Comparative Analysis of AI Models for Zone-Level Microclimate Prediction in a Partially Sealed Greenhouse

Hyun Tae Shin
Department of Smart Agriculture
Sunchon National University
Suncheon, Republic of Korea
sy4448441@naver.com

Hyeon Chang Jeong
Department of Smart Agriculture
Sunchon National University
Suncheon, Republic of Korea
gusckd555@naver.com

Yoe Hyun
Department of Artificial Intelligence Engineering
Sunchon National University
Suncheon, Republic of Korea
yhyun@scnu.ac.kr

Abstract— **This study presents a comparative analysis of AI-based virtual sensing models for zone-level microclimate prediction in a partially sealed greenhouse under limited sensor deployment. Temperature and humidity data from six measured zones were used to predict conditions in three unmeasured zones (Zones 7–9), simulating a realistic sensor-sparse environment. Four models—Random Forest, XGBoost, Multilayer Perceptron, and Long Short-Term Memory—were evaluated using RMSE and MAE. The results show that XGBoost achieved the highest accuracy for temperature prediction, while LSTM performed best for humidity prediction. Incorporating outdoor weather variables consistently improved prediction performance across all models. The findings confirm that AI-based virtual sensing can effectively estimate spatial microclimate conditions without dense sensor installation, offering a practical solution for intelligent greenhouse environment management.**

Keywords— **Virtual sensing, Partially sealed greenhouse, Zone-level microclimate prediction, Machine learning, Deep learning, Smart agriculture.**

I. Introduction (Heading 1)

The microclimate inside a greenhouse is a critical factor directly affecting crop growth, energy consumption, and environmental control efficiency. Particularly in semi-closed greenhouses, the limited ventilation structure and fluctuations in outdoor conditions cause significant variations in temperature and humidity distribution across different zones. To accurately capture these spatial microclimate variations, environmental data must be collected from multiple locations. However, in actual farming operations, deploying sensors throughout all zones is challenging due to sensor installation costs, equipment maintenance, and communication infrastructure limitations. Consequently, developing technology to predict the environment in unmonitored zones while sensors are installed only in some areas has emerged as a critical research topic for practical smart greenhouse operations.

Previous greenhouse environmental prediction studies have primarily focused on analyzing microclimate changes inside greenhouses using computational fluid dynamics or physics-based models. However, these approaches involve complex model development processes and extremely high computational costs, limiting their immediate applicability to actual farm operational systems. Recently, machine learning-based prediction utilizing meteorological and sensor data has gained attention as an alternative, with various machine learning and deep learning models being applied to temperature and humidity prediction. Nevertheless, most studies assume conditions where sufficient sensors are installed, and the problem of estimating the spatial environment in sensor-free zones has not been adequately addressed.

This study proposes an AI-based virtual sensing technique to predict the microclimate in unmeasured zones within semi-closed greenhouses. This approach utilizes external environmental variables and sensor data from certain internal zones to address these limitations. Data was collected from a total of nine zones. Zones 1 to 6 were used as model inputs, while Zones 7 to 9 were excluded from the input, simulating spaces without sensors. This approach aims to experimentally verify whether spatial microclimate prediction is feasible in a partially sensed environment.

II. Related Research

Various studies have been conducted on predicting microclimate inside greenhouses, primarily focusing on physics-based modeling and data-driven artificial intelligence techniques. Early research mainly utilized computational fluid dynamics (CFD) to reproduce thermal and flow characteristics and temperature distributions within greenhouses. Kim et al. [1] constructed a CFD model for a single-span glass greenhouse to simulate spatial temperature distributions, demonstrating the potential for high-resolution spatial analysis. Kim et al. [2] also predicted internal microclimate changes in a naturally ventilated greenhouse using thermal-hydraulic and ventilation models. However, such physical models are highly dependent on structural information and have high computational costs, limiting their real-time application.

To overcome such limitations, machine learning-based research for temperature and humidity prediction has recently been actively pursued. Hosseini Monjezi et al. [5] compared RBF neural networks, SVM, and GPR using outdoor temperature, humidity, wind speed, and solar radiation as inputs,

with the RBF model achieving the lowest RMSE. Choi [3] used a multilayer perceptron to make short-term predictions of greenhouse internal temperature and humidity, confirming that the neural network-based model demonstrated stable performance even in the actual measurement environment.
Deep learning-based research is also increasing. Oh et al. [6] applied an LSTM model to improve the prediction accuracy of smart greenhouse internal temperature, confirming the advantages of deep learning in processing time-dependent climate data. Furthermore, Wei et al. [4] quantitatively evaluated the prediction performance of greenhouse temperature and humidity by comparing multiple AI models, including BPPSO, LSSVM, and RBF, and reported that prediction capabilities vary significantly depending on the algorithm structure.
However, the common limitations of existing studies can be summarized as follows:

1. Most AI-based forecasting studies focus on temporal predictions and do not directly estimate the spatial distribution of microclimate within greenhouses.
2. Existing research assumes environments with sufficient sensor coverage, failing to address the challenge of predicting unmeasured zones commonly encountered in actual farms.

Therefore, in semi-closed greenhouses with limited sensor installations, AI-based virtual sensing research that estimates microclimate conditions in unmeasured zones by utilizing sensor data from some internal areas and outdoor information has not yet been sufficiently conducted. This study aims to address this research gap by comparing the prediction performance of various AI models (Random Forest, XGBoost, MLP, LSTM) to analyze the feasibility of zone-level spatial microclimate estimation.

III. MAIN BODY

A. Data Collection and Operational Environment

1) Research Greenhouse

This study developed a zone-based microclimate prediction model for semi-closed greenhouses used in protected horticulture.

This is a single-span greenhouse with a total floor area of approximately 256m². The interior is divided into nine zones for collecting sensor-based environmental data. The greenhouse employs a typical semi-closed operation mode utilizing natural ventilation through side vents and a heating system-based supplemental heating system (Fig. 1).

Figure 1 Panoramic view of the semi-enclosed single-span greenhouse under study.

The main specifications of the greenhouse are as shown in Table 1.

TABLE I. GREENHOUSE SPECIFICATIONS

Item	***Specification***
Structure type	Single-span polycarbonate greenhouse
Length(m)	32.0
Width (m)	8.0
Eavesheight(m)	2.3
Roof height (m)	4.2
Roof slope (°)	18
Ventilation method	Side-window natural ventilation + ceiling circulation fan
Heating method	Hot-water heating or FCU-based auxiliary heating
Exterior covering material	Polycarbonate panel
Internal zone configuration	9 zones (Zone 1–9)
Sensor-installed zones	Zone 1–6
Unmeasured zones (target zones)	Zone 7–9
Operation type	Partial sealed greenhouse

2) Circulation Fan

The semi-closed greenhouse used in this study is equipped with circulation ventilation fans to maintain stable internal airflow. The ventilation fans are arranged at regular intervals along the central section of the greenhouse ceiling (Fig. 2), preventing stagnation of internal air and mitigating temperature and humidity variations between zones. Particularly in semi-closed greenhouses, where side window ventilation is limited, insufficient internal air mixing can lead to localized temperature increases and humidity accumulation, making the role of ventilation fans even more critical.

This study collected temperature and humidity data from each zone under conditions where the ventilation fan was operating normally, aiming to incorporate the effect of fan operation on microclimate spatial distribution into the experimental data. Figure 2 shows the actual ventilation fan installed in the greenhouse under study.

Figure 2 Circulation ventilation fan installed in the semi-enclosed greenhouse under study

The general specifications of the ventilation fan are as shown in Table 2.

TABLE II. SPECIFICATIONS OF CIRCULATION FAN

Item	*Specification*
Fan diameter (mm)	Approximately 400–450 mm
Power consumption (W)	120–180 W
Width (m)	8.0
Air flow rate (m³/min)	70–110 m³/min
Rotational speed (RPM)	1,400–1,600 rpm
Installation position	Center ceiling line, spaced at 5–6 m intervals
Main function	Air mixing, temperature–humidity uniformity, removal of stagnant zones

3) *Data Collection and Operational Environment*

Figure 3 Outdoor weather station sensors (left) and indoor temperature–humidity sensor node (right) used for greenhouse data collection.

Figure 3 presents the sensor systems used to collect environmental data inside and outside the greenhouse. The outdoor weather station monitors external climatic factors, while the indoor temperature and humidity sensors measure the microclimate conditions within each designated zone. These datasets were used as model inputs and ground-truth values for evaluating the performance of microclimate prediction models.

a) *Outdoor Weather Station.*

The outdoor weather station was installed on the north side of the greenhouse to monitor external environmental conditions that directly affect internal microclimate variations. The station measured air temperature, relative humidity, wind speed, wind direction, solar radiation*, and rainfall*, which were used as model inputs to analyze the impact of outdoor climate on semi-closed greenhouse ventilation and temperature distribution. The measured variables and sensor specifications are summarized in Table 3.

TABLE III. OUTDOOR WEATHER DATA ITEMS

Cate gory	*Measurement Item*	*Unit*	*Sensor Type*	*Description*
Out door weat her	Air temperature (Ta)	° C	Thermistor / PT100	External air temperature
	Relative humidity (RH)	%	Capacitive RH sensor	External humidity
	Wind speed	m/s	3-cup anemometer	Airflow intensity
	Wind direction	°	Wind vane	Directional air movement
	Solar radiation	W/m^2	Pyranometer (optional)	Heating influence on greenhouse
	Rainfall	mm	Rain gauge	External precipitation conditions

b) *Temperature/Humidity Sensors.*

Indoor temperature and humidity sensors were installed in six zones (Zone 1–6) to capture the spatial microclimate distribution within the greenhouse. Each sensor node measured air temperature and relative humidity at fixed intervals and transmitted the data to a central data logger. These measurements served as input variables for model training, while data from unmeasured zones (Zone 7–9) were used as evaluation targets. The specifications of the indoor sensors are summarized in Table 4.

TABLE IV. INDOOR TEMPERATURE–HUMIDITY SENSOR SPECIFICATIONS

Item	*Specification*
Sensor type	Digital temperature–humidity sensor (e.g., SHT series)
Measurement variables	Air temperature, Relative humidity
Temperature range	−40 to 85 °C (typical)
Humidity range	0–100 %RH
Accuracy	±0.3 °C, ±2 %RH (typical)
Installation zones	Installation zones: Zone 1–9
Sampling interval	Fixed interval logging (e.g., 1–5 min)

c) Data Collection Settings.

Environmental data were recorded at fixed intervals throughout the experiment to obtain synchronized time-series data for model training and evaluation. All sensor readings were stored in a centralized logging system, and missing or noisy data points were removed during preprocessing. Table 5 summarizes the data collection settings, including sampling frequency, logging duration, and the variables used as model inputs and prediction targets.

TABLE V. DATA COLLECTION SETTINGS

Setting	*Description*
Sampling interval	1–5 min (fixed interval)
Recording duration	Full experimental period (continuous)
External input variables	Ta, RH, Wind speed, Wind direction, Solar radiation, Rain
Internal input variables	Temperature & RH from Zone 1–6

Setting	*Description*
Prediction target zones	Zone 7–9 (unmeasured zones)
Data preprocessing	Outlier removal, missing-value interpolation, normalization

d) Greenhouse Internal Zone Configuration and Sensor Placement.

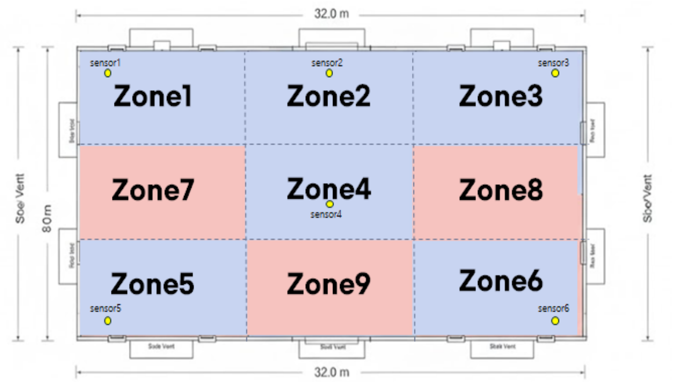

Figure 4 Configuration of the 9 Zones in the Study Greenhouse and Sensor Layout Diagram. The blue areas (Zones 1–6) are input zones where sensors are installed, while the red areas (Zones 7–9) are unmeasured zones (Virtual Sensing targets) without sensors.

This study divided the interior space into a total of nine zones (Zone 1–9) to analyze spatial microclimate variations within the greenhouse and predict unmeasured zones. Each zone was uniformly divided based on the greenhouse's length (32 m) and width (8 m), with the partitioned zones shown in Figure 4.

Environmental sensors (temperature/humidity) were installed in six zones: Zones 1, 2, 3, 5, 4, and 6. Data from these sensors was used as input for the AI model. Conversely, Zones 7, 8, and 9 were designated as unmeasured zones without sensors and were used for the Virtual Sensing evaluation. Sensor locations were chosen at representative points within each zone, considering the internal airflow within the greenhouse.

4) Virtual Sensing Scenario

The virtual sensing scenario was established to infer microclimate conditions in unmeasured zones under a limited sensing environment. Among the nine greenhouse zones, temperature and humidity data from Zones 1–6 were used as model inputs, while Zones 7–9 were excluded to simulate unmeasured areas. The input variables consisted of indoor temperature–humidity values from the six measured zones and optional outdoor weather variables, and the output variables were defined as the temperature and humidity of Zones 7–9.

The virtual sensing framework was constructed using time-series sensor datasets that were synchronized and preprocessed into fixed-interval records. The scenario follows a data-driven approach, in which no physical or CFD models are used; instead, all estimations are performed purely through machine-learning and deep-learning–based inference.
The input–output structure is illustrated in Figure 5.

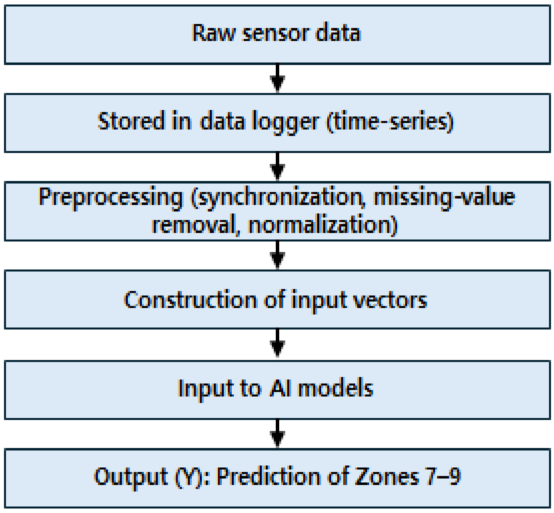

Figure 5 Panoramic view of the semi-enclosed single-span greenhouse under study.

5) AI Model Description

Four machine learning models were compared to evaluate their suitability for microclimate prediction under a partial sensing environment.

1) Random Forest (RF) and XGBoost represent tree-based ensemble models known for robustness against nonlinear relationships and limited data
2) Multilayer Perceptron (MLP) provides a baseline neural network structure for tabular environmental data.
3) Long Short-Term Memory (LSTM) was included to capture temporal dependencies in sequential sensor readings
4) All models were trained using identical input vectors and evaluated using RMSE and MAE for temperature and humidity predictions.

6) Data Preprocessing

Data preprocessing was conducted to ensure consistency and reliability of the time-series dataset. All sensor streams (indoor and outdoor) were synchronized to a 1-min interval using timestamp alignment. Missing values caused by communication delays were corrected by linear interpolation, and outliers were removed using a z-score–based filter. A 5-min moving average was applied to reduce high-frequency noise. All variables were normalized using min–max scaling to stabilize the training process.

7) Model Training Settings

The dataset was divided using an 80/20 train–test split while preserving temporal order to prevent data leakage. Random Forest and XGBoost hyperparameters (number of trees, depth, learning rate) were tuned via grid search. For MLP, the hidden layer size and activation functions were optimized, whereas LSTM models were trained using sliding windows (sequence length: 10–30 min). All models were trained using the Adam optimizer with early stopping to prevent overfitting.

8) Evaluation Metrics

Model performance was evaluated using Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE), defined as:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum(y_{i}-y_{i})^{2}} \quad (1)$$

$$\text{MAE} = \frac{1}{n}\sum|y_i - y_i| \quad (2)$$

Equation (1) represents the square-root average of the squared prediction errors and penalizes larger deviations more heavily, whereas Equation (2) expresses the average magnitude of absolute errors. Both metrics were calculated separately for temperature and humidity in the target Zones (7–9).

9) Experimental Scenario Design

Two experimental scenarios were designed to analyze the impact of external weather variables and limited sensing conditions.

1) Scenario 1 (Indoor only): Inputs consist of temperature and humidity from Zones 1–6.
2) Scenario 2 (Indoor + Outdoor): External weather variables (Ta, RH, wind speed) were added to the input vector.

For both scenarios, the prediction targets were the temperature and humidity of Zones 7–9.

All models used identical preprocessing and training pipelines to ensure fair comparison.

IV. Body Results and Discussion

This section presents the prediction results of temperature and humidity in the unmeasured zones (Zones 7–9) of a partially sealed greenhouse using the proposed virtual sensing–based AI models. Model performances were comparatively analyzed under identical input conditions. All models were trained using the same input vectors, consisting of temperature and humidity data from Zones 1–6 with optional outdoor weather variables, and were evaluated using RMSE and MAE metrics.

1) Temperature Prediction Results

Figure 6 illustrates the comparison between observed and predicted temperatures for Zones 7–9 across all models. When outdoor weather variables (outdoor temperature, relative humidity, and wind speed) were included as input features, prediction errors consistently decreased for all models. This result indicates that, due to the structural characteristics of partially sealed greenhouses, external weather conditions have a direct influence on internal microclimate behavior.

Among the evaluated models, XGBoost achieved the best overall performance. The average RMSE for Zones 7–9 ranged from approximately 0.45 to 0.60 °C, while the MAE ranged from 0.30 to 0.45 °C, demonstrating stable temperature estimation even in sensor-less zones.

The Random Forest model captured nonlinear relationships to a certain extent; however, a tendency toward overfitting was observed due to tree expansion. The MLP model showed relatively lower accuracy, likely because it does not explicitly incorporate temporal dependencies. The LSTM model outperformed MLP by modeling sequential patterns but exhibited slightly higher errors than XGBoost in temperature prediction.

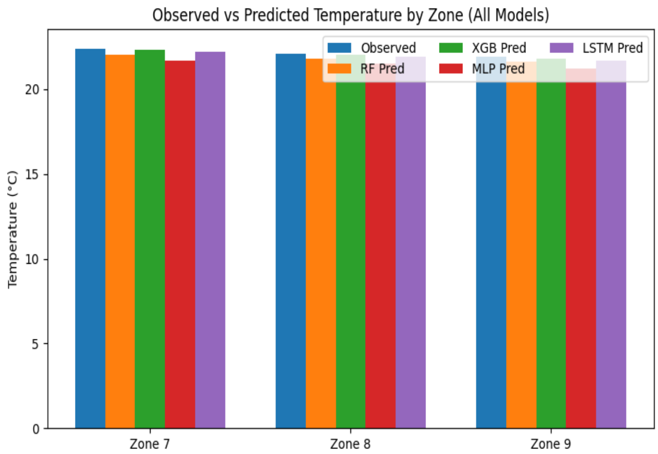

Figure 6 Panoramic view of the semi-enclosed single-span greenhouse under study.

2) Humidity Prediction Results

Figure 7 presents the observed and predicted humidity values for Zones 7–9. Humidity exhibited greater spatial and temporal variability than temperature, making it a more challenging variable to

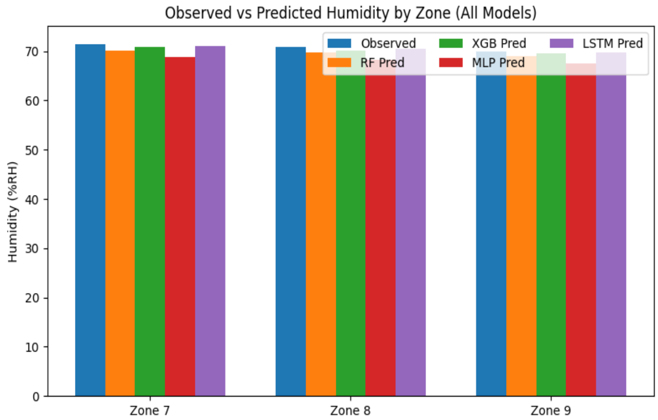

Figure 7 Panoramic view of the semi-enclosed single-span greenhouse under study.

predict. Nevertheless, the LSTM model demonstrated the highest prediction accuracy

The LSTM model achieved an average RMSE of approximately 1.5–2.0 %RH and an MAE of 1.0–1.4 %RH across Zones 7–9. This superior performance can be attributed to LSTM's ability to capture temporal dependencies and gradual transitions in humidity patterns among greenhouse zones.

XGBoost also showed competitive performance; however, slightly higher errors were observed during rapid humidity fluctuations compared to LSTM. Random Forest and MLP models exhibited relatively lower prediction stability under high humidity variability.

3) Analysis of Spatial Zone Characteristics

Although Zones 7–9 share similar geometric layouts, their microclimate behaviors differ depending on proximity to side vents, outdoor exposure, and ventilation fan effects.

1) Zone 7: Adjacent to side vents, highly influenced by outdoor conditions

2) Zone 8: Transitional zone influenced by both central and peripheral regions
3) Zone 9: Affected by exhaust airflow from circulation fans

Despite being excluded from model inputs, prediction errors across these zones remained comparable, indicating that spatial relationships among internal zones were effectively learned using data from Zones 1–6. In particular, Zone 7—despite its strong exposure to outdoor conditions—showed significantly improved prediction stability when outdoor variables were included.

4) *Effect of Outdoor Variables on Prediction Performance*

Figure 8 compares model performance (RMSE) with and without outdoor weather variables. The inclusion of outdoor variables consistently improved prediction accuracy across all models, with the most pronounced improvements observed for XGBoost and LSTM. This result highlights the critical role of outdoor weather information in microclimate prediction for partially sealed greenhouse environments.

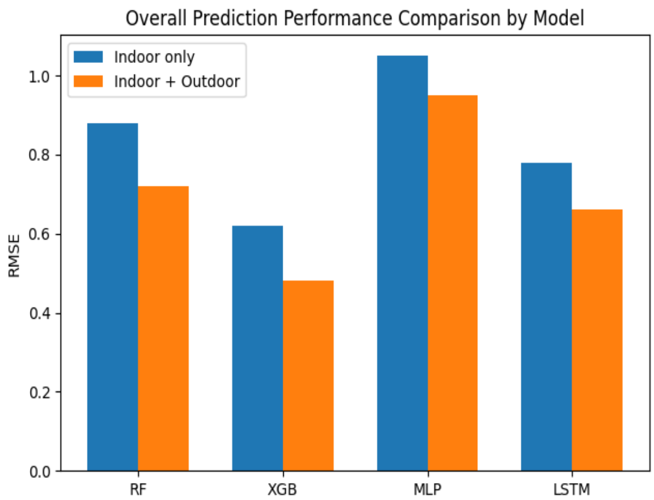

Figure 8 Overall prediction performance comparison by model with and without external weather variables.

V. Conclusion

This study investigated an AI-based virtual sensing approach for predicting microclimate conditions in unmeasured zones of a partially sealed greenhouse under limited sensor deployment. Temperature and humidity data collected from six measured zones, along with optional outdoor weather variables, were used to evaluate zone-level prediction performance of multiple machine learning models.

The experimental results indicated that XGBoost achieved the best performance for temperature prediction, while LSTM showed superior accuracy in humidity prediction. In addition, incorporating outdoor weather variables consistently improved prediction accuracy across all evaluated models, highlighting the importance of external environmental factors in partially sealed greenhouse microclimate dynamics.

These findings demonstrate that AI-based virtual sensing can effectively estimate zone-level microclimate conditions without dense sensor installation. The proposed approach provides a practical and scalable solution for intelligent greenhouse environment monitoring and control, particularly in real-world agricultural settings where sensor deployment is constrained.

Acknowledgment

This work was supported by Innovative Human Resource Development for Local Intellectualization program through the Institute of Information & Communications Technology Planning & Evaluation(IITP) grant funded by the Korea government(MSIT)(IITP-2025-RS-2020-II201489).

References

[1] S. H. Kim, H. S. Lee, and J. C. Han, "Development and validation of a microclimate prediction model in a single-span glass greenhouse using computational fluid dynamics," *Journal of Bio-Environment Control*, vol. 28, no. 2, pp. 95–104, 2019.

[2] Y. S. Kim and Y. H. Kim, "A model-based microclimate prediction model for ventilation control in a naturally ventilated greenhouse," *Journal of Bio-Environment Control*, vol. 31, no. 3, pp. 152–162, 2022.

[3] H. Choi, "Prediction of air temperature and relative humidity in greenhouse environments using multilayer perceptrons," *Computers and Electronics in Agriculture*, vol. 75, no. 2, pp. 235–243, 2011.

[4] X. Wei, Y. Zhang, L. Liu, and J. Chen, "Temperature and relative humidity prediction in a South China greenhouse based on machine learning," *Scientific Reports*, vol. 15, no. 1, pp. 1–12, 2025.

[5] P. Hosseini Monjezi, M. Khoshnevisan, and S. Rafiee, "Prediction of greenhouse indoor air temperature using artificial intelligence combined with sensitivity analysis," *Horticulturae*, vol. 9, no. 8, pp. 853–867, 2023.

[6] G. C. Oh, S. J. Kim, S. Y. Park, C. G. Lee, L. H. Cho, Y. K. Jeon, and D. H. Kim, "Development and verification of a smart greenhouse internal temperature prediction model using machine learning algorithms," *Journal of Bio-Environment Control*, vol. 31, no. 3, pp. 152–162, 2022.