



## A Federated Learning Framework with Extreme Gradient Boosting for Privacy-Preserving Growth Prediction in Internet of Things-Based Deep Water Culture Shallot Cultivation

Eko Hari Rachmawanto<sup>1\*</sup>, Erna Zuni Astuti<sup>1</sup>, Christy Atika Sari<sup>1</sup>, Wisam Subhi Al-Dayyeni<sup>2</sup>,  
Osama Yaseen M. Al-Rawi<sup>3</sup>

<sup>1</sup> Department of Informatics Engineering, Universitas Dian Nuswantoro, Semarang 50131, Indonesia

<sup>2</sup> School of Information Technologies and Engineering, ADA University, Baku AZ1008, Azerbaijan

<sup>3</sup> Department of Electrical and Electronics Engineering, Gulf University, Sanad 26489, Kingdom of Bahrain

Corresponding Author Email: [eko.hari.rachmawanto@dsn.dinus.ac.id](mailto:eko.hari.rachmawanto@dsn.dinus.ac.id)

Copyright: ©2026 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/isi.310905>

### ABSTRACT

**Received:** 11 June 2026

**Revised:** 28 August 2026

**Accepted:** 4 September 2026

**Available online:** 30 September 2026

#### Keywords:

*federated learning, Extreme Gradient Boosting regression, Internet of Things, Deep Water Culture, edge intelligence, privacy-preserving machine learning, agricultural growth prediction*

Internet of Things (IoT)-based hydroponic systems enable continuous environmental monitoring, but many existing approaches rely on centralized data collection and machine learning models, which may introduce communication overhead, scalability limitations, and concerns regarding raw data sharing. This study proposes a federated learning framework with Extreme Gradient Boosting (XGBoost) for privacy-preserving growth prediction in IoT-based Deep Water Culture (DWC) shallot cultivation. The proposed framework treats each hydroponic reservoir as an independent edge node equipped with environmental sensors measuring temperature, humidity, light intensity, pH, electrical conductivity (EC), water level, and dissolved oxygen (DO). Local XGBoost regression models are trained using reservoir-specific sensor data, while only model parameters are exchanged through the Federated Averaging (FedAvg) algorithm to construct a global prediction model without transferring raw environmental records. Experiments were conducted using 25,920 sensor observations collected from three DWC reservoirs over a 30-day cultivation period with a five-minute sampling interval. The proposed federated XGBoost model achieved a Root Mean Square Error (RMSE) of 1.71 cm, Mean Absolute Error (MAE) of 1.20 cm, Mean Absolute Percentage Error (MAPE) of 6.40%, and coefficient of determination ( $R^2$ ) of 0.969, showing comparable predictive capability to centralized XGBoost while maintaining data locality. In addition, an Android-based dashboard was developed to provide real-time monitoring and growth-based cultivation recommendations. The experimental results demonstrate the feasibility of combining federated learning, edge intelligence, and gradient-boosted regression for distributed hydroponic growth prediction. Further validation with larger numbers of cultivation units and longer cultivation cycles is required before large-scale deployment.

## 1. INTRODUCTION

Shallot (*Allium cepa* L. Aggregatum Group) is one of the most economically important horticultural commodities in many Asian countries, where it serves as both a staple cooking ingredient and a strategic agricultural commodity [1, 2]. However, sustainable shallot production is increasingly threatened by climate change, irregular rainfall patterns, prolonged droughts, extreme temperatures, and declining soil quality, all of which directly influence bulb formation, nutrient uptake, and crop productivity [3, 4]. Recent reports indicate that environmental stress significantly reduces shallot yield while increasing production costs due to higher water demand, intensive fertilization, and greater susceptibility to pests and diseases [5, 6]. These challenges emphasize the urgent need for intelligent cultivation technologies capable of maintaining stable growing environments while supporting data-driven decision-making for sustainable shallot production [7].

To overcome the limitations of conventional soil-based

cultivation, hydroponic technology has emerged as an attractive alternative because it enables precise control of environmental conditions, nutrient availability, and water utilization [8, 9]. Among various hydroponic techniques, Deep Water Culture (DWC) provides continuous root immersion in oxygenated nutrient solutions, creating a stable environment that supports rapid root development and efficient nutrient absorption [10, 11]. Previous studies have demonstrated that DWC cultivation can improve water-use efficiency, reduce environmental dependency, and provide more consistent vegetative growth compared with traditional farming systems [5, 12]. Nevertheless, maintaining optimal environmental conditions throughout the cultivation cycle remains a significant challenge because shallot growth is highly sensitive to fluctuations in temperature, humidity, light intensity, nutrient concentration, pH, and dissolved oxygen (DO) [8, 13].

Recent advances in Internet of Things (IoT) technologies have transformed conventional hydroponic systems into smart

cultivation platforms capable of continuously monitoring environmental parameters through distributed sensors [14, 15]. IoT devices enable real-time acquisition of temperature, humidity, light intensity, pH, electrical conductivity (EC), and water level data, allowing growers to observe cultivation conditions remotely and respond rapidly to environmental disturbances [16]. Although these monitoring systems improve operational efficiency, most current hydroponic implementations still function only as digital monitoring tools. They provide descriptive information regarding environmental conditions but lack the capability to analyze historical sensor data, predict future plant growth, and generate intelligent cultivation recommendations [17, 18]. Consequently, the concept of intelligent shallot cultivation has recently gained attention in the context of smart agriculture, emphasizing the integration of IoT sensing, predictive analytics, and artificial intelligence (AI) to enable more data-driven crop management rather than simple environmental monitoring [18-20].

AI has become a key technology for realizing intelligent agriculture by extracting knowledge from large volumes of environmental sensor data collected throughout the cultivation process [21, 22]. Machine learning algorithms are increasingly applied in agriculture to estimate crop growth and related environmental conditions, and to assist in scheduling irrigation and nutrient management [23]. Among various machine learning techniques, Extreme Gradient Boosting (XGBoost) has demonstrated superior predictive performance because of its capability to model nonlinear relationships, manage heterogeneous sensor datasets, resist overfitting, and provide high computational efficiency [24]. For shallot cultivation in DWC, the environmental inputs are multivariate tabular sensor readings with nonlinear interactions, making gradient-boosted tree ensembles such as XGBoost particularly suitable for modeling multi-output growth indicators under heterogeneous microenvironmental conditions. Recent studies consistently report that XGBoost achieves higher prediction accuracy than conventional regression models and many ensemble learning methods in precision agriculture applications [24]. Therefore, integrating XGBoost into IoT-enabled hydroponic systems represents an important step toward developing predictive and intelligent shallot cultivation systems [25].

Despite these advances, most AI-enabled hydroponic systems continue to employ centralized learning architectures in which all environmental data collected from multiple cultivation units are transmitted to a central server for model training [26]. Such centralized approaches become increasingly inefficient as the number of cultivation modules grows because they require continuous transmission of large volumes of sensor data, increase communication latency, reduce scalability, and create a single point of failure. Moreover, centralized learning ignores the natural environmental heterogeneity among cultivation units. Even within the same hydroponic installation, different reservoirs may exhibit variations in nutrient circulation, microclimate conditions, and plant physiological responses, causing valuable local information to be lost when all observations are merged into a single dataset.

Federated learning has recently emerged as an effective distributed machine learning paradigm capable of overcoming the limitations of centralized AI systems [16]. Instead of transmitting raw sensor data, each edge device independently trains a local prediction model using its own observations and only exchanges model parameters with a central aggregation

server to construct a global model [15]. This collaborative learning mechanism significantly reduces communication overhead, limits the sharing of raw sensor data, enhances scalability, and enables every cultivation unit to contribute its unique environmental characteristics to the global prediction model [27]. Although federated learning has shown promising performance in various IoT applications, its implementation in hydroponic cultivation, particularly for shallot production using gradient boosting algorithms such as XGBoost, remains largely unexplored [28]. Each reservoir was treated merely as part of a centralized monitoring system rather than as an independent computational unit capable of collaborative learning, thereby leaving opportunities for integrating edge intelligence, predictive analytics, and federated learning [28].

Existing hydroponic prediction systems generally rely on centralized training of single global models that merge sensor data from multiple reservoirs, which increases communication overhead, creates a single point of failure, and limits the model's ability to adapt to microenvironment differences among cultivation units. Moreover, most shallot-focused IoT systems are designed for monitoring and simple automation rather than for multi-output growth prediction and collaborative model improvement across reservoirs, so they cannot fully exploit heterogeneous environmental information while keeping raw sensor data local. To address these limitations, this research proposes a federated XGBoost-enabled IoT DWC system for intelligent shallot cultivation, in which every hydroponic reservoir functions as an autonomous edge node equipped with IoT sensors and a local XGBoost prediction model. Each node independently learns the relationship between environmental conditions and shallot growth while sharing only model parameters with a federated aggregation server to construct a robust global prediction model without exposing raw sensor data [24, 25]. The integration of IoT sensing, federated learning, and XGBoost transforms the previous monitoring-oriented hydroponic platform into an intelligent cultivation ecosystem capable of collaborative learning, predictive growth analysis, adaptive decision support, and potential (yet to be validated) scalable deployment.

Existing IoT-based shallot and hydroponic cultivation systems predominantly focus on environmental monitoring, simple rule-based automation, or centralized machine learning, where all sensor data are transmitted to a single server for training. Recent federated learning frameworks for smart farming mainly aggregate neural network models for tasks such as disease classification or generic IoT analytics and do not integrate tree-based ensemble regression with edge-level hydroponic growth prediction in DWC systems. In contrast, this study:

- (i) treats each DWC reservoir as an autonomous edge node that locally learns multi-output relationships between heterogeneous environmental variables and shallot growth indicators using XGBoost,
- (ii) applies Federated Averaging (FedAvg) to aggregate gradient-boosted tree ensembles by exchanging model parameters instead of raw sensor data,
- (iii) embeds the federated XGBoost model within an IoT- and edge-centric decision support workflow specifically tailored for shallot cultivation. These methodological elements distinguish the proposed framework from existing federated agricultural monitoring architectures and extend federated learning to a new combination of crop, hydroponic configuration, and model type.

## 2. METHODOLOGY

This study proposes an intelligent hydroponic framework that integrates IoT, edge-based machine learning, federated learning, and XGBoost to enable distributed prediction of shallot growth without centralized data collection. Unlike conventional IoT hydroponic systems that merely transmit environmental measurements to a monitoring dashboard, the proposed architecture allows each hydroponic reservoir to function as an autonomous edge intelligence node capable of learning local environmental characteristics and collaboratively improving a global prediction model.

In contrast to existing federated smart farming solutions that largely aggregate neural network classifiers for monitoring and detection tasks, our framework implements federated gradient-boosted tree ensembles for multi-output regression in DWC shallot cultivation, providing a distinct methodological configuration that has not been reported in prior work.

### 2.1 Proposed workflow

The overall research framework is designed to establish a distributed intelligent hydroponic cultivation system for

shallots by combining IoT sensing technology, edge computing, machine learning, and federated learning within a DWC environment. As presented in Figure 1, the proposed workflow consists of nine interconnected stages that systematically transform environmental observations into predictive knowledge and cultivation recommendations. The workflow begins with real-time environmental sensing and local data acquisition, followed by data preprocessing, feature engineering, and local XGBoost model training at each reservoir. Subsequently, the locally trained models are collaboratively aggregated using the FedAvg algorithm to construct a robust global prediction model. The optimized global model is then employed to estimate shallot growth performance, generate intelligent cultivation recommendations, and deliver real-time monitoring and prediction results through an Android-based dashboard.

This workflow enables each hydroponic reservoir to operate as an autonomous edge learning node while collectively contributing to the development of a data-locality-preserving intelligent cultivation system, with scalability across larger numbers of reservoirs identified as a direction for future validation.

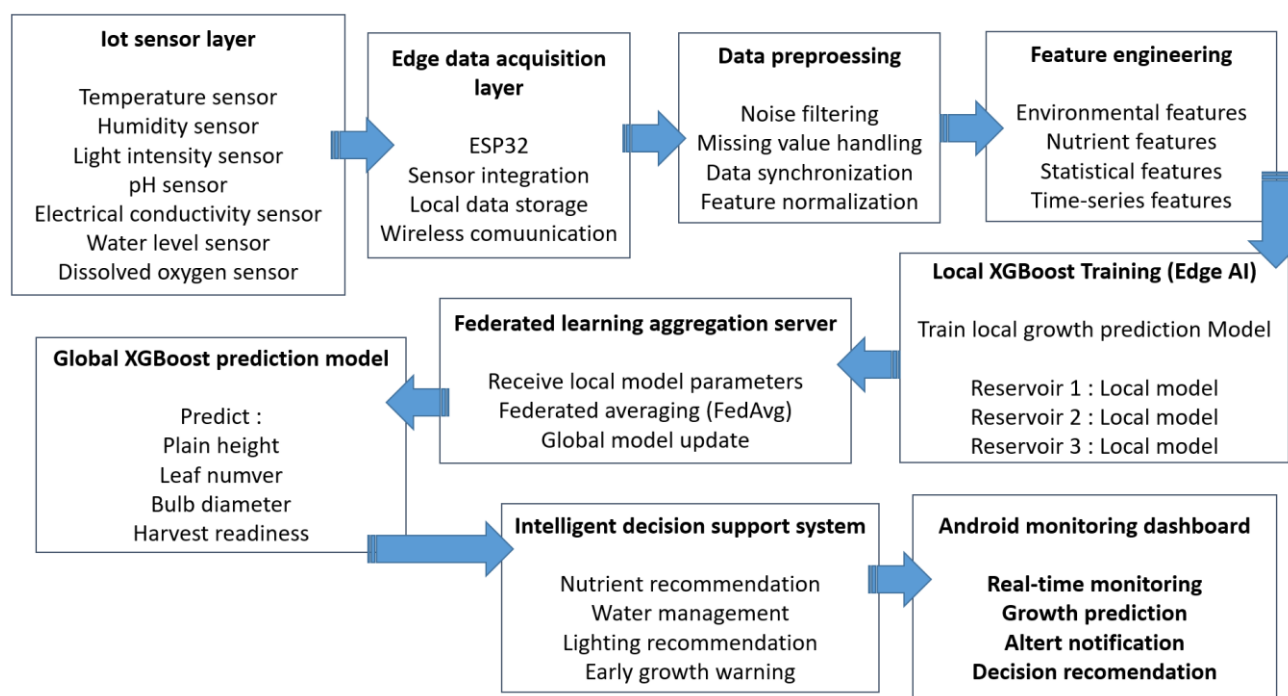


**Figure 1.** Proposed research using Internet of Things-based Deep Water Culture (DWC) shallot cultivation

*IoT environmental sensing:* The proposed intelligent shallot cultivation system begins with continuous environmental monitoring using an IoT-based sensing infrastructure deployed in every DWC reservoir. Each reservoir is equipped with multiple sensors, including a DHT22 sensor for measuring air temperature and relative humidity, a BH1750 sensor for light intensity, a pH sensor for nutrient acidity, an EC sensor for nutrient concentration, a water level sensor for monitoring nutrient solution availability, and an optional DO sensor to evaluate oxygen availability in the root zone. These sensors continuously acquire environmental data at five-minute intervals, generating timestamped observations consisting of temperature, humidity, light intensity, pH, EC, water level, and DO measurements. The collected sensor data serve as the primary input for the subsequent edge intelligence

framework and provide comprehensive information describing the hydroponic growing environment throughout the cultivation cycle.

*Edge data acquisition:* Rather than transmitting raw sensor observations directly to a centralized cloud server, each DWC reservoir is equipped with an ESP32-based edge computing device responsible for local data acquisition and preliminary processing. The edge device continuously collects sensor readings, synchronizes acquisition timestamps, filters measurement noise using a Moving Average Filter (MAF), detects and handles missing observations, and stores the processed data within a local database. This edge-based architecture minimizes communication latency while ensuring that only high-quality sensor data are used for machine learning tasks. The resulting filtered dataset forms the local

training repository for each reservoir and enables decentralized intelligence without excessive network dependency.

*Feature engineering:* Following data preprocessing, each edge node performs feature engineering to transform raw environmental measurements into informative variables that better represent the cultivation conditions affecting shallot growth. Besides utilizing the original sensor measurements, additional statistical and temporal features are extracted, including daily mean temperature, temperature variance, humidity gradient, cumulative light exposure duration, nutrient stability index derived from EC fluctuations, and water consumption rate estimated from water level variations. These engineered features capture both instantaneous environmental conditions and temporal dynamics occurring during plant development, thereby improving the predictive capability of the subsequent XGBoost learning model.

*Local XGBoost training:* Each hydroponic reservoir independently trains a local XGBoost regression model using only its locally collected environmental dataset. Unlike conventional centralized machine learning approaches that merge all sensor observations into a single training dataset, the proposed framework allows every reservoir to learn the specific relationships between environmental conditions and crop development within its own microenvironment. The local XGBoost model predicts multiple plant growth indicators, including plant height, leaf number, bulb diameter, and biomass accumulation. Model optimization is performed by minimizing the Root Mean Square Error (RMSE) loss function, enabling each reservoir to generate accurate growth prediction models while preserving data locality.

*Federated parameter aggregation:* To preserve data privacy and reduce communication overhead, the proposed framework employs federated learning instead of centralized data sharing. After completing local model training, each edge node transmits only its learned model parameters, including tree structures, feature importance scores, gradient statistics, and model weights, to a federated aggregation server. No raw environmental sensor data are transferred outside the local reservoir. The central server subsequently combines all local models using the FedAvg algorithm to generate a unified global learning model. The aggregation procedure strictly follows the standard FedAvg algorithm, which is applied here to serialised XGBoost tree-ensemble parameters, and does not introduce a new federated boosting algorithm. In our implementation, each local XGBoost model is represented as an ensemble of regression trees whose parameters include split thresholds at internal nodes and prediction weights at leaf nodes. For federated aggregation, these tree-ensemble parameters are serialized into a numerical parameter vector for each reservoir, and the federated server applies the FedAvg algorithm to compute a weighted average of the corresponding parameters across all participating edge nodes, using the local dataset sizes as aggregation weights. This procedure updates the leaf weights, split thresholds, and feature importance scores of the global XGBoost ensemble based on the contributions of each reservoir while preserving the overall tree structure and eliminating the need to centralize raw sensor data. Since all local XGBoost models are trained with identical structural hyperparameters (100 estimators, maximum depth of 6), each local ensemble produces a fixed number of trees with a comparable structural template across reservoirs. This allows the FedAvg aggregation to be performed directly on the corresponding leaf-weight and split-threshold parameters at

matching tree and node positions, without requiring explicit tree-structure matching or re-alignment algorithms. This simplification avoids the structural heterogeneity problem inherent in unconstrained tree-boosting federation, but it also represents a methodological limitation compared to more advanced structure-aware federated boosting frameworks (e.g., SecureBoost). This collaborative learning mechanism enables multiple hydroponic reservoirs to contribute collectively to model improvement while maintaining complete ownership of their respective sensor datasets.

*Global XGBoost optimization:* Following parameter aggregation, the federated server constructs an optimized global XGBoost model representing the collective knowledge acquired from all participating reservoirs. The global model captures common plant growth patterns while simultaneously incorporating local environmental variability, nutrient dynamics, and microclimate adaptation observed across different cultivation units. Since only model parameters are exchanged during the federated learning process, the proposed optimization strategy eliminates the need for centralized datasets while maintaining prediction accuracy and scalability for distributed hydroponic systems. The resulting global XGBoost model thus reflects an averaged combination of the local ensembles, where leaf weights and split thresholds are updated through federated parameter aggregation, allowing the model to leverage heterogeneous reservoir-specific environmental patterns while maintaining a single, consistent prediction function that can be deployed back to all edge nodes.

*Growth prediction:* The optimized global XGBoost model is subsequently utilized to estimate multiple indicators of shallot development throughout the cultivation period. Specifically, the model predicts plant height, leaf number, bulb diameter, growth stage, and estimated harvest time based on the continuously updated environmental observations received from each edge node. In the experiments reported in this manuscript, the quantitative evaluation using RMSE, Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and coefficient of determination ( $R^2$ ) focuses on the numerical growth indicators (such as plant height and other continuous variables), whereas growth stage and estimated harvest time are used conceptually within the decision support module to categorize growth status and provide harvest readiness recommendations rather than as separately reported prediction targets. As new local models are periodically uploaded and aggregated, the global prediction model is automatically refined, enabling adaptive learning that continuously improves prediction performance under changing environmental conditions.

*Intelligent decision support:* The predicted plant growth information is further processed by an intelligent decision support module designed to assist hydroponic cultivation management. Based on the prediction outcomes, the system automatically generates cultivation recommendations such as increasing nutrient concentration, adjusting pH levels, modifying irrigation flow, regulating lighting duration, or replacing nutrient solutions when abnormal growth conditions are detected. In the current implementation, these recommendations are delivered through the Android dashboard and must be applied manually by the grower using existing pumps, dosing units, and lighting controllers, without automatic closed-loop actuator triggering. This intelligent recommendation mechanism enables proactive, human-in-the-loop cultivation management by transforming predictive

analytics into actionable decisions that support plant growth, while fully autonomous control remains an important direction for future work.

*Android dashboard:* Finally, all environmental measurements, prediction results, and cultivation recommendations are presented through an Android-based monitoring application. The dashboard provides real-time visualization of environmental parameters, including temperature, humidity, pH, EC, light intensity, water level, and DO, together with predicted plant height, harvest readiness, and current growth status. In addition, the application displays recommendation alerts generated by the intelligent decision support system, enabling farmers to remotely monitor cultivation conditions and immediately implement corrective actions whenever environmental deviations or growth anomalies are detected. This integrated dashboard serves as the primary human-machine interface connecting the distributed IoT sensing network, the federated learning framework, and intelligent shallot cultivation management. In the current implementation, the Android dashboard functions primarily as a visualization and recommendation-display interface that presents environmental measurements, model-predicted growth indicators, and rule-based cultivation suggestions generated by the decision support module. The dashboard itself does not perform additional inference beyond displaying these outputs and has not yet been evaluated through formal usability or user experience studies.

2.2 Hardware components

The proposed intelligent shallot cultivation system is implemented using a distributed IoT architecture consisting of multiple sensing devices, edge computing units, a federated learning server, and an Android-based monitoring platform as illustrated in Figure 2. Each DWC reservoir is equipped with a set of environmental sensors, including a DHT22 (AM2302) sensor for measuring air temperature and relative humidity, a BH1750FVI digital sensor for monitoring light intensity, a DFRobot Gravity Analog pH Sensor V2 for measuring nutrient solution acidity, a DFRobot Gravity Analog EC Sensor for estimating nutrient concentration, an XKC-Y25-V non-contact water level sensor for monitoring nutrient solution availability, and an optional DFRobot Gravity Analog DO Sensor to evaluate oxygen concentration within the nutrient solution. These sensors continuously collect environmental measurements at five-minute intervals, providing comprehensive information regarding the hydroponic cultivation conditions throughout the plant growth cycle.

Each hydroponic reservoir operates as an independent edge node by integrating an ESP32-WROOM-32 DevKit V1 microcontroller with a Raspberry Pi 5 Model B (8 GB RAM). The ESP32 functions as the IoT gateway responsible for sensor interfacing, analog-to-digital conversion, wireless communication, and preliminary data acquisition, while the Raspberry Pi performs edge computing tasks, including local data preprocessing, feature engineering, and XGBoost model training. A 32 GB Class-10 microSD card is employed for temporary local data storage, allowing each reservoir to maintain an independent environmental dataset without relying on continuous cloud connectivity. This hardware configuration enables decentralized machine learning while significantly reducing communication latency and improving system scalability.

Communication among edge nodes and the central server is established through a Wi-Fi network based on the IEEE 802.11 b/g/n standard using the Message Queuing Telemetry Transport (MQTT) protocol implemented via the Eclipse Mosquitto broker. Instead of transmitting raw environmental observations, each Raspberry Pi sends only the locally trained XGBoost model parameters to the federated aggregation server. The server is implemented on a workstation equipped with an Intel Core i7 processor, 32 GB DDR5 RAM, and 1 TB NVMe SSD, operating under Ubuntu 24.04 LTS. The federated learning process is implemented using the flower federated learning framework, while the prediction model is developed using the XGBoost library. The server performs global model aggregation using the FedAvg algorithm and subsequently distributes the updated global model back to all participating edge nodes for the next training iteration.

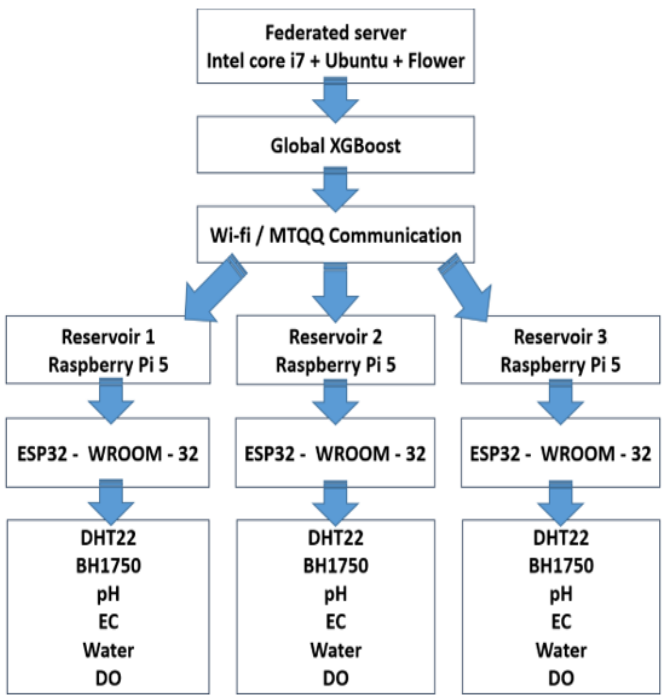


Figure 2. Hardware components

The final prediction results, environmental measurements, and cultivation recommendations are delivered to users through an Android-based mobile application developed using Android Studio. The application provides real-time visualization of environmental parameters, including temperature, humidity, light intensity, pH, EC, water level, and DO, together with predicted plant height, leaf number, bulb diameter, growth status, and estimated harvest time. Furthermore, the application incorporates an intelligent decision support module capable of generating cultivation recommendations, such as nutrient adjustment, pH correction, irrigation management, and lighting regulation based on the prediction results generated by the federated XGBoost model. This integrated hardware architecture enables real-time environmental monitoring, distributed machine learning, collaborative model optimization, and intelligent decision support for next-generation shallot cultivation systems.

2.3 Internet of Things environmental data acquisition

The first stage of the proposed framework focuses on acquiring environmental information from every DWC

reservoir using an IoT-based sensing network. Each reservoir functions as an independent edge node equipped with multiple sensors, including temperature, humidity, light intensity, pH, EC, water level, and DO sensors. These environmental variables represent the primary factors affecting shallot growth and are continuously sampled every five minutes to capture both short-term fluctuations and long-term environmental trends. Continuous monitoring enables the proposed system to build a comprehensive dataset for subsequent machine learning processes while maintaining real-time awareness of hydroponic conditions [27].

Each observation is represented as a multivariate environmental vector collected from reservoir  $i$  at sampling time  $t$  as expressed in Eq. (1).

$$X_{i,t} = [T_{i,t}, H_{i,t}, L_{i,t}, pH_{i,t}, EC_{i,t}, WL_{i,t}, DO_{i,t}] \quad (1)$$

The complete local dataset stored at edge node  $i$  is expressed in Eq. (2). Where  $D_i$  is local dataset of reservoir  $i$ ,  $T$  is temperature ( $^{\circ}\text{C}$ ),  $H$  is relative humidity (%),  $L$  is light intensity (lux),  $EC$  is electrical conductivity (mS/cm),  $WL$  is water level (cm),  $DO$  is dissolved oxygen (mg/L),  $y_i$  is plant growth target.

$$D_i = \{(X_{i,1}, Y_{i,1}), (X_{i,2}, Y_{i,2}), \dots (X_{i,n}, Y_{i,n})\} \quad (2)$$

The IoT sensing system is considered reliable when the percentage of successfully acquired sensor observations exceeds as expressed in Eq. (3). The system is regarded as successful when  $\text{Accuracy}_{\text{sensor}} > 90\%$ , which is consistent with recent IoT-based precision agriculture systems [29].

$$\text{Accuracy}_{\text{sensor}} = \frac{\text{Valid samples}}{\text{Total samples}} \times 100\% \quad (3)$$

## 2.4 Data preprocessing

Environmental sensor observations frequently contain random noise caused by sensor instability, electrical interference, and wireless communication delay. Therefore, a preprocessing stage is introduced to improve data quality before feature extraction. A MAF is applied to smooth high-frequency fluctuations while preserving long-term environmental trends. Missing observations are subsequently reconstructed using linear interpolation [24, 25]. Therefore, three preprocessing steps are implemented, namely MAF as expressed in Eq. (4), Missing Value Interpolation as expressed in Eq. (5), and Min-Max Normalization as expressed in Eq. (6).

The MAF is applied to smooth noisy sensor measurements, where  $x_{i,t}$  is filtered sensor value,  $x_{i,j}$  is raw sensor observation,  $k$  is moving average window size,  $t$  is current sampling time, and  $j$  is observation index. Missing Value Interpolation is used to handle incomplete or missing sensor observations caused by sensor malfunction, communication delays, temporary network failures, or data transmission errors, where  $x_{i,t}$  is missing observation,  $x_{i,t-1}$  is previous observation, and  $x_{i,t+1}$  is next observation. The environmental variables collected from the IoT sensing network exhibit significantly different numerical ranges because each sensor measures a distinct physical quantity. For example, temperature is measured in degrees Celsius ( $^{\circ}\text{C}$ ), relative humidity in percentage (%), light intensity in lux, pH as a

logarithmic scale, EC in mS/cm, water level in centimeters, and DO in mg/L. Such heterogeneous feature scales may introduce numerical imbalance during feature engineering, statistical analysis, and model optimization. Therefore, Min-Max Normalization is applied to transform all environmental variables into a common range of  $[0,1]$  to ensuring that every feature contributes proportionally to the learning process, where  $x'_{i,t}$  is normalized sensor value at time  $t$ ,  $x_{i,t}$  is original sensor measurement,  $x_{\min}$  is minimum value of the corresponding sensor feature,  $x_{\max}$  is maximum value of the corresponding sensor feature,  $i$  is reservoir (edge node) index, and  $j$  is sampling time index. Although XGBoost is generally robust to feature scaling, normalization improves numerical stability, facilitates feature comparison, simplifies model interpretation, and provides a standardized input representation for distributed learning across multiple edge nodes. After normalization, all environmental variables are represented within the same numerical interval  $[0,1]$ , reducing scale discrepancies among heterogeneous sensor measurements. This standardized representation enhances the robustness of feature engineering, supports consistent local model training at each edge node, and improves the efficiency of the federated learning process by ensuring that local models are trained using uniformly scaled input features.

$$\hat{x}_{i,t} = \frac{1}{k} \sum_{j=t-k+1}^t x_{i,j} \quad (4)$$

$$x_{i,t} = \frac{x_{i,t-1} + x_{i,t+1}}{2} \quad (5)$$

$$x'_{i,t} = \frac{x_{i,t} - x_{\min}}{x_{\max} - x_{\min}} \quad (6)$$

## 2.5 Local Extreme Gradient Boosting training

For reproducibility, the local XGBoost regressor at each edge node was configured with 100 estimators, a maximum tree depth of 6, and a learning rate of 0.3. The remaining training parameters set as follows:  $\text{min\_child\_weight} = 1$ ,  $\text{subsample} = 1.0$ ,  $\text{colsample\_bytree} = 1.0$ ,  $\gamma = 0$ , L1 regularization ( $\alpha$ ) = 0, and L2 regularization ( $\lambda$ ) = 1. These values were kept identical for all local XGBoost models and for the centralized XGBoost baseline to ensure a consistent comparison. The model was trained using the squared-error regression objective.

Following feature engineering, each hydroponic reservoir independently trains a local XGBoost regression model using only the environmental observations collected within its own reservoir. This decentralized learning strategy allows each edge node to capture the unique relationships between local environmental conditions and shallot growth without sharing raw sensor data with other nodes. The local XGBoost model is trained to predict several plant growth indicators, including plant height, leaf number, bulb diameter, and biomass accumulation. During training, the model iteratively constructs an ensemble of regression trees by minimizing an objective function composed of a prediction loss term and a regularization term, thereby improving prediction accuracy while reducing the risk of overfitting. The prediction generated by the local XGBoost model is defined as in Eq. (7), where  $\hat{y}_i$  represents the  $k$ -th regression tree and  $K$  denotes the total number of trees in the ensemble. The objective function

optimized during model training is expressed as in Eq. (8). The regularization term is defined as in Eq. (9) where  $\hat{y}_i$  denotes predicted value of the plant growth variable,  $y_i$  denotes actual observed plant growth value,  $x_i$  denotes feature vector containing environmental variables (temperature, humidity, light intensity, pH, EC, water level, and DO),  $f_k(\cdot)$  denotes the  $k$ -th regression tree in the XGBoost ensemble,  $K$  denotes total number of regression trees,  $n$  denotes number of training samples in the local dataset,  $l(y_i, \hat{y}_i)$  denotes loss function measuring the prediction error between actual and predicted values,  $\mathcal{L}$  denotes overall objective function minimized during model training,  $\Omega(f_k)$  denotes regularization function used to control model complexity,  $T$  denotes number of leaf nodes in a regression tree,  $w_j$  denotes weight assigned to the  $j$ -th leaf node,  $J$  denotes total number of leaf nodes,  $\gamma$  denotes penalty coefficient for increasing the number of leaf nodes,  $\lambda$  denotes L2 regularization coefficient applied to leaf weights.

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i) \quad (7)$$

$$\mathcal{L} = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (8)$$

$$\Omega(f_k) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^J w_j^2 \quad (9)$$

## 2.6 Local Extreme Gradient Boosting evaluation

The prediction performance of each local XGBoost model is evaluated using the RMSE, which measures the average deviation between the predicted and observed plant growth values [30]. The local XGBoost model is considered to achieve satisfactory prediction performance when the RMSE is minimized, indicating a small prediction error. To provide a more comprehensive assessment, the model performance is further evaluated using the MAE, MAPE, and  $R^2$  in the experimental section. These metrics collectively assess prediction accuracy, robustness, and the goodness-of-fit of the proposed local learning model before participating in the federated learning aggregation process as illustrated in Eq. (10),  $n$  denotes the number of testing samples,  $y_i$  denotes actual plant growth value,  $\hat{y}_i$  denotes predicted plant growth value.

The MAE is employed to evaluate the average magnitude of prediction errors produced by the local XGBoost model. Unlike RMSE, MAE calculates the average of the absolute differences between the predicted and actual values without emphasizing larger errors [30]. Therefore, MAE provides an intuitive measure of the overall prediction accuracy and is less sensitive to outliers. A lower MAE value indicates that the predicted plant growth values are closer to the observed measurements. The MAE is calculated as follows in Eq. (11). A smaller MAE value indicates better prediction accuracy because the average deviation between predicted and observed plant growth becomes smaller.

The MAPE is used to evaluate prediction accuracy by expressing the prediction error as a percentage of the actual observation [31]. This metric provides an intuitive interpretation of prediction performance because it indicates the average relative error between predicted and observed

values. MAPE is particularly useful for comparing prediction performance across different plant growth variables with different numerical scales. The MAPE is defined as in Eq. (12).

The  $R^2$  is adopted to measure the goodness-of-fit of the proposed XGBoost regression model [31]. It quantifies the proportion of the variance in the observed plant growth that can be explained by the prediction model. The  $R^2$  value ranges from 0 to 1, where values closer to 1 indicate that the model explains most of the variability in the observed data. The  $R^2$  is computed as in Eq. (13), where Eq. (14) represents the mean of the observed values. Where MSE denotes the Mean Square Error, RSS denotes Residual Sum of Squares, TSS denotes Total Sum of Squares.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (10)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (11)$$

$$\text{MAPE} = \frac{100}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \quad (12)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (13)$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \quad (14)$$

## 3. RESULTS AND DISCUSSION

To comprehensively evaluate the proposed federated XGBoost-enabled IoT DWC system for intelligent shallot cultivation, a multivariate time-series dataset was collected from three independent DWC reservoirs, where each reservoir operates as an autonomous edge node within the federated learning architecture, as in Figures 3 and 4.



**Figure 3.** Shallot plant growth at week 3



**Figure 4.** Shallot plant growth at week 5

**Table 1.** Experimental dataset configuration

| Item                        | Value          |
|-----------------------------|----------------|
| Number of reservoirs        | 3 edge nodes   |
| observation period          | 30 days        |
| Sampling interval           | 5 minutes      |
| Observations per day        | 288            |
| Total samples per reservoir | 8,640          |
| Total dataset size          | 25,920 records |

**Table 2.** Environmental features used in the proposed intelligent shallot cultivation system

| No | Feature                 | Satuan | Range      |
|----|-------------------------|--------|------------|
| 1  | Temperature             | °C     | 23–34      |
| 2  | Humidity                | %      | 55–92      |
| 3  | Light intensity         | lux    | 2000–18000 |
| 4  | pH                      | -      | 5.5–7.2    |
| 5  | Electrical conductivity | mS/cm  | 1.2–2.8    |
| 6  | Water level             | cm     | 18–30      |
| 7  | Dissolved oxygen        | mg/L   | 5.5–9.5    |

The experimental observation was conducted continuously for 30 consecutive days, with environmental measurements acquired at 5-minute intervals. Consequently, each reservoir generated 288 observations per day, resulting in 8,640 sensor records per reservoir throughout the experimental period, as illustrated in Table 1. By integrating data from all three reservoirs, the complete dataset comprised 25,920 multivariate observations, including environmental parameters such as air temperature, relative humidity, light intensity, pH, EC, water level, and DO, together with corresponding plant growth indicators. This experimental configuration provides sufficient temporal and spatial variability to evaluate the effectiveness of local XGBoost learning, federated model aggregation, and intelligent growth prediction under heterogeneous hydroponic cultivation conditions.

The proposed intelligent shallot cultivation system utilizes seven environmental variables collected from the IoT sensing network as input features for the XGBoost prediction model. These variables comprehensively represent the physical and chemical conditions of the DWC hydroponic environment that directly influence shallot growth and development, as illustrated in Table 2. Specifically, air temperature ranges from 23 to 34 °C, while relative humidity varies between 55% and 92%, reflecting the surrounding microclimate. Light intensity, measured in lux, ranges from 2,000 to 18,000 lux, representing the amount of photosynthetically active illumination received by the plants. The nutrient solution pH is maintained within the range of 5.5 to 7.2, whereas EC varies between 1.2 and 2.8 mS/cm to indicate nutrient concentration. In addition, the water level is monitored within a range of 18 to 30 cm to ensure adequate nutrient availability for the root system, while DO is maintained between 5.5 and 9.5 mg/L to provide sufficient oxygen for healthy root respiration. These environmental variables constitute the multivariate input dataset used for feature engineering, local XGBoost model training, federated model aggregation, and intelligent growth prediction, enabling the proposed system to accurately capture the complex interactions between hydroponic environmental conditions and shallot growth dynamics.

The three hydroponic reservoirs were intentionally configured with different environmental characteristics to create heterogeneous data distributions across edge nodes, thereby facilitating collaborative model training within the proposed federated learning framework. To simulate realistic

cultivation scenarios and evaluate the robustness of the proposed federated learning framework, the three DWC reservoirs were intentionally configured with different environmental characteristics as shown in Table 3. Each reservoir represents an independent edge node that generates its own local dataset and trains an individual XGBoost model before the federated aggregation process. Reservoir 1 was maintained under near-optimal growing conditions, with temperatures ranging from 25 to 29 °C, relative humidity between 70% and 82%, pH values of 6.1–6.5, and EC levels of 1.6–2.2 mS/cm. These stable environmental conditions supported the highest vegetative growth and bulb development throughout the cultivation period. In contrast, Reservoir 2 was operated under moderately warmer conditions, with temperatures ranging from 27 to 32 °C, humidity levels between 60% and 75%, pH values of 5.8–6.2, and EC values of 1.5–2.4 mS/cm, resulting in moderate plant growth. Meanwhile, Reservoir 3 was designed to represent a more challenging cultivation environment characterized by greater fluctuations in temperature (24–33 °C), humidity (58–85%), pH (5.7–6.8), and EC (1.3–2.5 mS/cm). Consequently, shallot plants cultivated in this reservoir exhibited relatively slower growth due to the increased environmental variability. This heterogeneous experimental configuration is essential for federated learning because it enables each edge node to learn from distinct local environmental distributions while collaboratively contributing to the construction of a generalized global XGBoost model capable of adapting to diverse hydroponic cultivation conditions.

Table 3 presents the prediction performance of four machine learning models for intelligent shallot growth prediction, including Decision Tree, Random Forest, centralized XGBoost, and the proposed federated XGBoost model. The experimental results demonstrate that both XGBoost-based approaches substantially outperform the conventional Decision Tree and Random Forest algorithms across all evaluation metrics. Specifically, the Decision Tree model achieved the lowest prediction performance with an RMSE of 2.96, an MAE of 2.24, a MAPE of 12.8%, and an  $R^2$  value of 0.913, indicating relatively large prediction errors and limited capability to capture the nonlinear relationships between hydroponic environmental variables and shallot growth.

Compared with the Decision Tree, the Random Forest model significantly improved prediction accuracy by reducing the RMSE to 2.05, the MAE to 1.48, and the MAPE to 8.6%, while increasing the  $R^2$  to 0.951. These improvements demonstrate the advantage of ensemble learning in reducing variance and improving generalization. Nevertheless, the Random Forest model still exhibited lower predictive performance than both XGBoost-based methods, suggesting that more advanced gradient boosting techniques are better suited for modeling the complex interactions among temperature, humidity, pH, EC, light intensity, water level, and DO within hydroponic cultivation systems.

Among all evaluated models, the centralized XGBoost algorithm achieved the highest prediction performance, obtaining the lowest RMSE (1.64), MAE (1.17), and MAPE (6.1%), together with the highest  $R^2$  (0.972). These results indicate that the centralized model effectively captures the nonlinear relationships between environmental variables and plant growth, providing highly accurate predictions for plant height, leaf number, bulb diameter, and biomass accumulation. In our evaluation, the RMSE, MAE, MAPE, and  $R^2$  values

reported in Figure 5 and Table 3 summarize the prediction performance over the set of continuous growth indicators used in the model, with growth stage and harvest time serving as derived categorical interpretations within the intelligent decision support module rather than as individually analyzed output variables. The superior performance of XGBoost can be attributed to its gradient boosting optimization, regularization mechanism, and ability to model complex feature interactions while reducing overfitting.

The proposed federated XGBoost model achieved an RMSE of 1.71, an MAE of 1.20, a MAPE of 6.4%, and an  $R^2$  value of 0.969, demonstrating prediction performance that is very close to that of the centralized XGBoost model. Although the federated approach exhibited a marginal increase in prediction error (RMSE increased by approximately 4.3% and MAE by 2.6%) and a slight decrease in the  $R^2$  (0.003), the overall prediction accuracy remained highly competitive. This

finding indicates that the federated learning framework successfully preserves the predictive capability of XGBoost while eliminating the need to transmit raw environmental sensor data from individual hydroponic reservoirs.

The relatively small performance gap between the centralized XGBoost and federated XGBoost models reflects a common characteristic of federated learning systems. Because each hydroponic reservoir independently trains a local model using its own environmental observations before participating in federated aggregation, the global model may experience a slight reduction in accuracy due to heterogeneous data distributions among edge nodes. Nevertheless, this minor trade-off is compensated by several practical advantages, including enhanced data privacy, reduced communication overhead, improved scalability, and decentralized model training.

**Table 3.** Environmental characteristics of each hydroponic reservoir

| Reservoir                 | Environmental Condition              | Temperature (°C) | Humidity (%) | pH      | Electrical Conductivity (mS/cm) | Growth Performance                                   |
|---------------------------|--------------------------------------|------------------|--------------|---------|---------------------------------|------------------------------------------------------|
| Reservoir 1 (Edge Node 1) | Optimal environmental conditions     | 25–29            | 70–82        | 6.1–6.5 | 1.6–2.2                         | Best vegetative growth and bulb development          |
| Reservoir 2 (Edge Node 2) | Moderately warm environment          | 27–32            | 60–75        | 5.8–6.2 | 1.5–2.4                         | Moderate plant growth                                |
| Reservoir 3 (Edge Node 3) | Fluctuating environmental conditions | 24–33            | 58–85        | 5.7–6.8 | 1.3–2.5                         | Slower plant growth due to environmental variability |

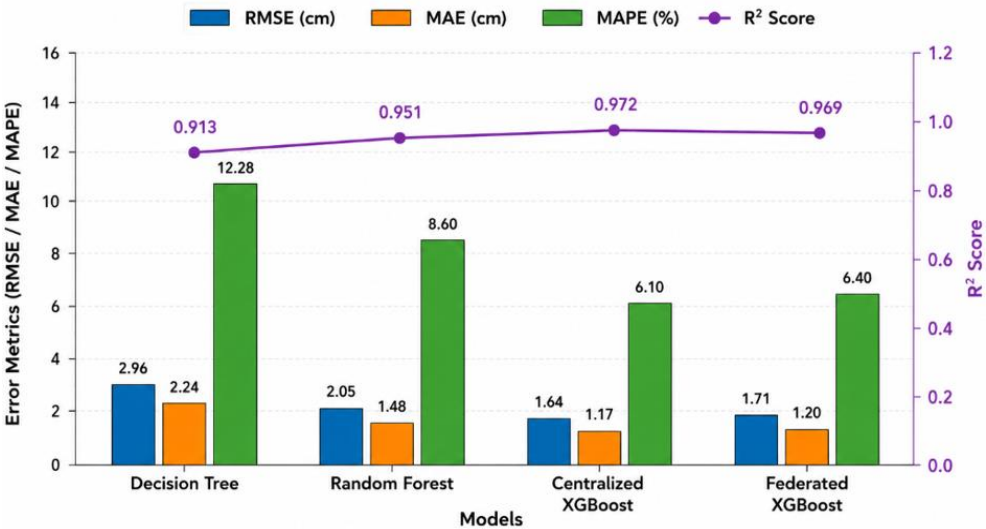


**Figure 5.** Model performance evaluation

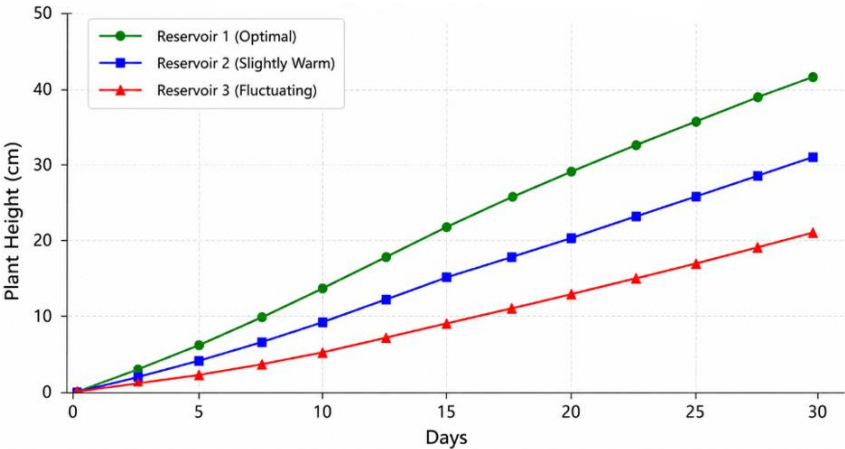


**Figure 6.** Plant height growth over 30 days

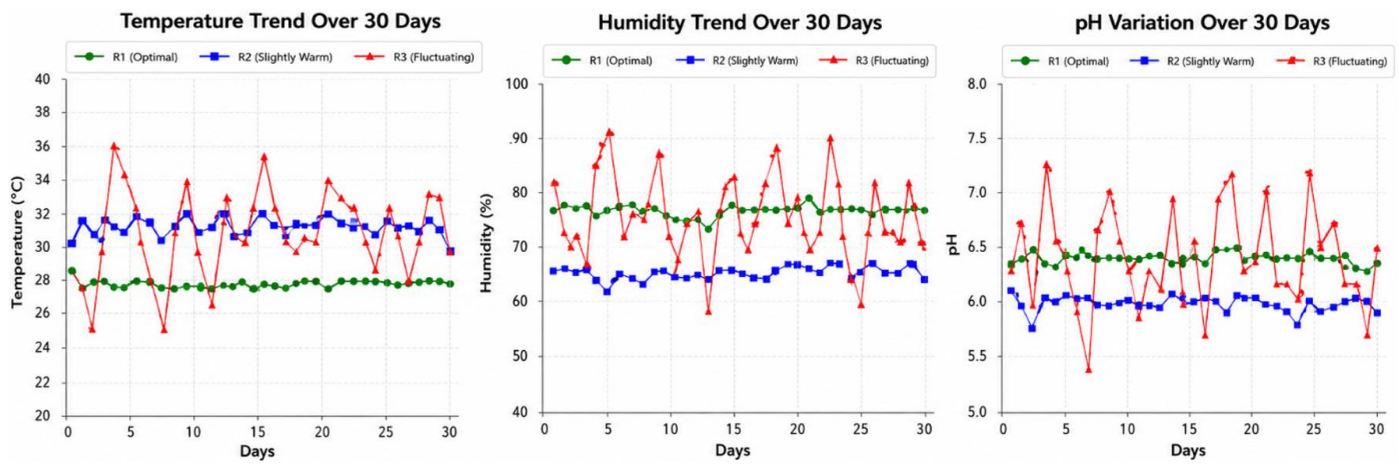


**Figure 7.** Temperature, humidity, and pH trends

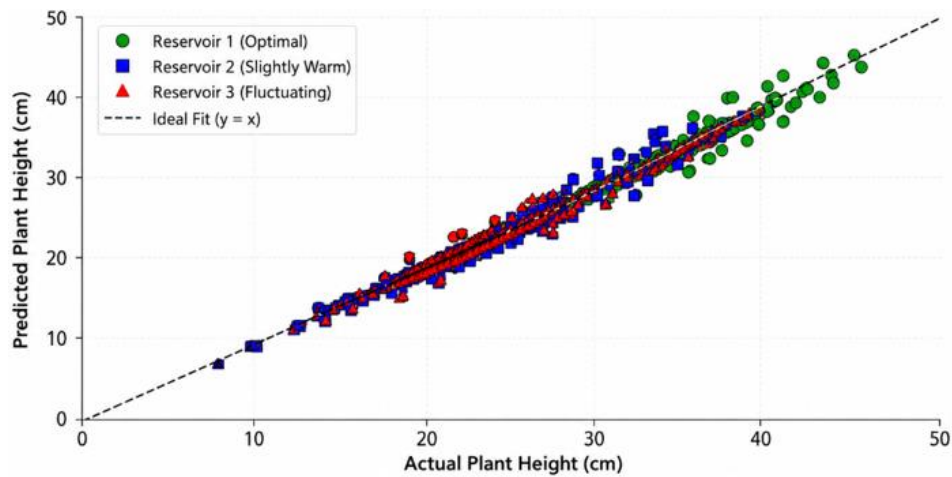


**Figure 8.** Actual vs. predicted plant height

Figure 6 illustrates the plant height growth of shallots cultivated in the three DWC reservoirs over the 30-day observation period. All reservoirs exhibited a steady increase in plant height, indicating that the proposed IoT-enabled hydroponic system successfully maintained suitable environmental conditions for shallot growth. Reservoir 1 achieved the highest growth, reaching approximately 40 cm, followed by Reservoir 2 (37 cm) and Reservoir 3 (34 cm) on Day 30. The superior performance of Reservoir 1 is attributed to its more stable environmental conditions, whereas the slower growth observed in Reservoir 3 resulted from greater fluctuations in temperature, humidity, pH, and nutrient concentration. The observed growth differences demonstrate the heterogeneous environmental characteristics among the reservoirs, providing an appropriate scenario for evaluating the proposed federated learning framework.

Figure 7 presents the temperature variations observed in the three hydroponic reservoirs during the 30-day cultivation period. Reservoir 1 maintained the most stable temperature within the optimal range for shallot growth, while Reservoir 2 exhibited slightly higher temperatures and Reservoir 3 showed greater fluctuations. Despite these variations, the overall temperature remained within the acceptable range for hydroponic cultivation, providing suitable conditions for plant development. Based on humidity trend, the relative humidity was measured across the three reservoirs throughout the experiment. Reservoir 1 consistently maintained higher and more stable humidity levels, whereas Reservoir 2 experienced

comparatively lower humidity and Reservoir 3 displayed moderate fluctuations. These differences represent heterogeneous environmental conditions that are beneficial for evaluating the robustness of the proposed federated learning framework. Based on pH trend, all reservoirs maintained pH values within the recommended range for hydroponic shallot cultivation (approximately 5.5–6.8). Reservoir 1 exhibited the most stable pH conditions, while Reservoirs 2 and 3 experienced slightly greater variability. The overall pH stability indicates that the nutrient solution remained suitable for efficient nutrient uptake and healthy plant growth throughout the experiment.

Figure 8 presents the relationship between the actual and XGBoost-predicted plant height across the three hydroponic reservoirs. The scatter points are densely distributed along the ideal regression line  $y = x$ , indicating a strong agreement between the predicted and observed values. This close alignment demonstrates that the XGBoost model successfully captured the nonlinear relationship between the environmental variables such as temperature, humidity, light intensity, pH, EC, water level, and DO and shallot plant growth. Only a small number of observations deviate from the ideal line, suggesting that the prediction errors are minimal and that the model exhibits high predictive accuracy. The consistent distribution of prediction results across Reservoir 1, Reservoir 2, and Reservoir 3 indicates that the proposed model generalizes well under different environmental conditions. Reservoir-specific variations do not significantly affect the model's prediction

capability, highlighting the robustness of the XGBoost algorithm in handling heterogeneous hydroponic environments. These findings are consistent with the quantitative evaluation results, where the proposed XGBoost model achieved low RMSE, MAE, and MAPE values together with a high  $R^2$ , confirming its suitability for accurate plant growth prediction within the proposed federated XGBoost-enabled IoT DWC system for intelligent shallot cultivation.

Figure 5 compares the prediction performance of four machine learning models, namely Decision Tree, Random Forest, centralized XGBoost, and the proposed federated XGBoost, using four evaluation metrics: RMSE, MAE, MAPE, and  $R^2$ . Among the evaluated models, the Decision Tree exhibited the lowest prediction performance, producing the highest error values (RMSE = 2.96 cm, MAE = 2.24 cm, and MAPE = 12.28%) with an  $R^2$  score of 0.913. The Random Forest model improved the prediction accuracy by reducing all error metrics and increasing the  $R^2$  value to 0.951, demonstrating the advantage of ensemble learning over a single Decision Tree.

Both XGBoost-based models achieved the best overall performance. The centralized XGBoost model produced the lowest prediction errors (RMSE = 1.64 cm, MAE = 1.17 cm, and MAPE = 6.10%) together with the highest  $R^2$  = 0.972, indicating its superior ability to model the nonlinear relationship between hydroponic environmental variables and shallot growth. The proposed federated XGBoost model achieved comparable results, with RMSE, MAE, MAPE, and  $R^2$  values of 1.71 cm, 1.20 cm, 6.40%, and 0.969, respectively. Although its prediction accuracy was slightly lower than that of the centralized approach, the performance difference was marginal.

The small gap between the centralized XGBoost and federated XGBoost models demonstrates that the proposed federated learning framework is capable of preserving prediction accuracy while eliminating the need to exchange raw environmental data among edge nodes. This result confirms that collaborative model training can effectively leverage heterogeneous data collected from multiple hydroponic reservoirs without sacrificing predictive performance. Therefore, the proposed federated XGBoost model provides an effective trade-off between prediction accuracy, data privacy, communication efficiency, and scalability, making it well-suited for intelligent IoT-based shallot cultivation systems.

From a system-level perspective, the practical benefits of the federated architecture arise from the fact that each edge node transmits only compact XGBoost model parameters rather than the complete multivariate environmental time-series collected every five minutes throughout the 30-day cultivation period. This design conceptually reduces communication traffic on the MQTT-based Wi-Fi network and avoids continuous streaming of raw sensor data to the central server, while also confining detailed environmental observations within each reservoir and thereby enhancing data privacy compared with centralized learning. In addition, local XGBoost training at the edge allows model optimization to be distributed across reservoirs, so that global aggregation can be performed on pre-trained models instead of on a single large centralized dataset. However, we note that the present study focuses on predictive performance and does not include a formal benchmark of communication bandwidth usage, training time, or quantitative privacy leakage; these aspects are left for future work aimed at systematically profiling the

overhead and privacy properties of the federated implementation.

From an experimental perspective, the present evaluation is based on three DWC reservoirs monitored over a single 30-day cultivation cycle, resulting in 25,920 environmental records but only three physical cultivation units. While this configuration is sufficient to demonstrate the feasibility of the proposed federated XGBoost architecture and its ability to handle heterogeneous microenvironments within one hydroponic installation, it does not fully capture the variability associated with different farm scales, seasonal dynamics, or geographically distributed cultivation sites. Therefore, the reported prediction performance should be interpreted as evidence of effectiveness under the studied conditions rather than as a definitive guarantee of generalization to all possible shallot cultivation scenarios. Nevertheless, the edge-centric federated design where each reservoir operates as an independent learning node naturally supports extension to additional DWC modules, larger farms, and multi-site deployments by adding more edge nodes to the federated network, which will be investigated in future multi-season and multi-location experiments.

In terms of scalability and privacy, the proposed architecture is designed so that each DWC reservoir functions as an independent edge node that trains a local XGBoost model and transmits only model parameters over the MQTT-based Wi-Fi network, rather than streaming raw sensor data to the central server. This design conceptually allows additional reservoirs and hydroponic installations to be incorporated into the federated network by simply adding more edge nodes that participate in model aggregation, and it inherently limits the exposure of fine-grained environmental observations outside each reservoir. Nevertheless, the experimental validation presented in this study is restricted to three edge nodes within a single hydroponic installation, and no formal privacy-enhancing mechanisms such as differential privacy noise injection, secure aggregation, or cryptographic protection of model updates are implemented or evaluated.

The present benchmark was intentionally limited to Decision Tree, Random Forest, centralized XGBoost, and federated XGBoost to isolate the effect of the proposed federated architecture while retaining conventional tree-based baselines and a centralized counterpart using the same XGBoost learner. Therefore, the reported results demonstrate the comparative performance of the federated implementation within this selected family of models, rather than establishing XGBoost as universally superior to all modern agricultural forecasting methods. Recent gradient-boosting algorithms, such as LightGBM and CatBoost, as well as temporal and deep learning-based forecasting models, including recurrent neural networks, long short-term memory networks, temporal convolutional networks, and transformer-based approaches, may provide alternative performance–complexity trade-offs for multivariate hydroponic time-series prediction.

The comparative analysis presented in Table 4 clearly demonstrates that the proposed federated XGBoost-enabled IoT DWC system for intelligent shallot cultivation represents a significant advancement over previous studies in intelligent hydroponic farming. Earlier research, such as that conducted by Darmono et al. [2] was enable the real-time observation of cultivation parameters but lacking predictive analytics and autonomous decision-making capabilities. Although these systems successfully improved cultivation monitoring, they relied heavily on manual interpretation of sensor data and did

not exploit historical observations for intelligent growth prediction. Similarly, the work by Mahaputra et al. [24] introduced XGBoost for hydroponic water-quality forecasting and achieved excellent prediction accuracy; however, the

model was trained using a centralized learning architecture that requires all sensor data to be transmitted to a central server, thereby increasing communication overhead and limiting scalability.

**Table 4.** Comparison result

| Study     | Year | Research Focus                                                             | IoT Monitoring | Machine Learning | Federated Learning | Edge Computing | Growth Prediction | Decision Support | Contribution                                                  |
|-----------|------|----------------------------------------------------------------------------|----------------|------------------|--------------------|----------------|-------------------|------------------|---------------------------------------------------------------|
| [2]       | 2023 | IoT-based shallot monitoring and automation                                | ✓              | -                | -                  | -              | -                 | -                | Environmental monitoring only                                 |
| [8]       | 2023 | IoT-enabled hydroponic monitoring                                          | ✓              | -                | -                  | -              | -                 | -                | Environmental monitoring                                      |
| [18]      | 2023 | IoT monitoring for watering and lighting control                           | ✓              | -                | -                  | -              | -                 | -                | Mobile-based monitoring only                                  |
| [24]      | 2026 | IoT + centralized XGBoost prediction                                       | ✓              | ✓ (XGBoost)      | -                  | -              | ✓                 | -                | RMSE = 1.64                                                   |
| Our study | 2026 | IoT + edge computing + federated XGBoost + intelligent shallot cultivation | ✓              | ✓ (XGBoost)      | ✓                  | ✓              | ✓                 | ✓                | RMSE = 1.71, MAE = 1.20, MAPE = 6.40%, R <sup>2</sup> = 0.969 |

Note: RMSE = Root Mean Square Error, MAE = Mean Absolute Error, MAPE = Mean Absolute Percentage Error, R<sup>2</sup> = coefficient of determination, IoT = Internet of Things, XGBoost = Extreme Gradient Boosting.

In contrast, the proposed framework integrates IoT sensing, ESP32-based edge computing, FedAvg, XGBoost regression, and an intelligent decision support system into a unified architecture specifically designed for DWC shallot cultivation. Each hydroponic reservoir operates as an autonomous edge node that independently trains a local XGBoost model and exchanges only model parameters with the federated server, eliminating the need to transmit raw environmental data. Experimental results demonstrate that the proposed federated XGBoost model achieves an RMSE of 1.71 cm, an MAE of 1.20 cm, a MAPE of 6.40%, and an R<sup>2</sup> value of 0.969, representing prediction performance that is nearly equivalent to the centralized XGBoost model while offering substantial advantages in terms of data privacy, communication efficiency, distributed intelligence, and system scalability. It should be noted that, in the context of this small-scale hydroponic experiment, the environmental sensor data themselves do not constitute highly sensitive personal information. The reported performance metrics (RMSE, MAE, MAPE, and R<sup>2</sup>) were obtained from a single experimental run under the studied 3-reservoir, 30-day configuration, without repeated trials, confidence intervals, or formal statistical significance testing. Consequently, while the observed performance gap between the centralized and federated XGBoost models is small and consistent with expectations for FedAvg-based aggregation, its statistical reliability has not been formally verified. Repeated experimental trials and statistical validation (e.g., paired significance tests across multiple runs) are identified as important directions for future work to strengthen the robustness of the reported comparisons.

The main privacy-related advantage of the proposed federated architecture lies in limiting the centralization of detailed farm-level cultivation patterns and operational strategies, rather than in providing strong cryptographic guarantees against adversarial attacks. Furthermore, unlike previous studies that mainly provide environmental monitoring or growth prediction, the proposed system transforms prediction outcomes into actionable cultivation recommendations, including nutrient adjustment, pH

regulation, irrigation management, and lighting optimization, all of which are accessible through an Android-based dashboard. These integrated capabilities establish the proposed framework as a more comprehensive, intelligent, and practically deployable solution for next-generation IoT-enabled hydroponic shallot cultivation.

Overall, these results demonstrate that the proposed federated XGBoost-enabled IoT DWC system achieves prediction performance comparable to centralized learning while offering the additional benefits of privacy preservation, communication efficiency, and collaborative edge intelligence. These characteristics make the proposed framework a promising solution for next-generation intelligent shallot cultivation systems that require accurate growth prediction without relying on centralized data collection.

## 4. CONCLUSIONS

This study proposed a federated XGBoost-enabled IoT DWC system for intelligent shallot cultivation, integrating IoT sensing, edge computing, federated learning, and the XGBoost algorithm into a unified intelligent hydroponic cultivation framework. The proposed architecture enables each hydroponic reservoir to function as an autonomous edge node that independently trains a local prediction model while collaboratively contributing to a global model through the FedAvg algorithm without sharing raw environmental data. Experimental results demonstrated that the proposed federated XGBoost model achieved competitive prediction performance, with an RMSE of 1.71 cm, an MAE of 1.20 cm, a MAPE of 6.40%, and an R<sup>2</sup> value of 0.969, which is comparable to the centralized XGBoost approach while providing significant advantages in terms of data privacy, communication efficiency, scalability, and distributed intelligence. At the same time, the current implementation has been validated on only three edge nodes and does not yet incorporate formal privacy-enhancing techniques such as differential privacy or secure aggregation, so future work will focus on extending the federated architecture to larger multi-

node deployments and conducting a rigorous security and privacy evaluation. Furthermore, the integration of an intelligent decision support module and an Android-based monitoring dashboard enables real-time environmental monitoring, accurate growth prediction, and adaptive cultivation recommendations, thereby transforming conventional hydroponic monitoring into an intelligent decision-making system. In its current form, the proposed framework provides intelligent, human-in-the-loop decision support for DWC shallot cultivation by combining IoT sensing, distributed XGBoost prediction, and actionable recommendations delivered through a mobile dashboard. The actual adjustment of nutrient concentration, pH, irrigation, and lighting remains under the grower's control and is not yet executed automatically by the system. Future work will focus on integrating automated actuator control and conducting long-term experiments to quantify the impact of closed-loop cultivation on yield, resource efficiency, and system robustness. It should also be emphasized that, although the proposed framework conceptually improves communication efficiency and data privacy by exchanging only model parameters over the federated MQTT network, a detailed quantitative analysis of communication cost, training time, and formal privacy metrics has not yet been conducted.

Although the proposed framework demonstrates promising performance, several opportunities remain for future research. First, the proposed system should be validated through long-term field experiments involving a larger number of hydroponic reservoirs and diverse environmental conditions to further assess its scalability and robustness. Second, future studies may investigate advanced federated optimization strategies, such as FedProx, FedNova, or adaptive aggregation algorithms, to better address data heterogeneity among edge nodes and further improve prediction accuracy. Finally, integrating additional sensing modalities, including machine vision for plant phenotyping and multispectral imaging for early stress detection, together with automated actuator control for nutrient dosing, irrigation, and lighting, would enable the development of a fully autonomous intelligent hydroponic cultivation system capable of real-time adaptive management for sustainable shallot production.

## ACKNOWLEDGMENT

This work was supported by Universitas Dian Nuswantoro (Grant No. 241/F.9/UDN-09/X/2025).

## REFERENCES

- [1] Boatwain Jacques, A.A., Adamchuk, V.I., Park, J., Cloutier, G., Clark, J.J., Miller, C. (2021). Towards a machine vision-based yield monitor for the counting and quality mapping of shallots. *Frontiers in Robotics and AI*, 8: 627067. <https://doi.org/10.3389/frobt.2021.627067>
- [2] Darmono, K., Ma'ruf, Y., Fadullah, A., Setiawan, B.P., Surono. (2023). Smart shallot farming: Prototype of Internet of Things (IoT)-based shallot farming system as a sustainable agricultural innovation. *East Asian Journal of Multidisciplinary Research*, 2(6): 2363-2372. <https://doi.org/10.55927/eajmr.v2i6.4444>
- [3] Mizik, T. (2021). Climate-smart agriculture on small-scale farms: A systematic literature review. *Agronomy*, 11(6): 1096. <https://doi.org/10.3390/agronomy11061096>
- [4] Sadek, N., Shehata, D. (2024). Internet of Things based smart automated indoor hydroponics and aeroponics greenhouse in Egypt. *Ain Shams Engineering Journal*, 15(2): 102341. <https://doi.org/10.1016/j.asej.2023.102341>
- [5] Ahmed, N., Zhang, B., Deng, L., et al. (2024). Advancing horizons in vegetable cultivation: A journey from age-old practices to high-tech greenhouse cultivation—A review. *Frontiers in Plant Science*, 15: 1357153. <https://doi.org/10.3389/fpls.2024.1357153>
- [6] Lakhiar, I.A., Yan, H., Syed, T.N., et al. (2025). Soilless agricultural systems: Opportunities, challenges, and applications for enhancing horticultural resilience to climate change and urbanization. *Horticulturae*, 11(6): 568. <https://doi.org/10.3390/horticulturae11060568>
- [7] Mulyono, I.U.W., Rachmawanto, E.H., Widyatmoko, K., Sarker, M.K., Yaacob, N.M., Doheir, M. (2025). Development of an automatic watering system and plant growth monitoring for hydroponic chili production using neural networks. *Ingénierie des Systèmes d'Information*, 30(1): 21-29. <https://doi.org/10.18280/isi.300103>
- [8] Kaur, G., Upadhyaya, P., Chawla, P. (2023). Comparative analysis of IoT-based controlled environment and uncontrolled environment plant growth monitoring system for hydroponic indoor vertical farm. *Environmental Research*, 222: 115313. <https://doi.org/10.1016/j.envres.2023.115313>
- [9] Rahman, M.A., Chakraborty, N.R., Sufiun, A., Banshal, S.K., Tajnin, F.R. (2024). An AIoT-based hydroponic system for crop recommendation and nutrient parameter monitorization. *Smart Agricultural Technology*, 8: 100472. <https://doi.org/10.1016/j.atech.2024.100472>
- [10] Sneineh, A.A., Shabaneh, A.A. (2023). Design of a smart hydroponics monitoring system using an ESP32 microcontroller and the Internet of Things. *MethodsX*, 11: 102401. <https://doi.org/10.1016/j.mex.2023.102401>
- [11] Duangpakdee, K., Thananta, G., Sukpancharoen, S. (2024). IoT enhanced deep water culture hydroponic system for optimizing Chinese celery yield and economic evaluation. *Smart Agricultural Technology*, 9: 100545. <https://doi.org/10.1016/j.atech.2024.100545>
- [12] Bijanto, B., Abidin, Z., Asy'ari, F.H., Ali, R.R. (2025). An intelligent IoT-based hydroponic irrigation system for strawberry cultivation using extreme gradient boosting decision model. *Jurnal Teknik Informatika*, 6(6): 5462-5473. <https://doi.org/10.52436/1.jutif.2025.6.6.5173>
- [13] Uziah, L., Samsumar, L.D., Subki, A. (2024). Implementation of Internet of Things (IoT) for automatic monitoring and control of shallot fertilization in Perampuan Village. *Journal of Computer Science and Informatics Engineering*, 3(4): 199-210. <https://doi.org/10.55537/cosie.v3i4.946>
- [14] Sari, C.A., Purwanto, P., Rachmawanto, E.H., Syukur, A. (2024). An integration of quantum systems using BB84 for enhanced security in aeroponic smart farming. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, 22(6): 1491-1501. <https://doi.org/10.12928/telkomnika.v22i6.26450>
- [15] Bachhal, P., Anand, V., Gargrish, S., et al. (2026). Optimizing federated learning for maize leaf disease classification using internet of agricultural things. *Cluster Computing*, 29(2): 100.

- <https://doi.org/10.1007/s10586-025-05888-8>
- [16] Killeen, P., Lin, C., Li, F., Kiringa, I., Yeap, T. (2025). IoT-based smart farming architecture using federated learning: A nitrous oxide emission prediction use case. *ACM Journal on Computing and Sustainable Societies*, 3(2): 1-38. <https://doi.org/10.1145/3723039>
- [17] Sari, C.A., Isinkaye, F.O. (2023). Secure text encryption for IoT communication using affine cipher and Diffie-Hellman key distribution on Arduino ATmega2560 IoT devices. *Jurnal Teknik Informatika*, 4(4): 849-855. <https://doi.org/10.52436/1.jutif.2023.4.4.1129>
- [18] Febriyani, V., Adityo, P., Wibowo, S. (2023). Application of IoT in mobile-based shallot watering and lighting monitoring system. *Journal Informatika dan Komputer*, 6(3): 221-226. <https://doi.org/10.33387/jiko.v6i3.7143>
- [19] Habibi, I.I.A., Wirayoga, S., Huda, M., Astono, G.Y. (2024). Modern transformation in agriculture for onion watering automation with solar cell and IoT technology. *Journal of Electrical Engineering and Computer Sciences*, 9(2): 149-158. <https://doi.org/10.54732/jeeecs.v9i2.7>
- [20] Imteaj, A., Thakker, U., Wang, S., Li, J., Amini, M.H. (2022). A survey on federated learning for resource-constrained IoT devices. *IEEE Internet of Things Journal*, 9(1): 1-24. <https://doi.org/10.1109/JIOT.2021.3095077>
- [21] Gupta, S., Arora, S., Qamar, S. (2025). Federated learning in smart farming: Applications and challenges. In *Convergence of AI, Federated Learning, and Blockchain for Sustainable Development*, pp. 121-146. [https://doi.org/10.1007/978-3-031-80949-1\\_7](https://doi.org/10.1007/978-3-031-80949-1_7)
- [22] Nguyen, D.C., Ding, M., Pathirana, P.N., Seneviratne, A., Li, J., Poor, H.V. (2021). Federated learning for Internet of Things: A comprehensive survey. *IEEE Communications Surveys and Tutorials*, 23(3): 1622-1658. <https://doi.org/10.1109/COMST.2021.3075439>
- [23] Aggarwal, M., Khullar, V., Goyal, N., Prola, T.A. (2024). Resource-efficient federated learning over IoAT for rice leaf disease classification. *Computers and Electronics in Agriculture*, 221: 109001. <https://doi.org/10.1016/j.compag.2024.109001>
- [24] Mahaputra, I.P.R., Yudistira, B.G.K., Shiddiq, J.F., Adnyana, G.D.W., Nugraha, S.Z.E.S. (2026). Short-term time-series forecasting of hydroponic water quality parameters using XGBoost based IoT sensor data. *Jurnal Informatika dan Teknik Elektro Terapan*, 14(2): 9458. <https://doi.org/10.23960/jitet.v14i2.9458>
- [25] Hung, R.J. (2026). An explainable XGBoost-based framework for IoT attack detection with unseen attack family evaluation. *Sensors*, 26(10): 3005. <https://doi.org/10.3390/s26103005>
- [26] Kushawaha, A., Shah, D., Vora, D., Zade, N., Iyer, K. (2024). Urban small-scale hydroponics: A compact, smart home-based hydroponics system. *MethodsX*, 13: 102998. <https://doi.org/10.1016/j.mex.2024.102998>
- [27] Zhang, S., Han, Q., Wang, H., Liu, J., Li, B. (2025). Federated learning with dual dynamic quantization optimization in smart agriculture. *Internet of Things*, 34: 101798. <https://doi.org/10.1016/j.iot.2025.101798>
- [28] Zheng, Z., Zhou, Y., Sun, Y., Wang, Z., Liu, B., Li, K. (2022). Applications of federated learning in smart cities: Recent advances, taxonomy, and open challenges. *Connection Science*, 34(1): 1-28. <https://doi.org/10.1080/09540091.2021.1936455>
- [29] Gong, R., Zhang, H., Li, G., He, J. (2025). Edge computing-enabled smart agriculture: Technical architectures, practical evolution, and bottleneck breakthroughs. *Sensors*, 25(17): 5302. <https://doi.org/10.3390/s25175302>
- [30] Arianto, A., Ramantoko, G., Hendriadi, A., et al. (2025). Shallot crop harvest time and yield prediction using machine learning based on farmers' tacit knowledge in Brebes Regency, Indonesia. *International Journal of Design and Nature and Ecodynamics*, 20(3): 663-672. <https://doi.org/10.18280/ijdne.200321>
- [31] Agus, M., Akhsa, A.T.P.D., Hasyim, M.F. (2026). Forecasting nutrient concentration dynamics in hydroponic lettuce cultivation using a hybrid fuzzy time series and long short-term memory approach for Internet of Things-based systems. *Jurnal Teknik Informatika*, 7(3): 2277-2293. <https://doi.org/10.52436/1.jutif.2026.7.3.5600>