OPEN ACCESS
International Journal of Innovations in Science & Technology

Green Growth: AI-Driven Intelligent Farming for Effective Resource Management

Javeria Rusool, Muhammad Afzal, Kishwar Rasool, Muskan Maryam
Department of Computer Science and IT, Ghazi University, D. G. Khan
*** Correspondence:** javeriarusool@gmail.com

Citation | Rusool. J, Afzal. M, Rasool. K, Maryam. M, "Green Growth: AI-Driven Intelligent Farming for Effective Resource Management", IJIST, Vol. 07 Issue. 03 pp 1677-1696, July 2025
Received | June 15, 2025 **Revised |** July 27, 2025 **Accepted |** July 28, 2025 **Published |** July 29, 2025.

Efficient fertilizer management plays a critical role in maximizing crop yield while reducing environmental harm and minimizing resource waste. This study presents an IoT-based intelligent fertilizer recommendation system designed to deliver accurate, real-time application guidance. The system integrates NPK sensors for soil nutrient detection, environmental sensors for humidity and temperature monitoring, and rain gauges to collect precipitation data. Data from the field is transmitted through an Arduino microcontroller to a cloud platform. A Random Forest classifier is used to determine the need for fertilization, while a CatBoost regressor estimates the required fertilizer quantity. The system was tested using real-time field data across 22 crop types, achieving 100% accuracy in classification and strong performance in regression tasks. Recommendations are automated and delivered via SMS to streamline field operations. The objective of this study is to develop an automated, sensor-driven fertilizer recommendation system using machine learning for precision agriculture. The novelty lies in the integration of real-time IoT sensing with hybrid AI models to optimize fertilizer use. This approach enhances productivity, reduces input waste, and supports environmentally sustainable farming.

Keywords: IoT Sensors 1; Microcontroller 2; Real-Time Monitoring 3; Cat Boost Regressor 4; Random Forest 5; and Crop Optimization 6;

Introduction:

During the 1980s and early 1990s, food and water supplies were generally adequate. However, rapid global population growth in recent decades has created increasing scarcity of these essential resources [1][2]. Agricultural systems now face two critical challenges: combating the impacts of climate change and meeting the nutritional demands of a growing population [3].

Traditional farming methods are struggling to keep up with rising demands for food production. These conventional practices lack the precision and data-driven insights required for efficient decision-making, leading to suboptimal use of inputs and reduced productivity. Food security depends on reliable harvests, which are influenced by fertilizer levels, crop growth stages, and soil nutrient content [4].

China has extensively implemented structured soil testing and fertilizer advisory techniques to improve agricultural efficiency [5]. Essential nutrients such as nitrogen, phosphorus, and potassium play significant roles in crop development. For example, phosphorus and nitrogen help reduce soil acidity and enhance ethylene synthesis, while potassium increases sugar content and ethylene production critical for the development of fruits like apples [6]. Hybrid rice and wheat productivity have significantly improved through the adoption of nutrient management techniques. The integration of chemical fertilizers, organic matter, and biological treatments has demonstrated the highest grain production and economic returns for these crops [7].

However, global population growth and rapid industrialization have led to reduced availability of arable land. In response, precision agriculture powered by the Internet of Things (IoT) and advanced data technologies has emerged as a promising solution to sustainably boost food production [8]. A growing trend in modern agriculture is the adoption of smart fertilization systems powered by IoT and sensing technologies, aimed at optimizing nutrient application and improving crop yields [9]. These systems are already in use for long-range monitoring of essential agricultural parameters such as soil conditions, irrigation efficiency, and water usage [10][11][12].

Technologies like distributed computing, cloud platforms, RFID, and wireless sensor networks are revolutionizing traditional farming by enabling real-time, data-driven operations [13][14]. IoT systems provide continuous environmental and soil data, allowing farmers to make informed decisions and increase overall productivity [15]. Nutrients such as nitrogen and phosphorus, which have been widely employed since the agricultural revolution, have significantly boosted crop yields but have also contributed to environmental concerns when misused [16][17].

Smart systems powered by machine learning analyze sensor-collected data to enhance both crop productivity and soil health [18]. These systems optimize nutrient delivery based on real-time environmental conditions. STFF (Soil Testing and Formula Fertilization) methods enable systematic, scientific fertilizer application [19]. Smart environments equipped with IoT adjust fertilization dynamically, accounting for moisture levels, crop needs, and weather conditions [20]. These systems also incorporate crop-specific nutrient demands, target yield levels, and fertilizer efficiency metrics determined through soil nutrient measurements [21].

Emerging digital agricultural technologies, including IoT, artificial intelligence (AI), and robotics, hold vast potential for improving productivity and ensuring sustainability [22]. Machine learning models such as Random Forest and CatBoost are increasingly applied in agriculture for tasks like drought prediction, crop classification, and yield forecasting [23][24]. These models are vital for reducing uncertainty in agricultural practices and optimizing resource use.

Agriculture is increasingly pressured to maximize crop yields while maintaining soil health and ensuring the efficient use of resources. Smallholder farmers, in particular, face challenges such as a lack of access to real-time information, incorrect fertilizer application, and poor utilization of available resources. These issues often lead to environmental degradation, low productivity, and high input costs. Compounded by erratic weather patterns and limited technological access, decision-making in agriculture becomes increasingly complex.

Existing machine learning (ML) models for predicting fertilizer needs typically lack precision and fail to provide timely, actionable recommendations, limiting their practical utility. To address these shortcomings, this study proposes an IoT-based, machine learning–driven fertilizer recommendation system that delivers accurate, real-time guidance for optimized nutrient use.

The proposed system integrates NPK sensors, temperature and humidity sensors, and rain gauges to collect field-level data, which is transmitted via Arduino microcontrollers to a cloud platform. A Random Forest classifier identifies the need for fertilization, while a CatBoost regressor predicts the optimal dosage. Fertilizer recommendations are then sent via SMS to farmers, facilitating timely and informed decisions.

Related Work:

This section focuses on existing literature in the above-mentioned subject and similar strategies in various fields.

Various studies have discussed the use of machine learning (ML) in optimizing fertilizer use and crop yield prediction. Conventional approaches rely on generalized recommendations, whereas machine learning methods utilize actual data to generate more accurate and tailored predictions for crops such as soybeans, rice, and maize. Extreme Random Tree (ERT) models have proven effective in yield prediction, achieving R^2 values of 0.744 for soybeans,0.775 for rice, and 0.749 for maize. Nevertheless, the research was characterized by high computational cost and variance errors, which decrease the accuracy of predictions [25]. Equally, Random Forest models have been applied when estimating the composition of fertilizers based on the study of color, and accuracy concerns still exist, along with issues regarding practical implementation in the real-time environment [26].

The Random Forest model was also implemented in sugarcane yield prediction, as it was trained under the meteorological and seasonal climate information. The model has attained 86.36 per cent precision in September preceding harvest, as well as 95.45 per cent in January of the harvest year. Nevertheless, due to the unusual weather conditions, it was not accurate enough, raising the question of adding climate resilience in the further models [27].

Real-time data collection has really enhanced the decision-making process in agriculture, all because of the Internet of Things (IoT). IoT-based Systems IoT-based solutions combine soil sensors, cloud computing technology, and ML models to automate and optimize fertilizer practices. For example, an IoT device powered by a CatBoost model was developed to monitor plant health and predict soil moisture levels, aiming to conserve water and boost crop productivity [28]. Another study uses CatBoost, LightGBM, and XGBoost to forecast agricultural productivity, with XGBoost achieving the highest accuracy at 99.12%. Nevertheless, uncertainties relating to the environment and changes in soil fertility present a difficulty when being handled in accounting [28].

One more IoT-related system created a synchronous data collection framework, and it collects real-time soil, humidity, and temperature data. Although this improved fertilizer recommendations and resource efficiency, the system requires further enhancement to become scalable, integrate satellite data effectively, and adapt to a variety of crops.[29]

Optimization algorithms have been employed to improve fertilizer efficiency. The combination of Extreme Random Tree (ERT) with the Cuckoo Search Algorithm (CSA) resulted in yield increases of 23.9% for maize, 13.3% for rice, and 20.3% for soybeans. Nevertheless, computational expenses and complicated data processing needs are serious problems [25].

Data scarcity in agricultural datasets has been addressed through the use of ensemble learning algorithms. A comparison of ML and ensemble learning methods revealed that XGBoost, CatBoost, and LightGBM are the best models, and in crop damage prediction, XGBoost corresponded to 89.56% accuracy. Nevertheless, the paper noted that imputation methods are largely affected by data distribution and the ratio of missingness. That is why more resistant procedures are necessary to deal with unfinished agricultural data and make the prediction process more precise [30].

Challenges in Prior Research and Proposed Improvements:

Despite significant advancements in emerging technologies such as machine learning (ML)- based agricultural systems and the Internet of Things (IoT), several limitations persist. Among these obstacles, computational complexity is a major one, and the cost of the models used, such as Extremely Randomized Trees (ERT) and Crow Search Algorithm (CSA), is high, which limits scalability. Environmental factors also pose a challenge, as many current models struggle to account for dynamic and intense climatic variations, thereby reducing the accuracy of their predictions. A third critical challenge is real-time data integration, as IoT-based models often struggle with the effective interpretation and assimilation of dynamic soil and climatic data. Lastly, generalizability remains a problem because fertilizer prescription algorithms very often fail to be adaptable to diverse types of crops and soils, which limits their applicability in any other farming settings. These problems should be addressed through optimization of computational efficiency, intake of trustworthy environmental modeling, enhancement of real-time data merging, as well as making recommendation systems more adaptive to different farming conditions.

Comparison of Related Work and Proposed Solution:

Table 1 below 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
ERT for crop yield prediction [25].	The ERT model applies to predict crop yield and smart fertilizer use.	The cost of computing is low, and errors related to variance lower the accuracy of the forecasts.	Create better parameter-tuned ERT models and use hybrid methods.
Random Forest for fertilizer estimation [26].	Used color analysis to estimate fertilizer composition.	Accuracy issues, real-time processing challenges, and compatibility limitations	Improve real-time processing, integrate spectral analysis for better estimation.
Random Forest for sugarcane yield [27].	Forecasted sugarcane yield using meteorological and seasonal data	Impact of climate change and extreme weather on accuracy	Incorporate climate resilience factors into the model
CatBoost-based IoT system [28].	Applied IoT sensors and CatBoost to forecast the level of soil moisture and to send fertilizers in a more optimized way.	Needed higher scalability, satellite data fusion, and not towards rigidity.	Make it more scalable, use satellite images, and transfer learning to have more use of family applications.

XGBoost for crop productivity [29].	Comparison of ML models for predicting crop productivity.	Accuracy was impacted by environmental uncertainties and differences in soil fertility.	Improve soil nutrition calculation, apply ensemble computing to improve robustness.
CSA + ERT for fertilizer optimization [25].	Optimization techniques for enhanced fertilizer application.	High computational costs.	Lightweight optimization of real-time algorithms.

Despite the significant achievements of earlier research, challenges such as computational complexity, unpredictable environmental conditions, and limited real-time adaptability persist. To enhance fertilizer recommendations and yield predictions, this study aims to bridge existing gaps by integrating an optimized machine learning framework with IoT-driven real-time data processing.

Precision agriculture has attracted growing interest for improving resource allocation and increasing crop yields; however, traditional fertilizer application methods still rely on manual

assessments, often leading to inefficient nutrient use. Machine learning models, such as Random Forest and CatBoost, have been extensively utilized to forecast soil fertility and suggest ideal fertilizer formulations by examining Nitrogen (N), Phosphorus (P), Potassium (K), and environmental variables like temperature, humidity, pH, and rainfall. IoT-driven technologies augment agricultural decision-making by facilitating real-time monitoring via sensors that relay soil data to cloud platforms for analysis, minimizing human intervention, and assuring accurate fertilizer application.

Figure 1 displays an IoT-ML-based fertilizer recommendation system in which sensor data is acquired, evaluated by a machine learning model to assess fertilizer requirements, and employed to activate an actuator for optimal nutrient distribution. Nevertheless, despite these advancements, several challenges persist such as limited model generalizability across regions, difficulties in integrating real-time data, and scalability constraints for large-scale agricultural operations. Numerous existing methods fail to integrate climate change and irregular weather patterns, consequently reducing their dependability. This study proposes a complete ML-IoT system to raise accuracy, optimize resource consumption, and improve agricultural yield via real-time soil analysis.

The system utilizes IoT sensors, including NPK, temperature, humidity sensors, and rain gauges, inside a structured five-layer IoT architecture for seamless data collecting, processing, and decision-making. Exploratory data analysis utilizing pair plots helps discover critical correlations between soil nutrients and environmental conditions, facilitating model development. Random Forest and CatBoost models are employed to estimate fertilizer requirements, and automated injectors connected with SMS messages provide exact and timely fertilizer application. The suggested approach is meant to be scalable and adaptable to varied agricultural situations, eventually boosting efficiency and sustainability in modern farming.

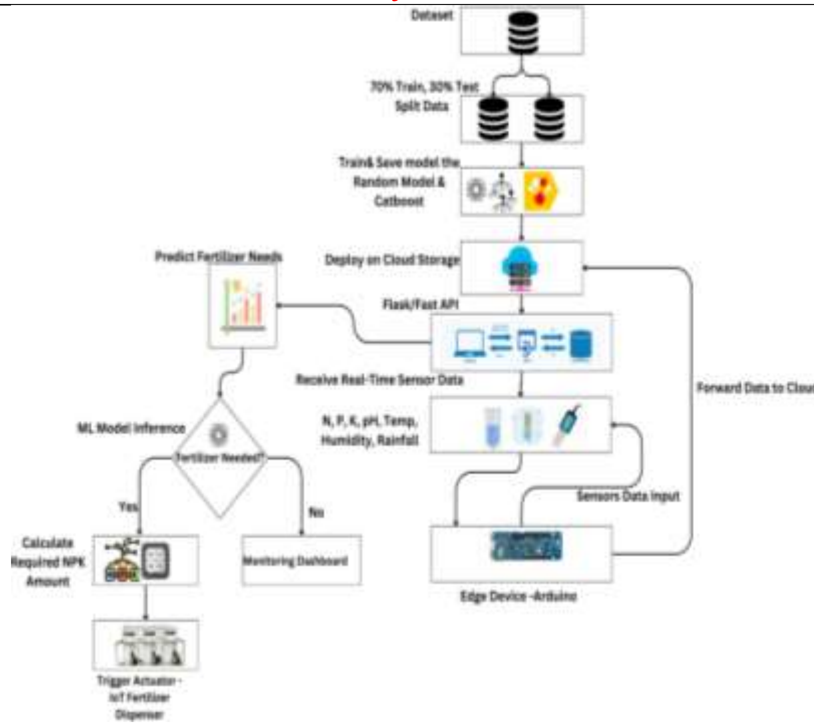

Figure 1. IoT-Enabled ML-Based Fertilizer Recommendation System Architecture

Novelty statement:

The novelty of this system lies in its integration of real-time multi-sensor data, hybrid ML models, and automated delivery of actionable insights. Unlike conventional approaches, it provides precise, location-specific guidance that enhances crop yield, reduces input waste, and supports environmentally sustainable farming. This research contributes to the emerging field of precision agriculture by connecting advanced technologies with practical, scalable solutions for smallholder farmers.

Material and Methods:

This study presents a Smart Fertilizer Recommendation System that provides real-time Nutrient requirement information for various crops. As shown in Figure 2, the system operates through three main steps: (1) IoT sensors collect soil quality data, (2) the data is pre-processed and analyzed, and (3) fertilizer requirements are calculated using machine learning models. Users get real-time fertilizer recommendation texts from the system, which provide personalized advice on how to produce maximum crop growth and soil health, depending on soil condition and the variety of the crop. It is extremely accurate, effective, and flexible to address the particular requirements of each fitting crop.

Collection of data with IOT Sensors:

The Arduino microcontroller is connected to the automated system of fertilization, and the configuration proposed to utilize the combination contains a variety of sensors. These sensors enable real-time monitoring of both ambient and soil conditions, focusing particularly on the most critical parameters for effective nutrient management. Arduino MKRFOX1200 is a mini power-efficient IoT board with the intended use of overseeing and managing the actual agricultural conditions, together with remote controls. It gathers data on real-time sensors and sends them to a cloud environment to process them. NPK sensor informs the level of nutrition in the soil of specific Nitrogen, Phosphorus, and Potassium nutrients by using the electrical conductivity method, thus indicating how nutritionally deficient the soil is, so that fertilizing can be done optimally. The DHT22 sensor accurately measures temperature and humidity using a single-wire connection, making it well-suited for monitoring the environmental conditions essential for crop growth. The high-accuracy data provided by a tipping bucket rain gauge allows evaluation of the soil moisture and the nutrient content. Moreover, the Arduino is tied up with an automation interface that connects an automated fertilizer injector to correctly apply fertilizer, and at the same time, it simulates the water levels to keep the supply steady.

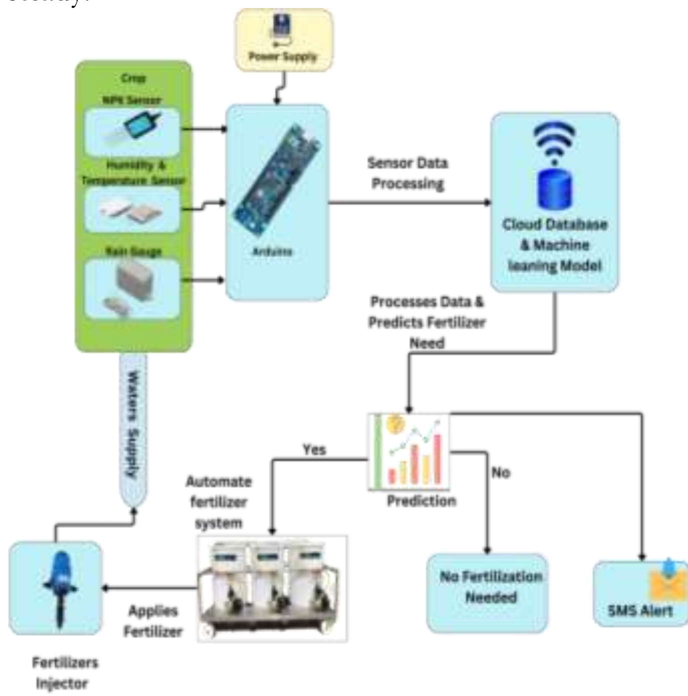

Figure 2. Proposed Layout of the Smart Fertilizer Recommendation Process.

Five-Layer IoT Network Layout for Smart Fertilizer Recommendation System:

Modern farming is advancing through the integration of the Internet of Things (IoT), forming a structured multi-level system that combines sensors, cloud computing, and communication technologies. This system increases the usage of resources and increases the productive agricultural land.

Figure 3 illustrates the five layers of the IoT network architecture used in smart agriculture, with each layer serving the following roles:

User Interface Layer: Helping farmers to control and track agricultural practices using dashboards, mobile notices, and automated suggestions via AI.

Process Layer: Processes aggregated data through cloud computing, machine learning algorithms, and predictive analytics to take data-informed decisions.

Connectivity Layer: Ensures that there is a seamless interconnection of the data between the devices and cloud platforms through the use of Wi-Fi and gateway solutions.

Computation Layer: Data preprocessing, filtering, and basic analysis are conducted on microcontrollers such as Arduino and other edge devices.

Sensor Layer: It records current sensory information about the environmental and soil reader conditions in the form of NPK sensor, temperature sensor, humidity sensor, and rain gauge.

This organized process enables real-time monitoring, effective decision-making, and targeted agricultural intervention, thus enhancing yield and sustainability.

Dataset Description and Preprocessing:

The dataset employed in this study plays a central role in developing the proposed fertilizer recommendation system. It was sourced from Kaggle and titled "Crop Recommendation Dataset" by author[31]. This dataset includes several key agronomic and environmental features;Nitrogen (N), Phosphorus (P), Potassium (K), temperature, humidity, pH, and rainfall that are crucial to crop development. It also covers 22 crop classes such as Apple, Banana, Watermelon, and Wheat.

The primary objective is to develop a predictive model that leverages Random Forest and CatBoost algorithms to recommend suitable fertilizers based on soil and environmental conditions. To train the model, we utilized the Kaggle dataset containing key features relevant to these conditions, enabling accurate fertilizer recommendations across diverse agricultural scenarios. The data model supports precision farming and enhances decision-making by farmers. Table 2 provides an overview of the key attributes and their importance in determining optimum fertilizer requirements.

Table 2. Dataset Description

	Nitrogen	Phosphorus	Potassium	Temperature	Humidity	PH_Value	Rainfall	Crop
0	5	136	195	22.356287	91.923605	6.264203	107.769741	Apple
1	24	128	196	22.750888	90.694892	5.521467	110.431786	Apple
2	10	136	204	21.198522	92.155951	6.276199	105.855435	Apple
3	82	78	46	29.148272	84.973237	5.738679	110.440880	Banana
4	91	84	52	24.900460	78.710248	6.390742	110.440880	Banana
61	109	21	55	24.900460	89.735242	6.770278	57.449421	Watermelon
62	118	15	45	24.214957	84.205770	6.538006	48.011385	Watermelon
63	31	76	82	20.824845	17.850571	7.599280	79.205092	wheat
64	24	55	78	17.302879	15.154059	6.649196	75.577904	wheat
65	56	67	78	17.574456	16.718266	8.255451	77.818914	Wheat

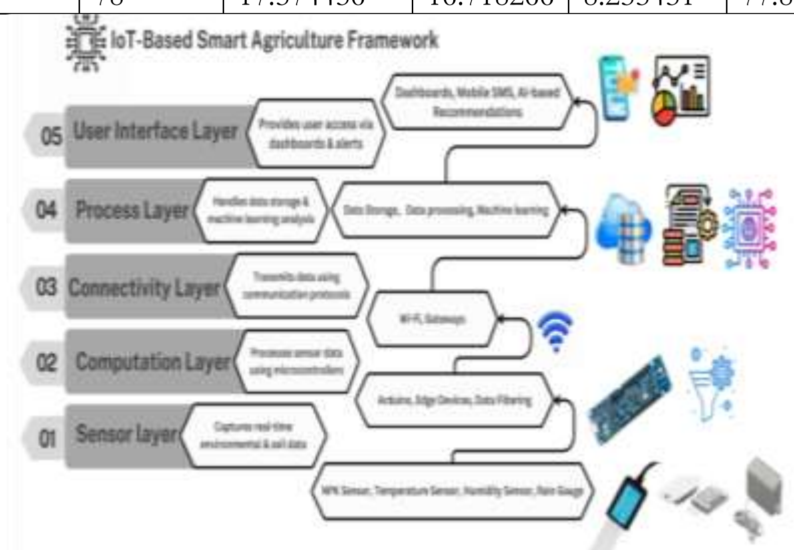

Figure 3. Five-Layer IoT Network Architecture for Smart Agriculture

Feature Engineering:

To improve the predictive capabilities of the proposed fertilizer recommendation system, two new features were engineered: **Fertilizer Needed** and **Fertilizer Quantity**. These features enable a hybrid modeling approach that supports both classification and regression tasks.

Fertilizer Needed (Binary Label):

This binary label determines whether a fertilizer application is necessary based on nutrient sufficiency thresholds. The following agronomic thresholds were established:

Nitrogen (N) < 100

Phosphorus (P) < 120

Potassium (K) < 130

Feature Engineering:

To improve the predictive capabilities of the proposed fertilizer recommendation system, two new features were engineered: Fertilizer Needed and Fertilizer Quantity. These features enable a hybrid modeling approach that supports both classification and regression tasks.

Fertilizer Needed (Binary Label):

This binary label determines whether a fertilizer application is necessary based on nutrient sufficiency thresholds. The following agronomic thresholds were established:

Nitrogen (N) < 100

Phosphorus (P) < 120

Potassium (K) < 130

If any of these conditions are met, the sample is labeled as requiring fertilizer ("Yes"); otherwise, it is marked as not required ("No"). These thresholds are grounded in agronomic research [32]. And were validated through exploratory analysis of the Kaggle dataset, which revealed that samples below these levels consistently exhibited suboptimal performance.

Fertilizer Quantity (Regression Target):

This continuous feature estimates the exact dosage of fertilizer required for each crop sample. It incorporates both the crop-specific baseline fertilizer requirement and nutrient deficiencies observed in the soil. The formulation is as follows:

Fertilizer Quantity=Base Amount_ crop +max (0, Deficiency in N) +max (0, Deficiency in P) +max (0, Deficiency in K)

Here, Base Amount_ crop refers to a predefined average fertilizer requirement for each crop species, and the max () functions ensure that only nutrient deficiencies influence the final recommendation. In crop-specific scenarios, this formulation allows the system to tailor fertilizer quantities according to the distinct needs of each crop while also addressing existing soil nutrient gaps.

Modeling Strategy:

These engineered features enable a two-stage machine learning pipeline:

A Random Forest classifier determines whether fertilization is needed (binary outcome).

A CatBoost regressor predicts the optimal quantity of fertilizer required (continuous output).

This integrated approach ensures both actionable decision support and precise dosage estimation, helping farmers make informed, data-driven interventions suited to real-time field conditions. A representative view of the engineered dataset is shown in Table 3.

Table 3. Fertilizer_ Prediction_ Data

	Nitrogen	Phosphorus	Potassium	Fertilizer Needed	Fertilizer Quantity
0	1	145	205	Yes	96
1	1	123	205	Yes	143
2	1	135	203	Yes	122
3	1	133	200	Yes	197
4	1	124	199	No	103

Pair Plot Analysis of Nutrients and Environmental Factors:

Figure 4 presents a pair plot that visualizes the relationships between key agricultural parameters, including Nitrogen, Phosphorus, Potassium, Temperature, Humidity, pH Value, and Rainfall. The diagonal elements of the plot display histograms, illustrating the distribution of each variable across the dataset. These histograms indicate varying distribution patterns, such as a multimodal distribution for nitrogen and a relatively uniform spread for rainfall. The off-diagonal elements of the pair plot provide scatter plots depicting pairwise correlations between variables. Several insights emerge from these visualizations. For instance, the interaction between Nitrogen, Phosphorus, and Potassium (NPK) nutrients shows distinct clustering, indicating possible grouping patterns among different crop types. Furthermore, the impact of environmental factors like Rainfall and Humidity on soil nutrients is evident, with certain variables exhibiting a dispersed pattern, suggesting a wide range of climatic conditions across different crop samples. Additionally, the presence of outliers in some feature combinations, such as extreme values of pH and Nitrogen, may indicate unique soil compositions or anomalies in data collection. By analyzing these patterns, the study aims to enhance crop yield prediction accuracy using machine learning models such as Random Forest and CatBoost. The insights gained from the pair plot help refine feature selection and model performance by highlighting key interactions among soil and environmental factors.

API Implementation for Real-Time Fertilizer Recommendation:

The Flask-based API development enabled real-time fertilizer recommendation access through web and mobile applications. The API functions by receiving soil sensor data, which it processes through machine learning models to produce exact fertilizer recommendations that relate to soil parameters. The API implements a RESTful architecture design to deliver multiple endpoints that support different functionalities. Soil parameter inputs such as nitrogen (N), phosphorus (P), potassium (K), temperature, humidity, pH, and rainfall can be provided to the /predict endpoint, which generates real-time fertilizer recommendations together with necessary fertilizer amounts. Through the /calculate-npk endpoint, the API determines NPK amounts needed for specific crops and deficient soil conditions to deliver optimal fertilizer doses. The /historical-data endpoint provides users access to archived predictions, which helps farmers evaluate fertilizer use patterns for making improved agricultural decisions. The /sensor-data endpoint establishes a database entry system that records live sensor data to generate valuable information for enhancing model accuracy. This API-based methodology allows the system to provide precise fertilizer advice, which drives both precise agricultural practices and sustainable soil care.

Model of Machine Learning for Fertilizer Prediction:

In this study, Random Forest Regression and CatBoost Regression have been selected for their strong performance in regression tasks, particularly in capturing complex nonlinear interactions. Those models were chosen because they are accurate, robust, and can be used for a range of datasets, including those containing categorical variables. Model Training and hyperparameter tuning.

To maintain the original class distribution across both sets, stratified sampling was employed to split the dataset in this study into training and testing sets in an 80:20 ratio. This procedure also helped to ensure that every subset reflected the entire diversity in the data, meaning that there is more trustworthy the assessment.

Random Forest: To produce predictions, we applied a machine learning technique referred to as Random Forest. The idea here is to combine many smaller decision-making models, such as trees, into one final option. We started from ($n_{ectimators}$100 trees of these, and found that increasing or otherwise modifying other parameters didn't bring much gain to the model's performance.

By averaging all trees' predictions, we can obtain a forecast yˆ for a given input x. The ultimate output is represented in equation 1

$$\hat{y} = \frac{1}{n_{estimator}} \sum_{t=1}^{n_{estimator}} h_t(x) \quad (1)$$

The performance of the Random Forest model was tested using the Mean Squared Error (MSE) metric, which is specified in equation 2

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (yi - \hat{y}i)^2 \quad (2)$$

For our Random Forest model, it was determined that the number of trees ($n_{ectimators}$) was to be set initially at 100 based on initial tuning. Since these changes had no observable effect on performance and the other hyperparameters stayed the same, I didn't make further modifications.

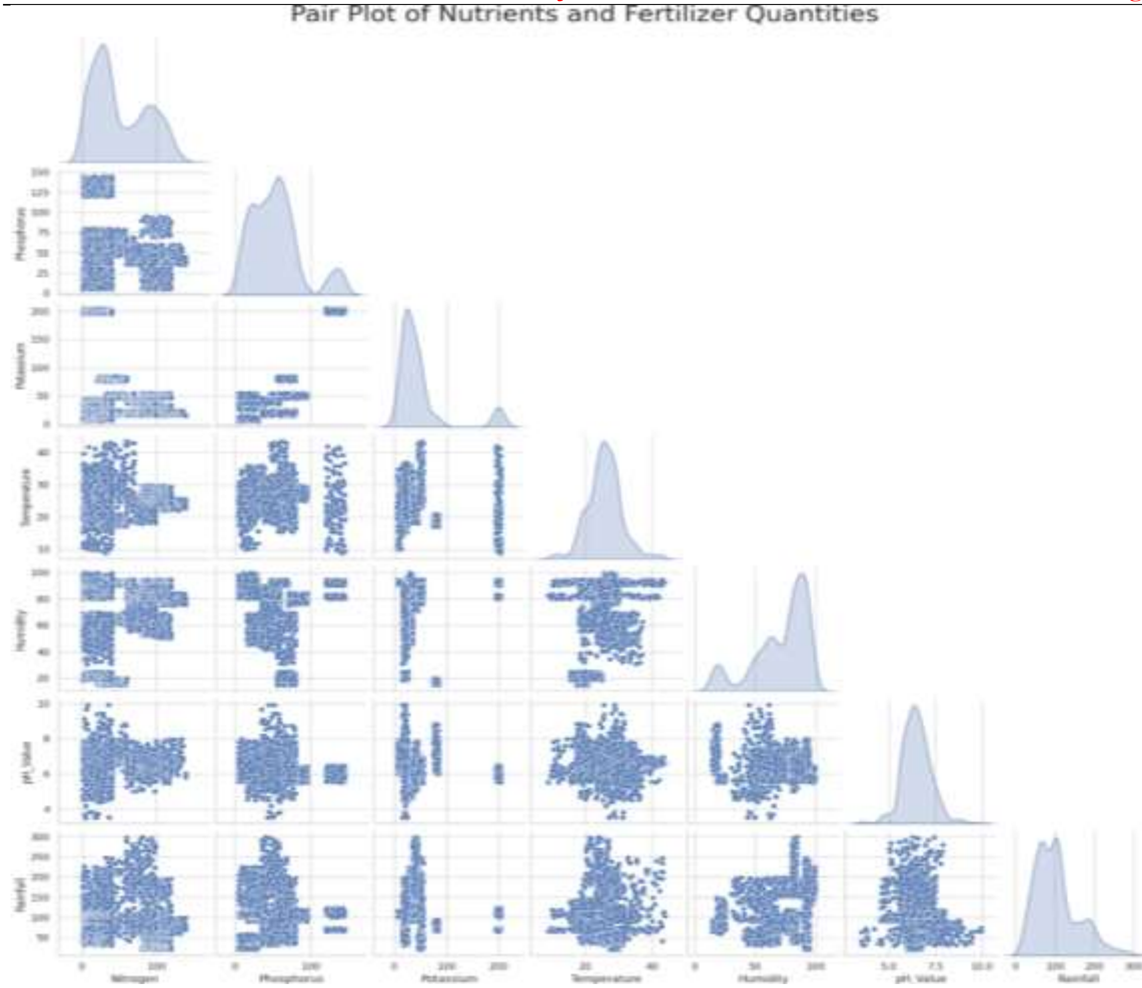

Figure 4. Pair Plot of Nutrients and Fertilizer Quantities Across Crops

Cat Boost Regressor:

Cat Boost, in the proposed method, is a gradient boosting algorithm selected for its efficiency in handling categorical data. To balance model complexity with training time, the CatBoost model was trained using 1,000 iterations, a tree depth of 6, and a learning rate of 0.1. Furthermore, an early stopping mechanism was implemented to minimize overfitting and ensure effective training. The model training was terminated if the performance on the validation set plateaued, meaning there was no substantial improvement over a defined number of iterations.

CatBoost builds the predictive model iteratively by adding trees that correct the residuals of previous iterations. The prediction at the t-th iteration is given by equation 3:

$$F_t(x) = F_{t-1}(x) + \eta \cdot gt(x) \quad (3)$$

Where η represents the learning rate and $gt(x)$ denotes the gradient of the loss function at iteration t. The model optimizes performance by minimizing the Root Mean Squared Error (RMSE), which is defined in equation 4:

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - F_T(x_i))^2} \quad (4)$$

Our models' validity and over-fitting prevention were achieved through k-fold cross-validation at a level of k=5. The k-fold approach splits the data into k sections so that the model trains k times, using different partitioned sections for validating the results and the remaining sections for training. The model's performance was evaluated by averaging the results from each fold in the cross-validation process, offering a more accurate representation of its generalization ability. The model's performance evaluation relied on the R^2 value computed through equation 5.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}i)^2} \quad (5)$$

yi is the true value, $\hat{y}$i is the predicted value, $\bar{y}$ is the mean of the true values across all samples.

"The overall workflow of the proposed fertilizer recommendation system is illustrated in Figure 5. It demonstrates the sequential flow from IoT-based data collection, cloud communication, preprocessing, machine learning prediction, and finally to user recommendation delivery."

Figure 5. Architecture of a Smart Agriculture System for Real-Time Fertilizer Recommendation Using Machine Learning"

Result and Discussion:

Random Forest Model Performance:

The Random Forest model provided a perfect predictive result on the amount of fertilizer required by various crops, depending on the parameters of the soil and the environment in which the crop would be located. Figure 6, which shows very closely the observed values (x-axis) and expected values (y-axis) along the 45-degree reference line, shows the correlation between the actual and predicted fertilizer levels. This demonstrates the potential to accurately assess fertilizer requirements across a wide range of crops and diverse environmental conditions.

The data points appear to almost perfectly line up, resulting in little variance between the actual from the expected. The model achieved a strong R^2 value of 0.999996, indicating that it captures nearly all the variability in fertilizer requirements. This conclusion is further supported by additional performance metrics. Furthermore, the high accuracy and reliability of the predictions are supported by the low MSE of 0.0052 of the models, suggesting that they reach very accurate fertilizer recommendations with a small amount of error. Table 4 below summarizes key performance metrics, including R^2, Mean Squared Error (MSE), and the model's overall alignment with the actual data.

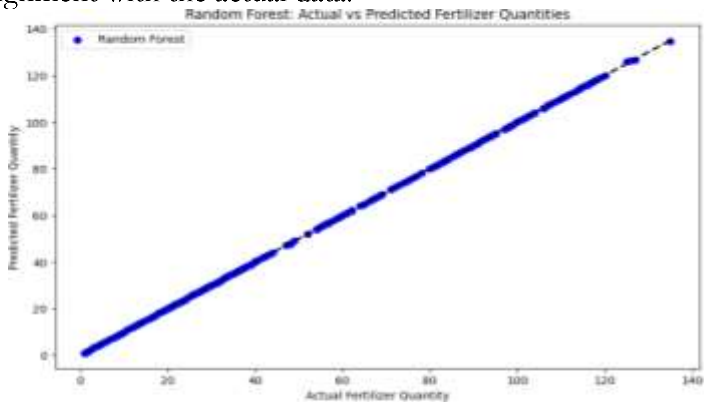

Figure 6. Relationship between Actual and Predicted Fertilizer Quantities (Random Forest Model.

Table 4. Summary of Performance Metrics for the Random Forest Model.

Metric	Value	Description
R^2 Score	0.999996	Indicates the proportion of variance in fertilizer quantity explained by the model's input features.
Mean Squared Error (MSE)	0.0052	Represents the average squared difference between actual and predicted fertilizer quantities.
Alignment with Actuals	High	Predicted values closely align with actual values along the 45-degree line, showing high accuracy.

Cat Boost Model Performance:

Figure 7 shows the performance of the CatBoost model, which has a high correlation between the actual and the predicted fertilizer quantities. Green dots can be viewed as a predicted value; the dotted line marks the possible 1:1 correlation. The closer the points match with the dashed line, the more controlling the predictions. The figure presents the difference between the actual and forecasted values of fertilizer through the CatBoost model. The graph shows a high level of linear correlation between the two sets of values, implying the high precision in the forecast of the model.

The green dots represent the predicted fertilizer levels, while the dashedline indicates a perfect prediction where the actual and predicted values are identical. The closer the data points are to the dashed line, the more accurate the model's predictions. The CatBoost model demonstrates a high level of accuracy in predicting fertilizer amounts, supporting the claim that it is an excellent predictive model.

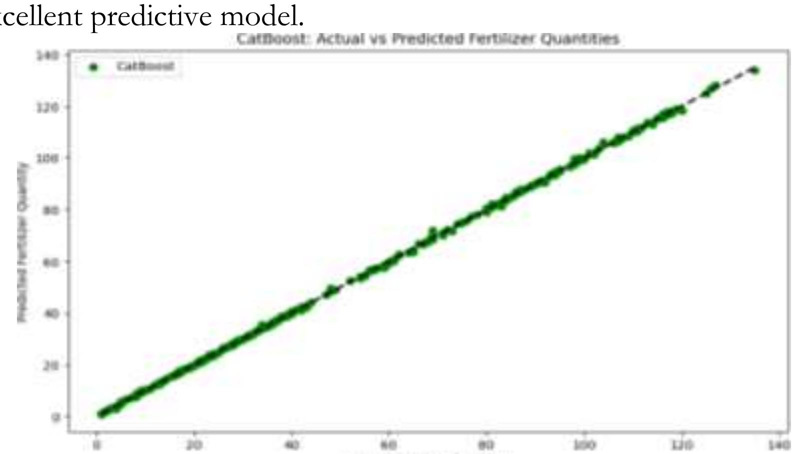

Figure 7. Cat Boost model

Cat Boost Model Training Progress:

Figure 8 displays a 3D image illustrating the training process of the CatBoost model, showing how total time (ms), test loss, learn loss, and best metrics evolve with the number of iterations. The blue line (learn) represents training loss as a function of iterations. The red line in the figure represents the lowest test loss observed up to the current iteration, indicating the model's best performance so far. As training progresses, the CatBoost model aligns with the graph, accurately reflecting the model's convergence behavior. Initially, both the training and test losses are high, but they decrease with each iteration. This occurs because, as long as the model remains imperfect, the best loss value continues to decline eventually reaching its minimum at iteration 999. As we iterate the model, total time (ms) increases, while the time before achieving optimal performance decreases. This reflects a good training process as the model converges toward the best solution. This 3D visualization complements the data in Table 5 by providing a more intuitive understanding of the model's performance and convergence.

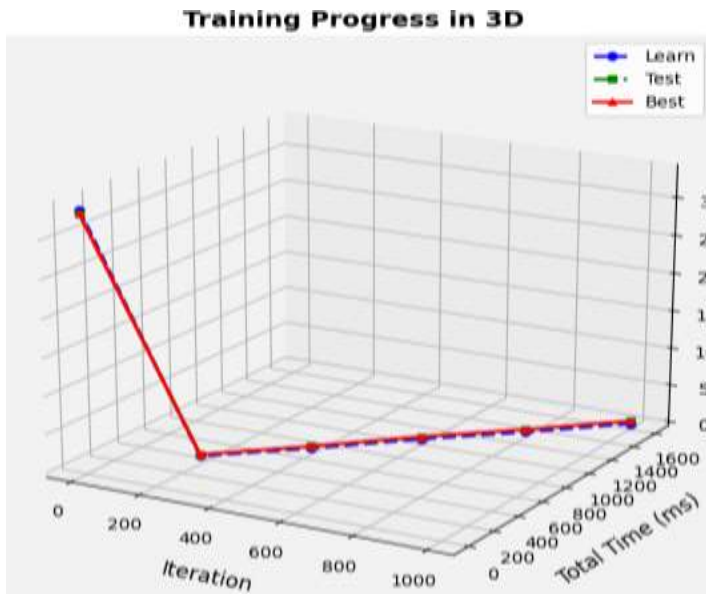

Figure 8. Cat Boost training progress

Model Performance Comparison:

As shown in Table 6 and Figure 9, A very low Mean Squared Error (MSE) of 0.005 and an R-squared value of 0.9999 proved the Random Forest model to be a near-perfect match. With an MSE of 0.238 and an R^2 value of 0.9998, CatBoost performed well; however, as illustrated

In Figure 7, Random Forest outperformed CatBoost in terms of prediction accuracy for this dataset.

Table 5. Training Progress

Iteration	Learn Loss	Test Loss	Best Test Loss	Total Time	Remaining Time
0	33.5865535	33.0503554	33.0503554	2.56	2.56
200	0.4995538	0.7009549	0.7009549	331	1.31
400	0.2954709	0.5746148	0.5746148	654	0.976
600	0.2042863	0.5242705	0.5242705	991	0.658
800	0.1522084	0.5010225	0.5010225	1290	0.321
999	0.1196821	0.4873802	0.4873802	1610	0

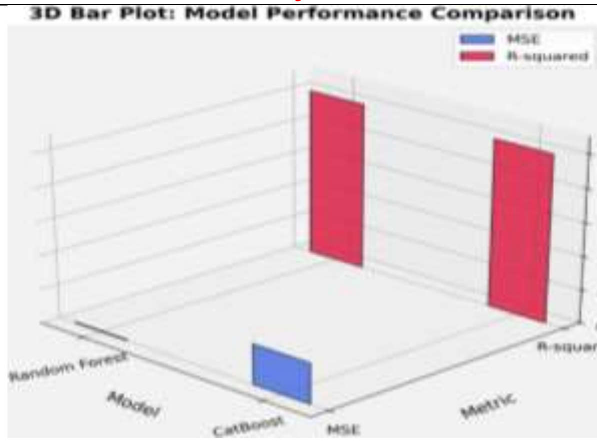

Figure 9. Comparison of model performance

Table 6. Comparison of model performance

Model	MSE	R-squared
Random Forest	0.005164545454545453	0.999961090572125
CatBoost	0.23753942961562133	0.999821038978844453

Visualization of Nutrition Requirements for Various Crops (in mL/ha):

Table 7 shows the requirements of nutrients in different crops under Nitrogen, Phosphorus, and Potassium (mL/ha). The figures show that there is a wide range of changes among the various crops, with cotton, pigeon peas, and coconut consuming the most nitrogen, followed by others like banana, rice, and watermelon that consume significantly less nitrogen. The requirement of phosphorus and potassium is quite constant among crops, but minor changes can be noticed.

A 3D representation of such nutrient requirements can be found in Figure 10, which shows the spatial distribution of Nitrogen (blue), Phosphorus (orange), and Potassium (green) for specific crop types.Visualization shows the importance of precision agriculture in improving the rational use of fertilizer so that crops receive the necessary nutrient amount according to their particular demands to increase their productivity and sustainability.

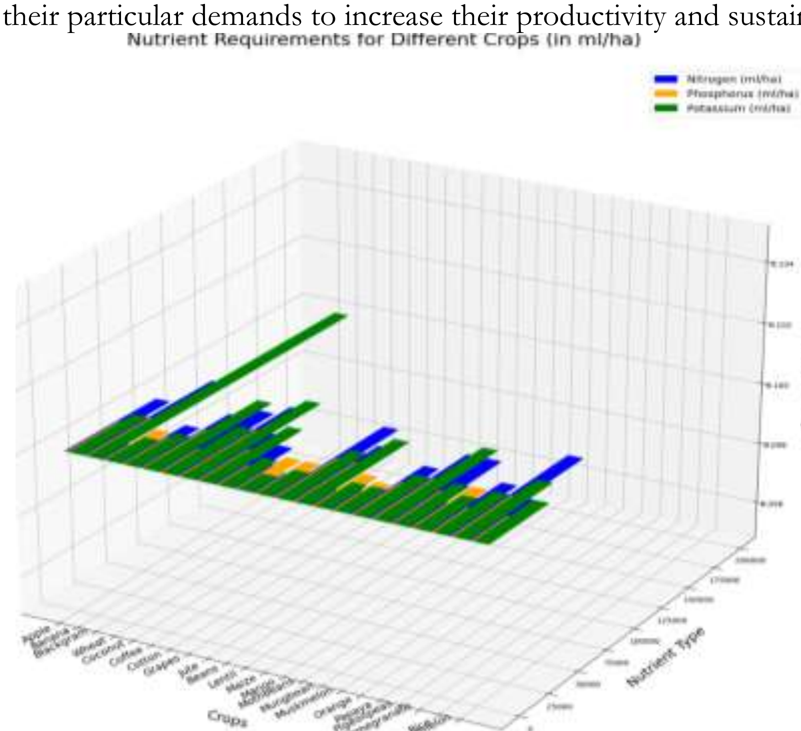

Figure 10. Nutrient Requirements for Different Crops

Table 7. Nutrient Requirements per Crop (ml/ha) for Nitrogen, Phosphorus, and Potassium.

Crop	Nitrogen (ml/ha)	Phosphorus (ml/ha)	Potassium (ml/ha)
Apple	100000	0.19	0.21
Banana	50000	0.1925	0.2075
Blackgram	75000	0.195	0.205
Wheat	125000	0.1975	0.2025
Coconut	150000	0.2	0.2
Coffee	175000	0.2025	0.1975
Cotton	200000	0.205	0.195
Grapes	100000	0.19	0.21
Beans	50000	0.1925	0.2075
Maize	75000	0.195	0.205
Mungbean	125000	0.1975	0.2025
Orange	150000	0.2	0.2
Papaya	175000	0.2025	0.1975
Pigeonpeas	200000	0.205	0.195
Pomegranate	100000	0.19	0.21
Rice	50000	0.1925	0.2075
Watermelon	75000	0.195	0.205

These findings underscore the importance of precision agriculture, emphasizing the need for crop-specific fertilizer recommendations to optimize agricultural productivity.

Matrix Analysis of Fertilizer Requirement Prediction:

The confusion parameters and the charts of the Random Forest and CatBoost models, as posted in Figure 11 and Figure 12, as well as the table of evaluation in Table 8, show that the models are perfect in predicting the number of fertilizers that should be used. Both models achieved a perfect fit of all the 261 instances where fertilizer was not needed and 179 where fertilizer was needed, and no instances of false positives or false negatives occurred. These findings suggest that the models are highly effective in distinguishing between cases where fertilizer is required and those where it is not.

Such an ideal classification implies that the models can be of great value to sustainable agriculture by maximizing the use of fertilizer. They can determine the fertilizer requirement and thus make the most appropriate choices and limit wastage of resources, alleviating environmental effects and promoting cost-effective agricultural behavior.

Table 8. Performance Evaluation.

Model	Predict: Not Required	Predict: Required
Actual Not Required	261	0
Actual Required	0	179

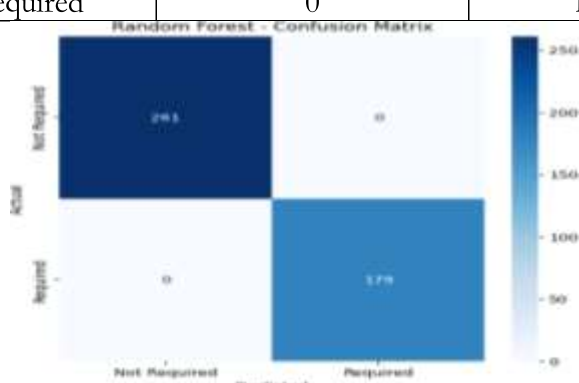

Figure 11. Random Forest Confusion Matrix.

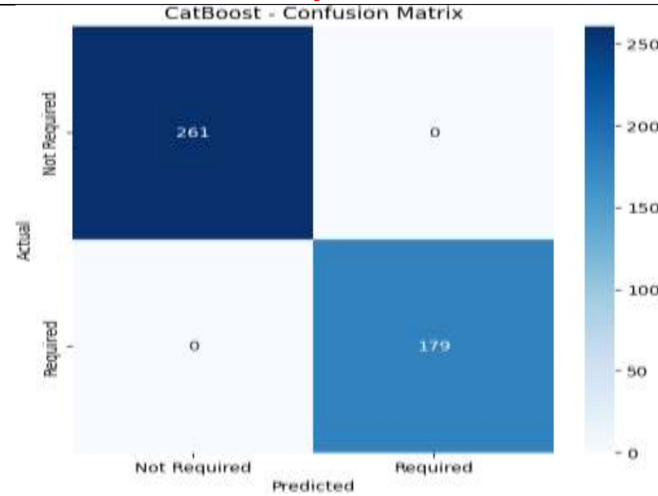

Figure 12. Cat Boost Confusion

Discussion:

This system illustrates the way classic planting can be altered under the unification of IoT and machine intelligence. Besides saving on cash, the intelligent fertilization system eliminates the environmental consequences of poor management of the fertilizer, such as nutrient loss through runoff and soil destruction. Approaching soil and environment conditions in this way, however, is the most appropriate means of providing specific recommendations that can be developed to enhance the health, productivity, and quality of the crops, which result in more successful agricultural operations , which, according to the appearance, seems to be more responsible.

The models are very accurate in their prediction, especially in the determination of fertilizer requirements, but are poor in their prediction of the quantity of fertilizers. The additional model fine-tuning, enhanced data gathering, and integration of more factors, like a type of soil, a specific stage of crop growth, and the production of the former crops, may make the system much more reliable and flexible to tackle this task. When we can work this system around additional crops and other geographical areas, then we could take part in sustainable farming by implementing local, data-informed decision-making in farms of many different types.

Lastly, this intelligent fertilizing system offers such an example of the metaphoric potential that IoT and AI can discover in building big-scale data-driven agricultural solutions. This finding indicates that it is just a prototype for a future in which precision agriculture can not only increase production but also efficiency in the management of the environment through farming practices that are agreeable to global sustainability principles.

Conclusion:

The smart fertilizer recommendation system proposed in the paper illustrates a data-driven method of precision agriculture with real-time collection of sensor data and its prediction using machine learning to optimize fertilizer recommendations. The five-layer approach to IoT allows fertilization to be carried out automatically and efficiently, so that the exact amount of nutrients is provided, and all the resources are used almost to the fullest. The study can prove through a comparative analysis of both the models of Random Forest and CatBoost that the former model performed significantly better than the latter because it had a high coefficient of fit of 0.9999, or nearly practically perfect, and a low error (known as MSE) of 0.0052, which is a very low error. Although no less effective, the CatBoost model demonstrated a slightly increased measure of MSE (0.238), showing enhanced and somewhat weaker predictive accuracy. Moreover, crop-specific nutrient breakdown looks into the significance of crop-specific fertilizer recommendations of soil properties and the environment. This system's strategic capabilities of machine learning and IoT integration can help derive in-time and data-based fertilizer recommendation decisions, further increasing agricultural efficiency, sustainability, and productivity.

Future Recommendations:

Future research can look into other deep learning algorithms, including even more sensors and adaptive learning, in order to further perfect the predictions and allow an autonomous smart farming system to make decisions. The study will help develop precision agriculture, increase sustainability in farming, and promote resource-efficient uses.

References:

[1] P. R. Harshani, T. Umamaheswari, R. Tharani, S. Rajalakshmi, and J. Dharani, "Effective Crop Productivity and Nutrient Level Monitoring in Agriculture Soil Using IOT," *ICSNS 2018 - Proc. IEEE Int. Conf. Soft-Computing Netw. Secur.*, Dec. 2018, doi: 10.1109/ICSNS.2018.8573674.

[2] S.-K. K. Divya Shanmugavel, Iryna Rusyn, Omar Solorza-Feria, "Sustainable SMART fertilizers in agriculture systems: A review on fundamentals to in-field applications," *Sci. Total Environ.*, vol. 904, p. 166729, 2023, doi: https://doi.org/10.1016/j.scitotenv.2023.166729.

[3] K. N. Bhanu, H. J. Jasmine, and H. S. Mahadevaswamy, "Machine learning Implementation in IoT based Intelligent System for Agriculture," *2020 Int. Conf. Emerg. Technol. INCET 2020*, Jun. 2020, doi: 10.1109/INCET49848.2020.9153978.

[4] N. S. R. Tamoghna Ojha, Sudip Misra, "Wireless sensor networks for agriculture: The state-of-the-art in practice and future challenges," *Comput. Electron. Agric.*, vol. 118, pp. 66–84, 2015, doi: https://doi.org/10.1016/j.compag.2015.08.011.

[5] H. M. S. K. Herath, D. Kumaragamage, and S. P. Indraratne, "Performance of an Optimized Nutrient Management Approach for Tomato in Central Sri Lanka," *Commun. Soil Sci. Plant Anal.*, vol. 44, no. 20, pp. 3049–3060, Nov. 2013, doi: 10.1080/00103624.2013.830122;REQUESTEDJOURNAL:JOURNAL:LCSS20;WGROUP:STRING:PUBLICATION.

[6] K. Kaack and H. L. Pedersen, "Effects Of Potassium, Phosphorus And Nitrogen Fertilization On Endogenous Ethylene And Quality Characteristics Of Apples (Malus Domestica L.)," *J. Plant Nutr.*, vol. 37, no. 7, pp. 1148–1155, 2014, doi: 10.1080/01904167.2013.868484.

[7] R. A. Shah and S. Kumar, "Direct and Residual Effect of Integrated Nutrient Management and Economics in Hybrid Rice Wheat Cropping System," *Environ. Sci*, 2014.

[8] M. M, P. N., B. Madhavi, T. Ramashri, and G. Haritha, "An Efficient Low Cost Iot Based Fertilizer Intimation System for Mulberry Plants," *SSRN Electron. J.*, Feb. 2023, doi: 10.2139/SSRN.4346092.

[9] Y. K. Raj Kishan Agrahari, "Smart fertilizer management: the progress of imaging technologies and possible implementation of plant biomarkers in agriculture," *Soil Sci. Plant Nutr.*, vol. 67, no. 3, 2021, doi: https://doi.org/10.1080/00380768.2021.1897479.

[10] V. T. P. Wei Zhao, Meini Wang, "Unmanned Aerial Vehicle and Geospatial Analysis in Smart Irrigation and Crop Monitoring on IoT Platform," *Mob. Inf. Syst.*, 2023, doi: https://doi.org/10.1155/2023/4213645.

[11] C. L. Wei Li, Muhammad Awais, Weimin Ru, Weidong Shi, Muhammad Ajmal, Saad Uddin, "Review of Sensor Network-Based Irrigation Systems Using IoT and Remote Sensing," *Adv. Meteorol.*, 2020, [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1155/2020/8396164

[12] I. W. M. Bastiaanssen, Wim G.M, David J Molden, "Remote sensing for irrigated

agriculture: examples from research and possible applications," *Agric. Water Manag.*, vol. 46, no. 2, pp. 137–155, 2000, doi: https://doi.org/10.1016/S0378-3774(00)00080-9.
[13] O. Diallo, J. J. P. C. Rodrigues, M. Sene, and J. Lloret, "Distributed database management techniques for wireless sensor networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 26, no. 2, pp. 604–620, Feb. 2015, doi: 10.1109/TPDS.2013.207.
[14] F. A. Alaba, M. Othman, I. A. T. Hashem, and F. Alotaibi, "Internet of Things security: A survey," *J. Netw. Comput. Appl.*, vol. 88, pp. 10–28, Jun. 2017, doi: 10.1016/J.JNCA.2017.04.002.
[15] R. Maheswari, H. Azath, P. Sharmila, and S. Sheeba Rani Gnanamalar, "Smart Village: Solar Based Smart Agriculture with IoT Enabled for Climatic Change and Fertilization of Soil," *2019 IEEE 5th Int. Conf. Mechatronics Syst. Robot. ICMSR 2019*, pp. 102–105, May 2019, doi: 10.1109/ICMSR.2019.8835454.
[16] M. de la L. M. Marcela Calabi-Floody, Jorge Medina, Cornelia Rumpel, Leo M. Condron, Marcela Hernandez, Marc Dumont, "Chapter Three - Smart Fertilizers as a Strategy for Sustainable Agriculture," *Adv. Agron.*, vol. 147, pp. 119–157, 2018, doi: https://doi.org/10.1016/bs.agron.2017.10.003.
[17] J. Sun, A. M. Abdulghani, M. A. Imran, and Q. H. Abbasi, "IoT Enabled Smart Fertilization and Irrigation Aid for Agricultural Purposes," *ACM Int. Conf. Proceeding Ser.*, pp. 71–75, Apr. 2020, doi: 10.1145/3398329.3398339;TOPIC:TOPIC:CONFERENCE-COLLECTIONS>CNIOT;PAGE:STRING:ARTICLE/CHAPTER.
[18] N. M. Basavaraju, U. B. Mahadevaswamy, and S. Mallikarjunaswamy, "Design and Implementation of Crop Yield Prediction and Fertilizer Utilization Using IoT and Machine Learning in Smart Agriculture Systems," *2nd IEEE Int. Conf. Networks, Multimed. Inf. Technol. NMITCON 2024*, 2024, doi: 10.1109/NMITCON62075.2024.10699184.
[19] W. X. TANG Han, WANG Jinwu, XU Changsu, ZHOU Wenqi, WANG Jinfeng, "Research Progress Analysis on Key Technology of Chemical Fertilizer Reduction and Efficiency Increase," *Nongye Jixie Xuebao/Transactions Chinese Soc. Agric. Mach.*, vol. 50, no. 4, 2019, [Online]. Available: https://www.nyjxxb.net/index.php/journal/article/view/792
[20] E. R. Youness Tace, Mohamed Tabaa, Sanaa Elfilali, Cherkaoui Leghris, Hassna Bensag, "Smart irrigation system based on IoT and machine learning," *Energy Reports*, vol. 8, no. 9, pp. 1025–1036, 2022, doi: https://doi.org/10.1016/j.egyr.2022.07.088.
[21] W. Z. Xinpeng Xu, Ping He a, Mirasol F. Pampolino, Adrian M. Johnston, Shaojun Qiu, Shicheng Zhao, Limin Chuan, "Fertilizer recommendation for maize in China based on yield response and agronomic efficiency," *F. Crop. Res.*, vol. 157, pp. 27–34, 2014, doi: https://doi.org/10.1016/j.fcr.2013.12.013.
[22] F. R. L. Tawseef Ayoub Shaikh, Tabasum Rasool, "Towards leveraging the role of machine learning and artificial intelligence in precision agriculture and smart farming," *Comput. Electron. Agric.*, vol. 198, p. 107119, 2022, doi: https://doi.org/10.1016/j.compag.2022.107119.
[23] E. M. Abdel-Rahman, F. B. Ahmed, and R. Ismail, "Random forest regression and spectral band selection for estimating sugarcane leaf nitrogen concentration using EO-1 Hyperion hyperspectral data," *Int. J. Remote Sens.*, vol. 34, no. 2, pp. 712–728, Jan. 2013, doi: 10.1080/01431161.2012.713142;JOURNAL:JOURNAL:TRES20.
[24] M. L. Junfei Chen, "Statistical Uncertainty Estimation Using Random Forests and Its Application to Drought Forecast," *Math. Probl. Eng.*, 2012, doi: https://doi.org/10.1155/2012/915053.

[25] W. Z. Jian Gao, "A Fertilization Decision Model for Maize, Rice, and Soybean Based on Machine Learning and Swarm Intelligent Search Algorithms," *Agronomy*, vol. 13, no. 5, p. 1400, 2023, doi: https://doi.org/10.3390/agronomy13051400.

[26] J. B. BAUTISTA, MARK JOHN B., CRUZ, JERSHWIN MARC A., PAGULAYAN, EUNICE VIELLE L, TABOR, DENNIS JUSTIN, FUENTES, JUSTINE5, SHEINN REYES, "UREA: A Case Study on Identifying Organic Fertilizer, Nutrients based on Color Characteristics Using Random Forest Algorithm for Small-scale Vegetable Farms," *IRE Journals*, vol. 7, no. 12, 2024, [Online]. Available: https://www.irejournals.com/formatedpaper/1705902.pdf

[27] D. S. & G. I.-B. Yvette Everingham, Justin Sexton, "Accurate prediction of sugarcane yield using a random forest algorithm," *Agron. Sustain. Dev.*, vol. 36, no. 27, 2016, doi: https://doi.org/10.1007/s13593-016-0364-z.

[28] R. S. Pavithra Mahesh, "Yield prediction for crops by gradient-based algorithms," *Plosone*, 2024, doi: https://doi.org/10.1371/journal.pone.0291928.

[29] K. O. Md Reazul Islam, "Machine learning enabled IoT system for soil nutrients monitoring and crop recommendation," *J. Agric. Food Res.*, vol. 14, p. 100880, 2023, doi: https://doi.org/10.1016/j.jafr.2023.100880.

[30] M. A. E.-R. Nermeen Gamal Rezk, Abdel-Fattah Attia, "An Efficient IoT-based Crop Damage Prediction Framework in Smart Agricultural Systems," *Res. Sq.*, vol. 8, 2024, [Online]. Available: https://www.researchsquare.com/article/rs-4697940/v1

[31] Atharva Ingle, "Crop Recommendation Dataset," *Kaggle*, 2025, [Online]. Available: https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset

[32] "cropsheets", [Online]. Available: https://aesl.ces.uga.edu/publications/soil/cropsheets.pdf