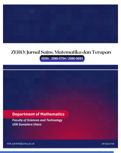
Zero : Jurnal Sains, Matematika, dan Terapan
E-ISSN : 2580-5754; P-ISSN : 2580-569X
Volume 9, Number 1, 2025
DOI: 10.30829/zero.v9i1.24421
Page: 150-161



Intelligent Actuator Control in Smart Agriculture through Machine Learning and Sensor Data Integration

[1] **Fadhillah Azmi**
Department of Electrical Engineering, Universitas Medan Area, Medan, Indonesia
[2] **M. Khalil Gibran**
Department of Computer Science, Universitas Islam Negeri Sumatera Utara, Medan, Indonesia
[3] **Insidini Fawwaz**
Department of Computer Science and Information Technology, Universitas Sumatera Utara, Medan, Indonesia
[4] **Rina Anugrahwaty**
Department of Telecommunication Engineering, Politeknik Negeri Medan, Medan, Indonesia
[5] **Amir Saleh**
Department of Computer Engineering and Informatics, Politeknik Negeri Medan, Medan, Indonesia

Article Info

Article history:

Accepted, 28 May 2025

Keywords:

Actuator Control Decisions;
Gradient Boosting;
Internet of Things (IoT);
Machine Learning;
Smart Argiculture.

ABSTRACT

Smart agriculture leverages Internet of Things (IoT) technology to develop intelligent greenhouses capable of monitoring and responding to environmental changes in real time. This study proposes the use of machine learning to analyze real-time sensor data—such as temperature, humidity, water level, and soil nutrient levels (N, P, K)—to determine the optimal timing for activating actuators, including fans, irrigation systems, and water pumps. In the initial stage, the study utilized the "IoT Agriculture 2024" dataset from Kaggle, which consists of 37,922 records and 13 attributes describing crop and environmental conditions. This dataset was used to train a robust machine learning model based on gradient boosting to support intelligent actuator control decisions. The model demonstrated strong predictive accuracy, achieving 99.62%. In the final stage, the model was evaluated in a simulated IoT-based agricultural system using synthetic sensor data designed to mimic real-world readings of temperature, humidity, soil moisture, and nutrient concentrations. The model achieved a high validation accuracy of 99.55%, indicating its reliability and robustness within the simulated environment. These results demonstrate that the integration of machine learning with real-time sensor data is an effective strategy for automating actuator control in smart greenhouses. The proposed approach has the potential to reduce manual intervention, optimize resource utilization, and improve overall agricultural productivity. This study contributes to the advancement of adaptive, data-driven precision agriculture systems that support long-term food security.

Corresponding Author:

Amir Saleh,
Department of Computer Engineering and Informatics,
Politeknik Negeri Medan, Medan, Indonesia
Email: amirsalehnst1990@gmail.comsollyaryzalubis@gmail.com

1. INTRODUCTION

Smart agriculture has been widely used to address increasingly complex challenges in the agricultural sector, such as unpredictable climate change, limited natural resources, and increasing global food needs [1], [2]. The most significant technology in the smart agriculture revolution is the smart greenhouse, which integrates Internet of Things (IoT) technology to monitor and control environmental conditions automatically and in real time. By utilizing various sensors installed in the greenhouse, this system can measure environmental parameters that greatly affect plant growth, such as temperature, soil moisture, water level, and soil nutrient content, such as nitrogen (N), phosphorus (P), and potassium (K) [3], [4], [5].

The data obtained from the sensors will be connected to the IoT system and then used to control various actuators in the greenhouse, such as fans, water pumps, and plant watering systems, with the aim of maintaining optimal conditions for plant growth. The fan will activate to cool the greenhouse temperature when it reaches a certain threshold. Likewise, if the soil moisture decreases, the watering system will automatically work to maintain the moisture needed by the plants. Thus, smart greenhouses reduce dependence on human intervention, allowing for more efficient and sustainable management [6].

However, although smart agriculture systems can collect large amounts of data and have automatic control capabilities, the main challenge faced is how to manage and analyze the data effectively to produce more accurate and efficient decisions [7]. The decision-making process based on sensor data is crucial because inappropriate decisions can lead to wastage of resources, such as energy and water, and reduce agricultural yields. Therefore, it is essential to develop an intelligent system that is capable of processing large and complex datasets to deliver reliable predictions and support optimal decision-making.

In this context, machine learning becomes a potential solution that can be used to help untangle the complexity of big data and assist in real-time data-based decision-making [8]. By using machine learning methods like KNN, SVM, NB, and Random Forest, these models can learn to predict whether an actuator should be turned on or off based on the environmental data collected [9], [10], [11]. Machine learning algorithms enable faster predictions and provide the ability to analyze patterns that may not be directly visible in the data, such as interactions between temperature, humidity, and soil nutrients that can affect actuator control decisions [12].

Various studies have been conducted to improve the efficiency of smart agriculture by integrating Internet of Things (IoT) technology and machine learning algorithms. Platero-Horcajadas et al. [13] developed a reinforcement learning-based climate control system connected to an IoT sensor network. This system is able to automatically control fans and water pumps based on temperature and humidity predictions, resulting in 15% energy efficiency and maintaining temperatures within ±1°C. Airlangga et al. [14] proposed a hybrid model of Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) to predict the status of fan actuators. This model utilizes the time pattern of temperature and humidity data and successfully achieves 96.8% accuracy in controlling greenhouse ventilation precisely.

Meanwhile, another approach was taken by Kaur et al. [11], who developed an IoT and machine learning-based automatic irrigation system. The study compared the Random Forest, Naïve Bayes, and KNN algorithms, with the best results achieved by Random Forest, which obtained 95.2% accuracy in classifying irrigation pump status based on soil moisture and weather. Meanwhile, Escamilla-García et al. [15] applied Backpropagation Neural Network (BPNN) to analyze temperature and humidity patterns in a greenhouse. This model was able to predict temperature with an error of less than 1.5°C and was used to automatically control heating in a simulated environment. The study by Farooq et al. [16] designed an IoT-based smart agriculture framework controlled by a deep neural network (DNN). This system was able to increase temperature control efficiency by up to 20% and operated with a latency of less than 5 seconds, allowing ventilation and sprinklers to work automatically based on current environmental conditions.

Based on previous research, machine learning-based approaches have enabled smart agriculture systems to operate more autonomously, reduce unnecessary resource usage, and support timely decision-making in environmental management. This study aims to design and validate a simulated control system utilizing the gradient boosting algorithm and other machine learning techniques to manage smart agricultural devices based on real-time environmental sensor data. By using parameters such as temperature, humidity, and soil nutrient levels, the proposed system is designed to determine the appropriate timing and actions for controlling actuators such as fans, water pumps, and irrigation systems.

In this architecture, environmental data collected by IoT-based sensors is represented in the form of synthetic data or simulated datasets, which is then transmitted to the machine learning model. The model analyzes the input and generates actuator control decisions that are then executed virtually within a simulation environment. The implementation of this system in a simulated scenario is expected to provide insights into its performance under real-world conditions, as well as help optimize the operation of agricultural devices, minimize waste, and improve productivity in a sustainable manner.

2. RESEARCH METHOD

This study aims to build a predictive model that can control smart agriculture actuators using machine learning algorithms, especially gradient boosting, based on data collected from various sensors in the greenhouse. The research methodology is divided into several main stages, which can be seen in Figure 1.

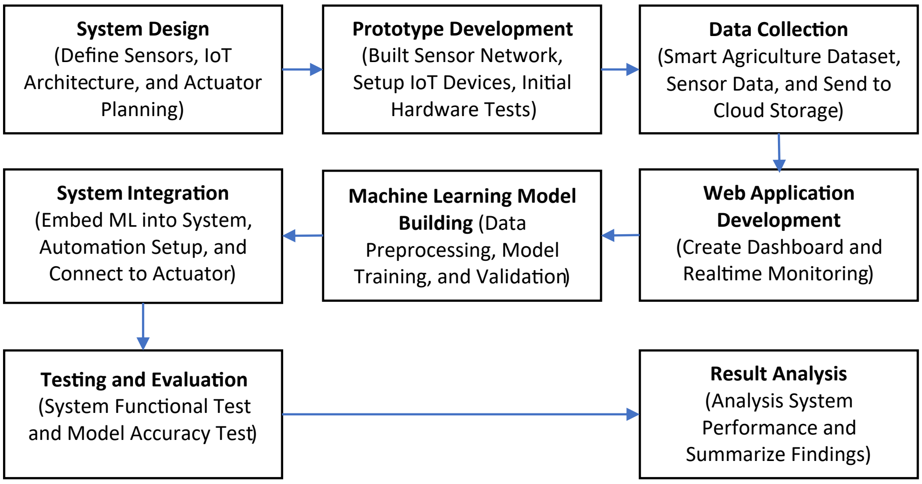


Figure 1. Flowchart of the Proposed Method

Each stage of Figure 1 can be described as follows:

2.1 System Design

At the system design stage, we developed an Internet of Things (IoT)-based architecture to enable intelligent monitoring and management in smart agriculture. The system employs various sensors with specific roles: a temperature sensor, which measures ambient temperature in degrees Celsius to monitor heat stress conditions; a humidity sensor, which measures air humidity in percentage to assess atmospheric moisture levels; a soil moisture sensor, which detects water content in the soil and outputs either a percentage or an analog value ranging from 0 to 255, used to determine irrigation needs; and a water level sensor, which measures the height of water in storage containers in centimeters, ensuring sufficient water supply. These sensors are connected to an ESP32 microcontroller, which reads and transmits the data in real time via Wi-Fi to a web-based database. The acquired data is used to formulate control rules and define precise thresholds for actuator activation.

In addition, a web application is developed as a monitoring dashboard, allowing users to monitor sensor data. The actuators used include fans, water pumps, and watering plant pumps, which can be operated automatically based on machine learning predictions from sensor data. With this design, the system is able to optimize plant environmental conditions adaptively, efficiently, and based on data. Figure 2 shows the system design proposed in this study.

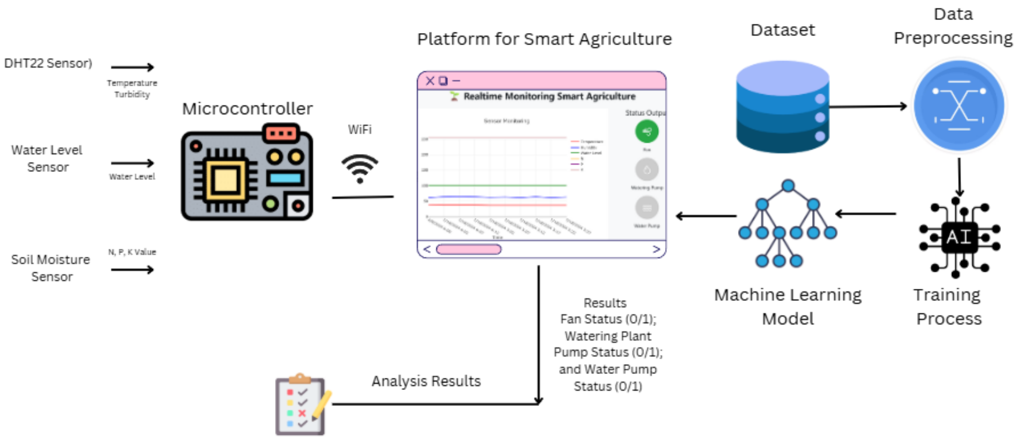


Figure 2. Proposed System Design

2.2 Prototype Development

After the system design stage is complete, the next process is prototype development, which realizes the design into a physical and functional form. At this stage, the selected sensors, such as temperature, humidity, soil moisture, and water level sensors, are assembled and connected to the main microcontroller, such as ESP32 or Arduino, using cable connections or breadboards. Additionally, the microcontroller installs and configures actuators such as fans, water pumps, and watering plant pumps for digital output control.

Next, initial programming is carried out on the microcontroller to read sensor data and test the basic capabilities of sending data to cloud storage using Wi-Fi connectivity. The system prototype is tested in a limited environment to ensure that all sensors are able to transmit data accurately and all actuators can function according to commands. This testing also includes sensor calibration and the preparation of environmental simulation scenarios to verify the actuator's response to various input conditions. Thus, the prototype development stage aims to ensure that the entire IoT system works stably and is ready to be integrated with advanced data processing and web-based applications. Figure 3 shows the development of a prototype of a smart agriculture system.

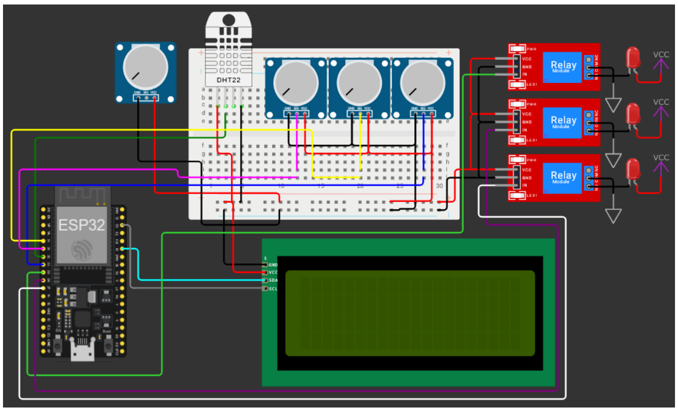


Figure 3. Prototype Development Design

The prototype in Figure 3 is designed to manage actuators in a smart agriculture system using sensor data integration and machine learning. This system uses various environmental sensors, such as temperature and humidity sensors (DHT22), water level sensors, and NPK soil sensors. Data collected from these sensors is sent to the ESP32 microcontroller via communication protocols such as I2C. The data is then sent to a server or IoT platform to be analyzed using the gradient boosting model. The results of this analysis determine whether actuators such as fans, water pumps, and watering plant pumps need to be activated or deactivated, allowing for more efficient resource management, increased crop productivity, and reduced environmental impact.

2.3 Data Collection

The data used in this study comes from a smart agriculture equipped with various sensors to monitor environmental conditions such as temperature, humidity, and soil nutrient content (nitrogen, phosphorus, and potassium). Additionally, the dataset includes the status of several greenhouse actuators, specifically fans, water pumps, and plant watering systems, which indicate whether each actuator is active (ON) or inactive (OFF). This dataset consists of 13 features with 37,922 rows of data obtained from Kaggle, which can be accessed on the internet page at the address https://www.kaggle.com/datasets/wisam1985/iot-agriculture-2024, which includes parameters such as those in Table 1 below.

Table 1. Dataset Used

No.	Feature Name	Description	Data Type
	Date	Timestamp of data collection	datetime
	temperature	Temperature inside the greenhouse	integer
	humidity	Humidity level inside the greenhouse	integer
	water_level	Soil moisture level	integer
	N	Nitrogen level in the soil (scale 0–255)	integer
	P	Phosphorus level in the soil (scale 0–255)	integer
	K	Potassium level in the soil (scale 0–255)	integer
	Fan_actuator_OFF	Fan actuator status indicator (OFF)	Boolean
	Fan_actuator_ON	Fan actuator status indicator (ON)	Boolean
0	Watering_plant_pump_OFF	Plant watering pump status indicator (OFF)	Boolean
1	Watering_plant_pump_ON	Plant watering pump status indicator (ON)	Boolean
2	Water_pump_actuator_OFF	Water pump actuator status indicator (OFF)	Boolean
3	Water_pump_actuator_ON	Water pump actuator status indicator (ON)	Boolean

The dataset in Table 1 contains environmental features such as temperature, humidity, water level, and soil nutrient content (N, P, K), which are the main inputs for building machine learning models in smart agriculture systems. These features determine the ideal conditions for crop growth and are used to predict the status of actuators (fan, watering plant pump, water pump) represented by the binary feature ON/OFF. Using algorithms such as Random Forest, Gradient Boosting, or Neural Networks, models can be trained to automatically manage temperature, humidity, and nutrient needs based on real-time data, thereby improving energy efficiency and crop productivity.

2.4 Machine Learning Model Building

At this stage, a machine learning-based system design is carried out with the aim of developing an intelligent prediction system for managing greenhouse actuators. The process begins with identifying relevant environmental parameters, such as temperature, air humidity, soil water level, and soil nutrient content (N, P, and K). The system is then designed to integrate data from these sensors into a decision-making architecture based on a predictive model.

The design process includes mapping the input data (from sensors) to the expected output (actuator responses), determining the types of actuators to be controlled (such as water pumps or ventilation fans), and adjusting the data structure for model training. Once the architecture and data flow are defined, the dataset is collected and divided into training and testing sets using a consistent split scheme (e.g., 80:20) to allow for objective model performance evaluation.

The resulting model will be used to predict when actuators should be activated based on environmental conditions detected by the sensors, enabling more efficient and responsive greenhouse management. Gradient boosting was chosen as the main algorithm in this study because of its ability to handle non-linear and complex relationships between features, as well as produce models with high accuracy [17]. Gradient boosting builds a robust prediction model by gradually combining several small decision trees, where each new tree tries to correct the errors of the previous tree [18], [19]. The model at the $M-th$ iteration can be expressed using the following in (1):

$$F_M(x) = F_0(x) + \sum_{m=1}^{M} \gamma_m h_m(x) \tag{1}$$

Where,

$F_0(x)$ is the initial model (bias, such as the target mean),
h_m is the model at the $m-th$ iteration (usually a small decision tree),
γ_m is the learning rate that controls the contribution of each model,
M is the number of iterations or trees.

In this study, other algorithms such as decision tree, random forest, KNN, SVM, logistic regression, and Naïve Bayes will also be used as comparison algorithms. Gradient boosting was chosen as the main approach because of its ability to produce more accurate and stable models in handling complex data [20], [21].

2.5 Testing and Evaluation

The testing and evaluation stage is carried out to test the overall performance of the system that has been built, both in terms of hardware functionality and the accuracy of machine learning predictions in managing actuators. Testing starts with a set of checks on the IoT system to make sure the sensor accurately reads environmental data, the data is sent to cloud storage smoothly, and the actuator reacts properly to commands from both manual control and machine learning predictions.

Next, the performance of the machine learning model was tested using a test dataset to measure accuracy, precision, recall, and F1-score in predicting actuator actions based on sensor data. The evaluation also includes a system reliability test during continuous operation under changing environmental simulation conditions to ensure that the system is able to respond to changing conditions quickly and accurately. In addition, the response time between sensor input, data processing, model prediction, and actuator activation was also analyzed to measure the operational efficiency of the system. This stage aims to ensure that the system works according to specifications, is stable, and is effective in supporting smart agricultural management.

In the final stage, we will check how well the model works by looking at its predictions against the current actuator prediction results, using measures such as accuracy (ACC), precision (PREC), recall (REC), and F1-score (F1), which is calculated with a specific formula: (2), (3), (4), and (5) [22], [23]:

$$ACC = \frac{TP+TN}{TP+FP+TN+FN} \tag{2}$$

$$PREC = \frac{TP}{TP+FP} \tag{3}$$

$$REC = \frac{TP}{TP+FN} \tag{4}$$

$$F1 = \frac{2 \times TP}{2 \times TP+FP+FN} \tag{5}$$

Where: TP (true positive) refers to events that are truly positive and predicted as positive. TN (true negative) refers to cases that are truly negative and predicted as negative. FP (false positive) refers to cases that are negative but predicted as positive, while FN (false negative) refers to cases that are positive but predicted as negative.

2.6 Web Application Development

The next stage in this research is the development of a web-based application that functions as the main interface for monitoring and controlling the smart agriculture system. The web application is designed so that users can monitor sensor data in real time, view historical data, and control actuators such as fans, water pumps, and planting lights remotely.

From the technology perspective, the web application was developed using Flask (Python) as the backend framework, which also integrates the machine learning model based on the gradient boosting algorithm. The front-end interface was designed using Bootstrap to ensure a responsive and user-friendly experience. MySQL is used as the database system to store and manage environmental and actuator data. The integration between the backend, database, and frontend is established through RESTful APIs, enabling dynamic data visualization and real-time actuator control. Figure 4 illustrates the architecture of the proposed web-based system.

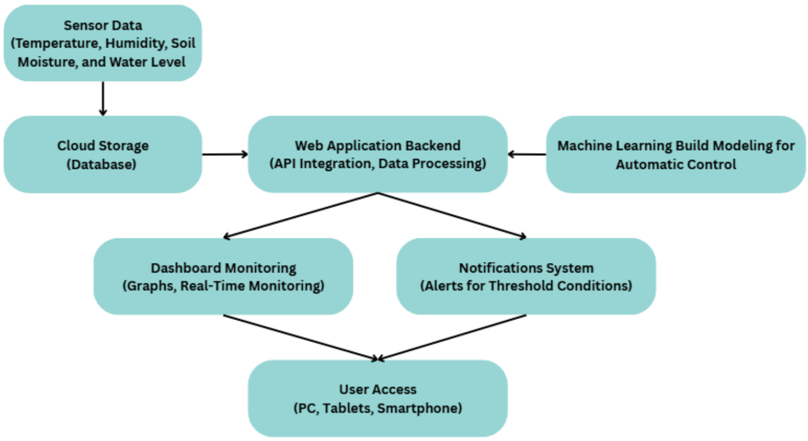


Figure 4. Proposed System Dashboard Development Design

The main features of this web application include a monitoring dashboard that displays graphs of changes in temperature, humidity, and soil nutrients, as well as an actuator page that is automatically controlled by activating or deactivating the actuator directly. The web application design is optimized to be responsive so that it can be accessed comfortably via computers, tablets, or smartphones. With this platform, greenhouse or agricultural area management can be done more efficiently, quickly, and based on actual data.

3. RESULT AND ANALYSIS

This study wants to create and test a machine learning model that can forecast when smart agriculture devices will turn on, using the gradient boosting algorithm. After conducting the machine learning model training stage, the results obtained from this study can be described as follows.

3.1. Machine Learning Model Performance

The dataset, consisting of 37,922 rows, will be cleaned and preprocessed prior to testing the machine learning model. The preprocessing steps include removing duplicate records, handling missing values using median imputation, normalizing numerical features using Min-Max scaling, and encoding boolean variables into binary format (e.g., True to 1, False to 0). These steps are essential to ensure the data is suitable for machine learning and to improve model accuracy and consistency.

The machine learning model is trained using 60% of the dataset for training, 20% for testing, and 20% for validation. The algorithm is trained with various sensor parameters used, such as temperature, humidity, nitrogen, phosphorus, potassium, and water level, to predict the actuator status (ON/OFF) of the fans, water pumps, and watering plant pumps. The next stage will be evaluated using the accuracy, precision, recall, and F1-score metrics for each output, which shows that the model has good performance in predicting the ON/OFF condition of the actuator. Table 2 shows the results of the comparison of machine learning method tests for each metric that has been carried out.

Table 2. Data Testing Results Using the Machine Learning Method

Model	Accuracy	Precision	Recall	F1-Score
Decision Tree	0.9954	0.9967	0.9965	0.9966
Random Forest	0.9955	0.9965	0.9969	0.9967
KNN	0.9897	0.9927	0.9933	0.9930
SVM	0.9934	0.9963	0.9940	0.9951
Logistic Regression	0.9925	0.9961	0.9930	0.9946

Naïve Bayes	0.9310	0.9456	0.9637	0.9541
Gradient Boosting	0.9962	0.9977	0.9967	0.9972

The evaluation results of seven machine learning algorithms show that the Gradient Boosting and Random Forest models have the best performance in predicting actuator status (such as fans, water pumps, and watering plant pumps) based on environmental sensor data. Gradient Boosting achieved an accuracy value of 0.9962, precision of 0.9977, recall of 0.9967, and F1-score of 0.9972. This value is very high and shows that the model is able to provide accurate and balanced predictions, both in recognizing positive and negative conditions. Random Forest showed results that were almost equivalent to Gradient Boosting, with an F1-score also reaching 0.9967 and an accuracy value of 0.9955.

The decision tree model, although simple, also provides excellent results with an accuracy of 0.9954, indicating that this model is able to provide fairly accurate predictions but with a higher possibility of overfitting when compared to ensemble models such as random forest and gradient boosting. Meanwhile, algorithms such as K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Logistic Regression show very decent performance, with an F1-score above 0.9900, so they remain an alternative if computational efficiency is a primary consideration. The Naïve Bayes model shows the lowest performance among all models tested, with an accuracy of 0.9310 and an F1-score of 0.9541. Although its recall is quite high, its precision is lower, indicating the model's tendency to produce excessive positive predictions.

Based on the overall results, the gradient boosting model was selected to be integrated into the IoT smart agriculture system. This model will be used to process sensor data in real-time and provide automatic decisions on actuator activation. Gradient boosting is a great choice for data-driven automation systems because it is very accurate and reliable, which helps improve the efficiency and flexibility of smart agricultural management.

3.2. Model Integration with IoT System

Integration of machine learning models with IoT systems is an important component of smart agriculture systems to ensure that sensor data can be collected, transmitted, and analyzed in realtime. In this study, IoT devices such as NodeMCU ESP32 are used to collect environmental data, including temperature, humidity, water level, and soil nutrients such as nitrogen (N), phosphorus (P), and potassium (K). This data is sent to the server using communication protocols such as HTTP, which is known to be fast and efficient for IoT data transfer. The display of the test results of IoT devices with machine learning integration can be seen in Figure 5 below.

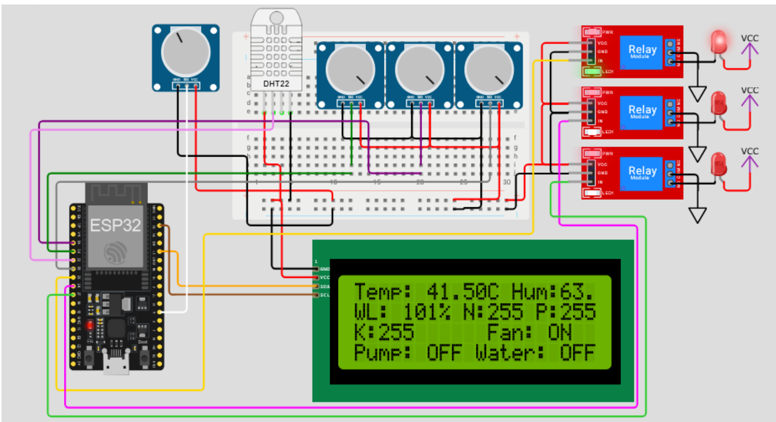


Figure 5. Prototype of IoT Device in Smart Agriculture System

The hardware shown in Figure 5 connects sensors directly to the microcontroller to measure environmental parameters in real time. To ensure the device functions properly, testing was performed on each component to measure the accuracy of sensor readings, Wi-Fi connection stability, and response time for sending data to the server. The test results indicate that this device is able to send data to the server in less than 1 second, with a communication success rate of over 90%.

After data is collected by the sensors, the IoT devices send the data to the server via a Wi-Fi connection, where a gradient boosting-based machine learning model is used to analyze the patterns and predict the status of the actuator, such as turning on the water pump, activating the fan, or opening the vents. Gradient boosting was chosen because of its ability to handle non-linear data, with high accuracy in recognizing complex patterns often found in environmental data.

Once the data is received by the server, an analysis process is carried out to determine whether the actuator needs to be activated or deactivated. The results of this analysis are then sent back to the IoT device to automatically control the actuator. In addition, this data is also forwarded to a web-based monitoring dashboard, allowing users to monitor environmental conditions in real time. This dashboard usually includes graphs of temperature, humidity, and soil nutrient levels, with actuator status indicators such as pumps and ventilation. Figure 6 shows an illustration of the results of real-time monitoring on the website dashboard display according to sensor input from the IoT devices used in the system.

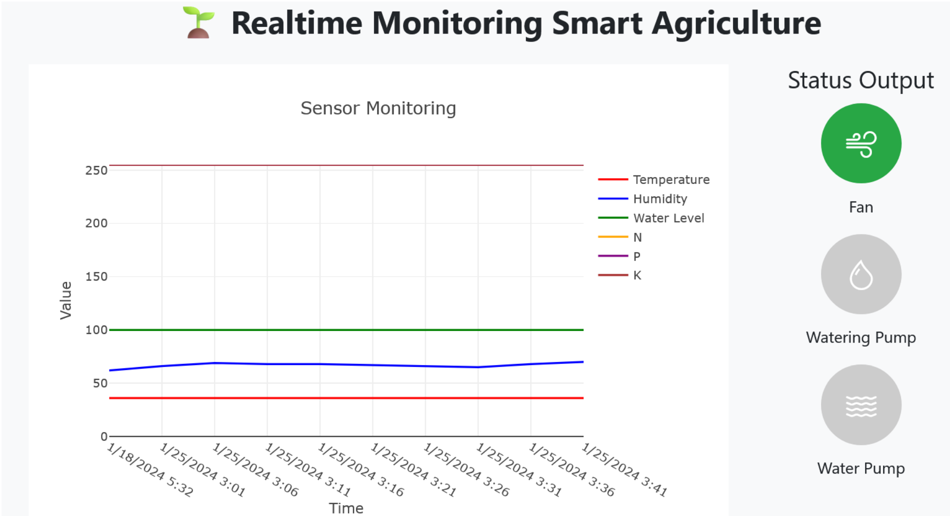


Figure 6. Monitoring Display on Website-Based Dashboard

The test results indicate that the integration of machine learning models with IoT devices in this system has an average prediction accuracy rate of 99%, with an optimal response time of around 2 to 3 seconds for each decision. This time depends on the speed of the Wi-Fi network and signal quality, as well as server configuration. By using gradient boosting, the system is able to recognize complex relationship patterns between features, such as interactions between temperature, humidity, and soil nutrients, resulting in more precise and efficient decisions.

The system also showed good stability during long-term testing, with a communication failure rate below 1%. This indicates that the gradient boosting-based approach is very effective in handling dynamic and varying sensor data, making it an ideal solution for smart agriculture systems.

3.3. Machine Learning Model Evaluation

Evaluating the performance of a machine learning model is essential to understanding how well it can accurately classify data. Some of the key metrics often used to measure model performance are accuracy, precision, recall, and F1-score. Accuracy measures the proportion of correct predictions to total predictions, precision shows how accurately the model predicts the positive class, recall measures how well the model captures all positive cases, and the F1 score is the harmonic mean between precision and recall, which provides a balanced picture of the accuracy and sensitivity of the model. Table 3 shows the results of evaluating several machine learning models based on these metrics.

Table 3. Evaluation Results Using Machine Learning Method

Model	Accuracy	Precision	Recall	F1-Score
Decision Tree	0.9941	0.9967	0.9946	0.9956
Random Forest	0.9950	0.9965	0.9962	0.9963
KNN	0.9884	0.9921	0.9927	0.9924
SVM	0.9920	0.9943	0.9938	0.9941
Logistic Regression	0.9912	0.9943	0.9927	0.9935
Naïve Bayes	0.9342	0.9496	0.9647	0.9566
Gradient Boosting	0.9955	0.9978	0.9956	0.9967

Table 2 shows the evaluation results of several machine learning models based on the accuracy, precision, recall, and F1-score metrics. The Gradient Boosting model has the best performance with an accuracy of 99.55% and an F1-score of 0.9967, indicating its excellent ability to capture complex patterns in the data. Random Forest also excels with an accuracy of 99.50% and an F1-score of 0.9963, reflecting its reliability in reducing overfitting compared to Decision Tree, which is slightly lower with an accuracy of 99.41%.

The KNN model has an accuracy of 98.84% and an F1-score of 0.9924, suitable for medium-sized data, although less than optimal for large data. SVM shows an accuracy of 99.20% and an F1-score of 0.9941, very good at avoiding false positives, but can be less efficient on large datasets. Logistic regression with an accuracy of 99.12% and an F1 score of 0.9935 shows competitive performance for linear data. Meanwhile, Naïve Bayes has a lower accuracy (93.42%) but is very fast and efficient, making it suitable for cases with many independent features, although it tends to produce more false positives.

To further highlight the consistency of the gradient boosting model, a detailed breakdown of its performance before and after the validation process is provided in Table 4. This table presents the values of accuracy, precision, recall, and F1-score, along with the percentage change for each metric. The results confirm that gradient boosting maintains excellent predictive capability with minimal performance fluctuation, reinforcing its reliability as the most robust model in this study.

Table 4. Gradient Boosting Model Performance Before and After Validation

Metric	Before Validation	After Validation	Improvement (%)
Accuracy	99.62%	99.55%	−0.07%
Precision	99.77%	99.78%	0.01%
Recall	99.67%	99.56%	−0.11%
F1-Score	99.72%	99.67%	−0.05%

The comparison results indicate that the Gradient Boosting model maintained the highest overall performance, with a post-validation accuracy of 99.55% and an F1-score of 0.9967, experiencing only a slight accuracy decrease of 0.07%, which demonstrates excellent consistency and generalization capability. This comparison reinforces the robustness of the gradient boosting model, which not only delivered the best performance during testing but also remained stable after undergoing model validation.

In addition, gradient boosting showed the smallest variation in performance across all metrics, with only a 0.05% drop in F1-score and a 0.11% drop in recall, while still slightly improving in precision by 0.01%. This consistency across metrics indicates that gradient boosting is not only accurate but also reliable in minimizing both false positives and false negatives. Therefore, it is considered the most effective and dependable model for actuator control prediction in smart agriculture systems.

3.4. Discussion

According to the results of the machine learning model evaluation, gradient boosting emerged as the most effective model for integration with IoT-based smart agriculture systems. This model achieved the highest accuracy (99.62%) along with the best precision, recall, and F1-score values among all evaluated models. Its outstanding predictive power demonstrates a strong capability to capture complex, non-linear relationships in environmental sensor data—such as variations in temperature, humidity, soil nutrient levels, and water level. These characteristics are critical in smart agriculture systems, where high-precision predictions are essential for managing actuators such as fans, water pumps, and irrigation systems. In this simulation, gradient boosting showed significant potential to optimize resource usage, reduce waste, and ultimately improve crop productivity and energy efficiency.

The superiority of gradient boosting is further validated by findings from previous studies. Airlangga et al. (2024) implemented the XGBoost algorithm in a smart greenhouse setting and achieved an accuracy of 94.47% for actuator control prediction [24]. Syed (2024) applied gradient boosting for smart agriculture using ensemble machine learning techniques in IoT environment and reported an accuracy of 99% [25]. Although the datasets and agricultural scenarios differ, the consistently high performance across these studies reinforces the robustness and reliability of gradient boosting as a leading model for real-time decision-making in data-driven agriculture.

For applications that require higher interpretability or operate under limited computational resources, alternative models such as Random Forest and Decision Tree may be preferable. These models also showed strong performance in this study, with accuracy levels exceeding 99.5%, and are inherently more interpretable—making it easier for farmers or agricultural managers to understand the logic behind automated decisions. For example, Random Forest can provide insights into which sensor features (e.g., soil moisture or air temperature) most influence actuator decisions like activating pumps or opening ventilation systems.

Ultimately, the choice of machine learning model for smart agriculture systems should be guided by the complexity of the data, hardware constraints, and the need for interpretability. While gradient boosting excels at modeling complex, non-linear systems with high accuracy, simpler models like decision trees or random forests offer greater transparency and faster computation, making them more suitable for real-time applications with limited resources. This study highlights the importance of aligning model selection with system requirements to achieve effective and scalable actuator management in IoT-enabled agriculture.

Although this research was conducted as a simulation, it provides clear evidence of the potential of gradient boosting for real-time predictive decision-making in irrigation scheduling, nutrient management, and microclimate control. The simulation results demonstrate how IoT-based intelligent systems can enhance resource efficiency, reduce waste, and contribute to sustainable agricultural practices.

The novelty of this study lies in the integration of real-time actuator control through simulation, environmental sensor fusion, and the application of the gradient boosting algorithm within a microcontroller-oriented, IoT-enabled smart agriculture architecture. Unlike previous approaches that mainly focused on static datasets or offline predictions, this study proposes a fully simulated architecture where sensor data—such as temperature, humidity, soil nutrients, and water levels—are processed in real time using a gradient boosting model embedded within a Flask-based backend. The predicted outcomes are then used to simulate autonomous control of actuators (e.g., fans, pumps, irrigation systems), enabling the system to respond adaptively and intelligently to environmental changes. This approach not only improves predictive accuracy and system responsiveness but also demonstrates a viable direction for enhancing operational efficiency and sustainability in precision agriculture.

4. CONCLUSION

This study aimed to develop an intelligent actuator management system for smart agriculture by integrating sensor data with machine learning techniques. The experimental results demonstrate that the use of the gradient boosting model significantly improves actuator control performance, achieving a test accuracy of 99.62% and an evaluation accuracy of 96.55%. These results outperform other machine learning algorithms tested, confirming the superiority of gradient boosting in handling complex, multivariate sensor data within IoT-based agricultural environments. The main contribution of this study lies in its integration of environmental sensing and predictive modeling to support real-time, data-driven decision-making in agriculture. The novelty of this work lies in the implementation of a high-performing, interpretable machine learning model—gradient boosting—within a simulated, microcontroller-oriented IoT system for real-time actuator control. Unlike many prior studies that focus solely on data analysis or offline prediction, this study simulates a fully connected framework that fuses multiple environmental sensor inputs—such as temperature, humidity, soil nutrients, and water levels—and processes them in real-time to drive actuator decisions. From a practical perspective, the proposed system can enhance agricultural productivity by enabling precise and timely actuator responses, reducing water and energy consumption, and minimizing environmental impact. Theoretically, this research provides insights into the selection and adaptation of machine learning models for IoT ecosystems in agriculture. To achieve optimal performance, model selection must consider system goals, computational capacity, and data complexity. Overall, this research demonstrates a promising direction for intelligent automation in agriculture and offers a scalable approach for sustainable resource management.

5. REFERENCES

[1] W. Kabato, G. T. Getnet, T. Sinore, A. Nemeth, and Z. Molnár, "Towards Climate-Smart Agriculture: Strategies for Sustainable Agricultural Production, Food Security, and Greenhouse Gas Reduction," Mar. 01, 2025, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/agronomy15030565.

[2] Q. Javaid, "MZ Journals Sustainable Solutions for a Warming Planet Climate Smart Agriculture as a Tool for Global Food Security." [Online]. Available: https://mzjournal.com/index.php/MZCJ

[3] M. N. Reza *et al.*, "Trends of Soil and Solution Nutrient Sensing for Open Field and Hydroponic Cultivation in Facilitated Smart Agriculture," Jan. 01, 2025, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/s25020453.

[4] M. K. Senapaty, A. Ray, and N. Padhy, "IoT-Enabled Soil Nutrient Analysis and Crop Recommendation Model for Precision Agriculture," *Computers*, vol. 12, no. 3, Mar. 2023, doi: 10.3390/computers12030061.

[5] P. Musa, H. Sugeru, and E. P. Wibowo, "Wireless Sensor Networks for Precision Agriculture: A Review of NPK Sensor Implementations," Jan. 01, 2024, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/s24010051.

[6] A. Hoque, A. S. Mazumder, S. Roy, P. Saikia, and K. Kumar, "Transformative Approaches to Agricultural Sustainability: Automation, Smart Greenhouses, and AI," *Int J Res Appl Sci Eng Technol*, vol. 13, pp. 1011–1023, Jan. 2025, doi: 10.22214/ijraset.2025.66494.

[7] R. Wacas, "SMART GREENHOUSE: AN IOT-BASED MONITORING AND CONTROL SYSTEM WITH PREDICTIVE ANALYTICS," vol. 8, pp. 2423–3080, Nov. 2024.

[8] R. Rayhana, G. Xiao, and Z. Liu, "Internet of Things Empowered Smart Greenhouse," *IEEE Journal of Radio Frequency Identification*, vol. PP, p. 1, Mar. 2020, doi: 10.1109/JRFID.2020.2984391.

[9] S. Islam *et al.*, "Image Processing and Support Vector Machine (SVM) for Classifying Environmental Stress Symptoms of Pepper Seedlings Grown in a Plant Factory," *Agronomy*, vol. 14, p. 2043, Sep. 2024, doi: 10.3390/agronomy14092043.

[10] P. Gourshettiwar and K. Reddy, "A Comprehensive Analysis with Machine Learning Algorithm and IoT Integration in Hydroponic Vegetable System for Nu-trition Management of Plants/Crops," Sep. 05, 2024. doi: 10.20944/preprints202409.0279.v1.

[11] A. Kaur, D. P. Bhatt, and L. Raja, "Developing a Hybrid Irrigation System for Smart Agriculture Using IoT Sensors and Machine Learning in Sri Ganganagar, Rajasthan," *J Sens*, vol. 2024, no. 1, p. 6676907, Jan. 2024, doi: https://doi.org/10.1155/2024/6676907.

[12] Y. Mekonnen, S. Namuduri, L. Burton, A. Sarwat, and S. Bhansali, "Review—Machine Learning Techniques in Wireless Sensor Network Based Precision Agriculture," *J Electrochem Soc*, vol. 167, no. 3, p. 037522, Jan. 2020, doi: 10.1149/2.0222003jes.

[13] M. Platero-Horcajadas, S. Pardo-Pina, J. M. Cámara-Zapata, J. A. Brenes-Carranza, and F. J. Ferrández-Pastor, "Enhancing Greenhouse Efficiency: Integrating IoT and Reinforcement Learning for Optimized Climate Control," *Sensors*, vol. 24, no. 24, Dec. 2024, doi: 10.3390/s24248109.

[14] G. Airlangga, J. Bata, O. I. Adi Nugroho, and B. H. P. Lim, "Hybrid CNN-LSTM Model with Custom Activation and Loss Functions for Predicting Fan Actuator States in Smart Greenhouses," *AgriEngineering*, vol. 7, no. 4, Apr. 2025, doi: 10.3390/agriengineering7040118.

[15] A. Escamilla-García, G. M. Soto-Zarazúa, M. Toledano-Ayala, E. Rivas-Araiza, and A. Gastélum-Barrios, "Applications of artificial neural networks in greenhouse technology and overview for smart agriculture development," Jun. 01, 2020, *MDPI AG*. doi: 10.3390/app10113835.

[16] M. S. Farooq, R. Javid, S. Riaz, and Z. Atal, "IoT Based Smart Greenhouse Framework and Control Strategies for Sustainable Agriculture," 2022, *Institute of Electrical and Electronics Engineers Inc.* doi: 10.1109/ACCESS.2022.3204066.

[17] H. Liu and X. Meng, "Explainable Ensemble Learning Model for Residual Strength Forecasting of Defective Pipelines," *Applied Sciences (Switzerland)*, vol. 15, no. 7, Apr. 2025, doi: 10.3390/app15074031.

[18] A. F. Ibrahim and M. A. Hussein, "Leveraging machine learning for prediction and optimization of texture properties of sustainable activated carbon derived from waste materials," *Sci Rep*, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-95061-3.

[19] Y. Song *et al.*, "High-performance prediction of soil organic carbon using automatic hyperparameter optimization method in the yellow river delta of China," *Comput Electron Agric*, vol. 236, p. 110490, 2025, doi: https://doi.org/10.1016/j.compag.2025.110490.

[20] X. Tang, M. Petković, G.-K. Chappa, L. Vieira, and N. Dropka, "Comparative analysis of machine learning approaches for predicting and interpreting Cz-sapphire growth," *J Cryst Growth*, vol. 664, p. 128185, 2025, doi: https://doi.org/10.1016/j.jcrysgro.2025.128185.

[21] Q. Y. Chen, S. M. Yin, M. M. Shao, F. S. Yi, and H. Z. Shi, "Machine learning-based Diagnostic model for determining the etiology of pleural effusion using Age, ADA and LDH," *Respir Res*, vol. 26, no. 1, Dec. 2025, doi: 10.1186/s12931-025-03253-2.

[22] M. Vakili, M. Ghamsari, and M. Rezaei, "Performance Analysis and Comparison of Machine and Deep Learning Algorithms for IoT Data Classification," 2020.

[23] Y. Yustikasari, H. Mubarok, and R. Rianto, “Comparative Analysis Performance of K-Nearest Neighbor Algorithm and Adaptive Boosting on the Prediction of Non-Cash Food Aid Recipients,” *Scientific Journal of Informatics*, vol. 9, no. 2, pp. 205–217, 2022, doi: 10.15294/sji.v9i2.32369.
[24] G. Airlangga, “Robust Fan Actuator Prediction in Smart Greenhouses Using Machine Learning: A Comparative Analysis of Ensemble and Linear Models,” *Journal of Information System Research (JOSH)*, vol. 6, no. 1, pp. 566–574, Oct. 2024, doi: 10.47065/josh.v6i1.6095.
[25] L. Syed, “Smart Agriculture using Ensemble Machine Learning Techniques in IoT Environment,” *Procedia Comput Sci*, vol. 235, pp. 2269–2278, 2024, doi: https://doi.org/10.1016/j.procs.2024.04.215.