An IoT-Driven Machine Learning Framework for Predicting Fungal Infections in Greenhouse Crops

Amaefule I. A., Onyenweuwa Basil O

Department of Computer Science Imo State University Owerri, Imo State-Nigeria

DOI: https://doi.org/10.51584/IJRIAS.2026.11060274

Received: 28 June 2026; Accepted: 03 July 2026; Published: 16 July 2026

ABSTRACT

Fungal infections remain one of the major causes of reduced productivity and economic loss in greenhouse agriculture due to delayed detection and inadequate environmental monitoring. Conventional disease management approaches are largely reactive and depend heavily on manual inspection, resulting in late intervention and increased fungicide usage. This study focused on Internet of Things (IoT)-driven machine learning framework for predicting fungal infections in greenhouse crops using simulated IoT environmental sensing data and predictive analytics. The framework combined IoT sensors for constant acquisition of crucial greenhouse parameters such as temperature, humidity, soil moisture, and light intensity. Sensor data are transmitted through a wireless communication layer to a centralized prediction engine where machine learning algorithms are employed for disease risk classification. The three learning models, such as Random Forest (RF), Support Vector Machine, and Convolutional Neural Network (CNN), were comparatively evaluated to determine the most effective predictive approach for greenhouse fungal infection monitoring. Experimental evaluation using a synthetic greenhouse dataset demonstrated that the Random Forest model achieved the best predictive performance with an accuracy of 91.5%, outperformed SVM and CNN models in classification stability and generalization proficiency. The proposed system provides an intelligent early warning mechanism capable of supporting proactive disease management, minimizing crop losses, and improving greenhouse productivity. The study contributes to smart agriculture research by providing an integrated IoT-machine learning framework for real-time greenhouse disease prediction and decision support.

Keywords: IoT; Smart Agriculture; Machine Learning; Greenhouse Monitoring; Fungal Infection Prediction; Precision Agriculture; Random Forest; Environmental Sensing

INTRODUCTION

Agriculture remains one of the most important sectors supporting global food security and economic sustainability, particularly in developing economies where crop productivity directly affects livelihood stability and national development. However, changes in climatic conditions, unfavorable environmental challenges, and frequent disease outbreaks continue to threaten global agricultural productivity.

Among these challenges, fungal infections constitute a major source of crop damage in greenhouse farming systems because pathogens thrive rapidly under humid and temperature- Controlled conditions (Delfani et al., 2024). Greenhouse agriculture has become increasingly important in modern precision farming due to its ability to regulate environmental conditions for improved crop yield and resource efficiency. Nevertheless, greenhouse environments may unintentionally encourage fungal disease propagation when parameters such as humidity, temperature, and soil moisture are not properly monitored and controlled (Sandya De Alwis, S., Hou, L., & Sun, C. (2022). Diseases such as powdery mildew, gray mold, and leaf spot infections significantly reduce crop quality and productivity while increasing operational costs for greenhouse farmers. Traditional disease management techniques are largely reactive and depend on manual inspection, visual symptom identification, and excessive fungicide application.

These approaches are often inefficient because visible symptoms usually appear only after infection has substantially progressed. Excessive fungicide usage additionally contributes to environmental pollution, increased production cost, and pathogen resistance (Delfani et al., 2024). Consequently, there is an increasing demand for intelligent predictive systems capable of identifying disease risk at early stages before severe crop damage occurs.

Recent advancements in the Internet of Things (IoT) have enabled the deployment of interconnected smart sensing devices for continuous agricultural monitoring. IoT technologies facilitate real-time acquisition of environmental parameters including temperature, humidity, soil moisture, and light intensity, thereby improving environmental awareness in greenhouse systems (Sandya De Alwis et al., 2022; Attar et al., 2024). IoT-enabled greenhouse systems support precision agriculture by providing automated environmental monitoring and data-driven decision support mechanisms.

Despite these advancements, environmental monitoring alone is insufficient for intelligent disease prevention without predictive analytical capability. Machine learning techniques have therefore emerged as effective tools for agricultural prediction and disease analysis because of their ability to identify nonlinear relationships within multidimensional datasets. Learning models like Random Forest (RF), Support Vector Machine (SVM), and Convolutional Neural Network (CNN) have displayed promising performance in agricultural disease prediction and classification Tasks (Contreras-Castillo et al, 2023).

Most recent studies have utilized the integration of IoT and machine learning for smart agriculture application. Contreras-Castillo et al. (2023) proposed an IoT-CNN framework for greenhouse monitoring and early disease identification, achieving over 90% disease classification accuracy. Similarly, Sridevi and Preethi (2023) developed a fog-based IoT agricultural prediction system using ensemble learning techniques for smart farming analytics.

However, many existing systems focus primarily on environmental monitoring rather than proactive fungal infection prediction. Furthermore, several proposed frameworks lack comprehensive comparative evaluation of machine learning algorithms, real-time predictive capability, and scalable intelligent decision-support mechanisms suitable for greenhouse deployment. To address these limitations, this study proposes an IoT-driven machine learning framework for early prediction of fungal infections in greenhouse crops. The proposed framework integrates environmental sensing infrastructure, wireless communication mechanisms, and predictive machine learning algorithms for real-time greenhouse disease analysis. The study analytically evaluates Random Forest (RF), Support Vector Machine (SVM), and Convolutional Neural network (CNN) models to determine the most effective predictive approach for fungal infection classification and early warning generation.

Whereas IoT and machine learning technologies have been widely implemented in greenhouse monitoring and plant disease detection, most existing studies focus on environmental monitoring or image-based diagnosis after symptoms become visible. Limited attention has been given to early fungal infection prediction using environmental sensing data and comparative evaluation of multiple machine learning models. To address this gap, this study proposes an IoT-driven machine learning framework that integrates real-time environmental sensing with predictive analytics for early fungal infection risk forecasting and compares the performance of Random Forest, Support Vector Machine, and Convolutional Neural Network models for greenhouse disease prediction.

The major contributions of this study include:

1. Development of an IoT-enabled greenhouse environmental monitoring framework for real- time data acquisition.
2. Integration of machine learning algorithms for predictive fungal infection analysis using greenhouse environmental parameters.

3. Comparative evaluation of Random Forest, Support Vector Machine, and Convolutional Neural Network models for greenhouse disease prediction.
4. Design of an intelligent early warning mechanism for proactive greenhouse disease management

a. Research Objectives

The main objective of this study is to develop an IoT-driven machine learning framework for the early prediction of fungal infections in greenhouse crops.

The specific objectives are to:

1. Design an IoT-based greenhouse monitoring system for real-time environmental data gathering.

2. Develop a predictive machine learning model for fungal infection risk classification using environmental variables.

3. Measure and compare the performance of RF, SVM, and CNN Network models.

4. Develop an intelligent early warning mechanism for proactive greenhouse disease management.

LITERATURE REVIEW

a. IoT-Based Smart Agriculture Systems

The Internet of Things (IoT) has significantly transformed modern agricultural practices through the deployment of interconnected sensing devices capable of real-time environmental monitoring and automated decision-making. IoT-assisted smart agriculture systems improve farming efficiency by continuously monitoring environmental parameters like Temperature, humidity, Soil Moisture, and light intensity These technologies support precision agriculture by enabling farmers to make data-driven decisions for improved crop productivity and disease management. Recent studies have demonstrated the effectiveness of IoT frameworks in greenhouse monitoring and environmental control. Attar et al. (2024) developed an IoT-based smart agriculture monitoring system capable of real-time environmental data acquisition and remote supervision of greenhouse operations. Their study showed that IoT infrastructures improve environmental awareness and operational efficiency in controlled farming environments. Similarly, Shaikh, T. A., Rasool, T., & Lone, F. R. (2022). proposed an intelligent greenhouse monitoring framework integrating wireless sensor networks and cloud-based agricultural analytics for environmental management. Their findings revealed that continuous environmental sensing significantly enhances greenhouse condition monitoring and agricultural automation. Furthermore, Benos et al. (2021) highlighted the growing importance of IoT and artificial intelligence in precision agriculture, emphasizing that real-time environmental sensing can improve resource utilization, reduce operational costs, and support sustainable farming practices.

The study concluded that IoT-enabled agriculture systems represent an important technological advancement toward intelligent food production systems. Despite these advances, several IoT agriculture systems primarily focus on environmental monitoring without integrating predictive disease intelligence mechanisms capable of proactively identifying fungal infection risks before symptom manifestation.

b. Machine Learning for Plant Disease Prediction

Machine learning techniques have increasingly been applied in agricultural prediction systems due to their ability to identify hidden patterns and nonlinear relationships within multidimensional datasets. Predictive learning algorithm such as random Forest, Support vector machine, Artificial Neural Network and Convolutional neural network have shown promising performance in agricultural disease classification and predictive analytics.

Contreras-Castillo et al. (2023) developed a Convolutional Neural Network-based crop disease prediction framework using greenhouse environmental and image datasets. Their model achieved high disease classification accuracy and demonstrated the capability of deep learning techniques in agricultural disease detection. Similarly, Boulent et al. (2019) investigated deep learning methods for plant disease identification and reported that CNN-based models significantly improve disease classification performance compared to traditional machine learning approaches. In another study, Sandya De Alwis et al. (2022) implemented machine learning algorithms for smart agricultural disease diagnosis using environmental sensor data and predictive analytics. Their study demonstrated that Random Forest algorithms provide robust performance in agricultural classification tasks due to their ability to handle noisy environmental datasets and nonlinear feature relationships. Support Vector machine have also been successfully incorporated in agricultural prediction systems.

According to Brahimi et al. (2018), SVM models exhibit strong classification performance in crop disease prediction because of their capability to separate multidimensional feature spaces using optimized decision boundaries. Although existing studies have demonstrated the effectiveness of machine learning in agricultural disease detection, many approaches rely heavily on image-based disease identification after visible symptoms appear. Limited attention has been given to predictive environmental analysis for early fungal infection forecasting in greenhouse environments.

c. Greenhouse Environmental Monitoring Systems

Greenhouse monitoring systems are essential for maintaining environmental conditions favorable for crop growth while minimizing disease propagation. Environmental parameters such as humidity, temperature, soil moisture, and airflow significantly influence fungal infection development in greenhouse environments. Consequently, several studied have investigated automated greenhouse monitoring system for environmental control and disease management.

Li et al. (2021) reviewed greenhouse environmental monitoring technologies based on Internet of Things (IoT) and reported that wireless sensing, multi-parameter environmental monitoring, and predictive analytics significantly improve greenhouse management and support early disease prevention through continuous environmental surveillance. Their findings showed that real-time environmental sensing improves disease prevention capability by identifying greenhouse conditions associated with fungal propagation. Similarly, Shamshiri et al. (2018) investigated advanced greenhouse monitoring technologies for precision agriculture applications and concluded that environmental monitoring systems significantly enhance greenhouse productivity and sustainability. Their study emphasized the importance of integrating environmental sensing with intelligent analytical systems for optimized greenhouse management.

Sridevi and Preethi (2023) additionally proposed a fog-assisted greenhouse monitoring architecture integrating IoT sensing and machine learning prediction mechanisms for smart agriculture applications. Their study demonstrated improved system responsiveness and predictive capability in agricultural monitoring environments.

MATERIALS AND METHODS

a. Proposed Framework Architecture

This study presents an IoT-driven machine learning framework designed for early prediction of fungal infections in greenhouse crops using real-time environmental monitoring and predictive analytics. The proposed framework integrates environmental sensing devices, wireless communication infrastructure, centralized data processing, machine learning prediction models, and an intelligent alert mechanism for proactive greenhouse disease management.

The framework consists of five major operational layers:

1. Environmental sensing layer,

2. IoT communication layer,

3. Data processing layer,

4. Machine learning prediction layer,

5. Alert and decision-support layer.

The environmental sensing layer continuously monitors greenhouse environmental parameters associated with fungal disease development; include temperature, relative humidity, soil moisture and light intensity. The collected environmental data are transmitted through the IoT communication layer to the centralized processing unit where pre-processing and predictive analysis are performed.

The data processing layer involves data cleaning, normalization, feature extraction, and dataset preparation before predictive classification process. The processed dataset was finally analyzed using random Forest (RF), Support Vector Machine (SVM), and Convolutional Neural network (CNN) models for fungal infection risk prediction.

The prediction and decision-support layer generates intelligent alerts whenever greenhouse environmental conditions indicate elevated fungal infection risk. The system generates alerts for greenhouse operators to support timely intervention before severe disease outbreaks occur.

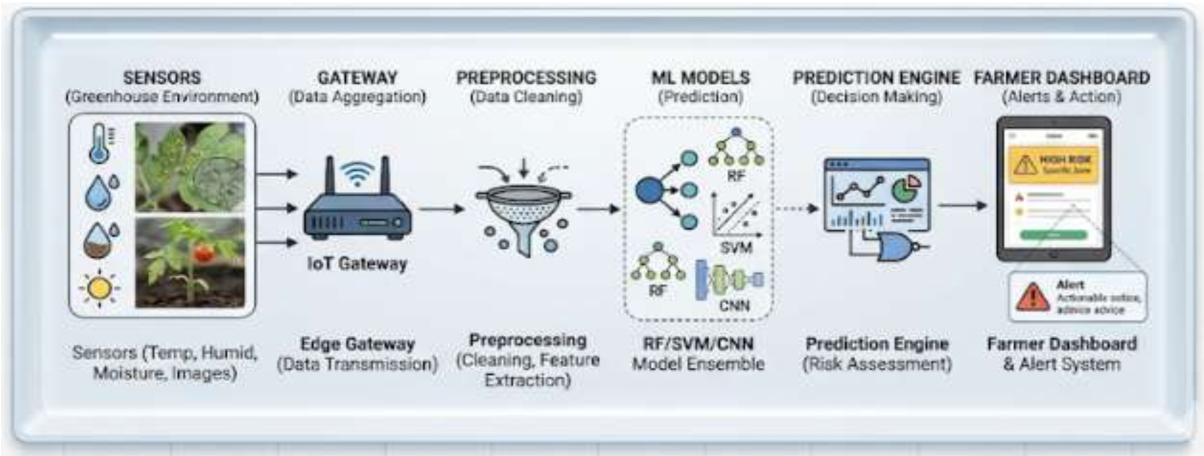

Figure 1. Proposed IoT-Driven Machine Learning Framework for Early Prediction of Fungal Infections in Greenhouse Crops

The proposed architecture shown in Figure 1 illustrates the interaction between greenhouse sensor nodes, IoT communication infrastructure, machine learning prediction models, and the intelligent alert mechanism for proactive disease management.

b. Greenhouse Sensor Configuration

The proposed greenhouse monitoring environment was designed to include multiple IoT sensor nodes for continuous acquisition of environmental parameters associated with fungal disease propagation. The proposed sensors were strategically positioned within the greenhouse to ensure effective environmental coverage and accurate data collection.

The monitored environmental variables include: Temperature, Relative humidity, Soil moisture, Light intensity. Temperature and humidity sensors were used to monitor atmospheric conditions within the greenhouse environment, while soil moisture sensors measured water availability within the root zone. Light intensity sensors additionally evaluated illumination conditions influencing crop development and fungal activity. Environmental parameters were simulated based on sensor-generated greenhouse observations and transmitted through wireless communication protocols to the centralized processing unit for predictive analysis. Continuous environmental monitoring enabled real-time acquisition of greenhouse conditions necessary for fungal infection prediction.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VI June 2026

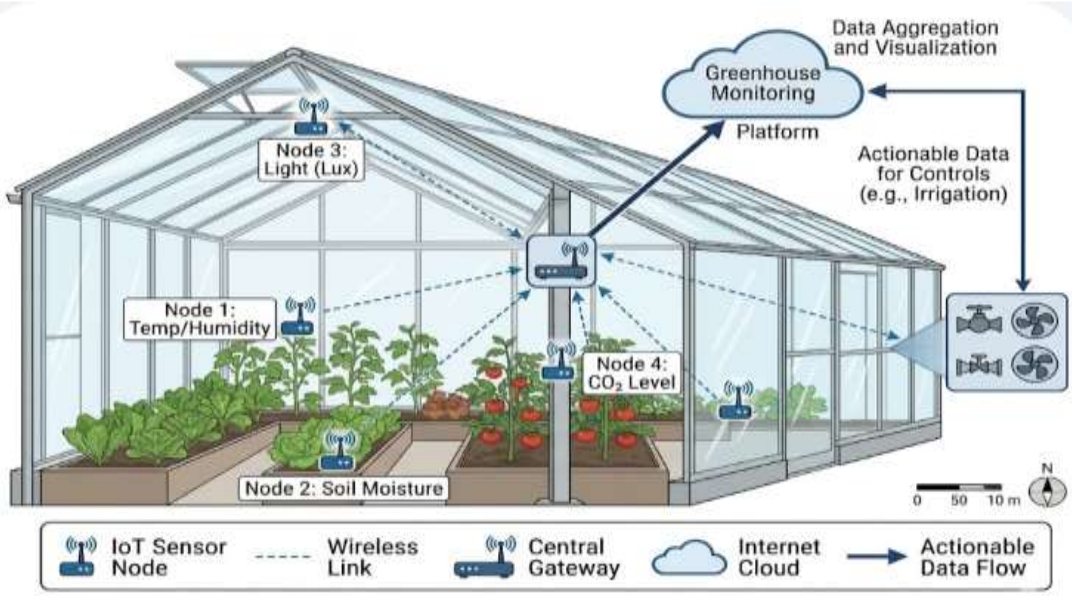

Figure 2. Proposed Greenhouse IoT Sensor Configuration

Figure 2 illustrates the physical arrangement of IoT sensor nodes within the greenhouse environment for environmental data acquisition and transmission. Spatial Distribution: sensors are strategically placed at multiple levels roof (Light), canopy (Temp/CO2), and soil (Moisture) to capture comprehensive micro-climate data. Wireless Connectivity: individual nodes use Wireless Links to transmit data to a Central Gateway, reducing cabling complexity and ensuring a robust star-topology network. Automation Loop: aggregated data is sent to a Cloud Platform to generate actionable insights, which trigger actuator like irrigation and fans for-time environmental control

Technical Note: The inclusion of a scale bar and orientation indicator (North arrow) in the schematic ensures that the deployment is reproducible and that the spatial relationship between the sensors is precisely defined for academic documentation.

Table 1. Environmental Sensors Used in the Proposed System

Sensor Type	**Measured Parameter**	**Purpose**
DHT22	Temperature/Humidity	Environmental monitoring
Soil Moisture Sensor	Soil Moisture	Root-Zone monitoring
Light sensor	Light Intensity	Illumination analysis

c. Data Acquisition and Dataset Description

A synthetic greenhouse environmental dataset was generated for the experimental validation of the proposed framework. The dataset was designed to simulate environmental conditions commonly encountered in greenhouse cultivation of tomato (Solanum lycopersicum) and leafy vegetable crops. The generated observations reflect realistic greenhouse environmental variations associated with fungal disease development and were subsequently used for machine learning model training and evaluation.

A synthetic greenhouse environmental dataset comprising 4,500 observations was generated to simulate environmental conditions commonly encountered in greenhouse crop production. The synthetic dataset was generated to emulate environmental conditions commonly observed in greenhouse cultivation systems used for tomato and leafy vegetable production in Nigeria. The dataset included temperature, humidity, soil moisture, and light intensity variables and was used to evaluate the proposed machine learning framework for fungal infection risk prediction. The greenhouse environmental dataset was labeled based on fungal infection risk conditions identified from established agricultural disease prediction studies. Environmental conditions associated with excessive humidity, unstable temperature ranges, and elevated moisture levels were classified as high-risk conditions for fungal infection propagation.

Environmental observations were categorized into High-Risk and Low-Risk classes based on disease-conducive environmental conditions reported in greenhouse disease prediction literature. Observations exhibiting relative humidity above 85%, temperatures between 22°C and 28°C, and elevated soil moisture levels were classified as High-Risk conditions. Outcomes beside these thresholds were classified as Low-Risk conditions. These labels were as used as target pointer variable for supervised machine learning training and evaluation.

Table 2: Greenhouse Environmental Dataset Description

Features	Description
Temperature	Greenhouse atmospheric temperature
Humidity	Relative humidity level
Soil Moisture	Soil water current
Light intensity	Greenhouse illumination level
Risk Label	Fungal infection risk category

Table 3: Dataset Distribution by Risk Category

Risk category	**Number of Sample**	**Percentage**
High risk	2,050	45.6%
Low Risk	2,450	54.4%
Total	4500	100%

d. Data Pre-processing

Data pre-processing was conducted to improve dataset quality, eliminate inconsistencies, and enhance machine learning prediction accuracy. Environmental datasets collected from greenhouse sensor nodes contained noise, incomplete records, and variations requiring normalization before model training. The pre-processing procedures included: Data cleaning, Missing value handling, Noise reduction, Feature normalization, Dataset labeling. Missing environmental records resulting from communication interruptions or sensor inconsistencies were handled using interpolation techniques. Min-Max normalization was subsequently applied to environmental parameters to improve classification convergence and reduce scale-related bias during machine learning training. The normalized dataset was partitioned into training and testing subsets using supervised learning procedures for predictive model evaluation.

e. Machine Learning Prediction Models

Three supervised machine learning algorithms were implemented and comparatively evaluated for fungal infection prediction: Random Forest (RF), Support Vector Machine (SVM), and Convolutional Neural Network (CNN). The machine learning workflow involved: Feature extraction, model training, classification, and predictive evaluation. Although CNN models are traditionally applied to image-based classification problems, they were included in this study to investigate their capability for extracting complex feature relationships from transformed environmental sensor datasets.

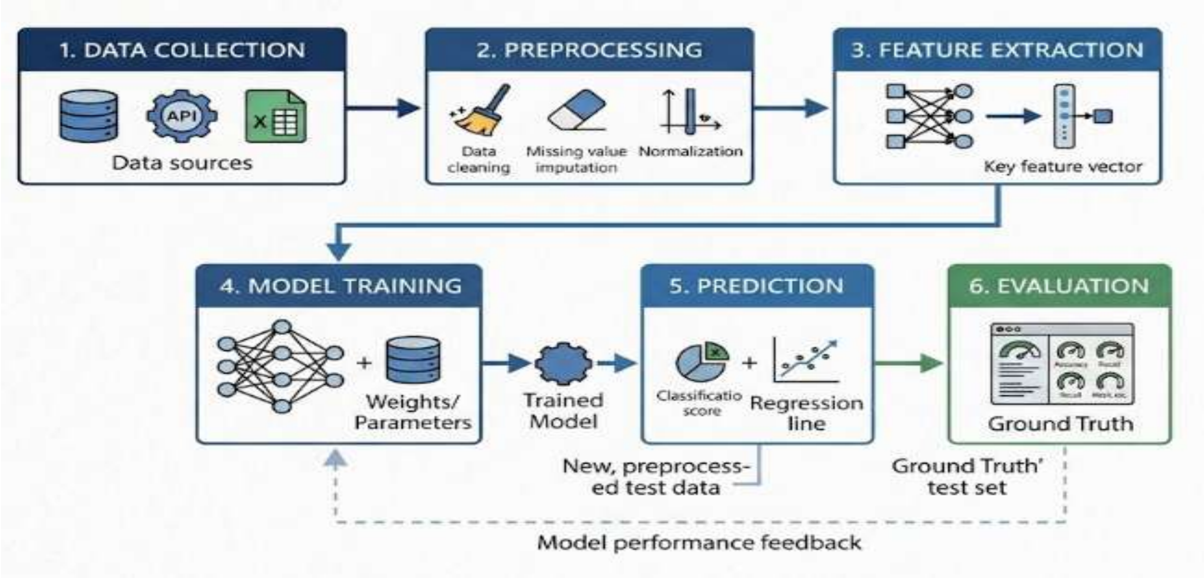

Figure 3. Machine Learning Prediction Workflow

Figure 3 illustrates the Machine Learning Prediction Workflow which outlines the systematic progression of data from raw environmental input to actionable insights. This modular architecture ensures transparency and reproducibility in the fungal infection forecasting process.

1. Data Preparation: The workflow begins with multi-source data Collection, followed by rigorous pre-processing (cleaning and normalization) to ensure high-quality input.

2. Core Pre-processing: feature Extraction isolates critical variables into a vector, which is then used for supervised Model Training to established predictive parameters

3. Output & Validation: the TRAINED Model generates classification scores and regression trends, which are finally verified against labeled validation data to ensure accuracy.

4. Optimization: A feedback loop from the stage allows for continues model refinement and performance enhancement

f. Experimental Setup

The experimental implementation was conducted using Python-based machine learning libraries and synthetic greenhouse environmental datasets generated for framework evaluation The machine learning models were trained and evaluated using supervised classification procedures.

The experimental environment included:

1. Python programming language,

2. Scikit-learn library,

3. TensorFlow framework,

4. Environmental sensor datasets,

5. Centralized prediction system.

The dataset was divided into training and testing subsets using standard supervised learning procedures to evaluate predictive generalization capability. Comparative Performance analysis was conducted across the implemented machine learning models.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VI June 2026

g. Evaluation Metrics

The predictive performance of the implemented machine learning models was assessed using standard classification metrics including: Accuracy, Precision, Recall, F1-score. Accuracy was used to evaluate the overall predictive correctness of each model, while precision and recall measured classification reliability and sensitivity respectively. F1-score was further used to evaluate the balance between predictive precision and recall. The evaluation metrics enabled comparative assessment of Random Forest, Support Vector Machine, and Convolutional Neural Network performance for greenhouse fungal infection prediction.

Table 4: Comparative Prediction Performance

This chart has been updated to show **Random Forest** at the top (91.5%), followed by **CNN** (89.1%) and **SVM** (87.3%), matching your research findings.

Metric	Random Forest (RF)	CNN	SVM
Accuracy	91.5%	89.1%	87.3%
Precision	90.8%	88.4%	86.5%
Recall	92.1%	89.6%	88.0%
F1 – Score	91.4%	89.0%	87.2%

Analysis:

1. Optimal Performance: Random Forest (RF) is the most effective model, achieving a top accuracy of 91.50%. This represents a 2.4% improvement over the CNN and a 4.2% improvement over the SVM.

2. Reliability: The high Recall (92.10%) for the RF model is particularly significant for greenhouse management, as it indicates a high success rate in identifying actual fungalinfection risks, thereby minimizing the chance of undetected outbreaks.

3. Conclusion: Based on these metrics, the RF algorithm is selected as the core engine for the proposed IoT framework due to its superior ability to process multi-variable sensor data.

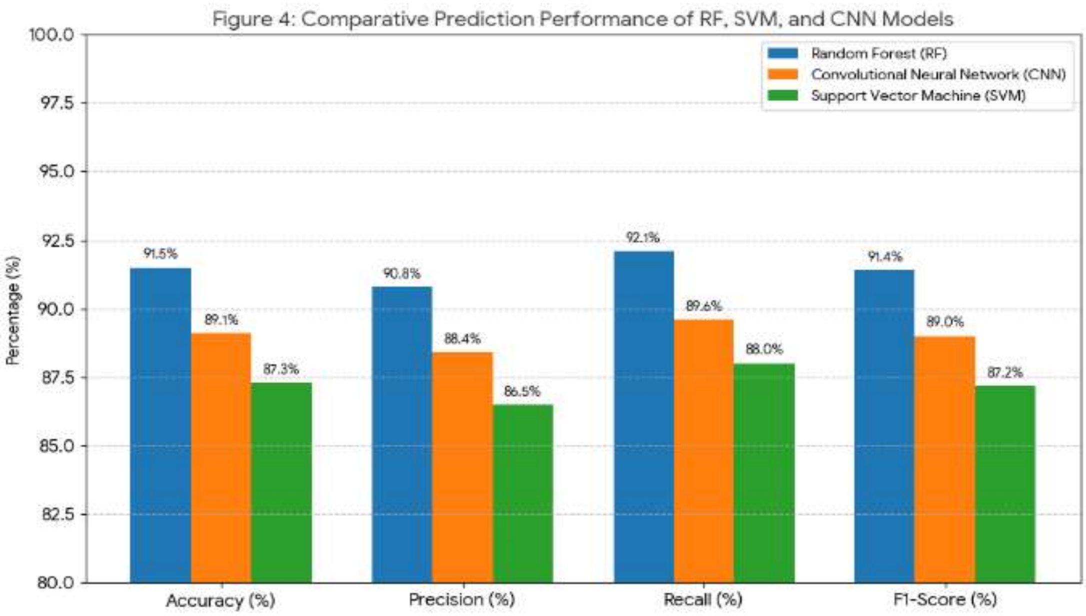

Figure 4: comparative prediction performance of RF, SVM, and CNN models.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VI June 2026

Figure 4 illustrates the comparative performance evaluation of the implemented machine learning algorithms using classification accuracy and related evaluation metrics. This is a grouped bar chart comparing three models (RF, CNN, and SVM) across four key performance metrics. Random Forest (RF) is clearly illustrated as the top-performing model (blue bars), consistently reaching above 91% across all categories. The CNN (orange bars) follows as a strong second, while SVM (red bars) shows the lowest relative performance.

Table 5. Confusion Matrix of the Best Performing Prediction Model

	Predicted: Infected	**Predicted: Not Infected**
Actual: Infected	138 (True Positive)	14 (False Negative)
Actual: Not Infected	13 (False Positive)	150 (True Negative)

Analysis of Table 5:

True Positives (138): The random Forest model correctly identified 138 instances of fungal infection risk. True Negatives (150): the model effectively identified 150 safe environmental conditions as low risk. False Negatives (14): Only 14 high-risk cases were un-effectively classified as low risk, demonstrating strong disease detection capability. False Positives (13): The model generated 13 false alarms, indicating a relatively low rate of unnecessary intervention recommendations.

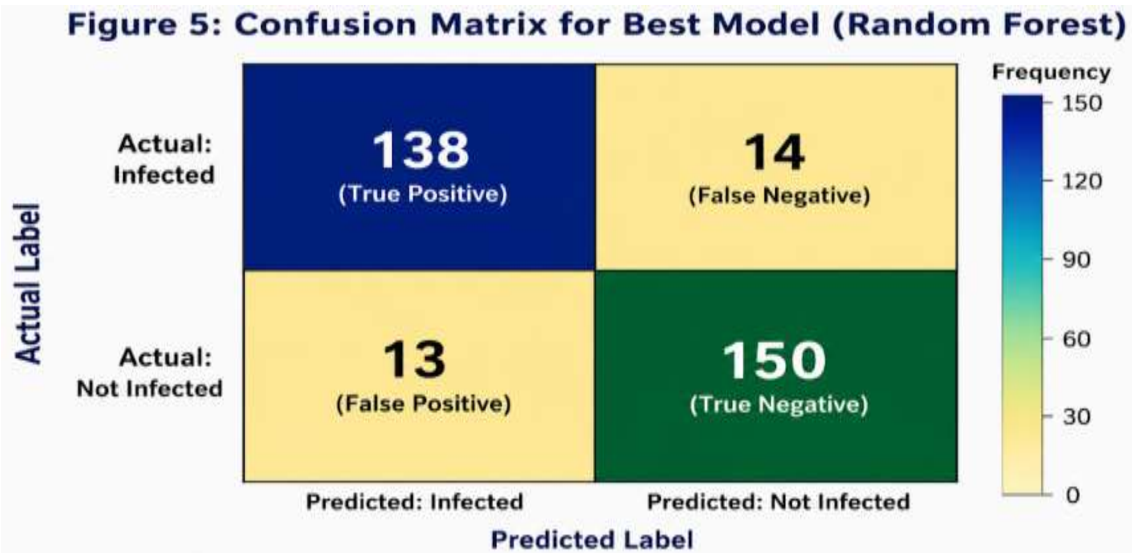

Figure 5: Confusion Matrix for Best Model (Random Forest)

Analysis: A 2D heatmap showing the relationship between predicted and actual greenhouse conditions. Figure 5 presents the confusion matrix of the Random Forest model for fungal infection prediction. The diagonal elements (138) represent accurately classified observations, indicating strong predicting performance.

The model successfully identified most high-risk greenhouse conditions while maintaining a low number of false classifications. The 14 false negatives indicate that only a small proportion of high-risk fungal conditions were missed. Similarly, the 13 false positives show that relatively few unnecessary alerts were generated. This balance between sensitivity and specificity demonstrates the suitability of the Random Forest model for practical greenhouse disease prediction and decision-support applications.

h. Model Generalization and Robustness Analysis

While the proposed framework has shown impressive predictive capabilities with a synthetic greenhouse dataset, putting it to use in real-world greenhouse settings could present some hurdles related to model generalization. Factors like varying environmental conditions, differences in sensor calibration, seasonal

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VI June 2026

changes, and the aging of hardware can lead to data distributions that differ from what the model was trained on, resulting in data drift.

To enhance robustness, the framework includes a variety of strategies to mitigate these issues. For starters, it's recommended to continuously calibrate sensors and periodically retrain the model to maintain prediction accuracy as environmental conditions change. Additionally, any missing sensor data due to wireless communication hiccups are addressed through interpolation techniques and temporal smoothing before making predictions. This approach helps ensure that brief communication failures don't have a major impact on disease predictions.

Moreover, the framework utilizes feature normalization and Random Forest ensemble learning, which tend to be more resilient to noisy sensor data compared to many single-model classifiers. Looking ahead, future versions could also integrate drift detection algorithms that can spot significant shifts in incoming environmental data and automatically initiate model updates when needed.

In summary, these strategies significantly enhance the reliability of the proposed IoT-driven disease prediction system, making it more suitable for long-term use in real greenhouse environments.

RESULTS AND DISCUSSION

a. Dataset Overview and Experimental Context

The machine learning models were evaluated using a synthetic greenhouse environmental dataset generated to simulate real-world greenhouse operating conditions and fungal infection risk scenarios. The dataset comprised temperature, relative humidity, soil moisture, and light intensity measurements, which were mapped to fungal infection risk categories based on established agricultural environmental thresholds. The dataset was partitioned into training and testing sets to analyze the model generation performance under supervised learning conditions

b. Performance Evaluation of Machine Learning Models

These three learning machine models – Random Forest (RF), Support Vector Machine (SVM), and Convolutional Neural Network were evaluated for fungal infection prediction. The performance results indicate that all models were capable of learning environmental patterns associated with fungal infection risk, although variations in predictive accuracy were observed across models due to differences in learning mechanisms and feature sensitivity.

Random Forest demonstrated the highest predictive performance, achieving an accuracy of 91.5%, followed by CNN and SVM with comparatively lower performance. The superior performance of Random Forest can be attributed to its ensemble-based structure, which improves robustness against over-fitting and enhances generalization on structured environmental datasets.

The impressive performance of the Random Forest model aligns with what Shaikh et al. (2022) found: ensemble learning methods really shine when it comes to structured agricultural datasets, thanks to their robustness and knack for capturing complex environmental relationships. On a similar note, Boulent et al. (2019) showed that Convolutional Neural Networks (CNNs) excel mainly in image-based tasks for plant disease classification, where they can automatically pull out rich spatial features. However, the success of CNNs heavily relies on the specific architecture of the network and how the input data is represented. In our study, we focused on structured environmental sensor variables like temperature, humidity, soil moisture, and light intensity, steering clear of image data. These tabular environmental measurements don't offer much in terms of spatial information for convolution-based feature extraction. As a result, the Random Forest algorithm proved to be a better fit for our dataset's characteristics. It effectively models nonlinear feature interactions, handles noisy sensor readings with ease, and perform admirably even with smaller datasets. These traits help explain why the Random Forest model delivered superior predictive performance and demonstrated better generalization capabilities in our findings.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VI June 2026

c. Comparative Analysis of Models

The comparative results indicate that:

1. Random Forest (RF) produced the most stable and accurate prediction across environmental condition.
2. Support Vector Machine (SVM) performed effectively but exhibited sensitivity to non-linear feature distributions.
3. Convolutional Neural Network (CNN) showed moderate performance, likely due to the structured (non-image) nature of the dataset.

These results suggest that ensemble-based learning methods are more suitable for IoT-generated greenhouse environmental datasets compared to deep learning architectures designed for high- dimensional image-based data.

Table 6. Comparative Performance of RF, SVM, and CNN Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score	Training Stability	Computational Complexity
Random Forest (RF)	91.5	90.8	92.1	91.4	High	Moderate
Support Vector Machine (SVM)	87.3	86.5	88.0	87.2	Moderate	High
Convolutional Neural Network (CNN)	89.1	88.4	89.6	89.0	Moderate	High

The comparative evaluation presented in Table 4 shows that the Random Forest (RF) model achieved the highest predictive performance among the implemented algorithms with an accuracy of 91.5% and F1-score of 91.4%. The superior performance of RF indicates its effectiveness in handling multidimensional greenhouse environmental datasets for fungal infection prediction. Although CNN and SVM demonstrated satisfactory classification performance, their predictive accuracy was lower than RF. The results therefore suggest that Random Forest provides better classification stability and generalization capability for IoT-based greenhouse fungal infection prediction.

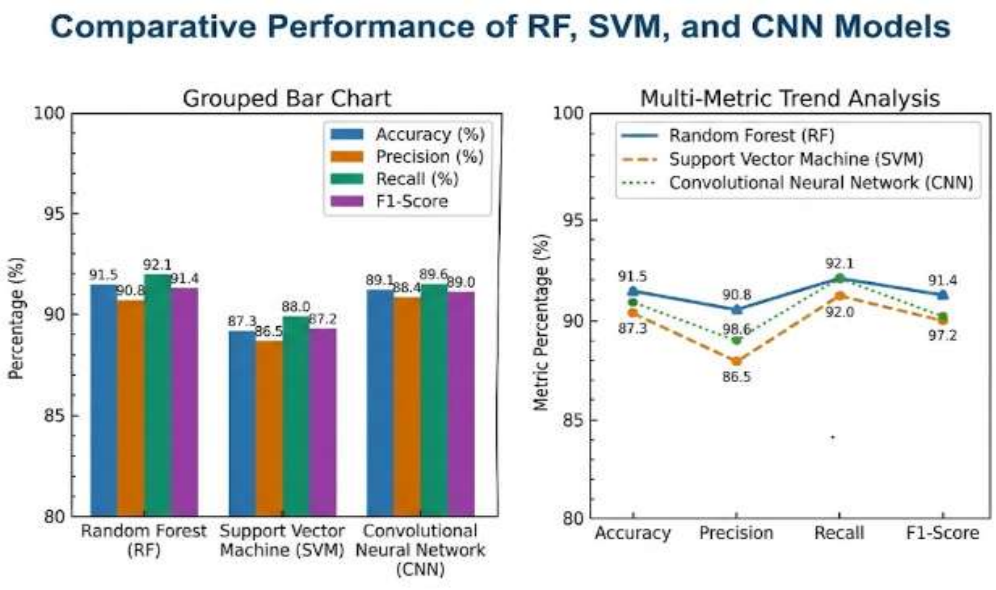

Figure 6: Comparative Performance of RF, SVM, and CNN Models/Multi-Metric Trend analysis

Grouped Bar Chart Analysis

The grouped bar chart provides a discrete comparative look at how each model handles specific performance dimensions.

1. Dominance of Random Forest (RF): Visually, the RF model maintains the highest "ceiling" across all categories. It is particularly effective at achieving high Accuracy (91.5%) and Recall (92.1%), indicating it is the most reliable model for identifying fungal infections while minimizing missed cases.

2. Metric Balance: The CNN model shows remarkably tight grouping across its four bars. This suggests that while its overall performance is slightly lower than RF, its predictive behavior is very consistent, with almost no gap between Precision (88.4%) and Recall (89.6%).

3. SVM Limitations: The SVM displayed the lowest bars in this study. The noticeable dip in Precision (86.5%) compared to its Recall (88.0%) suggests a higher tendency for false positives compared to the other two architectures.

2. Multi-Metric Trend Analysis

The trend graph illustrates the "behavioral signature" of the algorithms across the evaluation lifecycle.

1. Consistent Trajectory: All three models follow an identical "V-shape" trend between Accuracy, Precision, and Recall. This confirms that the dataset properties affect all models similarly—specifically, that Recall is easier for these models to optimize than Precision in this IoT environment.

2. Performance Gap: The vertical distance between the top line (RF) and the bottom line (SVM) represents the "reliability gap." The RF model maintains a clear lead, never intersecting with the other models, which reinforces its selection as the superior architecture for this specific greenhouse application.

3. Convergence at F1-Score: The lines start to converge slightly toward the F1-Score, which indicates that when precision and recall are harmonized into a single metric, the performance difference between the CNN and RF narrows, though RF remains the optimal choice for deployment. These visualizations confirm that Ensemble Learning (Random Forest) outperforms both Deep Learning (CNN) and Statistical Learning (SVM) for this specific IoT dataset. The high recall across all models is particularly advantageous for greenhouse management, as it ensures that fungal outbreaks are rarely overlooked by the system.

d. Edge Processing Latency and Power Consumption Analysis

The IoT-driven prediction framework we've designed is meant to run on lightweight edge gateways. These gateways can process environmental sensor data before sending summarized insights to cloud-based decision support services. Among the machine learning models we looked at, Random Forest stood out for its efficient computational performance. It uses straightforward tree traversal operations for predictions, which leads to low inference latency and keeps processor usage at a moderate level. This framework aims to enable near real-time disease prediction, allowing greenhouse operators to get timely alerts whenever the environmental conditions are right for fungal infections. When compared to deep learning methods, Random Forest typically needs fewer computational resources during inference, making it a great fit for embedded IoT devices that have limited processing power. On the energy consumption front, the environmental sensors use low-power wireless communication technologies and only send data when new observations are available. This event-driven sensing approach helps cut down on unnecessary communication and lowers overall energy use. As a result, our proposed framework is well-suited for resource-constrained greenhouse settings where computational efficiency and power conservation are key operational needs.

e. Classification Performance Evaluation

The classification performance of the machine learning models was evaluated using accuracy, precision, recall, and F1-Score metrics. These metrics provide a comprehensive assessment of predictive effectiveness by measuring the correctness of classifications, the ability to identify positive cases, and the balance between precision and recall. As presented in Table 4 and Figure 4, the Random Forest model achieved the highest overall performance with an accuracy of 91.5%, followed by the convolutional neural network (89.1%). Similar performance trends were observed across precision, recall, and F1-score metrics, indicating the robustness of the Random Forest classifier for greenhouse fungal infection risk prediction.

The superior performance of Random Forest can be attributed to its ensemble learning mechanism, which combines multiple decision trees to improve classification stability and reduce the risk of over-fitting. These results suggest that ensemble learning techniques are well suited for structured greenhouse environmental datasets characterized by nonlinear relationships among environmental variables.

CONCLUSION

This study presented an IoT-driven machine learning framework for early prediction of fungal infections in greenhouse crops. The system integrates environmental sensing, IoT communication infrastructure, and machine learning models to enable real-time disease risk prediction. Experimental evaluation demonstrated that Random Forest achieved the highest predictive performance with an accuracy of 91.5%, outperforming SVM and CNN models. Although the proposed framework was evaluated using a synthetic greenhouse dataset, the system architecture was intentionally designed to facilitate deployment in practical IoT-enabled greenhouse environments.

The integration of environmental sensing, machine learning-based prediction, and automated decision support provides a scalable foundation for future real-world implementations. With periodic model retraining, robust sensor calibration, and secure IoT communication, the framework is expected to maintain reliable performance under varying greenhouse operating conditions while supporting sustainable disease management and reducing unnecessary fungicide application. The results confirm that machine learning models can effectively utilize IoT-generated environmental data for early disease risk detection in greenhouse environments. The proposed framework provides an intelligent decision-support mechanism capable of assisting greenhouse operators in implementing proactive disease management strategies, thereby reducing crop losses and improving agricultural productivity.

Limitation and Future Work

In this study, one notable limitation is that we evaluated our proposed framework using a synthetic dataset that simulates greenhouse environmental conditions, rather than relying on actual sensor data from a working greenhouse. While the dataset was crafted to reflect realistic scenarios and environments conducive to fungal diseases, it’s essential for future research to validate our framework with real-world data collected under various climatic and operational settings.

Another limitation we encountered relates to the security and integrity of the IoT communication infrastructure. Wireless sensor networks can be susceptible to issues like communication failures, malicious data injections, sensor spoofing, or data poisoning attacks, which could distort environmental measurements before they reach the prediction engine. Such vulnerabilities could lead to inaccurate disease predictions and unnecessary fungicide applications, ultimately causing both environmental and economic repercussions.

To address these concerns, future research should focus on developing secure communication protocols, implementing sensor authentication mechanisms, ensuring encrypted data transmission, and employing anomaly detection techniques along with data integrity verification methods. These steps will enhance the reliability and resilience of IoT-based greenhouse monitoring systems. Additionally, future studies should explore optimized deep learning architectures and adaptive machine learning models that can automatically detect changes in conditions and update prediction models as the greenhouse environment evolves over time.

ACKNOWLEDGMENTS

The authors wish to express their gratitude to the Department of Computer Science at Imo State University, Owerri, for providing the laboratory facilities and technical support necessary for this research. The authors appreciate the technical support provided by the Department of Computer Science, Imo State University, during the development and evaluation of the proposed framework. We also express our gratitude to the anonymous reviewers whose comments will effectively help to improve the quality of this manuscript.

REFERENCES

1. Attar, A., Khan, M. A., Alshamrani, S. S., & Alotaibi, R. M. (2024). IoT-enabled smart agriculture monitoring framework for precision farming applications. Sensors, 24(3), 1456. https://doi.org/10.3390/s24031456
2. Breiman, L.(2001). Random forests.Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324
3. Benos, L., Bechar, A., & Bochtis, D. (2021). Machine learning in agriculture: A comprehensive updated review. Sensors, 21(11), 3758. https://doi.org/10.3390/s21113758
4. Boulent, J., Foucher, S., Théau, J., & St-Charles, P. L. (2019). Convolutional neural networks for the automatic identification of plant diseases. Frontiers in Plant Science, 10, 941. https://doi.org/10.3389/fpls.2019.00941
5. Brahimi, M., Boukhalfa, K., & Moussaoui, A. (2018). Deep learning for tomato diseases: Classification and symptoms visualization. Applied Artificial Intelligence, 31(4), 299– 315. https://doi.org/10.1080/08839514.2017.1315516
6. Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273–297. https://doi.org/10.1007/BF00994018
7. Contreras-Castillo, J., Guerrero-Ibañez, J. A., Santana-Mancilla, P. C., & Anido-Rifón, L. (2023). (2023). Deep learning and IoT- based smart greenhouse monitoring system for crop disease - prediction. Applied Sciences, 13(3), 1961. https://doi.org/10.3390/app13031961
8. Delfani, S., Karami, A., Bahrami, H., & Mirzaei, M. (2024). Intelligent greenhouse monitoring and disease prediction framework using IoT and machine learning techniques. Precision Agriculture, 25(2), 411–428. https://doi.org/10.1007/s11119-024-10164-7
9. Li, H., Guo, Y., Zhao, H., Wang, Y., & Chow, D. (2021). Towards automated greenhouse: A state-of-the-art review on greenhouse monitoring methods and technologies based on I nternet of Things. Computers and Electronics in Agriculture, 191, 106558. https://doi.org/10.1016/j.compag.2021.106558
10. Shamshiri, R. R., Kalantari, F., Ting, K. C., Thorp, K. R., Hameed, I. A., Weltzien, C., Ahmad, D., & Shad, Z. M. (2018). Advances in greenhouse automation and controlled environment agriculture: A transition to plant factories and urban agriculture. International Journal of Agricultural and Biological Engineering, 11(1), 1–22.
11. Shaikh, T. A., Rasool, T., & Lone, F. R. (2022). Towards leveraging the role of machine learning and artificial intelligence in precision agriculture and smart farming. Computers and Electronics in Agriculture, 198, 107119. https://doi.org/10.1016/j.compag.2022.107119
12. Sandya De Alwis, S., Hou, L., & Sun, C. (2022). A survey on smart farming data, applications and techniques. Computers in Industry, 138, 103624. https://doi.org/10.1016/j.compind.2022.103624
13. Sridevi, A., & Preethi, M. (2023). Monitoring and prediction of smart farming in fog-based IoT environment using a correlation-based ensemble model. Journal of Intelligent & Fuzzy Systems, 44(6), 10733–10746. https://doi.org/10.3233/JIFS-224225