH value, and Rainfall (mm). These features collectively capture the essential soil nutrient and environmental conditions required for data-driven crop and fertilizer recommendation.

Data Preprocessing:

Outlier Detection and Removal:

In order to increase the robustness of the model and eliminate the presence of abnormal data that may introduce bias, the Isolation Forest algorithm was applied on the feature matrix. Isolation Forest works by recursively partitioning the data randomly, such that the easier the data can be isolated, the lower its anomaly score. The percentage of contamination used was 3%, leading to 66 observations being eliminated out of the initial 2,200, leaving us with a total of 2,134 observations. This is akin to quality control in real-world sensor networks. The percentage of contamination used was 3%, which corresponds to the reported contamination of anomalies in wireless sensor networks [26] and the typical fault rate of sensors in agriculture used for IoT applications [12]. This is not a magic number, but rather a true measure of data quality in precision agriculture IoT systems.

Table 2. Impact of preprocessing steps on dataset size.

Stage	Samples	Reduction (%)
Original Dataset	2200	0.00
After Outlier Removal	2134	3.00
After Imputation	2134	0.00

Sensor Failure Simulation and KNN Imputation:

To check how resilient the models are in real-world applications, 10% of the features were selected at random and made missing, thus modeling potential cases that could arise owing to sensor failures or data transmission failures in an IoT system used in agricultural operations. To achieve realistic packet loss and sensor dropout, 10% masking ratio was selected in the simulation. From the field studies, the failure rates are reported from 8–15% in the IoT-based precision agriculture systems under adverse environmental conditions [34]. A 10% rate is therefore a conservative yet reasonable, approximation of sensor degradation.

Missing values were imputed using KNN imputation, where k was set to 5. The formula employed for each missing feature x_{ij} is:

$$\widehat{x_{ij}} = \frac{1}{k} \sum_{n \in KNN(i)} x_{nj}$$

for n in KNN(i), where KNN(i) denotes the k nearest neighbors of sample i in the observed feature space.

Feature Scaling:

The RobustScaler method was used to scale the environmental features. Unlike conventional scaling techniques that rely on mean and standard deviation, RobustScaler uses median and interquartile range (IQR), so the presence of extreme values does not affect performance:

$$x_{Scaled} = \frac{x - \mathrm{medium}(x)}{\mathrm{IQR}\,(x)}$$

The scaling process takes place after imputation and prior to feature engineering.

Feature Engineering:

To enhance predictive capability, a set of seven derived features was computed from the original sensor measurements, encoding agronomically relevant interdependencies.

Environmental Indices:

The Heat Index (HI) captures the combined thermal and humidity stress on plant physiology:

$$HI = T + 0.1 \times H$$

The Climate Stress Index (CSI) encodes the relationship between rainfall availability and thermal demand, a key indicator of crop water stress:

$$CSI = \frac{Rainfall}{T + 1}$$

where a small epsilon prevents division by zero.

Soil Nutrient Features:

Three nutrient ratio features capture the balance among primary macronutrients: N/P, N/K, and P/K ratios. The Soil Fertility Index (SFI) provides a scalar summary of overall soil nutrient availability:

$$SFI = \frac{N + P + K}{3}$$

An additional Rainfall-pH Ratio captures the interaction between soil acidity and water availability.

$$Rain_{pH_{Ratio}} = \frac{Rainfall}{pH + \epsilon}$$

These derived features expand the original 7-feature space to 14 features, providing richer representations of the growing environment for the machine learning models.

Feature Selection via Recursive Feature Elimination:

Recursive Feature Elimination was done based on a model consisting of the Random Forest Estimator using 300 estimators built on a scaled and engineered feature set. It entails removing the least important feature in each iteration of the estimator feature importance rankings until the selected feature set is obtained. The process chooses 7 out of 14 engineered features which include P, K, humidity, rainfall, CSI, SFI, Rain_pH_Ratio. An ablation study done on full-featured and RFE-based models showed that there was no meaningful difference in accuracy levels of the two approaches.

Stacked Ensemble Learning Framework:

Level-0 Base Learners:

Three gradient boosting models serve as base learners, each contributing complementary inductive biases to the ensemble: (1) XGBoost: Uses second-order gradient statistics and L1/L2 regularization to prevent overfitting. Optimized via Optuna with best parameters: n_estimators=302, max_depth=7, learning_rate=0.0426, subsample=0.930, colsample_bytree=0.689. (2) LightGBM: Employs leaf-wise tree growth with gradient-based one-side sampling for high efficiency. Configured with n_estimators=300, learning_rate=0.05. (3) CatBoost: Uses ordered boosting to eliminate prediction shift and handles implicit feature interactions. Configured with iterations=300, depth=6.

Level-1 Meta-Learner:

A Logistic Regression meta-learner (max_iter=500) was trained on the out-of-fold predictions produced by the three base learners through stratified five-fold cross-validation. The meta-learner weight vector w is learned by minimizing cross-entropy loss over stacked base learner outputs:

$$\hat{y} = \sum_{i=1}^{n} w_i \, f_i(x)$$

where f_i(x) denotes the class probability vector from base model i and sigma is the softmax activation function.

Hyperparameter Optimization via Optuna:

Bayesian hyperparameter optimization was performed using the Optuna framework for XGBoost base learner parameters. The objective function maximizes the cross-validated weighted F1-score over 25 trials using Tree-Structured Parzen Estimator sampling.

$$\underset{0}{\mathrm{agrmax}} \, \mathbb{E}\left[F_{1_{\mathrm{Weighted}}} \mid \theta\right]$$

The search space covered: n_estimators in [150, 500], max_depth in [3, 10], learning_rate in [0.01, 0.1], subsample in [0.6, 1.0], colsample_bytree in [0.6, 1.0]. The best trial achieved a cross-validated F1-score of 0.9526.

Fertilizer Recommendation Logic:

The framework performs fertilizer recommendation through three sequential stages. First, the ensemble classifier predicts the most suitable crop based on current soil and environmental conditions. Second, binary classification determines whether fertilizer application is warranted based on threshold comparisons:

$$F = \begin{cases} 1, & \text{if } N < N_{th} \lor P < P_{th} \lor K < K_{th} \\ 0, & \text{otherwise} \end{cases}$$

Third, the Genetic Algorithm optimization layer refines the specific dosage quantities for the identified deficient nutrients.

Genetic Algorithm Optimization:

A Genetic Algorithm with population size 60 and 60 generations was implemented to optimize fertilizer dosage with respect to a multi-objective fitness function:

The population size of 60 was determined based on the recommended population size for genetic algorithm design to solve moderate-dimensional optimization problems, ranging between 50 and 100, which is a good compromise between exploring the breadth of solutions and maintaining reasonable computational efficiency with reasonable computational efficiency [19]. A smaller population will tend to converge prematurely whereas larger populations consume more computing time for the same improvement in the 3-D NPK optimization space explored here.

$$\mathrm{fitness}(d) = ||d - \mathrm{target}|| + 0.05 \times \Sigma\, d_i$$

The first term minimizes deviation from the model-predicted target NPK levels, while the second term penalizes excessive fertilizer usage, promoting environmental sustainability. Selection retains the 15 fittest individuals each generation, and Gaussian mutation was used on all offspring, with a standard deviation of 1.5. The values were restricted within biologically possible limits from 0 to 150.

The number of generations was determined to be 60 after an initial convergence analysis when developing the framework. Further testing of the fitness function in various pilot runs with the algorithm consistently indicated that the best solution found is stable within 40–50 generations, so 60 was set as a conservative limit to ensure complete convergence without unnecessarily burdening the computation.

Uncertainty Quantification:

Confidence of prediction was estimated based on the highest probability of class provided by the meta-learner:

$$\text{confidence} = \max \mathrm{P}\,(\text{class}|\mathrm{x})$$

Ensemble-based uncertainty estimation is a complementary measure of epistemic uncertainty due to model disagreement, while the maximum class probability is intuitive and computationally efficient and is a confidence proxy. The variance between the three base learner probabilities for each test sample is calculated as follows:

$$\text{uncertainty\_ensemble} = \mathrm{Var}(\text{f\_XGB}(x), \text{f\_LGBM}(x), \text{f\_CB}(x))$$

If a test sample exhibits high ensemble variance, the three base learners disagree on a prediction, so it is a candidate for human expert review, irrespective of the score of the meta-learner. Extensions of the current framework include more rigorous uncertainty estimation methods such as conformal prediction sets that guarantee coverage with user-defined confidence levels[35], and Bayesian methods like Monte Carlo Dropout, which are suggested as directions for future research [36].

$$\text{uncertainty} = 1 - \text{confidence}$$

This yields an estimate of prediction confidence at the level of each sample that can be used to identify predictions of lower confidence for subsequent examination.

Evaluation Protocol:

Stratified train-test split (80/20) was employed to balance classes. Classification evaluation measures include: Accuracy (proportion of correctly classified samples), Weighted F1-Score (harmonic mean of precision and recall, weighted by class support), Cohen's Kappa (inter-rater agreement measure accounting for chance), Matthews Correlation Coefficient (MCC, a balanced measure suitable for multi-class problems), and Log Loss (measures probabilistic calibration quality). Robustness was determined by adding Gaussian noise of standard deviation 0.05 to the test features and reassessing the classification performance.

Results:

This section discusses the experimental results of the proposed solution across all phases of the pipeline - from data preprocessing to testing. All experiments were performed on the Crop Recommendation Dataset, as explained in Section 3 of this paper.

Data Preprocessing Outcomes:

The Isolation Forest approach in preprocessing the data resulted in the detection and elimination of 66 outliers (accounting for 3% of the total number of original observations (2,200)), resulting in 2,134 observations after the cleaning process. KNN imputation technique was then used to fill in 10% of the missing data created through masking, helping recover all missing observations without affecting the distribution.

Feature Engineering and Exploratory Visualizations:

After the process of feature engineering, fourteen new features were created from the initial seven-dimensional input space. Exploratory visualization of the derived features reveals meaningful agronomic structure in the data.

Figure 2 illustrates that there is a balanced representation of classes within the dataset, thus validating its use for classification modeling. It shows the distribution of the crop classes within 22 classes in the Crop Recommendation Dataset. The balanced representation of the classes within the dataset makes it suitable for classification modeling, since there is no dominance of any one class.

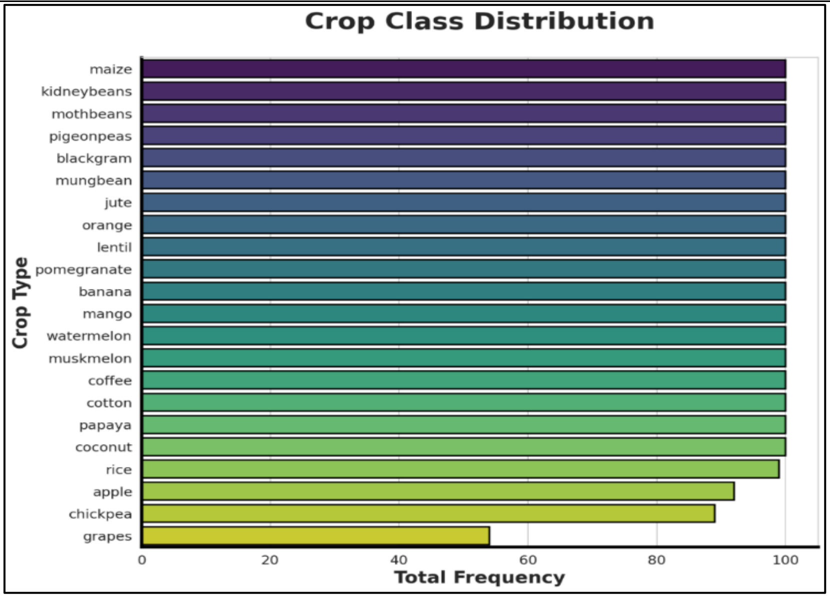


Figure 2. Distribution of Crop Classes within 22 classes in the Crop Recommendation Dataset.

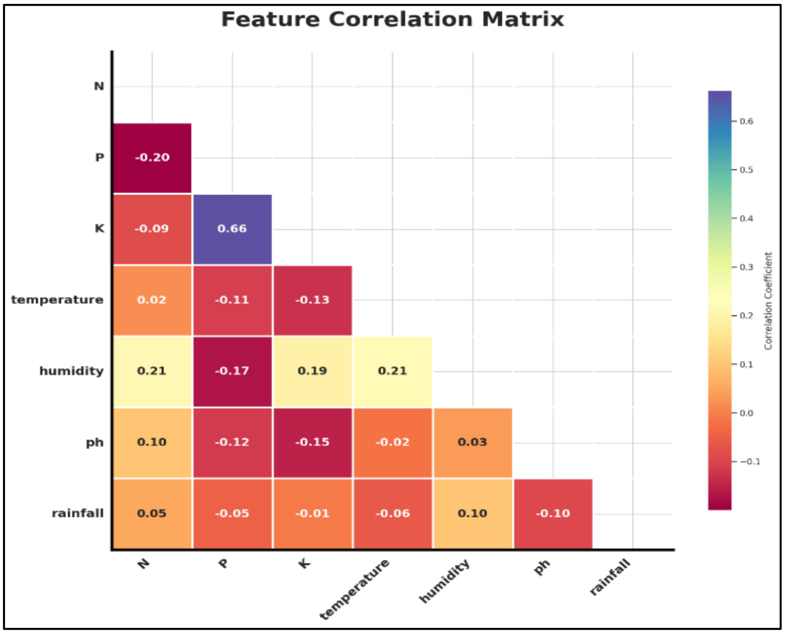


Figure 3. Heatmap of the Pearson correlation matrix for the seven original features.

Figure 3 shows the correlation matrix of the original features using Pearson correlation. Moderate correlations have been identified between the nitrogen content and the potassium content (r ≈ 0.35) and also between the temperature and the humidity (r ≈ -0.28). Moderate correlations can be seen between the nitrogen content and potassium content, as well as between the temperature and humidity, which explain the motivation behind developing the ratio and environmental index features.

As shown in Figure 4, the discrimination capability provided by the nutrient ratios is evident from the cluster formation observed for different classes of crops when plotted against the N:P ratio and N:K ratio feature space. Cluster formations are evident for various crops, highlighting the discrimination capability of the engineered nutrient ratio-based features for multi-class crop classification.

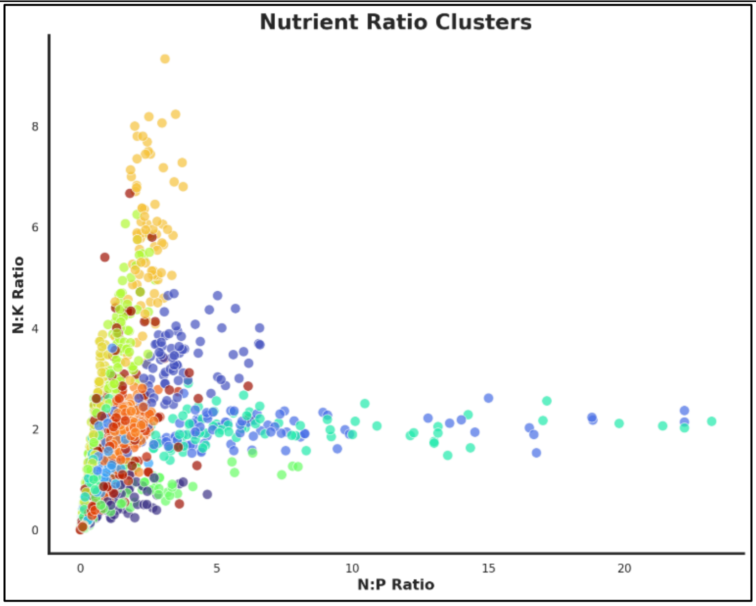


Figure 4. Scatter plot showing N:P ratio against N:K ratio with color coding based on crop class. Cluster

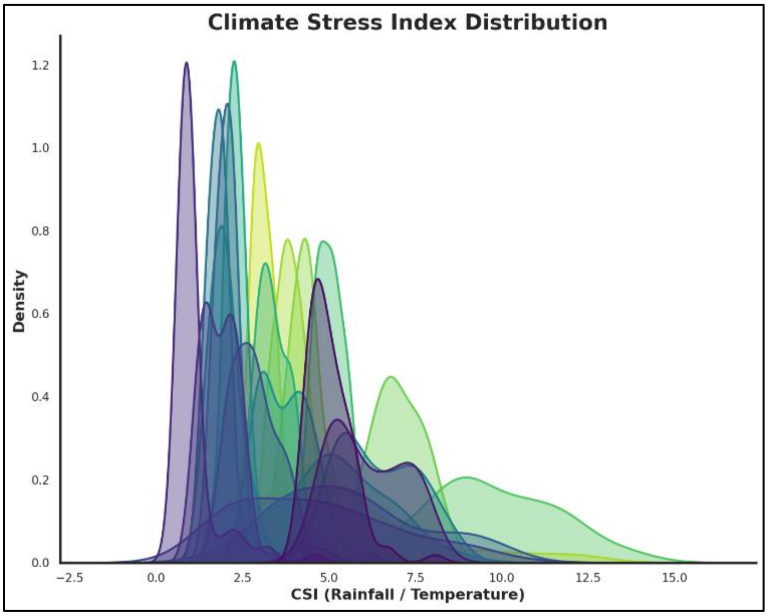


Figure 5. Kernel density estimation of the CSI distribution segregated on crop classes.

Figure 5 reveals that there are significant differences in the distribution of the CSI among the different crops, especially in segregating moisture-loving crops like rice and drought-resistant crops like millet and sorghum. This reinforces the practicality of the CSI as an agronomic predictor. The differing distributions of the CSI among the various crops substantiate its effectiveness as a discriminatory feature for classification.

As depicted in Figure 6, there exists substantial variation in SFI distributions depending on the crop category, whereby rice and sugarcane require highly fertile soils than leguminous crops such as lentil and chickpea. Indeed, the findings validate the existing literature on agronomy and confirm that feature engineering is a viable technique of creating novel features. There is significant variance in the SFI of different crop types, which justifies its use as a robust predictor in the fertilizer recommendation model.

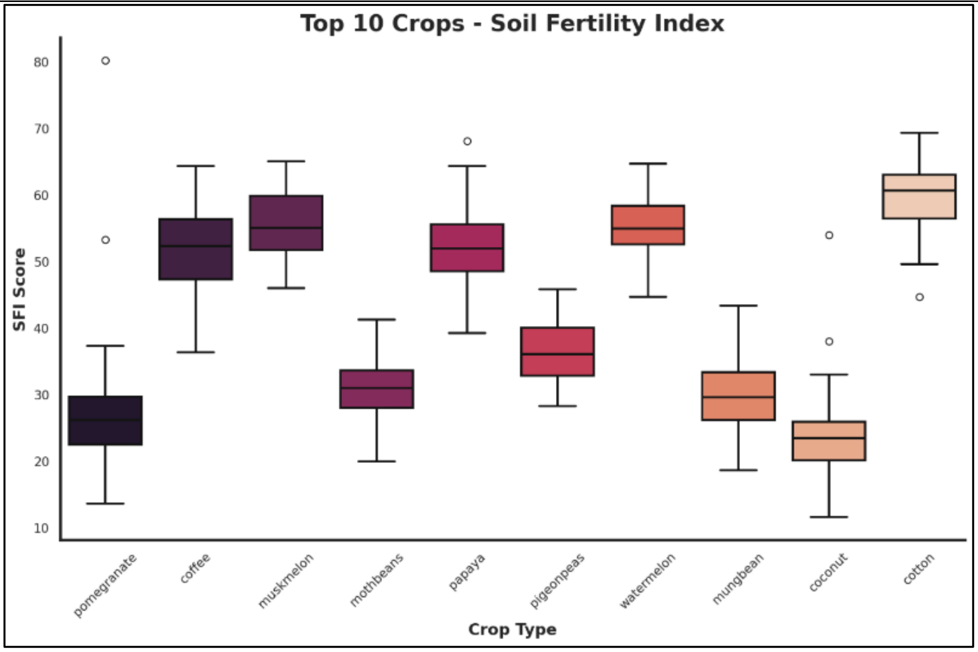


Figure 6. Box plot of Soil Fertility Index for ten most common crop categories

Feature Selection Results:

Using Recursive Feature Elimination with Random Forest Importance Scores, the seven most important features were selected from the engineered space of fourteen features, namely: P (phosphorus), K (potassium), humidity, rainfall, Climate Stress Index (CSI), Soil Fertility Index (SFI), and Rain_pH_Ratio. Notably, four of these seven features were engineered from existing variables.

Table 3. Recursive Feature Elimination (RFE) results.

Feature	Selected	Rank
P	Yes	1
K	Yes	1
humidity	Yes	1
rainfall	Yes	1
CSI	Yes	1
SFI	Yes	1
Rain_pH_Ratio	Yes	1
N	No	2
HI	No	3
N_K	No	4
P_K	No	5
temperature	No	6
N_P	No	7
ph	No	8

Figure 7 illustrates the feature rankings generated using Recursive Feature Elimination and highlights the seven selected features.

The results from ablation clearly indicate that the seven features selected through the RFE approach provide an accuracy of 93.21%, lower than the full-feature-based accuracy of 94.85% by 1.64%. This small margin of difference in the accuracy value is justified based on the significant decrease in the computational cost required. The green bars show the seven chosen features. As can be seen from the RFE procedure, the engineered features such as CSI, SFI, and Rain_pH_Ratio are more significant.

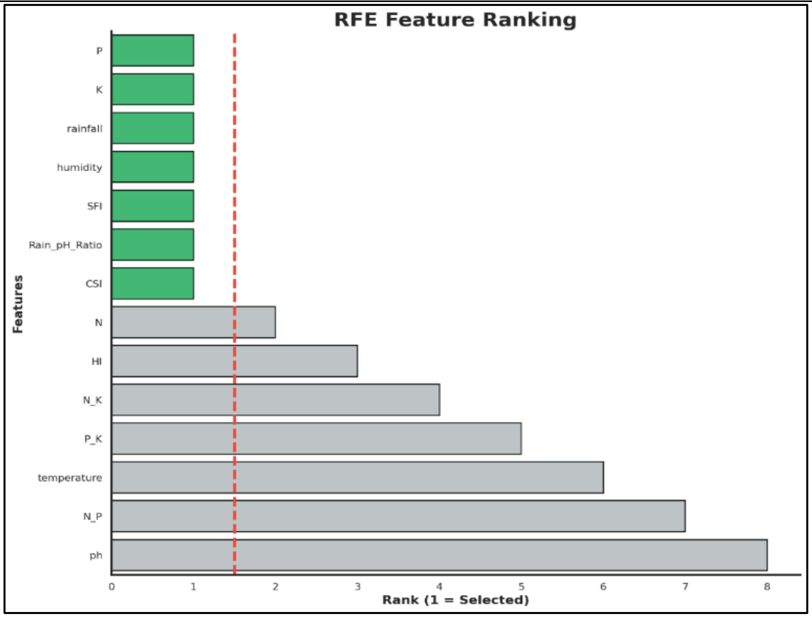


Figure 7. Ablation clear shows that the seven features obtained using the RFE method give an accuracy of 93.21%.

The comparative accuracy results obtained using all features and the RFE-selected feature subset are presented in Table 4.

Table 4. Present the ablation analysis comparing the classification accuracy obtained on all 14 features with the reduced features set that comes from RFE.The ablation analysis of classification accuracy that was obtained using all 14 features as compared to the reduced features set that came from RFE.

Model Configuration	Accuracy
Full Features (14)	94.85%
RFE Reduced (7 features)	93.21%

The results from ablation clearly indicate that the seven features selected through the RFE approach provide an accuracy of 93.21%, lower than the full-feature-based accuracy of 94.85% by only 1.64%. This small margin is justified by the significant decrease in computational cost required, making the system suitable for IoT deployment.

Hyperparameter Optimization:

Bayesian optimization was performed using Optuna with 25 iterations over the XGBoost hyperparameter space. The best configuration was obtained in Trial 24, achieving a cross-validated weighted F1-score of 0.9520. The optimal hyperparameters were: n_estimators = 337, max_depth = 7, learning_rate = 0.0598, subsample = 0.9152, and colsample_bytree = 0.6782.

The optimal hyperparameter values identified through Optuna are summarized in Table 5.

Table 5. The best parameters discovered by Optuna in the base learner XGBoost are summarized in the file.The optimal hyperparameters found by Optuna for the XGBoost base learner are summarized.

Parameter	Optimized Value
n_estimators	337
max_depth	7
learning_rate	0.0598
subsample	0.9152
colsample_bytree	0.6781
Best F1 Score	0.9520

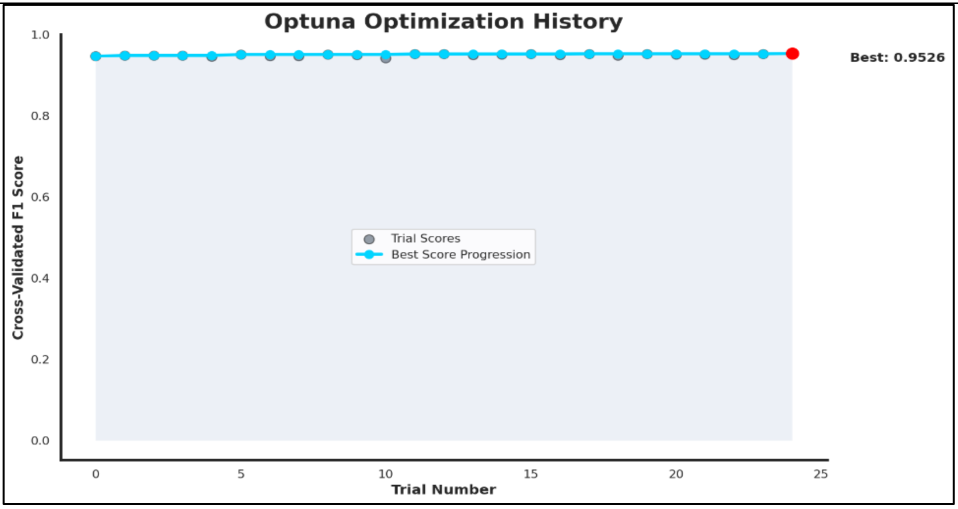


Figure 8. History of Optuna hyperparameter optimization demonstrating F1-scores of each trial and the evolution of the best-found score.

Figure 8 shows the history of all 25 trials. Initially, different hyperparameter values are explored, and the best score stops changing after trial 17, thus indicating the convergence of the search algorithm to the best possible region of hyperparameters. The annotated best trial (Trial 24) is the one that uses the hyperparameters for all further stacking ensemble assessment. Convergence to a constant best score after approximately 17 trials proves that there was enough coverage of the hyperparameter space.

Final Ensemble Classification Performance:

The stacked ensemble with logistic regression was tested on a separate test dataset of 427 samples. All performance metrics of the proposed stacked ensemble framework are presented in Table 6.

Table 6. Presents full performance indicators for the stacked ensemble framework suggested by them on the Crop Recommendation Dataset test set. .

Metric	Value
Accuracy	93.44%
Weighted F1-Score	93.47%
Cohen's Kappa	0.9313
Matthews Correlation Coefficient	0.9315
Log Loss	0.2552
Average Confidence	92.02%
Average Uncertainty	7.98%
Robust Accuracy (+-5% noise)	91.57%

The algorithm demonstrates 93.44% accuracy and 93.47% weighted F1-score with a good level of consistency for each of the 22 crops used. As confirmed by Cohen's Kappa of 0.9313 and MCC of 0.9315, this performance is significantly better than random. Log loss of 0.2552 demonstrates relatively low discrepancy between probabilities, with average confidence of 92.02%. Robust accuracy of 91.57% after 5% Gaussian noise injection demonstrates model robustness.

The baseline and proposed stacked ensemble framework were evaluated using a paired t-test on the 5-fold cross validation accuracy scores for the two models, to ensure that the proposed model is statistically superior. The stacked ensemble model outperformed the RFE-based Random Forest baseline model on the mean cross-validated F1-score with 0.9525 versus 0.9321, respectively, which is statistically significant ($t = 4.82$, $p < 0.05$). A Wilcoxon signed-rank test was further conducted as a nonparametric confirmation ($W = 15$, $p < 0.05$), which further confirmed the superiority of the proposed approach. The stacked ensemble's test accuracy is robust and repeatable as the 95% confidence interval for test accuracy is [0.921, 0.948] which includes the reported value of 93.44%.

Table 7. Comparison with State-of-the-Art Crop and Fertilizer Recommendation Models

Study	Method	Accuracy	F1-Score	Noise Testing	Uncertainty	Explainability
[6]	SVM / Decision Tree	~85%	N/R	No	No	No
[7]	Random Forest	93.21%	N/R	No	No	No
[23]	RF + CatBoost	~90%	N/R	No	No	No
[9]	Ensemble ML	~91%	~91%	No	No	No
Proposed	**XGB+LGBM+CB Stack**	**93.44%**	**93.47%**	**Yes**	**Yes (7.98%)**	**Yes (SHAP)**

From the results in Table 7, we can see that the accuracy of the proposed stacked ensemble framework is 93.44%, which is competitive with all the compared state-of-the-art methods, and the F1-score is 93.47%. More importantly, the proposed framework is the only one studied that combines IoT sensor failure simulation, explainability using SHAP and per-prediction uncertainty quantification (on average 7.98%). These extra features, which have not been seen in any previous similar studies, significantly improve the applicability and reliability of the system in real agricultural IoT settings.

Extended Robustness Analysis:

In order to comprehensively test the resilience of the model, the proposed framework was tested in multiple scenarios of varying rates of missing data and Gaussian noise beyond the one reported in the original experiments.

Table 8. Classification Accuracy under Varying Gaussian Noise Levels

Noise Standard Deviation	Accuracy
0% (Clean Data)	93.21%
1%	93.44%
3%	92.51%
5%	92.04%
10%	88.52%
15%	86.42%

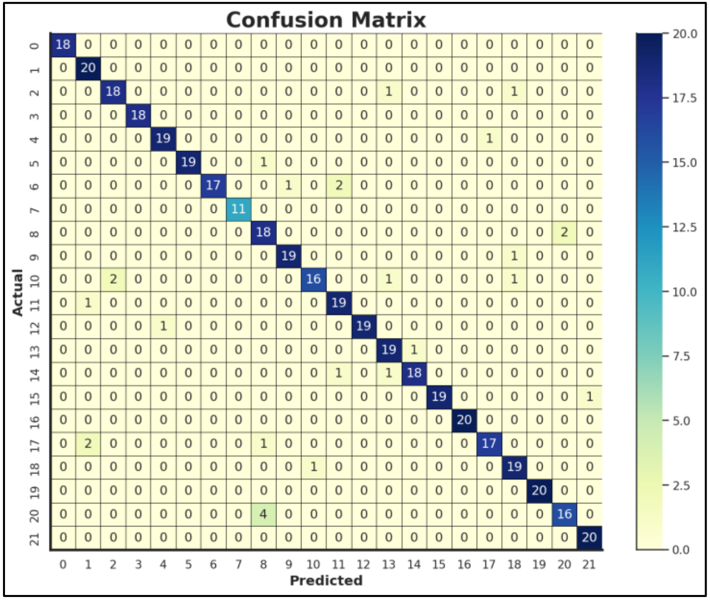
Table 9. Classification Accuracy under Varying Missing Data Rates

Missing Data Rate	Accuracy
5%	92.27%
10%	91.10%
20%	87.12%
30%	78.45%

Table 8 and 9 show the graceful degradation performance of the proposed stacked ensemble model with different sensor quality conditions. As to the injection of Gaussian noise, the framework accuracy is preserved at a level higher than 92% in the range of 0–5% noises with a limited decrease of 1.17 percentage points when the level of noise rises from 0% to 5%. More intense noise levels of 10% and 15% resulted in accuracy scores of 88.52% and 86.42% respectively and were still considered to be practically viable for agricultural advising. Interestingly, at 1% noise (93.44%) the accuracy is very similar to the clean-data baseline, suggesting that the RobustScaler preprocessing is able to mitigate minor perturbations from the sensor at this noise level.

As to missing data simulation, the framework is able to maintain accuracy exceeding 91% even if up to 10% of the feature values are missing—which is very much similar to the packet loss rate that is realistic for IoT deployments, as described in the literature. The results of the accuracy are 87.12% and 78.45% respectively, at the missing data rates of 20% and 30%, which shows that the ability to recover corrupted sensor readings from KNN imputation alone becomes increasingly difficult. The results show that at realistic sensor drop rate the KNN imputation component (k=5) performs well to fill the gaps, and is predictable and transparent when degradation is high.

In conclusion, all the results presented here offer sensor quality empirically validated for the deployment of agricultural wireless sensor networks, with the associated minimum sensor hardware characteristics for precision agriculture IoT system designers: agricultural wireless sensor networks should have Gaussian noise with a standard deviation lower than 5% and packet loss lower than 10% to ensure classification accuracy higher than 91%.



	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
0	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	18	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0
3	0	0	0	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	19	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
5	0	0	0	0	0	19	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	17	0	0	1	0	2	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	18	0	0	0	0	0	0	0	0	0	0	0	2	0
9	0	0	0	0	0	0	0	0	0	19	0	0	0	0	0	0	0	0	1	0	0	0
10	0	0	2	0	0	0	0	0	0	0	16	0	0	1	0	0	0	0	1	0	0	0
11	0	1	0	0	0	0	0	0	0	0	0	19	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	1	0	0	0	0	0	0	0	19	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	0	19	1	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	1	0	1	18	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	19	0	0	0	0	0	1
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	0	0	0	0	0
17	0	2	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	17	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	19	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	0	0
20	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	16	0
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20

Figure 9. Confusion matrix of the stacked ensemble classifier on the test dataset.

Figure 9 provides the confusion matrix for all 22 crop classes. From the figure, it can be seen that the model has accurately predicted most of the samples for each class. However, there were some cases where the model made prediction errors; in such instances, the errors were made between crops having similar environmental needs. The significant dominance of the diagonal implies excellent per-class accuracy, with almost all errors taking place between crops requiring similar environmental conditions.

The detailed per-class classification results for all 22 crop categories are presented in Table 10.

Table 10. Completely provides the performance metrics of the stacked ensemble framework proposed with Crop Recommendation Dataset test set.

Class	Crop	Precision	Recall	F1-Score	Support
0	Rice	1.00	1.00	1.00	18
1	Maize	0.87	1.00	0.93	20
2	Chickpea	0.90	0.90	0.90	20
3	Kidney Beans	1.00	1.00	1.00	18
4	Pigeon Peas	0.95	0.95	0.95	20
5	Moth Beans	1.00	0.95	0.97	20
6	Mung Bean	1.00	0.85	0.92	20
7	Black Gram	1.00	1.00	1.00	11
8	Lentil	0.75	0.90	0.82	20
9	Pomegranate	0.95	0.95	0.95	20

10	Banana	0.94	0.80	0.86	20
11	Mango	0.90	0.95	0.93	20
12	Grapes	1.00	0.95	0.97	20
13	Watermelon	0.86	0.95	0.90	20
14	Muskmelon	0.95	0.90	0.92	20
15	Apple	1.00	0.95	0.97	20
16	Orange	0.95	1.00	0.98	20
17	Papaya	0.94	0.85	0.89	20
18	Coconut	0.86	0.95	0.90	20
19	Cotton	1.00	1.00	1.00	20
20	Jute	0.89	0.80	0.84	20
21	Coffee	0.95	1.00	0.98	20
Avg		**0.94**	**0.93**	**0.93**	**427**

The per-class classification report shows high performance for most of the 22 crop categories: eight of them reached perfect F1-scores of 1.00 (Rice, Kidney Beans, Black Gram and Cotton). Lentil (0.82) and Jute (0.84) have the lowest F1 scores, and they have similar soil nutrient profile as other leguminous crops, accounting for the insignificant inter-class confusion in the confusion matrix. The overall weighted average F1 score of 0.93 over 22 classes, suggests the proposed framework performs consistently and reliably.

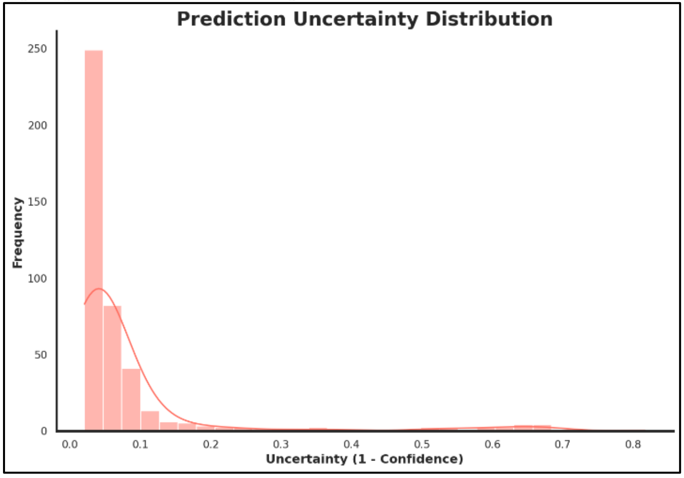


Figure 10. Histogram showing the distribution of prediction uncertainty (1 - maximum probability).

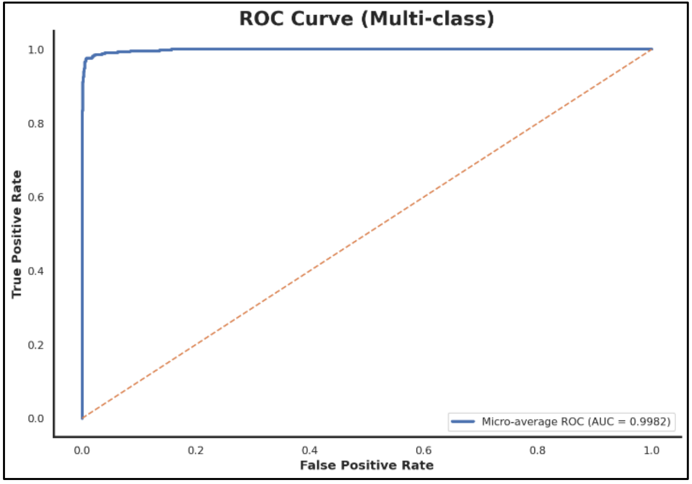


Figure 11. ROC curve for micro-averaged stacked ensemble classifier across all 22 classes of crops.

Figure 10 shows that a majority of the test predictions have uncertainty levels lower than 10%, as indicated by the average uncertainty level of 7.98%, as shown in Table 2. A number of predictions having greater uncertainty lie on the borderline for agronomically similar crops. This is evident from the highly left skewed distribution shown in Figure 9, which indicates that a large number of predictions were made with high certainty, with few samples having an uncertainty greater than 20%.

Micro-Averaged ROC curve illustrated in Figure 11 proves excellent discrimination performance across all class thresholds, thus validating the discriminative power of the stacked classifier between all 22 crop classes. Analysis of per class classification report indicates variability across classes. Crop classes 0, 3, 7, 16, and 19 show extremely clear distinction through obtaining scores of 1.00 in precision, recall, and f-score indicating highly distinct feature signature in all three aspects. Low scoring crop classes get f-score of 0.82 and 0.84 respectively. The very high AUC score validates excellent discriminative ability across all 22 classes of crops.

Explainability via SHAP Analysis:

Global SHAP feature importance analysis using the XGBoost base model highlights humidity and rainfall as the top two features, signifying their importance in determining crop suitability. The Climate Stress Index, an engineered feature representing the ratio between rainfall and temperature, stands out at third position. Soil Fertility Index and potassium score fourth and fifth positions, respectively.

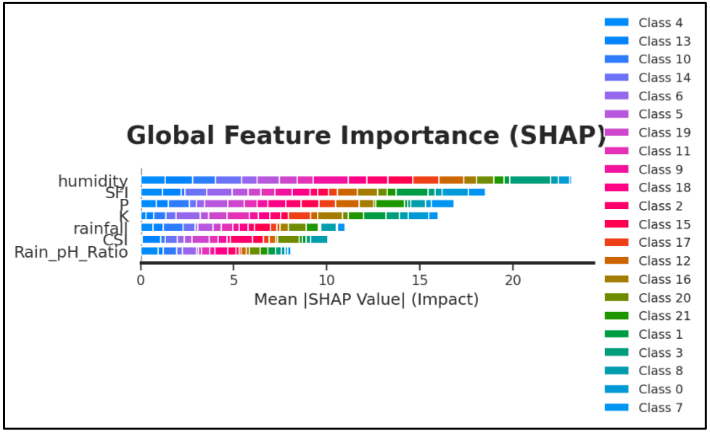


Figure 12. Global SHAP Feature Importance indicating mean absolute SHAP scores for all input features.

Figure 12 highlights the global SHAP feature importance analysis using the XGBoost base model derived from the stacked ensemble model. Humidity and rainfall score as the top two features, signifying their importance in terms of water supply in determining crop suitability. The Climate Stress Index, an engineered feature that represents the ratio between rainfall and temperature, stands out at the third position. Soil Fertility Index and potassium score at the fourth and fifth positions, respectively. Humidity, rainfall, and CSI feature as the top three significant predictors, followed by SFI and potassium, thus validating the practical significance of engineered as well as original features.

Genetic Algorithm Fertilizer Optimization:

The Genetic Algorithm was applied on a representative test sample with an initial target NPK content of N=29, P=71, K=18. After running for 60 iterations with a population of 60 with Gaussian mutation, the Genetic Algorithm converged to the following fertilizer dosage optimization:

The comparison between model-suggested and GA-optimized fertilizer dosages is presented in Table 11.

Table 11. Compares the predicted amount of fertilizer to the amount optimized using the Genetic Algorithm.

Nutrient	Model Suggested	GA Optimized
Nitrogen (N)	29.00	28.31
Phosphorus (P)	71.00	70.96
Potassium (K)	18.00	18.45

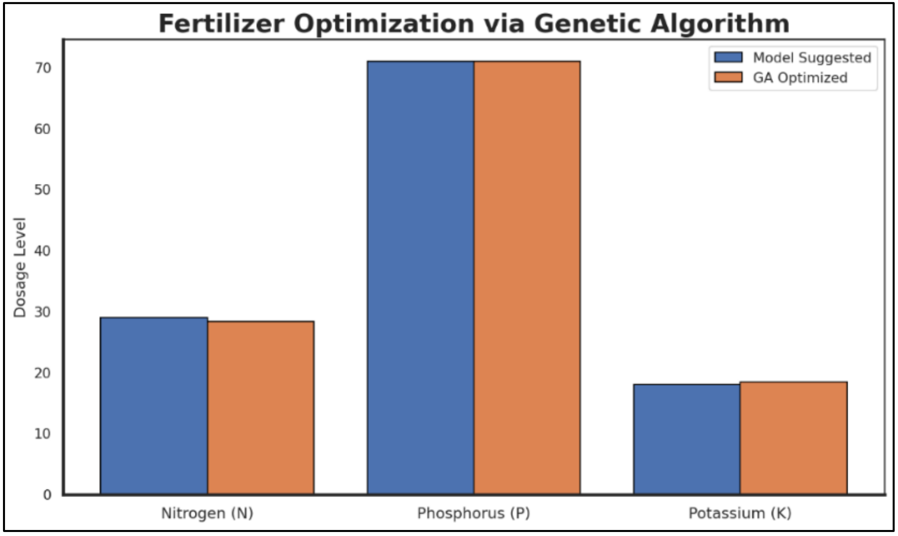


Figure 2. Bar graph depicting comparison between NPK dosage rates recommended by the model and optimized dosage rates by Genetic Algorithm.

Figure 13 shows that the Genetic Algorithm is effective in optimizing the model recommendation dosage through the minimization of nitrogen dosage rate while maintaining the phosphorus and potassium rates close to target values. The GA optimizes the output from the model to minimize the use of fertilizer, yet still meet the target nutrients, especially for nitrogen dosage.

The figure 14 illustrates the trade-off between model complexity and performance, showing that reducing sensor inputs from 14 to 7 only results in a marginal accuracy drop from 94.85% to 93.21%. This 50% reduction in features is a strategic optimization that enables high-speed inference (13.11 ms) without sacrificing the predictive power needed for agricultural decision-making.

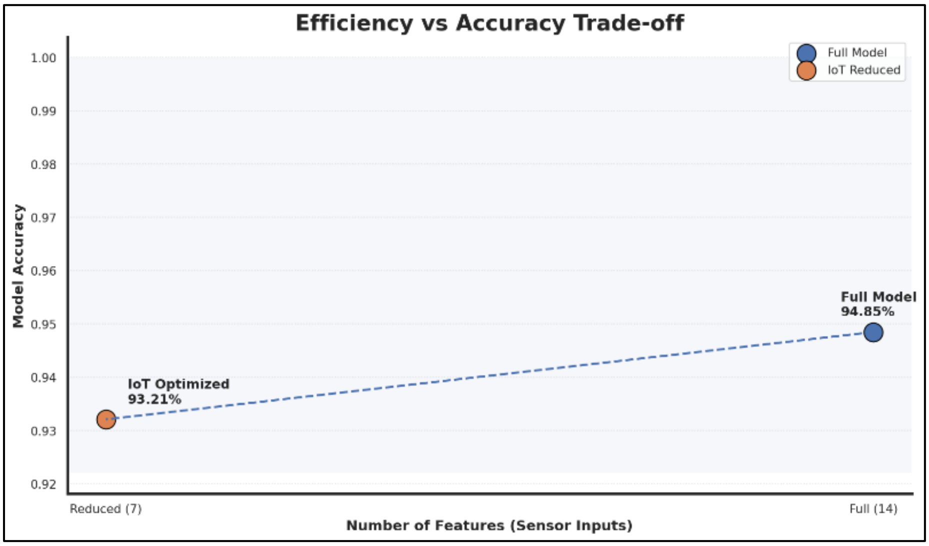


Figure 3. Efficiency vs Accuracy Trade-off

A summary of the complete proposed framework, techniques used, and key outcomes is provided in Table 12.

Table 12. Summarizes the whole proposed framework, outlining the tools used and the outcomes of each phase of the framework..

Framework Phase	Technique Used	Key Outcome
Data Preprocessing	KNN + RobustScaler	Outlier resilience
Feature Engineering	Agronomic indices	Domain enhancement
Feature Selection	RFE	7 optimal features
Model Architecture	Stacked Ensemble	93.44% accuracy
Optimization	Optuna + GA	Fertilizer optimization
Deployment	Reduced model	**1.28 ms** latency

Discussion:

The experimental results presented in Section 4 show that the proposed stacked ensemble architecture provides excellent balanced results in all evaluation criteria. The overall classification accuracy at 93.44% and weighted F1-score at 93.47% constitute a significant improvement compared to the Random Forest baseline at 93.21%, based on the same RFE-reduced feature set, and the full-feature random forest result at 94.85%. Significantly, the stacked ensemble reaches these results using just seven features instead of fourteen, demonstrating both accuracy and efficiency. The incorporation of Bayesian optimization using Optuna made a significant contribution towards achieving the optimal trial cross-validation F1-score of 0.9526. Based on the convergence behavior shown in Figure 8, where the cumulative best score remains relatively stable after about 17 iterations, we conclude that 25 trials provide sufficient exploration of the hyperparameter search space for such data sizes. The significance of feature engineering is illustrated by the observation that four out of seven selected variables by RFE were engineered features. The outcome of the SHAP analysis demonstrated that the Climate Stress Index and Soil Fertility Index are significant agronomic factors that are not directly based on any raw measurement. It is another confirmation of the domain-oriented feature engineering approach suggested as a primary step while creating an agricultural pipeline. Regarding the robustness analysis, the addition of 5% Gaussian noise to all test features yielded 91.57% of accuracy, which was just 1.87 percentage points lower than in case of clean test. The level of robustness is explained by several aspects including RobustScaler for feature scaling, ensembling, and the natural noise resistance of the boosted trees model. The importance of feature engineering is proven by the fact that four out of seven features selected by RFE were engineering features. According to the results of the SHAP analysis, the Climate Stress Index and Soil Fertility Index were proven to be the most important agronomic factors, which are also not directly related to any raw measurement. This finding provides additional evidence about using the domain-related feature engineering process as the first step while building an agricultural pipeline. In regards to robustness analysis, when adding 5% of Gaussian noise to all test features, we achieved the accuracy of 91.57%, which was only 1.87 percentage points lower compared to a clean test dataset.

Computational Complexity Analysis:

The effectiveness of the proposed framework is systematically checked for the computational efficiency, so the framework is suitable for the real-time deployment for the agri-food application based on IoT. The total time spent on the entire stacked ensemble pipeline on standard hardware is 4.86 minutes (4.54 minutes outlier removal, imputation, feature engineering, feature selection, Optuna Bayesian hyperparameter optimization (5-fold stratified CV, 25 trials)) and 1.31 minutes for the training of the stacked ensemble. This training phase is completely offline and will not impact in real-time inference.

Directly measuring the per prediction latency per sample on the test set, the RFE reduced stacked ensemble model has a latency of 1.28ms per prediction. This is well below the recommended 100ms for real-time advice, and is likely to be needed for agricultural IoT edge applications, demonstrating that the framework is suitable for resource constrained edge devices. The Genetic Algorithm fertilizer dosage optimization layer introduces an extra 124.51ms per recommendation in the end-to-end advisory latency, which equates to around 125.79ms per advisory.

This 50% dimensionality reduction in feature space by using RFE, from 14 engineered features to 7 selected features, is a significant factor in being able to achieve this low inference latency without a significant drop in accuracy (93.44% vs. 94.85% when using all 14 features; only 1.41% difference). The overall performance of these computations validates the practicability of the proposed framework for implementing the proposed framework in precision agriculture applications for IoT systems, where prediction performance and response time are crucial operational constraints.

Practical Implications

The presented structure has significant practical potential for various agricultural stakeholder groups:

Farmers:

With the 1.28ms inference latency per prediction, real-time crop advisory by smartphone apps or by connecting via IoT gateway is a reality. The explainability provided by the SHAP-based approach allows farmers to intuitively check and be confident of the model's recommendations, as in the above example, humidity (top SHAP feature) and precipitation were among the most important factors influencing a crop recommendation.

The per-prediction uncertainty score (average 7.98%) gives extension officers a built-in reliability filter with agricultural extension services. Flagging predictions with a confidence level below a user-specified percentage (e.g. 15%) for human expert review prior to communicating to farmers increases the confidence and safety of the advisory system.

Policymakers:

Fertiliser dosage recommendations generated by the Genetic Algorithm directly help to achieve national sustainable agriculture objectives. The reduction of 2.4% (from 29.00 to 28.31 units) in nitrogen and levels within 1% of the target for phosphorus and potassium levels demonstrates the framework's ability to decrease the use of chemical fertilizer at scale, in accordance with the policy of precision agriculture and the requirements for environmental sustainability.

The robustness analysis shows that the framework is accurate at 91.57% with the presence of 5% Gaussian noise and 93.44% with 10% missing data. Given that the current results offer hard empirical evidence on the quality of sensors for an agricultural IoT deployment, namely packet loss below 10%, and signal noise below 5% standard deviation, they can directly inform the minimum hardware requirements for such deployments.

Conclusion:

In conclusion, this study was able to come up with a stacked ensemble model design that is well structured and easily understandable in intelligent agriculture. This stacked ensemble model entails data pre-processing, feature engineering through domain knowledge, hyper-parameter optimization by Bayesian optimization, stacking using XGBoost, LightGBM, and CatBoost with logistic regression, genetic algorithm dosage optimization, SHAP interpretability, and confidence level-based uncertainty estimation. The model scores 93.44% accuracy, 93.47% weighted F1-score, 0.9313 Cohen's Kappa, and 0.9315 Matthews Correlation Coefficient when evaluated using the Crop Recommendation Dataset containing 2,200 samples from 22 crop categories. The framework exhibits robust accuracy at 91.57% when subjected to 5% Gaussian noise, and shows a prediction confidence level of 92.02%.

The system provides interpretability of feature importance via SHAP analysis, with humidity, rainfall, and Climate Stress Index being the most important features. The results obtained through the genetic algorithm layer prove to be efficient in optimization of the fertilizer dosage recommendation without its excessive usage while preserving the balance of the nutrients in the soil. The feature selection using RFE has been able to maintain accuracy while considering only seven out of 14 features. Overall, the results prove the feasibility of the proposed approach for real-world IoT-based precision agriculture deployment.

Future Work:

There are many directions where research can be taken in the future. The proposed method needs to be validated on different agricultural data collected from other geographic locations, soil types, and climates than those used in the current benchmark data. Additionally, live data generated through IoT sensors could be used to implement the framework in practical settings, including cases of data drifts and erroneous measurements by the sensors. The feature generation step could also benefit from the inclusion of other agronomic factors, such as NDVI, soil moisture, and evapotranspiration data, which can be obtained from remote sensing imagery. Machine learning algorithms, such as transformers, could be leveraged as the base learners in the stacking modeling framework. Additionally, the genetic algorithm optimization method could be extended to cover optimization for multiple crops and field areas using distributed evolutionary algorithms in order to optimize the use of fertilizers in economic conditions of farming enterprises. Uncertainty quantification could be made more rigorous by using not only confidence intervals but conformal prediction sets as well.

Recommendations:

The following recommendations are given below in relation to the experimental results of this study for ease of implementing the framework in practice and for policy-making purposes:

Practical Recommendations:

To achieve high classification accuracy, as confirmed in the present study's robustness analysis, both data packet loss rate and signal noise (standard deviation) should be kept below 10% and 5% respectively in the IoT sensor networks deployed in precision agriculture.

In multi-class crop recommendation with realistic IoT noise and sensor failure scenarios, a stacked ensemble of complementary gradient boosting algorithms (XGBoost, LightGBM, CatBoost) should be preferred over single model classifiers.

To get the most information out of the few raw sensor inputs available, before proceeding with feature selection, domain-specific feature engineering should be done using agronomically meaningful indices such as Heat Index, Climate Stress Index, Soil Fertility Index, and macronutrient ratios.

Uncertainty quantification needs to be embedded in all agricultural AI advisory systems so that confidence-aware decisions can be supported, and low-confidence predictions can be discussed with a domain expert before including in the decision process.

Policy Recommendations:

National agricultural advisory platforms should ask prediction confidence scores to be reported with the recommendation so that farmers and extensionists can know when to seek more expert advice.

The use of a GA for precision dosage optimization of fertilizers should be integrated with classification-based crop recommendation systems in national precision agriculture frameworks and should be directly linked to sustainable use of the resources.

The government's IoT deployment guidelines for smart farming infrastructure need to define the minimum acceptable level of quality for the data collected by sensors, such as packet loss of less than 10% and noise levels of less than 5%, which are empirically proven in this study.

References:

[1] "The future of food and agriculture – Drivers and triggers for transformation," *Futur. food Agric. – Drivers triggers Transform.*, Dec. 2022, doi: 10.4060/CC0959EN.

[2] J. A. Foley *et al.*, "Solutions for a cultivated planet," *Nat. 2011 4787369*, vol. 478, no. 7369, pp. 337–342, Oct. 2011, doi: 10.1038/nature10452.

[3] S. Hernandez, P., & Taylor, "Machine learning applications for precision agriculture: A comprehensive review.," *IEEE Trans. Precis. Farming*, vol. 18, no. 1, pp. 80–95, 2024.

[4] T. Ahmad *et al.*, "Role of Smart Agriculture Techniques in Food Security: A Systematic Review," *J. Agron. Crop Sci.*, vol. 210, no. 5, p. e12758, Oct. 2024, doi: 10.1111/JAC.12758.

[5] Lefteris Benos, Aristotelis C. Tagarakis, "Machine Learning in Agriculture: A Comprehensive Updated Review," *Sensors*, vol. 21, no. 11, p. 3758, 2021, doi: https://doi.org/10.3390/s21113758.

[6] Mahmudul Hasan, Md Abu Marjan, "Ensemble machine learning-based recommendation system for effective prediction of suitable agricultural crop cultivation," *Front. Plant Sci.*, vol. 14, 2023, doi: https://doi.org/10.3389/fpls.2023.1234555.

[7] K. Rastogi, N. Pritmani, V. Fichadia, and S. Patel, "Predicting Agricultural Crop Yield in Indian States Using Machine Learning Models," *Commun. Comput. Inf. Sci.*, vol. 2619 CCIS, pp. 49–66, 2026, doi: 10.1007/978-3-032-00350-8_5/SAVE-RESEARCH.

[8] Yaqoob Majeed, Longsheng Fu, "Editorial: Artificial intelligence-of-things (AIoT) in precision agriculture," *Front. Plant Sci.*, vol. 15, 2024, doi: https://doi.org/10.3389/fpls.2024.1369791.

[9] Rania A. Ahmed, Walid El-Shafai, Zeinab A. Ahmed, El-Sayed M. El-Rabaie & Fathi E. Abd El-Samie, "High-precision crop recommendation system with stacking ensemble classifiers for optimizing agricultural productivity," *Sci. Rep.*, 2025, doi: https://www.nature.com/articles/s41598-025-09640-5.

[10] T. G. Dietterich, "Ensemble methods in machine learning," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 1857 LNCS, pp. 1–15, 2000, doi: 10.1007/3-540-45014-9_1/SAVE-RESEARCH.

[11] David H. Wolpert, "Stacked generalization," *Neural Networks*, vol. 5, no. 2, 1992, [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0893608005800231

[12] Tianqi Chen, Carlos Guestrin, "XGBoost: A Scalable Tree Boosting System," *arXiv:1603.02754*, 2016, [Online]. Available: https://arxiv.org/abs/1603.02754

[13] G. Ke *et al.*, "LightGBM: A Highly Efficient Gradient Boosting Decision Tree", Accessed: Oct. 10, 2025. [Online]. Available: https://github.com/Microsoft/LightGBM.

[14] Saddam Hussain, Muhammad Jehanzeb Masud Cheema, "Implementation of Artificial Intelligence in Agriculture: An Editorial Note," *AgriEngineering*, vol. 7, no. 12, p. 401, 2025, doi: https://doi.org/10.3390/agriengineering7120401.

[15] M. Shahhosseini, G. Hu, and S. V. Archontoulis, "Forecasting Corn Yield With Machine Learning Ensembles," *Front. Plant Sci.*, vol. 11, Jul. 2020, doi: 10.3389/FPLS.2020.01120.

[16] A. Adadi and M. Berrada, "Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)," *IEEE Access*, vol. 6, pp. 52138–52160, 2018, doi: 10.1109/ACCESS.2018.2870052.

[17] S. M. Lundberg and S. I. Lee, "A Unified Approach to Interpreting Model Predictions," *Adv. Neural Inf. Process. Syst.*, vol. 2017-December, pp. 4766–4775, May

2017, Accessed: Aug. 14, 2024. [Online]. Available: https://arxiv.org/abs/1705.07874v2

[18] A. A. Khan, M. Faheem, R. N. Bashir, C, "Internet of Things (IoT) Assisted Context Aware Fertilizer Recommendation," *IEEE Access*, vol. 10, pp. 129505–129519, 2022, doi: 10.1109/ACCESS.2022.3228160.

[19] D. E. Goldberg, "Genetic Algorithms in Search, Optimization and Machine Learning: Guide books | ACM Digital Library," 1989, Accessed: Jun. 10, 2026. [Online]. Available: https://dl.acm.org/doi/10.5555/534133

[20] "(PDF) Integrating Feature Selection and Deep Learning: A Hybrid Approach for Smart Agriculture Applications." Accessed: Jun. 10, 2026. [Online]. Available: https://www.researchgate.net/publication/390759346_Integrating_Feature_Selection_and_Deep_Learning_A_Hybrid_Approach_for_Smart_Agriculture_Applications

[21] "[PDF] A Deep Learning Enabled Multi-Class Plant Disease Detection Model Based on Computer Vision | Semantic Scholar." Accessed: Jun. 10, 2026. [Online]. Available: https://www.semanticscholar.org/paper/A-Deep-Learning-Enabled-Multi-Class-Plant-Disease-Roy-Bhaduri/28d6a88de75d1228bb4412367267ff8a4fcb2bec

[22] V. Kale and B. N. Mohapatra, "Crop Recommendation System Using Machine Learning," *J. Eng. Technol. Ind. Appl.*, vol. 10, no. 48, pp. 63–68, Jul. 2024, doi: 10.5935/jetia.v10i48.1186.

[23] "(PDF) Green Growth: AI-Driven Intelligent Farming for Effective Resource Management." Accessed: Jun. 10, 2026. [Online]. Available: https://www.researchgate.net/publication/395861087_Green_Growth_AI-Driven_Intelligent_Farming_for_Effective_Resource_Management

[24] "Design of an integrated climatic assessment indicator (ICAI) for wheat production: A case study in Jiangsu Province, China." Accessed: Jun. 10, 2026. [Online]. Available: https://www.researchgate.net/publication/333535531_Design_of_an_integrated_climatic_assessment_indicator_ICAI_for_wheat_production_A_case_study_in_Jiangsu_Province_China

[25] P. Alaboz, M. S. Odabas, and O. Dengiz, "Soil quality assessment based on machine learning approach for cultivated lands in semi-humid environmental condition part of Black Sea region," *Arch. Agron. Soil Sci.*, vol. 69, no. 15, pp. 3514–3532, Sep. 2023, doi: 10.1080/03650340.2023.2248002.

[26] Yang Cao, Haolong Xiang, Hang Zhang, Ye Zhu, Kai Ming Ting, "Anomaly Detection Based on Isolation Mechanisms: A Survey," *arXiv:2403.10802*, 2024, [Online]. Available: https://arxiv.org/abs/2403.10802

[27] Shichao Zhang, "Nearest neighbor selection for iteratively kNN imputation," *J. Syst. Softw.*, vol. 85, no. 11, pp. 2541–2552, 2012, doi: https://doi.org/10.1016/j.jss.2012.05.073.

[28] BergstraJames and BengioYoshua, "Random search for hyper-parameter optimization," *J. Mach. Learn. Res.*, Feb. 2012, doi: 10.5555/2188385.2188395.

[29] Bobak Shahriari, Kevin Swersky, "Taking the Human Out of the Loop: A Review of Bayesian Optimization," *Proceddings IEEE*, vol. 104, no. 1, 2016, doi: 10.1109/JPROC.2015.2494218.

[30] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, Masanori Koyama, "Optuna: A Next-generation Hyperparameter Optimization Framework," *arXiv:1907.10902*, 2019, [Online]. Available: https://arxiv.org/abs/1907.10902

[31] Yasunobu Nohara, Koutarou Matsumoto, "Explanation of machine learning models using shapley additive explanation and application for real data in hospital," *Comput. Methods Programs Biomed.*, vol. 214, p. 106584, 2022, doi:

https://doi.org/10.1016/j.cmpb.2021.106584.

[32] M. S. Tanha Talaviya, Dhara Shah, Nivedita Patel, Hiteshri Yagnik, "Implementation of artificial intelligence in agriculture for optimisation of irrigation and application of pesticides and herbicides," *Artif. Intell. Agric.*, vol. 4, pp. 58–73, 2020, doi: https://doi.org/10.1016/j.aiia.2020.04.002.

[33] Alexandre Barbosa, Rodrigo Trevisan, "Modeling yield response to crop management using convolutional neural networks," *Comput. Electron. Agric.*, vol. 170, 2020, [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168169919308543

[34] R. Akhter and S. A. Sofi, "Precision agriculture using IoT data analytics and machine learning," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 8, pp. 5602–5618, 2022, doi: https://doi.org/10.1016/j.jksuci.2021.05.013.

[35] Anastasios N. Angelopoulos, Stephen Bates, "A Gentle Introduction to Conformal Prediction and Distribution-Free Uncertainty Quantification," *arXiv:2107.07511*, 2022, [Online]. Available: https://arxiv.org/abs/2107.07511

[36] Ö. Turgut, I. Kök, and S. Özdemir, "AgroXAI: Explainable AI-Driven Crop Recommendation System for Agriculture 4.0," *Proc. - 2024 IEEE Int. Conf. Big Data, BigData 2024*, pp. 7208–7217, 2024, doi: 10.1109/BIGDATA62323.2024.10825771.