Innovation in Computer and Data Sciences
2(2), 1–28
DOI: https://doi.org/10.64389/icds.2026.02296
https://sphinxsp.org/journal/index.php/icds/

Research article

IoT and Data Analytics for Greenhouse Optimization: A Critical Review of Methods, Limitations, and Future Directions

Shehu Mohammed[1,*], **Neha Malhotra**[1], **Jamila Muhammad Sani**[2], **Abubakar Jibo Magayaki**[1]

[1,*] School of Computer Applications, Lovely Professional University, Phagwara, Punjab, India

[2] Department of Computer Science, DIPS Institute of Management and Technology (DIPS IMT), Jalandhar, Punjab, India

* **Correspondence:** mohammedshehumafara@gmail.com

ARTICLE INFO

Keywords:
Climate Control disease
Digital Twin
IoT
Machine Learning
Precision Agriculture
Smart Greenhouse

Mathematics Subject Classification:
62H30, 90C90, 68T09, 92B05, 00A20

Important Dates:
Received: 19 May 2026
Revised: 16 June 2026
Accepted: 17 June 2026
Online: 28 June 2026

ABSTRACT

The integration of Internet of Things (IoT) technologies with advanced data analytics is transforming greenhouse agriculture by enabling greater precision, automation, and resource efficiency. This review critically examines recent developments in IoT-enabled greenhouse optimization, focusing on sensor technologies, communication infrastructures, and machine learning models, digital twins, and autonomous control systems. Drawing on peer-reviewed studies, recent reports, and emerging research trends, the review moves beyond descriptive analytics to evaluate practical deployment challenges and real-world applicability. Although predictive analytics, deep learning, and reinforcement learning have demonstrated strong performance in controlled environments, their adoption in commercial greenhouses remains limited by data scarcity, poor model generalization, simulation-to-reality gaps, interoperability issues, and the absence of standardized benchmarking frameworks. The review further highlights challenges related to cybersecurity, explainable artificial intelligence, federated learning, and economic scalability. While digital twin technologies show potential for optimization and decision support, their widespread implementation is constrained by calibration and synchronization complexities. By identifying these technological and research gaps, this review proposes a roadmap for developing robust, scalable, and autonomous greenhouse systems. The findings emphasize the need to shift from algorithm-centric research toward deployment-oriented solutions that support sustainable and intelligent protected agriculture.

1. Introduction

Population growth, climate change, water scarcity, and degradation of arable land are mounting pressures on global food security. The United Nations estimates that world food production must increase by approximately 70% by 2050 to meet the demands of a projected population of 9.7 billion. At the same time, the occurrence and severity of extreme weather events continue to destabilize traditional open-field agriculture, necessitating the development of more resilient and controlled production systems. Greenhouse horticulture has emerged as a promising solution, enabling year-round production of high-value crops under controlled environmental conditions that protect plants from climatic variability and optimize growth conditions [1], [2]. Unlike existing review studies, this work provides a critical evaluation of current approaches, focusing on real-world deployment limitations and bridging the gap between experimental performance and commercial greenhouse applications.

Nonetheless, conventional greenhouse control relies heavily on human expertise and manual intervention, introducing variability, scalability challenges, and operational inefficiencies. The introduction of the Internet of Things (IoT) has introduced a paradigm shift by enabling the deployment of dense sensor networks capable of continuously monitoring environmental conditions, crop physiological status, and resource utilization patterns. Combined with advanced data analytics and artificial intelligence (AI), these IoT systems transform raw sensor data into actionable insights that support automated climate control, precision irrigation, predictive maintenance, and yield optimization [1], [3], [4].

The combination of IoT and data analytics in greenhouse operations has demonstrated substantial benefits across multiple dimensions. Recent studies have reported improvements in water-use efficiency of up to 90%, energy savings of 15–30% through intelligent climate control algorithms, and yield increases of 10–25% through data-driven cultivation strategies [5], [6]. Furthermore, the global smart greenhouse market was valued at approximately USD 2.24 billion in 2025 and is projected to reach USD 3.77–4.59 billion by 2030, with a compound annual growth rate exceeding 8.5% [7], [8], [9].

This review provides an in-depth analysis of the technological ecosystem underpinning IoT-enabled greenhouse optimization. The discussion is organized around three fundamental pillars:

1. The underlying IoT infrastructure, including sensor technologies, communication protocols, and system architectures [10];

2. The analytical and algorithmic techniques used to process greenhouse data, including machine learning, deep learning, reinforcement learning, and digital twins [11];

3. The applications, challenges, and future directions of integrated IoT-enabled greenhouse systems [12].

By synthesizing recent developments in these areas, this review provides researchers, agricultural practitioners, and technology developers with a structured understanding of the current state of the field and the emerging research frontiers that are shaping the next generation of smart greenhouse systems. Unlike conventional review studies that primarily summarize technological advancements, this review adopts a critical deployment-oriented perspective by systematically examining the limitations that hinder the transition of IoT and AI technologies from controlled experimental environments to commercial greenhouse operations.

The novelty of this work lies in its comprehensive analysis of real-world deployment barriers, including poor model generalization, simulation-to-reality gaps, interoperability failures, limited explainability, cybersecurity vulnerabilities, and economic scalability constraints. Furthermore, this review integrates emerging concepts such as federated learning, digital twins, edge intelligence, and autonomous greenhouse systems into a unified research roadmap aimed at bridging the gap between theoretical innovation and practical implementation.

2. IoT Architecture for Smart Greenhouses

Modern IoT-based greenhouse systems are typically organized into hierarchical architectures that divide functionality across multiple layers, each responsible for specific tasks related to data acquisition, communication, processing, and decision-making. Such layered architectures provide modularity, scalability, and interoperability, enabling system designers to select technologies appropriate to the operational requirements of a greenhouse facility. The most widely adopted architectures in agricultural IoT applications consist of three or four layers, commonly including the perception layer, network layer, and application layer, while some frameworks incorporate an additional storage or processing layer to enhance data management and computational capabilities [3], [13].

2.1. Perception Layer

The sensing layer, also known as the perception layer, forms the physical foundation of a greenhouse IoT system. This layer provides direct interaction with the greenhouse environment and comprises a heterogeneous collection of sensors, actuators, cameras, and microcontrollers that support three primary functions: environmental monitoring, crop-state observation, and automated control implementation [14], [15].

Environmental monitoring within the perception layer involves the continuous measurement of key climatic variables that influence plant growth and development. Temperature sensors, including thermistors, resistance temperature detectors (RTDs), and thermocouples, provide air and substrate temperature measurements with accuracies ranging from ±0.1°C to ±0.5°C. Humidity sensors, typically based on capacitive or resistive technologies, are used to determine relative humidity and calculate vapor pressure deficit (VPD), a critical parameter affecting transpiration rates and disease susceptibility. Photosynthetically active radiation (PAR) is measured using photodiodes and quantum sensors to support decisions related to supplemental lighting. Carbon dioxide sensors, commonly based on non-dispersive infrared (NDIR) technology, are used to monitor CO_2 concentration levels within greenhouse environments to maximize photosynthetic efficiency through controlled CO_2 enrichment.

In addition to environmental monitoring, crop-directed sensing technologies are increasingly being integrated into the perception layer. Soil moisture and electrical conductivity (EC) sensors provide real-time information on substrate water content and nutrient availability. pH sensors enable accurate monitoring of nutrient solution acidity, while RGB and multispectral cameras generate visual data for crop health assessment, growth monitoring, and pest detection. Advanced greenhouse systems may also incorporate leaf wetness sensors, stem diameter sensors, and fruit growth sensors, providing direct physiological measurements of crop status.

The actuation component of the perception layer converts control decisions into physical actions. Actuators include relays and motor controllers that operate ventilation windows, heating systems, cooling pads, fogging units, supplemental lighting fixtures, irrigation valves, and carbon dioxide injection equipment. The integration of sensing and actuation elements enables the implementation of closed-loop control systems capable of maintaining environmental set points with minimal human intervention [16], [14].

2.2. Network Layer

The network layer, also referred to as the transmission or communication layer, is responsible for transferring data between greenhouse devices and higher-level computing platforms used for data management and analysis. This layer must address several challenges specific to greenhouse environments, including signal attenuation caused by dense crop canopies, the adverse effects of high humidity on electronic communication systems, and the need to support both high-bandwidth video streams and low-bandwidth sensor data across large and complex operational areas [17], [13].

Most greenhouse IoT network architectures employ heterogeneous communication technologies rather than relying on a single communication protocol. Instead, multiple technologies are integrated within a layered framework to satisfy diverse connectivity requirements. At the device level, short-range wireless communication protocols such as Zigbee, Bluetooth Low Energy (BLE), and Wi-Fi are commonly used to connect sensors and actuators to local gateways [17].

Zigbee is particularly well suited for greenhouse sensor networks due to its low power consumption, support for mesh networking, and ability to accommodate a large number of devices. Operating in the 2.4 GHz Industrial, Scientific, and Medical (ISM) band, Zigbee provides reliable indoor communication while extending network coverage through multi-hop routing [17], [13]. BLE offers even lower power consumption and is suitable for battery-powered devices, although it provides shorter communication ranges and lower data throughput. In contrast, Wi-Fi delivers the bandwidth required for applications such as video streaming and high-frequency data transmission but consumes significantly more energy [17].

For long-range communication between greenhouse facilities and cloud platforms, Low-Power Wide-Area Networks (LPWANs) such as LoRaWAN, NB-IoT, and Sigfox are widely employed. Among these technologies, LoRaWAN has gained significant attention in agricultural applications due to its energy efficiency, kilometer-scale communication range, operation in unlicensed frequency bands, and minimal infrastructure requirements [17]. NB-IoT leverages existing cellular infrastructure to provide stable and reliable connectivity; however, it requires network subscription services and adequate cellular coverage [13].

Beyond physical communication technologies, the network layer also encompasses the software protocols responsible for reliable data exchange among devices, gateways, and cloud services. Message Queuing Telemetry Transport (MQTT) is widely adopted because of its lightweight design, low bandwidth requirements, and publish-subscribe communication model, which facilitates scalable and reliable message delivery even under unstable network conditions. The Constrained Application Protocol (CoAP) provides another lightweight communication mechanism based on RESTful principles, making it suitable for resource-constrained IoT devices. Hypertext Transfer Protocol (HTTP) remains relevant for web-based services and cloud integration, particularly where compatibility with existing internet infrastructure is required [3], [13].

2.3. Application Layer

The application layer represents the highest level of the IoT architecture, where collected data are processed, analyzed, and transformed into actionable insights and control decisions. This layer encompasses cloud computing platforms, data analytics engines, machine learning models, visualization dashboards, and decision-support systems that collectively enable intelligent greenhouse management [1], [3].

Cloud computing platforms play a central role in large-scale data storage, processing, and analytics. Major cloud service providers, including Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP), offer dedicated IoT solutions that support device management, time-series data ingestion, stream processing, and integration with machine learning pipelines [3]. These platforms enable greenhouse operators to aggregate data across multiple facilities, perform large-scale historical trend analysis, and deploy advanced analytical models that may be computationally infeasible on edge devices [3], [18].

The application layer also provides the interfaces through which greenhouse operators interact with the IoT ecosystem. Web-based dashboards offer real-time visualization of environmental parameters, crop health indicators, and system alerts, facilitating data-driven decision-making. Mobile applications further enable remote monitoring and control, allowing operators to adjust environmental set points, acknowledge alarms, and review operational histories from virtually any location [16].

At a more advanced level, intelligent greenhouse systems incorporate automated decision-making capabilities that reduce dependence on continuous human supervision and support the transition toward autonomous greenhouse operations. These systems integrate predictive analytics, machine learning algorithms, and control strategies to optimize environmental conditions, resource utilization, and crop productivity with minimal human intervention [19], [20].

Figure 1 illustrates the interaction among the perception, network, and application layers, including sensors, gateways, cloud servers, analytics platforms, and actuators within a typical smart greenhouse architecture.

3. Sensor Technologies for Greenhouse Monitoring

The effectiveness of IoT-enabled greenhouse systems depends heavily on the quality, reliability, and diversity of sensor technologies deployed throughout the facility. Sensors serve as the primary interface between the physical greenhouse environment and digital management systems, enabling continuous monitoring of environmental conditions, crop status, and resource utilization. This section reviews the major sensor technologies employed in modern greenhouse operations, highlighting their specifications, functionalities, and contributions to data-driven crop production and decision-making [14], [15].

Environmental sensors represent the most mature and widely deployed category of greenhouse monitoring devices. Virtually all smart greenhouse IoT systems utilize temperature and humidity sensors to collect fundamental climatic information required for effective climate control [15], [16]. Commonly used digital sensors include the DHT22, which provides accurate temperature and humidity measurements over a broad operating range; the SHT30, which is factory-calibrated for improved thermal stability; and the BME280, which supports I^2C and SPI communication protocols for seamless integration with microcontroller-based platforms [15]. High-precision research-grade sensors can achieve accuracies of approximately ±0.1°C and ±1.5% relative humidity, whereas commercially deployed greenhouse systems typically operate with sensor accuracies of approximately ±0.5°C and ±3–5% relative humidity [14].

Light sensing plays a critical role in greenhouse management by supporting decisions related to supplemental lighting, shading, and crop growth optimization. Measurements of solar radiation and Daily Light Integral (DLI) provide valuable information regarding the quantity of light received by crops throughout the day [14]. Photosynthetically Active Radiation (PAR) sensors quantify light within the 400–700 nm wavelength range that drives photosynthesis and are typically expressed in $\mu mol, m^{-2}, s^{-1}$ [14]. Quantum sensors provide highly accurate PAR measurements, although at a higher cost, while photodiode-based sensors offer a cost-effective alternative with acceptable accuracy for many commercial greenhouse applications.

For greenhouses employing CO_2 enrichment strategies to enhance photosynthetic activity, continuous monitoring of carbon dioxide concentration is essential. Non-Dispersive Infrared (NDIR) sensors are the

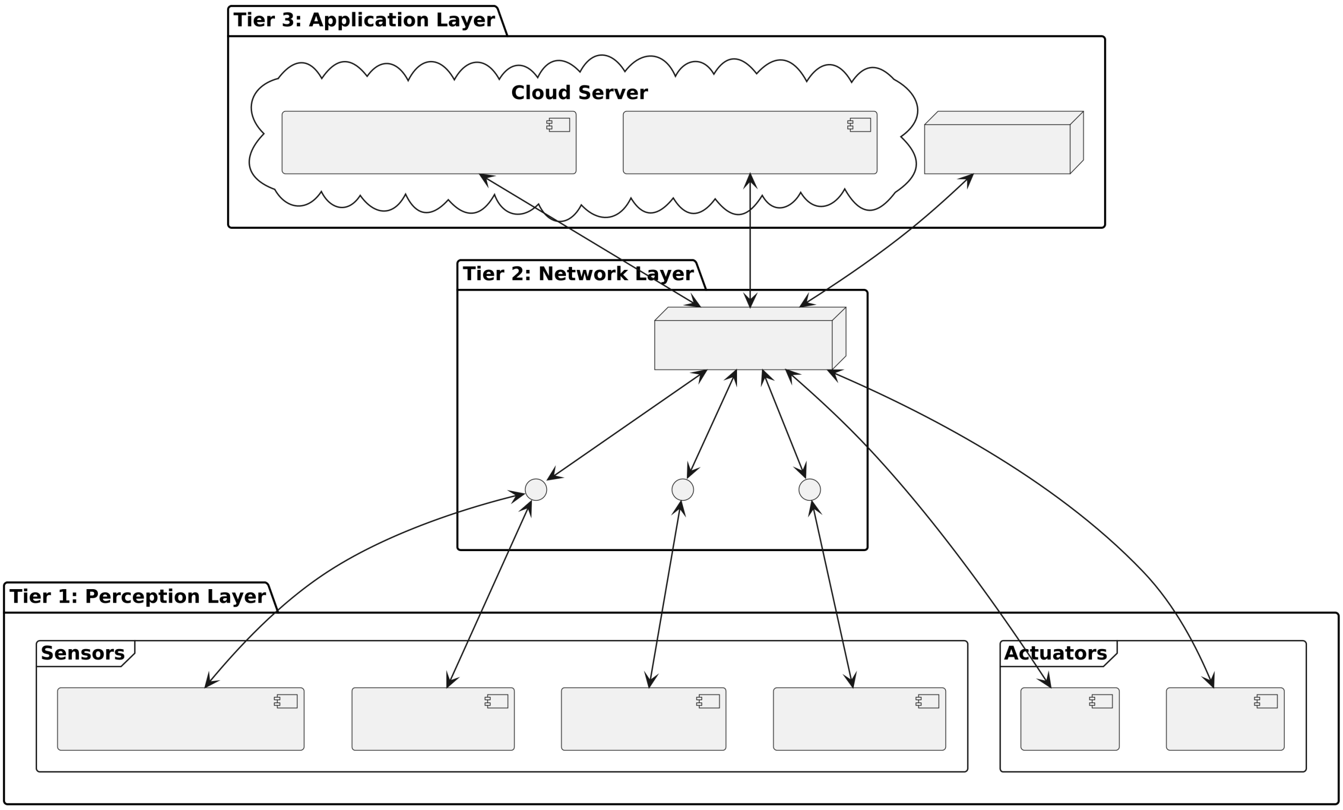

Figure 1. Three-Tier IoT Architecture for Smart Greenhouse Systems. The perception layer consists of environmental sensors and actuators responsible for data acquisition and control execution. The network layer facilitates communication between field devices and higher-level platforms through gateways and communication protocols. The application layer provides cloud-based storage, data analytics, machine learning, visualization dashboards, and decision-support services that enable intelligent greenhouse management and automated control.

dominant technology for this purpose because of their long-term stability, low maintenance requirements, and wide measurement ranges, typically spanning 0–2000 ppm or 0–5000 ppm in enrichment applications [15]. Accurate CO_2 monitoring is critical due to the significant operational costs associated with enrichment and the potential economic and environmental consequences of both under-dosing and over-dosing [6].

Table 1 summarizes the principal sensor technologies deployed in greenhouse IoT systems, together with their typical specifications and primary applications.

Table 1. Summary of Greenhouse Sensor Technologies

Sensor Type	**Parameters**	**Accuracy Range**	**Primary Application**
Temperature	Air temperature, substrate temperature	±0.1 to ±0.5°C	Climate control
Humidity	RH, VPD, dew point	±1.5–5% RH	Transpiration management
Light (PAR)	400–700 nm flux	±5%	Lighting control
CO_2 Concentration	400–1500 μmol mol^{-1}	±30 μmol mol^{-1}	Enrichment control
Soil Moisture	Volumetric water content	±2–3%	Irrigation scheduling
EC/pH	Ionic concentration, acidity	±2% EC, ±0.1 pH	Fertigation control
Multispectral Camera	VIS + NIR reflectance	Varies	Crop health monitoring
Thermal Infrared	Canopy temperature	±0.5°C	Water stress detection

Note: RH = Relative Humidity; VPD = Vapor Pressure Deficit; PAR = Photosynthetically Active Radiation; VIS = Visible Spectrum; NIR = Near Infrared.

Soil and substrate sensors provide critical insights into root-zone conditions and play an essential role in precision irrigation and fertigation management. Soil moisture sensors, including Time-Domain Reflectometry (TDR), Frequency-Domain Reflectometry (FDR), and capacitance-based sensors, enable continuous monitoring of water availability and facilitate accurate irrigation scheduling based on crop requirements rather than fixed watering intervals [14], [15]. Electrical Conductivity (EC) sensors measure ionic concentrations within the growing medium, providing information on nutrient availability, while pH sensors monitor nutrient solution acidity, which directly influences nutrient uptake efficiency. The integration of these sensing technologies supports automated fertigation systems capable of delivering water and nutrients according to real-time crop demands [21].

Imaging sensors are becoming increasingly important components of modern greenhouse monitoring systems. Conventional RGB cameras provide visual information for crop growth assessment, fruit counting, and quality evaluation. More advanced imaging technologies, including multispectral and hyperspectral cameras, capture information beyond the visible spectrum, incorporating near-infrared and shortwave-infrared wavelengths that enable early detection of plant stress, nutrient deficiencies, and disease symptoms before they become visually apparent [22]. Thermal infrared cameras further enhance monitoring capabilities by measuring canopy temperature, thereby facilitating the detection of water stress and disease conditions. In addition, three-dimensional depth cameras provide structural information about plant architecture, enabling detailed analysis of canopy development, plant morphology, and growth dynamics.

4. Communication Protocols and Connectivity

The selection of appropriate communication protocols and networking technologies is a critical design consideration in greenhouse IoT systems, as it directly influences system reliability, scalability, energy efficiency, and deployment costs. This section examines the principal wireless communication technologies and networking protocols employed in smart greenhouse environments, comparing their characteristics and suitability for various application requirements [13], [17].

Wireless communication technologies used in greenhouse IoT systems can generally be categorized into three groups based on communication range and power consumption: short-range communication protocols, cellular communication technologies, and Low-Power Wide-Area Network (LPWAN) protocols [17]. Short-range technologies such as Zigbee, Bluetooth Low Energy (BLE), Wi-Fi, and Z-Wave are commonly utilized for communication among sensors, actuators, and local gateway devices within greenhouse facilities [13], [17]. Cellular technologies, including 4G LTE, 5G, Narrowband IoT (NB-IoT), and LTE-M, provide extended connectivity for remote monitoring, cloud integration, and distributed greenhouse operations [13], [23]. LPWAN technologies such as LoRaWAN and Sigfox offer a compromise between communication range and energy consumption, enabling long-distance transmission of small data packets while maintaining low power requirements [17].

Among short-range communication technologies, Zigbee has emerged as one of the most widely adopted protocols for greenhouse sensor networks. Operating within the 2.4 GHz Industrial, Scientific, and Medical (ISM) frequency band, Zigbee supports data transmission rates of up to 250 kbps. A key advantage of Zigbee is its mesh networking capability, which enables multi-hop communication and self-healing network architectures that improve coverage and reliability [13], [17]. Furthermore, a single Zigbee coordinator can manage hundreds of connected devices, making the protocol particularly suitable for large-scale greenhouse deployments. Its low power consumption also allows sensor nodes to operate for extended periods using battery-powered configurations [17].

Wi-Fi provides substantially higher bandwidth than Zigbee, making it suitable for data-intensive applications such as video surveillance, image transmission, and high-frequency sensor reporting [17]. However, these advantages are accompanied by significantly higher energy consumption, limiting its suitability for battery-powered devices. In addition, signal attenuation caused by dense vegetation and greenhouse structures may reduce communication reliability and effective transmission range. Consequently, many greenhouse IoT architectures adopt hybrid communication strategies that utilize Zigbee or BLE for low-data-rate sensor nodes while reserving Wi-Fi connectivity for cameras, gateways, and high-bandwidth applications [17].

At the application layer, Message Queuing Telemetry Transport (MQTT) has become one of the most widely adopted messaging protocols in greenhouse IoT systems. MQTT employs a publish-subscribe communication model that decouples data producers, such as sensors, from data consumers, including analytics platforms and decision-support systems. This architecture enhances system flexibility, scalability, and maintainability [3]. The protocol's lightweight design, characterized by a fixed header size of only two bytes, enables efficient operation under constrained network conditions. Furthermore, MQTT supports multiple Quality of Service (QoS) levels, allowing system designers to balance communication reliability and bandwidth utilization according to application requirements [3], [13].

Table 2 presents a comparative summary of the major wireless communication technologies employed in greenhouse IoT systems, including their communication range, power consumption characteristics, and application suitability.

Table 2. Comparison of Wireless Communication Technologies for Greenhouse IoT

Technology	Data Rate	Range	Power Consumption
Zigbee	250 kbps	30–300 ft	Very Low
BLE	1 Mbps	100 ft	Low
Wi-Fi 6	< 9.6 Gbps	150 ft	High
LoRaWAN	< 50 kbps	1–9 miles	Extremely Low
NB-IoT	250 kbps	0.6–9 miles	Very Low
5G	1–10 Gbps	< 1 mile	Moderate–High

Note: BLE = Bluetooth Low Energy; NB-IoT = Narrowband Internet of Things.

As shown in Table 2, the selection of wireless communication technology depends on the specific operational requirements of the greenhouse system. Zigbee is widely adopted for sensor networks because of its low power consumption and mesh networking capability. BLE is particularly suitable for battery-operated devices, whereas Wi-Fi 6 supports high-bandwidth applications such as image and video transmission. LoRaWAN enables long-range environmental monitoring with minimal energy consumption, while NB-IoT provides reliable cellular connectivity for distributed IoT deployments. In contrast, 5G offers ultra-low latency and high data rates, making it suitable for robotics, autonomous control, and real-time video applications in next-generation smart greenhouses.

5. Data Analytics and Machine Learning Applications (Critical Perspective)

The rapid expansion of IoT sensor networks in greenhouse systems has resulted in the generation of large-scale time-series datasets that capture complex interactions among environmental variables, plant physiological responses, and control actions. These datasets provide valuable opportunities for the application of data analytics and machine learning techniques to support intelligent greenhouse management. However, despite substantial progress in predictive modeling and automated decision-making, many existing approaches remain constrained by challenges related to model generalization, data dependency, interpretability, and real-world deployment feasibility.

Although numerous studies have reported promising experimental results, a significant gap remains between algorithmic performance achieved under controlled research conditions and practical implementation in commercial greenhouse environments. Variations in crop types, environmental conditions, greenhouse infrastructure, and sensor configurations often limit the transferability of models across different operational settings. Furthermore, issues such as data scarcity, sensor noise, missing observations, and computational constraints continue to affect the reliability and scalability of machine learning-based greenhouse management systems [1], [3], [4].

Although these analytical approaches demonstrate considerable potential for enhancing greenhouse management, their practical deployment remains constrained by several technical, operational, and economic challenges. Many studies report high predictive performance under controlled experimental conditions; however, issues related to scalability, interoperability, computational resource requirements, model interpretability, data quality, and deployment robustness remain insufficiently addressed in commercial greenhouse environments. Furthermore, the transferability of machine learning models across different crops, climatic conditions, greenhouse infrastructures, and sensor configurations continues to pose a signif-

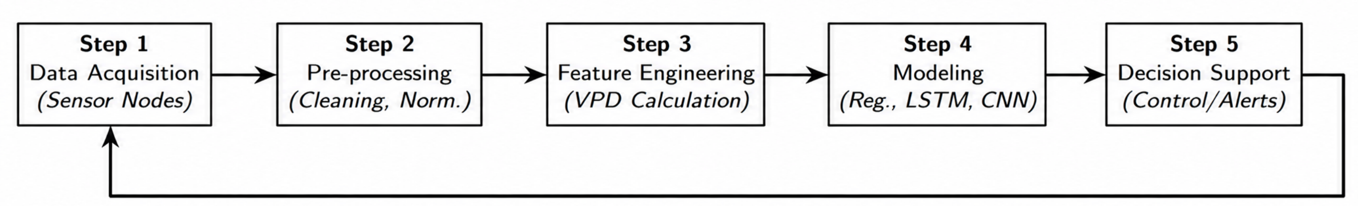

Figure 2. Data Analytics Workflow for Greenhouse Optimization. The workflow illustrates the sequential process through which greenhouse data are transformed into intelligent management decisions. Sensor nodes continuously acquire environmental, crop, and operational data, which undergo preprocessing procedures such as cleaning, normalization, and filtering. Relevant features, including climate indices and crop-specific indicators, are subsequently extracted and engineered for analysis. Machine learning and deep learning models, including regression algorithms, Long Short-Term Memory (LSTM) networks, and Convolutional Neural Networks (CNNs), are then employed to generate predictions and recommendations. The resulting outputs support intelligent decision-making processes, enabling automated climate control, precision irrigation, resource optimization, and real-time greenhouse management.

Table 3. Data Analytics Techniques for Greenhouse Optimization

Category	**Description**
Descriptive	Statistical analysis for historical trend reporting
Predictive	LSTM, ARIMA, and GRU for yield and weather forecasting
Prescriptive	Reinforcement learning for autonomous climate control
Diagnostic	CNN and SVM for pest and disease detection
Explainable AI	SHAP, LIME, and XAI models for transparent decision support

Note: LSTM = Long Short-Term Memory; ARIMA = AutoRegressive Integrated Moving Average; GRU = Gated Recurrent Unit; CNN = Convolutional Neural Network; SVM = Support Vector Machine; XAI = Explainable Artificial Intelligence.

icant challenge. As a result, the gap between experimental success and real-world implementation remains substantial. Tables 4 and 5 provide a critical comparison of the major artificial intelligence approaches employed in greenhouse optimization. Table 4 summarizes their primary applications and reported advantages, whereas Table 5 highlights deployment limitations and current adoption status.

Table 4. Applications and Advantages of AI Approaches for Greenhouse Optimization

AI Approach	Description
ARIMA/SARIMA	Climate forecasting; simple, interpretable, and low computational cost
LSTM/GRU	Climate prediction, irrigation forecasting, and yield estimation; captures temporal dependencies and nonlinear relationships
CNN-Based Vision Models	Disease detection, pest monitoring, and crop classification; high classification accuracy under controlled conditions
Reinforcement Learning (RL)	Autonomous climate control; learns adaptive control strategies and reduces manual intervention
Digital Twin Systems	Simulation, optimization, and predictive control; enables scenario testing and virtual optimization
Explainable AI (XAI)	Transparent decision support; improves interpretability and user trust
Federated Learning	Privacy-preserving multi-greenhouse learning; enables collaborative optimization without data sharing
Edge AI Systems	Real-time local inference; reduced latency and bandwidth usage

5.1. Predictive Analytics and Climate Control

Predictive analytics in greenhouses aims to forecast future environmental conditions, crop responses, and resource requirements using historical and real-time data [24], [25]. While this enables proactive greenhouse management, the reliability of such predictions is highly dependent on data quality, model assumptions, and environmental consistency.

Traditional time-series models such as ARIMA and SARIMA provide interpretable baselines for short-term forecasting. However, their linear assumptions limit their ability to capture the nonlinear and highly coupled dynamics of greenhouse environments. As a result, deep learning approaches such as LSTM and GRU models have been widely adopted due to their ability to model temporal dependencies and nonlinear relationships [24], [25], [26].

Despite their reported accuracy, these models suffer from poor generalization across different greenhouse settings. Most studies train and validate models using data from a single facility or controlled experimental setup, which does not account for variations in climate conditions, crop types, greenhouse structures, and sensor configurations. Consequently, models that perform well in research environments often degrade significantly when deployed in commercial operations [27]. This highlights the lack of cross-domain validation and standardized benchmarking in greenhouse predictive analytics.

Model Predictive Control (MPC) integrated with machine learning has shown promise for optimizing

Table 5. Deployment Limitations and Status of AI Approaches for Greenhouse Optimization

AI Approach	Description
ARIMA/SARIMA	Poor nonlinear modeling capability; limited to short-term forecasting
LSTM/GRU	Requires large datasets; poor cross-greenhouse generalization; mostly experimental or pilot-scale
CNN-Based Vision Models	Sensitive to lighting variation, occlusion, and dataset bias; limited commercial deployment
Reinforcement Learning (RL)	Simulation-to-reality gap; unsafe exploration; primarily simulation-based
Digital Twin Systems	Calibration complexity and synchronization challenges; mostly research-stage deployment
Explainable AI (XAI)	Limited real-time integration; early-stage adoption
Federated Learning	Data heterogeneity and synchronization difficulties; mostly conceptual implementation
Edge AI Systems	Hardware resource constraints limit model complexity; emerging deployment paradigm

climate control by combining predictive modeling with constraint-based optimization [25], [26]. However, MPC performance is highly sensitive to model accuracy, and errors in prediction can propagate into suboptimal or unstable control actions. While GRU-based predictive controllers have demonstrated improvements over conventional MPC in experimental studies [24], their robustness under real-world disturbances such as sensor noise, actuator delays, and environmental variability remains insufficiently validated.

5.2. Deep Learning for Crop Monitoring

Deep learning, particularly convolutional neural networks (CNNs), has significantly advanced computer vision applications in greenhouse agriculture, including disease detection, growth monitoring, and yield estimation [4], [22]. Reported accuracies exceeding 95% in disease classification demonstrate the strong potential of these models under controlled conditions [22].

However, these performance metrics often rely on curated and balanced datasets, which do not reflect real greenhouse environments characterized by varying lighting conditions, occlusions, plant overlaps, and sensor inconsistencies. As a result, model robustness and reliability in operational settings remain questionable.

Another limitation lies in the data dependency of deep learning models. Training high-performance models requires large volumes of labeled data, which are difficult and expensive to obtain in greenhouse contexts. Additionally, dataset bias, such as limited crop diversity or specific environmental conditions, further reduces model transferability across different greenhouse systems [27].

While edge computing has been proposed to enable real-time inference, computational constraints on edge devices limit the deployment of complex deep learning architectures. This creates a trade-off between model accuracy and real-time performance, particularly in resource-constrained greenhouse environments.

5.3. Reinforcement Learning for Autonomous Control

Reinforcement learning (RL) has emerged as a promising approach for autonomous greenhouse climate control, where an agent learns optimal control policies through interaction with the environment [26], [28]. Algorithms such as PPO, DDPG, SAC, and TD3 have demonstrated strong performance in simulation-based studies; with TD3 often showing improved stability and efficiency [28].

Despite these advances, RL-based control systems face significant barriers to real-world deployment. Most RL models are trained in simulated environments that simplify greenhouse dynamics and fail to capture real-world uncertainties such as sensor drift, equipment degradation, unexpected weather changes, and biological variability [10], [19]. This leads to the well-known simulation-to-reality gap, where policies learned in simulation do not transfer effectively to physical systems.

The Autonomous Greenhouse Challenge has demonstrated the feasibility of AI-driven greenhouse control systems [19], [20]; however, these systems rely heavily on controlled experimental setups and extensive pre-training. Their scalability and robustness in diverse commercial greenhouse environments remain largely unproven.

Furthermore, RL methods require extensive exploration during training, which can be unsafe or impractical in real greenhouse operations where incorrect actions may lead to crop damage or economic loss. Hybrid approaches combining RL with traditional control strategies show potential [28], but their practical adoption is still limited.

5.4. Explainable AI in Greenhouse Systems

As AI-driven decision-making becomes increasingly central to greenhouse management, the need for transparency and interpretability has gained significant attention. Explainable AI (XAI) techniques such as SHAP, LIME, and attention mechanisms aim to provide insights into model behavior and improve user trust [29], [30].

While these methods enhance interpretability, their practical integration into operational greenhouse systems remains limited. Most XAI implementations are evaluated in research settings rather than deployed in real-world decision-support systems. Additionally, explanations generated by these methods may not always align with agronomic reasoning, leading to potential confusion among end users.

For example, models such as Temporal Fusion Transformers combined with SHAP and LIME provide high predictive accuracy and feature attribution [31], but their complexity and computational requirements limit their usability in real-time greenhouse applications.

A key challenge lies in balancing predictive accuracy and interpretability. Highly accurate models are often complex and difficult to explain, while simpler interpretable models may lack sufficient predictive power. This trade-off continues to hinder the adoption of AI systems in greenhouse operations, where trust and reliability are critical for decision-making.

Fig. 3 compares major artificial intelligence algorithms based on predictive accuracy, computational cost, latency, and robustness.

6. Digital Twins for Greenhouse Simulation

One of the most innovative applications for greenhouse optimization is the concept of digital twins [32], [33]. A digital twin is a virtual representation of a physical greenhouse system that can be used to

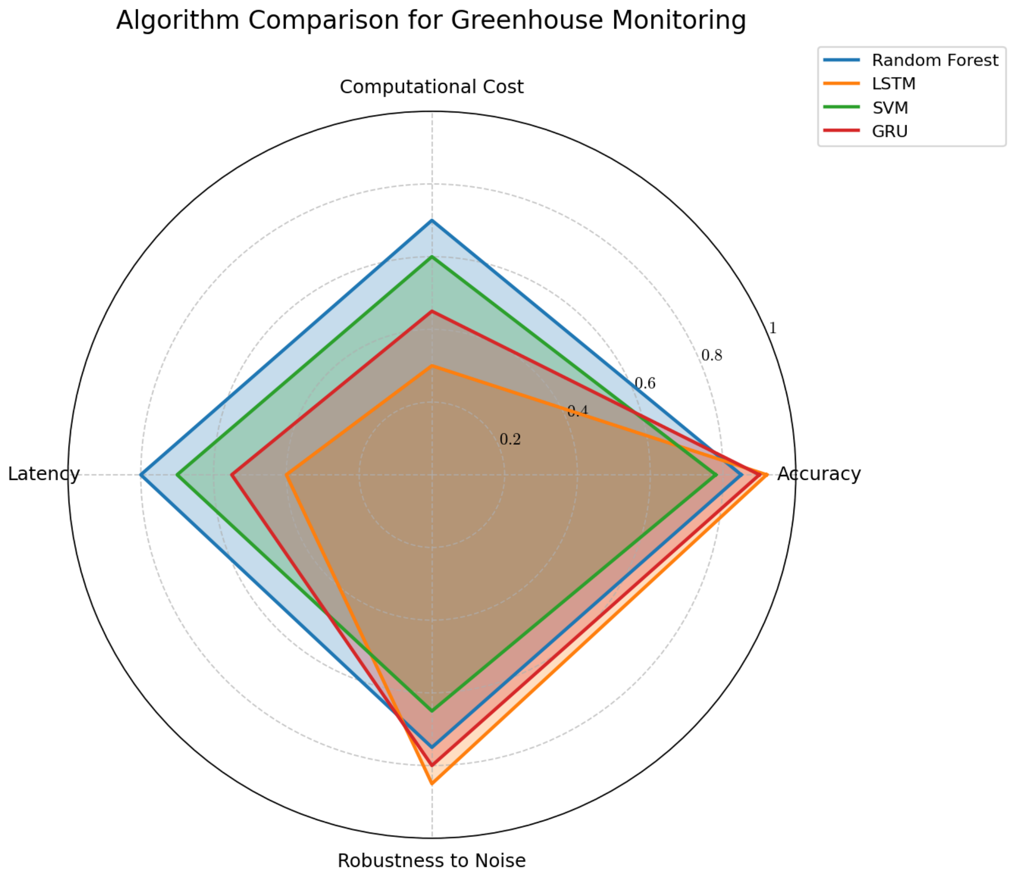

Figure 3. Comparative Performance of Machine Learning Algorithms for Greenhouse Monitoring. The radar chart compares Random Forest, LSTM, SVM, and GRU models based on four key evaluation criteria: predictive accuracy, computational cost, latency, and robustness to noise. LSTM and GRU achieve the highest predictive accuracy and robustness, making them suitable for time-series forecasting applications in greenhouse environments. Random Forest demonstrates balanced performance with moderate computational requirements, whereas SVM exhibits lower computational cost but reduced robustness and latency performance. The comparison highlights the trade-offs among predictive performance, computational efficiency, and operational reliability when selecting machine learning algorithms for greenhouse monitoring and decision-support applications.

simulate, analyze, and optimize operations without interfering with the actual environment [32], [33]. A greenhouse digital twin integrates real-time sensor data with physics-based and data-driven models to create a live virtual model that emulates the behavior of the physical greenhouse. This enables the simulation of management scenarios, training of AI-based control systems, and optimization of resource utilization [32], [34], [35].

Digital twins are more than visualization dashboards; they represent the complex interactions among climate control systems, plant physiology, and environmental conditions [32]. Researchers at Wageningen University have pioneered greenhouse digital twins that integrate sensor data with three-dimensional functional-structural plant models, enabling continuous assessment of how climate control and cultivation strategies affect plant growth [36]. These systems incorporate predictions of light distribution, water balance, and temperature dynamics to support forward-looking management decisions [32], [37].

Energy optimization represents one of the most significant benefits of digital twin technology in greenhouse production. The Greenhouse Industry 4.0 initiative in Denmark integrated IoT technologies, artificial intelligence, and cloud computing into a platform capable of simultaneously optimizing crop production, energy consumption, and labor costs [32], [38]. By responding to electricity market signals and demand-response programs, the platform reduced operational costs while maintaining crop yield and quality [38]. The digital twin further enabled scenario-based forecasting of greenhouse behavior, providing decision-support capabilities beyond those achievable through manual analysis.

Computational Fluid Dynamics (CFD) models constitute another important component of greenhouse digital twins. These models provide high-resolution spatial predictions of temperature, humidity, and airflow distributions that cannot be captured by conventional lumped-parameter approaches [37]. CFD-based analyses have identified optimal ventilation configurations capable of reducing average and maximum indoor temperatures under both normal and high heat-load conditions. However, CFD simulations are computationally intensive, often requiring up to 48 h for a single analysis. Consequently, reduced-order modeling techniques have been developed to achieve acceptable predictive accuracy while reducing response times to the order of seconds [37].

Digital twins also play a critical role in training reinforcement learning and other AI-driven control systems. The virtual environment allows agents to safely explore alternative control strategies without risking crop damage or operational disruptions. This significantly accelerates learning and enables the evaluation of strategies that would be impractical or unsafe in real greenhouse facilities [33]. The Wageningen Autonomous Greenhouse Challenge demonstrated the effectiveness of this approach, with participating teams extensively pre-training their control algorithms in simulation environments before deployment in real greenhouse settings [19], [20].

7. Integrated Pest and Disease Management

The Internet of Things (IoT) and data analytics have significantly enhanced greenhouse-based Integrated Pest Management (IPM), transforming conventional scouting practices into automated pest and disease detection and intervention systems [12], [39]. By enabling continuous monitoring and real-time data acquisition, IoT technologies address one of the most critical limitations of sustainable pest management: the lack of timely and localized information required for rapid intervention [12].

AIoT-based monitoring systems can automatically detect and identify common greenhouse insect pests with high accuracy. For example, an Intelligent and Integrated Pest and Disease Management (I2PDM) system deployed in commercial tomato and orchid greenhouses employed edge-computing devices and deep learning models to identify thrips (*Frankliniella intonsa*, *Thrips hawaiiensis*, *Thrips tabaci*) and whiteflies (*Bemisia argentifolii*, *Trialeurodes vaporariorum*) [12]. The system operated continuously over extended periods and generated long-term spatial-temporal pest population data, supporting data-driven IPM decisions and effective pest suppression strategies [12].

Computer vision constitutes a fundamental component of automated pest monitoring. Deep learning models trained on images captured from pest traps, plant surfaces, and crop canopies can detect pest occurrence, classify pest species, and estimate population density [22]. When integrated with robotics, computer vision enables autonomous greenhouse scouting. Mobile robots equipped with cameras and artificial intelligence can autonomously navigate greenhouse rows, continuously acquire images, perform real-time analysis, and identify localized pest hotspots [40].

Environmental monitoring data collected through IoT sensors further enhance IPM programs by supporting predictive models of pest and disease development. Many greenhouse pests and pathogens exhibit growth patterns strongly influenced by temperature and humidity conditions, and predictive outbreak models have demonstrated promising performance in previous studies [12], [39]. By integrating real-time sensor measurements with pest biology models, decision-support systems can alert growers when environmental conditions become favorable for pest outbreaks before severe infestations occur. Such proactive interventions reduce pesticide usage and promote biological control strategies, which remain central principles of sustainable IPM practices [39].

Pest detection has also become an important component of the Autonomous Greenhouse Challenge, where research teams develop deep learning models capable of identifying pests on small monitoring traps [19], [20]. These efforts demonstrate the considerable progress achieved in automated pest monitoring while simultaneously highlighting existing challenges, particularly the development of robust models capable of operating under varying environmental conditions, trap configurations, and pest life stages.

8. Challenges and Limitations

The transition toward Agriculture 4.0 has accelerated the adoption of IoT, artificial intelligence, robotics, and cloud computing technologies in greenhouse systems, creating substantial opportunities for greenhouse optimization and sustainable crop production [2], [41]. However, significant barriers continue to limit large-scale implementation and hinder the full realization of these technologies. Technical, economic, organizational, and policy-related challenges remain major obstacles that require coordinated efforts from researchers, technology developers, greenhouse operators, and policymakers [1], [3].

Cybersecurity remains one of the most critical challenges in IoT-enabled greenhouse systems. Security assessments of Controlled Environment Agriculture (CEA) platforms have revealed vulnerabilities associated with weak authentication mechanisms, inadequate encryption, and insecure communication protocols [10], [42]. As greenhouse operations become increasingly dependent on AI-driven decision-making systems, the risk of adversarial attacks, data poisoning, and malicious manipulation of control algorithms becomes more significant. Consequently, robust cybersecurity frameworks are essential for ensuring operational reliability and system resilience [10].

Another major technical challenge concerns data quality and availability. Machine learning models require large volumes of accurately labeled, high-quality data for effective training and deployment. Big-data-driven agricultural systems further emphasize the importance of scalable and structured datasets, as reliable decision-support systems depend heavily on accurate predictive inputs [27], [43]. However, greenhouse datasets remain limited due to proprietary data ownership, system heterogeneity, and the high cost of manual data annotation [27], [29]. Studies on greenhouse microclimate characterization have demonstrated that reliable monitoring of temperature, humidity, and vapor pressure deficit is essential for effective crop management, yet long-term environmental datasets remain scarce in commercial greenhouse operations [27], [44]. Furthermore, simulation-trained models frequently experience performance degradation when deployed in operational environments, a challenge commonly referred to as the simulation-to-reality gap [33].

Interoperability is another persistent challenge affecting greenhouse IoT systems. The absence of standardized data formats, communication protocols, and interfaces limits seamless integration among devices and often results in vendor lock-in [13]. Although standardization initiatives such as the Open Connectivity Foundation (OCF) and the OGC SensorThings API have been developed to address these issues, widespread adoption within the greenhouse industry remains limited [13].

Energy sustainability also presents a significant challenge. While intelligent greenhouse technologies can achieve energy savings ranging from 15–30%, AI computations, cloud processing, and automated control systems simultaneously increase overall energy consumption [38], [45]. Thermal energy storage systems and optimized cooling strategies have demonstrated substantial potential for improving greenhouse efficiency and reducing operating costs in hot and arid climates [45], [46]. Furthermore, integrating renewable energy resources and demand-response mechanisms can reduce energy demand and improve operational sustainability [39]. Nevertheless, limited computational resources at the edge continue to constrain the deployment of advanced AI models [18].

Economic considerations further restrict technology adoption, particularly among small and medium-sized greenhouse operators. The deployment of IoT infrastructure typically requires substantial investment in sensors, communication networks, computing platforms, and maintenance services [47], [48]. Although large-scale operations may achieve favorable returns on investment within a few years, smaller producers often face significant financial and technical barriers. Low-cost platforms such as Arduino and Raspberry Pi have reduced entry costs; however, advanced AI-driven optimization systems remain concentrated among technologically advanced producers [48].

Human factors represent another important but often overlooked dimension of technology adoption. The acceptance of AI-based decision-support systems depends largely on whether these technologies complement rather than replace human expertise and decision-making processes. Human–robot collaboration is becoming increasingly important in autonomous greenhouse operations, where robotic precision must be combined with human experience to achieve reliable and trusted outcomes [30], [49]. Explainable AI frameworks and human-in-the-loop approaches have been proposed as critical mechanisms for improving transparency, trust, and cooperation between growers and intelligent systems. More broadly, successful adoption of artificial intelligence in agriculture depends not only on predictive accuracy but also on user trust, interpretability, and operational feasibility [29], [30], [50].

Table 6 summarizes the major challenges affecting the deployment of smart greenhouse systems and highlights emerging IoT- and AI-driven technological solutions that may address these limitations in future greenhouse operations.

Future research should focus on developing interoperable greenhouse platforms, energy-efficient AI architectures, secure data-sharing mechanisms, and explainable decision-support systems. These advancements will be essential for enabling scalable, trustworthy, and economically viable smart greenhouse

Table 6. Future Directions and Challenges in Smart Greenhouse Systems

Challenge	Potential Solution
Data Heterogeneity	Standardized JSON and MQTT frameworks for interoperability
Energy Constraints	Energy harvesting through solar and thermal technologies
High Initial Cost	Low-cost open-source hardware platforms such as ESP32
Security Risks	Blockchain-based data integrity and secure communication
Limited Explainability	Explainable AI and human-in-the-loop decision support

Note: MQTT = Message Queuing Telemetry Transport; ESP32 = Low-cost microcontroller platform for IoT deployment.

ecosystems capable of supporting sustainable agricultural production.

9. Research Gaps and Future Research Directions

One of the most significant limitations in greenhouse IoT and AI research is the absence of standardized, publicly available datasets for model development and benchmarking [3], [18], [24]. Most studies rely on proprietary datasets collected from individual facilities, specific crop types, or short experimental periods, limiting reproducibility and cross-study comparison [1], [24]. The heterogeneity of greenhouse structures, climate conditions, crop varieties, and sensor configurations further complicates model transferability and generalization. Consequently, machine learning models that demonstrate strong performance in controlled experimental environments frequently experience significant performance degradation when deployed in commercial greenhouse operations [27], [29].

Unlike domains such as medical imaging and autonomous driving, where benchmark datasets have accelerated algorithmic development and standardized performance evaluation, greenhouse optimization lacks widely accepted reference datasets for climate control, disease detection, irrigation management, and yield forecasting [22], [29]. This fragmentation slows scientific progress and makes it difficult to determine whether newly proposed methods genuinely advance the state of the art. Future research should therefore prioritize collaborative multi-institutional dataset development, open benchmarking platforms, standardized evaluation protocols, and reproducible experimental frameworks to facilitate fair comparisons across algorithms and deployment environments [3], [18].

Beyond data availability challenges, several technical and operational barriers continue to impede large-scale adoption of AI-enabled greenhouse systems. Issues such as poor model generalization, simulation-to-reality gaps, limited explainability, interoperability failures, cybersecurity vulnerabilities, energy constraints, and economic scalability remain insufficiently addressed in current research. Although artificial intelligence models frequently demonstrate promising predictive performance under controlled laboratory conditions, their practical implementation in commercial greenhouse environments remains challenging due to these unresolved limitations.

Fig. 4 illustrates the major factors contributing to the gap between experimental AI performance and real-world greenhouse deployment. The figure highlights how data, technological, operational, and economic barriers collectively influence the successful transition of intelligent greenhouse systems from research prototypes to practical agricultural solutions.

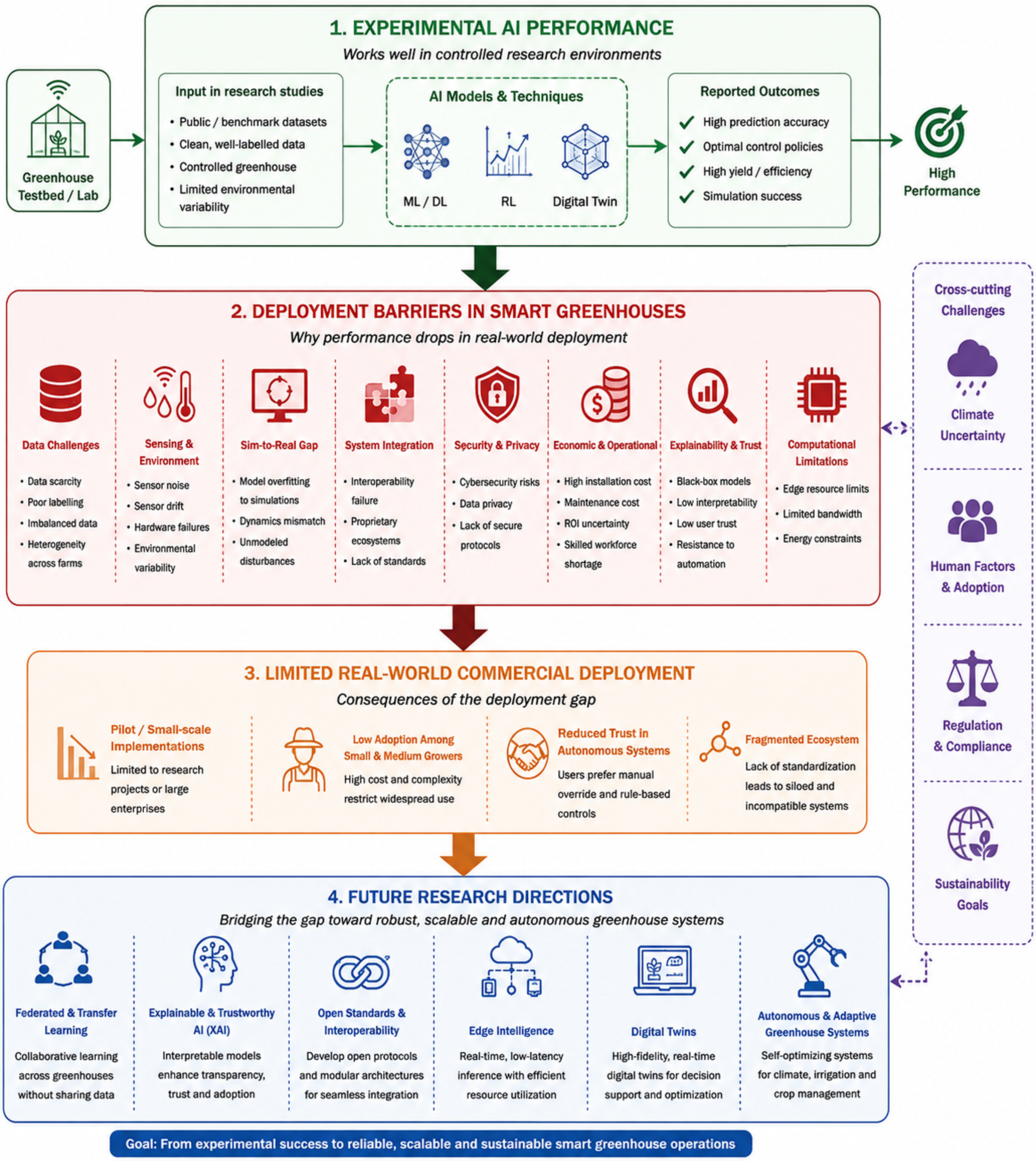

Figure 4. Conceptual framework illustrating the gap between experimental AI performance and real-world greenhouse deployment. The framework highlights key deployment barriers, including data limitations, simulation-to-reality gaps, interoperability challenges, security concerns, economic constraints, and limited explainability, together with future research directions for achieving scalable and reliable smart greenhouse systems.

9.1. Simulation-to-Reality Gap in Digital Twins and Reinforcement Learning

Digital twins and reinforcement learning (RL) have emerged as promising technologies for autonomous greenhouse optimization, enabling advanced climate control, resource management, and operational decision-making [32], [41], [49].Despite their considerable potential, practical deployment remains limited due to the persistent simulation-to-reality (Sim-to-Real) gap. Most RL controllers are developed and trained in virtual environments where greenhouse dynamics, crop responses, and actuator behaviors are represented using simplified and highly predictable models [26], [28]. While such environments facilitate rapid experimentation and algorithm development, they often fail to capture the complexity and uncertainty of real greenhouse operations.

Commercial greenhouse environments are influenced by numerous factors that are difficult to model accurately, including sensor drift, equipment degradation, unpredictable pest and disease outbreaks, labor interventions, communication delays, and sudden weather fluctuations [10], [19]. Consequently, control policies that demonstrate strong performance in simulation may experience substantial performance degradation when transferred to physical facilities. In some cases, poorly generalized policies may generate unstable control actions, inefficient resource utilization, or operational risks that negatively affect crop productivity and profitability.

Digital twins face similar challenges. Although they provide a virtual representation of greenhouse systems, maintaining accurate calibration and continuous synchronization with real-world conditions remains difficult [32], [49]. Model inaccuracies accumulate over time as greenhouse conditions evolve, reducing predictive reliability and limiting the effectiveness of optimization and decision-support functions. Furthermore, the computational requirements associated with high-fidelity digital twin models can restrict real-time deployment, particularly in resource-constrained greenhouse environments.

Current research largely evaluates digital twins and RL-based control systems under controlled experimental conditions, with relatively limited validation in long-term commercial greenhouse operations. As a result, the robustness, scalability, and transferability of these technologies remain insufficiently understood. Future research should focus on hybrid physics-informed AI models, domain adaptation techniques, uncertainty-aware reinforcement learning, transfer learning strategies, and continuous digital twin calibration frameworks. These approaches have the potential to reduce transfer failures, improve model adaptability, and enhance the reliability of autonomous greenhouse systems operating under real-world conditions [28], [33].

9.2. Explainability versus Predictive Accuracy Trade-Off

Deep learning models such as CNNs, LSTMs, GRUs, and transformer-based architectures have demonstrated remarkable predictive performance in greenhouse applications, particularly for climate forecasting, disease detection, and crop growth prediction [24], [27], [29]. However, these models frequently operate as black-box systems, making it difficult for growers and greenhouse operators to understand the rationale behind specific recommendations, predictions, or control actions [30], [31], [43].

This lack of transparency can hinder adoption, especially in high-risk agricultural environments where incorrect decisions may directly affect crop health, resource utilization, and economic outcomes. Although Explainable Artificial Intelligence (XAI) techniques such as SHAP, LIME, and attention-based visualization methods have emerged as promising solutions, most implementations remain confined to research settings and have not been widely integrated into operational greenhouse platforms [29], [30].

A fundamental challenge lies in balancing predictive accuracy and interpretability. Highly accurate models are often computationally complex and difficult to explain, whereas simpler interpretable models may provide lower predictive performance. Future research should therefore focus on practical explainability frameworks that preserve predictive accuracy while enhancing user trust, transparency, and decision support. Such systems should enable growers to interpret, validate, and, when necessary, override AI-generated recommendations [30], [31].

9.3. Cybersecurity and System Vulnerability in Smart Greenhouses

Cybersecurity remains one of the least explored yet most critical dimensions of greenhouse digital transformation [10], [42], [50]. Modern IoT-enabled greenhouse systems increasingly depend on cloud platforms, wireless communication networks, remote-access interfaces, and AI-driven control systems, all of which significantly expand the potential attack surface for malicious actors [2], [13].

Previous studies have identified several security vulnerabilities, including weak authentication mechanisms, insecure device firmware, insufficient encryption, data poisoning attacks, and manipulation of reinforcement learning reward functions [10], [42]. As greenhouse operations become increasingly autonomous, the potential consequences of cyberattacks extend beyond data breaches to include crop losses, operational disruptions, and economic damage.

Compared with industrial manufacturing sectors, greenhouse agriculture currently lacks comprehensive cybersecurity regulations, standardized certification procedures, and systematic incident-reporting mechanisms [10]. This creates a significant mismatch between technological dependence and security preparedness. Future research should therefore move beyond theoretical threat identification toward deployable defense mechanisms, including zero-trust architectures, secure-by-design greenhouse platforms, encrypted device communication, intrusion detection systems, and cybersecurity frameworks aligned with IEC 62443 standards for controlled environment agriculture [10], [41], [50].

9.4. Economic Barriers and Small-Farm Adoption Limitations

Although smart greenhouse technologies offer substantial benefits in terms of productivity, sustainability, and resource efficiency, adoption remains concentrated among large commercial operations with sufficient financial and technical capacity [7], [8]. Small and medium-sized growers continue to face significant barriers associated with sensor acquisition costs, communication infrastructure, maintenance requirements, software licensing fees, and the need for specialized technical expertise [5], [16].

Even when low-cost hardware platforms such as Arduino, Raspberry Pi, and ESP32 are available, integrating these technologies into reliable production-scale greenhouse systems remains challenging [35]. Consequently, a digital divide is emerging in which technologically advanced producers increasingly benefit from AI-driven optimization, while smaller growers continue to rely on conventional management practices.

Addressing this challenge requires future research to focus on scalable and cost-effective greenhouse architectures, modular plug-and-play IoT systems, open-source software platforms, and sustainable business models that reduce implementation costs. In addition, financial support mechanisms, cooperative technology-sharing models, and accessible training programs may play a critical role in expanding adoption among resource-constrained agricultural producers [16], [35].

9.5. Limited Real-World Deployment of Federated Learning

Federated learning has been widely proposed as a privacy-preserving approach for collaborative greenhouse optimization, enabling multiple facilities to benefit from shared model training without exposing sensitive operational data [11], [47], [44]. Despite its considerable potential, most existing studies remain conceptual, simulation-based, or limited to small-scale experimental settings, with relatively few demonstrations in commercial greenhouse networks [11].

Several practical barriers hinder real-world deployment. Greenhouse facilities often differ substantially in crop types, climate conditions, sensor configurations, management strategies, and data quality, creating significant challenges for model convergence and generalization. Additional constraints include unreliable network connectivity, heterogeneous computing resources, and difficulties associated with synchronizing local model updates across facilities operating under different production cycles and environmental conditions.

Trust also remains a critical concern. Although federated learning preserves data privacy, growers may still be reluctant to participate if model behavior lacks transparency or if the benefits of collaboration are unclear. Future research should therefore focus on hierarchical federated learning architectures, edge-assisted aggregation strategies, adaptive privacy-preserving mechanisms, and incentive models that encourage participation while maintaining commercial confidentiality and operational trust [11].

9.6. Interoperability Failure and Vendor Lock-In

Interoperability remains one of the most persistent obstacles to the large-scale adoption of IoT-enabled greenhouse systems. Although standards such as MQTT, OCF, IPSO Alliance protocols, and the OGC SensorThings API have been developed to improve device compatibility and data exchange, implementation across commercial greenhouse environments remains highly fragmented [13], [17]. Most greenhouse operators continue to rely on proprietary ecosystems in which sensors, gateways, cloud platforms, and control systems are designed to operate only within a specific vendor infrastructure.

This vendor-centric architecture creates significant lock-in effects. Growers who invest in a particular platform frequently face high switching costs when attempting to integrate third-party devices, upgrade infrastructure, or migrate to more advanced analytical systems. Incompatible data formats, closed APIs, limited data portability, and inconsistent communication protocols reduce operational flexibility and increase long-term costs [2], [13].

Interoperability limitations also affect scientific reproducibility and technology transfer. Models developed using one sensor ecosystem may not be readily transferable to other greenhouse environments employing different hardware and software configurations. This slows innovation and restricts the development of universally deployable AI solutions. Future research should prioritize open-architecture greenhouse platforms, standardized data models, interoperable middleware layers, and industry-wide standards that reduce dependence on proprietary ecosystems. Achieving true greenhouse intelligence will require not only more sophisticated algorithms but also seamless communication across devices, vendors, facilities, and operational scales [13], [17].

9.7. Future Research Roadmap Toward Autonomous Greenhouses

The ultimate objective of greenhouse optimization research is the realization of fully autonomous greenhouse systems capable of self-monitoring, self-learning, self-adaptation, and self-optimization with minimal human intervention [4], [19], [36], [44]. Achieving this vision will require the convergence of multiple enabling technologies, including advanced IoT infrastructures, explainable artificial intelligence, robotics, digital twins, 5G/6G communication networks, edge intelligence, and human-in-the-loop decision-support systems [23], [30], [36].

Future greenhouse systems should not be designed to replace growers but rather to augment human expertise by combining algorithmic precision with practical agronomic knowledge. Autonomous systems must remain transparent, trustworthy, secure, and adaptable to changing environmental and operational conditions. Consequently, future research should shift beyond isolated technology development and focus on integrated ecosystem design, where sensing, analytics, communication, control, and human interaction are treated as interconnected components of a unified intelligent system.

The next decade of research will likely determine whether greenhouse technologies evolve from partially automated systems to truly autonomous agricultural ecosystems. Progress in interoperability, explainability, cybersecurity, federated learning, digital twins, and edge intelligence will play a central role in bridging the gap between experimental innovation and practical deployment. Ultimately, this transition from technology adoption to system-level intelligence will define the future of autonomous protected agriculture.

10. Conclusion

This review critically examined the evolving role of the Internet of Things (IoT), data analytics, and artificial intelligence in greenhouse optimization, highlighting both the significant technological advances achieved over the past decade and the persistent challenges that continue to limit large-scale deployment. The integration of sensor networks, communication infrastructures, cloud platforms, and machine learning algorithms has transformed greenhouse management into a data-driven process capable of improving productivity, resource efficiency, and operational sustainability. However, despite substantial progress, much of the current research remains focused on algorithmic performance under controlled experimental conditions rather than practical implementation in commercial greenhouse environments.

A key observation from this review is that many widely adopted approaches, including LSTM-based forecasting, deep learning-based crop monitoring, and reinforcement learning-based climate control, often demonstrate strong predictive performance in isolated studies but exhibit limited generalization across diverse greenhouse settings. This limitation is primarily driven by data scarcity, environmental heterogeneity, lack of standardized benchmarking datasets, and insufficient cross-domain validation. Similarly, emerging technologies such as digital twins and federated learning offer considerable potential for intelligent greenhouse management but remain largely confined to research and pilot-scale implementations due to challenges associated with calibration, synchronization, data heterogeneity, and infrastructure requirements.

The review further identified several systemic barriers that constrain practical deployment, including cybersecurity vulnerabilities, interoperability failures, vendor lock-in, high implementation costs, limited explainability of AI models, and resource constraints associated with edge computing. Collectively, these challenges contribute to a significant gap between academic innovation and operational adoption, particularly for small and medium-scale greenhouse producers.

Future research should move beyond model-centric development and focus on deployment-oriented system design. Priority areas include the creation of standardized datasets and benchmarking frameworks, the development of explainable and trustworthy AI systems, the integration of physics-informed and data-driven modeling approaches, and the adoption of interoperable architectures capable of supporting heterogeneous greenhouse environments. Greater collaboration among researchers, technology developers, industry stakeholders, and policymakers will also be essential for addressing economic, regulatory, and operational barriers.

Ultimately, the transition from smart greenhouse systems to fully autonomous protected agriculture will depend not only on advances in artificial intelligence and IoT technologies but also on the successful integration of these innovations into secure, scalable, interoperable, and economically viable solutions. By identifying current limitations, deployment barriers, and future research opportunities, this review provides a foundation for advancing greenhouse intelligence from experimental innovation toward sustainable real-world agricultural impact.

Declarations

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethics Statement

This study did not involve human participants, animals, or identifiable personal data. All information presented in this review was obtained from previously published studies and publicly available sources.

Author Contributions

Author 1: Conceptualization, methodology, literature review, data curation, and writing—original draft preparation.

Author 2: Supervision, review, editing, and project administration.

Author 3: Visualization, validation, and manuscript review.

All authors contributed to manuscript revision, technical discussions, validation of content, and approval of the final version of the manuscript.

Declaration of Generative AI and AI-Assisted Technologies

During the preparation of this manuscript, the authors used AI-assisted tools to improve language clarity, organization, and readability. The authors carefully reviewed, edited, and validated all generated content and assume full responsibility for the content of the published work.

Acknowledgment

The authors acknowledge the contributions of researchers whose published work formed the foundation of this review. No external funding was received for this study.

References

1. K. Li, J. Shi, C. Hu, and W. Xue, “The intelligentization process of agricultural greenhouse: A review of control strategies and modeling techniques,” *Agriculture*, vol. 15, no. 20, p. 2135, 2025.
2. A. Tzounis, N. Katsoulas, T. Bartzanas, and C. Kittas, “Internet of things in agriculture, recent advances and future challenges,” *Biosystems engineering*, vol. 164, pp. 31–48, 2017.
3. P. Yu, F. Teng, W. Zhu, C. Shen, Z. Chen, and J. Song, “Cloud–edge–device collaborative computing in smart agriculture: architectures, applications, and future perspectives,” *Frontiers in Plant Science*, vol. 16, p. 1668545, 2025.

4. I. Hindi, A. Alsharkawi, M. Al-Ajlouni, and B. Qarallah, “Enhancing autonomous agriculture control systems in greenhouses for sustainable resource usage using deep learning techniques,” *Plos one*, vol. 21, no. 3, p. e0344946, 2026.

5. J. Falinski, “Msu scientists study how to increase water and nutrient-use efficiency in greenhouses and nurseries,” Sep. 2024, accessed: April 2026. [Online]. Available: https://msutoday.msu.edu/news/2024/09/msu-scientists-study-how-to-increase-water-nutrient-use-efficiency-in-greenhouses-nurseries

6. E. Kueneman, J. Hopmans, E. Raser, and C. S. Martin, “Water-smart agriculture: A biophysical-focused introduction: addressing needs and opportunities,” 2020.

7. Technavio, “Commercial greenhouse market analysis: Size and forecast 2024–2028,” Technavio, Tech. Rep. IRTNTR75762, 2024, accessed: April 2026. [Online]. Available: https://www.technavio.com/report/commercial-greenhouse-market-analysis

8. Grand View Research, “Smart greenhouse market size, share and trends analysis report,” 2026, accessed: April 2026. [Online]. Available: https://www.grandviewresearch.com/industry-analysis/smart-greenhouse-market-report

9. GlobeNewswire, “Deep learning market size to surpass $296 billion by 2031,” 2026, accessed: April 2026. [Online]. Available: https://www.globenewswire.com/news-release/2026/02/23/3242764

10. A. Vakhnovskyi, “Threat modeling and attack surface analysis of iot-enabled controlled environment agriculture systems,” *arXiv preprint arXiv:2604.13308*, 2026.

11. ——, “Hierfedcea: Hierarchical federated edge learning for privacy-preserving climate control optimization across heterogeneous controlled environment agriculture facilities,” *arXiv preprint arXiv:2604.13396*, 2026.

12. D. J. A. Rustia, L.-Y. Chiu, C.-Y. Lu, Y.-F. Wu, S.-K. Chen, J.-Y. Chung, J.-C. Hsu, and T.-T. Lin, “Towards intelligent and integrated pest management through an aiot-based monitoring system,” *Pest management science*, vol. 78, no. 10, pp. 4288–4302, 2022.

13. E. Roccatello, A. Pagano, N. Levorato, and M. Rumor, “State of the art in internet of things standards and protocols for precision agriculture with an approach to semantic interoperability,” *Network*, vol. 5, no. 2, p. 14, 2025.

14. T. C. de Araújo, L. T. Silva, and A. J. Moreira, “Deviation prediction and correction on low-cost atmospheric pressure sensors using a machine-learning algorithm.” in *SENSORNETS*, 2020, pp. 41–51.

15. Z. Li, J. Wang, R. Higgs, L. Zhou, and W. Yuan, “Design of an intelligent management system for agricultural greenhouses based on the internet of things,” in *2017 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC)*, vol. 2, 2017, pp. 154–160.

16. C. Volosciuc, R. Bogdan, B. Blajovan, C. Stângaciu, and M. Marcu, “Greenlab, an iot-based small-scale smart greenhouse,” *Future Internet*, vol. 16, no. 6, 2024. [Online]. Available: https://www.mdpi.com/1999-5903/16/6/195

17. South Dakota State University Extension, “Choosing the right wireless network technologies for agricultural iot,” 2024, accessed: April 2026. [Online]. Available: https://extension.sdstate.edu/

18.J. Shu and D. Yang, “Cnn-lstm-pot-based anomaly detection for smart greenhouse sensor data: A real-time edge deployment approach,” *Future Internet*, vol. 18, no. 4, 2026. [Online]. Available: https://www.mdpi.com/1999-5903/18/4/205

19.S. C. Maree, P. Zhang, B. M. van Marrewijk, F. de Zwart, M. Bijlaard, and S. Hemming, “Autonomous greenhouse cultivation of dwarf tomato: Performance evaluation of intelligent algorithms for multiple-sensor feedback,” *Sensors*, vol. 25, no. 14, 2025. [Online]. Available: https://www.mdpi.com/1424-8220/25/14/4321

20.S. Hemming, F. de Zwart, A. Elings, A. Petropoulou, and I. Righini, “Autonomous greenhouse challenge, (2019),” 2020.

21.M. Hasan, A. G. Bhat, V. K. S, M. R, A. K. Yadav, A. Saji, B. G, and S. A. Prakash, “Iot-based smart fertigation scheduling and wireless microclimate monitoring for a greenhouse dutch bucket hydroponic system,” *Irrigation and Drainage*, vol. 74, no. 5, pp. 1889–1903, 2025.

22.A. Upadhyay, N. S. Chandel, K. P. Singh, S. K. Chakraborty, B. M. Nandede, M. Kumar, A. Subeesh, K. Upendar, A. Salem, and A. Elbeltagi, “Deep learning and computer vision in plant disease detection: a comprehensive review of techniques, models, and trends in precision agriculture,” *Artificial Intelligence Review*, vol. 58, no. 3, p. 92, 2025.

23.Vertical Farm Daily, “5g in the greenhouse: When robots and bumblebees share the same signal,” Nov. 2025, accessed: April 2026. [Online]. Available: https://www.verticalfarmdaily.com/article/9781380/5g-in-the-greenhouse-when-robots-and-bumblebees-share-the-same-signal/

24.S. E. Arnaud, M. Calisti, and A. Polydoros, “Data-driven greenhouse climate regulation in lettuce cultivation using bilstm and gru predictive control,” *arXiv preprint arXiv:2507.21669*, 2025.

25.R. Sarayani, S. Kouravand, J. Massah, M. Dehghani-Soufi, T. Al-Shukaili, and M. Pasdar, “Design, development and technical evaluation of an intelligent drip irrigation system based on drainage percentage in soilless culture,” *Smart Agricultural Technology*, p. 101964, 2026.

26.B. van Laatum, S. Msaad, E. J. van Henten, R. D. Mcallister, and S. Boersma, “Stochastic model predictive control with reinforcement learning for greenhouse production systems under parametric uncertainty,” *Control Engineering Practice*, vol. 169, p. 106787, 2026.

27.J. Yu, C. Sun, J. Zhao, L. Ma, W. Zheng, Q. Xie, and X. Wei, “Prediction and control of greenhouse temperature: Methods, applications, and future directions,” *Computers and Electronics in Agriculture*, vol. 237, p. 110603, 2025.

28.H. M. Abbood, S. H. S. Alagheband, A. M. Imran, S. M. Ali, and M. A. Hassan, “Design adaptive non-linear pid control using reinforcement learning for optimal autonomous greenhouse microclimate regulation,” *Journal of Advanced Computational Intelligence and Intelligent Informatics*, vol. 29, no. 6, pp. 1464–1483, 2025.

29.Y. Li, T. Wang, H. Zu, S. Liu, and L. Zu, “Explainable ai-driven interpretation of environmental drivers of tomato fruit expansion in smart greenhouses using iot sensing,” *Scientific Reports*, vol. 15, no. 1, p. 40870, 2025.

30.D. G. Pai, M. Balachandra, and R. Kamath, “Explainable ai in agriculture: Review of applications, methodologies, and future directions,” *Engineering Research Express*, vol. 7, no. 3, p. 032202, 2025.

31. M. J. Bashir, S. Henna, and E. Furey, “Explainable ai for smart greenhouse control: Interpretability of temporal fusion transformer in the internet of robotic things,” in *2025 35th Irish Signals and Systems Conference (ISSC)*. IEEE, 2025, pp. 1–7.

32. D. A. Howard, Z. Ma, C. Veje, A. Clausen, J. M. Aaslyng, and B. N. Jørgensen, “Greenhouse industry 4.0–digital twin technology for commercial greenhouses,” *Energy Informatics*, vol. 4, no. Suppl 2, p. 37, 2021.

33. A. Khdoudi, T. Masrour, I. El Hassani, and C. El Mazgualdi, “A deep-reinforcement-learning-based digital twin for manufacturing process optimization,” *Systems*, vol. 12, no. 2, p. 38, 2024.

34. Y. Lyu, C. Cheng, X. Chen, S. Tang, S. Chen, X. Guan, L. Wu, Z. Liang, Y. Zhu, and G. Xia, “Growth simulation model and intelligent management system of horticultural crops: Methods, decisions, and prospects,” *Horticulturae*, vol. 12, no. 2, p. 139, 2026.

35. I. Fathi, “An iot-based low-cost smart greenhouse monitoring system using esp8266 and firebase for real-time environmental control,” *International Journal of Advanced Natural Sciences and Engineering Researches*, vol. 9, no. 5, pp. 268–286, 2025.

36. Wageningen University & Research, “Wur develops simulated greenhouse environment for faster robot development,” 2026, accessed: April 2026. [Online]. Available: https://www.wur.nl/en/news/wur-develops-simulated-greenhouse-environment-faster-robot-development

37. M. Jomaa, M. Abbes, F. Tadeo, and A. Mami, “Greenhouse modeling, validation and climate control based on fuzzy logic,” *Engineering, Technology & Applied Science Research*, vol. 9, no. 4, pp. 4405–4410, 2019.

38. Y. Cheng, S. Yang, S. Sun, P. Yu, and J. Xing, “Energy management for integrated energy system based on coordinated optimization of electric–thermal multi-energy retention and reinforcement learning,” *Processes*, vol. 13, no. 9, p. 2693, 2025.

39. M. A. Ali, R. K. Dhanaraj, and S. Kadry, “Ai-enabled iot-based pest prevention and controlling system using sound analytics in large agricultural field,” *Computers and Electronics in Agriculture*, vol. 220, p. 108844, 2024.

40. Agromedia Agency, “Ai-powered autonomous robots in greenhouse farming,” 2025, accessed: April 2026. [Online]. Available: https://agromediaagency.livejournal.com/120855729.html

41. M. Javaid, A. Haleem, R. P. Singh, and R. Suman, “Enhancing smart farming through the applications of agriculture 4.0 technologies,” *International Journal of Intelligent Networks*, vol. 3, pp. 150–164, 2022.

42. E. M. Pasca, D. Delinschi, R. Erdei, I. Baraian, and O. D. Matei, “A vulnerable-by-design iot sensor framework for cybersecurity in smart agriculture,” *Agriculture*, vol. 15, no. 12, p. 1253, 2025.

43. V. M. Ngo, T.-V. T. Duong, T.-B.-T. Nguyen, C. N. Dang, and O. Conlan, “A big data smart agricultural system: recommending optimum fertilisers for crops,” *International Journal of Information Technology*, vol. 15, no. 1, pp. 249–265, 2023.

44. L. M. Albuja-Illescas, M. Gómez-Cabezas, G. Jácome-Aguirre, J. P. Aragón-Suárez, R. Jiménez-Lao, A. Peña-Fernández, and M. T. Lao, “Microclimate characterization of a low-tech greenhouse during a tomato crop (solanum lycopersicum l.) production cycle in chaltura, imbabura,” *Plants*, vol. 14, no. 23, p. 3702, 2025.

45. A. Mehtab, H.-S. Mun, E. B. Lagua, H.-R. Park, J.-G. Kang, M. Sharifuzzaman, M. K. Hasan, Y.-H. Kim, S.-B. Ryu, and C.-J. Yang, "Thermal energy storage for sustainable smart agricultural facilities: Design, integration, control, environmental impacts, and future perspectives," *Sustainability*, vol. 18, no. 3, p. 1311, 2026.
46. S.-G. Choi, D.-Y. Park, D.-S. Choi, and Y.-H. Jung, "Performance analysis of thermal energy storage tanks and chillers for optimizing cooling efficiency in smart greenhouses in hot and arid climates," *Sustainability*, vol. 16, no. 24, p. 11136, 2024.
47. F. Hollinger and B. Sener, "Digital technologies for agriculture in türkiye," 2025.
48. V. Oguntosin, C. Okeke, E. Adetiba, A. Abdulkareem, and J. Olowoleni, "Iot-based greenhouse monitoring and control system," 2023.
49. L. Benos, V. Moysiadis, D. Kateris, A. C. Tagarakis, P. Busato, S. Pearson, and D. Bochtis, "Human–robot interaction in agriculture: a systematic review," *Sensors*, vol. 23, no. 15, p. 6776, 2023.
50. S. Sharma, K. Verma, and P. Hardaha, "Implementation of artificial intelligence in agriculture," *Journal of Computational and Cognitive Engineering*, vol. 2, no. 2, pp. 155–162, 2023.