



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

碩士學位論文

머신러닝을 활용한 IoT 작물재배시스템
내부온도 예측에 관한 연구

濟州大學校 大學院

메카트로닉스 工學科

許 守 美

2021 年 2 月

머신러닝을 활용한 IoT 작물재배시스템 내부온도 예측에 관한 연구




指導教授 康 喆 雄

許 守 美

이 論文을 工學 碩士學位 論文으로 提出함

2020 年 12 月

許守美의 工學 碩士學位 論文을 認淮함

審査委員長 임종환 
委 員 조계현 
委 員 장지우 

濟州大學校 大學院

2020 年 12 月

A Study on the prediction of Internal temperature
for IoT crop cultivation system using
machine learning methods

Su-Mi Heo
(Supervised by professor Chul-Ung Kang)

A thesis submitted in partial fulfillment of the requirement for
the degree of master of Mechatronics Engineering

2020. 12.

This thesis has been examined and approved.



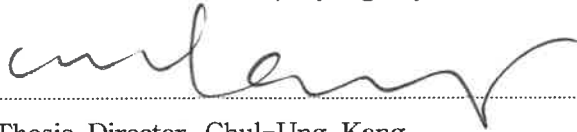
Thesis Committee Chair, Jong-Hwan Lim

Prof. of Jeju National University



Thesis Committee member, Kyung-Hyun Choi

Prof. of Jeju National University



Thesis Director, Chul-Ung Kang

Prof. of Jeju National University

.....
Date

Department of Mechatronics Engineering

GRADUATE SCHOOL

JEJU NATIONAL UNIVERSITY

목 차

제 1 장 서론	1
1. 연구 배경 및 필요성	1
2. 연구 목적과 방법	2
제 2 장 환경데이터 수집 및 전처리	4
1. 환경제어시스템 외형설계	4
2. 환경제어시스템 컨트롤러 설계	6
3. 환경제어시스템 통신프로토콜 설계	7
4. 모델 도출을 위한 데이터 수집	8
5. 모델 도출을 위한 데이터 처리	10
제 3 장 환경데이터 예측모델 도출	13
1. 예측모델 설정	13
2. 예측모델 도출방법	14
3. 예측 Model A 도출	15
4. 예측 Model B 도출	22
제 4 장 연구결과	32
1. 재배합 재질 영향 식 유무에 따른 모델 예측결과 비교	32
2. 재배합 재질 영향 식 유무에 따른 모델 성능평가 결과 비교	34
제 5 장 결론	40
참고문헌	42

LIST OF TABLES

Table 1 Environmental control system external components	11
Table 2 Components of the Environmental control system controller	13
Table 3 Example of WBML MQTT Server Information	13
Table 4 XML data address and parameters	15
Table 5 Prediction model Input variables	19
Table 6 Data set for Model A	20
Table 7 Data set for the influence formula by material	25
Table 8 Data set for Model B	27
Table 9 Data set for Model A & Model B	31

LIST OF FIGURES

Fig. 1 Environmental control system configuration diagram	4
Fig. 2 Environmental control system	5
Fig. 3 Environmental control system controller	6
Fig. 4 API provided by Agricultural Weather Information Service	9
Fig. 5 CSV file of System data	10
Fig. 6 Weather information data API	10
Fig. 7 Integrated dataset	11
Fig. 8 Train dataset	12
Fig. 9 Test dataset	12
Fig. 10 Multilayer Perceptrons, MLP architecture	15
Fig. 11 Model A Train dataset Input data(weather information data)	16
Fig. 12 Model A Train dataset Input data(system external temperature)	17
Fig. 13 Model A Train dataset Input data(controller temperature)	17
Fig. 14 Model A Test dataset Input data(weather information data)	18
Fig. 15 Model A Test dataset Input data(system external temperature)	18
Fig. 16 Model A Test dataset Input data(controller temperature)	19
Fig. 17 Model A Neural Network architecture	20
Fig. 18 Model A Weights and Bias values	20
Fig. 19 Comparison of Model A Predict data and Observation data	21
Fig. 20 Compare the details of Model A Predict data and Observation data	21
Fig. 21 The influence formula by material Dataset Input data	23
Fig. 22 Model B Train dataset Input data(weather information data)	25
Fig. 23 Model B Train dataset Input data(system external temperature)	26
Fig. 24 Model B Train dataset Input data(controller temperature)	26
Fig. 25 Model B Train dataset Input data(data predicted by the influence of materials)	27

Fig. 26 Model B Test dataset Input data(weather information data)	27
Fig. 27 Model B Test dataset Input data(system external temperature)	28
Fig. 28 Model B Test dataset Input data(controller temperature)	28
Fig. 29 Model B Test dataset Input data(data predicted by the influence of materials)	29
Fig. 30 Model B Neural Network architecture	30
Fig. 31 Model B Weights and Bias values	30
Fig. 32 Comparison of Model B Predict data and Observation data	31
Fig. 33 Compare the details of Model B Predict data and Observation data ·	31
Fig. 34 Comparison of Model A & Model B Predict data and Observation data	33
Fig. 35 Model A Correlation analysis results	35
Fig. 36 Model B Correlation analysis results	35
Fig. 37 Model A RMSE evaluation result	36
Fig. 38 Model B RMSE evaluation result	37
Fig. 39 Comparison of Model A & Model B RMSE evaluation result	38
Fig. 40 Compare the details of Model A & Model B RMSE evaluation result	39
Fig. 41 Observation data Outlier	39

ABSTRACT

This study focuses on efficient production, away from research that focused on automating cultivation facilities and increasing convenience. In this study, we discuss a model that predicts the internal temperature of the IoT crop cultivation system using machine learning, which derives the internal temperature prediction model of the crop cultivation system and ensures crop growth under optimal environmental conditions.

In this study, weather data were collected through external temperature data, internal temperature data, controller temperature data measured in crop farming systems and Agricultural Weather Information Service API before deriving the optimal prediction model. Afterwards, the collected data was refined to the appropriate form for the analysis. And preprocess of the data was carried out with process of parsing the information needed for the analysis and missing value. One of the learning methods of machine learning, Supervised Learning, is used to predict optimal internal temperatures by utilizing pre-processing data, and multi-layer perceptron algorithm is utilized as a neural network model to efficiently derive predictive models. In addition, two predictive models were classified according to the input variables. The effect of cultivation materials on the system, which is one of factors that affect internal temperatures data, is Formulated. And by comparing the results of predictive models that applied this formula with those of predictive models that did not apply the expression, we derive a model that predicts patterns similar to the temperature within the real system.

제 1 장 서론

1. 연구 배경 및 필요성

4차 산업혁명의 핵심기술인 사물인터넷(Internet of Things, IoT), 빅데이터(Big Data), 인공지능(Artificial Intelligence, AI) 등이 다양한 산업분야에 적용되고 있다. 사람이 반복적으로 하던 작업을 기계가 도우는 보조적인 측면에서의 개념에서 벗어나 데이터 및 알고리즘을 적용시켜 작업의 지능화를 통해 노동력과 사료비를 절감시키며 제품을 효율적으로 생산시키는 등의 변화를 보이고 있다.

[1]

과학기술예측조사(2012~2035)에 따르면 지구 평균온도의 지속적인 상승으로 인한 지구온난화의 심화, 대규모 자연재해 발생빈도 증가에 따른 이상기후현상 증가, 이상강우 기후변화 및 개발로 인한 수자원의 감소와 곡물가격의 지속적 상승으로 인한 식량 부족의 심화 등을 메가트렌드로 제시하고 있었다. 이를 통해 농업내부의 여건과 미래 환경변화는 스마트농업의 필요성을 더욱 강조하고 있으며, 스마트농업을 통해 국내 농업의 양적 질적 성장이 필요한 시점임을 알 수 있었다.[2]

이에 따라 농림수산식품부는 ‘농수축산분야 u-IT사업’을 통해 농업 경쟁력 향상을 목표로 하여 생산 정밀화 모델을 개발하며 기술개발도 진전되는 모습을 보이고 생산성 향상 등의 성과가 가시화 되었었다. 그러나 국내 기술은 R&D 수준에 머무르며 시설현대화와 제대로 연계되지 못하였다. 스마트팜 활성화를 위해 다양한 기술들을 적용하여 스마트팜 확산의 장애요인을 해소하고 효율적인 농작물 생산을 가능하도록 하는 연구의 필요성이 나타나고 있었다.[3]

이에 따라 (오정원, 2018)은 과수 생육환경 최적화 및 과수 수확시기 예측에 머신러닝 기술을 적용하여 기존 스마트팜 시스템에 수동제어를 최소화하고 과수 수확시기를 예측하는 연구를 진행하였다.[4] 또한 IoT기반 스마트 제어 운영 시스템을 연구한 (서신림, 2015)는 버섯 인공재배 시 생육상태에 큰 영향을 미치는

환경요소인 온도, 습도, CO₂ 등 생육환경 데이터를 측정하고 이를 실시간 모니터링 할 수 있는 IoT기반의 스마트 제어 시스템을 운영하여 최적의 벚꽃 생육환경을 유지시켰다.[5]

이와 같이 시설하우스 형태의 구조물 내 기술을 접목하여 온습도 및 풍향 등을 모니터링하고 이를 제어하는 단계의 연구에서 벗어나 최적의 작물생장환경을 구축하기위해 4차 산업혁명기술과 농업을 융합시켜 최적의 작물 재배환경을 예측하는 미래형 농업에 관련된 연구의 필요성이 나타나고 있다.[6]

2. 연구 목적과 방법

본 연구는 기존 재배시설 자동화 및 편리성 증대에 집중했던 연구에서 벗어나 효율성 있는 생산을 위해 머신러닝(Machine Learning)을 활용하여 작물재배 시스템 내부온도 예측모델을 도출해내 최적의 환경조건에서 작물의 생육이 이루어지도록 하는 머신러닝(Machine Learning)을 활용한 IoT 작물재배시스템 내부 온도 예측에 관한 연구를 하였다.

본 연구에서는 최적의 예측모델을 도출해내기 앞서 작물재배시스템에서 측정된 외부온도데이터, 내부온도데이터, 제어기온도데이터와 농업기상정보서비스 API를 통해 기상데이터를 수집하였다. 수집된 데이터를 분석에 알맞은 형태로 정제시킨 뒤 분석에 있어 필요한 정보들을 파싱하고 결측 데이터를 처리하며 분석에 적절한 형태로 데이터 전처리 작업을 진행하였다. 전처리 작업이 완료된 데이터를 가지고 최적의 시스템 내부온도를 예측하기위해 머신러닝(Machine Learning)의 학습방법 중 하나인 지도학습(Supervised Learning)의 회귀(Regression)방식을 사용하였고 효율적으로 예측모델을 도출하기위해 신경회로망 모델로 다층퍼셉트론(Multilayer Perceptrons, MLP) 알고리즘 구조를 활용하였다. 또한 입력변수에 따라 예측모델을 두 가지로 분류하였다. 시스템 내부온도데이터에 영향을 미치는 요소 중 하나로 선정한 재배함 재질이 시스템에 미치는 영향 식을 도출해내어 식을 적용시킨 예측모델과 식을 적용시키지 않은 예측모델의 결과를 비교하여 실제 시스템내부온도와 유사한 패턴의 데이터를 예측해내

는 모델을 도출해냈다.

본 논문은 총 5장으로 구성되어 있으며 각 장의 내용을 요약하면 다음과 같다.

제 2장은 작물재배시스템 및 농업기상정보서비스에서 데이터를 수집하고 전처리한 뒤 이를 학습데이터와 훈련데이터로 분할하는 과정을 기술하였다. 제 3장은 준비한 데이터를 통해 머신러닝 라이브러리를 사용하여 예측모델을 훈련하고 예측데이터를 도출해내는 과정과 재배함 재질이 시스템에 미치는 영향에 관한 식을 도출해내는 과정을 기술하였다. 제 4장은 재배함 재질이 시스템에 미치는 영향에 관한 식을 적용시킨 예측모델과 식을 적용시키지 않은 예측모델을 비교 분석하여 최적의 예측모델을 도출해내었다. 제 5장은 본 연구의 결론 및 향후 연구 방향을 제시하였다.

제 2 장 환경데이터 수집 및 전처리

최적의 시스템 내부온도 예측모델을 도출해내기 앞서 환경데이터를 수집하기 위해 작물재배시스템을 설계하였다. IoT 기술을 접목하여 생육환경을 제어하고 내부 외부 환경 센서 데이터를 수집해 LoRa통신망을 활용하여 데이터베이스에 저장하는 시스템을 제작하였고 이를 통해 시스템외부온도, 시스템내부온도, 제어기온도 데이터를 수집하였다. 다음은 작물재배환경제어시스템의 구성도를 나타낸 것이다.

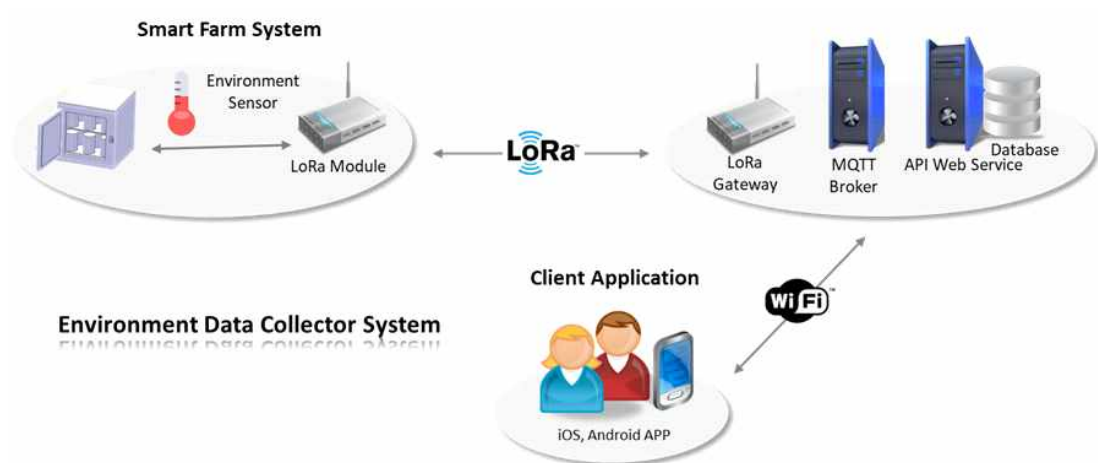


Fig. 1 Environmental control system configuration diagram

1. 환경제어시스템 외영설계

작물재배시스템은 내부의 온도를 일정하게 유지시켜줘야 하므로 저온저장고 형태로 제작하였다.

Table 1 Environmental control system external components

Categorize	Standard
External size	2,500mm × 1,000mm × 1,500mm
Internal size	2,400mm × 900mm × 1,400mm
External material	Sandwich panel 50T (Styrofoam stuffing inside) & External insulation
Finishing material	L-BAR molding
Processing method	<ul style="list-style-type: none"> - Use of rivets, use polyurethane foam for seams to improve heat insulation - The internal and external finishes are waterproofed using transparent silicone - Process the panel into a prefab



Fig. 2 Environmental control system

2. 환경제어시스템 컨트롤러 설계

환경제어시스템 컨트롤러는 측정파트와 제어파트로 나뉘며, 측정파트는 내부 온도센서 1개, 외부 온도센서 1개의 데이터를 처리하고 수집하였다. 센서 데이터 수집 및 전송을 위한 RS232 회로, 전원공급 및 제어를 위한 Regulator 회로, 동작상태를 확인하기 위한 LCD 회로 등 전체 동작에 필요한 필수 구성 회로를 통해 측정파트를 설계하였다.

제어파트는 냉난방기, 환풍기 등 총 2개의 Actuator의 전원공급 및 차단 제어로 구성하였다. 설정 값을 입력시켜 제어파트를 통해 생육환경을 제어하고 측정파트를 통해 시스템의 환경데이터를 수집하는 일련의 과정을 거치며, 이렇게 수집된 데이터는 LoRa 통신망을 활용하여 시스템에서 서버의 데이터베이스로 데이터를 전송하여 저장하게 되고 이를 사용자가 데이터베이스를 통해 확인할 수 있도록 하였다.



Fig. 3 Environmental control system controller

Table 2 Components of the Environmental control system controller

Categorize	Standard	Quantity
heating and cooling control	RFH-2000 Small heater(220V-400W)	2EA
	MWA04BTW WINIA Small air cooling (220V-550W)	1EA
	Temperature sensor for heating and cooling control	1EA
Ventilator	SHINILVentilator (220V-36W)	2EA
Temperature/humidity sensor	STA Temperature / humidity sensor 3V ~ 12V	2EA
Control board	AT Mega Process Enter the sensor value, environmental control value output	1EA

3. 환경제어시스템 통신프로토콜 설계

IoT 표준 프로토콜인 MQTT 프로토콜을 활용하여 데이터를 서버로 전송, 전송된 데이터를 데이터베이스로 저장하고 사용자가 확인할 수 있도록 하였다. 수집된 환경 센서 데이터를 통합모니터링하고 환경을 제어할 수 있도록 다음과 같이 통신 프로토콜을 설계하였다.

Table 3 Example of WBML MQTT Server Information

Categorize	MQTT Protocol	
Topic	WBML/ poriacocos_box/ [GATEWAY_ID]/ [UUID]/ down	Device -> Server
	WBML/ poriacocos_box/ [GATEWAY_ID]/ [UUID]/ up	Server-> Device
Message	ST command function type dataset ET	-

4. 모델 도출을 위한 데이터 수집

4.1 환경제어시스템 데이터 수집

작물재배시스템의 내부, 외부, 제어기 온도 데이터 수집을 진행하였다. 내부 온도센서 1개, 외부 온도센서 1개, 제어기 온도 데이터를 수집하였다.

파이썬(Python) 라이브러리인 pandas를 활용하여 csv파일 형태의 내부온도, 외부온도, 제어기 온도 데이터를 DataFrame형태로 읽어 들였다.

4.2 농업기상정보서비스 API 기상데이터 수집

지역평균기상데이터를 사용하기 위해 활용한 기상정보 데이터는 농업기상정보 서비스에서 2016년부터 2020년까지 일별 10분주기로 5년간의 제주도 김녕리 온도 기상데이터를 수집하였다.

농업기상정보서비스에서 제공하는 API(Application Programming Interface)는 XML(Extensible Markup Language) 형태로 데이터를 제공하고 있었다.

XML 형태의 데이터를 호출하기 위해 파이썬(Python) 라이브러리인 urllib.request를 활용하였다. 인증키와 사용자아이디를 XML 데이터 주소에 추가시켜 농업기상정보서비스에서 조회일자의 10분 자료를 호출하였다.

Table 4 XML data address and parameters

Categorize	Data
XML data address	<code>http://weather.rda.go.kr/openapi/inquiry_openapi_xml.jsp?mberid=UserID&regist_ky=authentication key&period=day&day=Year-Month-Day&gbn=1</code>
Parameter	mberid - UserID, regist_ky - authentication key, period - day, day - YYYY-MM-DD(Year-Month-Day), gbn - 1

This XML file does not appear to have any style information associated with it. The document tree is shown below.

```
<Root>
  <Info>
    <stncode>695971A001</stncode>
    <stnname>제주시 길남리</stnname>
    <day>2020-10-25 00:10</day>
    <temp_150>14.1</temp_150>
    <tmprt_50>13.9</tmprt_50>
    <tmprt_400>13.8</tmprt_400>
    <hd_150>59.1</hd_150>
    <hd_50>-%</hd_50>
    <hd_400>-%</hd_400>
    <wd_300>307.0</wd_300>
    <wd_150>-%</wd_150>
    <wd_1000>0.0</wd_1000>
    <arvly_300>3.0</arvly_300>
    <arvly_150>-%</arvly_150>
    <arvly_1000>0.0</arvly_1000>
    <afp>0.0mm</afp>
    <afv>mm</afv>
    <sunshnTime>0.0hr</sunshnTime>
    <solradQy>0.0MJ/m</solradQy>
    <dwcnTime>0.0hr</dwcnTime>
    <pnwgtp>hr</pnwgtp>
    <frfrTp>11.9</frfrTp>
    <udgrHeattCndctvt>-100.0</udgrHeattCndctvt>
    <udgrTp_10>18.5</udgrTp_10>
    <udgrTp_5>-%</udgrTp_5>
    <udgrTp_20>27.7</udgrTp_20>
    <soilMitr_10>31.2</soilMitr_10>
    <soilMitr_100cst/>
    <soilMitr_20>40.8</soilMitr_20>
    <soilMitr_20cst>-%</soilMitr_20cst>
    <soilMitr_30>-%</soilMitr_30>
    <soilMitr_30cst>-%</soilMitr_30cst>
  </Info>
```

Fig. 4 API provided by Agricultural Weather Information Service

제공 데이터 중 분석 시 활용할 데이터는 온도 데이터이므로 이를 위해 파이썬(Python) 라이브러리인 BeautifulSoup을 활용하였다. 이후 온도 데이터를 파싱하여 날짜데이터를 인덱스로 지정하여 연도별로 데이터를 정리하였다.

16 ~ 20년 5년간 10분주기의 데이터의 평균치를 구하기 위해 최솟값과 최댓값을 제외시킨 나머지 데이터의 평균을 구하였다.

5. 모델도출을 위한 데이터 처리

머신러닝(Machine Learning) 모델을 도출하기에 앞서 수집한 환경제어시스템 데이터와 농업기상정보서비스 API 기상데이터를 시간 순으로 정렬한 후 결측 데이터를 처리하였다. 이후 전처리된 데이터를 통합하여 하나의 DataFrame에 저장하였다. 통합된 데이터셋을 ‘단순홀드아웃 검증(hold-out validation)’ 분할방법을 활용하여 학습데이터셋(train dataset) 및 테스트데이터셋(test dataset) 두 가지로 분할하여 모델 도출을 위한 데이터 준비를 하였다.

5.1 데이터 전처리 및 통합

1분 주기로 입력되어있던 각 csv파일의 온도데이터를 10분주기로 데이터 과싱을 진행하고 날짜데이터를 인덱스로 지정하여 시간 순으로 데이터를 정렬하였다. 누락된 결측데이터는 바로 앞 데이터인 1분 전 데이터를 활용하여 결측치 대체를 진행하였다. 이후 시간 순으로 정렬한 기상정보데이터와 하나의 DataFrame 형태로 통합하여 저장하였다.

C	D
measured_value	measured_date
26	2020-10-25 0:01
26.3	2020-10-25 0:03
26.3	2020-10-25 0:05
26.3	2020-10-25 0:07
26.2	2020-10-25 0:09
26	2020-10-25 0:11
25.9	2020-10-25 0:13
25.8	2020-10-25 0:15
26.1	2020-10-25 0:17
26.3	2020-10-25 0:19
26.3	2020-10-25 0:21
26.2	2020-10-25 0:23
26.1	2020-10-25 0:26
26	2020-10-25 0:28

Fig. 5 CSV file of System data

This XML file does not appear to have any style information.

```

<?xml version="1.0" encoding="UTF-8" ?>
<root>
  <info>
    <stncode>695971A001</stncode>
    <stnname>제주시 김녕리</stnname>
    <day>2020-10-25 00:10</day>
    <temp_150>14.1</temp_150>
    <temp_50>13.9</temp_50>
    <temp_400>13.8</temp_400>
    <hd_150>59.1</hd_150>
    <hd_50>X</hd_50>
    <hd_400>X</hd_400>
    <wd_300>307.0</wd_300>
    <wd_150>X</wd_150>
    <wd_1000>0.0</wd_1000>
    <arvity_300>3.0</arvity_300>
    <arvity_150>X</arvity_150>
    <arvity_1000>0.0</arvity_1000>
    <atp>0.0</atp>
    <atv>X</atv>
    <sunshTime>0.0</sunshTime>
    <solrad0y>0.0</solrad0y>
    <descTime>0.0</descTime>
    <pnegTp>X</pnegTp>
    <trirTp>11.9</trirTp>
    <uderHeattCndctvt>-100.0</uderHeattCndctvt>
    <udgrTp_10>18.5</udgrTp_10>
    <udgrTp_5>X</udgrTp_5>
    <udgrTp_20>27.7</udgrTp_20>
    <soilMitr_10>31.2</soilMitr_10>
    <soilMitr_100>X</soilMitr_100>
    <soilMitr_20>40.8</soilMitr_20>
    <soilMitr_200>X</soilMitr_200>
    <soilMitr_30>X</soilMitr_30>
    <soilMitr_300>X</soilMitr_300>
  </info>
</root>

```

Fig. 6 Weather information data API

	weather	system_in	system_out	system
10-25 00:10	15.00	26.00	14.10	26.10
10-25 00:20	15.13	26.30	14.10	26.50
10-25 00:30	15.20	25.80	14.00	27.80
10-25 00:40	15.00	26.20	14.10	26.20
10-25 00:50	14.93	26.20	14.20	26.70
10-25 01:00	14.90	25.90	14.30	26.30
10-25 01:10	14.83	26.20	14.30	26.30
10-25 01:20	14.93	26.00	14.40	26.60
10-25 01:30	15.07	26.00	14.50	26.50
10-25 01:40	15.17	26.20	14.60	26.40
10-25 01:50	15.03	26.20	14.60	27.30
10-25 02:00	14.83	25.90	14.70	26.30
10-25 02:10	14.73	26.30	14.80	26.30
10-25 02:20	15.00	26.00	14.80	26.50
10-25 02:30	15.23	26.10	14.80	26.10

Fig. 7 Integrated dataset

5.2 단순홀드아웃 검증(hold-out validation) 분할

다양한 훈련검증방법 중 ‘단순홀드아웃 검증(hold-out validation)’이란 전체 데이터셋에서 일정부분을 학습데이터셋(train dataset)으로 지정하고 나머지부분의 데이터를 테스트데이터셋(test dataset)으로 지정하여 훈련하는 검증 방법이다. 이는 학습데이터셋(train dataset)으로 학습이 마친 뒤 정답(Label)이 정해지지 않은 데이터인 테스트데이터셋(test dataset)을 적용하여 학습된 알고리즘이 정확한 예측 값을 도출해냈는지 측정할 수 있는 방법이다.[7]

	train_weather	train_system_in	test_weather	test_system_in	
10-25 00:10	15.00	26.00	11-02 00:10	11.60	26.10
10-25 00:20	15.13	26.30	11-02 00:20	11.73	26.40
10-25 00:30	15.20	25.80	11-02 00:30	11.67	26.00
10-25 00:40	15.00	26.20	11-02 00:40	11.57	26.40
10-25 00:50	14.93	26.20	11-02 00:50	11.50	26.00
10-25 01:00	14.90	25.90	11-02 01:00	11.57	26.40
10-25 01:10	14.83	26.20	11-02 01:10	11.73	26.00
10-25 01:20	14.93	26.00	11-02 01:20	11.60	26.50
10-25 01:30	15.07	26.00	11-02 01:30	11.50	26.00
10-25 01:40	15.17	26.20	11-02 01:40	11.43	26.40
10-25 01:50	15.03	26.20	11-02 01:50	11.47	26.10
10-25 02:00	14.83	25.90	11-02 02:00	11.47	26.40
10-25 02:10	14.73	26.30	11-02 02:10	11.20	26.00
10-25 02:20	15.00	26.00	11-02 02:20	11.33	26.30
10-25 02:30	15.23	26.10	11-02 02:30	11.43	26.00

Fig. 8 Train dataset

Fig. 9 Test dataset

전체 데이터 10월 25일 00시 10분부터 11월 09일 23시 50분까지의 데이터 중 50%인 10월 25일 00시 10분부터 11월 01일 23시 50분까지 데이터를 학습데이터셋(train dataset)로 설정하고, 나머지 50%를 11월 02일 00시 10분부터 11월 09일 23시 50분까지의 데이터를 테스트데이터셋(test dataset)로 설정하였다.

제 3 장 환경데이터 예측모델 도출

작물재배시스템에 영향을 미치는 요소를 고려하고 입력변수에 따라 모델을 두 가지로 분류하여 선정하였다. 이후 머신러닝(Machine Learning) 라이브러리인 Tensorflow를 사용하여 두 예측모델을 구조화시키고 예측모델을 훈련한 뒤 예측데이터를 도출해냈다. 이후 재배함 재질이 시스템에 미치는 영향 식을 적용시킨 예측모델과 식을 적용시키지 않은 예측모델을 비교 분석하였다.

1. 예측모델 설정

Table 5 Prediction model Input variables

Categorize	Model A	Model B
Input variables	Weather information data, System external temperature, Controller temperature	
	-	System internal temperature predicted by the influence of materials

농업생산에 가장 큰 영향을 끼치는 다양한 기후요소인 기온, 강수량, 바람 등 중 온도와 관련된 요소인 기상데이터, 시스템외부온도, 제어기온도를 변수로 입력받으면 시스템 내부온도를 예측하는 모델을 Model A로 설정하였다,

비닐하우스 피복재질에 따른 비닐하우스 내부의 환경변화가 있듯이 재배함 재질에 영향을 받아 시스템 내부의 환경데이터에도 변화가 있을 것이라 추측하고, 입력변수 기상데이터, 시스템외부온도, 제어기온도 이외의 추가로 재배함 재질의 영향으로 인해 예측되는 시스템 내부온도데이터를 추가 변수로 입력받으면 시스템 내부온도를 예측하는 모델을 Model B로 설정하였다.

재배함 재질이 미치는 영향을 식으로 도출해내기 위해 시스템 제어가 없는 상태에서 측정된 데이터를 활용하였다. 이는 시스템 설정온도가 없는 환경에서 측정된 내부 온도데이터, 외부 온도데이터를 의미한다.

2. 예측모델 도출방법

예측모델을 도출하기 위해 4차 산업혁명 핵심기술 중 하나인 머신러닝(Machine Learning)을 활용하였다. 머신러닝(Machine Learning)이란 기계학습이라 부르기도 하며 인공지능(Artificial Intelligence, AI)을 구현하기 위한 한 방법으로서 컴퓨터가 인간이 학습하는 것과 같이 알고리즘과 프로그램을 이용한 데이터 학습을 통해 의사결정을 하거나 새로운 정보를 도출하는 것을 말한다.[8]

2.1 예측모델 학습방식

예측모델을 학습하는 방식으로는 지도학습(Supervised Learning)의 회귀(Regression)를 사용하였다. 지도학습(Supervised Learning)이란 머신러닝(Machine Learning)의 데이터 학습 방법 중 하나로 특정 입력(Input)에 따른 정답(Label)이 정해진 데이터를 학습시켜 컴퓨터가 스스로 결과를 예측하거나 결정하도록 하는 방식을 의미한다. 회귀(Regression)란 지도학습(Supervised Learning)의 방법 중 예측 값으로 연속적인 숫자 값을 도출해내는 방식을 의미한다.

2.2 예측모델 알고리즘

예측모델을 학습시키기 위한 알고리즘으로는 다층퍼셉트론 신경회로망을 사용하였다. 다층퍼셉트론(Multilayer Perceptrons, MLP)이란 입력층(Input layer)과 출력층(Output layer) 사이에 은닉층(Hidden layer)을 1개 이상의 두어 비선형적인 영역으로 분리되는 데이터에 대해서도 학습이 가능하도록 한 알고리즘이다. 다층퍼셉트론(Multilayer Perceptrons, MLP)은 입력층(Input layer)을 통해 받은 데이터를 중간단계를 구성하는 다수의 은닉층(Hidden layer)을 통해서 이를 여러 단계에 거쳐 학습을 진행한 뒤 출력층(Output layer)을 통해 최종적으로 예측 값

을 도출해낸다. 이러한 예측 값을 도출해내는 신경회로망 구조는 아래와 같이 표현할 수 있다.[9]

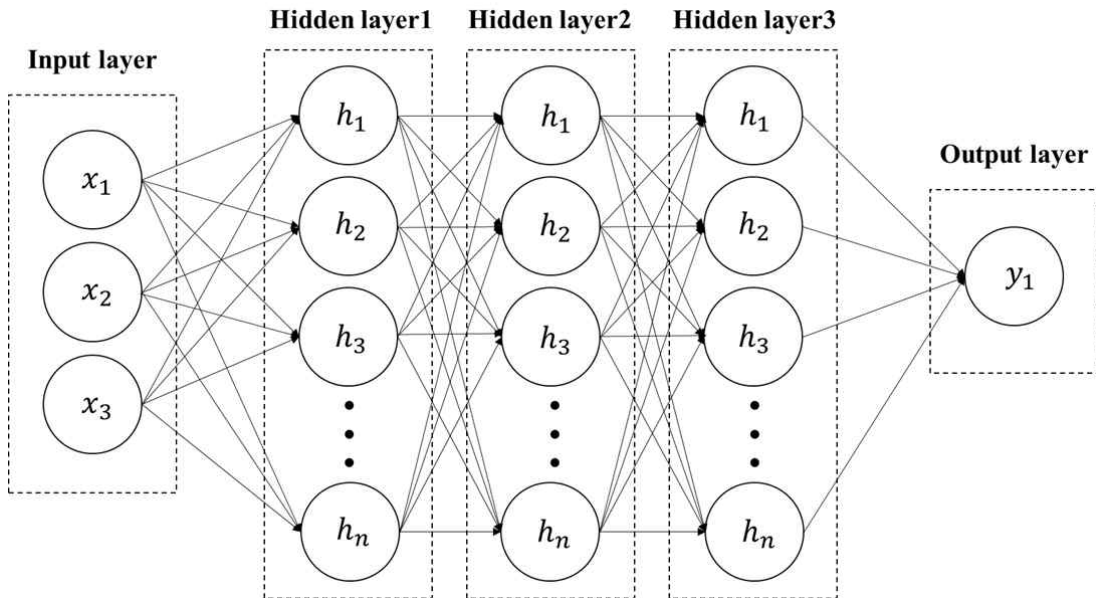


Fig. 10 Multilayer Perceptrons, MLP architecture

3. 예측 Model A 도출

시스템 내부온도 예측모델을 효율적으로 도출하고 훈련속도를 높이기 위해서 구글에서 오픈소스로 공개한 머신러닝(Machine Learning) 라이브러리인 Tensorflow를 설치하여 사용하였다. 이는 머신러닝 알고리즘을 구현하고 실행하기 위한 프로그래밍 인터페이스로서 확장이 용이하다.[10] 또한 모델을 효율적으로 구성하기 위해서 Tensorflow의 고수준의 API인 Keras 모듈을 활용하여 모델을 도출하였다.

3.1 예측 Model A 도출을 위한 데이터 준비

먼저 수집한 데이터를 학습데이터(train dataset) 및 테스트데이터(test dataset) 두 가지로 분할시켜 전처리시킨 데이터를 데이터셋으로 준비하였다.

Table 6 Data set for Model A

Categorize	Train dataset	Test dataset
Start date	10-25 00:10	11-02 00:10
End date	11-01 23:50	11-09 23:50
Input data	Weather information data, System external temperature, Controller temperature	
Target data	System internal temperature	-
Output data	-	Predicted system internal temperature

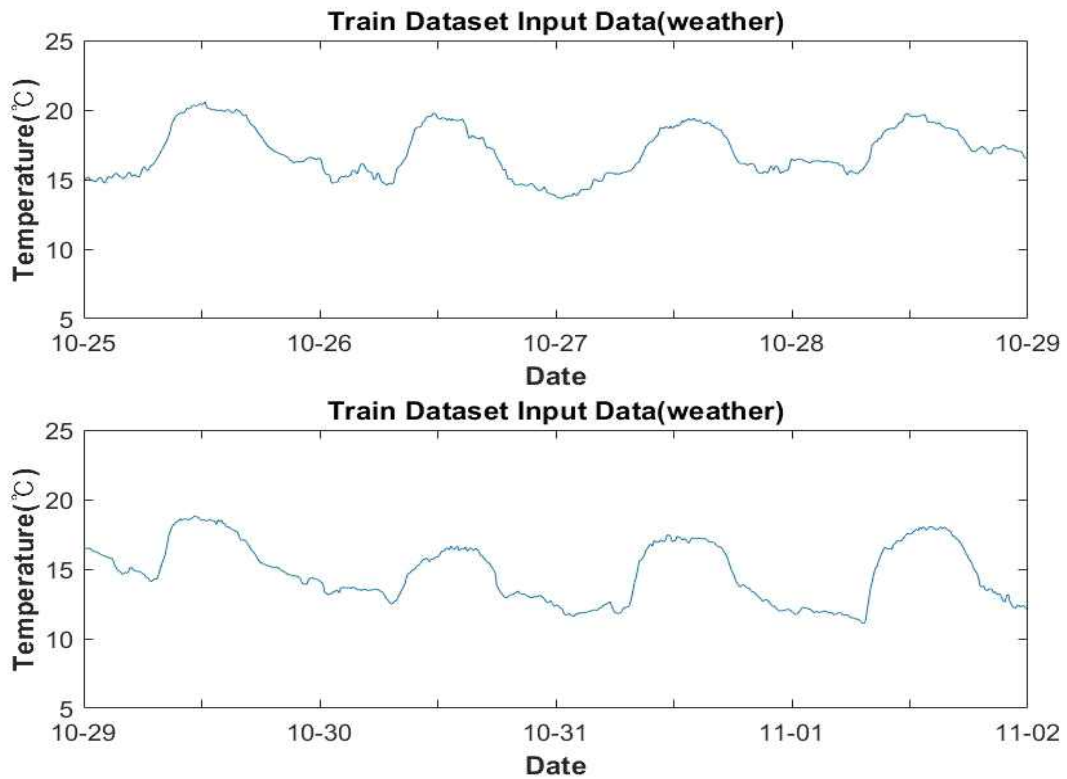


Fig. 11 Model A Train dataset Input data(weather information data)

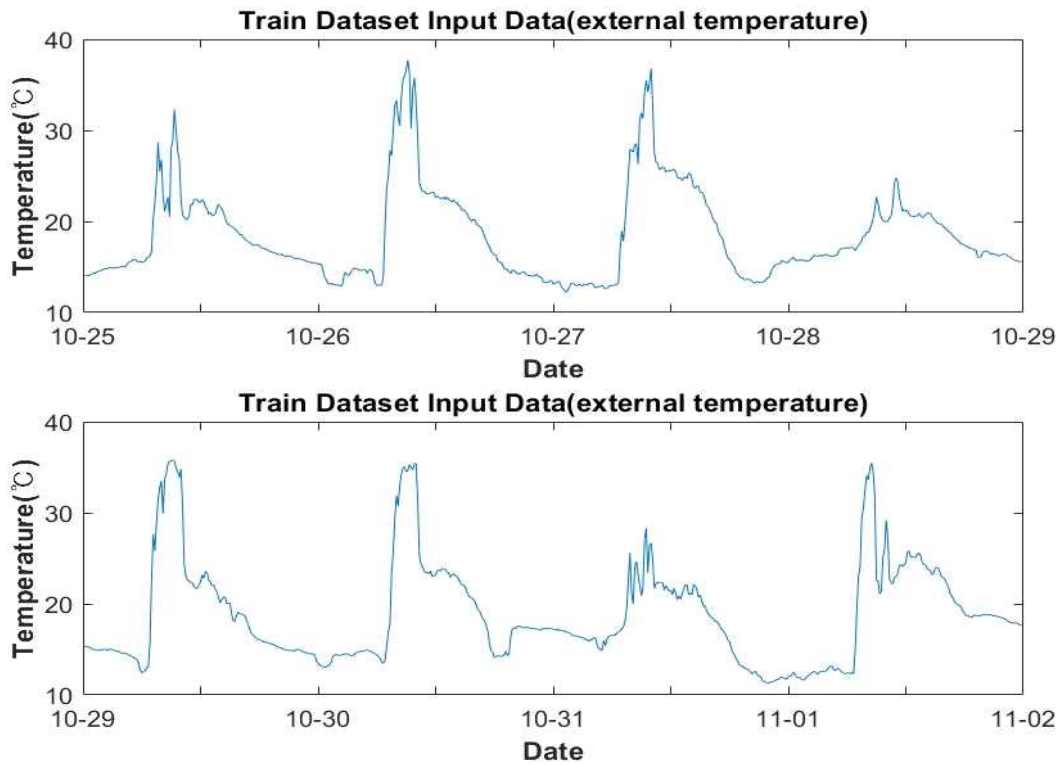


Fig. 12 Model A Train dataset Input data(system external temperature)

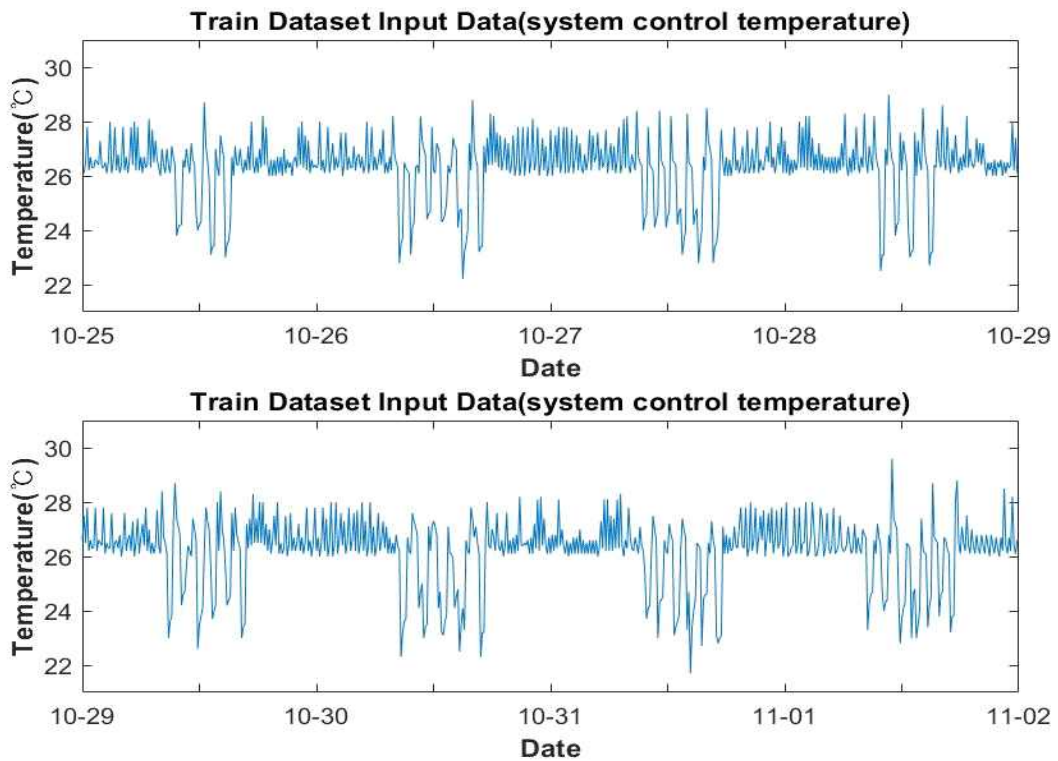


Fig. 13 Model A Train dataset Input data(controller temperature)

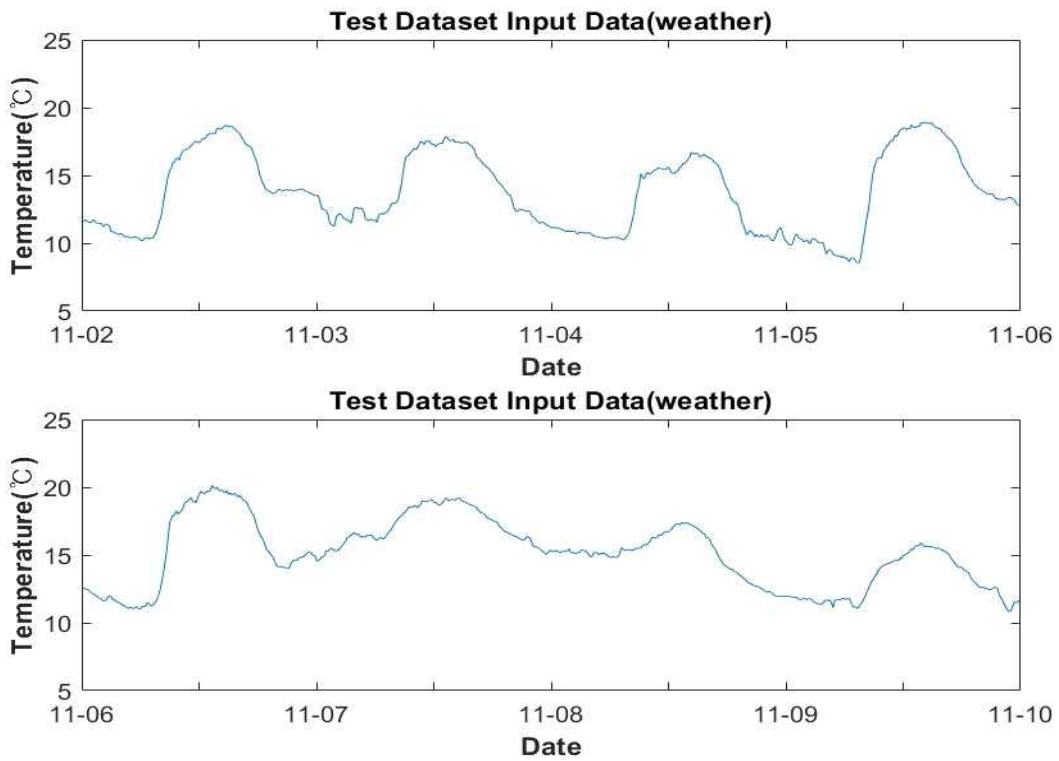


Fig. 14 Model A Test dataset Input data(weather information data)

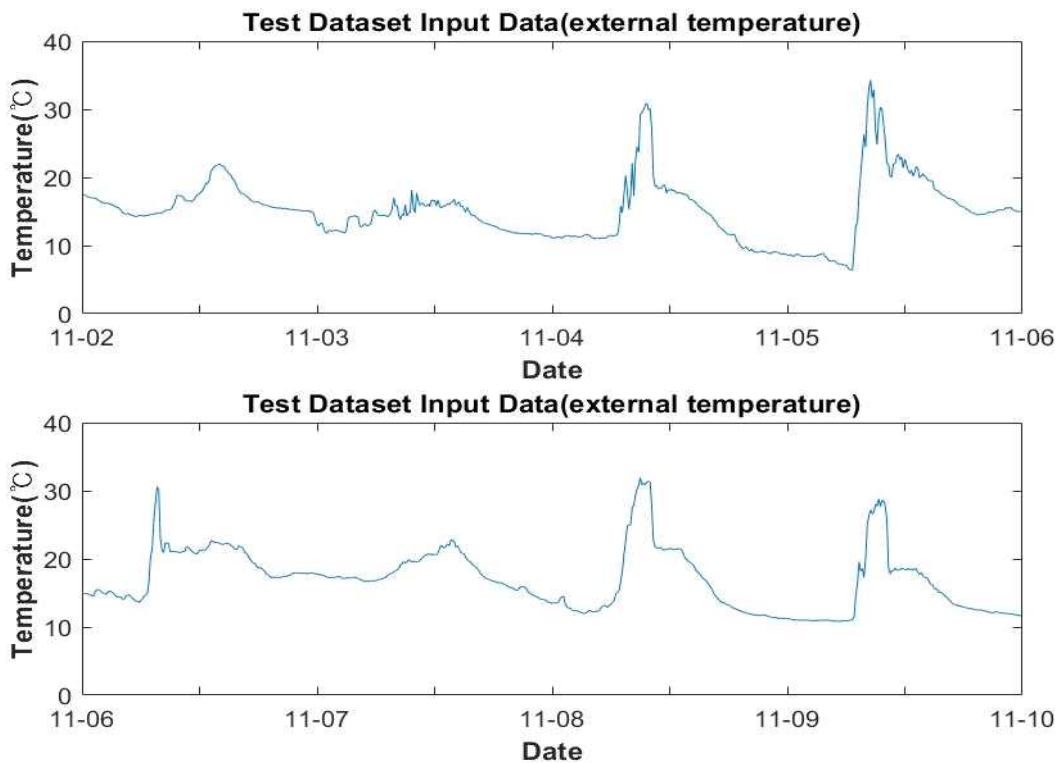


Fig. 15 Model A Test dataset Input data(system external temperature)

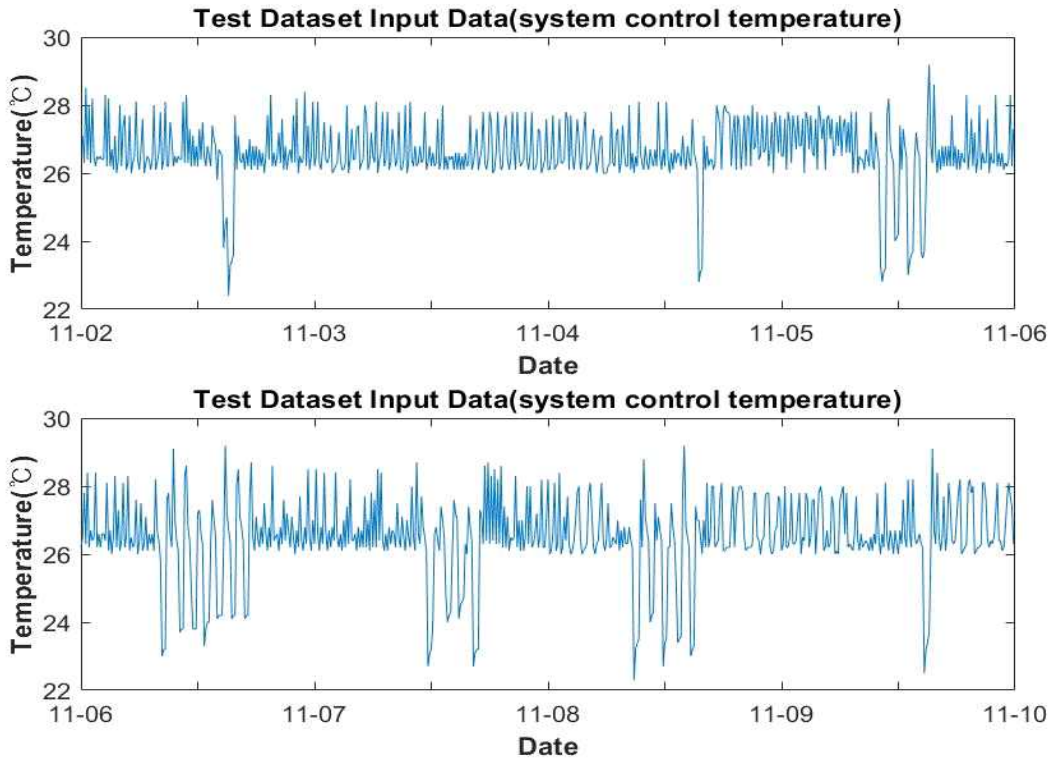


Fig. 16 Model A Test dataset Input data(controller temperature)

3.2 예측 Model A 구성

예측모델을 구성하기 위해서 텐서를 입력받고 출력하는 Keras의 함수형 API를 사용하여 모델을 만들었다.[11]

입력텐서(Input Tensor)로 기상청 평균기온, 시스템 외부온도, 제어기 온도 3개를 입력시킨 뒤 함수처럼 사용하기 위해 층 객체를 Dense를 통해 만들었다. 이후 텐서와 함께 층을 호출하면 은닉층 노드 100개로 구성된 은닉층(Hidden layer) 3개에서 이를 여러 단계에 거쳐 학습을 진행한 뒤 출력텐서(Output Tensor)가 시스템 내부온도를 반환하도록 하는 구조로 구성하였다.

이후 Model 클래스의 객체를 통해 함수형 API로 구조화시킨 모델의 입력텐서(Input Tensor)에서 출력텐서(Output Tensor)를 도출하기까지 필요한 모든 층의 출력을 추출한 뒤 반환하였다.

3.3 예측 Model A 훈련 및 도출

이후 구조화된 모델의 전체데이터를 반복하여 학습하는 횟수인 epochs를 100으로 지정하고 다층퍼셉트론의 알고리즘을 통해 다음의 그림에 나오는 모델의 가중치 및 편향치를 계산해내는 훈련을 시킨 뒤 예측 데이터를 도출하였다.

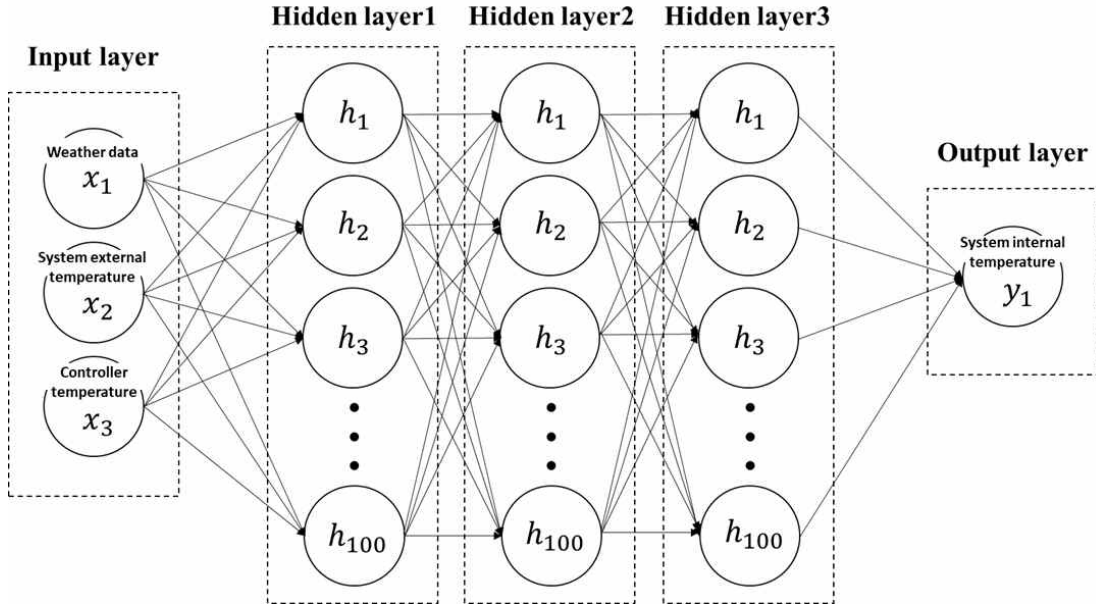


Fig. 17 Model A Neural Network architecture

```
[array([[ -9.00685042e-02,  2.16138232e-02,  7.87839517e-02,
        -1.38494531e-02, -4.96735424e-02,  2.04572394e-01,
         1.77888170e-01,  8.89839083e-02, -2.67906100e-01,
        -2.19781280e-01,  3.68746296e-02, -1.04326140e-02,
        -1.60553008e-01, -4.91895862e-02,  2.57924921e-03,
        -7.65559077e-02, -2.25045756e-01,  8.76750574e-02,
         1.11974113e-01, -1.90031528e-01,  1.84445567e-02,
        -3.09215169e-02,  2.99530085e-02, -4.27409373e-02,
         4.80752587e-02,  1.13088831e-01, -1.53332919e-01,
         1.54381141e-01, -6.12926409e-02, -2.47728646e-01,
        -2.29653597e-01, -1.03049427e-02,  1.33649379e-01,
        -2.48525113e-01,  1.29902393e-01,  2.98752487e-02,
        -1.22617036e-01, -1.89409792e-01,  1.16819337e-01,
        -1.77944392e-01, -9.08086672e-02,  1.26245081e-01,
        -2.46188149e-01,  9.69378054e-02, -8.30251649e-02,
        -2.03531440e-02,  3.82539928e-02, -1.62449241e-01,
         8.93295929e-02,  1.17257401e-01, -2.36341402e-01,
        -2.09440961e-02,  5.87656461e-02, -1.97951570e-02,
        -1.60609320e-01,  1.65186644e-01,  3.80670764e-02,
```

Fig. 18 Model A Weights and Bias values

훈련을 통해 Model A가 예측한 시스템 내부온도데이터와 실제 시스템 내부 온도데이터를 시각화하여 비교하였다. Model A가 예측한 데이터는 평균 25.14°C로 시스템내부온도의 평균과 0.77°C의 편차를 나타냄을 알 수 있었다.

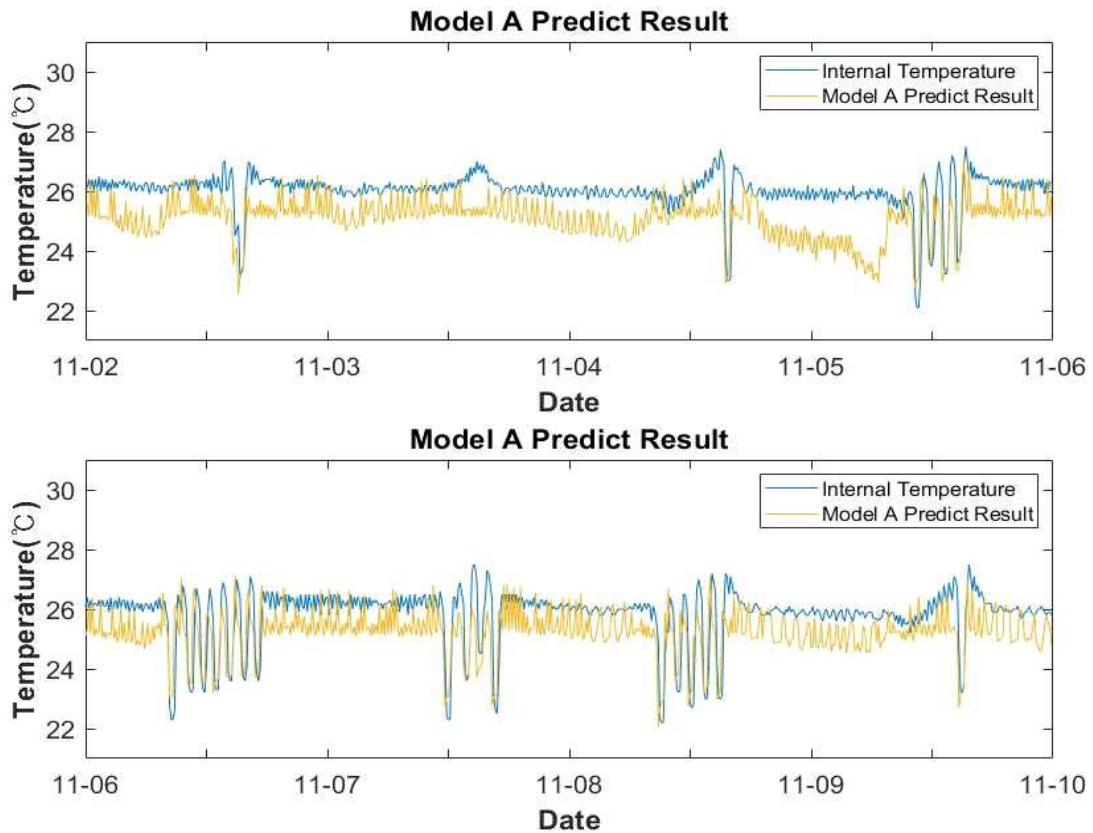


Fig. 19 Comparison of Model A Predict data and Observation data

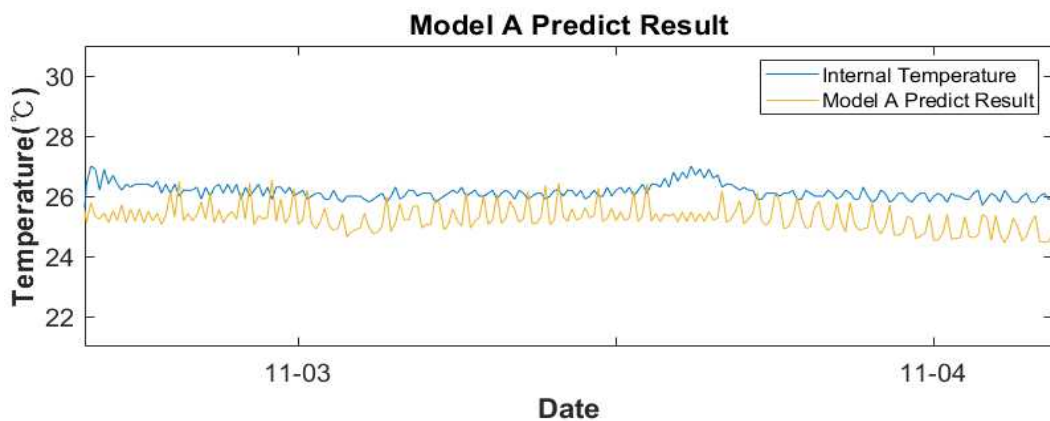


Fig. 20 Compare the details of Model A Predict data and Observation data

4. 예측 Model B 도출

예측 Model B를 도출 하는 과정에서는 먼저 재배합 재질의 영향으로 인해 예측되는 시스템 내부온도를 Model B의 변수로 활용하기 위해 이를 계산할 수 있는 식을 도출해내었다. 도출해낸 식을 활용하여 얻은 데이터와 Model B의 다른 변수 데이터셋을 가지고 앞서 도출해낸 예측 Model A과 동일한 방식인 머신러닝(Machine Learning) 라이브러리인 Tensorflow를 설치하여 사용하였다. 또한 모델을 효율적으로 구성하기 위해서 Tensorflow의 고수준의 API인 Keras 모듈을 활용하여 모델을 도출하였다.

4.1 예측 Model B 도출을 위한 변수요소 준비

4.1.1 재배합 재질 영향 식 도출을 위한 데이터 준비

작물재배시스템에 영향을 미치는 요소 중 하나로 선정된 재배합 재질이 시스템에 미치는 영향을 식으로 도출해내기 위해 시스템 제어가 없는 상태에서 측정된 데이터를 활용하였다. 이는 시스템 제어기온도가 없는 환경에서 측정된 내부 온도데이터, 외부 온도데이터 및 해당기간의 기상정보데이터를 의미한다.

08월 17일부터 08월 31일까지 시스템 제어기온도가 없는 환경에서 측정된 내부 온도데이터, 외부 온도데이터 및 해당기간 동안의 기상정보데이터를 수집하였다. 수집한 데이터를 10분주기로 데이터 과싱하고 결측 데이터를 처리하여 데이터를 전처리하여 이를 데이터셋으로 준비하였다.

Table 7 Dataset for the influence formula by material

Categorize	Dataset
Start date	08-17 00:10
End date	08-30 23:50
Input data	Weather information data, System external temperature
Output data	Weight and Bias used in the influence formula of material

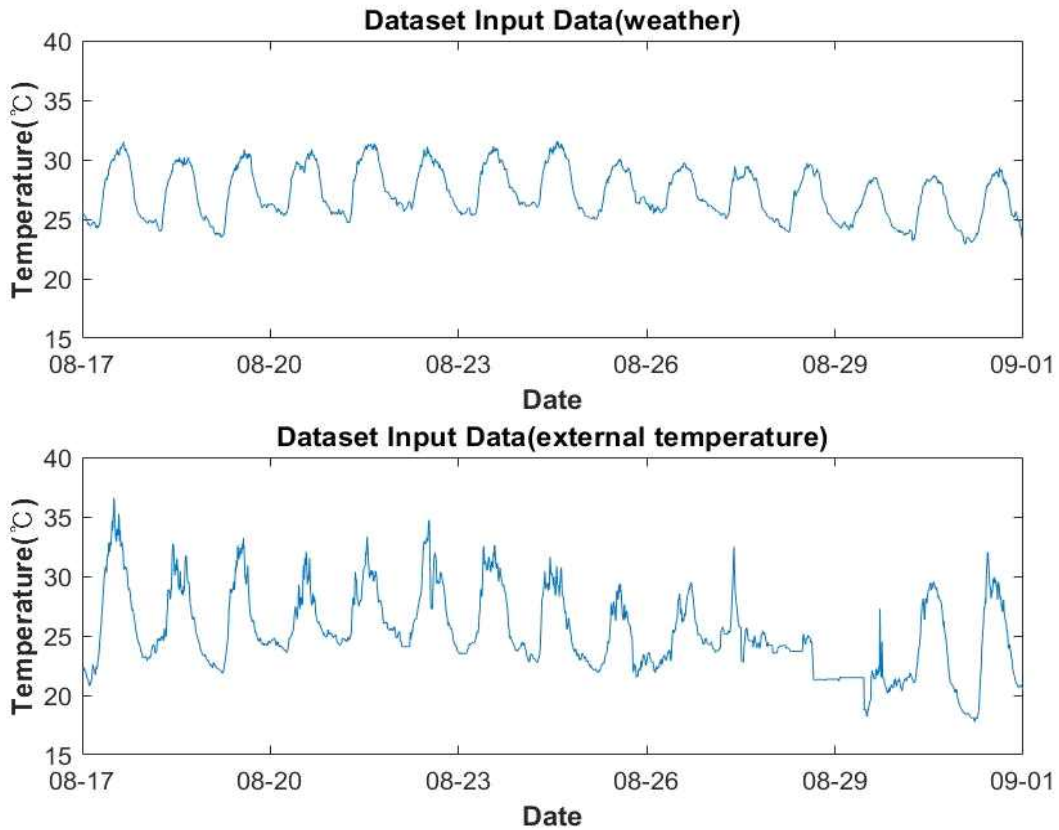


Fig. 21 The influence formula by material Dataset Input data

4.1.2 재배함 재질 영향 식 도출을 위한 모델 구성

앞서 예측모델을 구성하는데 사용한 방법과 같은 방법으로 Keras의 함수형 API를 사용하여 모델을 만들었다.

앞서 예측모델을 구성할 때와 달리 입력텐서(Input Tensor)로 시스템제어기 온도를 포함시키지 않은 기상청 평균기온, 시스템 외부온도 2개를 입력시킨 뒤 함수처럼 사용하기 위해 층 객체를 Dense를 통해 만들었다. 이후 텐서와 함께 층을 호출하면 출력텐서(Output Tensor)가 시스템 내부온도를 반환하도록 하는 구조로 구성하였다.

이후 Model 클래스의 객체를 통해 함수형 API로 구조화시킨 모델의 입력텐서(Input Tensor)에서 출력텐서(Output Tensor)를 도출하기까지 필요한 모든 층의 출력을 추출한 뒤 반환하였다.

4.1.3 재배합 재질 영향 식 도출을 위한 모델 훈련 및 식 도출

이후 구조화된 모델의 전체데이터를 반복하여 학습하는 횟수인 epochs를 100으로 지정하여 모델을 훈련하고 이후 모델이 반환하는 가중치 및 편향치를 읽어 들여 이를 식으로 도출해냈다.

$$\hat{y} = (0.3820 \times x_{[1]}) + (0.6588 \times x_{[2]}) + 1.6680 \quad (1)$$

$x_{[1]}$ 은 기상정보데이터 입력 값이며 $x_{[2]}$ 는 시스템외부온도 입력 값이고 0.3820은 기상정보데이터에 대한 가중치, 0.6588은 시스템외부온도에 대한 가중치, 1.6680은 모델이 반환시킨 편향치를 의미한다. 또한 제어기온도가 없는 환경에서 측정된 내부 온도데이터, 외부 온도데이터 및 해당기간 동안의 기상정보데이터만을 활용하여 도출해낸 본식의 \hat{y} 는 단순히 재배합 재질의 영향으로 인해 예측되는 시스템 내부온도데이터임을 알 수 있다. 이후 $x_{[1]}$ 에 10/25~11/02까지의 기상정보데이터, $x_{[2]}$ 에 10/25~11/02까지의 시스템외부온도를 입력시켜 해당기간에 예측되는 재배합 재질의 영향으로 예측되는 시스템내부온도를 도출해냈다.

4.2 예측 Model B 도출을 위한 데이터 준비

수집한 데이터를 학습데이터(train dataset) 및 테스트데이터(test dataset) 두 가지로 분할시켜 전처리시킨 데이터를 데이터셋으로 준비하였다.

Table 8 Dataset for Model B

Categorize	Train dataset	Test dataset
Start date	10-25 00:10	11-02 00:10
End date	11-01 23:50	11-09 23:50
Input data	Weather information data, System external temperature, Controller temperature, System internal temperature predicted by the influence of materials	
Target data	System internal temperature	-
Output data	-	Predicted system internal temperature

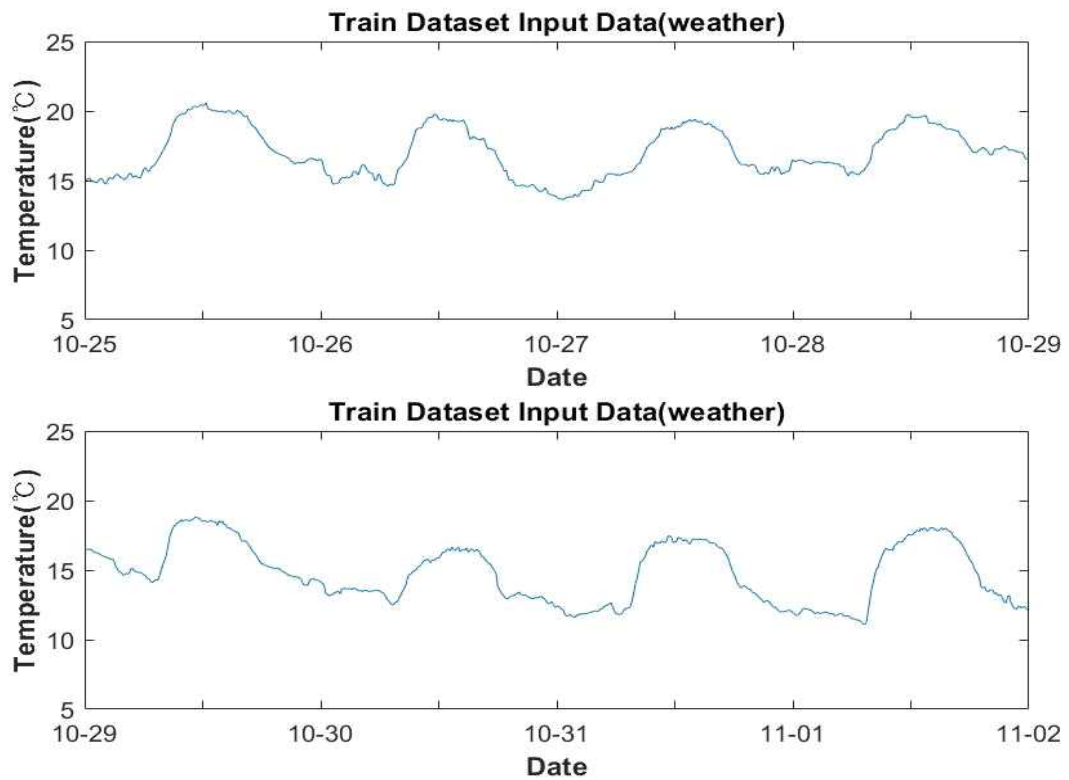


Fig. 22 Model B Train dataset Input data(weather information data)

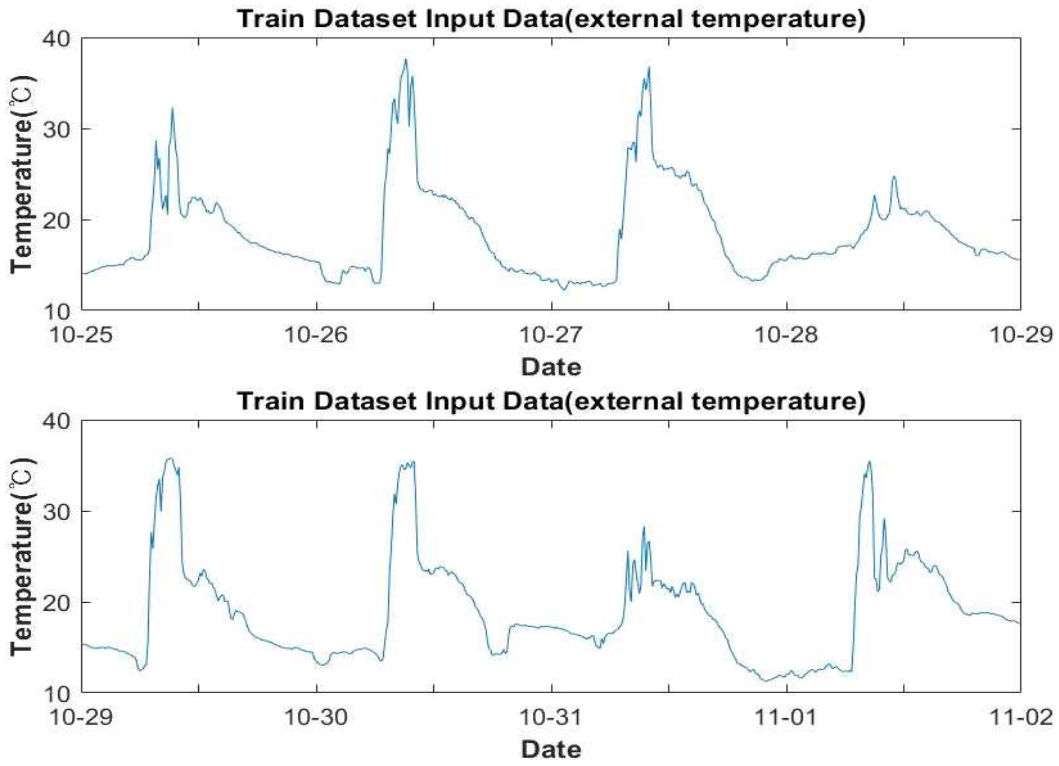


Fig. 23 Model B Train dataset Input data(system external temperature)

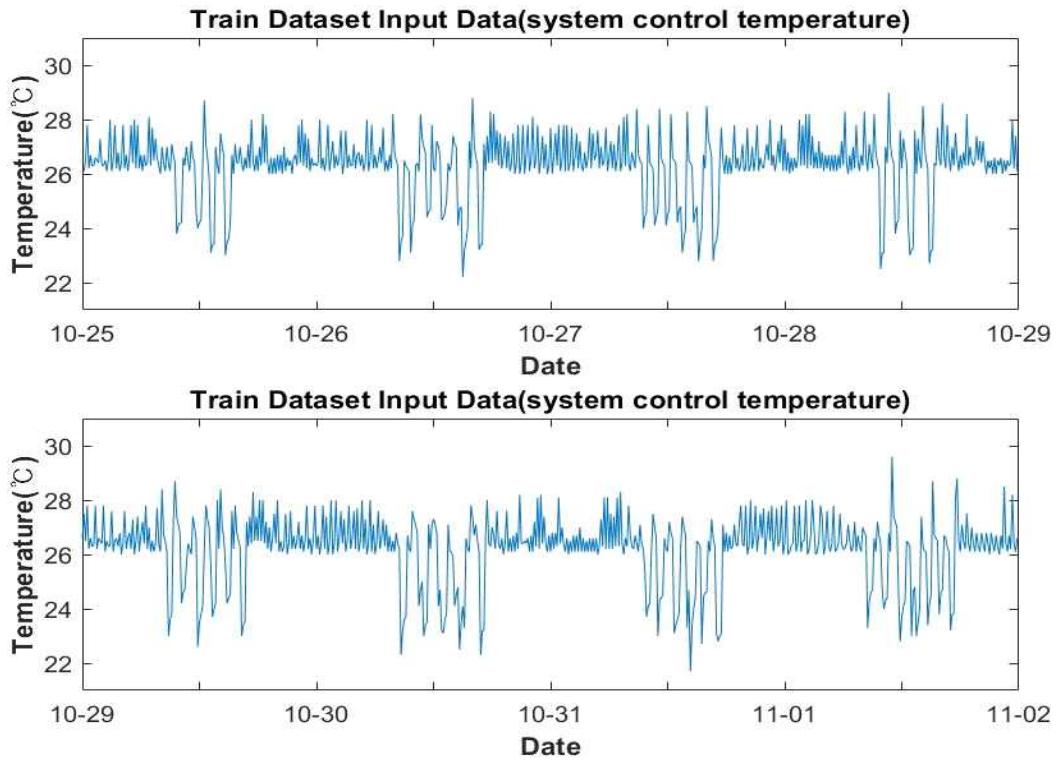


Fig. 24 Model B Train dataset Input data(controller temperature)

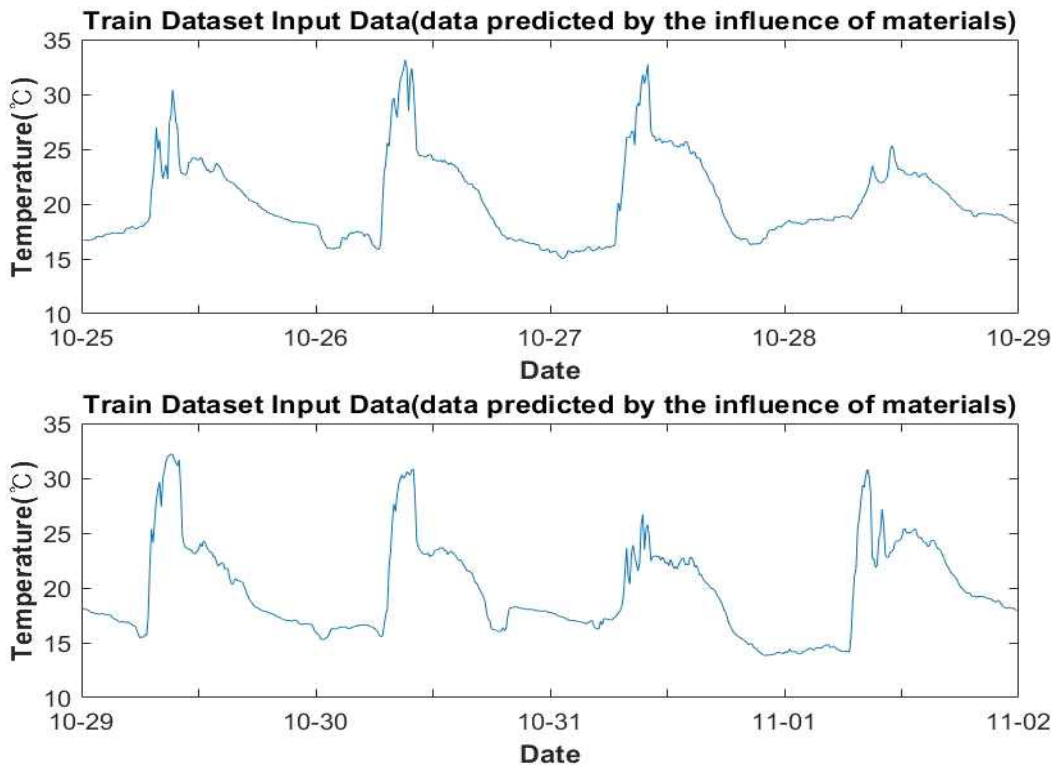


Fig. 25 Model B Train dataset Input data(data predicted by the influence of materials)

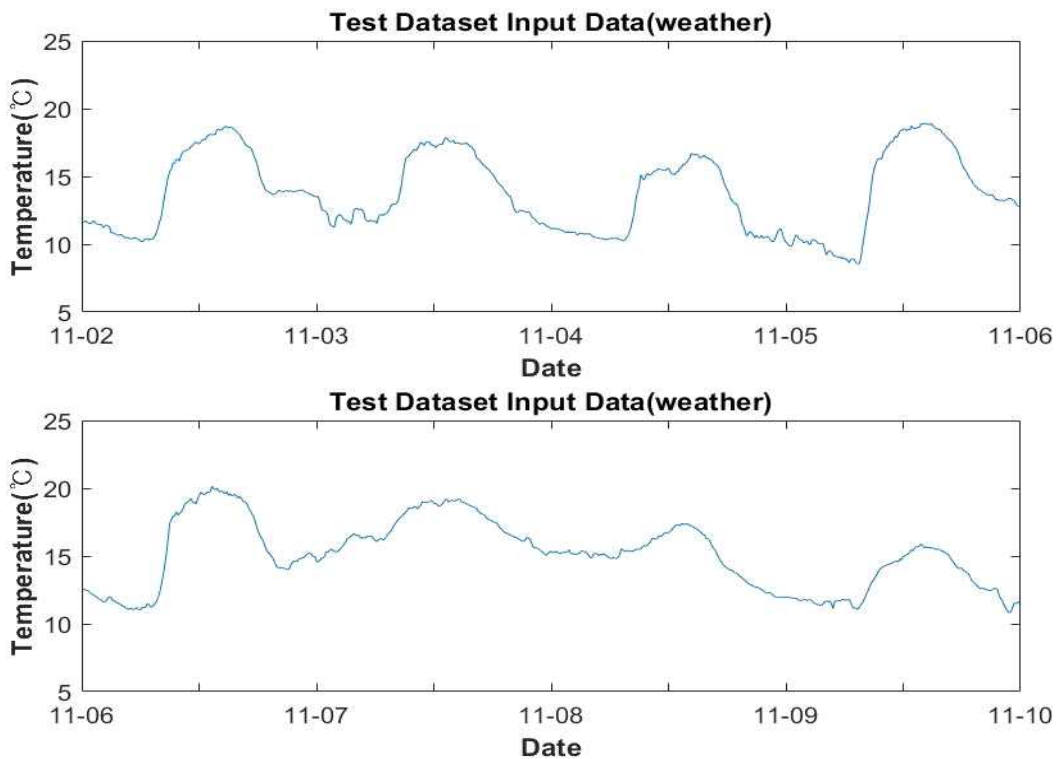


Fig. 26 Model B Test dataset Input data(weather information data)

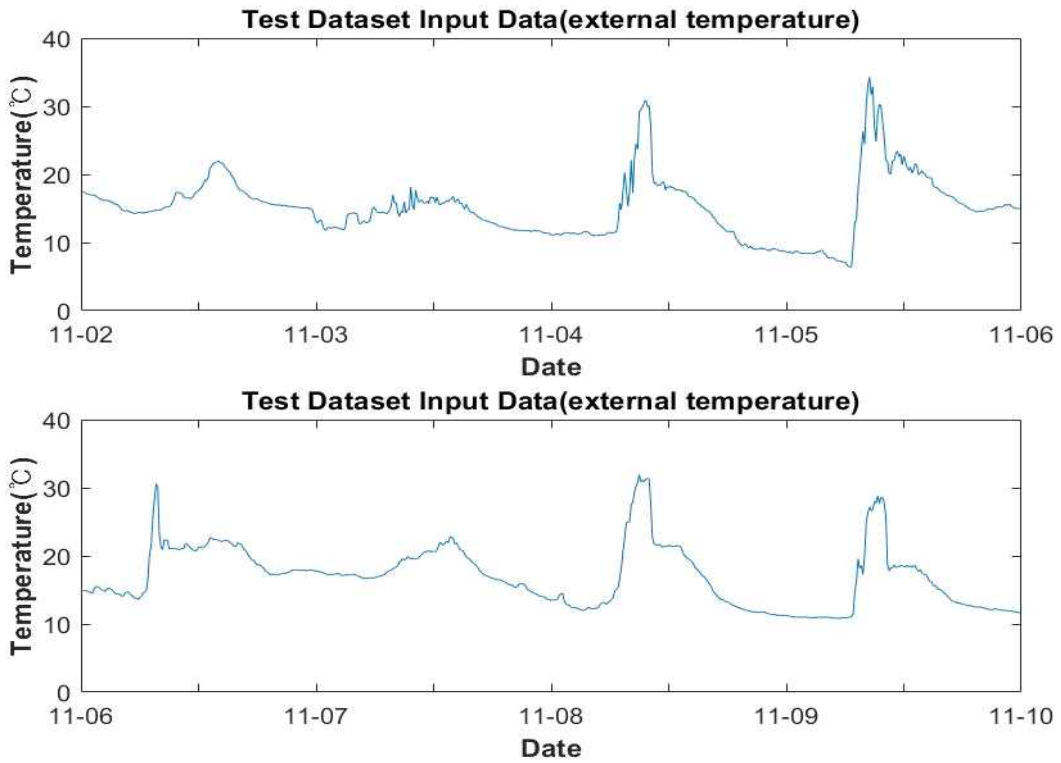


Fig. 27 Model B Test dataset Input data(system external temperature)

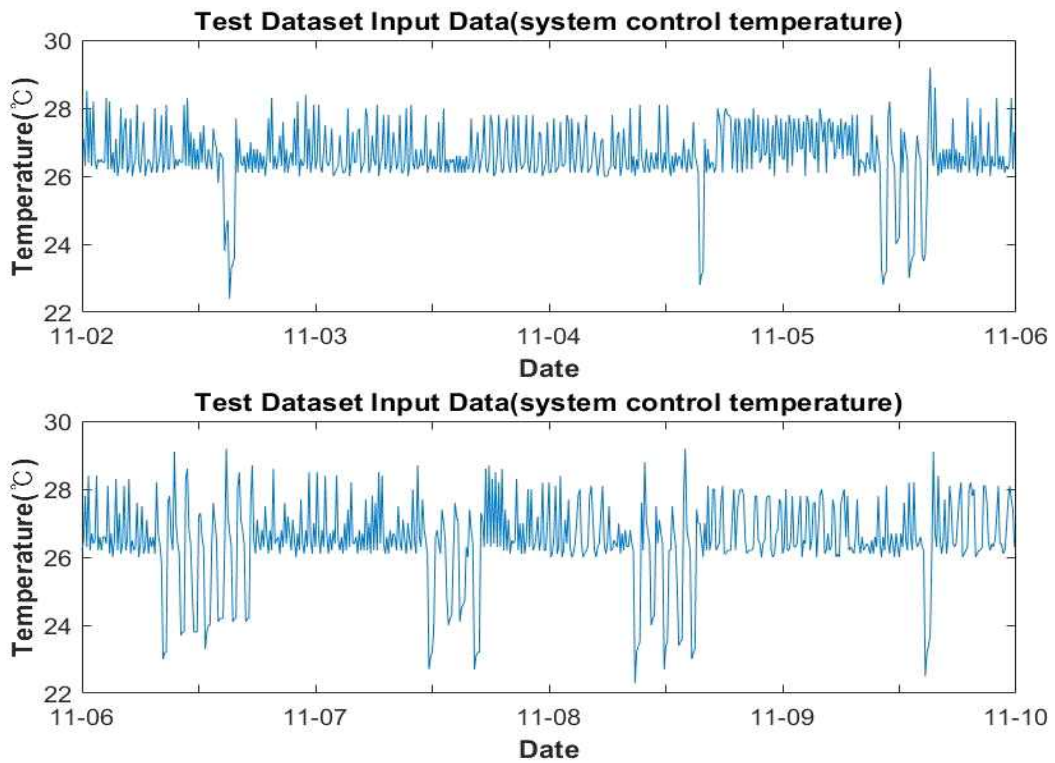


Fig. 28 Model B Test dataset Input data(controller temperature)

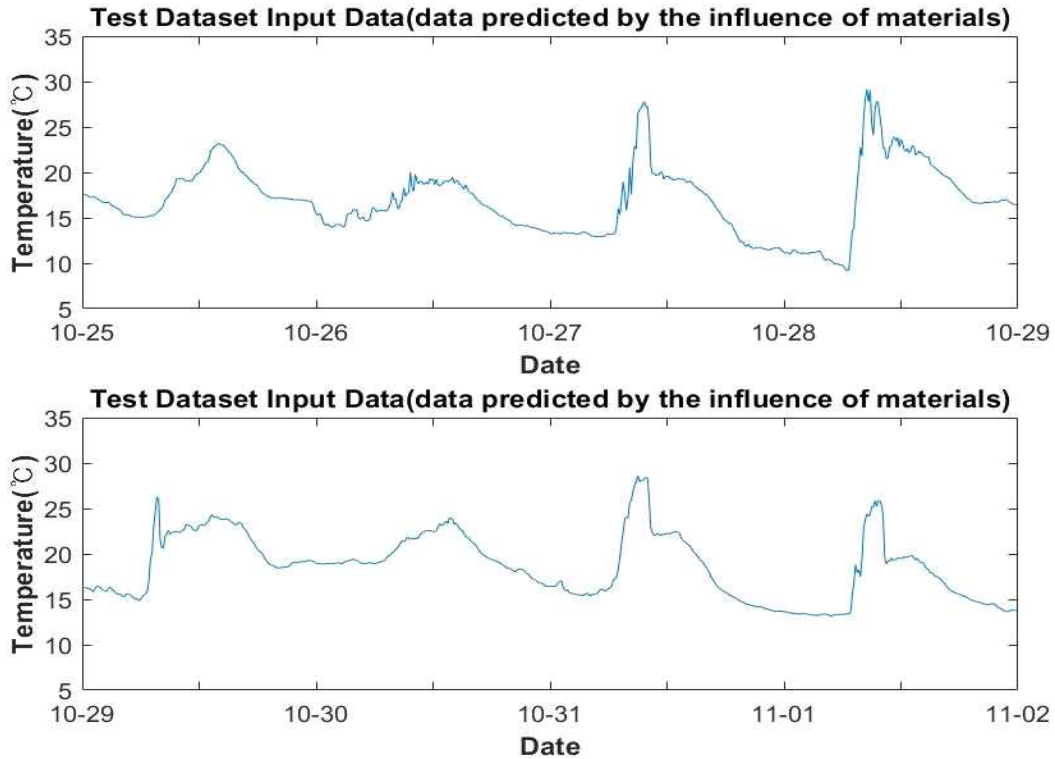


Fig. 29 Model B Test dataset Input data(data predicted by the influence of materials)

4.3 예측 Model B 구성

앞서 예측 Model A을 구성하는 방법과 동일한 방법으로 Keras의 함수형 API를 사용하여 모델을 만들었다.

입력텐서(Input Tensor)로 기상청 평균기온, 시스템 외부온도, 제어기온도, 재배함 재질의 영향으로 인해 예측되는 시스템 내부온도데이터 4개를 입력시킨 뒤 함수처럼 사용하기 위해 층 객체를 Dense를 통해 만들었다. 이후 텐서와 함께 층을 호출하면 은닉층 노드 100개로 구성된 은닉층(Hidden layer) 3개에서 이를 여러 단계에 거쳐 학습을 진행한 뒤 출력텐서(Output Tensor)가 시스템 내부 온도를 반환하도록 하는 구조로 구성하였다.

이후 Model 클래스의 객체를 통해 함수형 API로 구조화시킨 모델의 입력텐서(Input Tensor)에서 출력텐서(Output Tensor)를 도출하기까지 필요한 모든 층의 출력을 추출한 뒤 반환하였다.

4.4 예측 Model B 훈련 및 도출

이후 구조화된 모델의 전체데이터를 반복하여 학습하는 횟수인 epochs를 100으로 지정하고 다층퍼셉트론의 알고리즘을 통해 다음의 그림에 나오는 모델의 가중치 및 편향치를 계산해내는 훈련을 시킨 뒤 예측 데이터를 도출하였다.

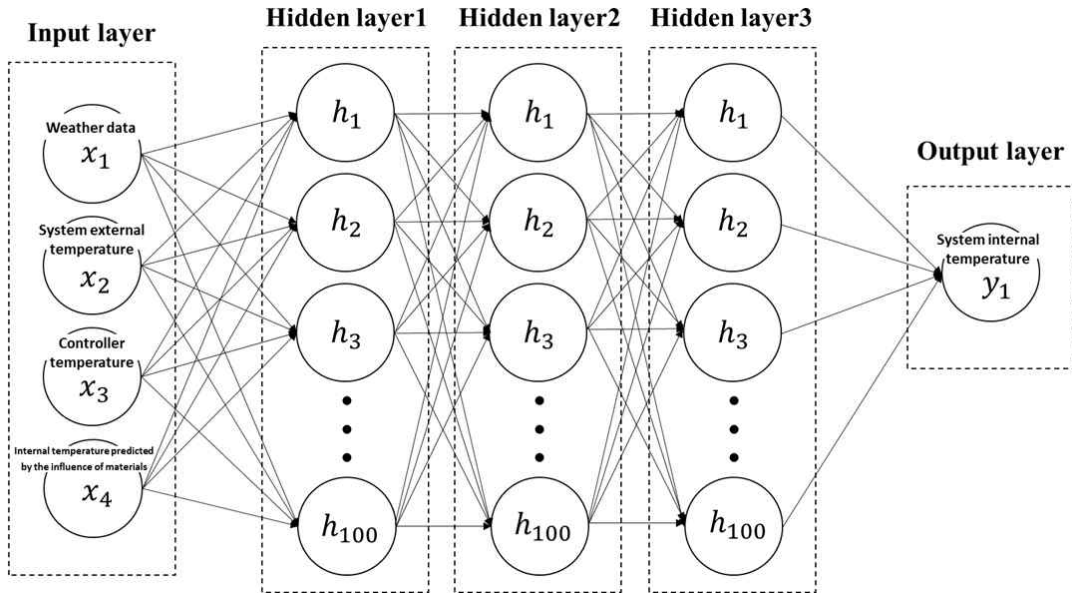


Fig. 30 Model B Neural Network architecture

```
[array([[ 1.62500501e-01,  1.22262605e-01,  6.45409971e-02,
          1.23013213e-01, -1.06179953e-01,  9.88079906e-02,
          -1.90108627e-01,  6.87537119e-02, -1.59312099e-01,
           8.92662331e-02, -2.75436968e-01,  1.00173317e-01,
          -2.68819975e-03,  8.43845084e-02,  1.08561158e-01,
           1.33511573e-01, -6.20288849e-02, -5.92179894e-02,
           1.39729023e-01, -2.43520841e-01, -1.12821676e-01,
          -2.11489290e-01, -2.42805347e-01,  2.94672102e-02,
          -2.38385841e-01, -2.79716879e-01,  1.21596955e-01,
          -1.61975935e-01, -2.11500973e-01, -2.29653865e-01,
          -1.02645151e-01, -1.60878804e-02, -9.79162306e-02,
          -1.31662533e-01, -4.46908996e-02,  7.22320527e-02,
           1.58941194e-01, -1.04721643e-01,  1.47297876e-02,
          -1.26135066e-01, -9.50489100e-03, -9.72964242e-02,
           4.10587005e-02, -1.48523539e-01,  1.30955681e-01,
           2.42917668e-02, -1.49523467e-01,  6.26029894e-02,
          -9.52965394e-02, -2.10776553e-01, -2.88844965e-02,
          -2.11489990e-01, -2.41884038e-01, -2.06636637e-01,
          -3.15177143e-01,  1.11149959e-01, -2.13474020e-01,
```

Fig. 31 Model B Weights and Bias values

훈련을 통해 Model B가 예측한 시스템 내부온도데이터와 실제 시스템 내부온도데이터를 시각화하여 비교하였다. Model B가 예측한 데이터는 평균 25.88°C로 시스템내부온도의 평균과 0.03°C의 편차를 나타냄을 알 수 있었다.

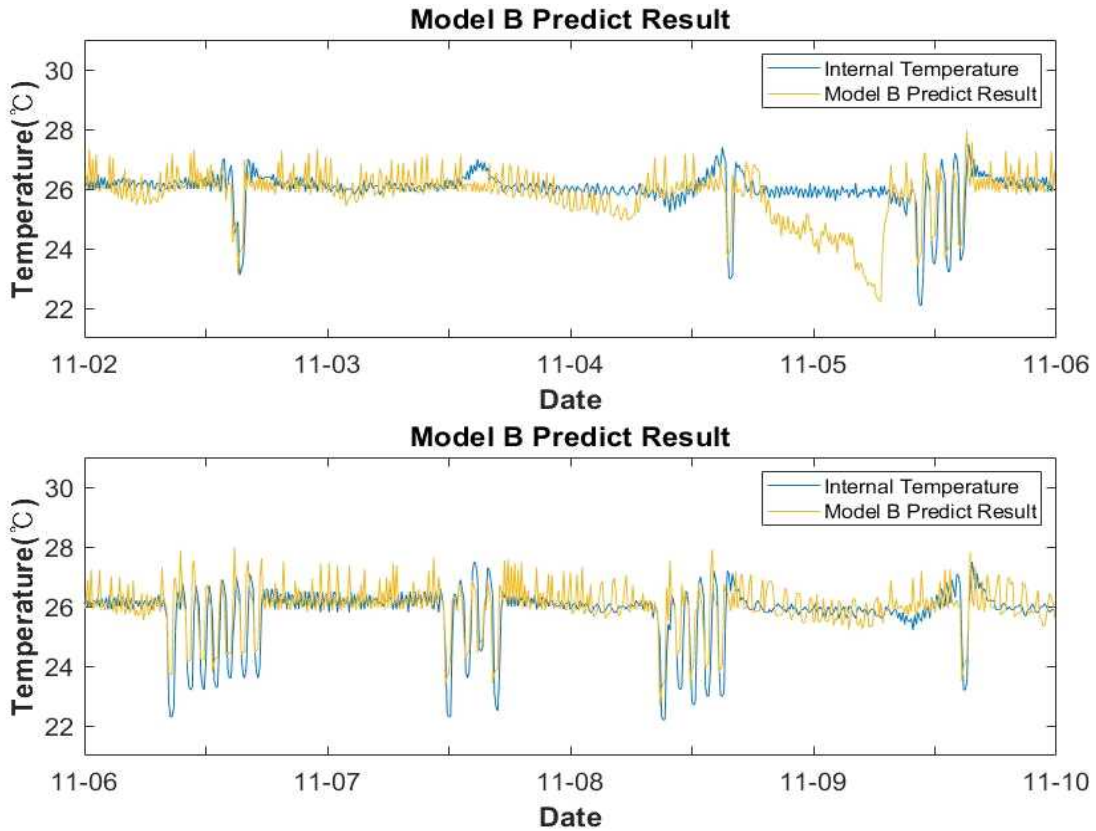


Fig. 32 Comparison of Model B Predict data and Observation data

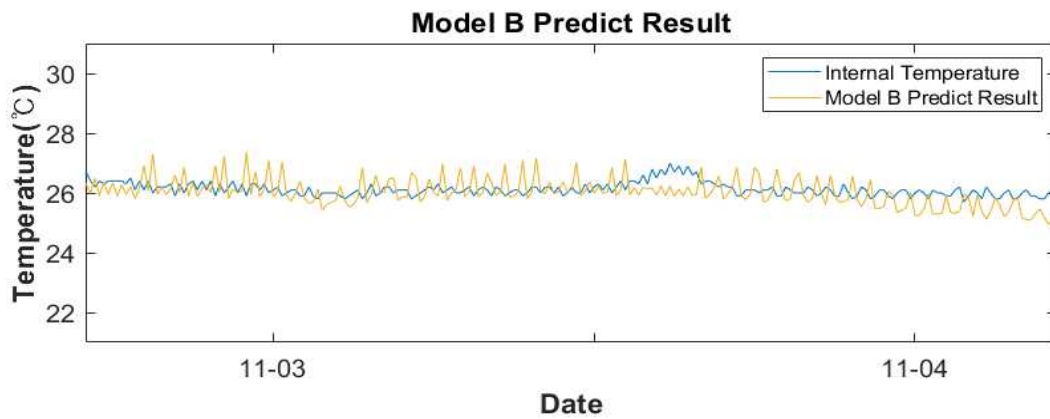


Fig. 33 Compare the details of Model B Predict data and Observation data

제 4 장 연구결과

1. 재배함 재질 영향 식 유무에 따른 모델 예측결과 비교

재배함 재질의 영향 식이 적용 되지 않은 모델인 Model A과 재배함 재질의 영향 식이 적용된 모델인 Model B가 예측한 시스템 온도데이터와 실제 측정된 시스템 온도데이터를 시각화하여 비교해보았다.

Table 9 Data set for Model A & Model B

Categorize	Model A	Model B
Train dataset Start date	10-25 00:10	
Train dataset End date	11-01 23:50	
Test dataset Start date	11-02 00:10	
Test dataset End date	11-09 23:50	
Input data	Weather information data, System external temperature, Controller temperature	
	-	System internal temperature predicted by the influence of materials
Output data	Predicted system internal temperature	

재배함 재질의 영향 식이 적용 되지 않은 모델인 Model A은 Input data로 기상정보데이터, 시스템외부온도, 제어기온도 3가지 변수를 입력시켜 학습한 모델이 Output data로 시스템내부온도를 예측하였고, 재배함 재질의 영향 식을 적용 시킨 모델인 Model B는 Input data로 기상정보데이터, 시스템외부온도, 제어기온도, 재배함 재질의 영향으로 예측되는 시스템 내부온도데이터 4가지 변수를 입력시켜 학습한 모델이 Output data로 시스템내부온도를 예측하였다. 두 모델이 예측한 결과는 다음과 같다.

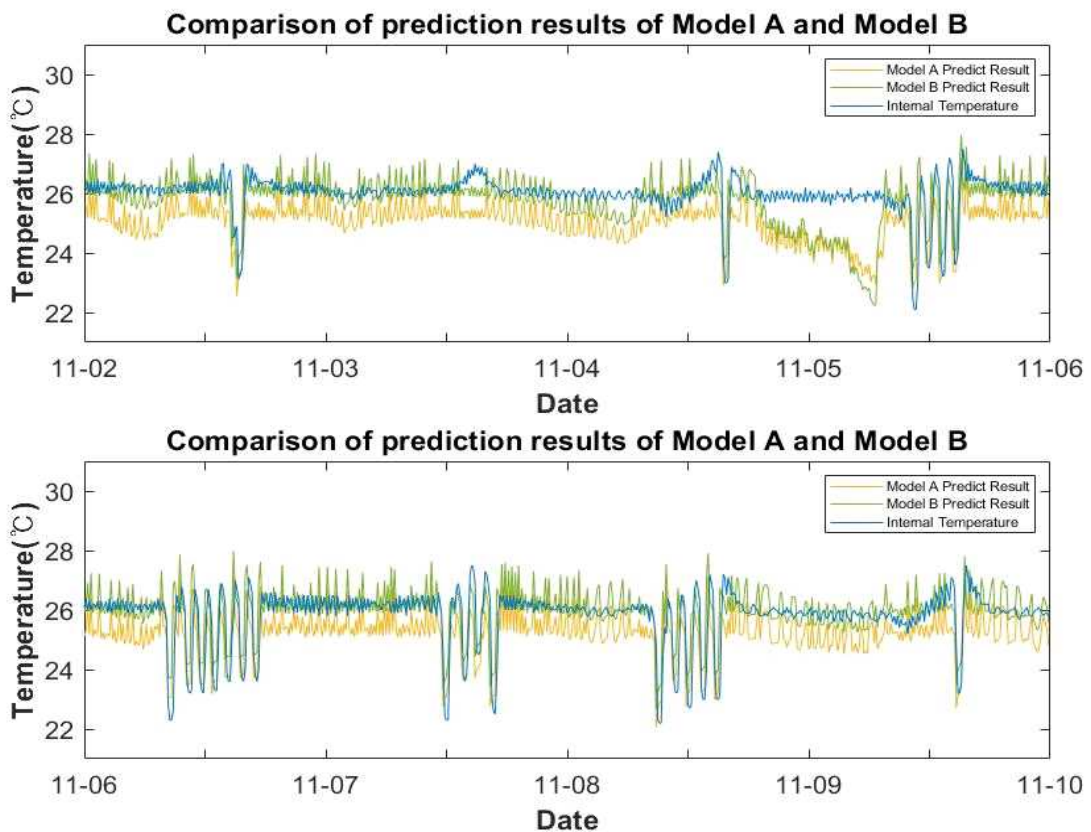


Fig. 34 Comparison of Model A & Model B Predict data and Observation data

시스템내부온도가 평균 25.91°C로 약 26°C의 온도를 나타내는 동안 Model A은 평균 25.14°C로 시스템내부온도의 평균과 0.77°C의 편차를 보였고, Model B는 평균 25.88°C로 시스템내부온도의 평균과 0.03°C의 편차를 나타냈다. 이를 통해 재배함 재질 영향 식이 적용된 모델인 Model B가 실측데이터와 유사한 패턴의 데이터를 예측함을 알 수 있다.

2. 제배함 제질 영향 식 유무에 따른 모델 성능평가 결과 비교

두 가지 타입으로 도출해낸 모델의 성능을 비교하여 최적의 내부온도 예측모델을 도출해내기 위해 모델 성능 평가인 RMSE 기법을 활용하였다.

2.1 상관분석

RMSE 기법을 활용하기에 앞서 상관분석을 통해 상관계수가 높은 변수에는 어떠한 변수가 있고 상관계수가 낮은 변수에는 어떠한 변수가 있는지 확인해보는 과정을 진행하였다.

본 과정에서는 피어슨 상관계수(Pearson correlation coefficient)를 사용하여 분석을 진행하였다. 피어슨 상관계수(Pearson correlation coefficient)란 두 변수가 각각 간격척도 혹은 비율척도로 측정된 경우의 상관관계의 크기를 나타내는 값으로서 변수들이 정규분포를 따른다는 가정 하에 적용하는 상관계수이다.

피어슨 상관계수(Pearson correlation coefficient)값의 범위는 -1부터 +1까지이며 계수의 부호는 관계의 방향을 나타내고 절댓값은 관계의 강도를 나타낸다. 즉, +1또는 -1에 가까울수록 상관관계가 있다고 말할 수 있으며 상관계수의 절대값이 일반적으로 0.4이상인 경우 상관관계가 있다고 해석한다.[12]

상관분석을 통해 전체적으로 도출해낸 예측모델식과 제어기온도와 시스템 내부온도가 상관계수가 높은 변수로 나타났다. Model A에서는 예측모델식과 제어기온도와의 상관계수가 0.74, 시스템 내부온도와의 상관계수가 0.68로 나타났으며 Model B에서는 예측모델식과 제어기온도와의 상관계수가 0.61, 시스템 내부온도와의 상관계수는 0.72으로 나타났다.

Model A은 예측모델식과 제어기온도와의 상관계수가 0.74로 높은 상관계수를 보이며 이는 제어기의 영향으로 인해 예측 온도데이터의 편차가 보다 높게 나타났음을 알 수 있었다.

Model B는 예측모델식과 시스템 내부온도와의 상관계수는 0.72로 높은 상관계수를 보이며 실제 시스템 내부온도와 유사한 데이터를 보다 많이 예측해냄을 알 수 있다.

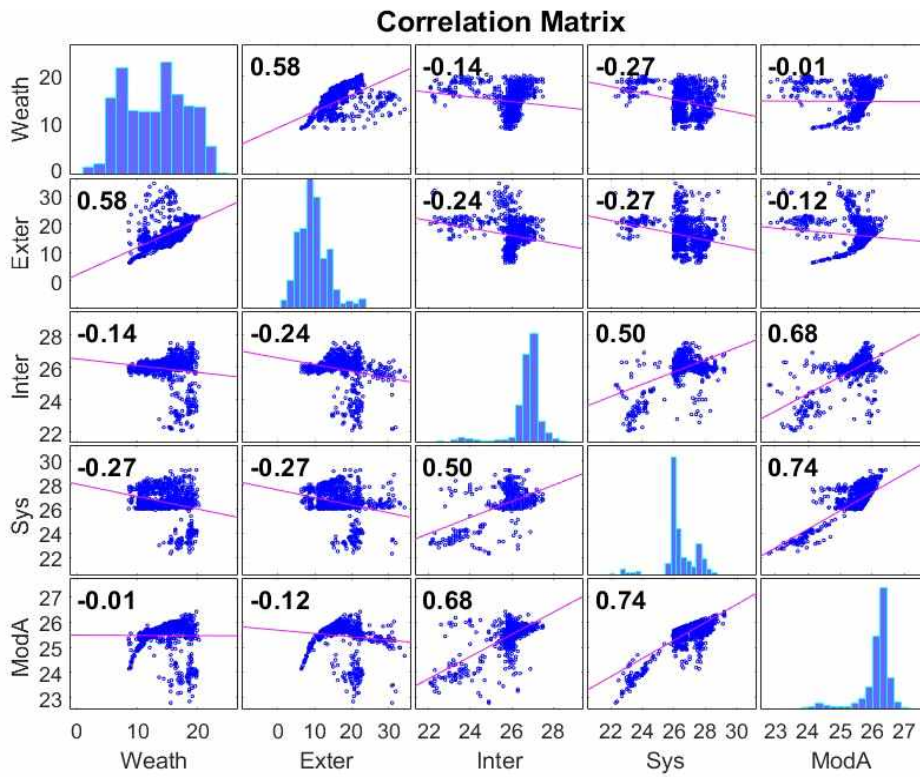


Fig. 35 Model A Correlation analysis results

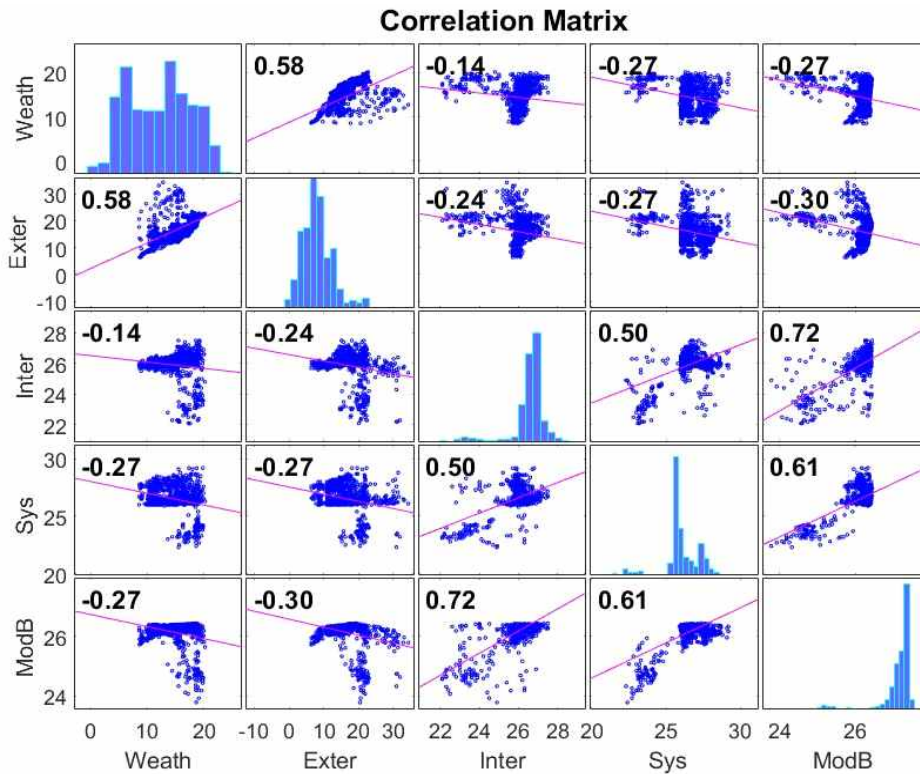


Fig. 36 Model B Correlation analysis results

2.2 RMSE 평가

예측모델평가는 모델이 실제 값과 예측 값의 차이를 표현하는 손실 함수로 평가를 진행한다. 본 과정에서는 모델의 평가를 위해 손실함수로 사용되는 평균 제곱근 오차(Root Mean Square Error, RMSE)를 이용하여 예측 성능을 평가하였다. RMSE는 모델이 예측한 값과 실제 값의 차이를 이용하여 정밀도를 표현하는 척도이며 아래 식과 같이 계산 할 수 있다.[13] 실제 측정된 시스템 온도데이터와 각 모델의 예측데이터를 가지고 평가를 진행하였다.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y - y_{pred})^2} \quad (2)$$

2.2.1 재배합 재질 영향 식 미적용 모델 RMSE 평가

RMSE 기법을 활용하여 재배합 재질 영향 식을 적용시키지 않은 모델인 Model A 평가결과이며 평균치는 0.94로 나타났다.

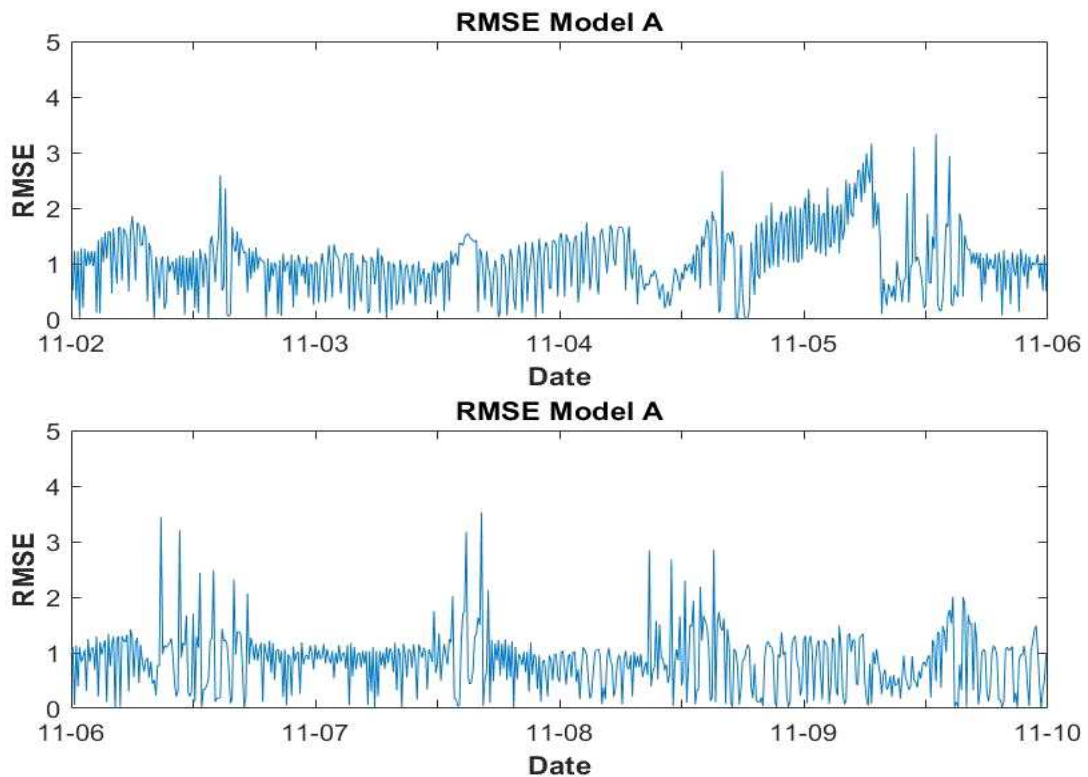


Fig. 37 Model A RMSE evaluation result

2.2.2 재배합 재질 영향 식 적용시킨 모델 RMSE 평가

RMSE 기법을 활용하여 재배합 재질 영향 식 적용시킨 모델인 Model B 평가결과이며 평균치는 0.61로 나타났다.

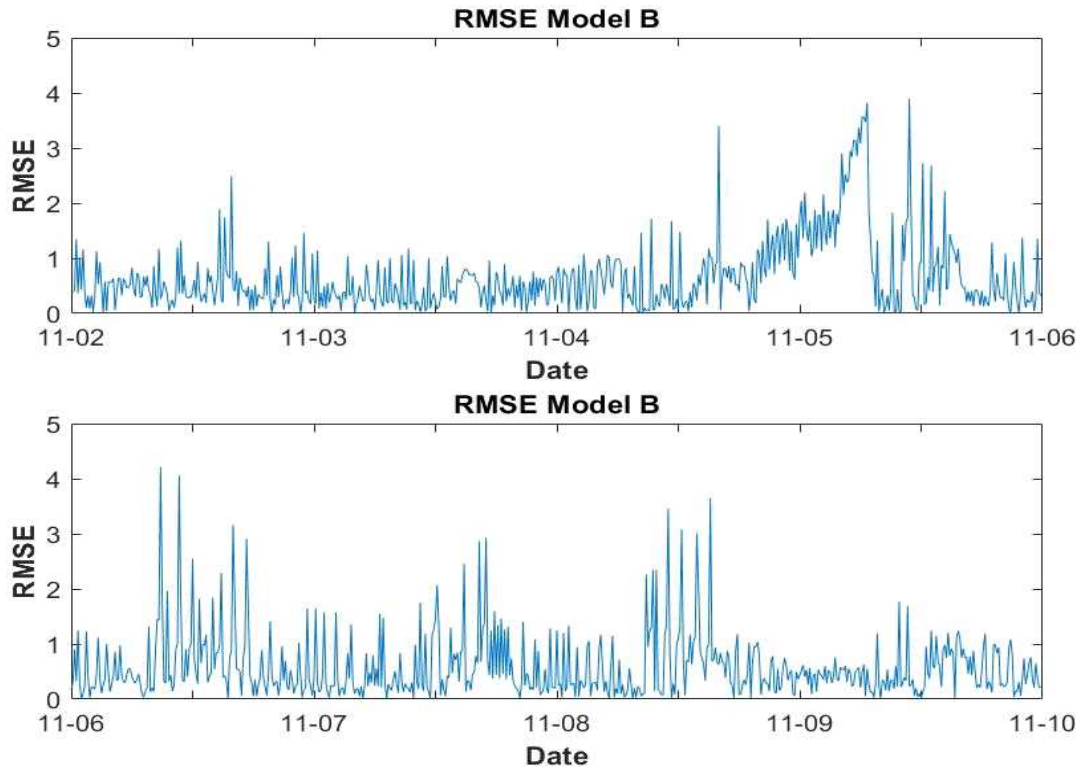


Fig. 38 Model B RMSE evaluation result

2.2.3 재배합 재질 영향 식 유무에 따른 모델 RMSE 평가 결과 비교

재배합 재질 영향 식이 적용된 모델인 Model B의 RMSE 평가 결과를 노란색 실선 그래프, 식을 적용하지 않은 모델인 Model A의 RMSE 평가 결과를 파란색 실선 그래프로 표현하였다.

재배합 재질 영향 식이 적용된 모델인 Model B가 예측한 시스템 온도데이터가 식을 적용하지 않은 모델인 Model A의 예측 데이터보다 평균 0.33의 낮은 오차를 나타내고 있음을 알 수 있다.

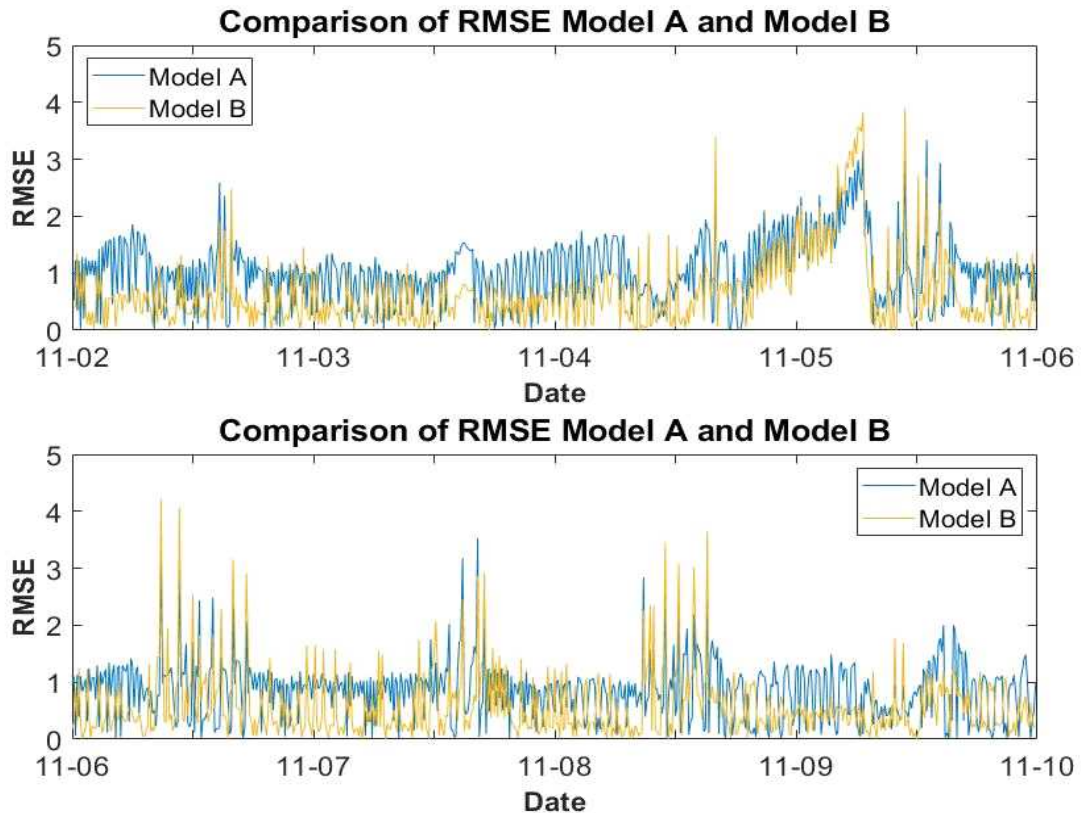


Fig. 39 Comparison of Model A & Model B RMSE evaluation result

Model A과 Model B의 RMSE평가 결과를 Fig.40에 보다 자세하게 나타내었다. Fig.40의 첫 번째 그래프에서 나타나듯 평균 0~1사이의 RMSE값을 보이다 두 번째 그래프와 같이 3~4사이의 값으로 이상치를 나타내는 부분이 있었다. 이러한 이상치를 나타내는 원인을 도출해내기 위해 Model A과 Model B가 예측한 실내온도데이터 및 실측데이터를 나타낸 그래프와 비교해보았다. RMSE값이 3~4사이의 이상치 데이터를 보이는 부분은 Fig.41의 두 번째 그래프에서 확인할 수 있듯이 평균 26도를 나타내는 첫 번째 그래프와 달리, 실제 실내온도가 급격히 변하여 이상치를 보였을 경우 이 흐름에 맞춰 Model A과 Model B의 예측데이터도 이상치를 보이며 이에 의해 RMSE값이 평균보다 높게 나오는 것을 확인할 수 있었다.

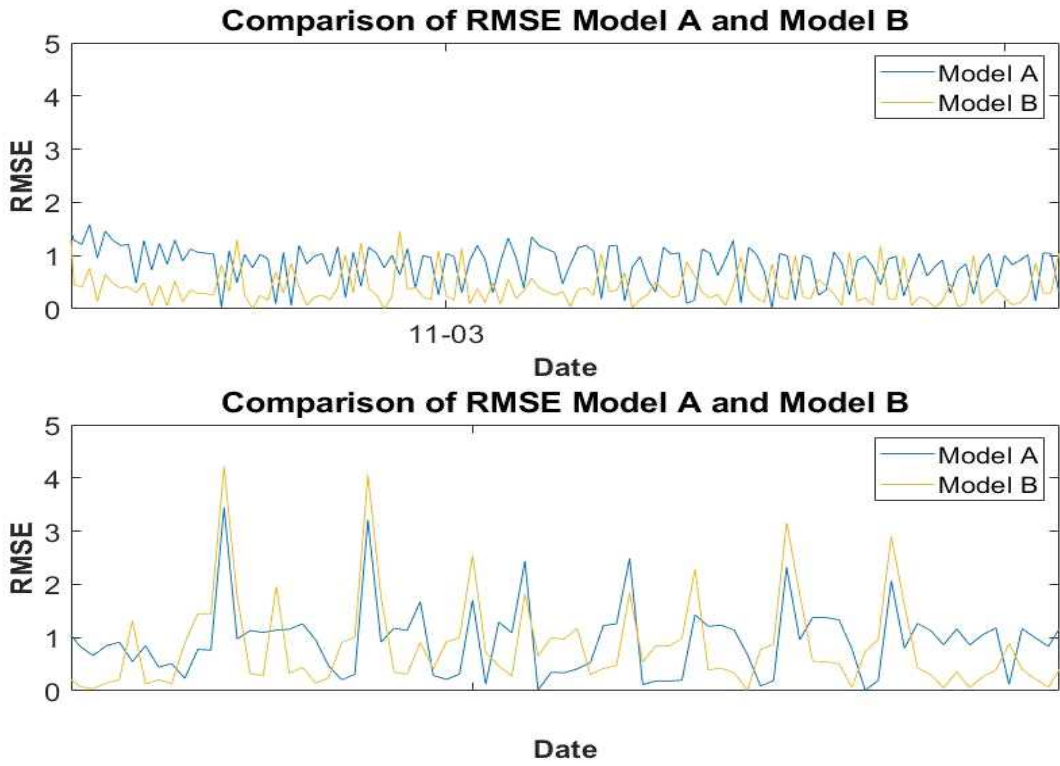


Fig. 40 Compare the details of Model A & Model B RMSE evaluation result

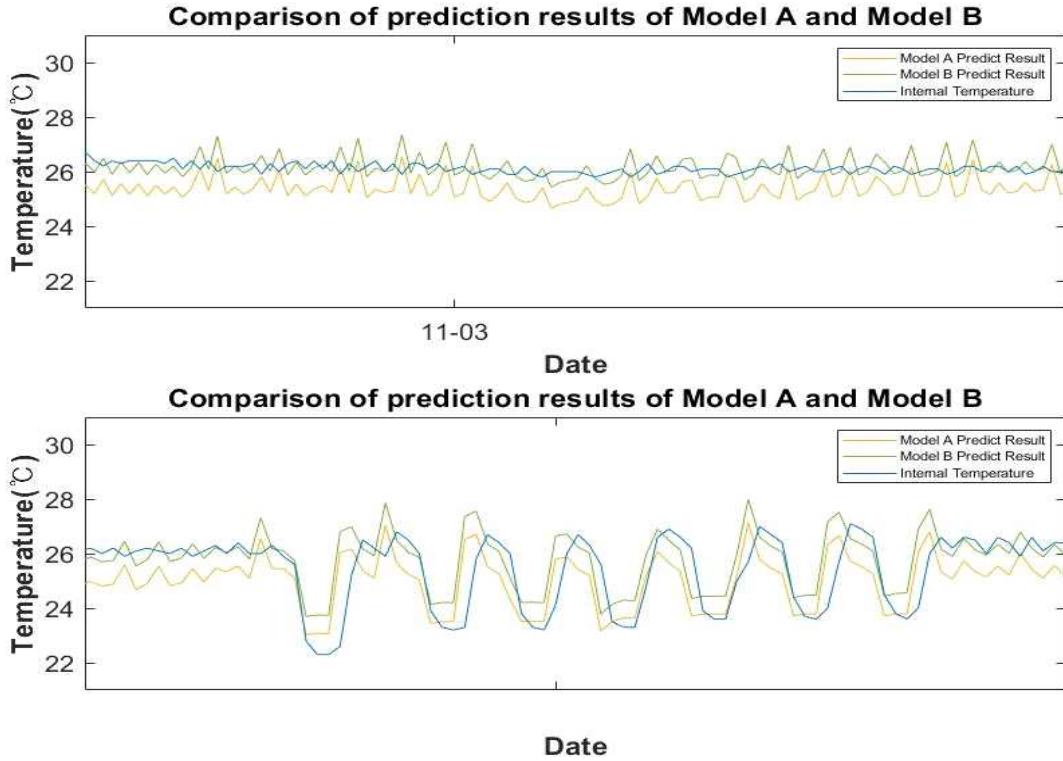


Fig. 41 Observation data Outlier

제 5 장 결론

본 연구는 기존 재배시설 자동화 및 편리성 증대에 집중했던 연구에서 벗어나 효율성 있는 생산성에 대한 방법을 찾기 위해 머신러닝(Machine Learning) 기법을 활용하여 최적의 환경조건에서 작물의 생육이 이루어지도록 하는 IoT 작물재배시스템 내부온도 예측모델을 도출해내는 것을 연구목표로 시작하였다.

본 연구에서는 최적의 예측모델을 도출해내기 위해 작물재배시스템에서 측정된 외부온도데이터, 내부온도데이터, 제어기온도데이터와 농업기상정보서비스 API를 통해 기상데이터를 수집하였다. 이를 머신러닝(Machine Learning)의 학습 방법 중 하나인 지도학습(Supervised Learning)의 회귀(Regression)방식을 사용하였고 효율적으로 예측모델을 도출하기위해 신경회로망모델로 다층퍼셉트론(Multilayer Perceptrons, MLP) 알고리즘 구조를 활용하여 도출하였다.

작물재배시스템에 영향을 미치는 요소를 고려하고 입력변수에 따라 모델을 두 가지로 분류하여 선정하였다. 입력변수로 기상정보, 시스템외부온도, 제어기온도 외 재배함 재질을 입력변수로 포함하지 않은 예측모델인 Model A, 재배함 재질을 입력변수로 포함한 예측모델인 Model B 두 가지 타입으로 예측모델을 나눠서 모델을 도출하고, 예측 데이터를 시각화 및 RMSE 평가 기법을 통해 성능평가를 진행하였다. RMSE 평가 결과 재배함 재질을 입력변수로 포함하지 않은 예측모델인 Model A은 평균 0.94, 재배함 재질을 입력변수로 포함한 예측모델인 Model B은 평균 0.61로 나타났다.

이는 재배함 재질 영향 식이 적용된 모델인 Model B가 예측한 시스템 온도 데이터가 식을 적용하지 않은 모델인 Model A의 예측 데이터보다 평균 0.33의 낮은 오차율을 나타내고 있음을 알 수 있으며 이를 통해 재배함 재질 영향 식이 적용된 모델인 Model B가 최적의 내부온도 조건에서 작물의 생육이 이루어지도록 하는 예측모델임으로 분석되었다.

갑작스럽게 변하는 내부온도 데이터에 영향을 받아 이상치를 보이는 예측모델의 한계를 극복하기위해 Anomaly Detection(이상탐지) 알고리즘을 활용하고,

보다 긴 기간 동안 데이터를 측정하여 다양한 계절과 날씨 등의 환경조건에서 측정된 데이터를 사용하여 모델을 학습시키며 예측 능력을 보완하고 향상시킬 필요성이 있다고 판단되었다.

향후 본 연구에서 도출한 예측모델을 활용하여 재배함에 폴리우레탄 (PU/PIR)판넬 이외에도 스티로폼(EPS)판넬 및 그라스울(G/W)판넬 등 다양한 재질을 적용하여 재질에 따른 모델식을 도출해내고, 판넬의 두께 및 재배함의 사이즈 등을 변수로 추가하여 다양한 두께와 사이즈에 따른 온도변화 실험을 통해 재질에 따른 작물재배시스템의 재배환경 기본값을 도출해낼 계획이다.

재질에 따른 작물재배시스템의 재배환경 기본값과 해당 지역의 기상정보, 외부기온값을 변수로 입력하면 해당 설치 장소에서의 작물재배시스템의 재배환경 기본값을 도출해낼 수 있을 것으로 예상된다. 이를 통해 다양한 지역과 특이한 환경을 갖춘 상황이라도 도출해낸 재배환경 기본값을 활용하면 각종 외부요인에 간접적으로 받으면서 작물이 성장하기 최적의 환경조건에서 생육이 이루어질 수 있을 것이라 기대할 수 있다.

참고 문헌

- [1] 김연중, “4차 산업혁명 대응 스마트팜 기술 및 정책동향“ 융합연구리뷰, 융합 연구정책센터, 2018. 04, p.4
- [2] 한국과학기술기획평가원, 과학기술예측조사 2012~2035 미래사회 전망과 과학 기술 예측. 2012. 02, p.16-17
- [3] 김연중, “4차 산업혁명 대응 스마트팜 기술 및 정책동향“ 융합연구리뷰, 융합 연구정책센터, 2018. 04, p.20-22
- [4] 오정원, “빅데이터 응용 머신러닝 기반 스마트팜 시스템의 설계 및 구현”, 박사학위논문, 대구가톨릭대학교 대학원 컴퓨터정보통신공학과, 2018. 08
- [5] 서신림, “버섯 재배를 위한 IoT기반 스마트 제어 운영 시스템”, 박사학위논문, 원광대학교 일반대학원 정보관리학과, 2015. 10
- [6] 김연중, 박지연, 박영구, “4차산업혁명에 대응한 스마트농업 발전방안”, 2017. 10, p.1-3
- [7] 프랑소와 솔레, “케라스 창시자에게 배우는 딥러닝”, 2019. 03,길벗, p.142-144
- [8] 배성완, “머신 러닝을 이용한 주택 가격 예측력 비교”, 박사학위논문, 단국대학교 대학원 도시계획 및 부동산학과 부동산학 전공, 2019. 06
- [9] 안드레아스 필러 , 세라 가이드, “파이썬 라이브러리를 활용한 머신러닝”, 2019. 03, 한빛미디어
- [10] 세바스찬 라시카, 바히드 미자리리, “머신러닝 교과서 with 파이썬, 사이킷 런, 텐서플로”,2019. 07,길벗, p.460
- [11] 프랑소와 솔레, “케라스 창시자에게 배우는 딥러닝”,2019. 03,길벗, p.315-316
- [12] 송은정, “도시부 교차로 간격과 서비스 수준의 상관분석 연구”, 석사학위논문, 인천대학교 대학원 건설환경공학전공, 2013. 08
- [13] 조경우, 정용진, 강철규, 오창현, “미세먼지 예측을 위한 기계 학습 알고리즘의 적합성 평가”, 한국정보통신학회논문지 Vol. 23, 2019. 01, p.20-26