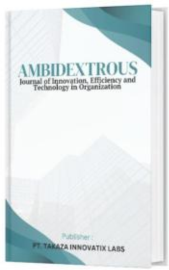
Ambidextrous : Journal of Innovation, Efficiency and Technology in Organization
https://journal.takaza.id/index.php/ambidextrous
Vol. 05, No. 01, 2026, pp. 61-70
E-ISSN: 3031-7002
E-mail: ambidex@takaza.id



A Framework on Optimizing Greenhouse Conditions for Plant Growth using Machine Learning Based on Time-Series Data

Nanang Fatchurrohman[1*], Yuli Panca Asmara[2]
[1]Universitas Putra Indonesia YPTK Padang
[2] INTI International University
*Corresponding Author e-mail: n.fatchurrohman@gmail.com

Article History:
Received: 25-05-2025
Revised: 31-05-2026
Accepted: 09-08-2026

Keywords: *MSMES; Plant Growth; Greenhouse Conditions; Machine Learning; Time-Series Data; Framework*

Abstract: *This study presents a machine learning (ML) framework for optimizing greenhouse environments to promote efficient plant growth. The approach utilizes time-series data to analyze and predict optimal growing conditions by integrating morphological traits—such as plant height, leaf size, and canopy structure—with dynamic environmental variables, including temperature, humidity, light intensity, and CO_2 levels. The proposed framework follows a structured, six-stage workflow comprising: data acquisition, data preprocessing, feature engineering, model selection and training, model evaluation, and prediction and decision support.. High-resolution sensors continuously collect environmental and plant-related data, which are then cleaned and formatted to handle missing values and temporal inconsistencies. Feature engineering techniques are applied to extract relevant growth indicators and environmental patterns. This framework lays the foundation for an intelligent decision support system that can assist growers in making informed environmental control decisions. By aligning greenhouse microclimates with the specific needs of crops, this approach aims to maximize yield, reduce energy consumption, and minimize resource waste. The adaptability of the framework also allows for its application across various plant species and cultivation setups, making it a scalable solution for modern agricultural operations.*

Introduction

The integration of machine learning (ML) into greenhouse operations has become a key innovation in advancing precision agriculture. Modern greenhouses are now equipped with Internet of Things (IoT) sensors that collect continuous, high-resolution time-series data related to environmental factors such as temperature, humidity, light intensity, and carbon dioxide concentration. When combined with morphological attributes like leaf size, plant height, and canopy coverage, this data serves as the foundation for predictive models that guide responsive greenhouse management.

ML models such as Random Forest Regression have proven capable of forecasting internal environmental variables with high accuracy, enabling more adaptive climate control compared to static, rule-based systems. These intelligent systems can optimize resource use while maintaining optimal conditions for plant development. As climate variability and global food demand increase, adopting predictive, data-centric approaches in agriculture is becoming increasingly necessary (Morales-García et al., 2023).

To operationalize these capabilities, a structured ML framework is essential. This includes distinct phases such as data acquisition, preprocessing, feature extraction, model training, and evaluation. The use of time-series sensor data enables models to capture temporal dynamics in both environmental conditions and plant responses, laying the groundwork for automated, real-time environmental adjustments tailored to specific crops and growth stages.A structured machine learning (ML) framework is essential for optimizing environmental conditions in smart greenhouses to enhance crop growth and efficiency.

The workflow typically involves six-stage workflow comprising: data acquisition, data preprocessing, feature engineering, model selection and training, model evaluation, and prediction and decision support. In the initial phase, IoT-based sensors gather time-series data on key environmental parameters such as temperature, humidity, soil moisture, and light intensity, along with plant-specific morphological indicators like leaf area and stem diameter (Mukherjee et al., 2024).

Preprocessing is critical to ensure clean and consistent datasets by addressing missing values, noise, and anomalies. Lim et al. (2023) highlight that customized preprocessing methods significantly influence model accuracy, especially when dealing with dynamic agricultural data. Feature engineering transforms raw input into meaningful metrics—such as growth rate patterns or light-to-temperature ratios—which improves the model's interpretability and predictive strength (Morales-García et al., 2023).

Various ML algorithms are then trained, including ensemble methods like random forest and gradient boosting, as well as advanced deep learning models such as long short-term memory (LSTM) networks. These models are evaluated using performance metrics like root mean square error (RMSE) and coefficient of determination (R^2). The best-performing model informs real-time decision-making for automated greenhouse control systems. According to Pugazh et al. (2024), integrating ML with IoT allows systems to autonomously adjust internal conditions, reducing resource waste while maintaining optimal growth. Such framework enable precision agriculture by aligning environmental control with specific crop requirements, ultimately contributing to yield improvement and sustainable farming practices.

Literature Revie

Evaluating the performance of machine learning (ML) models is a critical step in developing intelligent greenhouse systems for precision agriculture. Standard metrics such as root mean square error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2) are commonly used to quantify prediction accuracy. These indicators help determine how effectively a model captures the nonlinear relationships between environmental factors and plant responses. In recent studies, Random Forest Regression and other time-series models have demonstrated strong predictive capabilities for climate variables such as temperature and humidity within smart greenhouses, achieving MAE values below 1°C in some cases (Morales-García et al., 2023).

Lim et al. (2023) emphasized the importance of customized preprocessing techniques and model selection in enhancing prediction reliability, especially when dealing with heterogeneous and noisy agricultural data. Accurate predictions of internal conditions enable automated control systems to dynamically adjust temperature, lighting, irrigation, and CO_2 levels based on crop requirements. Such real-time optimization contributes to improved crop yields, better resource management, and reduced operational costs.

El Ouaham et al. (2023) highlight the growing role of AI in climate control, plant health monitoring, and robotic automation within modern greenhouse systems. Moreover, the integration of predictive analytics with automated actuators (e.g., fans, heaters, sprayers) enables intelligent systems to respond to suboptimal growing conditions with minimal human intervention. This transition from rule-based to data-driven decision-making represents a paradigm shift in greenhouse management, promoting sustainability and operational efficiency.

The development of intelligent decision support systems (DSS) is central to advancing greenhouse automation and precision agriculture. The proposed machine learning–based framework establishes a foundation for such systems by enabling real-time, data-driven environmental control. By aligning microclimatic conditions—such as temperature, humidity, CO_2 concentration, and light intensity—with crop-specific needs, this approach supports improved plant growth, enhanced yield, and greater energy efficiency. The core strength of this system lies in its ability to learn from historical data and adjust dynamically to new inputs, allowing continuous adaptation across various crop types and cultivation environments.

Wang et al. (2020) introduced a deep reinforcement learning model that accounted for multiple environmental variables to maximize cucumber yield while minimizing operational costs, showing the practical viability of learning-based control strategies. Unlike traditional rule-based systems, intelligent DSS can evaluate multiple factors simultaneously and provide actionable insights that optimize the greenhouse environment. Su and Xu (2022) demonstrated how setpoint decision support strategies, combined with adaptive hybrid control models, significantly improve both climate regulation and energy conservation in greenhouse settings.

Moreover, Ouammi et al. (2020) developed a comprehensive decision tool that regulates indoor greenhouse microclimates using predictive algorithms and real-time monitoring, further validating the benefits of integrating AI into greenhouse management. These advancements mark a shift toward scalable, autonomous systems capable of supporting sustainable agriculture in diverse operational contexts.

Optimizing plant growth within controlled environments such as greenhouses is a fundamental aspect of precision agriculture, aimed at increasing crop yields while minimizing resource consumption. Traditional greenhouse operations often rely on manual control and empirical knowledge, which can lead to inefficiencies and inconsistent outcomes. However, advancements in sensor technologies and artificial intelligence have enabled real-time monitoring of environmental conditions and plant responses. This opens new opportunities for integrating machine learning (ML) models into greenhouse management systems to make predictive and data-driven decisions.

Recent research highlights the effectiveness of ML algorithms, including support vector machines, random forests, and deep learning models like Long Short-Term Memory (LSTM), in accurately forecasting plant growth metrics based on time-series data collected from environmental sensors. For instance, Janaki et al. (2023) developed an LSTM-based model that successfully predicted stem growth and yield of ficus benjamina in a greenhouse environment, outperforming traditional regression models. Similarly, Rajathi et al. (2024) compared various ML models and found that XGBoost was most effective in predicting plant growth variables using IoT-based environmental data. These predictive models enable dynamic control of key parameters such as temperature, humidity, and light intensity, leading to more consistent and optimized growing conditions.

Furthermore, Jeevan et al. (2024) proposed a deep learning-based RNN model that offers high accuracy in yield prediction by incorporating microclimate data and plant physiological parameters. These advancements represent a shift from reactive to proactive greenhouse management, offering scalable solutions adaptable to different crops and cultivation setups. By leveraging predictive analytics, growers can achieve improved productivity, reduce energy and water use, and contribute to sustainable agriculture.

Recent studies validate the effectiveness of ML-based frameworks in managing greenhouse climate and predicting plant development. Wang et al. (2020) demonstrated that deep reinforcement learning (DRL) could optimize cucumber greenhouse conditions by modeling future rewards and plant responses to multiple control variables. Similarly, Chen and You (2022) introduced a data-driven robust model predictive control (DDRMPC) system that used machine learning to address uncertainties in weather forecasts and greenhouse dynamics, significantly reducing constraint violations and operational costs. In another study, Lachouri et al. (2022) applied an Adaptive Neuro-Fuzzy Inference System (ANFIS) to model nonlinear greenhouse environments, effectively predicting critical parameters like temperature, humidity, and CO_2 for tomato seedling growth.

This paper presents a machine learning (ML) framework developed to analyze time-series data comprising environmental parameters and plant growth metrics in greenhouse environments. The objective is to create predictive models that guide climate control decisions, enabling real-time adjustments tailored to the specific developmental stages of crops. By integrating these approaches, the proposed framework contributes to scalable, intelligent decision-making in smart agriculture, helping growers optimize plant growth, improve yield quality, and reduce energy and resource usage.

Proposed Framework

To develop an intelligent greenhouse management system, this study proposes a machine learning-based framework designed to analyze and predict plant growth responses under dynamic environmental conditions. The methodology follows a structured pipeline beginning with the collection of time-series data from sensors monitoring key environmental variables and plant growth indicators. This data is then preprocessed to ensure consistency and accuracy, followed by feature engineering to extract meaningful patterns that influence plant development. Multiple machine learning algorithms are trained and evaluated to identify the most effective model for predicting optimal climate conditions tailored to different crop stages. The resulting model is integrated into a decision support system capable of making real-time recommendations or automated adjustments to the greenhouse environment, thereby enhancing crop productivity and resource efficiency. The following presents the steps involved in the proposed framework.

1. Data Acquisition

The foundation of any intelligent greenhouse system lies in the accurate and continuous acquisition of relevant data. Environmental parameters such as temperature, humidity, CO_2 concentration, and light intensity are essential for understanding the growing conditions within a controlled environment. Simultaneously, plant growth indicators—including height, stem thickness, leaf area, and yield—provide insights into the biological response of crops to these external factors. To collect this data, advanced Internet of Things (IoT) sensor networks and monitoring devices are deployed throughout the greenhouse (Wang et al., 2020). These devices are capable of capturing real-time data at high frequency, which is then transmitted to centralized data storage systems, typically hosted in cloud platforms or secure databases. This continuous flow of high-quality data not only supports immediate monitoring but also forms the basis for downstream analytical processes. It enables historical tracking, model development, and predictive analytics, ultimately driving smarter, more responsive greenhouse management strategies

2. Data Preprocessing

Data preprocessing is a critical step in ensuring the reliability and usability of collected sensor data for machine learning applications. Raw environmental and plant-related data often contain missing values, sensor noise, or extreme outliers due to hardware malfunction, transmission errors, or environmental fluctuations. Therefore, the first stage involves cleaning the dataset by imputing or removing missing entries, filtering out noisy readings, and correcting any anomalies. Following this, normalization or scaling techniques such as min-max normalization or z-score standardization are applied to bring all input features onto a comparable scale, which helps improve model convergence and stability during training. Furthermore, time alignment of different data streams—collected from various sensors operating at different frequencies—is essential to maintain temporal coherence. By synchronizing environmental variables with corresponding plant growth metrics, the integrity of time-series relationships is preserved. These preprocessing steps collectively ensure that the dataset is both accurate and consistent, forming a reliable foundation for feature extraction and model training (Djordjevic et al., 2020).

3. Feature Engineering

Feature engineering enhances the model's ability to recognize complex relationships by transforming raw data into more informative representations. In the context of greenhouse agriculture, this involves generating new features that better capture plant-environment interactions. For instance, metrics like growth rate over time, light-use efficiency, and diurnal temperature variation provide deeper insights into the physiological responses of plants. These features are derived using domain knowledge and statistical methods applied to the cleaned dataset. In addition to creating new variables, feature selection techniques are used to identify which input factors have the most significant impact on plant growth. This step reduces dimensionality, improves computational efficiency, and enhances model interpretability. Common techniques include correlation analysis, recursive feature elimination, and model-based importance scoring. By retaining only the most relevant features, the resulting dataset becomes more focused and manageable, increasing the likelihood of accurate and robust predictions. Overall, feature engineering bridges the gap between raw data and effective machine learning modeling (Wang, 2021).

4. Model Selection and Training

The model development phase involves selecting appropriate machine learning algorithms based on the data characteristics and specific objectives of greenhouse optimization. Different algorithms offer distinct advantages depending on the type of prediction—whether it's forecasting plant growth trends or recommending environmental adjustments. Regression models such as Random Forest and XGBoost (Extreme Gradient Boosting) are popular for their ability to handle non-linear relationships and high-dimensional data. For time-series forecasting, Long Short-Term Memory (LSTM) and Recurrent Neural Networks (RNN) are employed due to their effectiveness in capturing temporal dependencies. In more advanced scenarios, reinforcement learning models are implemented to develop adaptive control strategies that respond to real-time environmental changes. Each model is trained on historical, labeled datasets where inputs consist of environmental conditions and outputs represent plant responses (Shiyale et al., 2020). Through iterative learning, these algorithms adjust their internal parameters to minimize prediction errors, enabling them to generalize well to unseen data. The trained models then serve as core components for decision-making systems in smart agriculture.

5. Model Evaluation

After training, machine learning models must be rigorously evaluated to ensure they are accurate, reliable, and generalizable. This is achieved using standard performance metrics, including Root Mean Square Error (RMSE), which measures the model's prediction error magnitude, and Mean Absolute Error (MAE), which assesses the average absolute difference between predicted and actual values. Additionally, the Coefficient of Determination (R^2) indicates how well the model explains the variability in the observed data. These metrics provide a comprehensive picture of model effectiveness. To prevent overfitting and ensure robustness, validation techniques such as k-fold cross-validation or a simple train-test split are applied. These methods test the model's performance on unseen subsets of the data, offering insights into how well it might perform in real-world settings (Guo et al., 2021). By thoroughly evaluating each model, researchers can select the most suitable approach for integration into greenhouse systems, ensuring dependable and scalable implementation for environmental control and crop prediction.

6. Prediction and Decision Support

Once a machine learning model has been successfully trained and validated, it can be used to make real-time predictions that inform greenhouse management decisions. These predictions may include forecasting future plant growth, identifying stress conditions, or recommending optimal environmental settings tailored to specific growth stages. The insights generated by the model are then integrated into automated or manual control systems, allowing for proactive adjustments in temperature, lighting, humidity, and CO_2 levels. This predictive capability reduces reliance on human intuition and allows growers to respond to changing conditions more effectively. Over time, these data-driven interventions lead to improved resource efficiency, reduced operational costs, and enhanced crop yield and quality. Additionally, the framework is adaptable, allowing the decision support system to be fine-tuned for different crop types, climate zones, and production goals. By closing the loop between data collection, prediction, and action, this approach represents a significant step toward intelligent and sustainable agriculture (Tomar & Saroha, 2024).

Discussion

This study highlights the pivotal role of intelligent data-driven systems in optimizing greenhouse operations through machine learning. The acquisition of environmental and crop-related parameters is foundational to smart greenhouse systems. Deploying distributed sensor networks enhances the granularity and frequency of real-time monitoring, enabling precise environmental control and data-rich decision making. These technologies collectively form a resilient infrastructure for precision agriculture, supporting both immediate interventions and long-term planning. However, the reliability of predictive models is directly linked to the quality of input data. Data preprocessing—such as filtering noise, handling missing values, and time-aligning multivariate sensor outputs—is critical to maintain data integrity and ensure robust analytical performance. Preprocessing also supports real-time processing, which is vital in dynamic agricultural environments where decisions must be made quickly and accurately (Lim et al., 2023).

Moreover, the extraction and selection of relevant features from raw data significantly influence model accuracy and efficiency. By deriving metrics like growth rate, thermal variation, or vapor pressure deficit, feature engineering allows machine learning models to capture complex plant-environment interactions more effectively. This step not only improves prediction quality but also reduces the computational burden by focusing on the most impactful variables. Model selection must align with the specific objectives of the greenhouse operation—be it yield forecasting, microclimate regulation, or anomaly detection. Algorithms such as Random Forest, XGBoost, and LSTM are commonly used due to their capability to model non-linear and temporal data effectively. Their deployment in smart agriculture has shown promising results in adapting to variable conditions and enabling autonomous decision-making systems (Mukherjee et al., 2024). Robust evaluation is essential before deploying these models in live systems. Performance metrics such as Root Mean Square Error (RMSE) and Coefficient of Determination (R^2), along with cross-validation techniques, are essential for assessing generalizability and avoiding overfitting. Well-validated models are more likely to deliver consistent results in fluctuating real-world greenhouse environments (Debdas et al., 2024).

Finally, integrating predictive models with real-time decision support systems completes the intelligent agriculture cycle. These systems enable automated adjustments for temperature, humidity, lighting, and irrigation, significantly reducing manual intervention and optimizing resource use. The adaptability of such systems across different crops and climatic zones demonstrates their potential for widespread, scalable deployment in sustainable agriculture.

Conclusion

This paper presents a comprehensive framework for optimizing greenhouse conditions using machine learning applied to time-series data. Central to the framework is the accurate and continuous acquisition of both environmental and crop-specific parameters. Leveraging distributed IoT sensor networks, greenhouse systems can now collect real-time data on temperature, humidity, light intensity, CO_2 levels, and plant growth indicators. This high-resolution data serves as the foundation for all predictive modelling and decision-making processes. Robust data preprocessing, addressing noise, temporal inconsistencies, and scaling, ensures data integrity and prepares it for effective model training. Feature engineering then transforms raw observations into actionable variables that capture complex interactions between plants and their microenvironments. Depending on the system's objectives, such as growth forecasting or resource optimization, then appropriate machine learning algorithms, including random forests or gradient boosting machines, are selected and trained using historical time-series data. Model evaluation, confirms the reliability and generalizability of predictions. Once validated, these models are integrated into decision support systems capable of real-time environmental control. This closed-loop feedback mechanism enables greenhouse automation that aligns climate conditions with plant needs, reducing manual oversight, conserving resources, and boosting crop quality and yield. The proposed framework demonstrates how integrating IoT infrastructure with time-series-based machine learning can create scalable, intelligent systems for sustainable and precision-driven agriculture. Positioning smart greenhouses as a key solution for addressing future global food security challenges.

References

Chen, W.-H., & You, F. (2022). Semiclosed greenhouse climate control under uncertainty via machine learning and data-driven robust model predictive control. IEEE Transactions on Control Systems Technology, 30(3), 1186–1197. https://doi.org/10.1109/TCST.2021.3094999

Debdas, S., Pandey, S., Gupta, S., Dash, R. K., Patnaik, M. C., & Singh, A. A. (2024). Optimizing resource efficiency in smart greenhouses through IoT. 2024 5th International Conference for Emerging Technology (INCET). https://doi.org/10.1109/INCET61516.2024.10593254

Djordjevic, M., Punovic, V., Danković, D., & Jovicic, B. (2020). Smart autonomous agricultural system for improving yields in greenhouse based on sensor and IoT technology. Journal of Applied Engineering Science. https://doi.org/10.5937/jaes0-26722

El Ouaham, W., Mouzouna, Y., & Sadik, M. (2023). A literature review on smart greenhouse and AI: Paradigms, opportunities and open issues. https://doi.org/10.1007/978-3-031-62488-9_13

Guo, C., Zi, Y., & Ren, W. (2021). A blockchain based framework for smart greenhouse data management. In Advances in Knowledge Management (pp. 299–310). Springer. https://doi.org/10.1007/978-3-030-82153-1_25

Janaki, K., Varun, N. V. S., Manohar, K., Kumar, K. M., Suresh, K., & Renuka, M. (2023). Deep learning to predict plant growth and yield in greenhouse environment. International Journal of Advanced Research in Science, Communication and Technology. https://doi.org/10.48175/ijarsct-9242

Jeevan, Y., Kuma, N., Chandan, R., Somanini, S. H., Vadtya, S., Ram, Y., ... & Kalra, R. (2024). Predictive modeling for enhanced plant cultivation in greenhouse environment. E3S Web of Conferences. https://doi.org/10.1051/e3sconf/202450701066

Lachouri, C. E., Mansouri, K., & Lafifi, M. M. (2022). Greenhouse climate modeling using fuzzy neural network machine learning technique. Revue d'Intelligence Artificielle. https://doi.org/10.18280/ria.360614

Lim, J., Yoo, J. S., Diop Emane, C. R., Kim, Y., & Beak, J. H. (2023). An effective smart greenhouse data preprocessing system for autonomous machine learning. Korean Institute of Smart Media. https://doi.org/10.30693/smj.2023.12.1.47

Morales-García, J., Bueno-Crespo, A., Martínez-España, R., & Cecilia, J. M. (2023).

Data-driven evaluation of machine learning models for climate control in operational smart greenhouses. Journal of Ambient Intelligence and Smart Environments, 15(1), 3–17. https://doi.org/10.3233/AIS-220441

Mukherjee, J., Mukherjee, D., Upadhyay, S., Basu, A., Shaw, L., Singh, D., & Manna, S. (2024). Non plant specific smart greenhouse with convective drying unit. 2024 International Conference on Trends in Quantum Computing and Emerging Business Technologies. https://doi.org/10.1109/TQCEBT59414.2024.10545221

Ouammi, A., Choukai, O., Zejli, D., & Sayadi, S. (2020). A decision support tool for the optimal monitoring of the microclimate environments of connected smart greenhouses. IEEE Access, 8, 212094–212105. https://doi.org/10.1109/ACCESS.2020.3039889

Pugazh, M. P., Xavier, A. V. M., & Shyni, S. S. (2024). Design of greenhouse system with Internet of Things and machine learning. 2024 10th International Conference on Advanced Computing and Communication Systems (ICACCS). https://doi.org/10.1109/ICACCS60874.2024.10716969

Rajathi, K., David, S. A., & Naveena, N. R. (2024). Comparative performance of machine learning models in predicting plant growth using IoT data. 2024 5th International Conference on Smart Electronics and Communication (ICOSEC), 636–640. https://doi.org/10.1109/ICOSEC61587.2024.10722597

Rokade, A., Singh, M., Malik, P., Singh, R., & Alsuwian, T. (2022). Intelligent data analytics framework for precision farming using IoT and regressor machine learning algorithms. Applied Sciences. https://doi.org/10.3390/app12199992

Shiyale, P. D., Thombre, H., Tyagi, G., & Chauhan, A. (2020). IOT-based smart greenhouse automation. International Journal of Scientific Research and Engineering Development.

Su, Y., & Xu, L. (2022). Setpoint decision support strategy and adaptive hybrid control of greenhouse climate: A simulation study. International Journal of Control, Automation and Systems, 20, 3740–3758. https://doi.org/10.1007/s12555-021-0638-3

Tomar, V., & Saroha, N. (2024). IoT-enabled smart greenhouse: Real-time monitoring and automated control for efficient agriculture. International Journal of Research and Review in Applied Science, Humanities, and Technology. https://doi.org/10.71143/48257x69

Wang, J., Chen, M., Zhou, J., & Li, P. (2020). Data communication mechanism for greenhouse environment monitoring and control: An agent-based IoT system. Information Processing in Agriculture, 7, 444–455. https://doi.org/10.1016/j.inpa.2019.11.002

Wang, L., He, X., & Luo, D. (2020). Deep reinforcement learning for greenhouse climate control. 2020 IEEE International Conference on Knowledge Graph (ICKG), 474–480. https://doi.org/10.1109/ICBK50248.2020.00073

Wang, Z. (2021). Greenhouse data acquisition system based on ZigBee wireless sensor network. Environmental Technology & Innovation, 22, 101689. https://doi.org/10.1016/j.eti.2021.101689