공학박사학위논문

DEVELOPMENT ON SENSING, PREDICTION AND CONTROL SYSTEM OF GREENHOUSE ENVIRONMENT USING MACHINE LEARNING

기계학습을 이용한 온실 환경의 센싱, 예측 및 제어 시스템 개발

2023년 2월

서울대학교 대학원
생태조경·지역시스템공학부
지역시스템공학전공

박 세 준

공학박사학위논문

DEVELOPMENT ON SENSING, PREDICTION AND CONTROL SYSTEM OF GREENHOUSE ENVIRONMENT USING MACHINE LEARNING

기계학습을 이용한 온실 환경의 센싱, 예측 및 제어 시스템 개발

2023년 2월

서울대학교 대학원
생태조경·지역시스템공학부
지역시스템공학전공

박 세 준

Development on sensing, prediction and control system of greenhouse environment using machine learning

기계학습을 이용한 온실 환경의 센싱, 예측 및 제어 시스템 개발

지도교수 이 인 복

이 논문을 공학박사 학위논문으로 제출함

2023년 2월

서울대학교 대학원

생태조경·지역시스템공학부 지역시스템공학전공

박 세 준

박세준의 공학박사 학위논문을 인준함

2023년 2월

위 원 장 _____(인)

부위원장 _____(인)

위 원 _____(인)

위 원 _____(인)

위 원 _____(인)

DEVELOPMENT ON SENSING, PREDICTION AND CONTROL SYSTEM OF GREENHOUSE ENVIRONMENT USING MACHINE LEARING

A DISSERTATION

SUBMITTED TO THE DEPARTMENT OF LANDSCAPE ARCHITECTURE AND RURAL SYSTEMS ENGINEERING AND THE COMMITTEE ON GRADUATE STUDIES OF SEOULD NATIONAL UNIVERSITY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

BY

SE-JUN PARK

FEBRUARY 2023

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as dissertation for the degree of Doctor of Philosophy.

_____

Chairman

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as dissertation for the degree of Doctor of Philosophy.

_____

Vice-Chairman

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as dissertation for the degree of Doctor of Philosophy.

_____

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as dissertation for the degree of Doctor of Philosophy.

_____

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as dissertation for the degree of Doctor of Philosophy.

_____

Abstract

Crop production through facility cultivation has developed continuously over the past several years by ensuring a year round stable and high-quality productivity. Currently, the area and total production of domestic crop cultivation facilities are 52,571 ha and 2,441,000 tons in 2020, respectively (MAFRA, 2022). Smart farm greenhouses with the provision of precise monitoring and regulation of the internal environment are rapidly expanding. However, one major problem encountered by many farmers is the difficulty to secure installation costs of smart farms. Also, most domestic greenhouses are not compatible for the installation of numerous sensors and control systems, so an appropriate sensors and control system for monitoring and controlling the internal environment of greenhouses are needed.

The goal of this study was to develop a system that could monitor and predict the internal environment of naturally ventilated greenhouses. With this three machine learning models have been developed to serve different purpose. The prediction current air temperature using optimal sensor machine learning model (PCTO-ML) of Chapter 3 selected the optimal sensor location of air temperature. The prediction future air temperature using optimal sensor machine learning model (PFTO-ML) of Chapter 4 predicted the air temperature inside the greenhouse in the near future for preemptive control of the air conditioning system. Additionally, the optimal sensor location of PCTO-ML was introduced into the second model to minimize the number of sensors required for future air temperature prediction. The prediction local ventilation rate by CFD driven machine learning model (PLV-CFD driven ML) predicted natural ventilation.

In Chapter 2, the necessity and direction of this dissertation were presented

through comprehensive review and analysis of the methodology and limitations of related previous studies. To establish the research methodology, research works related to importance of the greenhouse internal growth environment; greenhouse ventilation evaluation method; optimal sensor location for monitoring the greenhouse internal environment; and application of machine learning techniques in agriculture were reviewed.

In Chapter 3, the optimal sensor location was selected among the nine grid-shaped sensor locations for monitoring the internal air temperature of the naturally ventilated greenhouse using a machine learning model. To build a dataset of this model, environmental factors were collected and were preprocessed through linear interpolation to get rid of missing values. The machine learning model for the prediction of the internal air temperature of a natural ventilation greenhouse was developed from various models such as artificial neural networks (ANN), support vector regression (SVR), and long short-term memory (LSTM). Comparing the predictive performance of these machine learning models, the LSTM model had the highest accuracy (R^2 = 0.974, RMSE = 0.024, and P-RMSE = 0.458). The environmental data measured at sensor 5 showed the highest accuracy in predicting air temperature for each sensor installation location in the greenhouse (R^2 = 0.984, RMSE = 0.019, and P-RMSE = 0.365). Finally, in order to minimize the kind of sensor required for the machine learning model, the simplified LSTM models with reduced learning features were developed and it also proposed the optimal sensor as sensor 5.

In Chapter 4, the future air temperature inside the natural ventilation greenhouse was predicted using a machine learning model. The process of collecting and preprocessing learning data for model development followed the same procedure as

in Chapter 3. Therefore, the accuracy of the LSTM model according to the sequence length was evaluated, and 30 minutes was evaluated as the optimal sequence length. The results of predicting the future air temperature at each location in the greenhouse using the LSTM model with a sequence length of 30 minutes showed high prediction accuracy of $R^2 > 0.95$, and RMSE < 0.65. In order to minimize the installation of sensors in the greenhouse, the predictive accuracy of the LSTM model was evaluated by applying the optimal sensor location suggested in Chapter 3. The results showed a relatively large decrease in prediction accuracy when one optimal sensor was applied (R^2=0.918), but the prediction accuracy was similar to that of using all nine sensors (R^2=0.950) when three optimal sensors were applied (R^2=0.939). Therefore, it was recommended to apply at least three optimal sensor locations for future air temperature prediction.

In Chapter 5, the ventilation rate in naturally ventilated greenhouses by region was predicted using a machine learning model build from the results of Computational Fluid Dynamics (CFD) simulation which were used as learning features. In order to generate learning data, 210 cases of CFD simulation for 10 wind speeds, 7 wind directions, and 3 greenhouse ventilation window types were performed. Multi linear regression (MLR), SVR, Random Forest, and deep neural network (DNN) were developed for machine learning models, and the optimization of hyper-parameters for each machine learning model was performed. For each optimized machine learning model, the bootstrapping technique was applied to supplement a relatively small number of learning data. The results show that models such as RF, DNN and SVR using the rbf kernel function, which have already shown high accuracy ($R^2 >$ 0.9), were less accurate after applying bootstrapping. while MLR and SVR models which used poly-kernel functions have improved R^2. Finally, in order to minimize

the simulation cases, which significant amount of time, the accuracy of the machine learning model was evaluated according to the reduction of the CFD simulation. Result showed that most machine learning models have high accuracy for the learning data considered from 120 cases with only 4 cases of wind direction. The RMSE value was also evaluated low for the model indicating the lowest error for the predicted value.

The greenhouse environment prediction and control system proposed in this dissertation predicted the current (PCTO-ML) and future (PETO-ML) air temperature inside the greenhouse and local ventilation rate (PLV-CFD driven ML). When the three models constituting the system are linked, it is of great significance that it is possible to predict the internal air temperature of the natural ventilation greenhouse and to control the proper ventilation of the predicted air temperature. In particular, the optimal sensor location proposed through the PCTO-ML model is expected to contribute to improving farmers' income by reducing the initial and maintenance costs of the greenhouse. In addition, the PVL-CFD driven ML model was able to predict the natural ventilation rate for all wind environment conditions because it was possible to learn about environmental conditions that were difficult to measure due to the small frequency of occurrence. Since only minimum number of sensors are required for the development and operation of the system, it is economical to introduce these models and is expected to be applicable to most domestic greenhouses where air temperature sensors are installed.

Keyword: Computational fluid dynamics, environmental prediction, greenhouse, machine learning, natural ventilation, optimal sensor location

Student Number : 2015-30469

Table of Contents

List of Figures

List of Tables

Chapter 1. Introduction

1.1. Study Background

Crop production through facility cultivation is continuously developing over the past several decades by ensuring a year-round stable and high-quality production. Data from the Ministry of Agriculture, Food and Rural Affairs, South Korea (MAFRA) indicates that the scale of production is constantly increasing since the 1980s with 52,444 ha dedicated for domestic vegetable greenhouse in 2020 (MAFRA 2022). Meanwhile, the domestic agricultural population is aging and declining, while the demand for agricultural products is growing. In addition, there is an increase in automation and the number of large-scale greenhouses. The area of multi-span greenhouses in Korea increased from 5,227 ha in 2012 to 7,088 ha in 2019 (MAFRA, 2022). A large greenhouse enables easy automation and improved productivity.

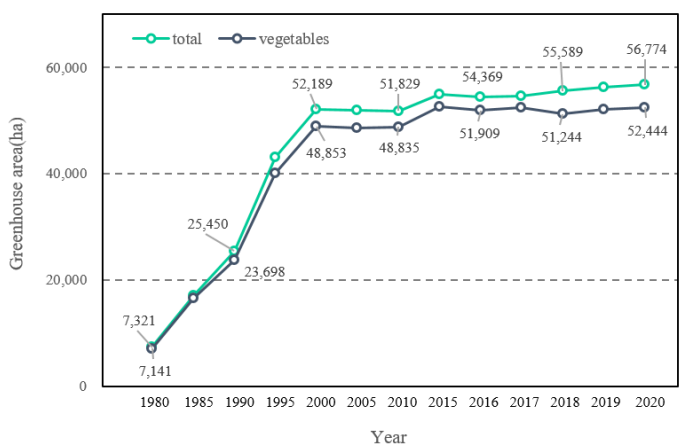


Figure 1-1 Installation area of domestic vegetable and horticultural greenhouses (MAFRA, 2022)

The development of ICT technology and the expansion of investments in smart farms are accelerating the demand of smart greenhouses which have many sensors and control systems. Smart farm employs the practice of precise management of productivity, accurate and timely monitoring of internal environments, analysis of big data using artificial intelligence, control of growing environment using ICT devices such as smartphone, and enhanced convenience (Yun, Lee et al. 2017). Smart farms can increase production by 20-30%, reduce labor, and improve convenience due to remote management (김태완 2019). The area of horticultural smart greenhouses in South Korea rapidly increased from 405 ha in 2014 to 6,485 ha in 2021. The global market for smart greenhouse has grown from 1.25 billion USD in 2020 to 1.85 billion USD in 2025 (MAFRA, 2022). More than 96% of the smart farms had been installed with temperature and humidity sensors to monitor the environment inside the greenhouse. Whereas, 40% of the smart farms were installed with environmental sensors such as carbon dioxide, solar radiation, geo temperature, and humidity. Additionally, various equipment and sensors to monitor the smart greenhouse such as PC systems, mobile systems, and CCTV were installed for further greenhouse management (MAFRA, 2022). These various equipment and systems enable precise environmental monitoring and control inside the greenhouse. However, it will certainly contribute to an increased installation and management costs of smart farms.

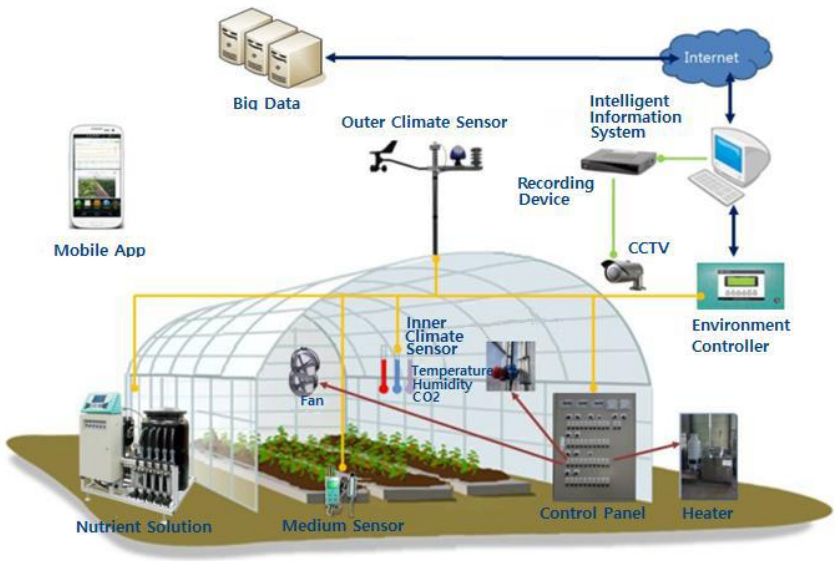


Figure 1-2 Greenhouse with numerous sensors and ICT equipment for greenhouse environment monitoring and control (Oh 2017)

In smart farms, the installation costs were rated by the respondents as the top most problem (38.3%) among the identified problems in installing smart farms as shown in Table 1-1 (MAFRA, 2022). In addition, 43.8% of the respondents from the horticultural smart greenhouses indicated that they had difficulties in using the smart farm. Among the difficulties, frequent failure of sensors and restrictions on connectivity in the existing installed equipment were often experienced (Table 1-2). It leads to the economic burden of greenhouse farmers. Especially, for application of smart greenhouses and advanced ICT technologies, modern greenhouses are equipped with several equipment such as various sensors and systems for environmental control. Another problem is the introduction large-scale smart farm equipment and number of sensors become challenging because most greenhouses in South Korea are single-span plastic greenhouses with weak durability (김태완 2019). Despite the strong and favorable advantages of smart farms such as reduction

of labor and improvement of productivity, still, there are several limitations that hinder the implementation of smart farms coupled with the use of ICT technologies due to the challenge brought by the difficulty in installation and relatively high management costs.

With the abovementioned challenges, it is therefore necessary to monitor the growing environment inside the greenhouse while minimizing the installation of excessive environmental sensors. For economic purposes, a technology and strategic technique for utilizing the sensors that were already installed in the existing greenhouse should be observed. Additionally, various sensors and control systems were needed to apply technologies to most existing greenhouses as well as modernized greenhouses.

With the recent development of machine learning algorithms and computer performance, many machine learning models were developed and are being used for industrial applications. Especially, the machine learning model could be applied to simple data measured from a general greenhouse and even to various data from a smart greenhouse. Machine learning models are generally of high accuracy with the capability to predict phenomena by identifying trends and patterns in a given range of gathered data without human intervention. In addition, it has an advantage that accuracy can be improved according to the accumulation of data. Since the machine learning model is a black box model, the physical relationship between the learning variable and the predictor is not essential. Therefore, machine learning model is well suited for predicting environments such as air temperature, humidity, etc. in greenhouses that are affected by various factors and have non-linear characteristics. Additionally, if the accuracy of the machine learning model is guaranteed, various variables can be predicted using minimal sensors. In other words, one air temperature

sensor can predict air temperatures at various locations as well as other types of environmental data such as relative humidity. Therefore, it may be used to minimize the sensor to be installed in the greenhouse, and as a result, the initial installation cost and maintenance cost can be reduced.

Table 1-1 Investigation of the common problems experienced in the smart farm introduction process (MAFRA, 2022) (unit : %)

Problems	Total	Crop						
		Tomato	Strawberry	Paprika	Cucumber	Citrus	Flower	Others
Installation cost	38.3	41.7	26.1	25.3	17.5	100	100	57.1
Low understanding of smart farm technology and equipment	28.8	22.5	35.9	40.5	66.0	-	-	19.2
Communication with installation company	16.0	15.2	20.7	14.5	-	-	-	17.4
Limitations of existing facilities for smart farm installation	8.3	10.6	9.6	5.3	16.5	-	-	-
Difficulty in preparing additional infrastructure	6.5	8.5	4.8	9.7	-	-	-	6.3
Others	2.1	1.5	3.0	4.7	-	-	-	-

Table 1-2 Investigation of difficulties in smart farm (MAFRA, 2022) (unit : %)

Type of crops	Difficulties					
	None	Lack of proactive response from companies	Frequent sensor and equipment failures	Difficulty in using installed smart farm equipment	Reduced utilization due to limited connectivity with existing facilities	Others
Horticulture	56.2	14.5	13.9	8.7	5.9	0.7
Open-field (fruit)	27.3	24.2	25.7	17.1	4.8	0.9
Open-field (vegetable)	23.2	29	24.5	8.3	15.1	-

1.2. Purpose of Research

The management of the internal environment is crucial to maintain an appropriate growing condition inside the greenhouse. The objective of this dissertation was to develop ML models for monitoring and predicting the internal environment of a greenhouse naturally ventilated. Machine learning model was developed to serve three different purposes. First, the model was developed to select the optimal sensor location for air temperature, which was one of the most important environmental factors inside the greenhouse. Second, the model was developed for preemptive control of HVAC systems by predicting the temperature inside the greenhouse as well as monitoring the current indoor air temperature. Third, the model was to quantitatively predict natural ventilation, which is one of the most basic method to control the optimal growing environment in the greenhouse such as air temperature and relative humidity. Through these three machine learning models developed in this study, the air temperature of a naturally ventilated greenhouse can be monitored on a real time basis and likewise can be predicted. The local ventilation rate of the naturally ventilated greenhouse according to external weather conditions could be predicted using the developed models. Figure 1-3 shows the overall research flow of this dissertation.

The overall background and the purpose of this dissertation were explained in Chapter 1. In Chapter 2, comprehensive literature reviews on a) importance of growing environment inside greenhouses; b) evaluation methods for ventilation of greenhouses; c) optimal sensor location for monitoring the internal environmnets; and d) application of machine learning in agricultural field were conducted to build the foundation and to establish the appropriateness of the dissertation.

In Chapter 3, the machine learning model was developed to suggest the optimal sensor location for monitoring the internal air temperature of the naturally ventilated greenhouse. The internal and external environmental data of naturally ventilated 8-span greenhouse such as air temperature, relative humidity, soil temperature, soil humidity, UV radiation were measured. Since the experimental greenhouse used in this study was naturally ventilated throughout the experimental period, a portable weather station was installed to consider the influence of the external environments such as external wind direction, wind speed, and air temperature. Meanwhile, since it is essential to establish reliable training data to develop a machine learning model with good performance data preprocessing such as linear interpolation and normalization was performed. Additionally, correlation analysis was conducted to select the learning feature of the machine learning models. For the development of machine learning models, artificial neural network, long short-term memory, and support vector regression were applied. The machine learning model with the best performance was selected by comparing the accuracies in predicting the air temperature inside the greenhouses according to several kind of machine learning models. The simplification of the learning data was analyzed to minimize the sensor installation for developing the machine learning model. The simplified model was developed by using the learning data selected through correlation analysis. Selection of the simplified machine learning model was based on minimum accuracy. Finally, the optimal sensor location for the temperature prediction inside the greenhouse was evaluated using the selected machine learning model.

In Chapter 4, the LSTM model for predicting the internal air temperature of naturally ventilated greenhouse was developed. The collection of training data for experimental greenhouse was conducted like similar with Chapter 3. For an LSTM

model that has strengths in processing time series data, the performance of the model is calculated differently depending on the sequence length, which is a unit of input and output of LSTM model. Therefore, the appropriate sequence length of the LSTM model was evaluated. The internal air temperature for the near future of the greenhouse was predicted and evaluated using the LSTM model with an appropriate sequence length. At this time, the sensor location of measuring the data used as the learning data was same with the location of the sensor for the predicted air temperature. However, the machine learning model proposed in this study have an objective to predict the air temperature inside the greenhouse minimizing the sensor installation. As such, the optimal sensor location presented in Chapter 3 was applied to the LSTM model of Chapter 4 and was analyzed. Consequently, the measured data at the optimum sensor location was used for learning the LSTM model to predict air temperatures at other sensor locations.

In Chapter 5, the machine learning model was developed to predict the local ventilation rate of the naturally ventilated greenhouse. CFD simulation, which can generate data for several environmental conditions designed by the researcher, was used to generate training data for the machine learning model. In this study, the CFD model validated in previous studies was used to accumulate the learning data. Using the validated CFD model, 10 wind speeds, 7 wind directions, and 3 vent openings were considered as training features of the machine learning model. Multiple linear regression, support vector regression, random forest, and deep neural network models were developed as machine learning models. The training data of the machine learning model proposed in this study was generated through CFD simulation. With this, it is important to secure the minimum preformance of CFD simulation for the machine learning model. Hence, the simplified machine learning

model was evaluated by anlayzing the accuracy of the machine learning model according to the number of CFD cases and applying the bootstrapping technique.

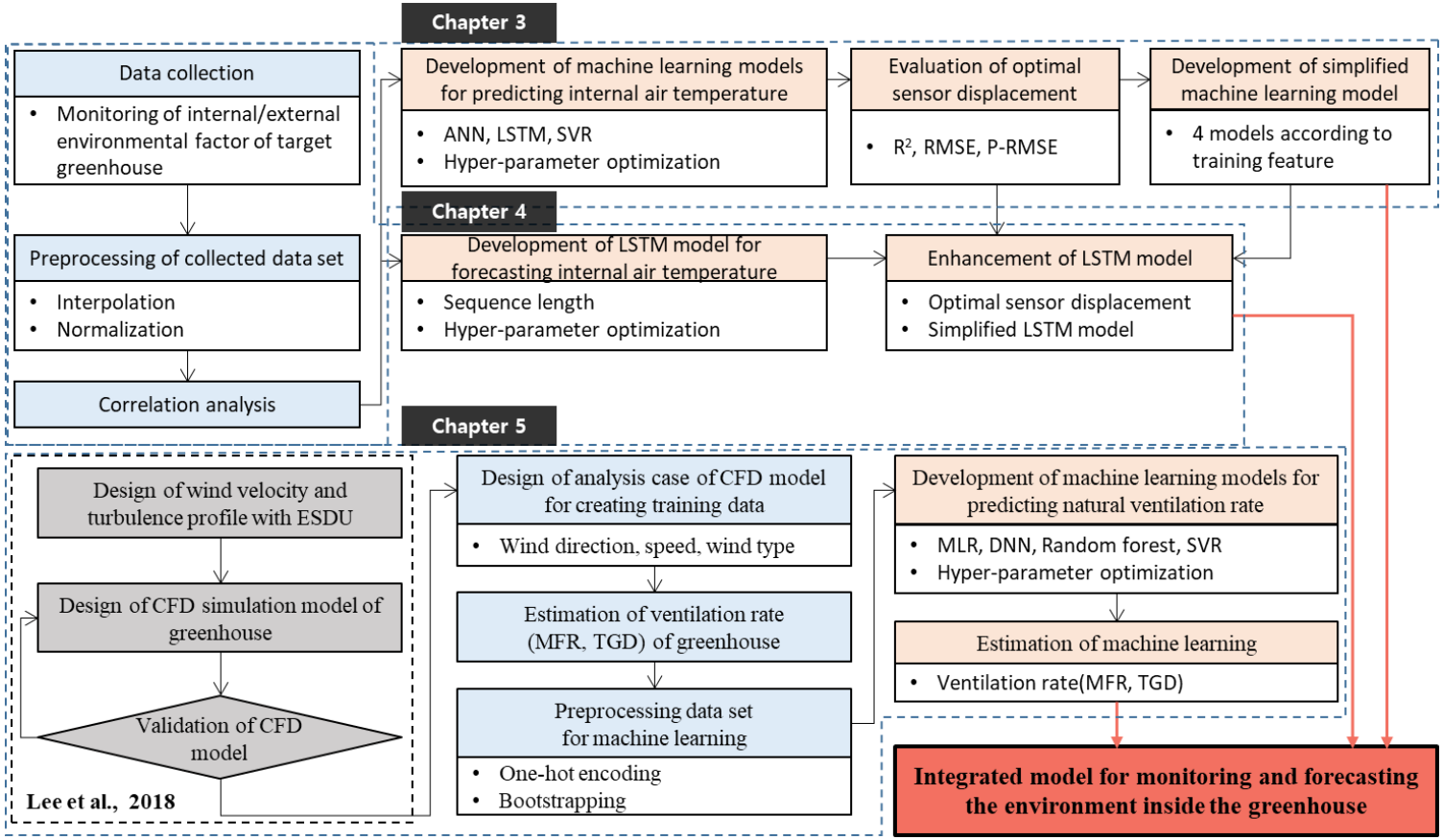


Figure 1-3 Research flowchart of this dissertation for greenhouse environment prediction and control model development

Chapter 2. Literature review

2.1. Importance of the growing environment inside the multi-span greenhouse

Facility agriculture offers the possibility to regulate environmental growing parameters such as temperature and relative humidity to meet the desired requirements of the crop. It offers the advantage to control the internal environment regardless of the external weather unlike field cultivation. The importance of facility agriculture has been intensified with its unique feature to allow and maintain supply of high quality crops throughout the year. For this reason the area of cultivation facilities and agricultural production had increased for the past several years. The need of larger facilities is emerging to improve productivity. However, the use of large facility is faced with technical problem such as non-uniformity of air temperature, humidity, and airflow in large. This can cause a decrease in productivity because of adverse effects on the growing environment for the crops (Hernández, Morales et al. 2002, Holcman and Sentelhas 2012, Kutta and Hubbart 2014, Marković, Pavlović et al. 2014, Costantino, Comba et al. 2019, Jia, Wei et al. 2019, Lee, Chung et al. 2019, Cayli 2020).

The improvement of the environmental growing parameters inside the enlarged greenhouse and the provision to maintain a uniform growing environment throughout the cropping season are important factors. For instance, the internal temperature of the greenhouse has a significant effect on flowering and fruiting depending on the growing season of crops. Coherent to the observation in other works, productivity differs in the uniformity of the internal environments (유인호,

조명환 et al. 2007, Cayli 2020). Additionally, the uniformity of the microclimate environment such as humidity and carbon dioxide related to the growth of crops affect productivity (Holcman and Sentelhas 2012, Lee, Chung et al. 2019).

The air temperature in the greenhouse is one of the factors affecting plant growth. Several studies were conducted to analyze the ventilation configuration and the thermal environment in greenhouses according by considering various environmental conditions (Boulard, Meneses et al. 1996, Okushima, Sase et al. 2000, Zhao, Teitel et al. 2001, Bartzanas, Boulard et al. 2004, Kacira, Sase et al. 2004, Baeza, Pérez-Parra et al. 2009, Fidaros, Baxevanou et al. 2010, Ishii, Okushima et al. 2014, 하정수 2015, Boulard, Roy et al. 2017, Fatnassi, Boulard et al. 2017, Geng, Zhang et al. 2019, Rasheed, Kwak et al. 2020, 박민정, 최덕규 et al. 2020, Cheng, Li et al. 2021, Lin, Zhang et al. 2021, Ogunlowo, Akpenpuun et al. 2021, Chauhan and Lunagaria 2022, Xia, Nan et al. 2022). (박민정, 최덕규 et al. 2020) investigated the specifications and ventilation windows of multi-span greenhouses in South Korea, and analyzed the airflow and temperature distribution inside the greenhouse according to the ventilation windows using CFD. In particular, the temperature of the height of the crop zone according to the type of roof window was analyzed. (Cheng, Li et al. 2021) developed a three-dimensional real-time greenhouse temperature simulator based on virtual sensors. Data such as temperature, humidity, and wind speed measured in real time inside and outside the greenhouse were analyzed, and regression equations were derived. The measured data were compared with CFD simulations. Finally, temperature monitoring at various points in the greenhouse was presented using the proposed virtual sensor. Fatnassi, Boulard et al. (2017) analyzed the thermal environment and airflow patterns in the

greenhouse according to the number of spans using CFD (Fatnassi, Boulard et al. 2017). The thermal environment and airflow patterns in the greenhouse according to the height of the greenhouse were also analyzed. The results in previous study showed that the heating cost increased exponentially with the height of the greenhouse. The internal environment become non-uniform as the span of the greenhouse increases. The study concluded that structure design of greenhouse such as the number of span and height have to be considered when designing large scales. (Ishii, Okushima et al. 2014) worked on the natural ventilation efficiency of an open-roof design as a mean to increase the natural ventilation efficiency of greenhouses. The temperature difference between the inside and the outside during summer was comparatively analyzed for an arch roof greenhouse and an open roof greenhouse. Likewise, (Baeza, Pérez-Parra et al. 2009) evaluated the ventilation efficiency of compared roof ventilation with combined roof and sidewall vents at the conditions of buoyancy effect using CFD. The effect on ventilation was analyzed according to insect screen on opposing sidewall vents. Ventilation efficiency was evaluated by analyzing the distribution of thermal environment in multi-span greenhouse using CFD simulation. (Chauhan and Lunagaria 2022) analyzed the thermal environment at condition of the open ventilated greenhouse by measuring the temperature at 12 sampling points. This study presented a regression model for predicting microclimate conditions inside a greenhouse relating relative humidity, air temperature, and soil temperature in the open ventilated greenhouse.

Previous studies conducted analysis of the thermal distribution utilizing data from specific point such as the central location of the greenhouse or around the ventilation openings for monitoring the temperature environment in the greenhouse. The analysis was performed based on the air temperature distribution and average

temperature for the entire greenhouse.

2.2. Estimation methods of ventilation for greenhouse

It is important to monitor the growing environment for proper control because the microclimate environment inside the greenhouse has a great effect on the growth of crops. Ventilation is the one of the best methods to control the micro-climate such as temperature, humidity, and CO_2 in the greenhouse. Ventilation in the greenhouse is a method for controlling the environment inside the greenhouse by exchanging the air inside and outside the greenhouse.

Generally, insufficient ventilation inside the greenhouse causes high temperature stress, diseases due to high humidity, and insufficient air flow near the leaves on the crops. As a result, the air exchange of the crop leaves and the concentration of carbon dioxide both decreases. It retards the photosyndissertation rate and decrease crop productivity (Kitaya, Shibuya et al. 1998, Shibuya and Kozai 2001, Shibuya, Tsuruyama et al. 2006, Holcman and Sentelhas 2012, Radojevic, Bjelogrlic et al. 2012, Kutta and Hubbart 2014, Marković, Pavlović et al. 2014).

The quantification and evaluation of the ventilation of greenhouse can be done through several methods such as field experiments, CFD simulation, and machine learning model. A number of studies were already conducted to quantitatively evaluate the ventilation of greenhouses following field experiment method using direct measurement and tracer gas method (Sase 1988, Fernandez and Bailey 1992, Kittas, Draoui et al. 1995, Boulard, Meneses et al. 1996, Wang, Boulard et al. 2000, Molina-Aiz, Valera et al. 2004, Mashonjowa, Ronsse et al. 2010), energy balance model (Kozai 1980, Chalabi and Bailey 1989, Fernandez and Bailey 1992, Boulard, Baille et al. 1993, Boulard and Draoui 1995, 남상운, 김영식 et al. 2012), and pressure difference model (Boulard, Meneses et al. 1996, Kittas, Boulard et al. 1996,

Papadakis, Mermier et al. 1996).

Previous studies on ventilation evaluation using field experiments were conducted, and a ventilation evaluation study was studied through theoretical equations using field experiments. (남상운, 김영식 et al. 2012) analyzed the effect of roof ventilation and installation standards of a single-span plastic greenhouse by quantifying buoyance driven ventilation, wind driven ventilation, and combined ventilation using energy balance model. (Kim, Lee et al. 2002) quantified the ventilation efficiency associated with side and roof window in a greenhouse with folding panel type windows based on an energy balance model. (Bot 1983) analyzed effect on ventilation of greenhouse by measuring the internal airflow relative to the shape of vent openings, opening angle, wind speed, wind direction, and temperature difference. As a result, this study presented that the ventilation of venlo-type greenhouse was affected by the change in pressure. (Sase, Takakura et al. 1983) conducted experiments to investigate changes in internal air flow and temperature distribution following the shape of the ventilation openings and external wind speed through a wind tunnel experiment. The study revealed that a wind speed was under 2 $m \cdot s^{-1}$, buoyancy and wind driven ventilation occur at the same time, and a stagnant zone in which the air inside and outside the greenhouse is not completely mixed, occurred resulting in an unstable distribution of temperature. Furthermore, wind speed above 2 $m \cdot s^{-1}$ was entirely governed by the external wind speed. (Fernandez and Bailey 1992) developed the equation for ventilation rate with respect to wind speed and angle of vent opening. The internal temperature of greenhouse was simulated using the energy balance model. (Boulard, Papadakis et al. 1997) investigated the ventilation of two-span greenhouse at the wind condition blowing parallel to the side windows. (Lee, Sase et al. 2003) examined the effect of roof vent

opening in a venlo-type greenhouse on ventilation efficiency through PIV tests and wind tunnels

The field experiment is one of the best methods to quantitatively analyze the ventilation and internal airflow pattern. However, this method may encounter several experimental errors because of the invisible air flow. This is especially that it is difficult to maintain a constant concentration of the tracer gas in the target space for the tracer-gas decay method. There are also inherent experimental difficulties in controlling invisible and intangible gases. Additionally, an experimental environment suitable for the researcher's experimental plan could not be controlled to quantify the ventilation rate because the external wind environment of the greenhouse could not be controlled artificially.

Therefore, simulations using computational fluid dynamics have been conducted as a supplement method for field experiment (Brugger, Short et al. 1987, Fatnassi, Boulard et al. 2002, 홍세운 and 이인복 2014, 조규정, 김기영 et al. 2015, Kwon, Jung et al. 2017, Lee, Lee et al. 2018, Li, Li et al. 2020, Kim, Kim et al. 2021). Computational fluid dynamics is a numerical analysis technique that estimates phenomena such as fluid flow, heat transfer, and related chemical reactions through computer-based simulations. The computational fluid dynamics as it simulates the flow of air and gas can eliminate errors that may be caused by experiments dealing with gas. Additionally, it is possible to simulate various environmental conditions suitable for the research purpose because all the conditions designed by the researcher can be adjusted for the simulation. (홍세운 and 이인복 2014) simulated the micro-climate inside the greenhouse and the ventilation according to various weather conditions and ventilation configuration of the greenhouse using

CFD. Finally, a numerical model was presented through sub-regression analysis. (조규정, 김기영 et al. 2015) evaluated the ventilation efficiency of the greenhouse by simulating the air flow in a single-span greenhouse according to the conditions of vent opening and the operation of the fans using CFD. Additionally, the ventilation efficiency was evaluated by simulating the temperature profile inside the greenhouse according to the insulation curtain. (Kwon, Jung et al. 2017) evaluated the ventilation rate of a naturally ventilated multi-span greenhouse using CFD. The ventilation efficiency of the greenhouse was evaluated by TGD method and MFR method. The results of the study indicated that the height of the side window of the greenhouse affects the ventilation rate, and that the most efficient ventilation is achieved at a height of 1 m. (Brugger, Short et al. 1987) improved the ventilation configuration of the greenhouse by performing CFD simulation in 2D steady state. (Fatnassi, Boulard et al. 2002) analyzed the ventilation performance of a large-scale Canarian-type greenhouse with insect screens installed on ventilation openings. The actual air exchange rate using tracer gas was compared with the predicted values from the CFD model. With respect to the wind direction, the ventilation rate increases proportionate to the wind speed and the size of the ventilation opening. The chimney effect could be neglected when the wind speed is greater than 2 m/s. The insect screen causes an additional large pressure drop through the ventilation openings, resulting in decrease of the ventilation rate and increase of the temperature in greenhouse. (Lee, Lee et al. 2018) quantitatively analyzed the ventilation rate of single-span greenhouse built on reclaimed land according to greenhouse types, vent openings, wind directions, and wind speed based on MFR and TGD methods. The chart for predicting the natural ventilation rate of greenhouses built in a reclaimed land was developed using the CFD simulation results according to various environmental conditions. (Kim, Kim

et al. 2021) analyzed the internal aerodynamic environment of 1-2W type greenhouse using CFD simulation. The CFD model was validated using the temperature and wind speed measured in the actual greenhouse. The natural ventilation rates were calculated according to the shading screen using the validated CFD model. The appropriate duct perforations and perforation angles were also suggested using the validated CFD model. (Li, Li et al. 2020) used CFD simulation to predict the airflow pattern and thermal behavior and to evaluate the ventilation efficiency of the arched greenhouse. This research suggested the position angle of the arched greenhouse and the arch chord angle, which represent the optimal ventilation efficiency. (Lee, Gwon et al. 2022) evaluated the ventilation efficiency of the Venlo-type advanced digital greenhouse using CFD. The amount of air inflow through the roof vent opening according to the height of the greenhouse was calculated.

Figure 2-1 Foreground of Venlo-type advanced digital greenhouse with high elevation (Lee et al., 2022)

Previous studies also evaluated the local ventilation by analyzing the air flow pattern inside the greenhouse using CFD aside from the optimal ventilation of the greenhouse according to various environmental variables. However, in most studies

using CFD, the number of CFD cases were limited, and only specific cases for the research subject were simulated. Furthermore, it is impossible to simulate environmental conditions in the field which are changing in real time using CFD simulation. Therefore, the studies of calculating the ventilation rate using CFD have been used for designing ventilation structures and presenting standards for ventilation control.

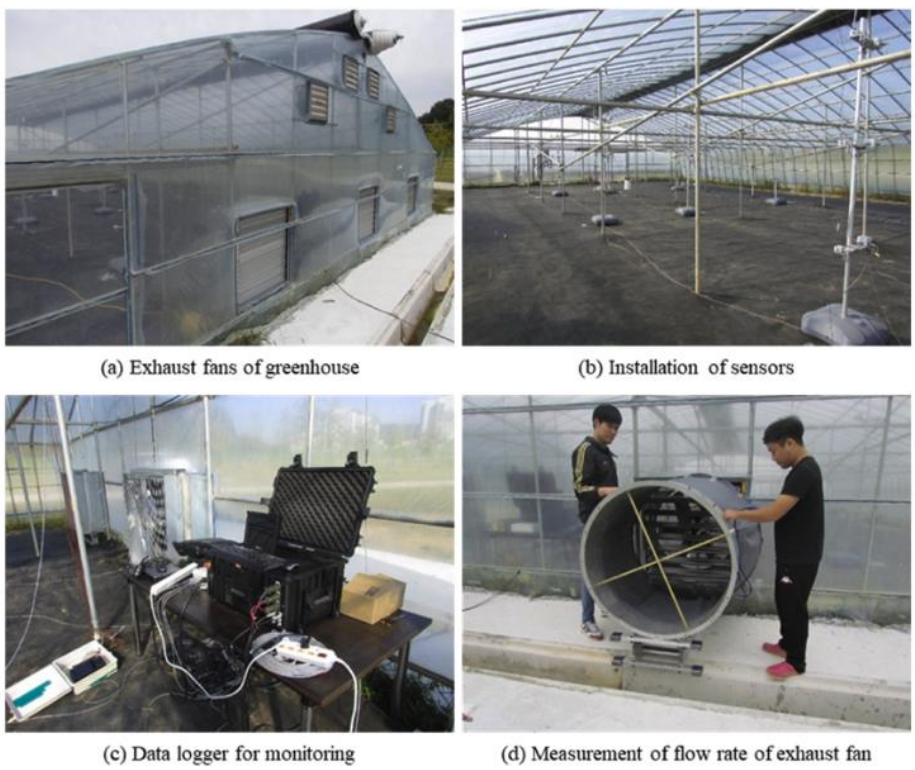


Figure 2-2 Measurement of air temperature and airflow in greenhouse for validation of CFD model (Kim et al., 2021)

Various studies have been conducted with the development of artificial intelligence algorithms and the development of various technologies for collecting and processing big data. Machine learning is for computer to learn information

processing ability by deriving patterns of data. Machine learning generally has the advantage of high accuracy and relatively fast computation time compared to several numerical analyses. With these advantages, studies were conducted to estimate the ventilation rate and internal environment of buildings (김상엽, 박경섭 et al. 2018, Tsai, Hsu et al. 2020, Park, Jeong et al. 2021, Saadon, Lazarovitch et al. 2021, Suo, Zhang et al. 2021).

(김상엽, 박경섭 et al. 2018) developed a model to predict the internal temperature for environmental control of a venlo-type greenhouse using a machine learning model. Several machine learning algorithms such as artificial neural network, recurrent neural network, and multiple regression models were developed and their prediction accuracy were compared. (Saadon, Lazarovitch et al. 2021) developed a machine learning model to estimate net radiation for a naturally ventilated greenhouse. Various measuring devices such as infrared gas analyzer system, thermocouple, 2D sonic anemometer, and pyrometer were used to measure the learning data. (Park, Jeong et al. 2021) estimated the natural ventilation rate of office buildings using eight machine learning models. The machine learning models were developed by learning the air temperature, relative humidity, wind speed and wind pressure difference measured in the office building in summer. The significant variable was evaluated for predicting the natural ventilation rate of the office building.

Figure 2-2 Outside view of the naturally ventilated greenhouse(GH-II) at the north position(a), lateral controller, and curtains closed in the semi-closed greenhouse(GH-I) at the south position (B). General view of the semi-closed greenhouse(GH-I) during an opening event(C) (Saadon et al., 2014)

Most studies that have been conducted, however, were only focused on the estimation of the internal air temperature of greenhouses and buildings. There were few research works that evaluated the ventilation rate of the greenhouses. Furthermore, most of the previous studies used the field measured data to develop machine learning model. For sufficient data accumulation in order to develop machine learning models, it is necessary to measure data over a considerable period. The variables for learning data consisted of the measurable data from the field experiments. In several previous studies, missing values and outliers were usually

encountered at the field experiments. Data preprocessing was performed to construct high-quality learning data. Because the performance of the machine learning model is absolutely dependent on the amount of learning data, a lot of effort was needed to accumulate the reliable data and preprocess the measured data.

Based on the results of reviewing various previous studies, the summary and characteristics of major studies are summarized and shown in Table 2-1.

Table 2-1 Previous researches for estimating the ventilation in greenhouses and the advantages and disadvantages of each method

Methods	Advantages	Disadvantages	Reference
Field experiment and energy balance model	It is the most intuitive and direct experimental method It is the most traditional method	There is time and economic cost required for experiment time and equipment Experimental error inevitably exists Experimental conditions such as the external wind environment cannot be artificially adjusted.	Sase 1988; Fernandez and Bailey 1992; Kittas; Draoui et al. 1995; Boulard; Meneses et al. 1996; Wang; Boulard et al. 2000; Molina-Aiz; Valera et al. 2004; Mashonjowa; Ronsse et al. 2010; Kozai 1980; Chalabi and Bailey 1989; Fernandez and Bailey 1992; Boulard; Baille et al. 1993; Boulard and Draoui 1995; Nam et al. 2012
CFD simulation	Experimental conditions can be adjusted artificially There are no errors due to field experiments	Computation time for simulation is usually long As the number of case operations increases, more time is consumed. It is impossible to analyze environmental conditions that change in real time in the field.	Brugger; Short et al. 1987; Fatnassi; Boulard et al. 2002; Hong and Lee 2014; Cho et al. 2015; Kwon; Jung et al. 2017; Lee et al. 2018; Li et al. 2020; Kim et al. 2021
Machine learning model	Computation time is relatively short. It generally has high accuracy.	It shows low accuracy on data other than the trained data.	Kim et al. 2018; Tsai; Hsu et al. 2020; Park; Jeong et al. 2021; Saadon; Lazarovitch et al. 2021; Suo; Zhang et al. 2021

2.3. Optimal sensor location for monitoring the internal environment of the greenhouse

Agricultural facilities have the advantage to be controlled so that parameters necessary for plant growth such as air temperature, relative humidity, radiation, and CO_2 concentration will be set to their optimum values (Medela, Cendón et al. , Pang, Chen et al. 2015, Vatari, Bakshi et al. , Brewster, Roussaki et al. 2017, Muangprathub, Boonnam et al. 2019, Rayhana, Xiao et al. 2020). Especially, the growth, quality, and productivity of crops are greatly affected by environmental factors such as air temperature and humidity. The growth environment of crops is primarily adjusted through natural and mechanical ventilation of greenhouses. The appropriate growing environment are also maintained by the heating and cooling systems. Therefore, it is important to monitor the environmental factors for the productivity and appropriate growing environmental parameters in the greenhouse to improve productivity. For this, environmental monitoring in the greenhouse is essential. Recently, smart farm greenhouses were integrated with the development of information and communication technology (ICT) and various sensors (Arif and Abbas 2015, Kodali, Jain et al. 2016, Mittal, Sarangi et al. 2018, Rayhana, Xiao et al. 2020). Many technologies and equipment have been applied to attain the appropriate growing environment in the greenhouses. Smart greenhouse is an intelligent farm that automatically maintains and manages the optimal growth environment without time and space restrictions by using ICT and data. The environmental control systems of the smart greenhouse can be operated using the measurement date of various environments in the greenhouses.

Previous studies have been conducted to apply ICT and various sensors to smart

greenhouse (Kodali, Jain et al. , Mittal, Sarangi et al. , Wu, Liu et al. 2014, Al-Bahadly and Thompson 2015, Arif and Abbas 2015, Nayyar and Puri 2016, Athani, Tejeshwar et al. 2017, Zhang, Zhang et al. 2017, Danita, Mathew et al. 2018, Marques and Pitarma 2018, Yimwadsana, Chanthapeth et al. 2018, Rayhana, Xiao et al. 2020). Various climate sensors and soil condition sensors were used in smart greenhouses (Kodali, Jain et al. , Mittal, Sarangi et al. , Wu, Liu et al. 2014, Al-Bahadly and Thompson 2015, Zhang, Zhang et al. 2017, Danita, Mathew et al. 2018, Marques and Pitarma 2018, Yimwadsana, Chanthapeth et al. 2018, Rayhana, Xiao et al. 2020). In addition, Plant growth monitoring systems were installed to manage crop field and production in smart greenhouses (Okayasu, Nugroho et al. , Slamet, Irham et al. , Hadabas, Hovari et al. 2019). Furthermore, plant diseases monitoring systems were established to take preemptive action against the spread of disease (Patil and Kale , Jumat, Nazmudeen et al. 2018, Kitpo and Inoue 2018, Materne and Inoue 2018). However, the installation of numerous sensors causes an increase in the initial installation cost of the greenhouse along with added cost for the maintenance of sensors. In fact, compared to low-tech greenhouse, high-tech greenhouse was calculated to be 4-8 times higher in construction cost per m^2 according to sensor and monitoring system installation (Tognoni, Pardossi et al. 1997, Pardossi, Tognoni et al. 2004, Rayhana, Xiao et al. 2020). With this, it is important to install minimum sensors in the greenhouse in the optimal location to monitor the entire environment inside the greenhouse and reduce initial and maintenance costs.

Related to the optimal sensor installation, the several previous studies were conducted to enhance the performance of remote sensor for stable data transmission and improvement of durability (Guerriero et al., 2011; Keskin et al., 2014; Toumpis & Tassiulas, 2006; Wang et al., 2005), to monitor the structural stability of

greenhouses (Hwang et al., 1996; J. Lee et al., 2009; Y. Lee et al., 2016; Yi et al., 2017; Lee, Kim, & Lee, 2016; Ren, Yan, & Jiang, 2001; Yi, Zhou, Li, & Wang, 2017), and to detect abnormal circumstances in the measured values of the sensors (Choi and Kim 2019, Ou, Chen et al. 2020). The previous studies, however, have little relevance on optimal environmental control in the greenhouses although information from the previous studies are important for the stable installation of sensors in the greenhouse and monitoring the durability of the greenhouse.

Several studies were also conducted to determine the optimal sensor locations for monitoring the internal air temperature and relative humidity of the greenhouse (C. Chen & Gorlé, 2022; Y. L. Chen & Wen, 2010; Feng et al., 2013; Hamel et al., 2006; Huang et al., 2014; X. Liu & Zhai, 2009; Y. Liu et al., 2014; Mazumdar & Chen, 2008; Waeytens et al., 2019; Wijaya et al., 2021; Zhang & Chen, 2007).

(이민구 and 정경권 2012) simulated the temperature change inside the building using CFD. The location with the least temperature change served as basis for the optimal sensor location. The optimal sensor location suggested by the CFD model was verified by installing 30 actual temperature sensors in the experimental building. (Guzmán, Carrera et al. 2018) suggested the development of a model-based virtual sensor based on CFD and control. Temperature analysis using CFD was performed for a small-scale greenhouse and one virtual sensor was proposed. (Utami, Yanti et al. 2021) simulated the heat distribution inside the building for various environmental conditions using CFD simulation. Finally, using the simulation results, the optimal sensor location was determined based on the thermal characteristics inside the building. (Chen and Gorlé 2022) optimized the placement of temperature sensors in buildings with buoyancy-driven natural ventilation using CFD simulations. By simulating the temperature distribution of the building, the optimal sensor

location was determined considering the average value and dispersion of temperature.

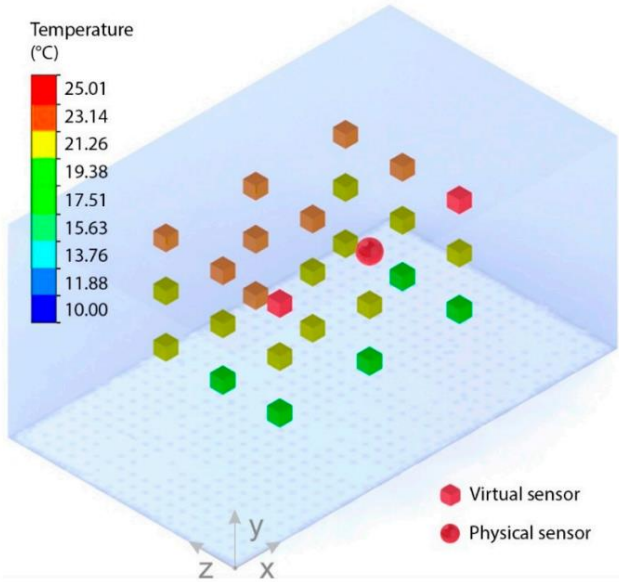
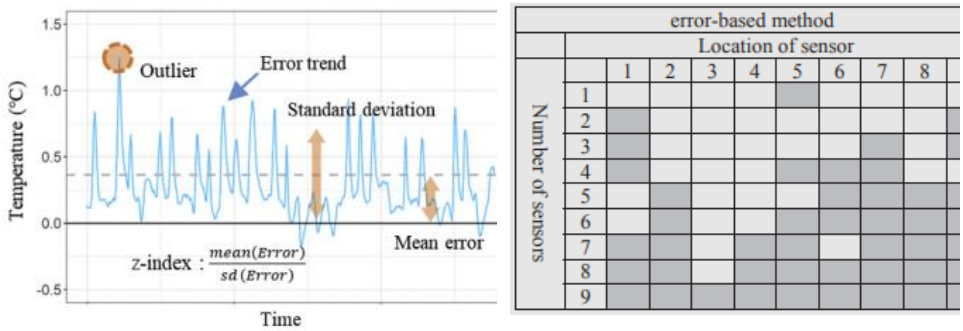


Figure 2-3 Three-dimensional (3D) real-time simulator (Guzmán, Carrea et al. 2018)

The studies concerning optimal sensor placement have also been explored with the application of theoretical formulas and algorithms. (Yoganathan, Kondepudi et al. 2018) proposed the optimal sensor location in the office building using clustering algorithms, information loss approach and paeto principle. Real-time data was analyzed for three indoor environmental parameters such as temperature, humidity, and luminance. As a result, the optimal sensor locations to minimize the loss of sensing data were analyzed and established in which 31 sensors initially used in the experiment were optimized to 6 sensors. (Chang, Ha et al. 2012) suggested the optimal sensor location to measure the pressure of the water distribution system using the entropy theory and genetic algorithm. Based on entropy theory, the sensor location that provides the most information from the entire system was selected as the optimal sensor location. (Lee, Lee et al. 2019) proposed an error-based and

entropy-based sensor placement. The sensor locations which could represent the entire environments inside the greenhouse were determined through error based method. Whereas, locations to detect the abnormal circumstance were selected through entropy based method. By suggesting the optimal sensor locations and number of installation points, it further increased the usability of such method.



(a) (b)

Figure 2-4 Optimal sensor locations using error-based method, (a) data processing for the error-based method, (b) optimal sensor locations according to nmber of sensors for summer season (Lee, Lee et al. 2019)

Referring to the two previous studies, the optimal sensor location was arbitrarily chosen, such as the on the basis of location that provides information that can represent the entire environments of the greenhouse. The applicability of these optimal sensor placements, however, may be affected when the target crops and the algorithm for controlling air conditioning system change. It is then necessary to select the optimal sensor location for monitoring the environment inside the greenhouse not merely by arbitrary decision but on the basis of reliable methods one of which could be clustering algorithms, information loss approach and paeto principle as proposed by (Yoganathan, Kondepudi et al. 2018) or the application of

CFD (Liu Y. et al. 2014; Guzman C. H. et al. 2018) and other proven approaches as presented in Table 2-2.

Table 2-2 Research on optimal sensor location of greenhouse and building

Author	Methods	Experimental target	Summary
Liu Y. et al. (2014)	CFD	Greenhouse	This study designed a CFD model to analyze the climate and thermal environment of greenhouses, simulating temperatures under mechanical ventilation conditions. The optimal sensor location proposed based on simulation results was determined to be in the center of the greenhouse.
Guzman C. H. et al. (2018)	CFD	Building	In this study, the optimal temperature sensor arrangement was proposed for a building with buoyancy-driven natural ventilation. For this, the air temperature distribution of the building was simulated using CFD simulation.
(Chen and Wen 2012)	CFD	Building	In this study, the sensor location was designed to minimize the detection time of pollutant concentration or exposure to pollutants in 12 test areas. CFD was used to simulate the internal airflow of the building for each scenario.
(Chen and Wen 2010)	CFD	Building	Based on airflow analysis using CFD for small office, large hall, and office suites, this paper proposed the optimal sensor location to minimize the diffusion and exposure time of pollutants.

(McGibney, Pusceddu et al. 2012)	CFD	Office	Using CFD, the thermal environment was analyzed considering all the heat sources in the office space. Finally, the minimum detection score for the temperature of the office space was calculated, and the suitability function for evaluating whether the tolerance limit was minimized.
Lee et al. (2019)	Error-based method Entropy-based method	Greenhouse	This study proposed error-based and entropy-based sensor locations using air temperatures measured at nine locations in eight-span greenhouses. Each method proposes a sensor location that can represent the air temperature of the entire greenhouse and an optimal sensor location for monitoring the areas most affected by the external wind environment.
(Yoganathan, Kondepudi et al. 2018)	Clustering algorithm Information loss approach Paeto principle	Building	This study showed the optimal sensor locations in office buildings using a clustering algorithm to select the optimal sensor locations. For 31 sensor data measured in an actual building. The results showed the optimal sensor position that could minimize data loss.
(Papadimitriou, Beck et al. 2000)	Entropy-based method	Building	This paper proposed the optimal sensor location for checking the structural safety of buildings. To this end, it proposed an entropy-based optimal sensor location to minimize instability based on Bayesian statistical methodology.
(Fontanini, Vaidya et al. 2016)	Perron-Frobenuous method	Building	This paper used a dynamic system approach for sensor placement in the air flow field. The optimal sensor location was proposed by calculating the contaminant tracking matrix using the discrete form of the Perron-Frobenius operator.

(Suryanarayana, Arroyo et al. 2021)	Data-driven model	Building	This research determined the optimal sensor placement for best control and monitoring of multi-zone buildings. Through a data-driven methodology, the research provided importance ranks for all sensor locations and suggested appropriate sensor placements for target buildings.
(Fu, Sha et al. 2014)	Data-driven model	Building	The control of the HVAC system of the building was optimized using sensor clustering and the data-driven simulation model. The environmental data such as air temperature were measured for a large space.

2.4. Application of machine learning models to agriculture

Machine learning was first coined by Samuel in 1959 as the research field that provides learning capabilities to computers without being explicitly programmed. The method of learning games by defining an evaluation function as the sum of the product of a weight and a feature defined as a pattern on a game board and changing the weight.

Early machine learning models failed to solve the XOR problem, so they failed to develop further and entered a period of stagnation. Since then, with the development of computer performance and algorithms was explored and is now being applied in numerous fields. Machine learning means the field of algorithm development that enables computers to infer and analyze phenomena by learning from data. Machine learning models could be divided into supervised learning models and unsupervised learning models. Supervised learning is further divided into classification and regression. Classification algorithm of machine learning is an algorithm that classify given data by class and it is used when predicting categorical variables. Regression is algorithm that predict variables by estimating the relationship between given data and it is used for predicting discrete variables. There are other regression models of machine learning to include linear regression, logistic regression, k-nearest neighbor regression, and decision tree.

In greenhouses, several studies using machine learning were conducted for predicting crop yield (Sengupta and Lee 2014, Ali, Cawkwell et al. 2016, Senthilnath, Dokania et al. 2016, Ramos, Prieto et al. 2017, Kulkarni, Mandal et al. 2018, Cai, Guan et al. 2019, Gümüşçü, Tenekeci et al. 2020), disease detection (Chung, Huang

et al. 2016, Amara, Bouaziz et al. 2017, Pantazi, Moshou et al. 2017, Pantazi, Tamouridou et al. 2017), crop quality analysis (Maione, Batista et al. 2016, Hu, Pan et al. 2017), and intelligent harvesting (Horng, Liu et al. 2019, Zhang, Karkee et al. 2020).

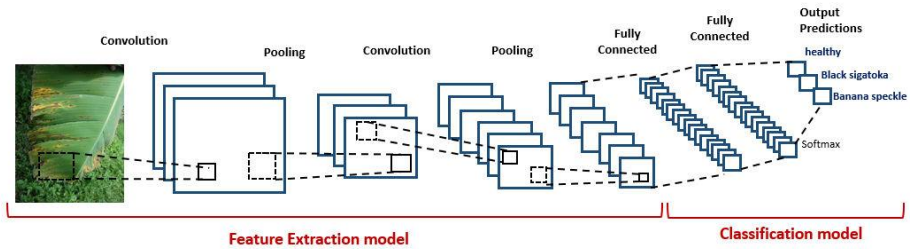


Figure 2-5 A graphical depiction of LeNet model (Amara, Bouaziz et al. 2017)

The CNN-based machine learning models and classification models were generally used as machine learning models for predicting crop yields. (최호길, 안희학 et al. 2019) collected farm bio-sensor data and developed a machine learning model that diagnoses diseases of crops grown on farms and predicted harvests for the year. A neural network model was developed, and accuracy was compared with random forest and gradient boosting tree. (Senthilnath, Dokania et al. 2016) predicted the maximum tomato yield using high-resolution RGB images obtained from unmanned aerial vehicles. For image data, k-means, expectation maximization, and self-organizing map algorithms were used to distinguish tomatoes and non-tomatoes. (Cai, Guan et al. 2019) used LASSO, support vector machine, random forest and neural network algorithms to predict wheat yield across Australia. Wheat yield was predicted using satellite data, and the performance of wheat yield prediction was improved during the learning phase of the data by integrating climate

data. (Gümüşçü, Tenekeci et al. 2020) estimated the seeding date of agricultural products using kNN, SVM, and decision tree algorithms. The machine learning model was trained using climate data for the past 300 days, and the kNN model showed the highest prediction accuracy.

Several studies have also been conducted to develop the machine learning models for detecting diseases in crops. Diseases were detected by learning the leaf images of crops. (우현준, 조주연 et al. 2017) used CNN to determine whether a crop was infected with a disease. A model for classifying disease and non-disease and disease was provided using the disease sign model. (Amara, Bouaziz et al. 2017) came up with machine learning model to classify diseases by learning the leaf images of bananas in advance. The machine learning model proposed in this study was highly accurate but at challenging conditions such as illumination, complex background, and different resolution conditions. Red, green, and blue color represent information of healthy and infected leaves. Red, green, and blue color information was extracted and preprocessed to resize pixels and convert to grayscale. (Horng, Liu et al. 2019) proposed a harvesting system by developing an algorithm for a robot arm using internet of things technology and smart image recognition. The robot arm was trained to recognize crops using a neural network model, and the accuracy was 84%. Machine learning models have been applied in various agricultural fields. Especially, many studies on the cognitive and classification models have been performed by learning the image data. Despite numerous studies about machine learning, only few studies are available relative to monitoring or predicting the internal environment such as air flow, temperature, and humidity related to the growing environment that has a great influence on the crop production in the greenhouse.

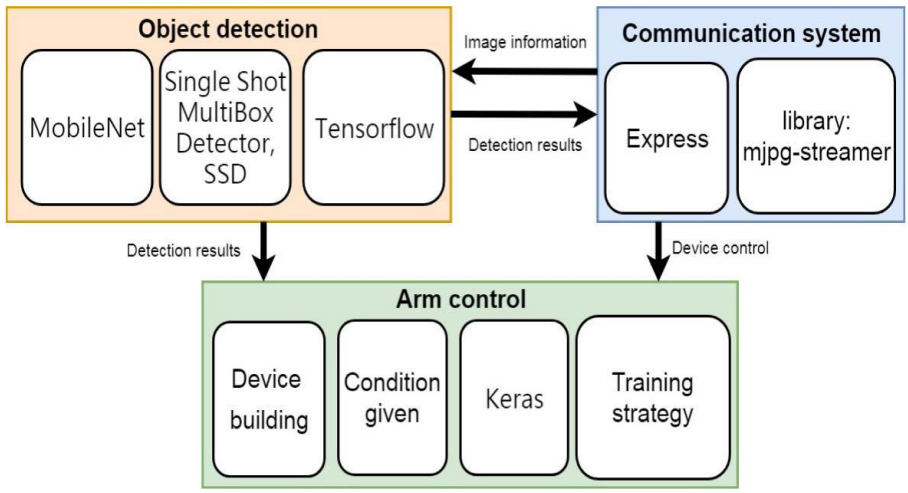


Figure 2-6 Model of designed remote harvesting system (Horng, Liu et al. 2019)

From the results of previous studies, various machine learning models predicted several variables for agricultural facilities and crops. Especially, the regression models of machine learning were expected to be applied for estimating internal environments such as air temperature and humidity because it is possible to predict the environmental variables in the future or in other locations within the agricultural facility under the specific conditions by learning data.

Table 2-3 Previous research on greenhouse industry using machine learning algorithms.

Author	Machine learning algorithm	Aim
(문태원, 박준영 et al. 2020)	CNN	In this study, the weight of paprika was estimated by learning the paprika image at a specific time in the raw data state using a convolutional neural network.
(홍성은, 박태주 et al. 2020)	Multi Linear Regression Random Forest ConvLSTM	Measurement date, specimen number, growth, flower cluster height, stem diameter, leaf length, leaf width, number of leaves, flowering group, fruit-bearing group, number of fruits, harvest group growth-related information was collected, and the amount of tomato growth and production was collected.
(오정원, 김행곤 et al. 2019)	LinearRegression ML	Machine learning was applied to the harvest season to ensure that the fruit was harvested at the highest quality and shipped at a higher price. The harvest time is predicted by the hour, and FPSML is used to make it possible to check on PCs and smartphones.
(최강인, 노혜민 et al. 2019)	DNN	Hyperspectral images used for analytical studies of plant growth states were applied with DNN to resolve the limit points affected by factors such as physical factors and data complexity. The leaves were classified into 4 types, including normal leaves and leaves damaged by pests, and the background.
(김근식, 구종회 et al. 2020)	YOLO	This study designed a system that could monitor mushroom growth using deep learning and sensor networks. The system automatically measured the size and growth rate of mushrooms.

(나명환, 조완현 et al. 2020)	CNN	Tomatoes' pests were detected using CNN and developed as a web application platform.
(최호길, 안희학 et al. 2019)	CNN, ANN	The crop yield was predicted using the current weather and soil microbial content.
(김예슬, 곽근호 et al. 2018)	SVM, 2D-CNN, 3D-CNN	This research compared the hyperparameters of the machine learning model and the accuracy of the model according to the size of the training data. The machine learning model was a model for classifying crops.
(우현준, 조주연 et al. 2017)	CNN	For the diagnosis of pests of crops, a CNN model using image processing was used. Information was acquired using an app or web in a smartphone environment.
(Fuentes, Yoon et al. 2017)	Faster R-CNN, SSD, R-FCN	This study uses technology-based data annotation and augmentation methods to provide better performance of deep learning-based detectors for real-time tomato disease and pest recognition.
(Rustia, Chao et al. 2021)	CNN	This study used insect-adhesive paper traps and wireless imaging devices to construct a greenhouse data set for detecting and recognizing pests in a stationary environment.

(Zhang, Xu et al. 2020)	CNN	This study achieves accurate estimation of growth-related traits for greenhouse lettuce. CNNs are used to model the relationship between RGB images of greenhouse lettuce and their growth-related characteristics (LFW, LDW, and LA).
(Durmuş, Güneş et al. 2017)	CNN	This research uses Raspberry Pi deep learning to achieve disease detection and automatic irrigation, and uses NB-IoT technology to connect the two information to Alibaba IoT cloud for real-time monitoring, automation and intelligent realize the transformation.
(Cao, Cui et al. 2020)	AlexNet, SqueezeNet	Deep learning was used to detect leaf diseases in tomato plants. Two different deep learning network architectures, AlexNet and SqueezeNet, have been trained and tested on tomato images from the Plant Village dataset.
(Afonso, Barth et al. 2019)	MaskRCNN	Development of deep learning model that distinguishes sweet pepper fruits from other parts
(López-Aguilar, Benavides-Mendoza et al. 2020)	ANN	This study used ANN to simulate the accumulated aerial dry matter (leaf, stem and fruit) and fresh fruit yield of a tomato crop
(Alhnaity, Pearson et al. 2019)	SVR, Random forest, LSTM	Variable parameter such as CO_2, humidity, radiation, outside temperature, inside temperature, actual yield, stem diameter variation measurements were used to predict ficus stem diameter or tomato yield problems

(Bakay and Ağbulut 2021)	Deep learning, SVM, ANN	This study forecasted GHG emissions (CO_2, CH_4, N_2O, F-gases and total GHG) by machine learning models which was trained on electricity production shares of energy resources by years (year, coal, liquid fuels, natural gas, renewable energy and wastes, and total)
(Flores, Zhang et al. 2021)	SVM, Random forest	Corn and soybean seedlying images in greenhouse were collected for successive five days after germination and dataset was generated after image segmentation and noise removal
(Guo, Juan et al. 2017)	Random forest, SVM, Neural network	A machine learning model was developed to determine the moisture state of the plant root zone in the greenhouse. The prediction accuracy and calculation time of each model were compared.
(Hamrani, Akbarzadeh et al. 2020)	SVM, LASSO, FNN, CNN, LSTM, etc.	Nine ML models were compared to predict soil greenhouse gas (GHG) emissions in the agricultural sector. The accuracy and calculation time of all models were compared, and the advantages and disadvantages of each model were analyzed.
(Lee, Gottschlich et al. 2018)	LSTM	describe a novel time-series anomaly detection system called Greenhouse. Our key goal in Greenhouse is to combine state-of-the-art machine learning and data management techniques for efficient and accurate prediction of anomalous patterns over high volumes of time-series data.
(Ge, Zhao et al. 2022)	Linear regression, SVR, K neighbor regression, Random forest, etc.	Machine learning models were developed for the evapotranspiration of crops by using meteorological factors such as net solar radiation, daily average temperature, daily minimum temperature, and daily maximum temperature.

Chapter 3. Optimal sensor location for predicting air temperature inside greenhouse based on machine learning models: Sensing

3.1. Introduction

Since the internal environment of the facility in protected agriculture can be arbitrarily adjusted, it can provide stable production all year round. Hence, its importance was amplified and increased due to climate change and an increase in natural disasters. The greenhouse horticulture area in South Korea has increased from 700 ha in 1970 to 52,444 ha in 2020 (MAFRA 2022). Large-sized greenhouses have advantages in automating production facilities and increasing productivity, but failure to maintain a uniform and optimum growing environment inside leads to low productivity. The development of ICT technology has introduced systems to precisely monitor the internal environment of greenhouses and control the appropriate growth environment of crops. However, for large-sized greenhouses, it is necessary to install a large number of sensors to monitor the environment inside the greenhouse for maintaining uniformly optimum growing environment. As a result, it is true that it acts as a heavy burden on the part of farmers due to the increased initial and maintenance costs relative to the installation cost of sensors.

Therefore, it is important to install a minimum sensor in a place that represents the entire greenhouse environment. In general, sensor location is determined by the subjective experience of the greenhouse designer. Typically, many studies presented

center of facilities as monitoring locations for environmental factors such as air temperature and humidity (Kittas, Karamanis et al. 2005, Feng, Li et al. 2013, Park, Lee et al. 2022). However, the internal environment of a greenhouse is generally not uniform, and environmental factors at the center of the greenhouse may not be representative of the entire greenhouse. Consequently, optimal sensor installation for minimal sensor installation and proper monitoring is extremely important in terms of increasing farmers' profits.

Previous studies for optimal sensor location selection have focused on the security of wireless sensor communication networks or the lifetime of wireless sensors (Wang, Basagni et al. 2005, Toumpis and Tassiulas 2006, Guerriero, Violi et al. 2011, Keskin, Altınel et al. 2014) and on monitoring the structural stability of buildings (Hwang, Kim et al. 1996, Lee, Kim et al. 2009, Lee, Kim et al. 2016, Yi, Zhou et al. 2017). Research using theoretical formulas (Papadimitriou, Beck et al. 2000, Fontanini, Vaidya et al. 2016, Yoganathan, Kondepudi et al. 2018, Lee, Lee et al. 2019, Suryanarayana, Arroyo et al. 2021) and CFD (Hamel, Chwastek et al. 2006, Zhang and Chen 2007, Mazumdar and Chen 2008, Liu and Zhai 2009, Chen and Wen 2010, Feng, Li et al. 2013, Huang, Zhou et al. 2014, Liu, Chen et al. 2014, Waeytens, Durand et al. 2019, Wijaya, Utami et al. 2021, Chen and Gorlé 2022) was conducted to select the optimum sensor position for internal environmental monitoring of the facility.

Studies on selecting the optimal sensor location using theoretical formulas and algorithms suggested a location representing the average value or minimizing the error value as the optimal sensor location using the field measured data. (Yoganathan, Kondepudi et al. 2018) proposed the optimal sensor location of the building using the clustering algorithm, information loss approach, and paeto principal. The optimal

sensor location was designed to minimize the loss of sensing data, and as a result, 31 sensors were optimized with 6 sensors. In (Lee, Lee et al. 2019)'s research, two method : error-based and entropy-based sensor location were introduced to select the optimal sensor location for monitoring the internal air temperature of a greenhouse. Both methods were quantitative methods that suggested optimal sensor locations based on average or highly variable regions of greenhouse air temperature. However, these studies also applied researchers' subjective judgments and criteria of optimal sensor location, which may not apply well when new conditions for greenhouse control were presented.

Studies related to optimal sensor positioning using CFD have also been conducted. In (Chen and Wen 2010)'s study, three test environments such as small office, large hall and office suite were simulated for sensor system design to detect pollutant dispersion using multi-zonal, zonal, and CFD models. After investigating a variety of models, CFD was proposed as the most accurate model. A CFD that considers the factors affecting greenhouse microclimate such as wet curtain, inner fan was also used to simulate air temperature and flow distributions in greenhouses (Feng, Li et al. 2013). As a result, the optimum sensor location was presented as the point where there was no significant change in air temperature and flow. On the other hand, (Liu, Chen et al. 2014) designed a CFD model that could analyze the climatic and thermal environments of a greenhouse, and simulated the temperature under mechanical ventilation conditions. Based on the simulation results, proposed optimal sensor location was determined to be at the center of the greenhouse. CFD-based optimal sensor location selection research has the advantage of being able to visualize and quantitatively analyze invisible air temperature, humidity, and flow distributions because it uses a simulation model. It has the limitation of requiring a lot of

computing time due to additional environmental conditions such as changing ventilation conditions, and pollutant generation.

However, most previous studies selected one or several sensors as representative sensors for the entire facility in terms of optimal sensor location selection. Also, in the selection of the optimal sensor location, the place representing the average value or the place with the smallest or greatest environmental change was selected as the representative point according to the researcher's judgment without quantitative criteria. In this case, for spaces with relatively large environmental changes, single or several sensors have a limit in accurately monitoring the internally entire space. Especially in naturally ventilated greenhouses, there are internally regional differences on local ventilation rate due to various external wind speeds and directions during natural ventilation resulting in that the air temperature distribution within the greenhouse may vary widely. As a result, there is concern about performance degradation with respect to the optimum sensor position.

Meanwhile, with the recent development of artificial intelligence and machine learning (ML) models, various applications are being made in the agricultural field (Zou, Yao et al. 2017, Hongkang, Li et al. 2018, Molano-Jimenez, Orjuela-Cañón et al. 2018, Gorczyca 2019, López-Aguilar, Benavides-Mendoza et al. 2020). In particular, various studies have been conducted for monitoring and predicting the internal environment, which can be a solution for controlling the internal environment of a large greenhouse. ML models generally have the advantage of not only showing high prediction accuracy, but also being able to learn from data alone without human subjective intervention. In addition, since it is easy to continuously improve the model by adding learning data, it has the advantage of being easy to learn and respond to new environmental conditions (Dahiya, Gupta et al. 2022).

Since the machine learning model is a black box model, it is not necessary to select the optimal sensor location based on a specific standard, and it is possible to directly estimate the optimal sensor location through the estimation of environmental data at a specific location.

The purpose of this study was to develop the Prediction Current Temperature using Optimal sensor Machine Learning model (PCTO-ML) for selecting the optimal sensor location that provided learning data for ML models with the highest accuracy. The PCTO-ML predicted the air temperature at different points inside a naturally ventilated multi-span greenhouse. For this purpose, artificial neural network (ANN), long short-term memory (LSTM), and support vector regression (SVR) ML models were developed and compared its accuracy of predicting point-wise air temperature in the greenhouse. Then, the environmental data measured at each sensor location in the greenhouse was used to learn the selected ML model that showed the highest accuracy and air temperatures at other points in the greenhouse were predicted. As a result, the sensor location that provided learning data for the ML model with the highest accuracy was evaluated as the optimal sensor location. This means that the point-by-point air temperature in the greenhouse could be estimated by installing environmental sensors at a minimum number of locations. However, many learning features of ML models mean installing a large number of sensors such as air temperature and humidity, CO_2, etc. Therefore, in this study, a simplified ML model which was defined as PCTO-ML was developed to reduce kind of learning features and evaluate the optimal sensor location.

3.2. Materials and methods

The study flowchart is illustrated in Figure 3-1. A machine learning (ML) model was developed to enable optimal sensor location for the air temperature prediction at nine points inside naturally ventilated greenhouse. First, the data on greenhouse indoor and outdoor conditions were collected for the modelling. The data for environmental factors including air temperature and relative humidity were collected from nine sites inside the naturally ventilated greenhouse for the indoor conditions. To reflect the changes in indoor conditions caused by the outdoor conditions, the data on factors including air temperature, relative humidity, and wind direction and speed were collected outside the greenhouse. The collected data were pre-processed to ensure a high level of performance by the developed ML model. Linear interpolation was performed on the missing values during the data collection period and outliers in the collected data, and normalization was performed on each feature to prevent potential bias of a specific feature in ML model training. In addition, a correlation analysis was performed on the collected data for the prediction of experimental greenhouse indoor air temperature, and a training set prepared using features exhibiting a high correlation. Three ML models which were ANN, LSTM and SVR were constructed, and the one with the most outstanding performance of predicting the greenhouse indoor air temperature was selected. The final selected ML model used training data measured at each point in the naturally ventilated greenhouse to predict air temperatures at other points. As a result, the sensor location that provided the learning data of the ML model that showed the best prediction performance was evaluated as the optimal sensor location. Nevertheless, the final selected ML model required a large number of sensors because it required many

kinds of learning features. Subsequently, to minimize the kind of sensor installations, three simple ML models were suggested by gradually excluding the features of the final selected ML model which was PCTO-ML.

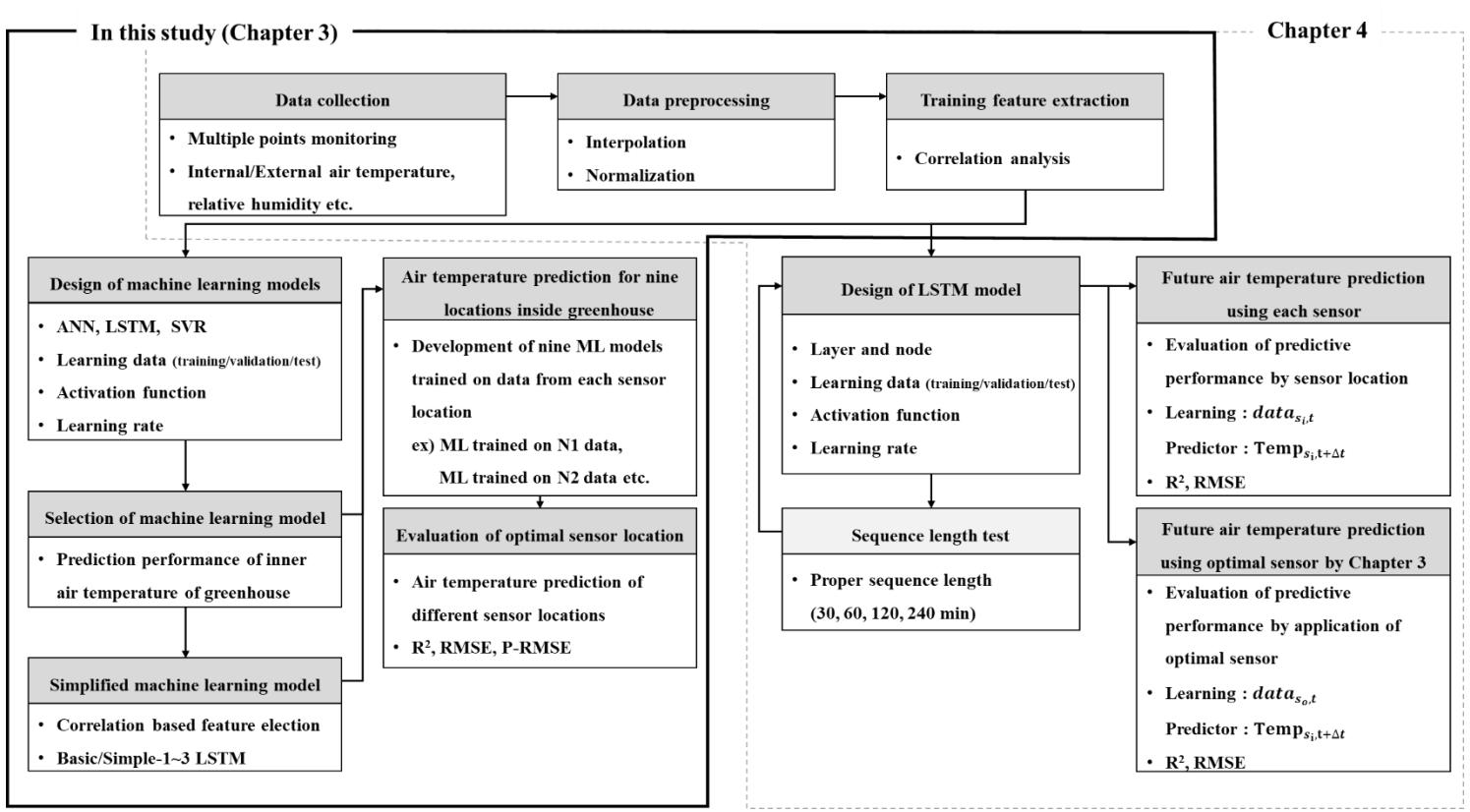


Figure 3-1 Flowchart of developing a machine learning model for selecting the optimal sensor location for experimental greenhouse

3.2.1. Experimental greenhouse

The target greenhouse was a 1-2W type, 8-span plastic film greenhouse, located approximately 1 km to the south of the Boryeong Thermal Power Plant in Gojeong-ri, Jugyo-myeon, Boryeong-si, Chungcheongnam-do, South Korea (126°29'E, 36°23'N). The greenhouse structure was 4.0 m in width and 30.0 m in length with eaves and ridges at a height of 4.5 m and 5.7 m, respectively. The interior of the greenhouse was divided into work and cultivation spaces (Figure 3-2). The workspace, which was at a 4.0 m distance from the entrance, is a place for simple tasks inside the greenhouse and the control of indoor mechanical systems. The cultivation space is 768 m^2 in floor area with a single cover of agricultural polyolefin film of 0.15 mm thickness on all sides from the floor to the walls and the front and back sides (Figure 3-2). One hundred trees of Irwin mango were being cultivated inside the greenhouse (Figure 3-3(a)). The Irwin mango trees were cultivated in pots, each of which had a diameter of 0.8 m, and there were 13 lines of pots in total. Irwin mango trees in the experimental greenhouse have a canopy of 0.4 to 0.9 m.

Natural ventilation was performed in the target greenhouse through the side and roof vents, and a heat pump (ADF-SLX12WHB, A-San Inc., Korea) was used to run the heating and cooling system. The side vents of the greenhouse were located at a height of 0.5 ~ 2.1 m and 2.7 ~ 3.5 m from the floor. The roof vents were located vertically on the uppermost part of the greenhouse roof, each approximately 0.6 m in width, to be opened and closed using a cover. The capacity of the heat pump inside the greenhouse was 43,276W for cooling and 36,786W for heating, and a duct of approximately 60 cm diameter was used to ensure uniform heating and cooling inside the greenhouse.

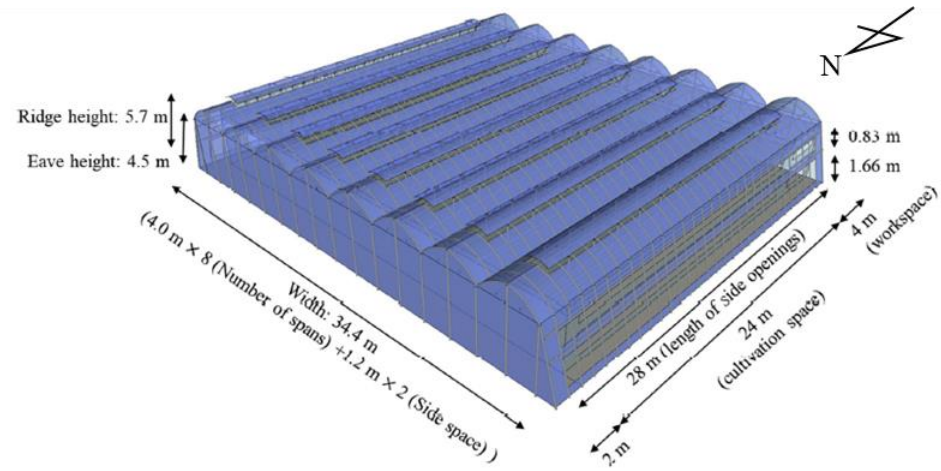


Figure 3-2 Schematic diagram of the target greenhouse

(a) Irwin mangoes

(b) Roof vents

(c) Side vents

(d) Workspace

(e) Heat pump

(f) Air duct

Figure 3-3 Crop and components of experimental greenhouse located on the west coast of South Korean in Jugyo-myeon, Boryeong-si, Chungcheongnam-do Province (126˚29'E, 36˚23'N). (Irwin mangoes, windows, heat pump, air duct and workspace)

3.2.2. Machine learning models

3.2.1.1. Artificial neural network (ANN)

The ANN is a model created to resemble the human brain. The basic principle of operation is to connect several neurons in a mutually complementary manner, with an optimal output produced and predicted for an input. As with neurons, each node is divided into an input, a hidden, and an output layer (Figure 3-4(a)). When a value is supplied to the input layer at the anterior most part among the circular nodes, that value gets transferred to the hidden and output layers. The ANN receives an input, modifies the internal state according to that input, and produces an output according to that input and the level of activation. As a non-linear structure of statistical data comprising the modelling tools, the ANN is characterized by its applicability in simulating the complex relationship between the input and output that cannot be expressed via other functions. The output of each node is referred to as the activation or node value, and the system output is produced at the final step of computation. ANN has been studied in various fields due to its ability to learn all nonlinear functions (Wanjawa and Muchemi 2014, Al-Shawwa, Al-Absi et al. 2018, Cabaneros, Calautit et al. 2019). As shown in eq.3-1, ANN calculates the output(y) by multiplying the input layer $\mathrm{X}(x_1, x_2, \cdots, x_n)$ and the weight of each layer $\mathrm{W}(w_1, w_2, \cdots, w_n)$ and adding constant b.

$$\mathrm{y} = \sum_{i=1}^{n} x_i w_i + b \qquad \text{(Eq. 3-1)}$$

3.2.1.2 Support vector regression (SVR)

The support vector machine was proposed by Vapnik in 1979 as a statistical learning theory whereby the entities of two different categories are classified through the detection of the hyperplane (Figure 3-4(b)). In contrast to conventional methods based on the principle of empirical risk minimization, the SVR is based on structural risk minimization to reduce the upper bounds of the generalization error to the lowest level and thus exhibit an outstanding performance (Smola and Schölkopf 2004). The SVR can be applied in regression prediction models through the use of a training set. An SVR model enables prediction by converting a non-linear regression problem than cannot be solved at the input data using the Kernel function to a linear regression problem through mapping (Dibike, Velickov et al. 2001). SVR with radial basis kernel functions (SVR_{rbf}) uses following equation :

$$SVR_{rbf} = \exp(-y\{x_i - x_j\}^2) \quad \text{(Eq. 3-2)}$$

Where y is a constant value according to the radial basis function. The advantage of SVR models, therefore, lies in preventing overloading by controlling the error range through the use of the loss function to boost up the generalization performance (Vapnik 1995). SVR has been actively used in remote sensing, hydrology, agriculture, etc., due to its rapid and accurate prediction performance (Chevalier, Hoogenboom et al. 2011, Mountrakis, Im et al. 2011, Deka 2014, Ichii, Ueyama et al. 2017).

3.2.1.3 Long short-term memory (LSTM)

Long short-term memory is a recurrent neural network (RNN) structure proposed by Hochreiter and Schmidhuber in 1997. It is an RNN model capable of learning the

long-term dependency of data, which was proposed to overcome the vanishing gradient problem of conventional RNN models. As shown in figure 3-4(c), an LSTM model has an input, a forget, and an output gate for maintaining the cell state. The forget gate functions to determine which data to exclude from the input data and the output of the previous cell, whereas the input gate determines which among the newly supplied data to store or memorize in the cell state, and the output gate determines which between the input and the memory to transfer to the output. Although the LSTM uses the same methods as the RNN, whereby the final output is computed through hidden variables, the appropriate use of the gates in the process of dealing with the hidden variables to control the flow of data, is what makes the LSTM applicable to the learning and prediction of sequential or time-series data such as in text generation, voice recognition, and text translation. The principle of an LST M layer can be described as following equations:

$$f_t = \sigma_t\left(W_f x_t + U_f h_{t-1} + b_f\right) \quad \text{(Eq. 3-3)}$$

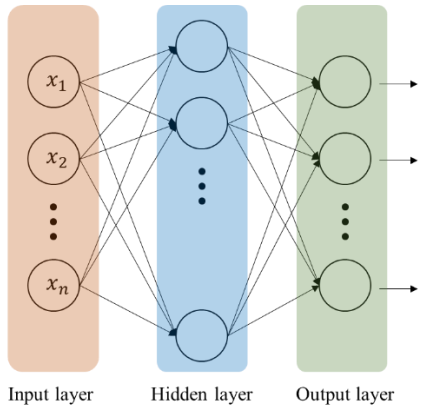
$$i_t = \sigma_i(W_i x_t + U_i h_{t-1} + b_i) \quad \text{(Eq. 3-4)}$$

$$o_t = \sigma_o(W_o x_t + U_o h_{t-1} + b_o) \quad \text{(Eq. 3-5)}$$

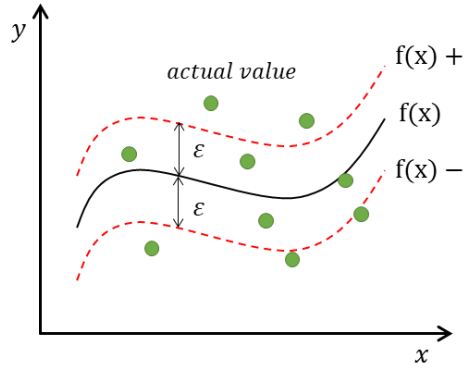
$$C_t = f_t \odot C_{t-1} + i_t \odot tanh(W_c x_t + U_c h_{t-1} + b_c) \quad \text{(Eq. 3-6)}$$

$$h_t = o_t \odot tanh(C_t) \quad \text{(Eq. 3-7)}$$

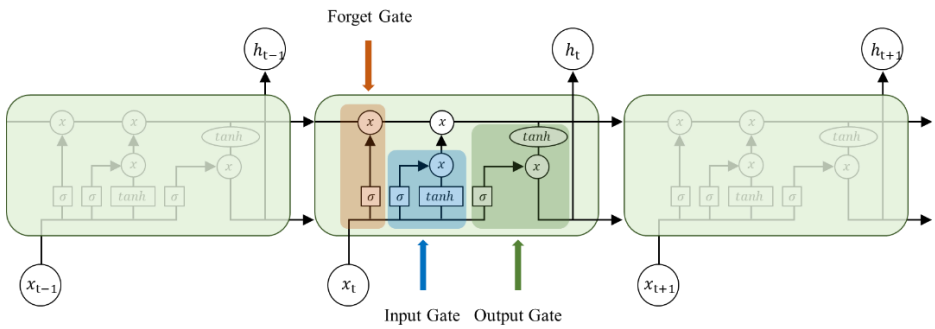
The model has a concept of cell status C_t in accordance of time t, where f_t, i_t, o_t are forget gate, input gate and output gate at time t respectively. σ_t, σ_i, σ_o are sigmoid function of each gate, $\odot$ indicates the Hadamard product, h_t is the hidden layers and W, U and b are the model parameters that is learned each step (Lippi, Montemurro et al. 2019, Li, Ding et al. 2021, Lee, Lee et al. 2022).



(a) ANN



(b) SVR



(c) LSTM

Figure 3-4 Conceptual diagram of machine learning models (ANN, SVR, LSTM)

3.2.3. Experimental procedure

3.2.3.1. Data acquisition for modelling machine learning models

In this study, a ML model to estimate the indoor air temperature at each site inside the naturally ventilated greenhouse was developed towards optimal sensor location for the greenhouse indoor air temperature monitoring. Regarding thermodynamics of the thermal conditions of the experimental greenhouse, factors such as the greenhouse indoor and outdoor air temperatures and the heat exchange owing to crop, soil, and solar radiation are critical (Vanthoor, Stanghellini et al. 2011, Choab, Allouhi et al. 2019, Cai, Wei et al. 2022). Moreover, as natural ventilation has a direct effect on the micro-climate inside the greenhouse (Villagrán, Gil et al. 2012), the outdoor wind direction and speed should also be considered in the estimation of the greenhouse indoor conditions. Therefore, sensors were installed at nine sites inside the greenhouse as shown in figure 3-5, to collect the data on greenhouse indoor conditions. Each sensor was installed at a height of 0.9 m from the floor, which is the height of the crop zone. The measurements taken by each sensor included the air temperature, relative humidity, soil temperature, soil humidity, soil electrical conductivity (EC), CO_2, atmospheric pressure, and ultraviolet (UV radiation) and illuminances. A temporary weather station was also installed outside the greenhouse to measure outdoor air temperature, relative humidity, UV radiation, wind direction, wind speed, and atmospheric pressure. The data were collected in July 2017 for the summer season with the high temperature, all vents of the greenhouse were opened. Each measurement was taken at 1-s intervals, and the mean of 10 min data was recorded.

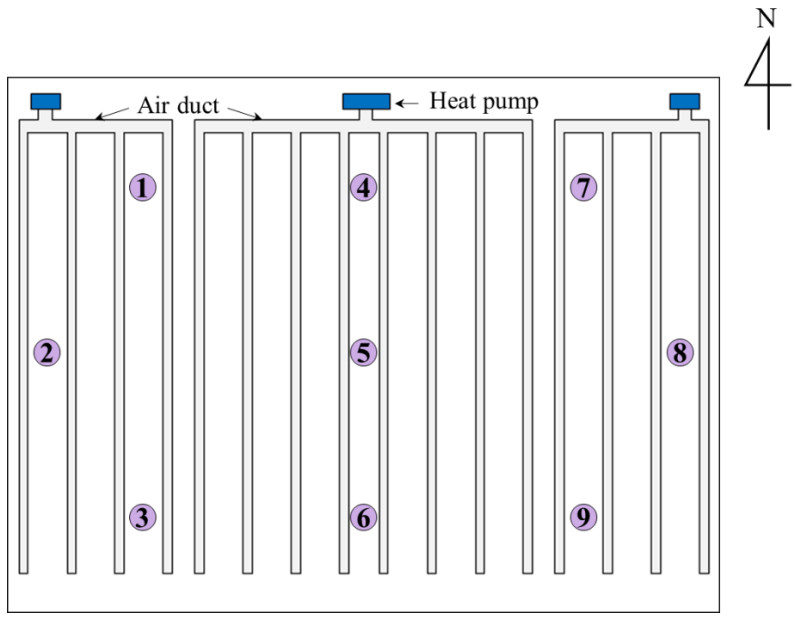


Figure 3-5 Measuring points to acquire learning dataset of machine learning models

3.2.3.2. Pre−processing of the dataset

In general, data pre-processing is critical to ensure a high level of performance of an ML model (Setiawan, Djanali et al. 2019, Singh and Singh 2020). Based on the quality of training data, the performance and prediction accuracy of the model can vary significantly. In the presence of multiple outliers and missing values in the training set or coexistence of textual and numerical data, a moderate level of performance cannot be expected of the ML model. Moreover, data normalization is an essential step in data pre-processing, which includes feature scaling within a shared range to prevent the dominance of a larger number feature over a smaller one. When the range of a specific training feature is relatively large compared to that of other training features, the contribution of the feature regarding the ML model may be biased. Among the features used in the training set in this study, the range of air temperature was approximately 24 ~ 40 °C, whereas radiation ranged between 0 and

several tens of thousands in $\mathrm{w} \cdot \mathrm{m}^{-2}$. Hence, as shown in Eq. 3-8, the Min Max normalization scale was used to set the range of [0, 1] for all measured environmental factors, to ensure equal contribution of each feature in the data through the process of learning.

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad \text{(Eq. 3-8)}$$

where, x' is the normalized sample data, x is each individual sample data, x_{min} is the minimum value of sample data, and x_{max} is the maximum value of sample data.

As the Min Max normalization is based on the minimum and maximum values, there is a drawback in which the outliers in the data are not adequately treated. In addition, missing values and outliers were intermittently observed in the data of each sensor installed at the target greenhouse. Hence, to resolve this, linear interpolation was performed on the outliers and missing values.

The selection of appropriate learning data in the design of a machine learning model can increase the accuracy of the model. Furthermore, eliminating learning data that can reduce the accuracy of the model not only improves the accuracy of the model, but also minimizes the kind of sensors for collecting the learning data. Therefore, in this dissertation, correlation analysis was performed between the data processed with normalization and linear interpolation and the internal air temperature which would be predicted by the machine learning model. The correlation analysis is a method of analyzing the relationship between two variables. Two variables can be related or correlated independently of each other, where the strength of the relationship between the two variables is defined as correlation. In

correlation analysis, ρ is used as a population correlation coefficient and r is used as a sample correlation coefficient as a unit representing the degree of correlation. In this study, correlation analysis of learning data with predictor variable was performed to select learning data and improve the accuracy of the prediction model, and Pearson's correlation coefficient was used for the correlation analysis of learning data. The Pearson correlation coefficient is commonly used in bivariate correlation analysis to find associations between variables. The value of the correlation coefficient represents +1 if X and Y are exactly equal, 0 if they are completely different, and -1 if they are exactly equal in opposite directions. Therefore, in this study, a correlation analysis was performed using the Pearson correlation coefficient as an index for air temperature, humidity, soil temperature, soil humidity, EC, CO_2, atmospheric pressure, UV radiation, and illuminance at each point in the naturally ventilated greenhouse. The performance of machine learning models was improved by using variables that show high positive or negative correlations in machine learning models.

3.2.3.3. Design of machine learning models

Towards optimal sensor location inside the naturally ventilated greenhouse, an ML model was developed in this study to estimate the greenhouse indoor air temperature. For this, the data of each sensor inside the greenhouse were used as the training set in developing an ML model to predict the air temperature at all other sensors. To illustrate, the ML model trained on the data of Sensor #1 was used to predict the temperatures at Sensors #2 to #9. To ensure that the developed ML model exhibits a high level of performance, the ANN, LSTM, and SVR that are widely applied in the

prediction of building indoor conditions were used in developing the ML model in this study (Singh and Tiwari 2017, Glad 2020, Jung, Kim et al. 2020, Fan, Ji et al. 2021).

Optimization of hyper-parameters used in the algorithm is extremely important in enhancing the level of ML model performance. There is no set method of hyper-parameter optimization in general, and repeated designing and trial and error are necessary to define the optimal hyper-parameters. Although there is no absolute best hyper-parameter value, suitable values can be identified for each type of data and model (Reimers and Gurevych 2017, Raschka 2018). Thus, to ensure efficient and accurate model computations, in this study, the hyper-parameters were defined for each ML model through repeated designing and trial and error.

The main hyper-parameters constituting an ANN model include the Loss Function to evaluate the model performance and the Optimizer, a function to reduce the Loss Function. In this study, the mean squared error (MSE) for the Loss Function was used in evaluating the model performance, and Adam was used for the Optimizer for efficient training. In addition, the tanh function was used as the model activation function, and the batch size and epoch were set to 64 and 200, respectively, for adequate computations. The main hyper-parameters constituting an LSTM model include the Sequence Length as the length of the data to be used in a single training, the number of hidden units in the cell, Loss Function, and Optimizer. In this study, the Sequence Length was set to 1, and the hidden units were set to 32 and 16 in the first and second layers, respectively. For the Loss Function, Optimizer, activation function, batch size, and epoch, the same values as those in the ANN model were used. The main hyper-parameters constituting an SVR model include the kernel function, degree, gamma, and C-value. The kernel function allows the linear

classification of non-linear data as it is a function where the data of low-dimension are mapped to the high-dimension. The degree determines the order of the polynomial kernel, whereas gamma determines the flexibility of the decision boundary, and the C-value indicates the acceptable limit of errors. If gamma or C-value are excessively high, overfitting may result. In this study, the poly function was used as the kernel function, and the degree was set to 2, gamma to 1, and C-value to 10,000. For the data used in designing the ML model, a training set of 70% and test set of 10% were applied, and the remaining 20% was the validation set during the training to increase the reliability of the developed model.

3.2.3.4. Evaluation of optimal sensor location

The PCTO-ML model developed in this study is applicable to the management of greenhouse indoor conditions through the prediction of temperatures at different sites inside the greenhouse. Here, an adequate control of temperature for crop growth could be the basis of greenhouse management. It is thus important to reliably predict the temperature at each sensor site inside the greenhouse, especially for the monitoring of the peak value of high or low temperature, so as to control the heating, ventilation, air conditioning (HVAC) system to ensure adequate temperature for crop growth. Thus, in this study, the R^2 for the trend of indoor air temperature prediction and RMSE to test the errors in prediction values were used as the indicators of ML model evaluation. In addition, the prediction performance for the peak values representing the maximum and minimum values was evaluated using the peak-weighted RMSE (P-RMSE) that assigns weights on simulated peak values. The P-RMSE is an objective function developed by the United States Army Corps of

Engineers, Hydrologic Engineering Centre (USACE-HEC) for use in the HEC-1, whereby a simulated value greater than the mean of observed values is given a weight >1 and a simulated value lower than the mean of observed values is given a weight <1.

$$\mathrm{R}^2 = \left(\frac{\sum_{i=1}^{n}(O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^{n}(O_i - \bar{O})^2}\sqrt{\sum_{i=1}^{n}(P_i - \bar{P})^2}} \right)^2 \quad \text{(Eq. 3-9)}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}[O_i - P_i]^2} \quad \text{(Eq. 3-10)}$$

$$\mathrm{P-RMSE} = \left[\frac{1}{n}\left\{ \sum_{i=1}^{n}(O_i - P_i)^2 \left(\frac{O_i + \bar{O}}{2\bar{O}} \right) \right\} \right]^{\frac{1}{2}} \quad \text{(Eq. 3-11)}$$

where, O and P are the observed and simulated values, respectively, and $\bar{O}$ and $\bar{P}$ indicate the mean of observed and simulated values, respectively.

In this study, the ML models to predict the naturally ventilated greenhouse indoor air temperature at each defined site were developed for optimal sensor location. The ML models learned the previously selected learning features among the environmental data measured by each sensor. First, ANN, LSTM, SVR models were developed and the model with the highest prediction accuracy was selected. To select the optimal model, the aforementioned three kind of statistical indicators were used to evaluate the air temperature inside the greenhouse to determine the level of prediction. Developed ML models predicted the air temperature at nine locations inside greenhouse per sensor used as learning data. As a result, nine statistical index values were calculated per sensor for the evaluation of the optimal sensor location. Therefore, the mean values were estimated and compared with the result of

predicting the air temperatures at other sensor sites based on the learning of data at a given sensor site. The sensor site allowing the most accurate prediction of the air temperature at each of the remaining sites was determined using the final selected ML model.

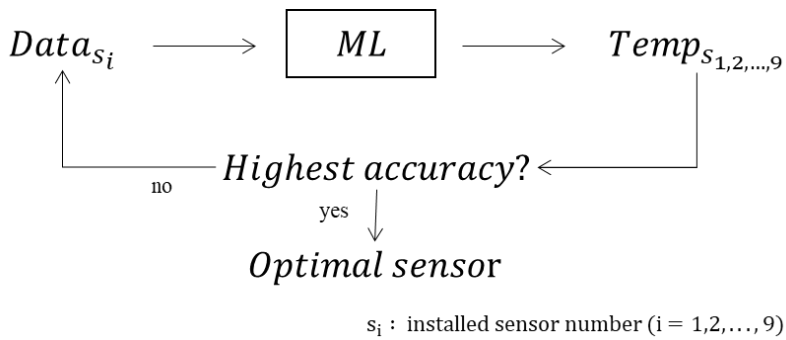


Figure 3-6 Brief description of optimal sensor location selection for air temperature prediction inside naturally ventilated greenhouse using machine learning model (Data means selected learning features such as air temperature, relative humidity and UV radiation)

3.2.3.5. Design of simplified machine learning models for optimal sensor location

When there are many learning features of the machine learning model, the type of sensor required for data collection increases. Therefore, in this study, correlation analysis was performed and a simplified model was developed to reduce the number of sensors installed.

A correlation analysis was performed on the data collected for naturally ventilated greenhouse indoor and outdoor conditions, and the variables with a high correlation for the estimation of greenhouse indoor air temperature were used in the training set. As a result, the training set included the air temperature, relative humidity, and UV

radiation and illuminances at each sensor site inside the greenhouse as well as the air temperature, relative humidity, and wind direction and speed outside the greenhouse. However, applying all these variables in an actual farm would necessitate the installation of too many sensors, which would increase the installation and maintenance costs. Thus, to improve the practical application of the model proposed in this study, a goal was set to minimize the required sensors, and to simplify the model, an attempt was made to reduce the training features to a minimum. Hence, a baseline model was designed to incorporate solely the temperature sensors inside the greenhouse, and with the sequential addition of a training feature displaying a high correlation coefficient, a total of four steps were designed (Table 3-1). For the model of each step, the air temperature prediction accuracy at each site inside the greenhouse was evaluated, and the proposed model was the one with an optimal level of accuracy.

Table 3-1 Training features of the simplified model to reduce the kind of installed sensor

Simplified model	Training features
Simple LSTM-1	Internal air temperatures
Simple LSTM-2	Internal air temperatures, external air temperatures, wind direction, wind speed
Simple LSTM-3	Internal air temperatures and relative humidity, external air temperatures and relative humidity, wind direction, wind speed
Basic LSTM	Internal air temperatures, relative humidity, UV radiation and illuminance, external air temperatures, relative humidity and UV radiation, wind direction, wind speed

3.3. Results and discussion

3.3.1. Statistical analysis of measured environmental data

Air temperature data measured in summer season (July 2017) at nine locations inside the naturally ventilated greenhouse and 1 location outside the greenhouse are shown in figure 3-7 as highest, lowest, average, 1st and 3nd quantile values. The average air temperatures at the nine locations inside the greenhouse were similar to the outside air temperature, which may be affected to continuous natural ventilation during the summer. However, due to the illuminance in summer, the highest air temperature inside the greenhouse was 41.6℃, which was higher than the highest air temperature outside, which is 33.3℃. Considering that the optimum growing temperature range for Irwin mango was 22 ~ 30°C, it was necessary to introduce an additional cooling system in the daytime. The mean air temperature values of nine locations in the greenhouse showed a maximum difference of 1°C, but the maximum difference was 1.8°C in the case of tertiles. The average soil temperature in the greenhouse was 29.4°C, which was similar to the average air temperature in the greenhouse of 28.4°C. Average, maximum, minimum and standard deviation values for other soil relative humidity, UV radiation and illuminance are shown in the Table 3-2. On the other hand, the collected data were greatly affected by the external wind environment because the target greenhouse was constantly in natural ventilation during the measurement period. Therefore, the external wind environment was analyzed and described as wind rose (Figure 3-8). The average wind speed in the area where the greenhouse was located was 43.1% for 2.71 $m \cdot s^{-1}$, and 0 to 2 $m \cdot s^{-1}$ and 22.0% for over 4 $m \cdot s^{-1}$. Wind directions were mainly analyzed in north-northeast (NNE), northeast (NE), east-northeast (ENE) and east (E). Since the average wind

speed in the study area was relatively high, the influence of natural ventilation was relatively large, and it was judged that the sensors at positions #7 to #9 were greatly affected by the outside air temperature.

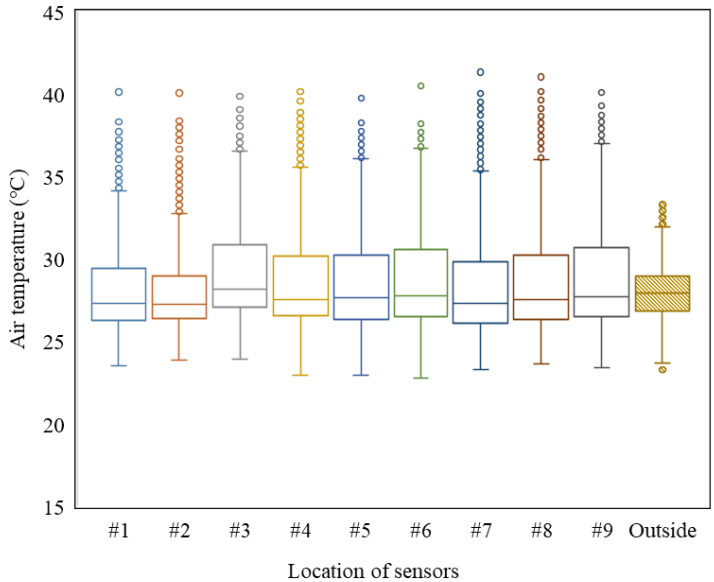


Figure 3-7 The boxplot for the temperature measured at each sensor location containing average, highest, lowest, 1st quartile, 3nd quartile value in July 2017

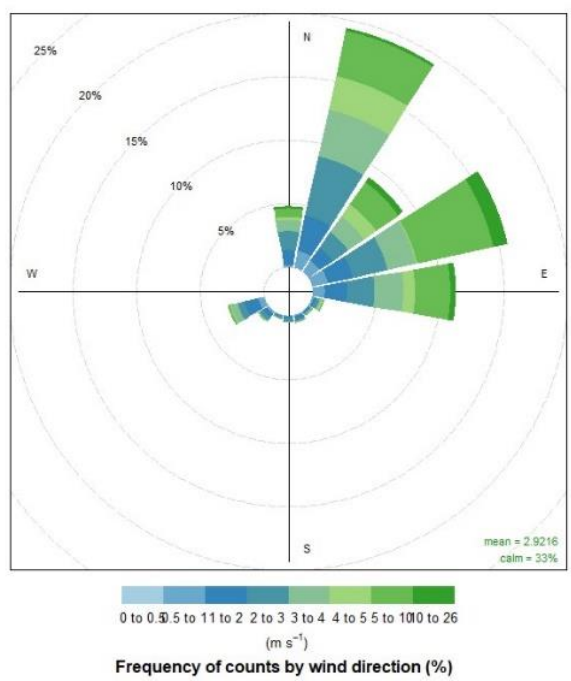


Figure 3-8 Wind rose of the wind conditions outside the greenhouse in July 2017

Table 3-2 Statistical analysis of measured environmental factors inside and outside the greenhouse (Air temp. : air temperature; Air RH: air relative humidity; Soil temp.: soil temperature; Soil RH: soil relative humidity; UV rad. : UV radiation; Illum.: illuminance)

		Air temp. (℃)	Air RH (%)	Soil temp. (℃)	Soil RH (%)	UV rad. ($W \cdot m^2$)	Illum. (lux)
Ave.	inside	28.4	80.3	29.4	18.6	26.9	7337.0
	outside	28.0	67.4	-	-	10.1	-
Max.	inside	41.6	100.0	34.2	43.6	511.0	65535.0
	outside	33.3	77.0	-	-	73.0	-
Min.	inside	22.8	41.9	26.3	2.5	0	0
	outside	23.3	44.0	-	-	0	-
Std.	inside	2.8	9.7	1.5	9.4	45.2	16.9
	outside	1.8	5.9	-	-	16.9	-

3.3.2. Correlation analysis of the dataset

The appropriate configuration of learning data is good for efficient computation and accuracy improvement of machine learning models, as well as minimizing the kind of sensors for collecting learning data. A typical appropriate feature selection method is correlation analysis. Through correlation analysis, a machine learning model can be designed by selecting learning data showing a high correlation with the predicted factor.

In this dissertation, for minimum sensor installation and maximum efficiency of the ML model computation, a correlation analysis was performed to select the learning feature. The analysis focused on the correlation of the air temperature at each sensor site inside the naturally ventilated greenhouse with the environmental

factors (indoor air temperature, relative humidity, soil temperature, soil humidity, EC, CO_2, atmospheric pressure, and UV radiation and illuminances). The values of the environmental factors at all sensor sites and the temperature at each sensor site were used in the correlation analysis (Figure 3-9). This means that environmental data with high correlation coefficient (p) should be considered first in the design of machine learning models.

The estimated correlation coefficient for the greenhouse indoor air temperature was 0.883, with a strong positive correlation. In most cases, the correlation was strong at $p \geq 0.85$. The correlation coefficient for the indoor relative humidity was -0.765, with a strong negative correlation. This is presumed to be because the relative humidity decreases as the temperature increases during summer. In addition, the correlation coefficients for UV radiation and illuminances were 0.659 and 0.644, respectively, with a strong positive correlation. Thus, the data on air temperature, relative humidity, and UV radiation and illuminances were selected for the training set in the ML model development.

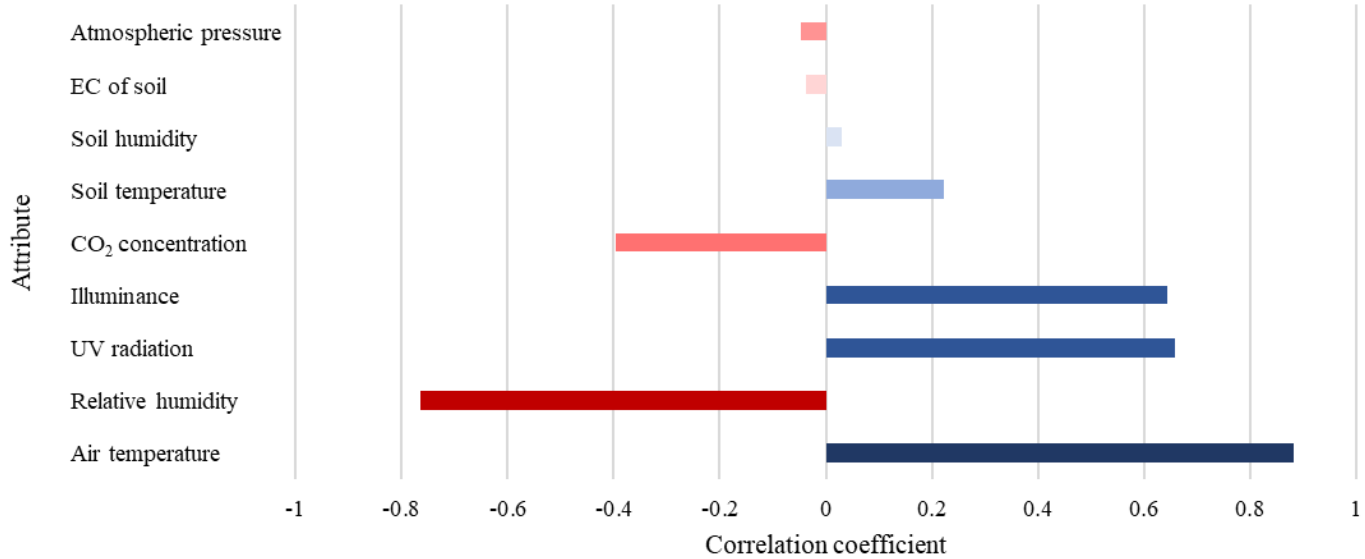


Figure 3-9 Correlation analysis between internal air temperature of greenhouse and each environmental factors (Air temperature, Relative humidity, UV radiation, Illuminance, Soil temperature, Soil humidity, EC of soil, CO_2 concentration, Atmospheric pressure)

3.3.3. Evaluation of machine learning models and selection of best performing model

The ANN, LSTM, and SVR models were developed as the ML models for optimal sensor location to predict the naturally ventilated greenhouse indoor air temperature, and the prediction accuracy of each model was evaluated to select the one with the highest performance level. The data collected at each sensor site were used as the training set and the air temperature at all other sensor sites was predicted, with the evaluation based on R^2, RMSE, and P-RMSE. In other words, the data at Sensor #1 as the training set were used to predict the indoor air temperatures at Sensors #2 to #9 through eight models, and so on for all nine sensors, to obtain statistical indicators for the prediction models. Thus, by estimating the mean of these indicators for each model, the performance of greenhouse indoor air temperature prediction was compared across the ML models (Table 3-3).

First, based on R^2, the value of which was 0.955, 0.974, and 0.943 for ANN, LSTM, and SVR, respectively, the LSTM model was evaluated as the one with the highest prediction accuracy for the trend in indoor air temperature. Next, the RMSE values were 0.031, 0.024, and 0.034 for ANN, LSTM, and SVR, respectively, and the LSTM model was evaluated as the one producing the least errors. Next, the P-RMSE values, which allow the analysis of prediction errors on maximum or minimum values, were 0.599, 0.458, and 0.649 for ANN, LSTM, and SVR, respectively, and the LSTM model was again the model producing the least errors. The results thus indicated that the LSTM model was the one with the highest prediction accuracy on peak values. This is presumed to be because the training set used in this study was a time-series data so that each dataset was under the influence

of temporal order, and as the LSTM model most adequately reflects such aspects of the data, the performance could have been the highest.

Table 3-3 Prediction accuracy of air temperature inside greenhouse according to the machine learning models

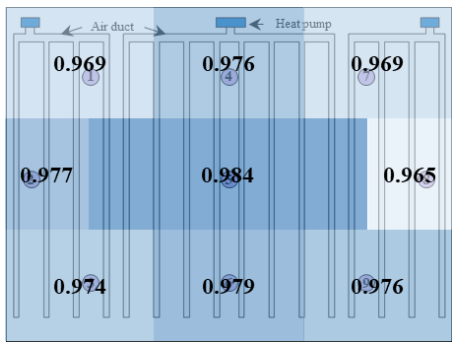
Statistical index	ANN	LSTM	SVR
R^2	0.955	0.974	0.943
RMSE	0.031	0.024	0.034
P-RMSE	0.599	0.458	0.649

3.3.4. Evaluation of predicting internal air temperature according to sensor location

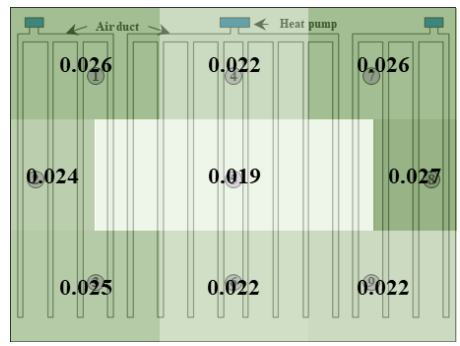
A machine learning model was developed for predicting the air temperature inside the naturally ventilated greenhouse, and LSTM was selected as the model that showed the best performance. The LSTM model predicted the air temperature at the other 8 sensor locations using the measured data from sensors at each location inside the greenhouse. For example, an LSTM model trained on data measured by sensor #1 predicts air temperatures at sensor locations 2 through 9. In this study, the average accuracy of predicting the air temperature at sensor location #2 to #9 was calculated in order to evaluate the sensor #1 as the optimal sensor location. In this way, the optimum sensor location was evaluated using each statistical index for sensor positions #1 to #9 (Figure 3-10).

First, the #5 sensor location, which was the center position of the experimental greenhouse, was evaluated as the optimal sensor location about all statistical indicators. On the other hand, sensors #7 and #8 showed relatively low prediction trends (R^2) and high error levels (RMSE, P-RMSE). This was because the main wind

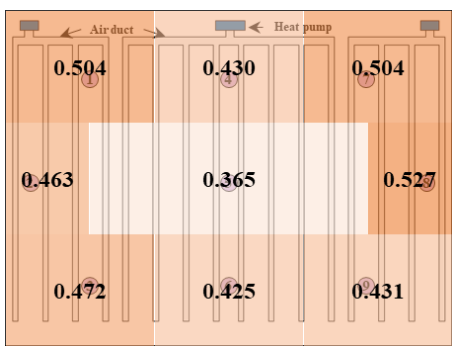
direction in the target greenhouse was the northwest wind, and sensors #7 and #8 were greatly affected by the outside air. Therefore, it was found that the prediction performance of the greenhouse air was degraded. Therefore, it was interpreted that the prediction performance for the indoor air of the greenhouse has deteriorated. In the case of #1 sensor location, when the northwest wind that was the main wind direction blows, it corresponded to an area where the air flow was relatively weak compared to the other sensor locations. Therefore, due to insufficient exchange with greenhouse air, it also showed low prediction performance. So, when selecting the optimal sensor location, it was judged that an area mainly affected by external air or an area with stagnant airflow inside the greenhouse should be excluded.



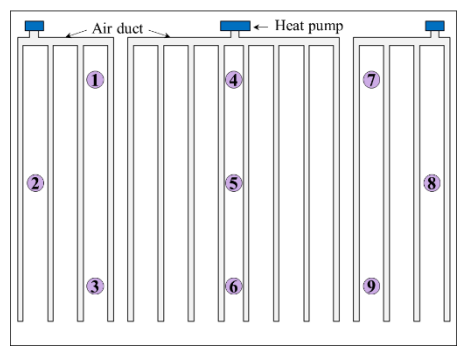
(a) Predictive accuracy based on R^2



(b) Predictive accuracy based on RMSE



(c) Predictive accuracy based on P-RMSE



(d) Top view of greenhouse and sensor locations

Figure 3-10 (a), (b), (c) accuracy of the LSTM model that learned data from each sensor location and predicted air temperatures at eight different locations according to the statistical indices and (d) the distribution of each sensor

3.3.5. Evaluation of optimal sensor location in the LSTM and simplified LSTM models

In this study, an ML model was developed towards optimal sensor location for predicting the naturally ventilated greenhouse indoor air temperature. Among the three ML models developed, the LSTM model with the highest accuracy was selected as the final model. The training set for the selected model included the

air temperature, relative humidity, and UV radiation and illuminances at each sensor site inside the greenhouse that had been identified in a correlation analysis for the respective measurements and the greenhouse indoor air temperature data. Although a training set including numerous features could increase the model accuracy, when considering the application of the model in practice, it could result in high sensor installation and maintenance costs. Thus, a simplified LSTM model was designed with an aim to minimize the sensor installation and the model accuracy was evaluated at each step of reducing the learning features. Table 3-4 lists the optimal sensor location indicated by the model at each step of simplification with the respective statistical indicators, which were mean values of the estimated air temperatures at eight sensor sites using the data of air temperature measured at the selected site in the training set.

For the simple LSTM-1 model using the training set including solely the air temperature at each indoor sensor site and for the simple LSTM-2 model using the training set including the air temperature at each indoor sensor site as well as the outdoor air temperature, and wind direction and speed, the R^2 was 0.856 and 0.863, RMSE was 0.057 and 0.049, and P-RMSE was 0.664 and 0.945. In contrast, for the simple LSTM-3 model using the training set to which the data on indoor and outdoor relative humidity were added, the R^2, RMSE, and P-RMSE were 0.982, 0.020, and 0.380, respectively, and for the basic LSTM model using the training set to which the data on indoor and outdoor UV radiation and illuminances were added, the R^2, RMSE, and P-RMSE were similar to those of the simple LSTM-3 model. It was thus determined that the simple LSTM-3 model using the training set including the indoor and outdoor air temperature and relative humidity and the outdoor wind direction and speed led to an adequately high accuracy.

The optimal sensor location after training set minimization mostly led to the selection of Sensor #5. In the case of applying the P-RMSE of the simple LSTM-2 model, Sensor #9 was selected as the optimal sensor site. In addition, the simple LSTM-1 model indicated Sensor #6 or #9 as the optimal sensor site based on the statistical indicators. Thus, if a limited set of sensors can be installed, it would be appropriate to install the sensors at sites #6 or #9, but in consideration of the model prediction accuracy, site #5 was determined to be appropriate based on the simple LSTM-3 model. These findings may prove valuable in optimal sensor location according to the sensor being applied and the target levels and values of statistical indicators.

Additionally, the optimal sensor locations for using two or three sensors were evaluated using the same method. This is because even though a single optimal sensor has the advantage of requiring minimal cost, it poses the risk that monitoring of the entire interior of the greenhouse is not possible in the event of a sensor failure. When using two sensors, sensors at sites #3 and #9 were used as learning sensors showed the best prediction accuracy for all statistical indicators (R^2=0.977, RMSE=0.022 and P-RMSE = 0.424). Sensors #3 and #9 were sensors located in both side windows of the greenhouse, respectively, and had the advantage of considering both the air temperature that was inflow and outflow during natural ventilation. When using three sensors, sensors at sites #3, #4, and #7 showed the highest R^2 (0.980) and the second lowest RMSE (0.024) and P-RMSE (0.464), respectively. Sensors at sites #1, # 2, and #7 showed the second highest R^2 (0.979), while showing the lowest RMSE (0.022) and P-RMSE (0.428). Therefore, sensors #1, #2, and #7 were selected as optimal sensor locations based on RMSE and P-RMSE, which showed significant differences. Sensors #1, #2, and #7 included both congested areas

of the experimental greenhouse and locations mainly affected by the external wind environment. Therefore, it was judged that the environment of various locations in the greenhouse could be considered.

Table 3-4 Optimal sensor location selection result of basic LSTM model and simple LSTM models according to statistical indicators (The basic LSTM model was the model that has learned all features with high correlation, and the simple LSTM models were the models that has reduced train features according to the correlation)

ML model		Simple LSTM-1			Simple LSTM-2			Simple LSTM-3			Basic LSTM		
Statistical index		R^2	RMSE	P-RMSE	R^2	RMSE	P-RMSE	R^2	RMSE	P-RMSE	R^2	RMSE	P-RMSE
Optimal value		0.863	0.057	0.664	0.863	0.049	0.945	0.982	0.020	0.380	0.984	0.019	0.365
Sensor location	1	0.790	0.070	1.363	0.856	0.058	1.118	0.973	0.024	0.459	0.969	0.026	0.504
	2	0.380	0.113	1.993	0.762	0.080	1.533	0.980	0.022	0.429	0.977	0.024	0.463
	3	0.822	0.066	1.073	0.840	0.059	1.137	0.977	0.023	0.447	0.974	0.025	0.472
	4	0.835	0.066	0.926	0.821	0.055	1.062	0.972	0.024	0.455	0.976	0.022	0.430
	5	0.848	0.060	0.862	0.863	0.049	0.951	0.982	0.020	0.380	0.984	0.019	0.365
	6	0.856	0.057	0.863	0.856	0.051	0.991	0.978	0.022	0.425	0.979	0.022	0.425
	7	0.843	0.058	1.013	0.853	0.052	0.996	0.969	0.026	0.492	0.969	0.026	0.504
	8	0.732	0.092	0.792	0.764	0.067	1.296	0.966	0.027	0.512	0.965	0.027	0.527
	9	0.841	0.068	0.664	0.851	0.049	0.945	0.978	0.022	0.414	0.976	0.022	0.431

3.4. Conclusion

In this study, ML modelling was applied to determine optimal sensor location for predicting the air temperature at each site inside the naturally ventilated greenhouse. For this, linear interpolation was performed on missing values and outliers followed by normalization of the training set, and ANN, LSTM, and SVR models were developed for comparing the respective accuracy of predicting the greenhouse indoor air temperature. The results indicated that the LSTM model had the highest level of prediction accuracy based on R^2 = 0.974, RMSE = 0.024, and P-RMSE = 0.458. The LSTM model was selected as the PCTO-ML, and the consequent optimal sensor location was evaluated for the prediction of greenhouse indoor air temperature. As a result, when using one sensor, sensor #5 was shown to be the optimal sensor site. In addition, by evaluating the prediction accuracy of the PCTO-ML in line with the reduction in the training set, the number and type of required sensors were minimized. As a result, the use of the indoor and outdoor air temperature and relative humidity and the outdoor wind direction and speed were shown to be the minimum required set to ensure an adequate accuracy (R^2 = 0.982, RMSE = 0.0.020, and P-RMSE = 0.380). Additionally, the optimal sensor locations when using two sensors were sensors #3 and #9, and the optimal sensor locations when using three sensors were evaluated as sensors #1, #2, and #7.

The optimal sensor location based on the PCTO-ML proposed in this study will allow the prediction of the air temperature at each site inside the greenhouse using a minimum number of sensors and to a high level of accuracy. Thus, the reduction in sensor installation and maintenance costs can be expected. This positive effect is predicted to be maximized when the model is applied to a large-scale greenhouse

requiring numerous monitoring sensors. In addition, as an PCTO-ML is characterized by a continuous improvement through additional learning of data, the prediction accuracy could likely be continuously increased regarding different environmental conditions. The PCTO-ML proposed in this study allows not only optimal sensor location but also the prediction of air temperature at each site inside the greenhouse, and with the advancement of an information and communication technology (ICT), the real-time data of greenhouse indoor conditions at each optimal sensor site are anticipated to contribute to the real-time monitoring at each site inside the greenhouse.

Chapter 4. Time series forecasting for air temperature inside a naturally ventilated greenhouse with optimal sensor location based on LSTM: Prediction

4.1. Introduction

Crop cultivation under greenhouse technologies offers the advantage and convenience to artificially and even remotely control and adjust the internal environmental conditions to meet and maintain an appropriate growing environment favourable for crops toward an improved productivity. Environmental factors such as air temperature, relative humidity, and CO_2 greatly affect plant growth and quality of produce. This justifies the importance of proper control and management of the growing environment.

Controlling the internal environment of the greenhouse could be challenging due to complex environmental systems, including nonlinearity of response variables, time variation, and uncertainty (Wang, Wang et al. 2009, Hamza and Ramdani 2020). It is therefore important to understand and manage the complexity of the internal environment of the greenhouse. The greenhouse consists of crops that perform photosynthetic activities that biologically breathe inside. Also, it consists of several components such as floors which may interact with plants by either diffusing or absorbing solar radiations. The greenhouse is then comprised of various components and processes that perform heat exchange in the form of sensible and latent heat (Singh, Singh et al. 2006, Sethi, Sumathy et al. 2013, Yeo, Lee et al. 2022). During

summer, the air temperature inside the greenhouse is relatively high, so even when ventilation is started, heat is emitted into the greenhouse by various latent heat elements, including the floor that absorbs solar radiation. This phenomenon will result to temporal rise in the air temperature inside the greenhouse especially during day time even though ventilation has started. If the rise of air temperature inside the greenhouse is predicted efficiently by available models, then natural ventilation may be performed in advance to avoid extreme rise of air temperature inside. Compared to traditional ventilation where it only starts to operate the moment the air temperature reached its peak or level depending on the setting of the ventilation. The former way of ventilation can be more efficient as compared to the later one. Therefore, it is deemed necessary to have a pre-emptive environmental control of the internal environment of the greenhouse. This way, the energy input for managing the internal environment can be reduced, which can lead to improved productivity and sustainability of the overall setup.

Research methods for predicting future situations using time series data include method using a statistical model and a method using machine learning. The use of various statistical models that included regression analysis method (Okamoto and Koshi 1989, Amral, Ozveren et al. 2007); Kalman filtering method(Heemink and Segers 2002, Louka, Galanis et al. 2008, Nobrega and Oliveira 2019, Zhou, Guo et al. 2020); and Box-Jenkins's autoregressive integrated moving average (ARIMA) (Ho, Xie et al. 2002, Cadenas, Rivera et al. 2016, Sen, Roy et al. 2016, Alsharif, Younes et al. 2019, Shadab, Said et al. 2019, Lai and Dzombak 2020) were explored in previous works. ARIMA, the most common in particular, is a model that considers both autoregression and moving average as methods of statistical analysis of traditional time series data. Statistical models, including ARIMA, have great

strengths in predicting regular data, but generally have limitations in predicting environmental data measured in natural environments with nonlinearity characteristics (Nie, Liu et al. 2012). Furthermore, the ARIMA model should have the attribute of stationary for time series data or be converted into data with stationary.

Since the machine learning model is a black box model, there is no need to consider the physical relationship between the data, and the stationary of the data is not guaranteed like statistical models such as ARIMA. With the recent development of algorithms and the sustained growth of computer performance, prediction research is being conducted using machine learning, which has strength in nonlinear data prediction (Bontempi, Ben Taieb et al. 2012, Guo, Juan et al. 2017, Khosravi, Machado et al. 2018, Alhnaity, Pearson et al. 2019, Liu, Wang et al. 2019, 오정원, 김행곤 et al. 2019, Hutapea, Pratiwi et al. 2020, 홍성은, 박태주 et al. 2020, Gong, Yu et al. 2021, Ozbek, Sekertekin et al. 2021, Seng, Zhang et al. 2021, Zarinkamar and Mayorga 2021, Ge, Zhao et al. 2022, Lee, Lee et al. 2022). (Gong, Yu et al. 2021) predicted tomato yields in greenhouses using the temporary convolutional network (TCN) and recurrent neural network (RNN). Tomato yields were predicted based on time series environmental data such as air temperature, relative humidity, carbon dioxide, and radiation in the target greenhouse. (Lee, Lee et al. 2022) predicted the internal air temperature and relative humidity of naturally-ventilated duck house using the RNN model. The RNN model was designed by learning time series data on the air temperature and humidity inside the duck house and information such as the type of duck houses. The appropriate sequence length of the RNN model was evaluated, and finally, an RNN model with less than 1% was presented. Despite numerous studies about machine learning, only few studies are

available relative to monitoring or predicting the internal environment such as air flow, temperature, and humidity related to the growing environment that has a great influence on the crop production in the greenhouse.

Machine learning models are expected to perform well in predicting the internal environment of the greenhouse, but various kinds of learning data are needed for the development of machine learning models with high predictive performance. In addition, for the internal uniformity of greenhouses, data collection at various points is essential for predicting the internal environment of each location inside the greenhouse. As a result, it may be necessary to install a large number of sensors to introduce a machine learning model. Especially, one of the biggest challenges of smart greenhouses with many sensors and control systems has been pointed out as initial and maintenance costs (MAFRA, 2022). Therefore, it is important to install an efficient monitoring sensor and to consider the optimum number of the sensors. The optimum number of sensors can be determined when the appropriate sensor locations in the greenhouse were already determined.

The purpose of this study was to develop the prediction future air temperature using optimal sensor machine learning model (PFTO-ML) for predicting the future air temperature of each of the nine sampling locations in the greenhouse considering the optimized (minimum) number of sensors using the PCTO-ML. Specifically, an PFTO-ML was developed to predict the air temperature of the sensor in advance by learning the environmental data and external environmental data measured from one of the sensors installed in the greenhouse. Furthermore, in order to minimize the number of sensors to be installed in the greenhouse, the future air temperature for the nine locations in the greenhouse was predicted by learning environmental data and external environmental data measured at the optimal sensor location. For the

development of the PFTO-ML, various environmental data such as air temperature, relative humidity, and CO_2 inside the experimental greenhouse were collected. Missing and outliers were corrected through data pre-processing, and it was selected as a feature of the PFTO-ML through correlation analysis. The PFTO-ML, which is a prediction model using time series data, is affected by prediction accuracy according to the sequence length, which is a unit of input data. The appropriate sequence length was evaluated through a test for the sequence length. Future air temperature prediction for each sensor location inside the greenhouse was performed using the developed PFTO-ML, and the derived results were evaluated. Additionally, for minimizing sensor installation, future air temperatures for the nine locations inside the greenhouse were predicted using only the data measured at the optimal sensor location which was suggested by PCTO-ML.

4.2. Materials and methods

This study followed the flow chart presented in Figure 4-1. The main purposes of this study were 1) development of prediction model for future air temperature inside naturally ventilated greenhouse 2) application of optimal sensor location on the prediction model. The temperature of each location inside the naturally ventilated greenhouse was predicted using environmental data measured from the sampling points inside the greenhouse. The environmental data such as air temperature and relative humidity, electrical conductivity (EC) and illuminance were collected from the nine locations in the target greenhouse during summer. In order to consider the effect of natural ventilation, a portable weather station was installed to collect external air temperature, relative humidity, wind direction, and wind speed data. An LSTM model with strengths in prediction using time series data was developed in this study. First, data pre-processing was performed to develop a high-performance LSTM model. Linear interpolation of missing and outliers was performed on the environmental data collected for a month of July in 2017. Various data such as air temperature, relative humidity, EC, illuminance, etc. inside the greenhouse were collected. Normalization was performed among the data, since the units of the environmental data were different, in order to avoid bias in the specific learning features and was reflected in LSTM model learning. By using the min max normalization scale, the range of all environmental data was normalized to [0, 1]. After which, upon pre-processing the learning data the LSTM model was developed to predict the air temperature for the near future of each location inside the greenhouse. At this time, the main hyper-parameter of the LSTM model used the design information of previous studies that predicted the internal air temperature of

the greenhouse while using the same data. Meanwhile, the sequence length which is the hyper-parameter of the LSTM model means the time unit of future prediction and affects the accuracy of the model. Therefore, the appropriate sequence length was calculated by evaluating the model accuracy for the sequence length values of 30, 60, 120, and 240 min. The final developed PFTO-ML could predict the future air temperature for the sensor locations used as training data among the data collected from the nine different locations in the greenhouse. In this study, the optimal sensor location from PCTO-ML was applied to minimize the installation of sensors in large greenhouses. Finally, the PFTO-ML was developed using the environmental data of the optimal sensor location proposed by PCTO-ML, and compared with the basic LSTM model.

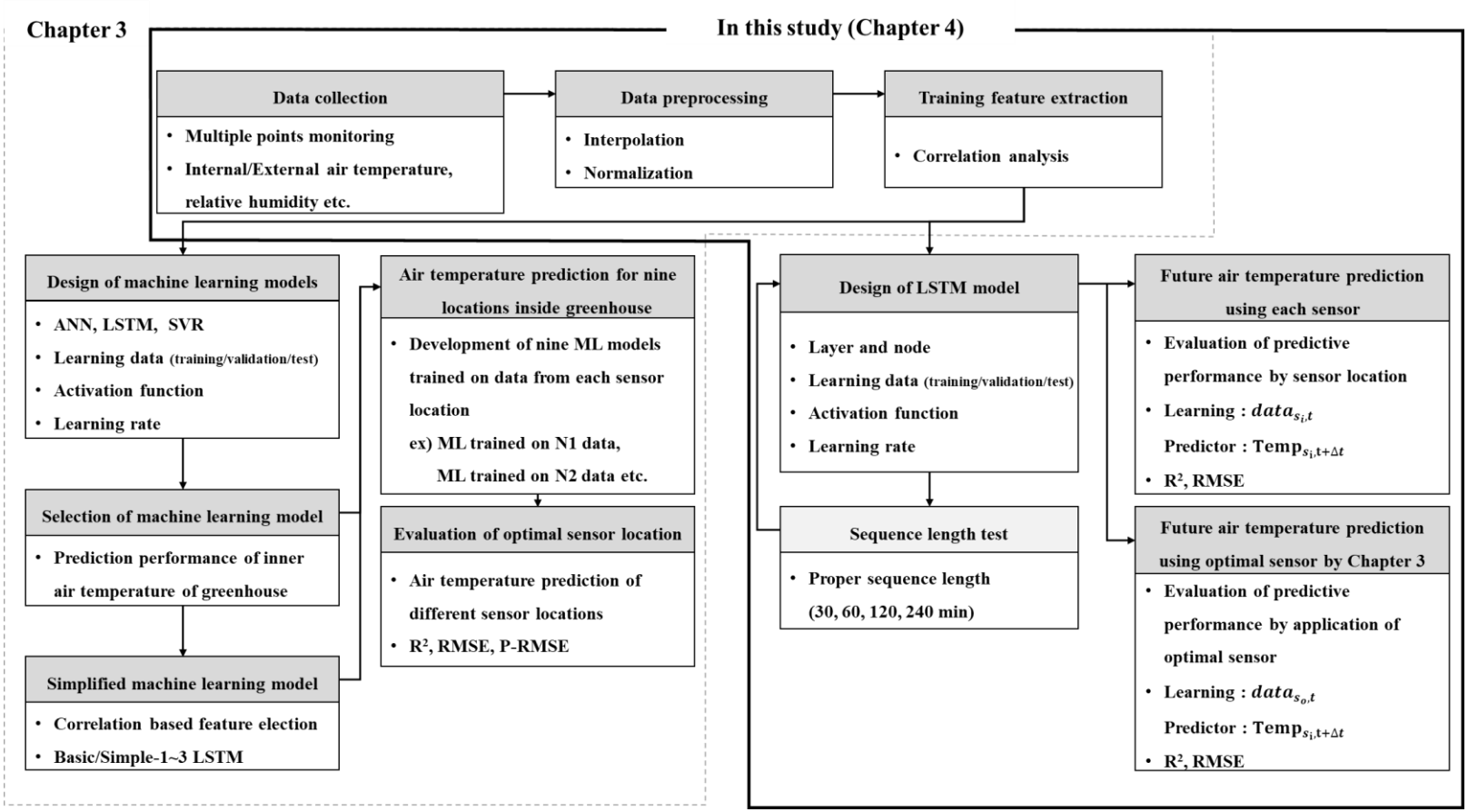


Figure 4-1 Flowchart of the experimental procedure

4.2.1. Target greenhouse

The target greenhouse was a 1-2W type, 8-span plastic film greenhouse, located approximately 1.0 km to the south of the Boryeong Thermal Power Plant in Gojeong-ri, Jugyo-myeon, Boryeong-si, Chungcheongnam-do, South Korea (126°29'E, 36°23'N). The greenhouse structure was designed with a width of 4.0 m, a length of 30.0 m, a ridge height of 4.5 m, and an eave height of 5.7 m. The interior space of the greenhouse was largely divided into work space and cultivation space. The work space was 4.0 m from the entrance for the purpose of simple work and mechanical system control inside the facility, and the other area consisted of a crop growth space (Figure 4-2). The cultivation area had a floor area of 768.0 m^2. The cultivation space was covered with a single 0.15 mm thick agricultural polyolefin film on the ceiling, sides, front and rear. Inside the greenhouse, 100.0 Irwin mangoes were grown in a total of 13.0 rows as shown in Figure 4-2. The air environment inside the greenhouse was controlled by natural ventilation and heat pumps (ADF-SLX12WHB, A-San Inc., Korea) through side and roof vent openings. The side window of the greenhouse was installed at a vertical height and was located at a height of 0.5 to 2.1 m and 2.7 to 3.5 m from the floor. The roof vent opening was opened and closed by a cover of about 0.6 m in the vertical direction at the top of the greenhouse roof. The heat pump capacity in the greenhouse was 43, 276.0 W for cooling and 36,786.0 W for heating, respectively, and supplies uniform cooling and heating inside the greenhouse through a duct with a diameter of approximately 60.0 cm

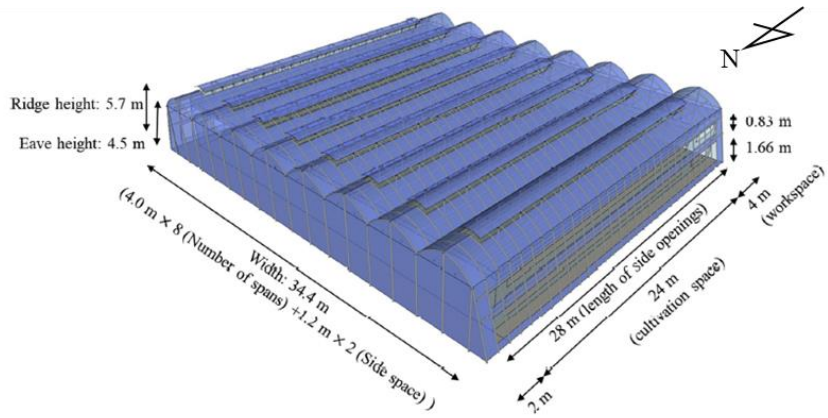


Figure 4-2 Schematic drawing of target greenhouse

(a) Irwin mangoes

(b) heat pump

(c) roof vent opening

(d) side vent opening

Figure 4-3 Crops and air conditioning facilities in the experimental greenhouse located on the west coast of South Korean in Jugyo-myeon, Boryeong-si, Chungcheongnam-do Province (126˚29'E, 36˚23'N). (Irwin mangoes, roof and side vent opening and heat pump)

4.2.2. Long Short-Term Memory (LSTM)

A Recurrent Neural Network (RNN) is a type of neural network that contains loops, allowing information to be stored within the network (Rumelhart, Hinton et al. 1986, Werbos 1990). RNN is an artificial neural network that is suitable for handling time series data, and unlike general neural networks, repetitive learning is possible through the memory inside the neural network. Memory is a suitable algorithm for the field of processing time series data because it can save information obtained from previous steps of learning and provides feedback with this information to consider previous steps of input data. RNN is optimized for training and forecasting time series data, but have poor long-term predictability due to long-term dependency problems that make it difficult to connect past information to current tasks. This problem was addressed through the development of an LSTM model with improved long-term dependence through memory cells (Hochreiter and Schmidhuber 1997). In this study, time series environmental data were collected at nine sensor locations in the experimental greenhouse. Therefore, it was judged that LSTM was suitable as a machine learning model that showed excellent performance in prediction using time series data.

The memory cell computes the input state (i) with the input data (x) at the current time step (t) and the hidden state (h) at the previous step (t-1) (Eq. 4-1). Through the input state, input gate (g) (Eq. 4-2), forget gate (f) (Eq. 4-3), and output gate (o) (Eq. 4-4) are calculated, and the internal state (m) (Eq. 4-5) of the current time step (t) is updated using these values and the internal state value of the previous time step (t - 1). This is followed by calculation of the hidden state (h) (Eq. 4-6) of the current step to be used in the next step (t + 1) (Wang and Raj 2017). LSTM has the advantage of

being able to make longer-term predictions by considering short-term information and old past information.

$$i^{t} = \sigma(W_{ix}x^{t} + W_{ih}h^{t-1}) \quad \text{(Eq. 4-1)}$$

$$g^{t} = \sigma(W_{gi}i^{t}) \quad \text{(Eq. 4-2)}$$

$$f^{t} = \sigma(W_{fi}i^{t}) \quad \text{(Eq. 4-3)}$$

$$o^{t} = \sigma(W_{oi}i^{t}) \quad \text{(Eq. 4-4)}$$

$$m^{t} = g \odot i^{t} + f^{t}m^{t-1} \quad \text{(Eq. 4-5)}$$

$$h^{t} = o^{t} \odot m^{t} \quad \text{(Eq. 4-6)}$$

where, σ is activation function, W_{ix} is the weight of the input data in the input state, W_{ih} is the weight for the hidden state in the input state, $\odot$ is wise-element product.

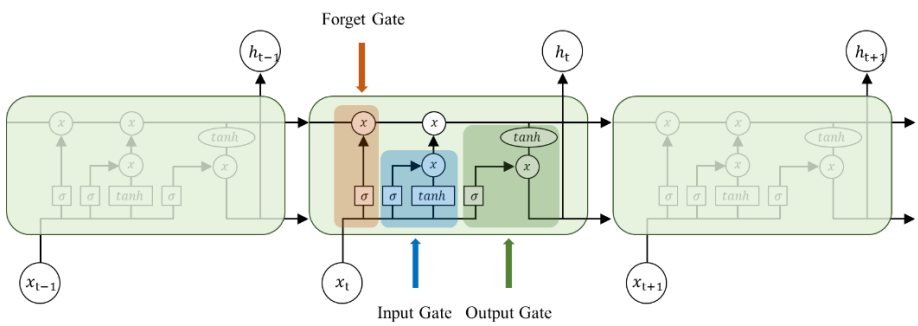


Figure 4-4 Conceptual diagram of LSTM model

4.2.3. Experimental procedure

4.2.3.1. Data collection for modelling PFTO-ML

Microclimate factors in greenhouses, such as the internal airflow, air temperature, relative humidity, CO_2 concentration, and illuminance, have an important influence on the growth and quality of crops. Therefore, microclimate factors within the greenhouse need to be monitored and conditioned through the air conditioning system. In this study, the internal microclimate factors were monitored at 1-s interval from the nine sampling points, as shown in Figure 4-5. Each sensor was installed at a height of 0.9 m from the floor. The observed data was processed into 10-minute average data and used as learning data for the machine learning model, and used to predict future air temperatures for each sensor location. The data was measured from a month of July in 2017 with all ventilation windows fully open for the target period. This means that it had a great influence on the microclimate factors inside the greenhouse depending on the external wind environment. A portable weather station was installed outside the greenhouse to monitor the external environment such as wind direction, wind speed, relative humidity, and air temperature every 10 minutes.

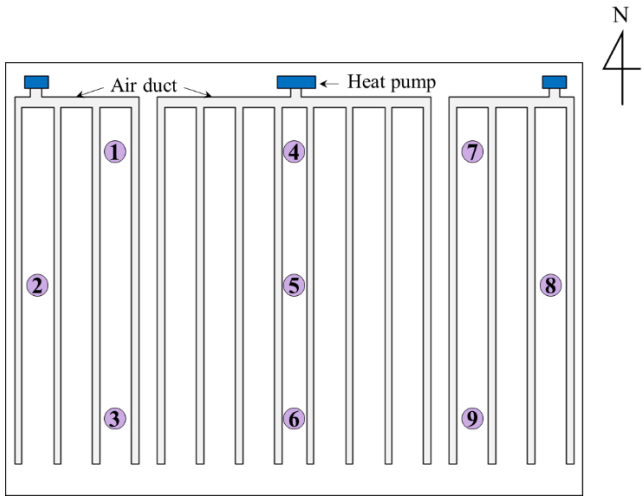


Figure 4-5 Location of installed sensors to measure environmental factors inside target greenhouse. The sensors were installed at the height of 0.9m above the floor

4.2.3.2. Pre-processing for dataset

When multidimensional values are considered in designing a machine learning model, the contribution to the machine learning model may be biased for specific learning data. The variables considered for developing the PFTO-ML in this study are air temperature, air relative humidity, soil temperature, soil humidity, EC, CO_2, atmospheric pressure, UV, and illuminance at different sensor locations in the naturally ventilated greenhouse. These variables were multidimensional values of varied characteristics and ranges. Among the features used as learning data in this study, the air temperature has a range of 24 to 40 degrees, while the radiation has a range of zero to tens of thousands of $W \cdot m^{-2}$. At this time, the contribution of radiation with a large unit may be biased toward the machine learning model compared to other features. To prevent the occurrence of bias, the dimensions of the measured dataset among independent variables were normalized before designing the PFTO-

ML. Hence, as shown in Eq. 4-7, the min max normalization scale, a technique for normalizing all datasets for training machine learning models, was applied by normalizing all values in the data to the range of [0, 1].

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad \text{(Eq. 4-7)}$$

where, x' is the normalized sample data, x is each individual sample data, x_{min} is the minimum value of sample data, and x_{max} is the maximum value of sample data.

The learning data used in this study were measured values obtained by installing sensors inside and outside the target greenhouse. There was a missing value according to the abnormal malfunction of the sensor, and linear interpolation was performed through interpolation. Linear interpolation is an interpolation method that assumes that two points on a one-dimensional straight line are linearly proportional to the straight distance to estimate the value between them.

4.2.3.3. Design of PFTO-ML

Machine learning models can improve their predictive performance when trained on many types of data, but the performance of machine learning models can decrease if they learning features with low data quality or low correlation with predictors. In addition, indiscriminately many kinds of learning data require many sensors for the field application of machine learning models, resulting in a decrease in the usability of machine learning models. Therefore, proper selection of learning data is important. Therefore, this study referred to the results of previous studies that completed the

correlation analysis between the data measured in the experimental greenhouse for the predicted air temperature inside the greenhouse in Chapter 3. As a result, the air temperature and relative humidity inside the greenhouse with the highest Pearson correlation coefficient were considered as learning feature, and the external air temperature, relative humidity, wind direction, and wind speed were also used as learning feature for the PFTO-ML to consider the effect of natural ventilation (Table 4-1).

Table 4-1 Correlation analysis between internal air temperature of greenhouse and each environmental factors from Chapter 3 (Air temperature, Relative humidity, UV radiation, Illuminance, Soil temperature, Soil humidity, EC of soil, CO_2 concentration, Atmospheric pressure)

Parameters	Correlation coefficient	Parameters	Correlation coefficient
Air temperature	0.883	Soil temperature	0.222
Relative humidity	-0.765	Soil humidity	0.031
UV radiation	0.659	EC of soil	-0.037
Illuminance	0.644	Atmospheric pressure	-0.047
CO_2 concentration	-0.395		

The main hyper-parameters constituting an LSTM model include the Sequence Length as the length of the data to be used in a single training, the number of hidden units present in the LSTM cell, loss function to evaluate the model's performance, and optimizer which is a function to reduce loss function. For the design of the LSTM model in this study, the design information of the LSTM model in Chapter 3, which used the same data and showed high accuracy in predicting the internal air temperature of the naturally ventilated greenhouse, was used. In previous studies,

the same environmental factors were measured in the same experimental greenhouse as this study. Previous studies used hyper-parameters optimized through iterative design and trial-and-error for hyper-parameters of LSTM models as design information (Table 4-2).

Meanwhile, sequence length, which is one of the main hyper-parameters of LSTM, which has strengths in predicting time series data, means the size of the input and output of the LSTM model. In other words, it is the unit of learning for LSTM models, and is generally known to affect the predictive performance of the machine learning model (Lee, Jun et al. 2019, Lee, Lee et al. 2022). Therefore, in this study, the appropriate sequence length was calculated by evaluating the accuracy of the future air temperature prediction of the PFTO-ML according to the sequence length. Sequence length was tested for 30, 60, 120, and 240 min (Table 4-2).

Table 4-2 Hyper-parameter of LSTM model from Chapter 3 and design of sequence length

Hyper-parameter	Value	Hyper-parameter	Value
number of layer	2	activation function	tanh
layer node	first layer: 32 second layer : 16	optimizer	adam
		batch size	64
loss function	MSE	epoch	200
Sequence length	30, 60, 120, 240 min		

4.2.3.4. Statistical indices for evaluating LSTM model

In this study, an LSTM model was used to predict point-wise air temperature inside a naturally ventilated greenhouse in the near future. Time series-type learning

data were used for various environmental variables measured in the experimental greenhouse during the LSTM model development. The predicted trend of air temperature inside the greenhouse was evaluated, and the error of the predicted value was evaluated. The correlation coefficient (R^2) value that identifies the tendency of internal air temperature prediction as an evaluation index of the LSTM and the RMSE was used to evaluate the error of the prediction value.

$$R^2 = \left(\frac{\sum_{i=1}^{n}(O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^{n}(O_i - \bar{O})^2}\sqrt{\sum_{i=1}^{n}(P_i - \bar{P})^2}} \right)^2 \quad \text{(Eq. 4-8)}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}[O_i - P_i]^2} \quad \text{(Eq. 4-9)}$$

where, O and P are the observed and simulated values, respectively, and $\bar{O}$ and $\bar{P}$ indicate the mean of observed and simulated values, respectively.

4.2.3.5. Application of optimal sensor location to minimize sensor installation

In this study, an PFTO-ML was designed to predict the future air temperature for the location inside the naturally ventilated greenhouse using time series environment data measured in the target greenhouse. The developed PFTO-ML can predict future air temperatures inside the greenhouse. However, to predict the air temperature distribution inside the greenhouse, a number of sensors were inevitable, which leads to an increase in the number of sensor installations and cost. Thus, to minimize the number of sensors, an LSTM-basic model was developed using the environmental

data measured at the optimal sensor location proposed by PCTO-ML as a learning data. Then, the air temperature at other sensor locations in the greenhouse was predicted using the Simple LSTM-3 model. PCTO-ML proposed sensor #5 in the center of the greenhouse as the optimal sensor location when one sensor was installed. In addition, #3 and #9 sensors were proposed as optimal sensor locations when two sensors were installed, and #1, #2, and #7 sensors were proposed as optimal sensor locations when three sensors were installed. With this, the PFTO-ML(#5), PFTO-ML(#3#9) and PFTO-ML(#1#2#7) models were developed using the environmental data measured by the optimal sensors as the learning data was used in this study. Finally, the accuracy of LSTM-basic and PFTO-MLs for predicting the future air temperature inside the greenhouse were compared.

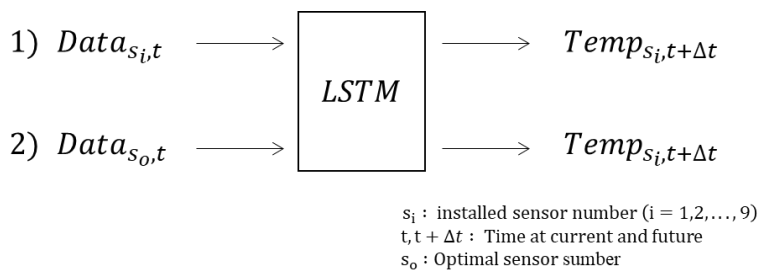
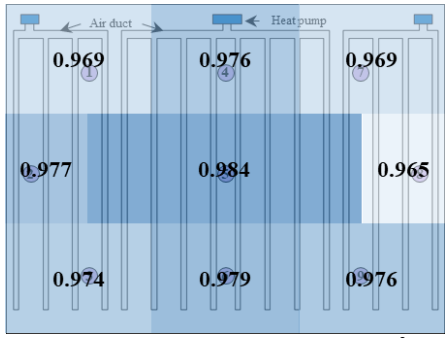
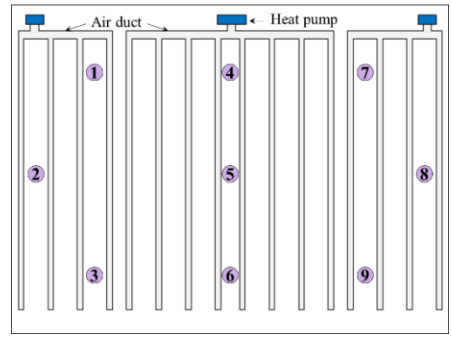


Figure 4-6 Brief description for predicting the future air temperature of a naturally ventilated greenhouse using LSTM, 1) Predictions using individual sensors 2) Predictions using optimal sensors



(a) Predictive accuracy based on R^2



(b) Top view of greenhouse and sensor location

Figure 4-7 The optimal sensor location for predicting the internal air temperature of the naturally-ventilated greenhouse using the LSTM-basic model from PCTO-ML (R^2 value at each point represents the accuracy in predicting the air temperature at other points using the sensor data at the corresponding location).

4.3. Results and discussion

4.3.1. Descriptive analysis of air temperature and wind environment in measured time series data

The LSTM model was designed in this study to predict the future air temperature for the location inside the naturally ventilated greenhouse. The characteristics of the air temperature inside the greenhouse, which was a predictor, was analysed through descriptive analysis by considering the measured air temperature data. The highest air temperature inside the greenhouse during the daytime was higher than the external highest air temperature due to the warming effect by the greenhouse cover, and the daily temperature difference was also greater than that outside. The highest air temperature inside the greenhouse was 41.6℃, which was higher than the highest air temperature outside, which is 33.3℃. The average air temperature inside the greenhouse was 28.4℃, but it often exceeded 30℃.

During the experimental period, the greenhouse remained naturally ventilated, so the external wind environment had a great influence on the inside air temperature. Therefore, wind roses describing the external wind environment were analyzed. (Figure 4-9). The average wind speed in the area where the greenhouse was located was 43.1% for 2.71 $m \cdot s^{-1}$, and 0 to 2 $m \cdot s^{-1}$ and 22.0% for over 4 $m \cdot s^{-1}$. The main wind direction was analyzed to the northeast. Since the average wind speed in the study area was relatively high, the influence of natural ventilation was relatively large, and it was determined that the sensors at positions #7 to #9 were greatly affected by the outside air temperature.

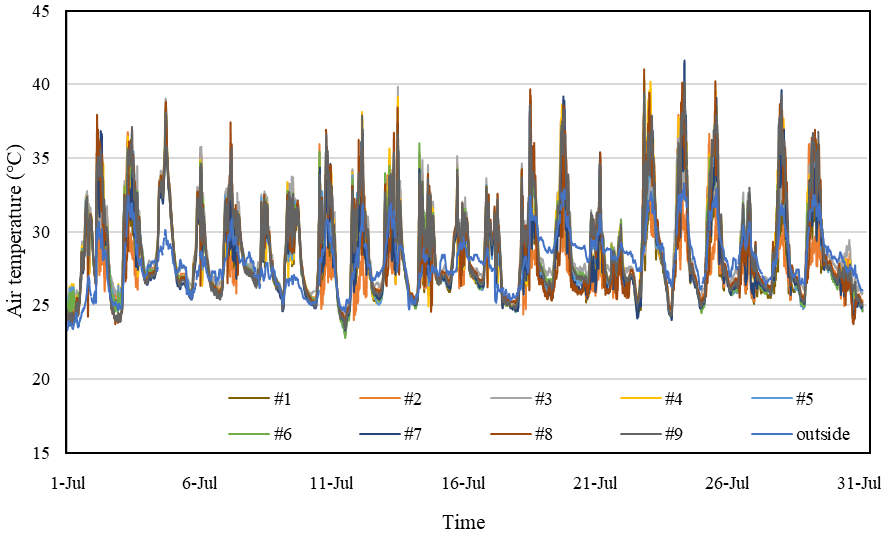


Figure 4-8 Air temperature collected inside and outside the experimental greenhouse during July, 2017

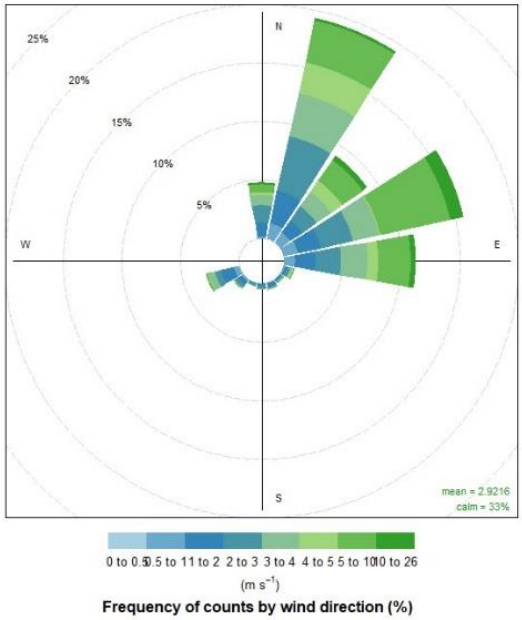


Figure 4-9 Wind rose of the wind conditions outside the experimental greenhouse in July, 2017

4.3.2. Model performance of LSTM according to sequence length

The sequence length means a unit of input data of the LSTM model, and the prediction accuracy of the model can be evaluated differently depending on the sequence length. This study attempted to calculate the appropriate sequence length by evaluating the accuracy of the LSTM model according to the sequence length, and a sequence length test was conducted on 30, 60, 120, and 240 min (Table 4-3). The accuracy of the LSTM model according to the sequence length was evaluated by averaging the static indexes of nine LSTM models that predicted future air temperatures with data measured at each sensor location. Result revealed that as the sequence length was increased, the prediction accuracy, indicated by R^2 decreased and the RMSE, which corresponds to the degree of error in the predicted air temperature, increased. Furthermore, when the sequence length was increased, the input data at distant locations on the time series data had fewer chances to learn, and the prediction accuracy decreased. On the other hand, from the results of this study, the shorter the sequence length, the higher the prediction accuracy, but the future situation of the short unit was predicted. Therefore, a sequence length that was too short might not be suitable from a prediction point of view. From this, the optimal sequence length was evaluated at 30 min.

Table 4-3 Air temperature prediction accuracy (R^2, RMSE) of LSTM according to sequence length

Statistical index	Sequence length			
	30 min	60 min	120 min	240 min
R^2	0.950	0.944	0.931	0.901
RMSE	0.642	0.691	0.795	0.929

4.3.3. Evaluation of air temperature prediction performance for

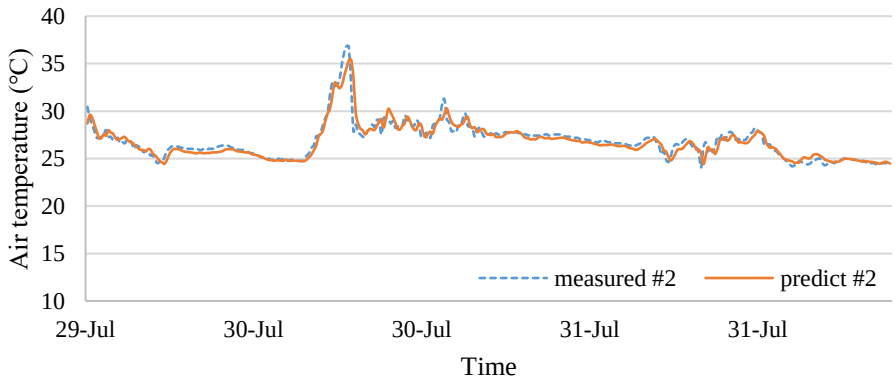
each sensor location in the greenhouse

The air temperature of each sensor location of the naturally ventilated greenhouse was predicted using the LSTM model with a sequence length of 30 min calculated earlier, and average prediction accuracy of LSTM models in each sensor location was shown in Table 4-4. Overall, the prediction of future air temperatures inside the greenhouse using the LSTM model showed high accuracy for most sensor locations ($R^2 > 0.95$, RMSE < 0.65). However, it showed relatively low prediction accuracy for some sensor locations, such as sensor 7 and sensor 2. This can be explained by the main wind direction in the target greenhouse which was formed in the northeast during the data measurement period. The experimental greenhouse was naturally ventilated through side and roof vent openings, and since the data measurement period was during hot and humid season, natural ventilation was always performed. Therefore, the location of sensor 7 was greatly affected by the external wind environment, and it was believed that the complexity of the time series data had been increased by the external wind occurring at irregular intervals (Fig. 4-7). For the same reason, the RMSE value at the location of sensor 8 was relatively high. The sensor 2 was located at exit point where the air in the greenhouse escapes to the outside. Given the wind direction of the warm air in the lab, and although R^2 is underestimated, the RMSE values in sensor 2 were similar to other sensor locations. R^2 was evaluated low, but the value of RMSE, which indicates the degree of error in the predicted value, was evaluated at a level similar to the result of other sensor locations.

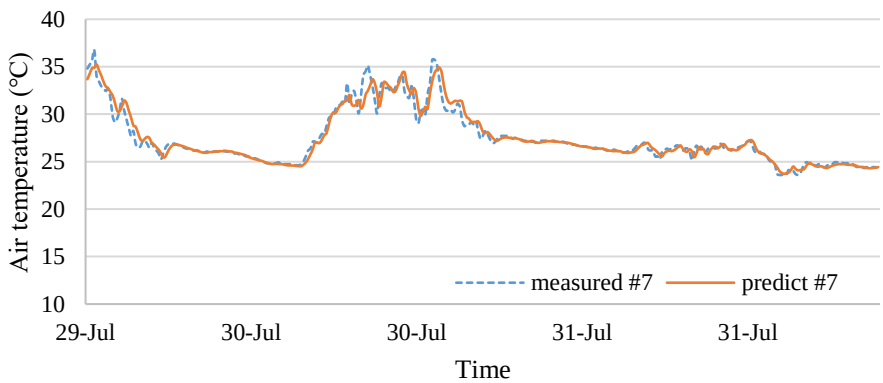
Table 4-4 Prediction accuracy (R^2, RMSE) of LSTM models in each sensor

location

Sensor location	#1	#2	#3	#4	#5	#6	#7	#8	#9
R^2	0.935	0.887	0.959	0.964	0.967	0.970	0.940	0.957	0.976
RMSE	0.607	0.643	0.592	0.638	0.640	0.579	0.701	0.755	0.624



(a) Time series prediction results for LSTM models : sensor #2



(b) Time series prediction results for LSTM models : sensor #7

Figure 4-10 Time series prediction results for LSTM models by sensor location (measured means collected data for experimental greenhouse, and predicted means of value predicted from the LSTM; #2 and #7 mean sensors 2 and 7, respectively)

4.3.4. Applying optimal sensor location to predicting air temperature inside naturally ventilated greenhouse

In this study, the future air temperature inside the naturally ventilated greenhouse was predicted by learning the internal environmental data and external weather data of the natural ventilation greenhouse measured in time series data. At this time, each sensor was required to predict the air temperature at nine locations inside the greenhouse. An important objective of this research was to reduce the initial installation and maintenance costs by minimizing the installation of sensors. Therefore, the future air temperatures inside the greenhouse were predicted using the optimal sensor location presented in the study of Chapter 3. Chapter 3 presented sensor 5 position, which was the center location of the greenhouse as the optimal sensor location of the experimental greenhouse, and thus predicted future air temperatures for nine points inside the greenhouse by learning environmental data measured from sensor 5 location. As a result, the prediction result of the future air temperature for each sensor location was shown in Table 4-5.

When sensor location 5, which was the optimal sensor location when using one sensor, was applied as learning data of the LSTM prediction model, the future air temperature prediction performance was relatively reduced at most sensor locations.

Although, most of the values of R^2 calculated was 0.9 or higher which indicates that the trend toward future temperatures was well predicted, the prediction results at sensors 1 and 2 still showed a big difference. However, this trend had improved as the number of optimal sensors increases. When two optimal sensors (#3 and #9) were applied to the LSTM prediction model, the average prediction accuracy for all sensor locations was R^2=0.915, RMSE=0.793, and when three optimal sensors were applied, the average prediction accuracy for all sensor locations was R^2=0.939, and RMSE=0.772. According to Chapter 3, when predicting the current air temperature of the naturally ventilated greenhouse, only one optimal sensor showed excellent prediction performance, but it was recommended to use at least three optimal sensor locations for future air temperature prediction.

Table 4-5 Predictive accuracy for each sensor location in the LSTM model with data from the optimal sensor location (#5)

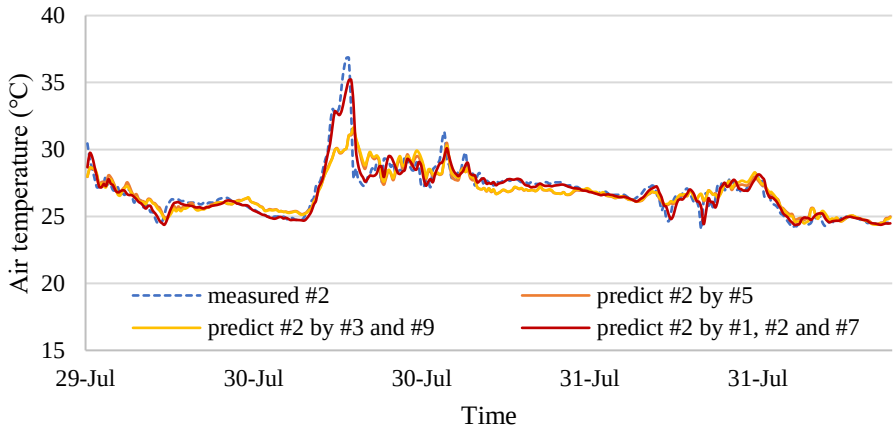
Sensor location	#1	#2	#3	#4	#5	#6	#7	#8	#9
R^2	0.832	0.563	0.879	0.932	0.974	0.953	0.929	0.890	0.960
RMSE	1.028	1.295	1.059	1.070	0.514	0.651	0.869	1.318	0.793

Table 4-6 Predictive accuracy for each sensor location in the LSTM model with data from the optimal sensor location (#3 and #9)

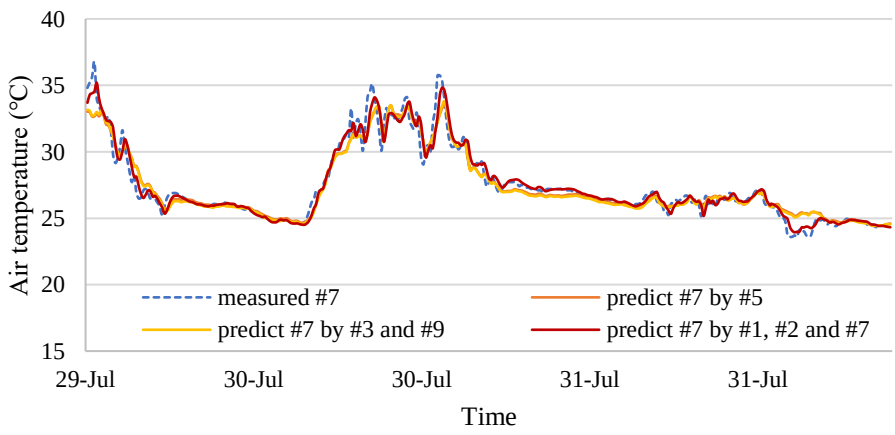
Sensor location	#1	#2	#3	#4	#5	#6	#7	#8	#9
R^2	0.905	0.714	0.954	0.946	0.958	0.960	0.928	0.899	0.976
RMSE	0.765	1.027	0.595	0.862	0.801	0.628	0.778	1.129	0.552

Table 4-7 Predictive accuracy for each sensor location in the LSTM model with data from the optimal sensor location (#1, #2 and #7)

Sensor location	#1	#2	#3	#4	#5	#6	#7	#8	#9
R^2	0.938	0.897	0.931	0.944	0.960	0.953	0.950	0.922	0.960
RMSE	0.602	0.611	0.766	0.854	0.914	0.869	0.627	0.969	0.739



(a) Prediction of future air temperature at sensor 2 location



(b) Prediction of future air temperature at sensor 7 location

Figure 4-11 Prediction of future air temperature at sensor locations 2 and 7 using LSTM with data from optimal sensor locations

4.4. Conclusion

In this study, the future air temperature for each location inside the naturally-ventilated greenhouse was predicted using the PFTO-ML using time series data. For this, environmental data such as air temperature, relative humidity, and soil temperature for each internal location were collected in the experimental greenhouse. The collected environmental data were interpolated with outliers and missing values through data preprocessing, and highly correlated data were selected as learning feature. A sequence length test was performed that affects the accuracy of the PFTO-ML that learned the time series data, and 30 min was calculated as an appropriate value. The sequence length test, which affects the accuracy of the PFTO-ML, was performed, and 30 min was evaluated as an appropriate value. As a result of predicting future air temperature by location inside the greenhouse using the finally developed LSTM model, high accuracy (R^2>0.95, RMSE<0.65) was obtained for most locations. In addition, for economical sensor installation, the optimal sensor location suggested by PCTO-ML was applied. The prediction of future air temperatures by location inside the greenhouse using data collected from optimal sensor locations showed that the prediction accuracy was reduced in most locations. However, the accuracy was improved when multiple optimal sensor locations were applied. The results showed a relatively large decrease in prediction accuracy when one optimal sensor was applied (R^2=0.918), but the prediction accuracy was similar to that of using all nine sensors (R^2=0.950) when three optimal sensors were applied (R^2=0.939). Therefore, it was recommended to apply at least three optimal sensor locations for future air temperature prediction.

Chapter 5. Ventilation Rate Prediction in Naturally Ventilated Greenhouses Using a CFD–Driven Machine Learning Model: Control

5.1. Introduction

Crop production through greenhouse cultivation has been increasing over several years by securing stable productivity and producing high-quality crops throughout the year. The greenhouse cultivation area and total production of vegetables in Korea have consistently increased since the 1970s and reached 82,810 ha and 2,312,000 tons in 2021 (MAFRA 2022). In particular, the area of single-span greenhouses is close to 86% of the total domestic greenhouse area, and most of single span greenhouse adopt natural ventilation (Lee, Lee et al. 2018). Ventilation in a greenhouse not only controls the internal air temperature rise in summer season by exchanging indoor and outdoor air, but also maintains appropriate environmental conditions such as proper humidity, carbon dioxide concentration, and constant air flow. Therefore, ventilation has a great influence on the growing environment of crops. If proper greenhouse ventilation is not performed in summer, crops are subjected to high-temperature injury and high humidity causes diseases and insect pests and the air flow near the leaves of the crop is insufficiently formed. As a result, the gas exchange of crop leaves is reduced, and the concentration of carbon dioxide decreases, leading to a decrease in the rate of photosynthesis, which leads to a decrease in crop productivity (Kitaya, Shibuya et al. 1998, Shibuya and Kozai 2001, Shibuya, Tsuruyama et al. 2006, Holcman and Sentelhas 2012, Radojevic, Bjelogrlic

et al. 2012, Kutta and Hubbart 2014, Marković, Pavlović et al. 2014). Therefore, it is important to maintain proper growth environment through ventilation. In particular, natural ventilation is very important for energy saving in farms because it requires no additional energy.

Field experiments concerning analysis of internal airflow in greenhouses including quantification of ventilation rates have already been given emphasis in many research works (Sase 1988, Fernandez and Bailey 1992, Kittas, Draoui et al. 1995, Boulard, Meneses et al. 1996, Wang, Boulard et al. 2000, Molina-Aiz, Valera et al. 2004, Mashonjowa, Ronsse et al. 2010, He, Chen et al. 2015). Most of the previous studies have used tracer decay method to measure airflow in greenhouses through field experiments and at the same time to quantify ventilation rates. However, although ventilation rate measurement research by field experiments is the most intuitive method, still, it faced with certain degree of challenge as it has to deal with the invisible air flow. Besides, air flow regulation is associated with several difficulties in addition to the limitations brought by the uncontrollable external wind environment.

Recent studies have used CFD simulations to evaluate the ventilation rate with the aim to overcome the aforementioned challenges and limitations concerning air flow measurement and regulation (Kacira, Sase et al. 2004, Baeza, Pérez-Parra et al. 2006, Campen 2006, Lee, Hong et al. 2006, Hong, Lee et al. 2008, Baeza, Pérez-Parra et al. 2009, Romero-Gómez, Choi et al. 2010, He, Chen et al. 2015, Benni, Tassinari et al. 2016, Akrami, Javadi et al. 2019, Villagran, Romero et al. 2019, Li, Li et al. 2020, Villagrán and Bojacá 2020, Park, Lee et al. 2022). (Park, Lee et al. 2022) evaluated the ventilation efficiency according to the eave height of the Korean venlo-type greenhouse. Simulations with the aid of CFD were used to analyze the temperature

at crop height and the airflow inside the greenhouse with external wind direction and speed factored in. (Li, Li et al. 2020) used CFD simulation to optimize the ventilation structure of an arch shape greenhouse. The study evaluated the natural ventilation rate of the structural elements such as ventilation structures and arch chord angles. (Hong, Lee et al. 2008) evaluated the TGD ventilation rate inside the greenhouse according to the structural type and natural climate factors such as wind direction of a multi-span greenhouse using CFD. Thus, CFD simulations were used to evaluate the ventilation rate of various variables and to optimize the ventilation structure. However, although research on ventilation using CFD can analyze the conditions desired by researchers, it requires specialized simulation knowledge and technology, high-end computer and significant amount of time for model computation. Therefore, studies using CFD are costly and time-intensive, and the number of simulation cases can be limited when considering model computation time.

On the other hand, with the recent development of big data and computer science, application of machine learning has been expanded. Machine learning refers to algorithms and statistical models that allow machines to learn on their own from data. Generally, it has high accuracy and requires relatively short computation time compared to CFD simulation. With this, in the field of agriculture, many studies have been conducted relevant to production prediction, environmental monitoring, animal welfare, disease detection, and so on (Sengupta and Lee 2014, Pantazi, Moshou et al. 2017, Ramos, Prieto et al. 2017, Çerçi and Daş 2019, Moon, Hong et al. 2019, Allouhi, Choab et al. 2021, Sujatha, Chatterjee et al. 2021, Fuentes, Viejo et al. 2022). Since the machine learning model learns the data measured in the field, the learning rate is low for conditions with low frequency of occurrence. That is, the predictive performance of the machine learning model sharply decreases for the untrained

condition. Hence, it is essential to acquire a sufficient range of training data for good prediction performance of machine learning models and application to various environmental conditions.

In this study, the prediction local ventilation rate CFD-driven machine learning model (PLV-CFD driven ML) was developed for predicting the local ventilation rate of a naturally-ventilated greenhouse by learning the results of CFD simulation. Data on the wind direction, wind speed, and vent opening of greenhouse were used for the calculation of the CFD models and for the generation of learning data for CFD-driven machine learning models. In this case, the verified CFD model of previous studies was used to generate reliable learning data (Lee, Lee et al. 2018). The tracer gas decay method was used to estimate the local ventilation rate of the greenhouse, and the TGD ventilation rate was calculated for 27 regions of crop height in the greenhouse. The training data generated from CFD simulations were used to train machine learning models such as multiple linear regression, support vector machine, random forest and deep neural network after data pre-processing. Each machine learning model has optimized hyper-parameters to improve prediction accuracy.

The CFD-driven machine learning models presented in this research require CFD results to generate learning data, but there are many difficulties in calculating numerous cases. Therefore, in order to supplement the limited number of training data, the prediction accuracy of the ML model by applying the bootstrapping technique was evaluated. Next, the minimum CFD case calculations for training data generation were evaluated by evaluating the prediction accuracy of the ML model by reducing the CFD case calculations for training data generation. As a result, the ML model with the highest prediction accuracy was defined as PLV-CFD driven ML.

5.2. Materials and Methods

Figure 5-1 shows the flowchart of this study. In this study, machine learning (ML) models were developed to predict the ventilation rate of a naturally ventilated greenhouse by zone. CFD simulation, which enables the analysis of arbitrary environmental conditions designed by researchers, was used to create training data for the ML models. To this end, case operations to create training data for the ML models were performed using a CFD model from a previous study (Lee et al.), which was verified by conducting the wind tunnel test and PIV test. The case operations of the CFD model were performed using the external wind speed, wind direction, and vent type in greenhouses as variables. In addition, the ventilation rate at the height of the crop group in the greenhouse was calculated by zone. The created CFD model simulation results were subjected to data preprocessing to be used as training data for ML models based on regression models. Multiple linear regression (MLR), support vector machine (SVM), random forest (RF), and deep neural network (DNN) models were developed as ML models to predict the ventilation rate of naturally-ventilated greenhouses, and hyper-parameter optimization was performed for each model to improve prediction accuracy. The bootstrapping technique was applied to the training data to supplement the limited number of CFD simulation cases. To minimize the number of CFD simulation cases, the prediction accuracy of the ML models was evaluated according to the training data reduction. Finally, the ML model with the highest prediction accuracy was defined as PLV-CFD driven ML.

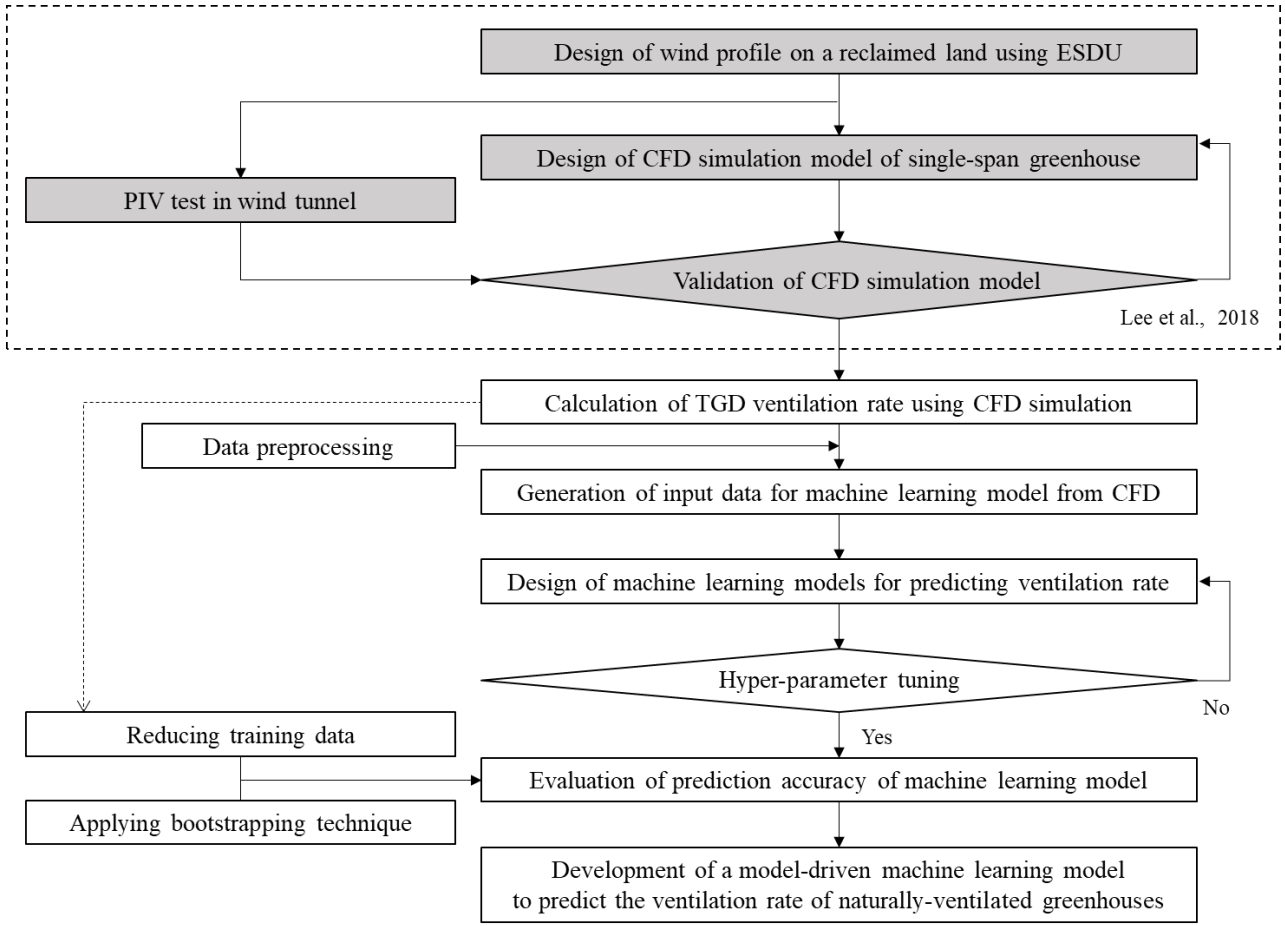


Figure 5-1 Flowchart of the experimental procedure

5.2.1. Target greenhouse

In this study, the naturally-ventilated even-span greenhouse (RDA 3-1 S Type, 2001) developed in Korea for dissemination to farms was selected as the target model. This model has been commonly used in Korea to overcome high temperature as a greenhouse model with the ventilation area ratio increased from 13 to 21% through structural improvement (RDA, 2014), and it has a structure that facilitates analysis according to various ventilation methods. Figure 5-2 shows the specifications and geometry of the greenhouse. The widths of the side and roof vents are 1.35 and 1.68 m, respectively, and a high natural ventilation rate can be expected due to the use of a large vent area. In this study, the natural ventilation rate of the target greenhouse was calculated for ventilation structures that use side vents, roof vents, and both side and roof vents. Internal crops were not considered to predict the natural ventilation rate according to the greenhouse structure as well as the external wind direction and wind speed.

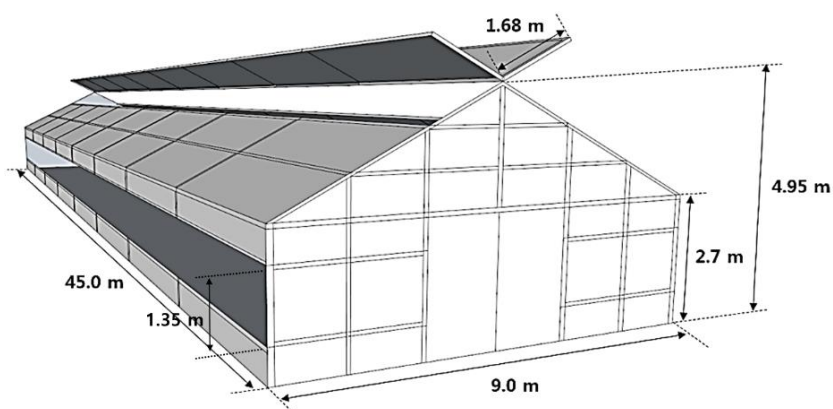


Figure 5-2 Schematic diagram of Even-span type greenhouse model for ventilation analysis (Lee et al., 2018)

5.2.2. Computational Fluid Dynamics

CFD is a numerical analysis tool that can interpret fluid flow, heat transfer, and chemical reactions in a system containing fluids through computer simulations. CFD uses Navier-Stokes equations, which are nonlinear differential equations, as governing equations and conducts numerical analysis using the finite difference method. The spatial domain is transformed into a small-volume mesh logarithmic equation, and a numerical algorithm is applied to qualitatively and quantitatively analyze the fluid flow phenomenon. It has been actively used in various fields, including machinery, aviation, chemical engineering, manufacturing, civil engineering, architecture, and environment, and active research has also been conducted in the agricultural field, such as livestock facility and greenhouse environment analysis (Bournet and Boulard 2010, Bjerg 2011, Hong, Lee et al. 2011, Ramponi and Blocken 2012, Boulard, Roy et al. 2017, Kim, Lee et al. 2022, Park, Lee et al. 2022, Yeo, Lee et al. 2022).

In this study, a three-dimensional (3D) grid network was designed using commercial CFD software (ver. 18.2, ANSYSInc, PA, USA) to analyze the local ventilation rate of naturally-ventilated even-span greenhouse. In addition, boundary conditions were set and computations were performed for the designed target space. Fluent, a numerical analysis software program for fluid flow in an analysis space designed with a two-dimensional or 3D grid network, was created based on mass, energy, and momentum conservation laws. The mass, momentum, and energy conservation equations used in calculations are as follows.

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \vec{v}) = S_m \quad \text{(Eq.5-1)}$$

$$\frac{\partial}{\partial t}(\rho \vec{v}) + \nabla \cdot (\rho \vec{v} \vec{v}) = -\nabla \mathrm{P} + \nabla \vec{\tau} + \rho \vec{g} + \vec{F} \quad \text{(Eq.5-2)}$$

$$\frac{\partial}{\partial t}(\rho h) + \nabla \cdot (\vec{v}(\rho h + P)) = \nabla \left(k_{eff} \nabla T - \sum_j h_j \overrightarrow{J_j} + (\bar{\tau} \vec{v}) \right) + S_h \quad \text{(Eq.5-3)}$$

Where, ρ is the density of the fluid (kg·m^{-3}), $\vec{v}$ is the flow velocity of the fluid (m·s^{-1}), P is the static pressure (Pa), $\vec{\tau}$ is the stress tensor (Pa), and $\vec{g}$ is the acceleration due to gravity. (m·s^{-2}), $\vec{F}$ is the external force (N·m^{-3}), S_m is the mass source term of the mass (kg·m^{-3}), k_{eff} is the effective conductivity (kg·m^{-2}·s^{-1}), T is the temperature (K), E is the specific enthalpy indicating the enthalpy per unit mass (J·kg^{-1}), t is the time (s), $\overrightarrow{J_i}$ is the diffusion flux of i type (kg·m^{-1}·s^{-1}), S_h is the enthalpy rise based on the chemical reaction or radiation (kg·m^{-1}·s^{-3}).

5.2.3. Machine learning models

5.2.3.1. Multiple Linear Regression

MLR is a linear regression model for predicting the dependent variable by modeling the linear relationship between several independent variables and one dependent variable. As it predicts the dependent variable using several features, better performance than conventional linear regression can be expected.

$$\mathrm{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \epsilon \quad \text{(Eq.5-4)}$$

where y is the dependent variable and x represents independent variables. β is the coefficient of each independent variable and ϵ is the residual.

MLR models have relatively faster learning speed and prediction than other ML

models, and they operate relatively well even with very large datasets and sparse datasets. They may, however, exhibit low prediction accuracy in some cases and are sensitive to outliers because the relationship between dependent and independent variables is limited to a linear relationship.

5.2.3.2. Support Vector Regression (SVR)

SVM is an ML model proposed by Vapnik in 1979. It is a classifier for finding decision boundaries that are as far apart as possible for objects in two different categories, and classifies classes while satisfying specific conditions. (Figure 5-3). Data and SVM training algorithms aim to get a hyperplane that divides the dataset into a predefined number of individual classes in a manner consistent with the training examples (Mountrakis, Im et al. 2011). It determines classes through decision boundaries, and it is a supervised learning model with powerful performance that can also be used for linear or nonlinear classification, regression, and outlier detection. It is particularly appropriate for complex classification and suitable for small or medium-sized datasets. SVR is a generalized method to predict random real values by introducing the ε-insensitivity loss function to SVM that is used for the classification prediction of training data (Vapnik 1999). The nonlinear expansion of SVR is performed by converting the original feature space that is not linearly separable to a new space with a higher dimension as with SVM. In other words, the SVR model enables prediction by using the kernel function and converting a low-dimensional nonlinear regression problem in the input space into a high-dimensional linear regression problem through mapping (Dibike, Velickov et al. 2001). SVR has been actively used in remote sensing, hydrology, agriculture, etc., due to its rapid and accurate prediction performance (Chevalier, Hoogenboom et al.

2011, Mountrakis, Im et al. 2011, Deka 2014, Ichii, Ueyama et al. 2017).

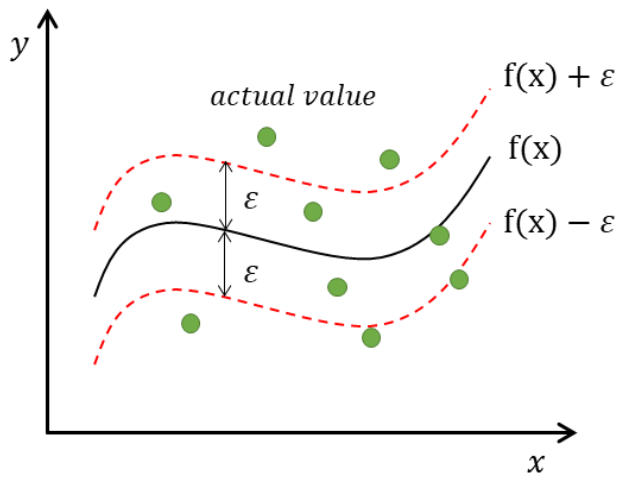


Figure 5-3 Conceptual diagram of support vector regression

5.2.3.3. Random forest

The RF algorithm creates several decision trees by randomly sampling n data from a given dataset and then determines the final prediction by a majority vote based on the prediction results of each decision tree. When each decision tree model is trained, the bagging method, which trains individual decision tree models with the dataset sampled from the entire train dataset by allowing overlapping, is used. Random forest uses the ensemble learning method, so it has the advantage of preventing the overfitting problem in which the accuracy drops sharply when other data other than the training data comes. RF has an algorithm suitable for regression analysis as well as classification. As the number of the decision trees generated in RF increases, the accuracy of the prediction result by a majority vote increases with excellent generalization performance. As the number of generated decision trees increases, however, the space required for analysis increases and higher performance is

required for analysis equipment. It also solves the problem of overfitting, but has the disadvantage of not explaining the process by which the results are derived. Random forest has the disadvantage of consuming a large amount of memory, but it is used in various fields as an advantage of being effective in large-capacity data processing (Philibert, Loyce et al. 2013, Abuella and Chowdhury 2017, Lin, Wu et al. 2017, Senagi, Jouandeau et al. 2017, Varma and Anand 2021).

$$ni_j = w_j C_j - w_{left(j)} C_{left(j)} - w_{right(j)} C_{right(j)} \qquad \text{(Eq.5-5)}$$

In a binary tree, in which a decision tree has two nodes, the nodes importance is calculated using Gini importance as show in eq.5-5. ni_j is the importance of node j, while w_j, c_j is weighted number of samples and the impurity value of node j.

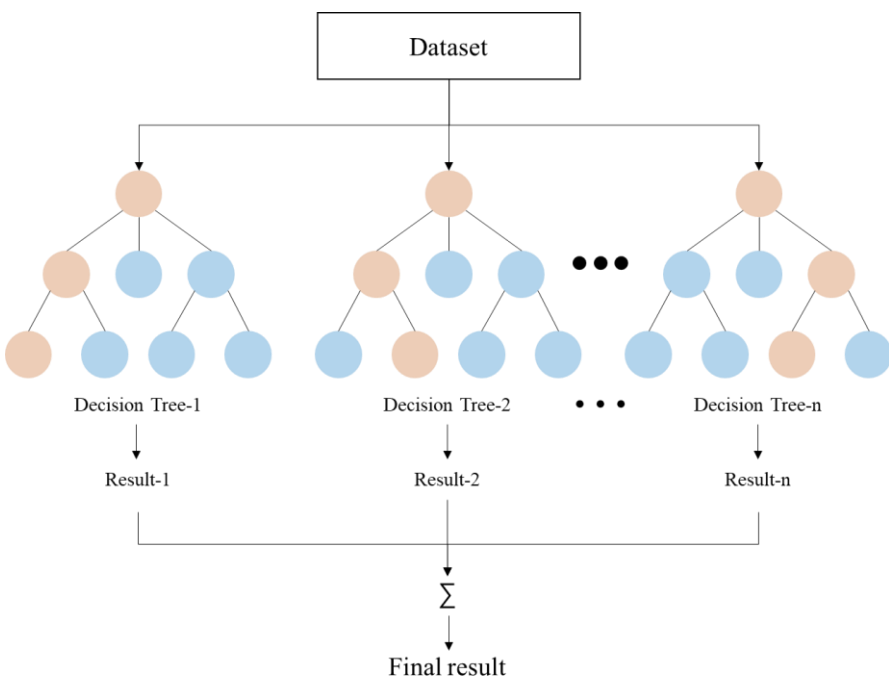


Figure 5-4 Conceptual diagram of random forest

5.2.3.4. Deep Neural Network

Artificial neural network has recently been used in many industrial fields, based on hypotheses and mathematical models of how the human brain solves given problems (Philemon, Ismail et al. 2019, Shahid, Rappon et al. 2019, Escamilla-García, Soto-Zarazúa et al. 2020, Abdolrasol, Hussain et al. 2021, Kola, Bojja et al. 2021, Li, Delpha et al. 2021, Lee, Lee et al. 2022). Deep neural network (DNN) is an artificial neural network (ANN) with a certain level of complexity that contains hidden layers between the input layer and output layer. ANN is an algorithm that imitates the way the human brain recognizes patterns. The algorithm makes it possible to classify various input data, interpret clusters, and recognize specific patterns from data. As with conventional ANN, DNN can model complex nonlinear relationships. A deep neural network can have a deeper network structure than a single neural network and can utilize more complex structures, so it can handle complex nonlinear relationships between input and output data more efficiently (Bengio, Courville et al. 2013). DNN can identify the potential structure of data because it can learn the features of different levels for each layer with multiple depths. Therefore, when appropriate structures are used in combination in the learning stage, higher performance can be expected compared to a neural network with a single structure. DNN can learn various nonlinear relationships as it includes multiple hidden layers, but it may cause such problems as overfitting, which increases the error for actual data, and high time complexity due to a large amount of computation for learning and excessive learning. Various studies have also been conducted in the agricultural field (Gorczyca, Milan et al. 2018, Choi, Park et al. 2021, Grimberg, Teitel et al. 2022).

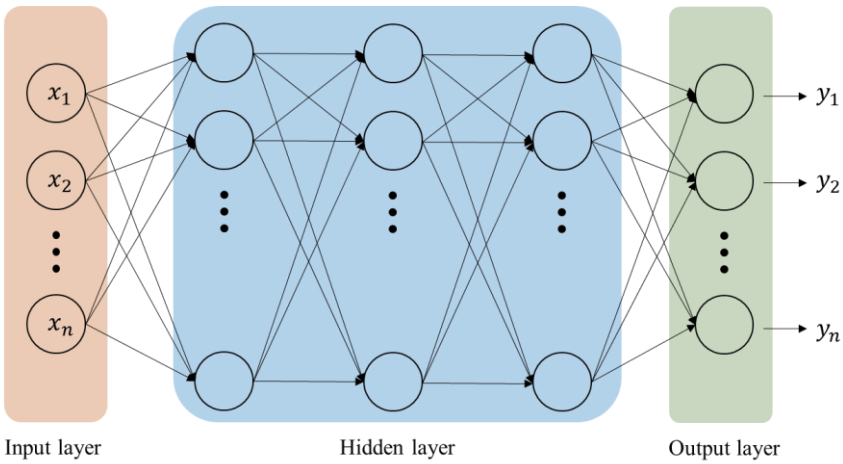


Figure 5-5 Conceptual diagram of deep neural network

In the stage of DNN, the activation function, a function for deciding each node to be activated or not, were variously developed such as sigmoid, tanh, ELU and etc. The most commonly used activation function is RELU, and the equation is as follows:

$$\mathrm{f(x)} = \begin{cases} 0 \ (x < 0) \\ x \ (x \geq 0) \end{cases} \qquad \text{(Eq.5-6)}$$

5.2.4. Estimation method of natural ventilation rate

In order to prompt an appropriate growing environment in the greenhouse, it is essential to consider the appropriate air conditioning system design, and for that purpose it is important to quantitatively evaluate ventilation. In most studies on ventilation analysis, the amount of air exchange inside the target facility has been calculated with respect to the volume replacement based on the mass conservation law. The equation to calculate the ventilation rate is as follows:

$$\mathrm{AER_{MFR}} = \frac{v_i A_i}{V} \times 60 = \frac{v_0 A_0}{V} \times 60 \qquad \text{(Eq. 5-7)}$$

where AER_{MFR} is the air exchange rate of MFR per minute ($AER \cdot min^{-1}$), $v_{i,0}$ are the air velocity of inflow and outflow, $A_{i,0}$ are the area of the inflow and outflow vents (m^2) and V is the volume of the greenhouse (m^3). In this case, a value for the entire facility is presented, but the local ventilation effect cannot be considered. In addition, when calculation is performed based on the mass conservation law, the fact that ventilation performance varies depending on the characteristics and geometry of inflow and outflow vents even for facilities with the same volume cannot be reflected.

To supplement these shortcomings, the ventilation rate calculation method by the tracer gas concentration method, which can calculate overall and local ventilation rates by injecting tracer gas into the target facility and calculating real-time changes in tracer gas concentration, has been used. The trace gas concentration decay method is an experimental method to calculate the ventilation rate by quantitatively analyzing the trace gas concentration decrease in the facility with time from the start of ventilation after uniformly filling the greenhouse with trace gas. It can consider not only the external wind environment but also the structural characteristics of the greenhouse, the ventilation form, etc., and can analyze the local ventilation rate inside the greenhouse. It is mainly classified into the following methods depending on the tracer gas injection method: the step-up method in which tracer gas is injected with the start of ventilation until a certain concentration (i.e. the time that can be assumed as a steady state) is reached, the step-down method in which ventilation is started when the injection of tracer gas is complete and the degree of attenuation of the tracer gas concentration is used, and the pulse method in which tracer gas is injected into the target area where ventilation is performed within a short period of time and then injection is interrupted again (Sandberg and Sjöberg 1983). After determining the tracer gas injection method, the age of air can be obtained by

creating a time-concentration trend curve and calculating the first moment of the concentration curve based on the tracer gas data measured during the experiment. Depending on the tracer gas injection method, the local average age and local residual residence time at any point P in indoor space can be calculated using the step-up method, step-down method, and pulse method. In this study, the step-up method was used as follows.

$$\mathrm{AER_{TGD}} = \int_0^{\infty} (1 - \frac{C_p^{sup}(t)}{C(\infty)})\,\mathrm{dt} \qquad \text{(Eq. 5-8)}$$

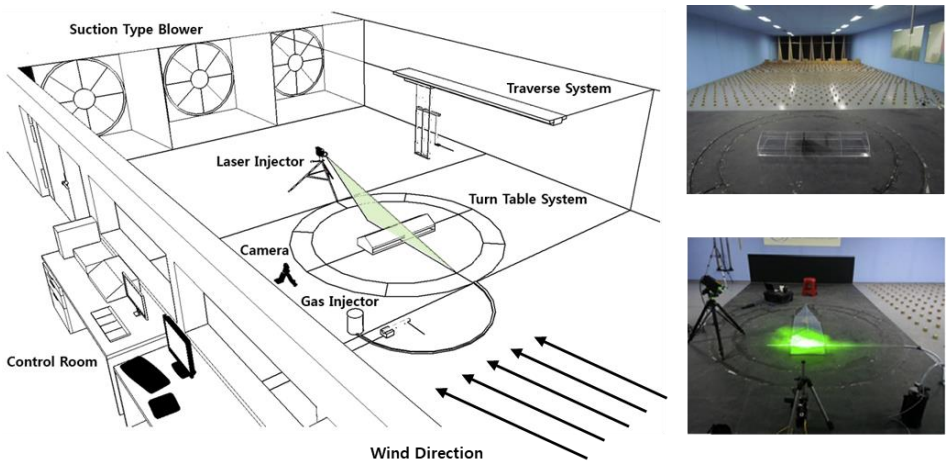
where, $\mathrm{AER_{TGD}}$ is air exchange rate computed using the TGD method (AER·min^{-1}) C is concentration of tracer gas (ppm), $\mathrm{C}^{sup}(\mathrm{t})$ is the tracer gas concentration at time t, and $\mathrm{C}(\infty)$ is the concentration at the exhaust vent after reaching the steady state.

5.2.5. Experimental procedure

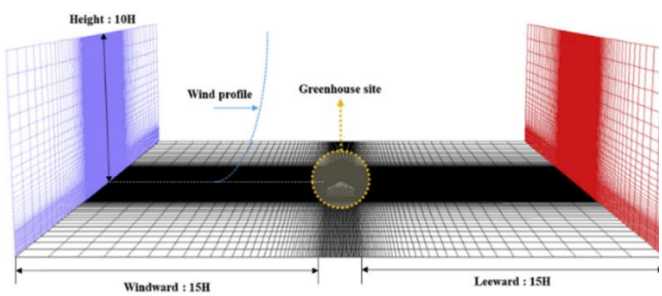
5.2.5.1. Training data generation using CFD model verified from previous research (Lee et al., 2018)

In this study, ML models were developed to predict the natural ventilation rate of naturally-ventilated even-span greenhouses, and training data for the ML models were generated through CFD simulation. A CFD simulation model enables model computations and analysis under the conditions desired by the researcher, but special efforts are required to secure the reliability of the model. The grid design of a CFD model can affect the accuracy of the simulation results (Rong, Nielsen et al. 2016, Hong, Exadaktylos et al. 2017, Yeo, Decano-Valentin et al. 2020), and selection of a

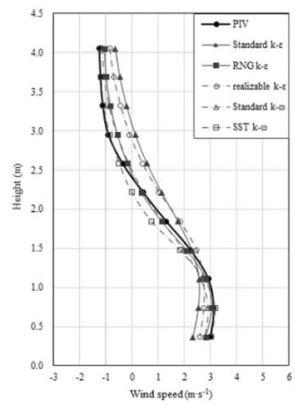
turbulence model also significantly affects the model results (Norton, Sun et al. 2007, Ramponi and Blocken 2012, Bjerg, Norton et al. 2013, Kwon, Lee et al. 2016). Therefore, it is very important to select a proper grid size and a turbulence model. Since the natural ventilation of a greenhouse is directly affected by the external wind environment, consideration of the external wind environment is essential for natural ventilation analysis and the proper design of the flow area outside the greenhouse is required (Bournet, Khaoua et al. 2007, Kim, Lee et al. 2017, Lee, Lee et al. 2018). Therefore, in this study, the CFD model of (Lee, Lee et al. 2018) which secured sufficient reliability for the analysis of natural ventilation in greenhouses through proper grid and turbulence model design, external wind environment consideration, and external flow area design, was used (Figure 5-6). The information on CFD domain design, grid design, turbulence model, and boundary conditions was used as shown in Table 5-1.



(a) Overall view of the PIV test in the wind tunnel



(b) CFD simulation design of the external domain and greenhouse



(c) Validation results of turbulence model

Figure 5- 6 Design and validation process of previous CFD models for estimating natural ventilation rates in greenhouses; PIV test, external zone design and turbulence model validation results (Lee et al., 2018)

Table 5-1 Design factors of the naturally ventilated greenhouse CFD model (Lee et al. 2018)

Design factor	Value
Turbulence model	RNG $\mathbf{k-\epsilon}$
Grid size	0.2 m
Solver	Pressure-based solver
Numerical algorithm	SIMPLE algorithm
Discretization	Second-order
Time condition	Steady-stated, and transient state
Dimension	Tree-dimensional simulation
Operating pressure	101,325 Pa
Gravitational acceleration	9.81 m·s-1
Air density	1.225 kg·m-3
Air viscosity	1.7894×10-5 kg·m-1·s-1

The CFD technique can simulate arbitrary environmental conditions and calculate results for arbitrary physical quantities. In the case of such physical quantities as the amount of natural ventilation, in particular, it is difficult to measure stable values due to wind environment conditions that vary in real time during field experiments. Therefore, the CFD technique that can calculate the amount of natural ventilation under arbitrary environmental conditions, such as the wind direction and wind speed, was utilized to construct training data for ML. Simulation was performed on important influence factors, i.e. the wind direction (WD), wind speed (WS), and vent opening conditions of the target greenhouse, to predict the amount of natural ventilation. Seven cases (0, 15, 30, 45, 60, 75, and 90°) were analyzed for the wind direction and ten cases (1,.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0, and 5.5 $m \cdot s^{-1}$) for the wind speed. In the case of vent opening conditions of the greenhouse, three cases

(a case of opening only roof vents, a case of opening only side vents, and a case of opening both side and roof vents) were analyzed. The mass flow rate (MFR) for the total amount of ventilation in the greenhouse was calculated through simulation, and the TGD-computed ventilation rate was calculated for the ventilation rate by zone. As for the ventilation rate by zone, the greenhouse was divided into 27 areas with a width of 3 m and a length of 5 m and the ventilation rate by zone was calculated at the height of the crop group (0 to 1 m) (Figure 5-7). Coordinate values with respect to the center of the greenhouse as the origin were provided to each calculated local ventilation rate. 80% of the training data were used for the development of ML models and 20% for the verification of the developed models.

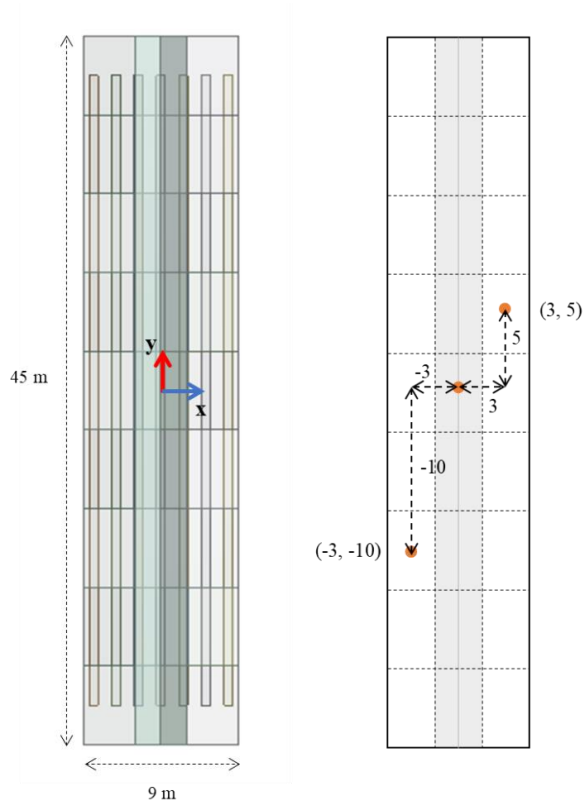


Figure 5-7 Defining the compartment design and x, y-coordinates inside the greenhouse to calculate the local ventilation rate computed by TGD method

5.2.5.2. Pre-processing for dataset of machine learning models

In this study, training data for ML models that predict the ventilation rate of a naturally-ventilated greenhouse were created through CFD simulation. Among the CFD simulation results, the factors considered as learning features for ML models were the external wind direction, wind speed, vent type, MFR, TGD, and x and y coordinates. Here, the x and y coordinates were given in the width direction and the length direction with reference to the center of the greenhouse in order to reflect the positional information of the local ventilation rate computed by TGD method. The coordinate means the center position of each section. Meanwhile, the types of the data used for the training of ML models can be divided into numerical data that have discreet or continuous values and categorical data that have such values as names and sequence. Numerical data can be used for regression analysis as they have numerical values while categorical data can be used for classification models. For utilizing categorical data in a regression model, preprocessing that assign numbers to variables with the meaning of letters is required. One-Hot encoding, one of the techniques to process letters as numbers for natural language processing, is a data preprocessing technique that can be applied to only categorical data. It expresses n categorical data as n sparse vectors by adding a new feature according to the type of the feature value, assigning 1 only to the column corresponding to the eigenvalue, and assigning 0 to the remaining columns. In this study, the feature for the vent type of the naturally-ventilated greenhouse has values that correspond to the opening of side vents, roof vents, and both side and roof vents and these values correspond to categorical variables. Therefore, label encoding was performed for the use of categorical data in ML models, and values of 1, 2, and 3 were assigned to the opening

of side vents, the opening of both side and roof vents, and the opening of roof vents (Table 5-2). Since this cannot be applied to ML models designed based on a regression model, however, One-Hot encoding was performed as a data preprocessing process for this and the results are shown in Table 5-3.

Table 5-2 Label encoding applied data preprocessing results for use in regression ML models of categorical data (WD: wind direction; WS: wind speed; WT: vent opening type of greenhouse such as side open(1), both side and roof open (2) and roof open(3); x, y-coordinate: x, y coordinates of TGD which is the local ventilation rate; MFR: ventilation rate computed MFR method; TGD: ventilation rate computed TGD method)

WD	WS	WT	x-coordinate	y-coordinate	MFR	TGD
0	1.0	1	-3	-20	0.13	0.23
0	1.0	2	-3	-20	0.38	0.25
0	1.0	3	-3	-20	0.29	0.23
0	1.5	1	-3	-20	0.44	0.33
⋮	⋮	⋮	⋮	⋮	⋮	⋮
90	5.5	1	3	20	6.73	0.67
90	5.5	2	3	20	14.46	8.99
90	5.5	3	3	20	6.98	3.09

Table 5-3 One-hot encoding applied data preprocessing results for use in regression ML models of categorical data

WD	WS	Type Window _Side	Type Window _SideRoof	Type Window _Roof	x-coordinate	y-coordinate	MFR	TGD
0	1.0	1	0	0	-3	-20	0.13	0.23
0	1.0	0	1	0	-3	-20	0.38	0.25
0	1.0	0	0	1	-3	-20	0.29	0.23
0	1.5	1	0	0	-3	-20	0.44	0.33
⋮	⋮			⋮	⋮	⋮	⋮	⋮
90	5.5	1	0	0	3	20	6.73	0.67
90	5.5	0	1	0	3	20	14.46	8.99
90	5.5	0	0	1	3	20	6.98	3.09

5.2.5.3. Design of machine learning models

The prediction accuracy of an ML model is significantly affected by the design of the model's hyper-parameters. There is no absolutely best value for hyper-parameters and the optimization method is not determined, but the optimal hyper-parameter values can be found through iterative design and trial and error (Reimers and Gurevych 2017, Raschka 2018). Therefore, the hyper-parameters of the SVR, RF, and DNN models developed in this study were optimized to improve the prediction accuracy of the models. First, the main hyper-parameters of the SVR model are techniques that enable linear classification of nonlinear data, including the kernel function that performs linear classification by mapping low-dimensional data to a higher dimension, degree that determines the order of the polynomial kernel function, gamma that determines the curvature of decision boundaries, and C that determines the degree of allowing data samples to be placed in different classes. In this instance, overfitting may occur if the values of gamma and C are set to be excessively large. In this study, rbf and poly were considered as the kernel functions of SVR, and hyper-parameter tuning was performed for the values of C and gamma. The main hyper-parameters of the DNN model are each layer and node that constitute the neural network, the loss function that evaluates the performance of the model, the optimizer that is a function to reduce the loss function, and the activation function that applies the input values to each node of DNN for output to the next layer. In this study, the performance of the model was evaluated using MSE for Loss function, and Adam was used as Optimizer for efficient learning. relu, which is commonly used due to fast learning and low computation cost, was used as the activation function. For an improvement in the performance of the DNN model, an

attempt was made to derive the optimal neural network structure by tuning the number of layers and the number of nodes. The main hyper-parameters of the RF model are n-estimators, which is the number of decision trees, and min_samples_split, which is the minimum number of sample data for splitting nodes. In this study, the tuning of the two hyper-parameters was performed to improve the prediction accuracy of the RF model.

Table 5-4 Hyper-parameters of machine learning models and tuning range design for hyper-parameter optimization

Model	Hyper-parameter	Range
SVR (kernel = rbf)	C	10^n (n=0-4)
SVR (kernel = poly)	C	10^n (n=0-4)
	gamma	10^{-n} (n=0-5)
	degree	2
RF	n-estimators	5-11
	min samples split	2-6
DNN	layer node	2^n (n=1-7)
	number of layer	1-4
	loss function	MSE
	activation function	relu
	optimizer	adam

5.2.5.4. Bootstrapping training data

In this study, data were created using CFD simulation to calculate the ventilation rate of the naturally-ventilated greenhouse, which is the training data for the ML models. CFD simulation enables computations and analysis under the environmental conditions desired by researchers. Therefore, when the training data of an ML model

are created using CFD simulation, the limitation of the ML model that the prediction performance decreases for data that exceeded the range of the training data can be supplemented. The computations of the CFD simulation model, however, may involve an increase in the number of grids or an increase in the complexity of model computations to improve the accuracy of numerical analysis, which requires high-performance computation equipment and long computation time. Since absolute time is required for simulation model computations, a limited number of computations are possible in reality. Therefore, it is important to create an appropriate number of training data for training the ML model.

Meanwhile, bootstrapping, a technique used in statistics, is a method of applying random sampling before hypodissertation testing or the calculation of evaluation indicators (Eforn 1979). In other words, it is a procedure to obtain statistics and construct a model using the samples generated from given samples through random sampling. The average of N data randomly sampled from the existing data is created as new data, and this process is repeated several time to amplify the existing data M times. Assuming that separate samples are $\mathrm{y_n} = (y_1, y_2, y_n)$, the resampled units of $\mathrm{y_n^*} = (y_1^*, y_2^*, y_n^*)$ can be obtained in large quantities using the Monte Carlo bootstrap approximation for the mean and variance of the given parameters. Therefore, in this study, the bootstrapping technique was used to supplement the generation of training data using CFD simulation, which enables only a limited number of case operations. For the data generated by CFD simulation, ten data were randomly extracted, and the existing data were amplified by 2, 3, 5, and 10 times to select the optimal number of restoration extractions.

5.2.5.5. Reducing data set

CFD simulation can arbitrarily adjust the range of the training data of an ML model because it enables the simulation and analysis of the model under the environmental conditions desired by the researcher. Depending on the simulation model design, however, computation time from several hours to several days can be required even for one case. Therefore, it is important to design the minimum number of simulation model case operations that secures the appropriate prediction accuracy of the ML model. Therefore, in this study, the prediction accuracy of the developed ML models was evaluated while the number of CFD simulation cases was reduced (Table 5-5). Under the condition of maintaining the maximum and minimum ranges for the external wind direction and wind speed, which were considered as CFD simulation cases, a reduction in the number of cases was considered. For the external wind speed, ten cases (1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0, and 5.5 $m \cdot s^{-1}$) were considered in the existing model while five cases (1.0, 2.0, 3.0, 4.0, and 5.5 $m \cdot s^{-1}$) were considered in the reduced model. In the case of the external wind direction, seven cases (0, 15, 30, 45, 60, 75, and 90°) were considered in the existing model while four cases (0, 30, 60, and 90°) were considered in the reduced model.

Table 5-5 Reduction of CFD-driven learning data for designing a simple machine learning model for predicting local ventilation rate in naturally ventilated greenhouses (WS : wind speed; WD : wind direction; WT : vent opening type of greenhouse)

CFD simulation cases for training data of simple machine learning models	Total case
WS: 10 case (1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0, 5.5 $m \cdot s^{-1}$) WD: 7 case (0, 15, 30, 45, 60, 75, 90°) WT: 3 case (Side, roof, side and roof vents)	210
WS: 10 case (1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0, 5.5 $m \cdot s^{-1}$) WD: 4 case (0, 30, 60, 90°) WT: 3 case (Side, roof, side and roof vents)	120
WS: 5 case (1.0, 2.0, 3.0, 4.0, 5.5 $m \cdot s^{-1}$) WD: 7 case (0, 15, 30, 45, 60, 75, 90°) WT: 3 case (Side, roof, side and roof vents)	105
WS: 5 case (1.0, 2.0, 3.0, 4.0, 5.5 $m \cdot s^{-1}$) WD: 4 case (0, 30, 60, 90°) WT: 3 case (Side, roof, side and roof vents)	60

5.3. Results and discussion

5.3.1. Dataset from CFD simulation results

CFD simulation was performed to develop ML models for predicting the ventilation rate of the naturally-ventilated greenhouse by zone. The ventilation rate computed by the TGD method and the ventilation rate computed by the MFR method were calculated for ten external wind speed cases ranging from 1.0 to 5.5 $m \cdot s^{-1}$, seven external wind direction cases ranging from 0 to 90°, and three greenhouse vent type cases. Their averages were found to linearly increase with the external wind speed. The MFR-computed ventilation rate was calculated to be between 1.17 and 6.46 $AER \cdot min^{-1}$ according to the external wind speed and the TGD-computed ventilation rate ranged from 0.45 to 4.36 $AER \cdot min^{-1}$. Since the MFR-computed ventilation rate considers only the flow rates at the inflow and outflow vents of the greenhouse and the TGD-computed ventilation rate considers the flow inside the greenhouse, the former is generally calculated to be higher than the latter under the same environmental conditions (Lee, Lee et al. 2018).

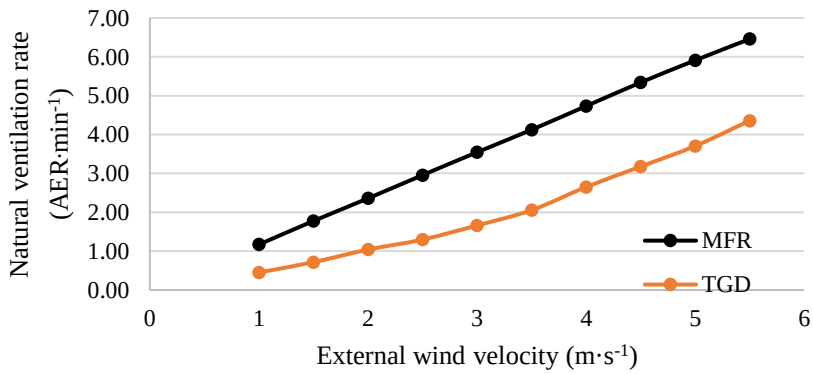


Figure 5-8 Natural ventilation rate based on the MFR and TGD method according to external wind velocity

Figure 5-9 showed the airflow distribution at a height of 1m inside the greenhouse according to the external wind direction when only the side windows are open and the outside wind speed is 5.5 m·s^{-1}. In the case of 0 degree wind parallel to the length of the greenhouse, the external wind blowing parallel to the side windows of the greenhouse did not flow into the greenhouse. As a result, the overall air flow was low inside the greenhouse.

When the external wind direction was 45 degrees, the strongest flow velocity was formed at the end of the greenhouse where the external air entered, and the relatively low flow velocity was formed at the opposite greenhouse end. When the outside wind direction was 90 degrees perpendicular to the side windows of the greenhouse, a uniform and strong airflow was formed throughout the greenhouse. Due to the even distribution of wind to the side windows of the greenhouse, it is expected that the local ventilation rate of the greenhouse will also be calculated equally.

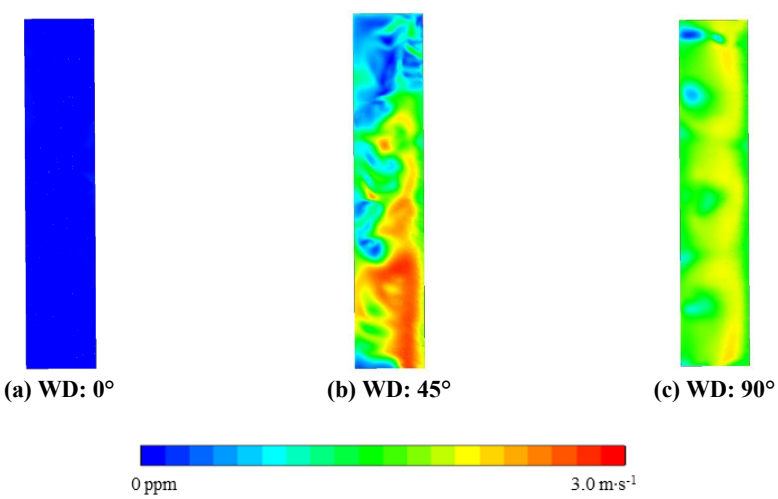
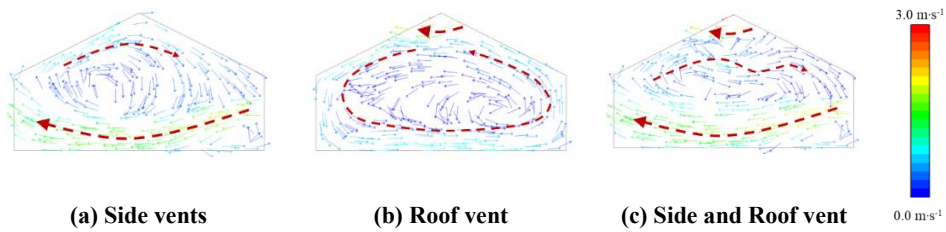


Figure 5-9 CFD computed airflow distribution in greenhouse with side vents opening according to wind direction (WD) when the external wind speed was 5.5 m·s^{-1}

Figure 5-10 shows the airflow inside the greenhouse according to the vent type when the external wind speed was 2.5 $m \cdot s^{-1}$. When ventilation is performed through side vents, the airflow introduced from the windward side forms the main airflow in the lower part of the greenhouse and is discharged through the vent on the opposite side. A clockwise backward flow is observed in the upper part of the greenhouse, but it is relatively weak compared to the main airflow. When ventilation is performed through roof vents, the air introduced from a roof vent is discharged through the opposite vent, and some of the introduced air forms weak airflow in the counterclockwise direction inside the greenhouse. Finally, when ventilation is performed through both side and roof vents, the airflow introduced from the windward side forms the main airflow in the lower part of the greenhouse as with ventilation through side vents, and the air introduced from a roof vent forms airflow that presses the counterclockwise airflow formed in the upper part of the greenhouse.



(a) Side vents **(b) Roof vent** **(c) Side and Roof vent**

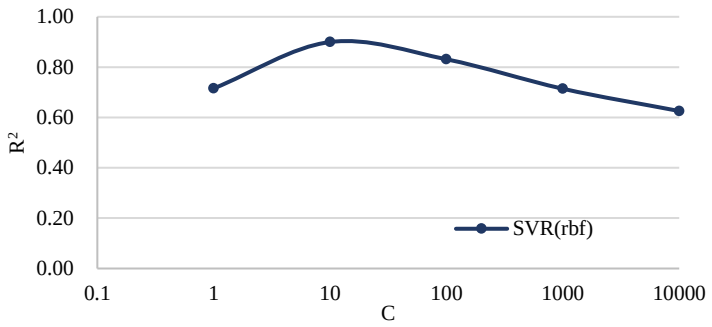
Figure 5-10 Airflow inside the greenhouse according to the vents opening when the external wind speed was 2.5m·s^{-1}

5.3.2. Hyper-parameter optimization of machine learning models

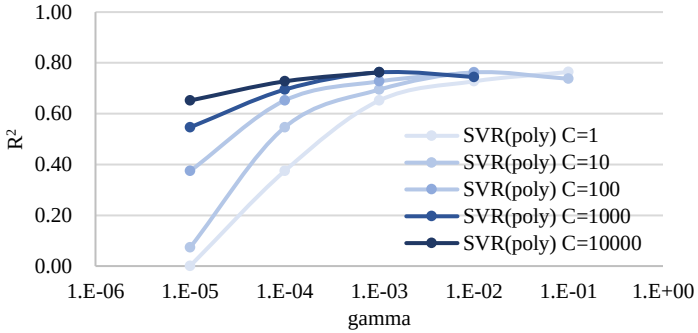
To improve the prediction accuracy of the ML models, the hyper-parameters of each ML model were optimized through the trial-and-error method (Table 5-6). The

kernel function of the SVR model enables linear classification by mapping low-dimensional data to a higher dimension. Thus, an appropriate kernel function must be selected according to the training data. Therefore, in this study, the RBF and poly functions, which are the representative kernel functions of SVR, were used and the hyper-parameters for each were optimized. As for C and gamma, which are used in the two kernel functions, C means the degree of tolerance for the error and gamma determines the flexibility of decision boundaries. In general, the complexity of the algorithm increases as these two values increase and the complexity decreases as the value decrease. First, when the rbf kernel function was used and the prediction accuracy for the local ventilation rate of the naturally-ventilated greenhouse according to the C value was evaluated, the highest prediction accuracy of R^2=0.90 and RMSE = 0.643 AER·min^{-1} was observed at C=10 (Figure 5-11(a)). When the poly kernel function was used, the prediction accuracy showed a tendency to increase as the values of C and gamma increased. The highest prediction accuracy of R^2=0.76 and RMSE = 0.963 AER·min^{-1} was observed at C=10000 and gamma = 10^{-3}. The RF model is an ensemble model that performs prediction by training multiple decision trees and analyzing the results. The hyper-parameters of the model are n-estimator, which is the number of decision trees, and min samples split, which is the minimum number of sample data for splitting nodes. The tuning results for the two hyper-parameters exhibited high accuracy for most design values (R^2 > 0.93, RMSE < 0.51 AER·min^{-1}), and the hyper-parameter values that showed the highest accuracy were n-estimator = 11 and min sample split = 2 (R^2=0.95, RMSE = 0.43 AER·min^{-1}). Finally, the prediction accuracy of DNN was evaluated according to the number of layers and the node design of each layer. The accuracy was found to increase as the number of layers increased, and the highest accuracy was observed when the layer

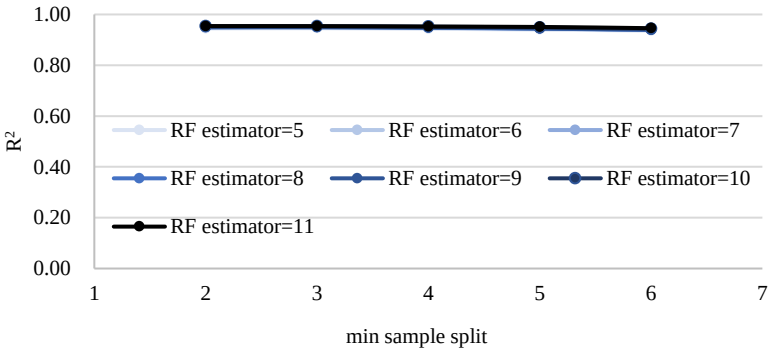
nodes from the first layer to the last layer were set to 64, 128, 128, and 1 at Layer = 4 (R^2 = 985, RMSE = 0.242 AER·min^{-1}).



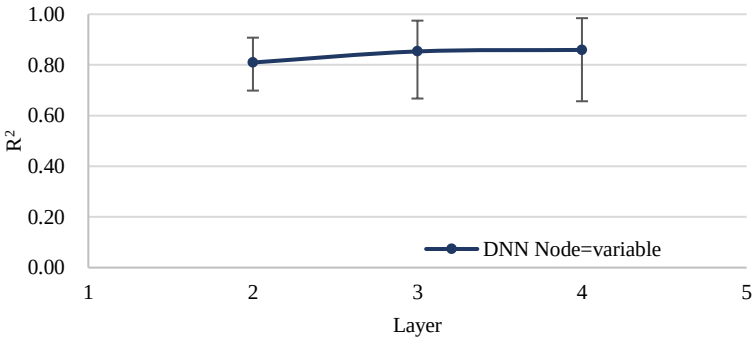
(a) SVR (rbf)



(b) SVR (poly)



(c) RF



(d) DNN

Figure 5-11 Prediction accuracy (R^2) results according to hyper-parameter tuning test for each machine learning mode

Table 5-6 Hyper-parameter optimization values for each machine learning model

Model	Target hyper-parameter	Optimal hyper-parameter
SVR (kernel = rbf)	C	10
SVR (kernel = poly)	C	10000
	gamma	10^{-3}
RF	n-estimators	11
	min samples split	2
DNN	layer node	64 / 128 / 128 / 1
	Number of layer	4

5.3.3. Evaluation of prediction accuracy of machine learning model by applying bootstrapping

The proposed ML models that use the CFD simulation results as training data can artificially adjust their learning range because CFD simulation can be performed under arbitrary environmental conditions set by the researcher. The creation of training data, however, is limited due to the CFD case simulation time. Therefore, the bootstrapping technique, which can supplement the number of training data by regenerating samples from the given samples for the CFD simulation results through random sampling, was applied.

When the bootstrapping technique was applied to each ML model, R^2 decreased for the RF and DNN models as well as the SVR model that uses the rbf kernel function, which exhibited high prediction performance (R^2>0.9) before the application of the technique, and it increased for the MLR model and the SVR model that uses the poly kernel function, which showed relatively low prediction performance (R^2<0.9). This appears to be because the models that exhibited

sufficiently high prediction accuracy showed overfitting for a relatively small number of training data but the overfitting was mitigated through the application of the bootstrapping technique. The RMSE value was found to decrease for all of the ML models except for the DNN model. While the RMSE value significantly decreased by more than 0.2 for the SVR model that uses the poly kernel function and MLR, there was no significant change in RMSE for the RF model that had already applied the ensemble model technique.

On the other hand, in the case of the remaining ML models except for the MLR model, there was no significant improvement when 10 times restoration extraction was performed. A lot of restoration extraction has the disadvantage that the complexity of the model increases and the bias of prediction may increase. Therefore, in this study, five times was suggested as the number of appropriate restoration extractions.

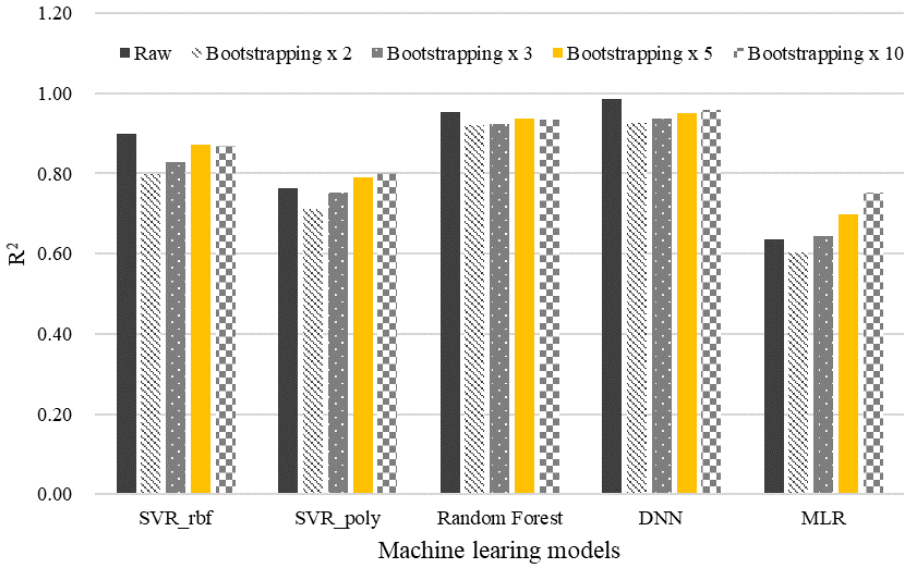


Figure 5-12 R^2 calculation results for each machine learning model by applying bootstrapping

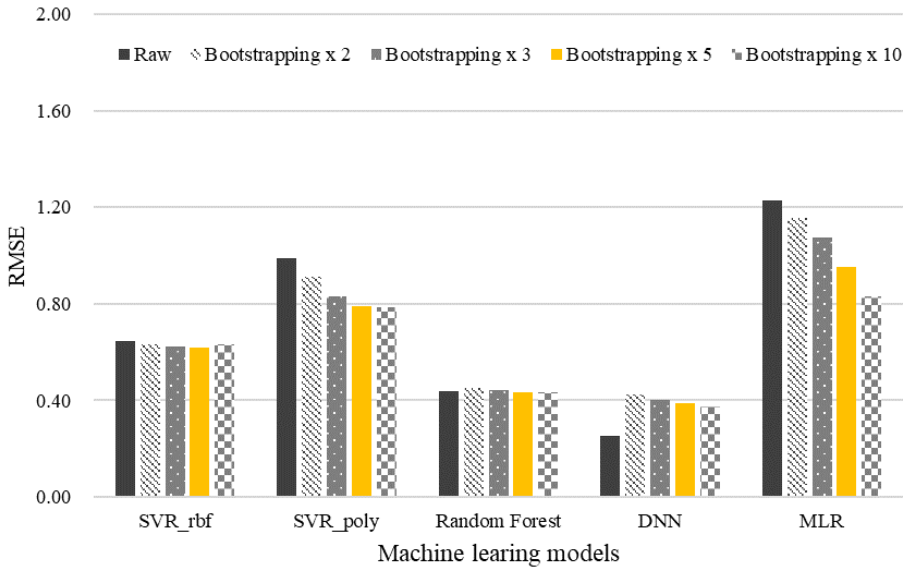


Figure 5-13 RMSE calculation results for each machine learning model by applying bootstrapping

5.3.4. Optimize the machine learning models for reducing data set

In this study, CFD-driven ML models were developed to predict the local ventilation rate of a naturally-ventilated greenhouse. To this end, CFD case operations were performed for ten external wind speed cases, seven external wind direction cases, and three greenhouse vent type cases. Since computations for 210 cases are required, however, much time is needed to construct training data for the ML models. It took about 12 hr using a computational computer (Intel(R) Xeon(R) CPU E5-2620 v3 @ 2.40GHz, RAM 32.0 GB) to calculate the CFD model of 1 case, and it took 105 days to calculate when using one computational computer. Therefore, in this study, the prediction accuracy of the CFD-driven ML models was evaluated by reducing the number of cases for the external wind speed and wind direction (Figure 5-14). Consequently, most of the ML models exhibited the highest accuracy for the training data created from 120 cases that considered only four wind direction cases. The error for the predicted value was also found to be lowest because the RMSE value was evaluated to be low. The DNN and RF models exhibited high R^2 and the change in R^2 depending on the number of CFD simulation cases was not relatively significant, but the RMSE value decreased by 0.07 to 0.13 and 0.11 to 0.16 AER·min^{-1} depending on the number of cases, respectively. All of the SVR, RF, and DNN models showed low R^2 values and high RMSE values for the training data created from 210 cases that considered ten wind speed cases and seven wind direction cases. This indicates that the prediction accuracy of the ML models decreased when all cases were considered because the model complexity relatively increased. Finally, the RF model trained on 120 cases of CFD results showed the

highest accuracy. This was because the RF model was an Ensemble model, which was a method of deriving the final result by synthesizing the results of various decision tree algorithms. In addition, the RF model did not require normalization of data, and bootstrap were applied by default. As a result, RF model has been generally known to have high performance and minimize the overfitting problem. Therefore, it was judged that the RF model showed the highest accuracy as the final model for CFD-driven ML with a relatively small amount of data and the risk of overfitting. However, when building a PLV-CFD driven ML model for other facility, careful consideration of CFD case for generating learning data should be made.

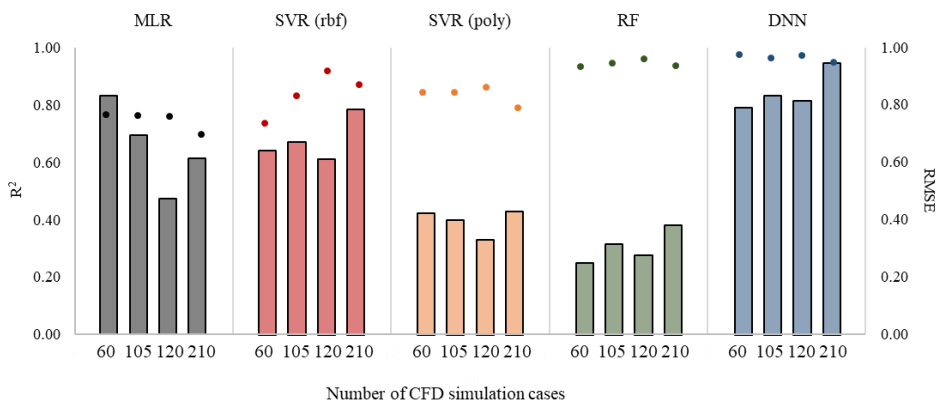


Figure 5-14 Calculation of prediction accuracy for each machine learning model according to the number of CFD simulation cases (line graph means R^2, bar graph means RMSE)

5.4. Conclusions

In this study, machine learning (ML) models that predict the local ventilation rate of a naturally-ventilated greenhouse using the computational fluid dynamics (CFD) simulation results were developed. First, a verified CFD model (Lee, Lee et al. 2018) was used to create reliable training data, and CFD simulation was performed for ten external wind speed cases, seven external wind direction cases, and three greenhouse vent type cases. For the simulation results, the ventilation rate by zone at the height of the crop group in the greenhouse was calculated, and training data for the ML models were created through the preprocessing process. Multiple linear regression (MLR), support vector regression (SVR), random forest (RF), and deep neural network (DNN) models were designed using the created training data, and hyper-parameter tuning was performed to improve the prediction accuracy of each model. Consequently, the DNN model exhibited the highest accuracy (R^2 = 985, RMSE = 0.242 AER·min^{-1}). Since the CFD-driven ML models proposed in this study require considerable cost and time for the creation of training data, a relatively small number of training data were supplemented by applying the bootstrapping technique. Consequently, R^2 was improved for ML models (MLR and SVR (poly)) that exhibited relatively low accuracy ($R^2 < 0.9$), and the RMSE value decreased for most of the ML models. Ten external wind speed cases and seven external wind direction cases, which were considered for the creation of training data, were reduced to five and four cases, respectively, to minimize CFD simulation cases, and the created training data were used for the ML models. It was found that most of the ML models exhibited the highest prediction accuracy when five external wind speed cases and seven external wind direction cases were considered. Finally, the ML model with the

highest prediction accuracy was defined as PLV-CFD driven ML. The proposed PLV-CFD driven ML that use the CFD simulation results are expected to supplement the limitations of ML models that can perform prediction only in the range of training data and have very low prediction accuracy for outliers that exceed the range.

Chapter 6. Conclusions

In this study, the of greenhouse environment prediction and control system was developed to monitor and predict the internal environment and local ventilation rate of the naturally ventilated greenhouse. In particular, this work proposed the optimal sensor location for monitoring the air temperature at the greenhouse and used it in the machine learning model to minimize the number of sensors required for the application of the machine learning models.

Machine learning models in the greenhouse environment prediction and cotrol system according following three purposes as previously discussed were developed. In Chapter 4, the PCTO-ML was developed for monitoring the internal air temperature of the natural ventilation greenhouse, and the optimal sensor location was selected through the model. Internal environment data were collected from nine different sampling points from a naturally ventilated greenhouse, and missing values and outliers were resolved through interpolation during the data preprocessing process. In addition, the training feature of the machine learning model was selected by analyzing the correlation between the measured and predicted air temperature inside the greenhouse, which is a predictor of the machine learning model. Machine learning models of ANN, SVR, and LSTM were developed, and LSTM was evaluated to show the highest accuracy (R^2 = 0.974, RMSE = 0.024, and P-RMSE = 0.458). So, final LSTM model was defined as the PCTO-ML. Therefore, the optimal sensor location for monitoring air temperature in the greenhouse was evaluated using the LSTM model, and sensor 5, which was the center location of the greenhouse, was selected as the optimal sensor location (R^2 = 0.984, RMSE = 0.019, and P-

RMSE = 0.365) in case of using one sensor. In order to minimize the type of sensor required for the machine learning model, the Simplified LSTM model with reduced learning features was developed as the PCTO-ML, and the optimal sensor locations were evaluated as sensor 5. The optimal sensor locations when using two sensors were evaluated as sensors 3 and 9, and the optimal sensor locations of the places to use three sensors were evaluated as sensors 1, 2, and 7.

In Chapter 4, the future air temperature inside the naturally ventilated greenhouse was predicted using PFTO-ML. The PFTO-ML was developed, and the process of collecting and preprocessing learning data for model development was the same as in Chapter 3. The highest prediction accuracy of the PFTO-ML, sequence length, was evaluated at 30 min. The results of predicting the future air temperature at each location in the greenhouse using the PFTO-ML showed high prediction accuracy of $R^2 > 0.95$, and RMSE < 0.65. In order to minimize the installation of sensors in the greenhouse, the predictive accuracy of the PFTO-ML was evaluated by applying the optimal sensor location suggested in PCTO-ML. As a result, the prediction accuracy was calculated relatively low when using one optimal sensor, but when the number of optimal sensors increased, the prediction accuracy improved.

In Chapter 5, the PLV-CFD driven ML was developed for predicting the ventilation rate in naturally ventilated greenhouses by region through learning the results of CFD simulation. To compensate for the limitations of machine learning models that prediction range was determined by the range of learned data, CFD simulation results that can be calculated for arbitrary environmental conditions were used as learning data for machine learning models. In order to generate learning data, 210 CFD cases of 10 wind speeds, 7 wind directions, and 3 greenhouse ventilation window types were performed. MLR, SVR, Random Forest, and DNN were

developed for PLV-CFD driven ML, and the optimization of hyper-parameters for each machine learning model was performed. For each optimized machine learning model, the bootstrapping technique was applied to supplement a relatively small number of learning data. Finally, the accuracy of the machine learning model according to the decrease in the case of the CFD simulation used as learning data was evaluated.

The greenhouse environment prediction and control system proposed in this dissertation predicted the current (PCTO-ML) and future (PETO-ML) air temperature inside the greenhouse and local ventilation rate (PLV-CFD driven ML). When the three models constituting the system are linked, it is of great significance that it is possible to predict the internal air temperature of the natural ventilation greenhouse and to control the proper ventilation of the predicted air temperature. In addition, since only minimum number of sensors are required for the development and operation of the model, it is economical to introduce this system for domestic greenhouses application where air temperature sensors are installed. In addition, it is expected that it can be applied to other agricultural facilities if the model design pipeline proposed in this study is used in consideration of additional factors that can consider the ventilation characteristics of facilities.

References

Ministry of Agriculture Food and Rural Affairs. 2022. Available online: https://www.mafra.go.kr (accessed on 10 June 2022)

Abdolrasol, M. G., S. S. Hussain, T. S. Ustun, M. R. Sarker, M. A. Hannan, R. Mohamed, J. A. Ali, S. Mekhilef and A. Milad (2021). "Artificial neural networks based optimization techniques: A review." Electronics **10**(21): 2689.

Abuella, M. and B. Chowdhury (2017). Random forest ensemble of support vector regression models for solar power forecasting. 2017 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), IEEE.

Afonso, M., R. Barth and A. Chauhan (2019). Deep learning based plant part detection in Greenhouse settings. 12th EFITA International Conference, EFITA.

Akrami, M., A. Javadi, M. Hassanein, R. Farmani, G. Tabor, A. Negm and H. E. Fath (2019). "Analysing greenhouse ventilation using Computational Fluid Dynamics (CFD)."

Al-Bahadly, I. and J. Thompson (2015). Garden watering system based on moisture sensing, IEEE.

Al-Shawwa, M. O., A. A.-R. Al-Absi, S. A. Hassanein, K. A. Baraka and S. S. Abu-Naser (2018). "Predicting temperature or humidity in the surrounding environment using artificial neural network."

Alhnaity, B., S. Pearson, G. Leontidis and S. Kollias (2019). Using deep learning to predict plant growth and yield in greenhouse environments. International Symposium on Advanced Technologies and Management for Innovative Greenhouses: GreenSys2019 1296.

Ali, I., F. Cawkwell, E. Dwyer and S. Green (2016). "Modeling managed grassland biomass estimation by using multitemporal remote sensing data—A machine learning approach." IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing **10**(7): 3254-3264.

Allouhi, A., N. Choab, A. Hamrani and S. Saadeddine (2021). "Machine learning

algorithms to assess the thermal behavior of a Moroccan agriculture greenhouse." Cleaner Engineering and Technology **5**: 100346.

Alsharif, M. H., M. K. Younes and J. Kim (2019). "Time series ARIMA model for prediction of daily and monthly average global solar radiation: The case study of Seoul, South Korea." Symmetry **11**(2): 240.

Amara, J., B. Bouaziz and A. Algergawy (2017). A Deep Learning-based Approach for Banana Leaf Diseases Classification. BTW (workshops).

Amral, N., C. Ozveren and D. King (2007). Short term load forecasting using multiple linear regression. 2007 42nd International universities power engineering conference, IEEE.

Arif, K. I. and H. F. Abbas (2015). "Design and implementation a smart greenhouse." International Journal of Computer Science and Mobile Computing **4**(8): 335-347.

Athani, S., C. H. Tejeshwar, M. M. Patil, P. Patil and R. Kulkarni (2017). Soil moisture monitoring using IoT enabled arduino sensors with neural networks for improving soil management for farmers and predict seasonal rainfall for planning future harvest in North Karnataka—India, IEEE.

Baeza, E., J. Pérez-Parra, J. Lopez and J. Montero (2006). CFD study of the natural ventilation performance of a parral type greenhouse with different numbers of spans and roof vent configurations. International Symposium on Greenhouse Cooling 719.

Baeza, E. J., J. J. Pérez-Parra, J. I. Montero, B. J. Bailey, J. C. López and J. C. Gázquez (2009). "Analysis of the role of sidewall vents on buoyancy-driven natural ventilation in parral-type greenhouses with and without insect screens using computational fluid dynamics." Biosystems engineering **104**(1): 86-96.

Bakay, M. S. and Ü. Ağbulut (2021). "Electricity production based forecasting of greenhouse gas emissions in Turkey with deep learning, support vector machine and artificial neural network algorithms." Journal of Cleaner Production **285**: 125324.

Bartzanas, T., T. Boulard and C. Kittas (2004). "Effect of vent arrangement on windward ventilation of a tunnel greenhouse." Biosystems Engineering

88(4): 479-490.

Bengio, Y., A. Courville and P. Vincent (2013). "Representation learning: A review and new perspectives." IEEE transactions on pattern analysis and machine intelligence **35**(8): 1798-1828.

Benni, S., P. Tassinari, F. Bonora, A. Barbaresi and D. Torreggiani (2016). "Efficacy of greenhouse natural ventilation: Environmental monitoring and CFD simulations of a study case." Energy and Buildings **125**: 276-286.

Bjerg, B. (2011). CFD Analyses of Methods to Improve Air Quality and Efficiency of Air Cleaning in Pig Production. Chemistry, emission control, radioactive pollution and indoor air quality, IntechOpen.

Bjerg, B., T. Norton, T. Banhazi, G. Zhang, T. Bartzanas, P. Liberati, G. Cascone, I.-B. Lee and A. Marucci (2013). "Modelling of ammonia emissions from naturally ventilated livestock buildings. Part 1: Ammonia release modelling." Biosystems engineering **116**(3): 232-245.

Bontempi, G., S. Ben Taieb and Y.-A. L. Borgne (2012). Machine learning strategies for time series forecasting. European business intelligence summer school, Springer.

Bot, G. P. (1983). Greenhouse climate: from physical processes to a dynamic model, Wageningen University and Research.

Boulard, T., A. Baille and B. Draoui (1993). "Greenhouse natural ventilation measurements and modelling."

Boulard, T. and B. Draoui (1995). "Natural ventilation of a greenhouse with continuous roof vents: measurements and data analysis." Journal of Agricultural engineering research **61**(1): 27-35.

Boulard, T., J. Meneses, M. Mermier and G. Papadakis (1996). "The mechanisms involved in the natural ventilation of greenhouses." Agricultural and forest meteorology **79**(1-2): 61-77.

Boulard, T., G. Papadakis, C. Kittas and M. Mermier (1997). "Air flow and associated sensible heat exchanges in a naturally ventilated greenhouse." Agricultural and Forest Meteorology **88**(1-4): 111-119.

Boulard, T., J.-C. Roy, J.-B. Pouillard, H. Fatnassi and A. Grisey (2017). "Modelling of micrometeorology, canopy transpiration and

photosyndissertation in a closed greenhouse using computational fluid dynamics." Biosystems Engineering **158**: 110-133.

Bournet, P.-E. and T. Boulard (2010). "Effect of ventilator configuration on the distributed climate of greenhouses: A review of experimental and CFD studies." Computers and electronics in agriculture **74**(2): 195-217.

Bournet, P., S. O. Khaoua, T. Boulard, C. Migeon and G. Chasseriaux (2007). "Effect of roof and side opening combinations on the ventilation of a greenhouse using computer simulation." Transactions of the ASABE **50**(1): 201-212.

Brewster, C., I. Roussaki, N. Kalatzis, K. Doolin and K. Ellis (2017). "IoT in agriculture: Designing a Europe-wide large-scale pilot." IEEE communications magazine **55**(9): 26-33.

Brugger, M., T. Short and W. Bauerle (1987). "An evaluation of horizontal air flow in six commercial greenhouses." American Society of Agricultural Engineers (USA).

Cabaneros, S. M., J. K. Calautit and B. R. Hughes (2019). "A review of artificial neural network models for ambient air pollution prediction." Environmental Modelling & Software **119**: 285-304.

Cadenas, E., W. Rivera, R. Campos-Amezcua and C. Heard (2016). "Wind speed prediction using a univariate ARIMA model and a multivariate NARX model." Energies **9**(2): 109.

Cai, W., R. Wei, L. Xu and X. Ding (2022). "A method for modelling greenhouse temperature using gradient boost decision tree." Information Processing In Agriculture **9**(3): 343-354.

Cai, Y., K. Guan, D. Lobell, A. B. Potgieter, S. Wang, J. Peng, T. Xu, S. Asseng, Y. Zhang and L. You (2019). "Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches." Agricultural and forest meteorology **274**: 144-159.

Campen, J. (2006). Ventilation of small multispan greenhouse in relation to the window openings calculated with CFD. III International Symposium on Models for Plant Growth, Environmental Control and Farm Management in Protected Cultivation 718.

Cao, Y., L. Cui and Q. Lv (2020). Research and Implementation of Organic

Cucumber Intelligent Greenhouse Monitoring System Based on NB-IoT and Raspberry Pi. 2020 9th International Conference on Applied Science, Engineering and Technology (ICASET 2020), Atlantis Press.

Cayli, A. (2020). "Temperature and relative humidity spatial variability: An assessment of the environmental conditions inside greenhouses."

Çerçi, K. N. and M. Daş (2019). "Modeling of heat transfer coefficient in solar greenhouse type drying systems." Sustainability **11**(18): 5127.

Chalabi, Z. and B. Bailey (1989). "Simulation of the energy balance in a greenhouse." Divisional Note, AFRC Institute of Engineering Research (UK).

Chang, D.-E., K.-R. Ha, H.-D. Jun and K.-H. Kang (2012). "Determination of optimal pressure monitoring locations of water distribution systems using entropy theory and genetic algorithm." Journal of Korean Society of Water and wastewater **26**(1): 1-12.

Chauhan, K. and M. Lunagaria (2022). "Interpolation of microclimatic parameters over capsicum under open ventilated greenhouse." International Journal of Economic Plants **9**(1): 95-100.

Chen, C. and C. Gorlé (2022). "Optimal temperature sensor placement in buildings with buoyancy-driven natural ventilation using computational fluid dynamics and uncertainty quantification." Building and Environment **207**: 108496.

Chen, L. and J. Wen (2010). "Comparison of sensor systems designed using multizone, zonal, and CFD data for protection of indoor environments." Building and Environment **45**: 1061-1071.

Chen, L. and J. Wen (2012). "The selection of the most appropriate airflow model for designing indoor air sensor systems." Building and Environment **50**: 34-43.

Chen, Y. L. and J. Wen (2010). "Comparison of sensor systems designed using multizone, zonal, and CFD data for protection of indoor environments." Building and Environment **45**(4): 1061-1071.

Cheng, X., D. Li, L. Shao and Z. Ren (2021). "A virtual sensor simulation system of a flower greenhouse coupled with a new temperature microclimate model using three-dimensional CFD." Computers and Electronics in

Agriculture **181**: 105934.

Chevalier, R. F., G. Hoogenboom, R. W. McClendon and J. A. Paz (2011). "Support vector regression with reduced training sets for air temperature prediction: a comparison with artificial neural networks." Neural Computing and Applications **20**(1): 151-159.

Choab, N., A. Allouhi, A. El Maakoul, T. Kousksou, S. Saadeddine and A. Jamil (2019). "Review on greenhouse microclimate and application: Design parameters, thermal modeling and simulation, climate controlling technologies." Solar Energy **191**: 109-137.

Choi, H.-M. and J.-M. Kim (2019). "Anomaly Detection System of Smart Farm ICT Device." The Journal of the Institute of Internet, Broadcasting and Communication **19**(2): 169-174.

Choi, K., K. Park and S. Jeong (2021). Classification of Growth Conditions in Paprika Leaf Using Deep Neural Network and Hyperspectral Images. 2021 Twelfth International Conference on Ubiquitous and Future Networks (ICUFN), IEEE.

Chung, C.-L., K.-J. Huang, S.-Y. Chen, M.-H. Lai, Y.-C. Chen and Y.-F. Kuo (2016). "Detecting Bakanae disease in rice seedlings by machine vision." Computers and electronics in agriculture **121**: 404-411.

Costantino, A., L. Comba, G. Sicardi, M. Bariani and E. Fabrizio (2019). Thermal environment inside mechanically ventilated greenhouses: results from a long-term monitoring campaign. International Mid-Term Conference of the Italian Association of Agricultural Engineering, Springer.

Dahiya, N., S. Gupta and S. Singh (2022). "A Review Paper on Machine Learning Applications, Advantages, and Techniques." ECS Transactions **107**(1): 6137.

Danita, M., B. Mathew, N. Shereen, N. Sharon and J. J. Paul (2018). IoT based automated greenhouse monitoring system, IEEE.

Deka, P. C. (2014). "Support vector machine applications in the field of hydrology: a review." Applied soft computing **19**: 372-386.

Dibike, Y. B., S. Velickov, D. Solomatine and M. B. Abbott (2001). "Model induction with support vector machines: introduction and applications." Journal of Computing in Civil Engineering **15**(3): 208-216.

Durmuş, H., E. O. Güneş and M. Kırcı (2017). Disease detection on the leaves of the tomato plants by using deep learning. 2017 6th International conference on agro-geoinformatics, IEEE.

Eforn, B. (1979). "Bootstrap methods: another look at the jackknife." The Annals of Statistics **7**: 1-26.

Escamilla-García, A., G. M. Soto-Zarazúa, M. Toledano-Ayala, E. Rivas-Araiza and A. Gastélum-Barrios (2020). "Applications of artificial neural networks in greenhouse technology and overview for smart agriculture development." Applied Sciences **10**(11): 3835.

Fan, L., Y. Ji and G. Wu (2021). Research on Temperature Prediction Model in Greenhouse Based on Improved SVR. Journal of Physics: Conference Series, IOP Publishing.

Fatnassi, H., T. Boulard, H. Benamara, J. C. Roy, R. Suay and C. Poncet (2017). "Increasing the height and multiplying the number of spans of greenhouse: How far can we go." Acta Hortic **1170**: 137-144.

Fatnassi, H., T. Boulard, H. Demrati, L. Bouirden and G. Sappe (2002). "SE—Structures and Environment: Ventilation Performance of a Large Canarian-Type Greenhouse Equipped with Insect-Proof Nets." Biosystems Engineering **82**(1): 97-105.

Feng, L., H. Li and Y. Zhi (2013). Greenhouse CFD simulation for searching the sensors optimal placements. 2013 Second International Conference on Agro-Geoinformatics (Agro-Geoinformatics), IEEE.

Fernandez, J. and B. Bailey (1992). "Measurement and prediction of greenhouse ventilation rates." Agricultural and Forest Meteorology **58**(3-4): 229-245.

Fidaros, D., C. Baxevanou, T. Bartzanas and C. Kittas (2010). "Numerical simulation of thermal behavior of a ventilated arc greenhouse during a solar day." Renewable Energy **35**(7): 1380-1386.

Flores, P., Z. Zhang, C. Igathinathane, M. Jithin, D. Naik, J. Stenger, J. Ransom and R. Kiran (2021). "Distinguishing seedling volunteer corn from soybean through greenhouse color, color-infrared, and fused images using machine and deep learning." Industrial Crops and Products **161**: 113223.

Fontanini, A. D., U. Vaidya and B. Ganapathysubramanian (2016). "A

methodology for optimal placement of sensors in enclosed environments: A dynamical systems approach." Building and Environment **100**: 145-161.

Fu, Y., M. Sha, C. Wu, A. Kutta, A. Leavey, C. Lu, H. Gonzalez, W.-N. Wang, B. Drake and P. Biswas (2014). Thermal Modeling for a HVAC Controlled Real-life Auditorium.

Fuentes, A., S. Yoon, S. C. Kim and D. S. Park (2017). "A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition." Sensors **17**(9): 2022.

Fuentes, S., C. G. Viejo, E. Tongson, F. R. Dunshea, H. H. Dac and N. Lipovetzky (2022). "Animal biometric assessment using non-invasive computer vision and machine learning are good predictors of dairy cows age and welfare: The future of automated veterinary support systems." Journal of Agriculture and Food Research **10**: 100388.

Ge, J., L. Zhao, Z. Yu, H. Liu, L. Zhang, X. Gong and H. Sun (2022). "Prediction of greenhouse tomato crop evapotranspiration using XGBoost machine learning model." Plants **11**(15): 1923.

Geng, X., Q. Zhang, Q. Wei, T. Zhang, Y. Cai, Y. Liang and X. Sun (2019). "A mobile greenhouse environment monitoring system based on the internet of things." Ieee Access **7**: 135832-135844.

Glad, K. (2020). Neural Network Predictive Control for Greenhouse Climate.

Gong, L., M. Yu, S. Jiang, V. Cutsuridis and S. Pearson (2021). "Deep learning based prediction on greenhouse crop yield combined TCN and RNN." Sensors **21**(13): 4537.

Gorczyca, M. (2019). "Machine learning applications for monitoring heat stress in livestock."

Gorczyca, M. T., H. F. M. Milan, A. S. C. Maia and K. G. Gebremedhin (2018). "Machine learning algorithms to predict core, skin, and hair-coat temperatures of piglets." Computers and Electronics in Agriculture **151**: 286-294.

Grimberg, R., M. Teitel, S. Ozer, A. Levi and A. Levy (2022). "Estimation of Greenhouse Tomato Foliage Temperature Using DNN and ML Models." Agriculture **12**(7): 1034.

Gümüşçü, A., M. E. Tenekeci and A. V. Bilgili (2020). "Estimation of wheat planting date using machine learning algorithms based on available climate data." Sustainable Computing: Informatics and Systems **28**: 100308.

Guerriero, F., A. Violi, E. Natalizio, V. Loscri and C. Costanzo (2011). "Modelling and solving optimal placement problems in wireless sensor networks." Applied Mathematical Modelling **35**(1): 230-241.

Guo, D., J. Juan, L. Chang, J. Zhang and D. Huang (2017). "Discrimination of plant root zone water status in greenhouse production based on phenotyping and machine learning techniques." Scientific reports **7**(1): 1-12.

Guzmán, C. H., J. L. Carrera, H. A. Durán, J. Berumen, A. A. Ortiz, O. A. Guirette, A. Arroyo, J. A. Brizuela, F. Gómez and A. Blanco (2018). "Implementation of virtual sensors for monitoring temperature in greenhouses using CFD and control." Sensors **19**(1): 60.

Hadabas, J., M. Hovari, I. Vass and A. Kertész (2019). "IoLT smart pot: an IoT-cloud solution for monitoring plant growth in greenhouses."

Hamel, D., M. Chwastek, B. Farouk, K. Dandekar and M. Kam (2006). A computational fluid dynamics approach for optimization of a sensor network. 2006 IEEE International Workshop on Measurement Systems for Homeland Security, Contraband Detection and Personal Safety, IEEE.

Hamrani, A., A. Akbarzadeh and C. A. Madramootoo (2020). "Machine learning for predicting greenhouse gas emissions from agricultural soils." Science of The Total Environment **741**: 140338.

Hamza, A. and M. Ramdani (2020). "Non-PDC interval type-2 fuzzy model predictive microclimate control of a greenhouse." Journal of Control, Automation and Electrical Systems **31**(1): 62-72.

He, K.-s., D.-y. Chen, L.-j. Sun, Z.-l. Liu and Z.-y. Huang (2015). "The effect of vent openings on the microclimate inside multi-span greenhouses during summer and winter seasons." Engineering Applications of Computational Fluid Mechanics **9**(1): 399-410.

Heemink, A. and A. Segers (2002). "Modeling and prediction of environmental data in space and time using Kalman filtering." Stochastic Environmental

Research and Risk Assessment **16**(3): 225-240.

Hernández, J., M. Morales, I. Escobar, N. Castilla and T. Soriano (2002). Radiation transmission differences in east-west oriented plastic greenhouses. XXVI International Horticultural Congress: Protected Cultivation 2002: In Search of Structures, Systems and Plant Materials for 633.

Ho, S.-L., M. Xie and T. N. Goh (2002). "A comparative study of neural network and Box-Jenkins ARIMA modeling in time series prediction." Computers & Industrial Engineering **42**(2-4): 371-375.

Hochreiter, S. and J. Schmidhuber (1997). "Long short-term memory." Neural computation **9**(8): 1735-1780.

Holcman, E. and P. C. Sentelhas (2012). "Microclimate under different shading screens in greenhouses cultivated with bromeliads." Revista Brasileira de Engenharia Agrícola e Ambiental **16**: 858-863.

Hong, S.-W., V. Exadaktylos, I.-B. Lee, T. Amon, A. Youssef, T. Norton and D. Berckmans (2017). "Validation of an open source CFD code to simulate natural ventilation for agricultural buildings." Computers and Electronics in Agriculture **138**: 80-91.

Hong, S.-W., I.-B. Lee, H.-S. Hwang, I.-H. Seo, J. Bitog, J.-I. Yoo, K.-S. Kim, S.-H. Lee, K.-W. Kim and N.-K. Yoon (2008). "Numerical simulation of ventilation efficiencies of naturally ventilated multi-span greenhouses in Korea." Transactions of the ASABE **51**(4): 1417-1432.

Hong, S., I. Lee, H. Hwang, I. Seo, J. Bitog, K. Kwon, J. Song, O. Moon, K. Kim and H. Ko (2011). "CFD modelling of livestock odour dispersion over complex terrain, part II: Dispersion modelling." Biosystems Engineering **108**(3): 265-279.

Hongkang, W., L. Li, W. Yong, M. Fanjia, W. Haihua and N. Sigrimis (2018). "Recurrent neural network model for prediction of microclimate in solar greenhouse." IFAC-PapersOnLine **51**(17): 790-795.

Horng, G.-J., M.-X. Liu and C.-C. Chen (2019). "The smart image recognition mechanism for crop harvesting system in intelligent agriculture." IEEE Sensors Journal **20**(5): 2766-2781.

Hu, H., L. Pan, K. Sun, S. Tu, Y. Sun, Y. Wei and K. Tu (2017). "Differentiation of deciduous-calyx and persistent-calyx pears using hyperspectral

reflectance imaging and multivariate analysis." Computers and Electronics in Agriculture **137**: 150-156.

Huang, G., P. Zhou and L. Zhang (2014). Optimal location of wireless temperature sensor nodes in large-scale rooms. 13th international conference on indoor air quality and climate, indoor air.

Hutapea, M. I., Y. Y. Pratiwi, I. M. Sarkis, I. K. Jaya and M. Sinambela (2020). Prediction of relative humidity based on long short-term memory network. AIP Conference Proceedings, AIP Publishing LLC.

Hwang, J.-S., T.-W. Kim and S.-M. Hong (1996). "Optimal location of controller and sensor for vibration control of building structure." JOURNAL-ARCHITECTURAL INSTITUTE OF KOREA **12**: 173-179.

Ichii, K., M. Ueyama, M. Kondo, N. Saigusa, J. Kim, M. C. Alberto, J. Ardö, E. S. Euskirchen, M. Kang and T. Hirano (2017). "New data-driven estimation of terrestrial CO_2 fluxes in Asia using a standardized database of eddy covariance measurements, remote sensing data, and support vector regression." Journal of Geophysical Research: Biogeosciences **122**(4): 767-795.

Ishii, M., L. Okushima, H. Moriyama, S. Sase, N. Fukuchi and A. Both (2014). Experimental study of natural ventilation in an open-roof greenhouse during the summer. XXIX International Horticultural Congress on Horticulture: Sustaining Lives, Livelihoods and Landscapes (IHC2014): 1107.

Jia, W., Z. Wei, Y. Zhang, L. Zhang, Q. Zhang and P. Feng (2019). Effects of Different Irrigation and Fertilization Treatments on Tomato Growth and Water Distribution and Fertilizer Distribution Uniformity in Greenhouse. 2019 ASABE Annual International Meeting, American Society of Agricultural and Biological Engineers.

Jumat, M. H., M. S. Nazmudeen and A. T. Wan (2018). "Smart farm prototype for plant disease detection, diagnosis & treatment using IoT device in a greenhouse."

Jung, D.-H., H.-J. Kim, J. Y. Kim, T. S. Lee and S. H. Park (2020). "Model predictive control via output feedback neural network for improved multi-window greenhouse ventilation control." Sensors **20**(6): 1756.

Kacira, M., S. Sase and L. Okushima (2004). "Effects of side vents and span numbers on wind-induced natural ventilation of a gothic multi-span greenhouse." Japan Agricultural Research Quarterly: JARQ **38**(4): 227-233.

Kacira, M., S. Sase and L. Okushima (2004). "Optimization of vent configuration by evaluating greenhouse and plant canopy ventilation rates under wind-induced ventilation." Transactions of the ASAE **47**(6): 2059-2067.

Keskin, M. E., I. K. Altınel, N. Aras and C. Ersoy (2014). "Wireless sensor network lifetime maximization by optimal sensor deployment, activity scheduling, data routing and sink mobility." Ad Hoc Networks **17**: 18-36.

Khosravi, A., L. Machado and R. Nunes (2018). "Time-series prediction of wind speed using machine learning algorithms: A case study Osorio wind farm, Brazil." Applied Energy **224**: 550-566.

Kim, J.-g., I.-b. Lee, S.-y. Lee, S.-j. Park, D.-y. Jeong, Y.-b. Choi, C. Decano-Valentin and U.-h. Yeo (2022). "Development of an Air-Recirculated Ventilation System for a Piglet House, Part 1: Analysis of Representative Problems through Field Experiment and Aerodynamic Analysis Using CFD Simulation for Evaluating Applicability of System." Agriculture **12**(8): 1139.

Kim, J.-Y., S.-Y. Lee, H.-H. Kim, H. Chun and I.-H. Yun (2002). "Ventilation effect of the greenhouse with folding panel type windows." Journal of Bio-Environment Control **11**(1): 5-11.

Kim, R.-w., J.-g. Kim, I.-b. Lee, U.-h. Yeo, S.-y. Lee and C. Decano-Valentin (2021). "Development of three-dimensional visualisation technology of the aerodynamic environment in a greenhouse using CFD and VR technology, part 1: Development of VR a database using CFD." Biosystems Engineering **207**: 33-58.

Kim, R.-w., I.-b. Lee and K.-s. Kwon (2017). "Evaluation of wind pressure acting on multi-span greenhouses using CFD technique, Part 1: Development of the CFD model." Biosystems engineering **164**: 235-256.

Kitaya, Y., T. Shibuya, T. Kozai and C. Kubota (1998). "Effects of light intensity and air velocity on air temperature, water vapor pressure, and CO_2 concentration inside a plant canopy under an artificial lighting

condition." Life Support & Biosphere Science **5**(2): 199-203.

Kitpo, N. and M. Inoue (2018). Early rice disease detection and position mapping system using drone and IoT architecture, IEEE.

Kittas, C., T. Boulard, M. Mermier and G. Papadakis (1996). "Wind induced air exchange rates in a greenhouse tunnel with continuous side openings." Journal of Agricultural Engineering Research **65**(1): 37-49.

Kittas, C., B. Draoui and T. Boulard (1995). "Quantification of the ventilation of a greenhouse with a roof opening." Agricultural and Forest Meteorology **1**(77): 95-111.

Kittas, C., M. Karamanis and N. Katsoulas (2005). "Air temperature regime in a forced ventilated greenhouse with rose crop." Energy and buildings **37**(8): 807-812.

Kodali, R. K., V. Jain and S. Karagwal IoT based smart greenhouse, IEEE.

Kodali, R. K., V. Jain and S. Karagwal (2016). IoT based smart greenhouse, IEEE.

Kola, R. R. K., P. Bojja and P. R. Kumari (2021). Optimal Technique of Tumor Detection and Prediction of Livestock by Deep Neural Network with TensorFlow and Keras. Journal of Physics: Conference Series, IOP Publishing.

Kozai, T. (1980). "Sase and M. Nara, A modeling approach to greenhouse environmental control by ventilation." Acta Hort **106**: 125-136.

Kulkarni, S., S. N. Mandal, G. S. Sharma and M. R. Mundada (2018). Predictive analysis to improve crop yield using a neural network model. 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI), IEEE.

Kutta, E. and J. Hubbart (2014). "Improving understanding of microclimate heterogeneity within a contemporary plant growth facility to advance climate control and plant productivity." Plant Sci **2**(5): 167-178.

Kwon, K.-s., I.-b. Lee and T. Ha (2016). "Identification of key factors for dust generation in a nursery pig house and evaluation of dust reduction efficiency using a CFD technique." Biosystems engineering **151**: 28-52.

Kwon, S.-H., S.-W. Jung, S.-G. Kwon, J.-M. Park, W.-S. Choi and J.-S. Kim (2017). "Comparative study on efficiencies of naturally-ventilated multi-span greenhouses in korea." Journal of the Korean Society of Industry

Convergence **20**(1): 8-18.

Lai, Y. and D. A. Dzombak (2020). "Use of the autoregressive integrated moving average (ARIMA) model to forecast near-term regional temperature and precipitation." Weather and Forecasting **35**(3): 959-976.

Lee, C. K., M. Chung, K.-Y. Shin, Y.-H. Im and S.-W. Yoon (2019). "A study of the effects of enhanced uniformity control of greenhouse environment variables on crop growth." Energies **12**(9): 1749.

Lee, D.-W., K. Jun, S. Lee, J.-K. Ko and M. S. Kim (2019). Abnormal gait recognition using 3D joint information of multiple kinects system and RNN-LSTM. 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE.

Lee, G.-J., J.-G. Gwon, S.-Y. Lee, T.-G. Gang and I.-C. Choe (2022). "미래 시설원예 연구를 위한 테스트베드 구축: 농촌진흥청 농업공학부 첨단디지털온실." Magazine of the Korean Society of Agricultural Engineers **64**(1): 64-73.

Lee, I., S. Sase, L. Okushima, A. Ikeguchi, K. Choi and J. Yun (2003). "A wind tunnel study of natural ventilation for multi–span greenhouse scale models using two–dimensional particle image velocimetry (PIV)." Transactions of the ASAE **46**(3): 763.

Lee, J., H. Kim, J. Park and Y. Jang (2009). "Selection of optimal measurement location using Kalman filtering." Journal of the Architectural Institute of Korea **29**(1): 193-196.

Lee, L., S. Hong, H. Hwang and L. Seo (2006). Study on ventilation efficiencies of naturally ventilated multi-span greenhouses in Korea. International Symposium on Greenhouse Cooling 719.

Lee, S.-y., I.-b. Lee and R.-w. Kim (2018). "Evaluation of wind-driven natural ventilation of single-span greenhouses built on reclaimed coastal land." Biosystems Engineering **171**: 120-142.

Lee, S.-y., I.-b. Lee, U.-h. Yeo, J.-g. Kim and R.-w. Kim (2022). "Machine Learning Approach to Predict Air Temperature and Relative Humidity inside Mechanically and Naturally Ventilated Duck Houses: Application of Recurrent Neural Network." Agriculture **12**(3): 318.

Lee, S.-y., I.-b. Lee, U.-h. Yeo, R.-w. Kim and J.-g. Kim (2019). "Optimal sensor placement for monitoring and controlling greenhouse internal environments." Biosystems Engineering **188**: 190-206.

Lee, T. J., J. Gottschlich, N. Tatbul, E. Metcalf and S. Zdonik (2018). "Greenhouse: A zero-positive machine learning system for time-series anomaly detection." arXiv preprint arXiv:1801.03168.

Lee, Y., J. Kim and S. Lee (2016). "The Optimal Placements and Number of Sensors for Dynamic Monitoring of Tall Buildings, J." Wind Eng. Inst. Korea **20**(2): 99-105.

Li, B., C. Delpha, D. Diallo and A. Migan-Dubois (2021). "Application of Artificial Neural Networks to photovoltaic fault detection and diagnosis: A review." Renewable and Sustainable Energy Reviews **138**: 110512.

Li, G., S. Ding, Y. Li and K. Zhang (2021). Music generation and human voice conversion based on LSTM. MATEC Web of Conferences, EDP Sciences.

Li, H., Y. Li, X. Yue, X. Liu, S. Tian and T. Li (2020). "Evaluation of airflow pattern and thermal behavior of the arched greenhouses with designed roof ventilation scenarios using CFD simulation." PloS one **15**(9): e0239851.

Lin, D., L. Zhang and X. Xia (2021). "Model predictive control of a Venlo-type greenhouse system considering electrical energy, water and carbon dioxide consumption." Applied Energy **298**: 117163.

Lin, W., Z. Wu, L. Lin, A. Wen and J. Li (2017). "An ensemble random forest algorithm for insurance big data analysis." Ieee access **5**: 16568-16575.

Lippi, M., M. A. Montemurro, M. Degli Esposti and G. Cristadoro (2019). "Natural language statistical features of LSTM-generated texts." IEEE Transactions on Neural Networks and Learning Systems **30**(11): 3326-3337.

Liu, P., J. Wang, A. K. Sangaiah, Y. Xie and X. Yin (2019). "Analysis and prediction of water quality using LSTM deep neural networks in IoT environment." Sustainability **11**(7): 2058.

Liu, X. and Z. J. Zhai (2009). "Protecting a whole building from critical indoor contamination with optimal sensor network design and source

identification methods." Building and Environment **44**(11): 2276-2283.

Liu, Y., J. Chen, Y. Lv and X. Li (2014). "Temperature simulation of greenhouse with CFD methods and optimal sensor placement." Sensors & Transducers **26**: 40.

López-Aguilar, K., A. Benavides-Mendoza, S. González-Morales, A. Juárez-Maldonado, P. Chiñas-Sánchez and A. Morelos-Moreno (2020). "Artificial neural network modeling of greenhouse tomato yield and aerial dry matter." Agriculture **10**(4): 97.

Louka, P., G. Galanis, N. Siebert, G. Kariniotakis, P. Katsafados, I. Pytharoulis and G. Kallos (2008). "Improvements in wind speed forecasts for wind power prediction purposes using Kalman filtering." Journal of Wind Engineering and Industrial Aerodynamics **96**(12): 2348-2362.

MAFRA. (2022). Retrieved 30. Nov., 2022, from https://www.mafra.go.kr.

Maione, C., B. L. Batista, A. D. Campiglia, F. Barbosa Jr and R. M. Barbosa (2016). "Classification of geographic origin of rice by data mining and inductively coupled plasma mass spectrometry." Computers and Electronics in Agriculture **121**: 101-107.

Marković, D. B., R. M. Pavlović, U. M. Pešović and S. S. Ranđić (2014). "System for monitoring microclimate conditions in greenhouse." Acta Agriculturae Serbica **19**(38): 105-114.

Marques, G. and R. Pitarma (2018). Agricultural environment monitoring system using wireless sensor networks and IoT, IEEE.

Mashonjowa, E., F. Ronsse, J. Milford, R. Lemeur and J. Pieters (2010). "Measurement and simulation of the ventilation rates in a naturally ventilated Azrom-type greenhouse in Zimbabwe." Applied Engineering in Agriculture **26**(3): 475-488.

Materne, N. and M. Inoue (2018). IoT monitoring system for early detection of agricultural pests and diseases, IEEE.

Mazumdar, S. and Q. Chen (2008). "Influence of cabin conditions on placement and response of contaminant detection sensors in a commercial aircraft." Journal of Environmental Monitoring **10**(1): 71-81.

McGibney, A., D. Pusceddu, S. Rea, D. Pesch, M. Geron and M. Keane (2012). A methodology for sensor modeling and placement optimization to support

temperature monitoring.

Medela, A., B. Cendón, L. Gonzalez, R. Crespo and I. Nevares (2013). IoT multiplatform networking to monitor and control wineries and vineyards, IEEE.

Mittal, A., S. Sarangi, S. Ramanath, P. V. Bhatt, R. Sharma and P. Srinivasu IoT-based precision monitoring of horticultural crops—A case-study on cabbage and capsicum, IEEE.

Mittal, A., S. Sarangi, S. Ramanath, P. V. Bhatt, R. Sharma and P. Srinivasu (2018). IoT-based precision monitoring of horticultural crops—A case-study on cabbage and capsicum, IEEE.

Molano-Jimenez, A., A. D. Orjuela-Cañón and W. Acosta-Burbano (2018). Temperature and Relative Humidity Prediction in Swine Livestock Buildings. 2018 IEEE Latin American Conference on Computational Intelligence (LA-CCI), IEEE.

Molina-Aiz, F. D., D. L. Valera and A. J. Álvarez (2004). "Measurement and simulation of climate inside Almeria-type greenhouses using computational fluid dynamics." Agricultural and Forest Meteorology **125**(1-2): 33-51.

Moon, T., S. Hong, H. Y. Choi, D. H. Jung, S. H. Chang and J. E. Son (2019). "Interpolation of greenhouse environment data using multilayer perceptron." Computers and Electronics in Agriculture **166**: 105023.

Mountrakis, G., J. Im and C. Ogole (2011). "Support vector machines in remote sensing: A review." ISPRS Journal of Photogrammetry and Remote Sensing **66**(3): 247-259.

Muangprathub, J., N. Boonnam, S. Kajornkasirat, N. Lekbangpong, A. Wanichsombat and P. Nillaor (2019). "IoT and agriculture data analysis for smart farm." Computers and electronics in agriculture **156**: 467-474.

Nayyar, A. and V. Puri (2016). Smart farming: IoT based smart sensors agriculture stick for live temperature and moisture monitoring using Arduino, cloud computing & solar technology.

Nie, H., G. Liu, X. Liu and Y. Wang (2012). "Hybrid of ARIMA and SVMs for short-term load forecasting." Energy Procedia **16**: 1455-1460.

Nobrega, J. P. and A. L. I. Oliveira (2019). "A sequential learning method with

Kalman filter and extreme learning machine for regression and time series forecasting." Neurocomputing **337**: 235-250.

Norton, T., D.-W. Sun, J. Grant, R. Fallon and V. Dodd (2007). "Applications of computational fluid dynamics (CFD) in the modelling and design of ventilation systems in the agricultural industry: A review." Bioresource technology **98**(12): 2386-2414.

Ogunlowo, Q. O., T. D. Akpenpuun, W.-H. Na, A. Rabiu, M. A. Adesanya, K. S. Addae, H.-T. Kim and H.-W. Lee (2021). "Analysis of heat and mass distribution in a single-and multi-span greenhouse microclimate." Agriculture **11**(9): 891.

Oh, S.-j. (2017). "A Design of intelligent information system for greenhouse cultivation." Journal of Digital Convergence **15**(2): 183-190.

Okamoto, H. and M. Koshi (1989). "A method to cope with the random errors of observed accident rates in regression analysis." Accident Analysis & Prevention **21**(4): 317-332.

Okayasu, T., A. P. Nugroho, A. Sakai, D. Arita, T. Yoshinaga, R.-i. Taniguchi, M. Horimoto, E. Inoue, Y. Hirai and M. Mitsuoka Affordable field environmental monitoring and plant growth measurement system for smart agriculture, IEEE.

Okushima, L., S. Sase, I.-B. Lee and B. Bailey (2000). Thermal environment and stress of workers in naturally ventilated greenhouses under mild climate. V International Symposium on Protected Cultivation in Mild Winter Climates: Current Trends for Suistainable Technologies 559.

Ou, C.-H., Y.-A. Chen, T.-W. Huang and N.-F. Huang (2020). Design and implementation of anomaly condition detection in agricultural IoT platform system. 2020 International Conference on Information Networking (ICOIN), IEEE.

Ozbek, A., A. Sekertekin, M. Bilgili and N. Arslan (2021). "Prediction of 10-min, hourly, and daily atmospheric air temperature: comparison of LSTM, ANFIS-FCM, and ARMA." Arabian Journal of Geosciences **14**(7): 1-16.

Pang, Z., Q. Chen, W. Han and L. Zheng (2015). "Value-centric design of the internet-of-things solution for food supply chain: Value creation, sensor portfolio and information fusion." Information Systems Frontiers **17**(2):

289-319.

Pantazi, X. E., D. Moshou, R. Oberti, J. West, A. M. Mouazen and D. Bochtis (2017). "Detection of biotic and abiotic stresses in crops by using hierarchical self organizing classifiers." Precision Agriculture **18**(3): 383-393.

Pantazi, X. E., A. A. Tamouridou, T. Alexandridis, A. L. Lagopodi, G. Kontouris and D. Moshou (2017). "Detection of Silybum marianum infection with Microbotryum silybum using VNIR field spectroscopy." Computers and Electronics in Agriculture **137**: 130-137.

Papadakis, G., M. Mermier, J. Meneses and T. Boulard (1996). "Measurement and analysis of air exchange rates in a greenhouse with continuous roof and side openings." Journal of Agricultural Engineering Research **63**(3): 219-227.

Papadimitriou, C., J. L. Beck and S.-K. Au (2000). "Entropy-based optimal sensor location for structural model updating." Journal of Vibration and Control **6**(5): 781-800.

Pardossi, A., F. Tognoni and L. Incrocci (2004). "Mediterranean greenhouse technology." Chronica horticulturae **44**(2): 28-34.

Park, J., B. Jeong, Y.-T. Chae and J.-W. Jeong (2021). "Machine learning algorithms for predicting occupants' behaviour in the manual control of windows for cross-ventilation in homes." Indoor and Built Environment **30**(8): 1106-1123.

Park, S.-J., I.-B. Lee, S.-Y. Lee, J.-G. Kim, Y.-B. Choi, C. Decano-Valentin, J.-H. Cho, H.-H. Jeong and U.-H. Yeo (2022). "Numerical Analysis of Ventilation Efficiency of a Korean Venlo-Type Greenhouse with Continuous Roof Vents." Agriculture **12**(9): 1349.

Patil, K. A. and N. R. Kale A model for smart agriculture using IoT, IEEE.

Philemon, M. D., Z. Ismail and J. Dare (2019). "A review of epidemic forecasting using artificial neural networks." International Journal of Epidemiologic Research **6**(3): 132-143.

Philibert, A., C. Loyce and D. Makowski (2013). "Prediction of N2O emission from local information with Random Forest." Environmental pollution **177**: 156-163.

Radojevic, N., B. Bjelogrlic, V. Aleksic, N. Rancic, M. Samardzic, S. Petkovic and S. Savic (2012). "Forensic aspects of water intoxication: four case reports and review of relevant literature." Forensic Science International **220**(1-3): 1-5.

Ramos, P., F. A. Prieto, E. Montoya and C. E. Oliveros (2017). "Automatic fruit count on coffee branches using computer vision." Computers and Electronics in Agriculture **137**: 9-22.

Ramponi, R. and B. Blocken (2012). "CFD simulation of cross-ventilation for a generic isolated building: impact of computational parameters." Building and environment **53**: 34-48.

Raschka, S. (2018). "Model evaluation, model selection, and algorithm selection in machine learning." arXiv preprint arXiv:1811.12808.

Rasheed, A., C. S. Kwak, W. H. Na, J. W. Lee, H. T. Kim and H. W. Lee (2020). "Development of a building energy simulation model for control of multi-span greenhouse microclimate." Agronomy **10**(9): 1236.

Rayhana, R., G. Xiao and Z. Liu (2020). "Internet of things empowered smart greenhouse farming." IEEE Journal of Radio Frequency Identification **4**(3): 195-211.

Reimers, N. and I. Gurevych (2017). "Optimal hyperparameters for deep lstm-networks for sequence labeling tasks." arXiv preprint arXiv:1707.06799.

Romero-Gómez, P., C. Y. Choi and I. L. Lopez-Cruz (2010). "Enhancement of the greenhouse air ventilation rate under climate conditions of central Mexico." Agrociencia **44**(1): 1-15.

Rong, L., P. V. Nielsen, B. Bjerg and G. Zhang (2016). "Summary of best guidelines and validation of CFD modeling in livestock buildings to ensure prediction quality." Computers and Electronics in Agriculture **121**: 180-190.

Rumelhart, D. E., G. E. Hinton and R. J. Williams (1986). "Learning representations by back-propagating errors." nature **323**(6088): 533-536.

Rustia, D. J. A., J. J. Chao, L. Y. Chiu, Y. F. Wu, J. Y. Chung, J. C. Hsu and T. T. Lin (2021). "Automatic greenhouse insect pest detection and recognition based on a cascaded deep learning classification method." Journal of Applied Entomology **145**(3): 206-222.

Saadon, T., N. Lazarovitch, D. Jerszurki and E. Tas (2021). "Predicting net radiation in naturally ventilated greenhouses based on outside global solar radiation for reference evapotranspiration estimation." Agricultural Water Management **257**: 107102.

Sandberg, M. and M. Sjöberg (1983). "The use of moments for assessing air quality in ventilated rooms." Building and environment **18**(4): 181-197.

Sase, S. (1988). "The effects of plant arrangement on airflow characteristics in a naturally ventilated glasshouse." Engineering and Economic Aspects of Energy Saving in Protected Cultivation 245: 429-435.

Sase, S., T. Takakura and M. Nara (1983). Wind tunnel testing on airflow and temperature distribution of a naturally ventilated greenhouse. III International Symposium on Energy in Protected Cultivation 148.

Sen, P., M. Roy and P. Pal (2016). "Application of ARIMA for forecasting energy consumption and GHG emission: A case study of an Indian pig iron manufacturing organization." Energy **116**: 1031-1038.

Senagi, K., N. Jouandeau and P. Kamoni (2017). "Using parallel random forest classifier in predicting land suitability for crop production." Journal of Agricultural Informatics **8**(3): 23-32.

Seng, D., Q. Zhang, X. Zhang, G. Chen and X. Chen (2021). "Spatiotemporal prediction of air quality based on LSTM neural network." Alexandria Engineering Journal **60**(2).

Sengupta, S. and W. S. Lee (2014). "Identification and determination of the number of immature green citrus fruit in a canopy under different ambient light conditions." Biosystems Engineering **117**: 51-61.

Senthilnath, J., A. Dokania, M. Kandukuri, K. Ramesh, G. Anand and S. Omkar (2016). "Detection of tomatoes using spectral-spatial methods in remotely sensed RGB images captured by UAV." Biosystems engineering **146**: 16-32.

Sethi, V., K. Sumathy, C. Lee and D. Pal (2013). "Thermal modeling aspects of solar greenhouse microclimate control: A review on heating technologies." Solar energy **96**: 56-82.

Setiawan, B., S. Djanali, T. Ahmad and I. Nopember (2019). "Increasing accuracy and completeness of intrusion detection model using fusion of

normalization, feature selection method and support vector machine." Int. J. Intell. Eng. Syst **12**(4): 378-389.

Shadab, A., S. Said and S. Ahmad (2019). "Box–Jenkins multiplicative ARIMA modeling for prediction of solar radiation: a case study." International Journal of Energy and Water Resources **3**(4): 305-318.

Shahid, N., T. Rappon and W. Berta (2019). "Applications of artificial neural networks in health care organizational decision-making: A scoping review." PloS one **14**(2): e0212356.

Shibuya, T. and T. Kozai (2001). "Light-use and water-use efficiencies of tomato plug sheets in the greenhouse." Environment Control in Biology **39**(1): 35-41.

Shibuya, T., J. Tsuruyama, Y. Kitaya and M. Kiyota (2006). "Enhancement of photosyndissertation and growth of tomato seedlings by forced ventilation within the canopy." Scientia Horticulturae **109**(3): 218-222.

Singh, D. and B. Singh (2020). "Investigating the impact of data normalization on classification performance." Applied Soft Computing **97**: 105524.

Singh, G., P. P. Singh, P. P. S. Lubana and K. Singh (2006). "Formulation and validation of a mathematical model of the microclimate of a greenhouse." Renewable Energy **31**(10): 1541-1560.

Singh, V. and K. Tiwari (2017). "Prediction of greenhouse micro-climate using artificial neural network." Appl. Ecol. Environ. Res **15**(1): 767-778.

Slamet, W., N. M. Irham and M. S. A. Sutan IoT based growth monitoring system of guava (Psidium guajava L.) Fruits, IOP Publishing.

Smola, A. J. and B. Schölkopf (2004). "A tutorial on support vector regression." Statistics and computing **14**(3): 199-222.

Sujatha, R., J. M. Chatterjee, N. Jhanjhi and S. N. Brohi (2021). "Performance of deep learning vs machine learning in plant leaf disease detection." Microprocessors and Microsystems **80**: 103615.

Suo, D., C. Zhang, P. Gradu, U. Ghai, X. Chen, E. Minasyan, N. Agarwal, K. Singh, J. LaChance and T. Zajdel (2021). "Machine learning for mechanical ventilation control." arXiv preprint arXiv:2102.06779.

Suryanarayana, G., J. Arroyo, L. Helsen and J. Lago (2021). "A data driven method for optimal sensor placement in multi-zone buildings." Energy and

Buildings **243**: 110956.

Tognoni, F., A. Pardossi and G. Serra (1997). Strategies to match greenhouses to crop production.

Toumpis, S. and L. Tassiulas (2006). "Optimal deployment of large wireless sensor networks." IEEE Transactions on Information Theory **52**(7): 2935-2953.

Tsai, Y.-Z., K.-S. Hsu, H.-Y. Wu, S.-I. Lin, H.-L. Yu, K.-T. Huang, M.-C. Hu and S.-Y. Hsu (2020). "Application of random forest and ICON models combined with weather forecasts to predict soil temperature and water content in a greenhouse." Water **12**(4): 1176.

Utami, S. S., R. J. Yanti, E. Nurjani and R. Wijaya (2021). "Optimal thermal sensors placement based on indoor thermal environment characterization by using CFD model." Journal of Applied Engineering Science **19**(3): 628-641.

Vanthoor, B., C. Stanghellini, E. J. Van Henten and P. De Visser (2011). "A methodology for model-based greenhouse design: Part 1, a greenhouse climate model for a broad range of designs and climates." Biosystems Engineering **110**(4): 363-377.

Vapnik, V. (1995). "The Nature of Statistical Learning Theory, Springer-Verlag; New York, Inc."

Vapnik, V. (1999). Three remarks on the support vector method of function estimation. Advances in kernel methods: support vector learning**:** 25-41.

Varma, P. S. and V. Anand (2021). "Random forest learning based indoor localization as an IoT service for smart buildings." Wireless Personal Communications **117**(4): 3209-3227.

Vatari, S., A. Bakshi and T. Thakur (2016). Green house by using IOT and cloud computing, IEEE.

Villagrán, E. and C. Bojacá (2020). "Study using a CFD approach of the efficiency of a roof ventilation closure system in a multi-tunnel greenhouse for nighttime microclimate optimization." Revista Ceres **67**: 345-356.

Villagrán, E. A., R. Gil, J. F. Acuña and C. R. Bojacá (2012). "Optimization of ventilation and its effect on the microclimate of a colombian multispan greenhouse." Agronomía colombiana **30**(2): 282-288.

Villagran, E. A., E. J. B. Romero and C. R. Bojacá (2019). "Transient CFD analysis

of the natural ventilation of three types of greenhouses used for agricultural production in a tropical mountain climate." Biosystems Engineering **188**: 288-304.

Waeytens, J., S. Durand and S. Sadr (2019). Experimental validation of a CFD-based air quality sensor placement strategy to localize indoor source emissions. Building Simulation 2019.

Wang, D., M. Wang and X. Qiao (2009). "Support vector machines regression and modeling of greenhouse environment." Computers and electronics in agriculture **66**(1): 46-52.

Wang, H. and B. Raj (2017). "On the origin of deep learning." arXiv preprint arXiv:1702.07800.

Wang, S., T. Boulard and R. Haxaire (2000). Measurement and analysis of air speed distribution in a naturally ventilated greenhouse. International Conference and British-Israeli Workshop on Greenhouse Techniques towards the 3rd Millennium 534.

Wang, Z. M., S. Basagni, E. Melachrinoudis and C. Petrioli (2005). Exploiting sink mobility for maximizing sensor networks lifetime. Proceedings of the 38th annual Hawaii international conference on system sciences, IEEE.

Wanjawa, B. W. and L. Muchemi (2014). "ANN model to predict stock prices at stock exchange markets." arXiv preprint arXiv:1502.06434.

Werbos, P. J. (1990). "Backpropagation through time: what it does and how to do it." Proceedings of the IEEE **78**(10): 1550-1560.

Wijaya, R., S. S. Utami, R. J. Yanti and E. Nurjani (2021). "OPTIMAL THERMAL SENSORS PLACEMENT BASED ON INDOOR THERMAL ENVIRONMENT CHARACTERIZATION BY USING CFD MODEL." Applied Engineering Science.

Wu, G. H., F. Liu, J. X. Li and W. Wang (2014). Environmental monitoring system designing: A Internet of Things approach, Trans Tech Publ.

Xia, S., X. Nan, X. Cai and X. Lu (2022). "Data fusion based wireless temperature monitoring system applied to intelligent greenhouse." Computers and Electronics in Agriculture **192**: 106576.

Yeo, U.-H., C. Decano-Valentin, T. Ha, I.-B. Lee, R.-W. Kim, S.-Y. Lee and J.-G. Kim (2020). "Impact analysis of environmental conditions on odour

dispersion emitted from pig house with complex terrain using CFD." Agronomy **10**(11): 1828.

Yeo, U.-H., S.-Y. Lee, S.-J. Park, J.-G. Kim, J.-H. Cho, C. Decano-Valentin, R.-W. Kim and I.-B. Lee (2022). "Rooftop Greenhouse:(2) Analysis of Thermal Energy Loads of a Building-Integrated Rooftop Greenhouse (BiRTG) for Urban Agriculture." Agriculture **12**(6): 787.

Yeo, U.-H., S.-Y. Lee, S.-J. Park, J.-G. Kim, Y.-B. Choi, R.-W. Kim, J. H. Shin and I.-B. Lee (2022). "Rooftop Greenhouse:(1) Design and Validation of a BES Model for a Plastic-Covered Greenhouse Considering the Tomato Crop Model and Natural Ventilation Characteristics." Agriculture **12**(7): 903.

Yi, T. H., G. D. Zhou, H. N. Li and C. W. Wang (2017). "Optimal placement of triaxial sensors for modal identification using hierarchic wolf algorithm." Structural Control and Health Monitoring **24**(8): e1958.

Yimwadsana, B., P. Chanthapeth, C. Lertthanyaphan and A. Pornvechamnuay (2018). An IoT controlled system for plant growth, IEEE.

Yoganathan, D., S. Kondepudi, B. Kalluri and S. Manthapuri (2018). "Optimal sensor placement strategy for office buildings using clustering algorithms." Energy and Buildings **158**: 1206-1225.

Yun, N.-G., J.-S. Lee, G.-S. Park and J.-Y. Lee (2017). "한국형 스마트팜 정책 및 기술개발 현황." Magazine of the Korean Society of Agricultural Engineers **59**(2): 19-27.

Zarinkamar, R. and R. Mayorga (2021). "Outdoor Relative Humidity Prediction via Machine Learning Techniques."

Zhang, J., M. Karkee, Q. Zhang, X. Zhang, M. Yaqoob, L. Fu and S. Wang (2020). "Multi-class object detection using faster R-CNN and estimation of shaking locations for automated shake-and-catch apple harvesting." Computers and Electronics in Agriculture **173**: 105384.

Zhang, L., Z. Xu, D. Xu, J. Ma, Y. Chen and Z. Fu (2020). "Growth monitoring of greenhouse lettuce based on a convolutional neural network." Horticulture research **7**.

Zhang, T. and Q. Chen (2007). "Identification of contaminant sources in enclosed spaces by a single sensor." Indoor air **17**(6): 439-449.

Zhang, X., J. Zhang, L. Li, Y. Zhang and G. Yang (2017). "Monitoring citrus soil moisture and nutrients using an IoT based system." Sensors **17**(3): 447.

Zhao, Y., M. Teitel and M. Barak (2001). "SE—structures and environment: Vertical temperature and humidity gradients in a naturally ventilated greenhouse." Journal of Agricultural Engineering Research **78**(4): 431-436.

Zhou, Y., S. Guo, C.-Y. Xu, F.-J. Chang and J. Yin (2020). "Improving the reliability of probabilistic multi-step-ahead flood forecasting by fusing unscented Kalman filter with recurrent neural network." Water **12**(2): 578.

Zou, W., F. Yao, B. Zhang, C. He and Z. Guan (2017). "Verification and predicting temperature and humidity in a solar greenhouse based on convex bidirectional extreme learning machine algorithm." Neurocomputing **249**: 72-85.

김근식, 구종회 and 차의영 (2020). "심층 학습 기반의 버섯 성장 모니터링 시스템 구현." 한국통신학회 학술대회논문집: 757-758.

김상엽, 박경섭, 이상민, 허병문 and 류근호 (2018). "머신러닝 기반의 온실 제어를 위한 예측모델 개발." 한국디지털콘텐츠학회 논문지 **19**(4): 749-756.

김예슬, 곽근호, 이경도, 나상일, 박찬원 and 박노욱 (2018). "작물분류에서 기계학습 및 딥러닝 알고리즘의 분류 성능 평가: 하이퍼파라미터와 훈련자료 크기의 영향 분석." Korean Journal of Remote Sensing **34**(5): 811-827.

김태완 (2019). "ICT 기반 스마트팜 온실 현황과 전망." 한국통신학회지 (정보와통신) **36**(3): 3-8.

나명환, 조완현 and 김상균 (2020). "딥러닝 알고리즘을 이용한 토마토에서 발생하는 여러가지 병해충의 탐지와 식별에 대한 웹응용 플랫폼의 구축." 품질경영학회지 **48**(4): 581-596.

남상운, 김영식, 고기혁 and 성인모 (2012). "단동 플라스틱 온실의 천창 환기효과와 설치기준 분석." Korean Journal of Agricultural Science **39**(2): 271-277.

문태원, 박준영 and 손정익 (2020). "합성곱 신경망을 이용한 온실 파프리카의 작물 생체중 추정." 시설원예· 식물공장 **29**(4): 381-387.

박민정, 최덕규, 손진관, 윤성욱, 김희태, 이승기 and 강동현 (2020). "연동형 비닐하우스의 환기창 형태 조사 및 자연환기 효과 분석." 시설원예· 식물공장 **29**(1): 36-42.

오정원, 김행곤 and 김일태 (2019). "머신러닝 적용 과일 수확시기 예측시스템 설계 및 구현." 스마트미디어저널 **8**(1): 74-81.

우현준, 조주연, 이주선 and 한동일 (2017). "토너먼트식 구조를 적용한 병해충 진단 시스템 개발." 대한전자공학회 학술대회: 1316-1319.

유인호, 조명환, 이시영, 전희 and 이인복 (2007). "순환팬이 온풍난방 온실의 기상분포 균일화에 미치는 영향." 생물환경조절학회지 **16**(4): 291-296.

이민구 and 정경권 (2012). "CFD 해석을 이용한 실내 온도 최적 측정 위치 선정 방법." 한국정보통신학회논문지 **16**(7): 1559-1566.

조규정, 김기영 and 양원모 (2015). "온실 ICT 융복합 실태조사와 복숭아형 랙피니언천창 적용 단동온실 및 CFD 유동해석." 시설원예· 식물공장 **24**(4): 308-316.

최강인, 노혜민, 정성환 and 유철중 (2019). "초분광 영상과 심층 신경망을 이용한 작물 생육상태 분류 기법: 파프리카잎 사례 연구." 한국정보기술학회논문지 **17**(12): 1-12.

최호길, 안희학, 정이나 and 이병관 (2019). "IoT 및 딥 러닝 기반 스마트 팜 환경 최적화 및 수확량 예측 플랫폼." 한국정보전자통신기술학회 논문지 **12**(6): 672-680.

하정수 (2015). Evaluation of Natural Ventilation Efficiency of Protected Cultivation System in Reclaimed Land using Aerodynamic Simulation, 서울대학교 대학원.

홍성은, 박태주, 방준일 and 김화종 (2020). "ConvLSTM 을 사용한 토마토 생산량및 성장량 예측 모델에 관한 연구."

한국정보기술학회논문지 **18**(1): 1-10.

홍세운 and 이인복 (2014). "자연 환기식 온실의 모델 기반 환기 제어를 위한 미기상 환경 예측 모형." 시설원예· 식물공장 **23**(3): 181-191.

국 문 초 록

시설 재배를 통한 작물 생산은 연중 안정된 생산성과 고품질의 작물을 생산함으로써 지난 수년간 지속적으로 발전해오고 있다. 현재 국내 채소 재배 시설 면적과 총 생산량은 각각 2020년 52,571 ha와 2,441,000 ton으로 1970년대부터 현재까지 그 규모가 꾸준히 증가해 오고 있는 추세이다 (MAFRA, 2022). 또한 ICT 기술의 발전과 각종 센서 및 제어 시스템의 개발로 인해 온실 내부 환경에 대해 정밀한 모니터링과 제어가 가능한 스마트팜 온실이 빠르게 보급되고 있다. 하지만 스마트팜을 도입한 많은 농가에서 설치비용의 확보에 어려움을 느끼고 있으며, 센서 및 장비의 잦은 고장과 스마트팜 활용의 어려움을 가지고 있는 것으로 나타났다.

본 논문의 최종 목표는 자연환기 온실의 내부 환경을 모니터링하고 예측할 수 있는 모델을 개발하는 것이다. 따라서 온실 내부의 가장 중요한 환경 요소인 공기 온도의 최적 센서 위치를 선정하기 위한 기계학습 모델을 개발하였다. 현재의 온실 내부 공기 온도를 모니터링하는 것 뿐만 아니라 가까운 미래의 온실 내부 공기 온도를 예측하여 공조시스템의 선제적 제어에 활용될 수 있는 기계학습 모델을 개발하였다. 또한 예측된 온실 내부 공기 온도를 작물의 적정 생육 환경으로 조절할 수 있는 가장 기본적인 방법인 자연환기를 예측한 기계학습 모델을 개발하였다.

2장에서는, 선행 연구의 방법론, 한계점의 분석을 통해 연구의 필요성과 연구 방향성의 기틀을 다지고, 연구 방법론의 적립을 위해

“온실 내부 생육 환경의 중요성”, “온실 환기 평가 방법”, “온실의 내부 환경 모니터링을 위한 최적 센서 위치”, “농업 분야에서의 기계학습 기법 적용”에 대한 연구사를 검토하였다.

3장에서는 기계학습 모델을 이용하여 자연환기 온실의 내부 공기 온도 모니터링을 위한 최적 센서 위치를 선정하였다. 이를 위하여 현장 실험을 통해 자연환기식 온실 내부의 환경 요소들을 수집하였으며, 기계학습 모델의 예측 정확도를 높이기 위하여 일련의 전처리 과정을 수행하였다. 자연환기 온실의 내부 공기 온도 예측을 위한 기계학습 모델은 ANN, SVR, LSTM을 개발하였다. 기계학습 모델들의 예측 성능을 비교한 결과 LSTM이 가장 높은 정확도(R^2 = 0.974, RMSE = 0.024, P-RMSE = 0.458)를 보이는 것으로 평가되었다. 따라서 LSTM 모델을 이용하여 온실 내 공기 온도 모니터링을 위한 최적 센서 위치를 평가하였으며, 온실 중앙 위치인 5번 센서 위치에서 측정한 환경데이터를 학습할 경우, 온실 내부의 각 센서 설치 위치에 대한 공기 온도 예측에 가장 높은 정확도를 보였다 (R^2 = 0.984, RMSE = 0.019, P-RMSE = 0.365). 마지막으로 기계학습 모델에 필요한 센서의 종류를 최소화하기 위하여 학습 인자를 줄인 Simplified LSTM 모델을 개발하였으며, 이때의 최적 센서 위치는 5번으로 평가되었다.

4장에서는 기계학습 모델을 이용하여 자연환기 온실 내부의 미래 공기 온도를 예측하였다. LSTM 모델이 개발되었으며, 모델 개발을 위한 학습 데이터의 수집과 전처리 과정은 3장과 동일하다. 단, LSTM 모델의 미래 상황 예측 시 입력값의 단위에 해당하는 시퀀스 길이에 따른 LSTM

모델의 정확도를 평가하였으며, 30 분이 최적 시퀀스 길이로 산정되었다. LSTM 모델을 이용하여 온실 내 각 위치에서의 미래 공기 온도를 예측한 결과 대부분 $R^2 > 0.85$, RMSE < 0.65의 높은 예측 정확도를 나타내었다. Chapter 3에서 제시한 최적 센서 위치를 적용하여 LSTM 모델의 예측 정확도를 평가하였다. 그 결과 1개의 최적 센서를 적용할 경우 예측 정확도 (R^2=0.918)의 상대적으로 큰 감소가 나타났지만, 3개의 최적 센서를 적용할 경우 (R^2=0.939) 센서 9개를 모두 이용한 경우 (R^2=0.950)와 유사한 예측 정확도를 나타냈다. 따라서 미래 공기 온기 예측을 위해 최소 3개의 최적 센서 위치를 적용할 것이 추천된다.

5장에서는 CFD 시뮬레이션의 연산 결과를 학습한 기계학습 모델을 이용하여 자연환기 온실의 지역별 환기량을 예측하였다. 학습 데이터의 생성을 위하여 10 가지 풍속, 7 가지 풍향, 3 가지 온실 환기창 타입에 대한 CFD 시뮬레이션을 수행하였다. 기계학습 모델은 MLR, SVR, Random forest, DNN이 개발되었으며, 각각의 기계학습 모델에 대한 hyper-parameter의 최적화를 수행하였다. 그 후 상대적으로 적은 수의 학습 데이터를 보완하기 위하여 Bootstrapping 기법을 적용하여 학습하였다. 그 결과 Bootstrapping 기법 적용 전 높은 예측 성능(R^2>0.9)을 나타내었던 RF, DNN, rbf kernel 함수를 이용하는 SVR 모델은 R^2가 감소한 반면, 상대적으로 예측 성능이 낮은(R^2<0.9) 모델인 MLR과 poly kernel 함수를 이용하는 SVR 모델의 경우 R^2가 개선되는 것으로 나타났다. 최종적으로 상대적으로 연산에 많은 시간을 필요로하는 CFD 시뮬레이션의 연산 case의 최소화를 위하여 CFD 시뮬레이션 case의 감소에 따른 기계학습

모델의 정확도를 평가하였다. 그 결과 대부분의 기계학습모델에서 풍향을 4 case만을 고려하여 총 120 case가 고려된 학습데이터에 대하여 가장 높은 정확도를 나타냈다. 뿐만 아니라 RMSE 값 또한 낮게 평가되어 예측값에 대한 오차가 가장 낮은 것으로 나타났다.

본 논문은 3가지 기계학습 모델(PCTO-ML, PETO-ML, PLV-CFD driven ML)로 구성된 온실 내부환경 예측 및 제어 시스템을 통하여 온실 내부의 현재 (PCTO-ML)와 미래 공기온도(PETO-ML)를 예측하며, 온실 내부의 지역적 환기율을 예측(PVL-CFD driven ML)함으로써 예측한 공기 온도의 적절한 제어가 가능하다는 것에 큰 의의가 있다고 판단된다. 특히 PCTO-ML 모델을 통하여 제안한 온실 내부 공기 온도 예측에 필요한 최적 센서 위치는 온실의 초기비용 및 유지보수 비용을 줄이므로써 농가의 소득 향상에 기여할 수 있을 것으로 기대된다. 또한 PVL-CFD driven ML 모델은 발생 빈도가 작아 실측이 어려운 환경조건에 대해서도 학습이 가능했기 때문에 모든 풍환경 조건에 대한 자연환기율을 예측할 수 있었다. 본 논문에서 제안한 온실 내부환경 예측 및 제어 시스템은 시스템의 개발과 운영을 위해 최소한의 온도 센서만을 필요로 하기 때문에 시스템 도입에 경제성을 가지며, 온도 센서가 설치되어 있는 국내 대부분의 온실에도 적용이 가능할 것으로 판단된다.

주요어: 전산유체역학, 환경 예측, 온실, 기계학습, 자연환기, 최적센서 위치

학 번: 2015-30469

감사의 글

먼저 석사, 박사과정 동안 학문적으로 큰 가르침을 주시고, 인간적으로도 잘 성숙할 수 있도록 가르침을 주신 이인복 지도교수님께 깊은 감사를 드립니다. 연구실에 처음 들어왔을 때부터 연구를 배우는 과정, 병역특례를 마치고 연구실에 돌아왔을 때, 박사논문을 준비하기까지 매번 반겨 주시고 좋은 기회를 마련해 주셔서 항상 감사드립니다.

더불어 학부와 대학원 과정의 오랜 기간 동안 학문 내외적으로 많은 도움과 가르침을 주신 교수님들께도 감사의 말씀을 전하고 싶습니다. 뵐 때마다 좋은 말씀해주시는 김준 교수님 감사드립니다. 항상 여러가지 조언을 주시고, 학문을 공부함에 있어 마음가짐을 가르쳐 주신 최진용 교수님 감사드립니다. 항상 웃는 모습으로 따뜻한 격려를 해주시는 김성배 교수님 감사드립니다. 많은 자리에서 여러가지 인싸이트를 주신 강문성 교수님 감사드립니다. 항상 웃는 모습으로 격려해 주시면서, 부족한 제 논문이 좋은 방향으로 갈 수 있도록 따끔한 지도로 심사위원장을 맡아 주신 손영환 교수님 감사드립니다. 고민이 있을 때 가까이서 들어주시고, 따뜻한 격려를 해주신 송인홍 교수님 감사드립니다. 항상 가까이서 응원과 조언을 주셨고, 먼 타지에서도 제 논문이 완성도를 갖출 수 있도록 많은 고견을 주신 최원 교수님 감사드립니다. 마지막으로 저로서는 도전이었고 그렇기에 두려움이 많았던 논문 주제에 대해서 많은 조언과 도움을 주신 서정욱 교수님,

조성균 박사님 진심으로 감사드립니다.

대학원 생활 동안 누구보다 많은 시간을 보내며 같이 고생하고 서로 응원했던 연구실 식구들에게도 감사드립니다. 졸업논문을 준비하는 힘든 시기에 많은 의지가 되고 다방면으로 조언을 주신 정득영 형님, 연구실에서 항상 든든한 동료이자 친구가 되어준 준규, 서툰 영어에도 즐겁게 대화하며 많은 도움을 준 크리스티나 모두 고마움을 전하며, 모두 멋진 논문을 완성한 것을 축하드립니다. 편하게 지내다가도 연구 내외적으로 진지한 이야기를 많이 나눈 정화, 후배이지만 누구보다 깊은 생각을 갖고 여러가지 힘이 되어준 효혁이, 궂은 일도 묵묵히 도맡으며 응원해준 솔뫼, 짓궂은 장난에도 항상 웃어주며 많은 도움을 준 다인이, 연구에 집중할 수 있도록 많은 신경을 써주시는 서하 선생님에게도 고마움을 전합니다. 연구실 생활 동안 가장 많은 격려와 충고, 기쁨과 슬픔을 같이한 김락우 교수님, 연구실 내외 생활 하나하나 세심하게 응원해준 여욱현 박사님, 후배이지만 배울점이 많은 이상연 박사님에게도 무엇보다도 큰 고마움을 전합니다. 졸업하신 홍세운 교수님, 서일환 교수님, 권경석 박사님, 하태환 박사님께도 아무것도 모르던 석사 때부터 여러가지로 많은 것을 알려주셔서 진심으로 감사드립니다. 그리고 연구실 생활을 같이 하였던 정수형, 예슬누나, 승노, 영배, 민형이, 관용이, 유미에게도 감사의 마음을 전합니다.

긴 대학원 과정 동안 의지가 되어준 학과 내 많은 선후배님들과 특히 09학번 동기들에게도 감사드립니다.

마지막으로 대학원 석사와 박사의 긴 기간 동안 공부에 집중하여

최선을 다할 수 있도록 물심양면으로 응원해주신 부모님과 형에게 깊은 감사의 마음과 사랑한다는 말을 전하고 싶습니다. 기다려주신 만큼 지금보다 더 사랑하고, 효도하는 아들과 동생이 되겠습니다.