

기계학습 알고리즘을 이용한 스마트 온실 내부온도 예측 모델 개발 및 검증

오광철¹ · 김석준² · 박선용² · 이충건¹ · 조라훈² · 전영광² · 김대현^{3*}

¹강원대학교 농업생명과학연구원 연구원, ²강원대학교 스마트농업융합학과 대학원생, ³강원대학교 바이오시스템기계공학 교수

Development and Verification of Smart Greenhouse Internal Temperature Prediction Model Using Machine Learning Algorithm

Kwang Cheol Oh¹, Seok Jun Kim², Sun Yong Park², Chung Geon Lee¹, La Hoon Cho²,
Young Kwang Jeon², and Dae Hyun Kim^{3*}

¹Researcher, Agriculture and Life Science Research Institute, Kangwon National University, Chuncheon 24341, Korea

²Graduate Student, Department of Interdisciplinary Program in Smart Agriculture, Kangwon National University, Chuncheon 24341, Korea

³Professor, Department of Biosystems Engineering, Kangwon National University, Chuncheon 24341, Korea

Abstract. This study developed simulation model for predicting the greenhouse interior environment using artificial intelligence machine learning techniques. Various methods have been studied to predict the internal environment of the greenhouse system. But the traditional simulation analysis method has a problem of low precision due to extraneous variables. In order to solve this problem, we developed a model for predicting the temperature inside the greenhouse using machine learning. Machine learning models are developed through data collection, characteristic analysis, and learning, and the accuracy of the model varies greatly depending on parameters and learning methods. Therefore, an optimal model derivation method according to data characteristics is required. As a result of the model development, the model accuracy increased as the parameters of the hidden unit increased. Optimal model was derived from the GRU algorithm and hidden unit 6 ($r^2 = 0.9848$ and $RMSE = 0.5857^\circ\text{C}$). Through this study, it was confirmed that it is possible to develop a predictive model for the temperature inside the greenhouse using data outside the greenhouse. In addition, it was confirmed that application and comparative analysis were necessary for various greenhouse data. It is necessary that research for development environmental control system by improving the developed model to the forecasting stage.

Additional key words : artificial intelligence, machine learning, simulation model, smart greenhouse, time series data

서 론

우리나라의 작물 재배 농가는 2000년부터 5년 단위로 약 8%가 지속해서 감소되어 왔다(Fig. 1). 이 중 대부분을 차지하고 있는 노지 작물(논벼, 콩, 감자, 고구마, 고추, 배추 등) 재배 농가는 2015년 94만 9천 가구에서 2020년 87만 7천 가구로 7만 2천 가구(7.6%)가 감소되었지만, 시설 작물(고추, 수박, 딸기, 토마토, 포도 등) 재배 농가는 2015년 12만 5천 가구에서 2020년 12만 6천 가구로 약 1천 가구(1%)가 증가하였다. 이러한 시설 작물 재배 농가의 증가로 인하여 2015년에서 2020

년 전체 작물 재배 농가 수는 증감률은 약 -5%로 약화되었다.

시설농업은 통제된 시설 안에서 빛, 온도, 습도 등의 재배 환경을 인위적으로 조성하여 연중 내내 농산물을 생산하는 농업으로 자동화를 통하여 보다 효율적으로 농산물을 생산할 수 있으며 이에 대한 연구는 지속적으로 증대되고 있다(2020 농림어업총조사, 통계청). 또한 국내 농가의 고령 인구 비중이 2015년 38.3%에서 2020년 42.3%로 늘어나 노동집약적 작업에서 자동화 통한 농작물 생산방안이 요구되고 있다.

이러한 시설농업의 자동화를 위하여 다양한 연구가 수행되어왔다. 실시간 생육환경제어를 위한 농가 온실 내부 자동 온도 조절 환경 시스템을 구축하여 농산물 생산효율을 향상시켰으며(Ahn 등, 2016), 온실 환경의 변화에 따라 즉시 제어가 가능한 웹 기반 시스템을 구축하여 온실 환경에 대한 실시간 확

*Corresponding author: daekim@kangwon.ac.kr

Received March 21, 2022; Revised June 10, 2022;

Accepted June 20, 2022

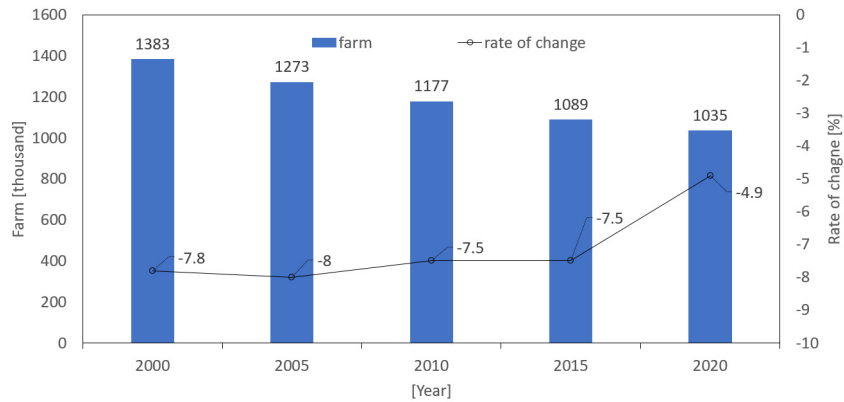


Fig. 1. Change in the number of farms by year (2020 Agriculture, Forestry and Fisheries Census, Statistics Korea).

인 및 데이터수집 시스템을 구현하였다(Choi 등, 2012). 또한 이러한 농업 자동화 연구를 효과적으로 활용하기 위해서는 온실 내부 환경변화에 대한 정확하고 정밀한 예측기법이 요구된다. 따라서 기존의 전산유체역학(computational fluid dynamics, CFD) 시뮬레이션 기법을 활용한 내부 환경 예측 연구들이 진행되어 왔다(Cho 등, 2015; Fatnassi 등, 2015; Hong와 Lee, 2014; Kim, 2001; Lee와 Short 1999; Lee 등 2006a, 2006b; Tadj 등, 2010; Yu 등, 2014).

하지만 온실 시스템은 가외 변인(extraneous variable)인 태양광, 외기 온-습도를 포함하고 있어 기존의 시뮬레이션 분석의 한계점이 존재한다. 최근 이러한 문제를 극복하기 위하여 빅데이터를 기반으로 한 인공지능(artificial intelligence, AI) 기술이 활용하고자 한다. 인공지능은 4차 산업혁명의 핵심 기술로 음성인식, 이미지 분류, 회귀 분석, 데이터 시각화, 추천 시스템, 강화학습 등 다양한 분야에서 활용되고 있으며 (Hwang 등, 2021; Park 등, 2018; Na 등, 2017), 이중 인공지능 전망을 통한 시계열 자료 처리 및 예측 관련 연구 또한 활발히 이루어지고 있다(Kwon 등, 2021). 특히 최근의 인공지능 기술 발달로 인하여 비선형이며 다차원의 복잡한 시스템에 대한 실시간 분석이 가능해졌으며, 산업 분야에서 스마트공장의 공정 최적화를 위하여 활용되는 등 다양한 연구가 선행되었다 (Oh 등, 2020). 이러한 빅데이터를 활용한 인공지능 기법을 적용하기 위해서는 대상 시스템의 환경 데이터가 필수적으로 요구된다.

국내 시설농업 농가중 자동화 시설이 설치된 농가는 유형별로 비닐하우스 56.6 %, 유리온실 24.8 %로 가장 많이 차지 (2020 농림어업총조사, 통계청)하고 있으며 여기서 얻어지는 다양한 데이터를 활용한 기계학습모델 적용이 유리하다. 이러한 특성을 바탕으로 인공지능 기술을 활용한 온실 내부 환경 예측을 위한 다양한 연구가 수행되었는데 다원적 회기 인

공신경망 모델을 개발하여 온실 내부온도 및 습도 변화 예측에 활용될 수 있음을 검증하였으며(Song 등, 2019), LSTM (long short term memory) 알고리즘과 데이터 특성 분석을 통하여 온실 내부 CO₂ 농도 분포 예측이 이루어졌다(Lee 등, 2020). 또한 ANN(artificial neural network), RNN(recurrent neural network), MRM(multi regression model) 알고리즘을 통하여 개발된 기계학습모델 예측결과를 활용하여 온실 환경을 효과적으로 제어함으로써 에너지를 효율적으로 관리할 수 있음을 검증하였다(Kim과 Jung, 2017; Choi 등, 2019; Kim 등, 2001). 이처럼 다양한 방법으로 인공지능 모델을 활용한 연구가 수행되고 있지만 모델 구성에 필수적인 매개변수 및 데이터 적용 방법에 대한 명확한 방법이 제시되어 있지 않다.

최종적으로 본 연구에서는 기계학습모델 개발을 위한 데이터 수집 및 분석, 모델 구성을 위한 파라미터 선정과 사례연구(case study)을 통하여 최적 모델 개발 및 평가를 수행하였다. 초매개변수에 변화에 따른 인공지능 기계학습모델의 특성 분석이 진행되었으며, 비선형이며 다차원의 복잡한 시설농업 시스템 비닐하우스의 내부온도 예측을 위한 모델 개발 연구가 이루어졌다.

재료 및 방법

1. 온실 시스템 구성

본 연구는 강원대학교 설치된 유리온실(300m², 강원대학교 춘천시 효자동 192-1)의 환경제어 시스템에 대하여 분석이 이루어졌다. 대상 유리온실은 자연환기 시스템이며 내부에서 발생하는 추가적인 가외변인을 최대한 제거하기 위하여 아무런 식물을 배치하지 않은 조건에서 실험이 수행되었다. 또한 예측하고자 하는 내부온도와 가장 유의미한 관계가 있는 외부 기상데이터(외부온도, 외부습도, 태양광, 풍향, 풍속, 감우데

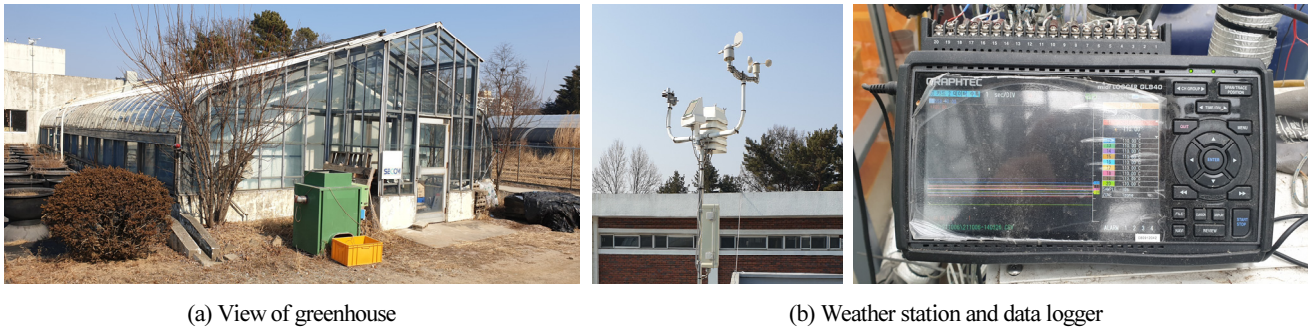


Fig. 2. Configuration of experimental equipment.

이터)가 기상관측장치(JNGW100, Jinong, Anyang-si, Korea)를 통하여 수집되었으며 온실 내부온도는 지상 1m 위치에서 열전대(Thermocouple, GTPK-02-17, GILTRON, Seoul, Korea)를 통하여 수집되었다. 최종적으로 데이터로거(GL840, GRAPHTEC Co., Kanagawa, Japan)를 통하여 분당 1개의 데이터가 수록되었다(Fig. 2).

2. 모델 개발

2.1 데이터 전처리

기계학습모델 개발에 첫 단계는 대규모의 데이터베이스 안에서 일정한 규칙을 찾아내어 분석하는 데이터 마이닝(data mining)으로 시작된다. 각기 다른 단위로 수집된 데이터는 정규화를 통하여 기계학습에 적용 가능한 데이터로 전환한다. 또한, 측정 오차 및 신뢰도를 확보하기 위하여 데이터 전처리가 필요하며, 대상 시스템의 정상범위 밖으로 측정된 값은 필요에 따라 제거한다(Charu, 2018). 또한 기계학습모델의 성능은 데이터의 양과 질에 의존적이다. 따라서 모델 적용을 위한 데이터는 정확하며, 적절한 개수로 구성되어 있어야 한다. 이러한 데이터 변환 과정을 차원축소(dimensionality reduction)라고 하며 기계학습의 가장 핵심적인 전처리 과정 중 하나이다. 차원축소는 단순히 데이터의 압축이나 잡음(noise)을 제거하는 것이 아니라 수집된 데이터를 가장 잘 나타낼 수 있는 잠재공간(latent space)을 도출하는 것(Gabriel 등, 1971; Huh, 2017)으로 결과예측에 영향이 미미한 데이터는 전체 특징 집합에서 제거할 수 있으며, 변수 삭제를 통하여 데이터의 차원을 축소시킬 수 있다. 데이터 전처리는 기존 입력값을 기반으로 새로운 데이터를 생성하기 때문에 모델 학습 전에 수행되어야 하며, 입력 데이터 간의 절댓값 차이가 크면 모델 학습에 많은 영향을 미칠 수 있으므로 각 데이터는 전처리 후 분석이 이루어진다. 본 연구에서는 특징 정규화(feature normalization)의 방법을 통하여 데이터 전처리가 수행되었다(Charu, 2018) (식 1).

$$\frac{x_{ij} - \min_j}{\max_j - \min_j} \rightarrow x'_{ij} \quad (1)$$

x_{ij} : i번째 데이터의 j번째 차원의 값
 \min_j : j번째 특징의 최솟값
 \max_j : j번째 특징의 최댓값
 x'_{ij} : 정규화된 x값

2.2 학습모델 선정

전처리를 통하여 정규화된 데이터를 활용하기 위하여 대상 시스템의 특성에 알맞은 모델선정이 필요하다. 시간의 흐름에 따라 실시간으로 주변 환경이 달라지고 이에 따라 데이터 특성이 변화되기 때문에 이전 데이터와 인공신경망을 활용한 학습을 통하여 모델이 개발되며 온실 내부온도 예측이 이루어진다. 따라서 대상 온실 시스템의 시계열 데이터는 특정 수준의 정상성(stationarity)을 가진다고 가정하여 분석이 이루어졌으며, 지도학습(supervised learning)의 회귀 분석(regression analysis)방법을 통하여 연구가 수행되었다(Arigo 등, 2018; Hope 등, 2017; Kim와 Jung, 2017; Muller와 Guido, 2017). 이러한 지도 학습을 통한 기계학습모델의 개발은 레이블된 데이터(labeled data)를 이용하여 모델에 입력값에 대한 출력값을 함께 학습하는 과정이 반드시 요구된다.

이러한 회귀분석의 가장 기본적인 분석방법인 RNN(recurrent neural network) 알고리즘은 데이터의 시퀀스가 길어지면 베니싱 문제(vanishing gradient problem)가 발생하여 모델 정확도가 감소할 수 있다. 이를 해결하기 위하여 LSTM(long short term memory)과 GRU(gated recurrent units) 알고리즘은 제안되었으며 이를 활용한 모델 개발이 이루어졌다. LSTM 알고리즘(Fig. 3)은 hidden layer와 input gate, output gate, forget gate의 세 가지 게이트로 구성된 메모리 블록과 추가적인 cell을 생성하여 베니싱 문제를 개선하였다(Arigo 등, 2018; Chang 등, 2019; François, 2017). 또한 GRU 알고리즘

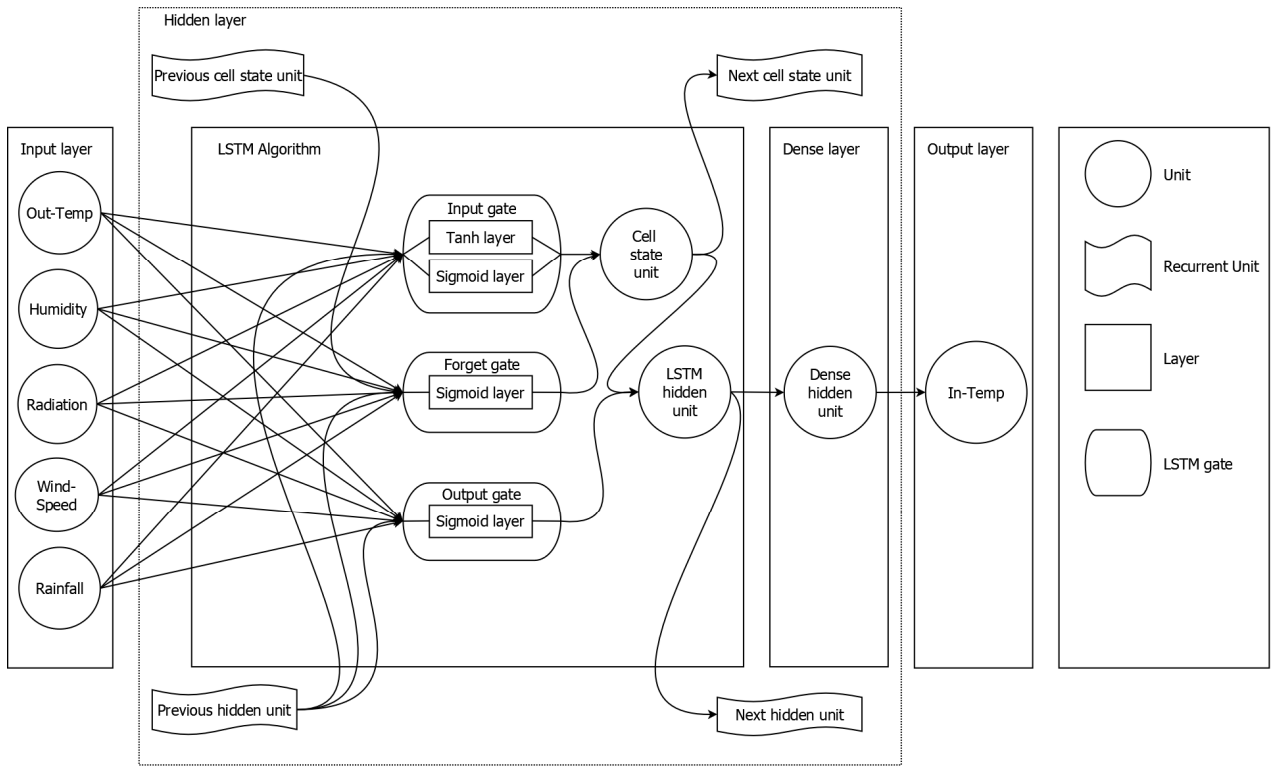


Fig. 3. Algorithm of long short-term memory (LSTM).

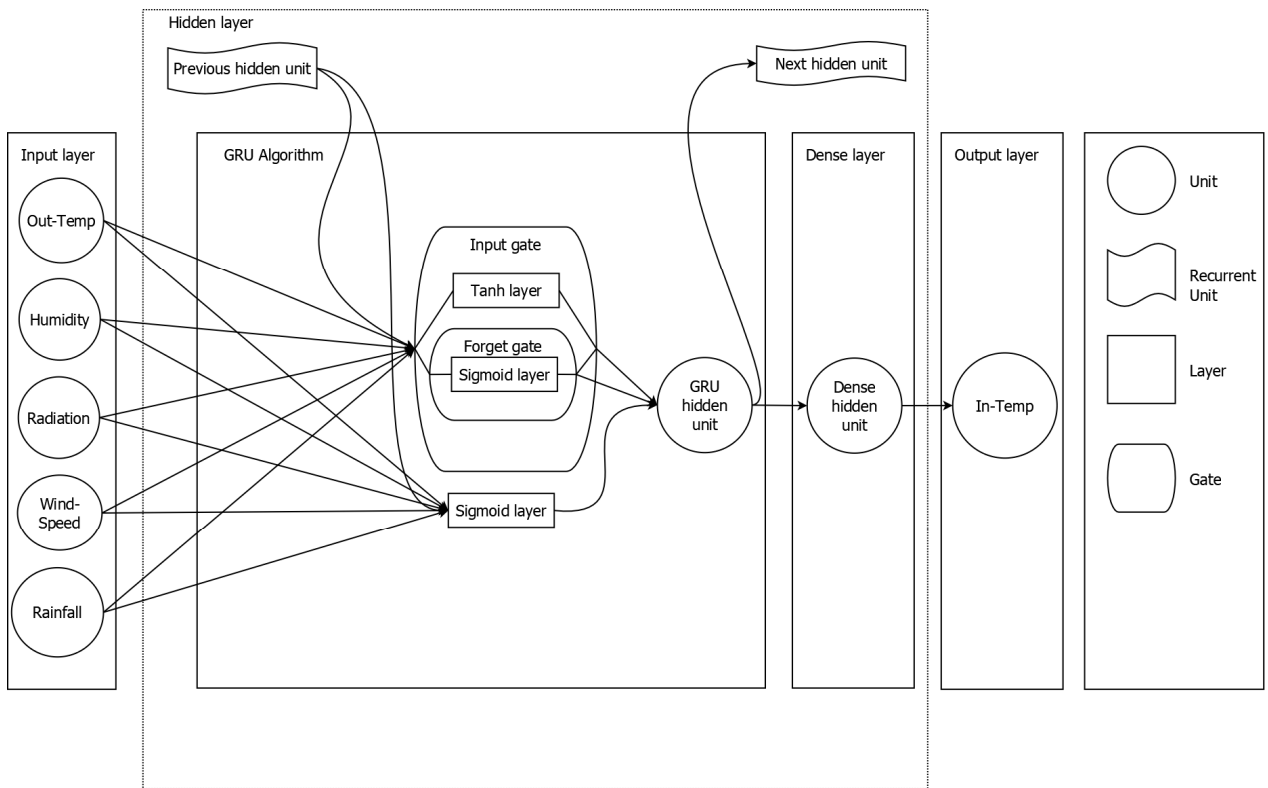


Fig. 4. Algorithm of gated recurrent units (GRU).

(Fig. 4)은 LSTM 알고리즘에서 게이트를 1개 축소한 reset gate, update gate로 이루어져 있으며 cell의 제거를 통하여 계산 부하를 감소시킨 알고리즘이다(Cho 등, 2014).

3. 기계학습모델 개발방법

기계학습모델은 알고리즘의 확장, 반복학습횟수 및 숨은층(hidden layer)의 증가를 통하여 정밀도 및 정확도 상승시켜 모델 최적화를 달성할 수 있다. 숨은층은 기계학습모델에서 입력레이어와 출력레이어 사이에 위치하는 모든 레이어를 뜻하며, 내부의 숨겨진 층을 1-2로 학습하는 방법의 경우 얇은 학습(shallow learning)이라 부른다. 숨은층의 개수를 3개 이상으로 두껍게 쌓아 분석하는 기법이 딥러닝(deep learning)이며 연속된 층으로 구성하여 점진적으로 가치 있는 값을 학습하는 데 강점이 있다(Charu, 2018). 모델을 개발하는데, 얼마만큼의 층(layer)을 활용하여 학습이 이루어졌는지에 따라 모델의 깊이가 결정되며, 이러한 층들은 매개변수를 통하여 연결되어 있다.

3.1 매개변수

기계학습의 매개변수는 가중치와 같이 모델이 학습되면서 업데이트되는 매개변수(parameter)와 학습 속도나 정칙화에 관여하며 기계학습모델 학습을 위하여 설정 가능한 초매개변수(hyper-parameter)로 구분된다. 학습이 진행됨에 따라 매개변수는 제어가 불가능하지만 초매개변수는 입력값에 따른 다양한 결과가 도출된다. 초매개변수 설정에는 다양한 방법이 있다. 가장 기본적으로 grid search가 있으며 다른 대표적인 방법으로 manual search, random search, bayesian search가 있다. Manual Search는 연구자 임의대로 초매개변수를 찾는 방법이며, grid search는 초매개변수의 범위와 간격을 미리 정해 각 경우의 수에 모두 대입하여 최적의 경우의 수를 찾는다. Random search는 정해진 범위에서 난수를 생성해 최적의 초매개변수를 찾으며, bayesian search는 surrogate 모델을 따로 설정하고 이 모델의 매개변수를 업데이트하며 초매개변수를 찾는다(Charu, 2018). 이 중 본 연구는 온실 시스템에 대한 기초적인 분석 결과를 위하여 가장 기본적인 grid search 방법을 통하여 초매개변수 최적화를 수행하였다.

3.2 최적화 함수

최적화 함수는 앞서 설명한 매개변수의 가중치(weight)와 편향(bias)을 최적화하기 위해 사용되는 방법으로 다양한 방법들이 존재한다. 일반적으로 딥러닝은 학습되는 가중치가 많으므로 학습에 계산 부하가 크게 발생한다. 따라서 확률적

경사 하강법을 활용하여 일부 데이터의 샘플만 무작위로 추출해 학습 속도를 개선할 수 있다. 초매개변수 중 하나인 학습률이 너무 크면 학습이 빠르게 진행되지만, 최적화 지점을 지나쳐버릴 수 있고 학습률이 너무 작으면 학습 시간이 증가하며 과소 적합의 문제가 발생한다. 따라서 본 연구에서는, 아담(adam) 최적화 함수를 사용하여 모델을 구성하였으며, 학습률 초깃값을 0.001로 설정하여 모델학습을 수행하였다.

3.3 손실함수

손실함수(loss function)는 모델을 통하여 예측된 결과와 실제값과의 차이를 계산하여 모델 최적화를 수행하는 방법으로, 손실함수의 종류는 MSE(mean squared error), MAE(mean absolute error), binary cross entropy, categorical cross entropy sparse, categorical cross entropy 등이 있다. 기계학습에서 신경망 성능을 측정하는 손실함수는 모델이 올바른 방향으로 학습될 수 있는 역할을 수행한다(Charu, 2018). 본 연구에는 대부분의 회기 예제에 활용되는 MSE를 적용하여 분석이 이루어졌으며, 오차가 증가할수록 큰 값이 도출되며 오차가 감소할수록 작은 값이 도출된다.

3.4 모델 개발방법

신경망의 초기화는 모델의 안정성 문제들과 관련이 있으며, 초기 입력값이 변화할 경우 동일한 데이터로 반복 학습을 수행하더라도 다른 결과가 도출될 수 있다. 따라서 항상 일정한 초깃값이 입력되도록 “Seed()” 함수를 통하여 입력시드를 고정하여 모델 개발이 이루어졌다. 이를 통하여 반복 학습에 대한 모델의 재현성을 달성하였으나 반복계산횟수의 증가 및 최적화 함수의 내부 알고리즘에 따라 변화하는 학습률로 인하여 완벽히 동일한 결과를 도출하지 않는다. 또한 이전 수행된 시계열 데이터에 대한 연구결과(Kwon 등, 2021)를 바탕으로 아래 Table 1의 조건으로 모델 개발이 수행되었다.

기계학습모델은 반복계산(iteration)을 통하여 매개변수의

Table 1. Model development conditions (Kwon et al., 2021).

| Item | Value |
|------------------|--------------|
| Algorithms | RNN/LSTM/GRU |
| Hidden layer | 1-20 |
| Optimizer | Adam |
| Learning rate | 0.001 |
| Loss function | MSE |
| Number of epochs | 200 |
| Batch size | 256 |

가중치가 업데이트되어 학습이 이루어지고 모델 성능이 개선된다(Muller와 Guido, 2017; Kim와 Jung, 2017). 모델 개발은 Python 기반의 TensorFlow를 이용하여 수행되었으며 Keras 딥러닝 라이브러리를 사용하였다.

또한 인공지능 지도학습을 통하여 모델 개발 시 훈련데이터와 검증데이터를 명확히 분리하지 않으면 모델의 예측 신뢰성이 낮아지며, 시계열데이터의 경우 미래의 데이터가 과거를 예측하는 데 사용될 수 없다. 따라서 예측 모델 8일간 데이터를 이용하여 학습하고 이후 1일 데이터를 통한 학습의 검증(validation)에 활용하였으며, 이후 훈련에 사용되지 않는 1일 데이터를 활용하여 최종적인 모델 확인(verification)이 이루어졌다(Fig. 5). 대상 유리온실의 2021년 9월 17일에서 9월 27일까지 10일간의 1분 단위의 데이터(14,400개)를 이용하여 이루어졌다.

4. 실험 및 시뮬레이션 결과 비교·분석 방법

개발 모델은 결정계수(coefficient of determination)(Kim 등, 2007)를 통하여 실험값과 시뮬레이션 값의 정확도(accuracy) 비교분석(식 2)이 이루어졌으며 평균 제곱근 오차(root mean square error, RMSE)를 통하여 모델 정밀도(precision)가 분석되었다(식 3).

$$r^2 = \frac{n \left(\sum_{i=1}^n X_i Y_i \right) - \left(\sum_{i=1}^n X_i \right) \left(\sum_{i=1}^n Y_i \right)}{\sqrt{\left[n \sum_{i=1}^n X_i^2 - \left(\sum_{i=1}^n X_i \right)^2 \right] \left[n \sum_{i=1}^n Y_i^2 - \left(\sum_{i=1}^n Y_i \right)^2 \right]}} \quad (2)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (d_i^2)} \quad (3)$$

d_i : difference between i th estimated and i th measured values

Y_i : i th measured value

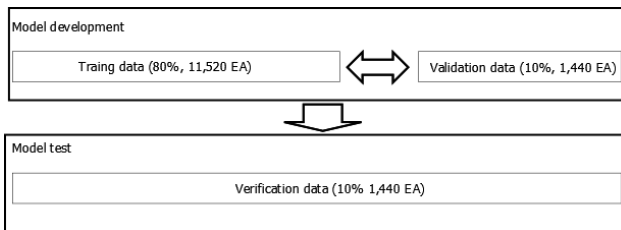


Fig. 5. Split data of development model and test model.

X_i : i th estimated value
 R^2 : coefficient of determination
 RMSE : root mean square error

결과 및 고찰

1. 실험 데이터

기계학습모델을 개발하기에 앞서 측정 데이터의 분석이 이루어졌다. 대상 온실의 10일간의 외부온도는 최저 20°C에서 최대 30°C까지 나타났으며, 온실 내부온도는 온실의 보온특성으로 인하여 최대 40°C까지 상승하였다. 또한, 감우데이터를 살펴보면 2일간(21일-22일) 비로 인하여 높은 습도 유지 및 비교적 낮은 온도 특성이 나타났다(Fig. 6). 이렇게 다양한 데이터 특성이 기계학습모델에 학습되어 내부온도 예측 모델 개발이 수행되었다.

2. 개발 모델 사례분석 결과

대상 데이터에 대하여 각 세 가지 알고리즘(RNN, LSTM, GRU)과 숨은층(hidden layer)의 변화에 따른 사례분석이 수행되었으며 결과는 Table 2와 Fig. 7로 나타났다. Table 2을 살펴보면 모든 개발 모델은 숨은층 증가에 따라 평균 파라미터의 개수가 증가하였으며 RNN 알고리즘의 파라미터가 평균 184개로 가장 적고 GRU 530개, LSTM 734개 순서로 많게 나타났다. 개발 모델의 평균 기본적이고 단순한 알고리즘인

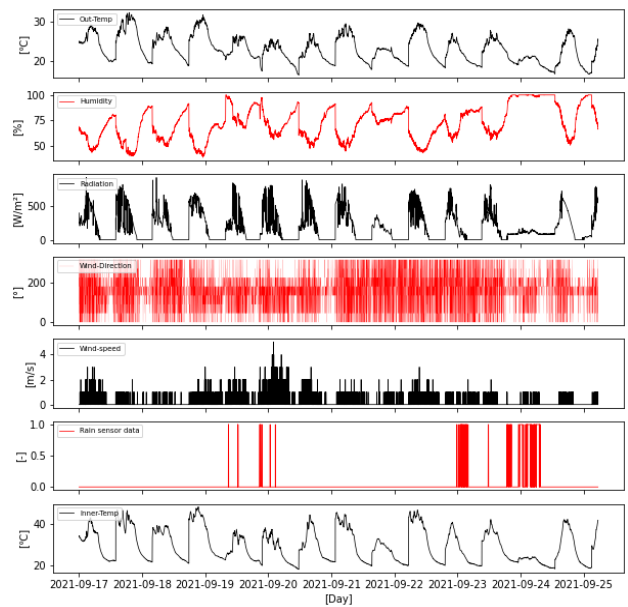


Fig. 6. Measurement data of greenhouse system.

Table 2. Model parameter change according to the hidden unit and the algorithm.

| Hidden layer | RNN algorithm | | | LSTM algorithm | | | GRU algorithm | | | | |
|--------------|----------------------|---------------------------|----------------------|----------------------|---------------------------|----------------------|----------------------|---------------------------|----------------------|--------|--------|
| | Number of parameters | r ² validation | RMSE validation [°C] | Number of parameters | r ² validation | RMSE validation [°C] | Number of parameters | r ² validation | RMSE validation [°C] | | |
| 1 | 6 | 0.7445 | 0.8488 | 2.5394 | 0.8946 | 1.5196 | 18 | 0.6414 | 0.6992 | 3.2246 | 6.7777 |
| 2 | 14 | 0.9020 | 0.9075 | 1.4922 | 0.8829 | 1.6426 | 42 | 0.9730 | 0.9209 | 1.4465 | 2.9592 |
| 3 | 24 | 0.8431 | 0.9172 | 2.1898 | 0.7972 | 0.8930 | 72 | 0.9101 | 0.9156 | 1.6974 | 3.2127 |
| 4 | 36 | 0.9642 | 0.9661 | 0.9180 | 0.9443 | 0.8735 | 108 | 0.9723 | 0.9620 | 0.9473 | 2.1083 |
| 5 | 56 | 0.8621 | 0.9597 | 1.7510 | 0.9057 | 0.8723 | 150 | 0.9655 | 0.9469 | 0.8835 | 2.5263 |
| 6 | 66 | 0.9885 | 0.9681 | 0.5650 | 0.9769 | 0.9641 | 198 | 0.9848 | 0.9771 | 0.5857 | 1.7417 |
| 7 | 84 | 0.9750 | 0.9511 | 1.1733 | 0.9687 | 0.9567 | 252 | 0.9860 | 0.9781 | 0.6621 | 1.6671 |
| 8 | 104 | 0.9454 | 0.9603 | 1.3385 | 0.9754 | 0.9724 | 312 | 0.9873 | 0.9777 | 0.5822 | 1.6281 |
| 9 | 126 | 0.9405 | 0.9210 | 1.4889 | 0.9847 | 0.9669 | 378 | 0.9879 | 0.9683 | 0.9221 | 1.8285 |
| 10 | 135 | 0.9780 | 0.9654 | 0.7257 | 0.9744 | 0.9668 | 395 | 0.9642 | 0.9660 | 1.1494 | 1.8836 |
| 11 | 151 | 0.9811 | 0.9753 | 0.9101 | 0.9909 | 0.9642 | 440 | 0.9857 | 0.9506 | 0.6944 | 2.2622 |
| 12 | 167 | 0.9697 | 0.9797 | 0.8745 | 0.9890 | 0.9595 | 485 | 0.9790 | 0.9707 | 0.8840 | 1.7369 |
| 13 | 183 | 0.9111 | 0.9398 | 2.1175 | 0.9806 | 0.9726 | 530 | 0.9832 | 0.9736 | 0.9854 | 1.7160 |
| 14 | 199 | 0.8815 | 0.9584 | 1.6106 | 0.9888 | 0.9780 | 575 | 0.9887 | 0.9708 | 0.8209 | 1.7710 |
| 15 | 215 | 0.9812 | 0.9787 | 0.6438 | 0.9849 | 0.9709 | 620 | 0.9905 | 0.9785 | 0.4570 | 1.5159 |
| 16 | 231 | 0.9902 | 0.9847 | 0.7402 | 0.9768 | 0.9530 | 665 | 0.9928 | 0.9843 | 0.5046 | 1.2912 |
| 17 | 247 | 0.9920 | 0.9821 | 0.4348 | 0.9919 | 0.9726 | 710 | 0.9855 | 0.9788 | 0.9606 | 1.5409 |
| 18 | 263 | 0.9892 | 0.9582 | 0.5189 | 0.9874 | 0.9759 | 755 | 0.9864 | 0.9847 | 0.5460 | 1.2618 |
| 19 | 279 | 0.9399 | 0.9446 | 1.4737 | 0.9913 | 0.9766 | 800 | 0.9791 | 0.9674 | 1.0168 | 1.8526 |
| 20 | 295 | 0.9553 | 0.9002 | 1.7534 | 0.9881 | 0.9590 | 845 | 0.9889 | 0.9772 | 0.6276 | 1.5957 |
| 21 | 311 | 0.9879 | 0.9830 | 0.6611 | 0.9870 | 0.9676 | 890 | 0.9881 | 0.9662 | 0.8133 | 2.0352 |
| 22 | 327 | 0.9901 | 0.9774 | 0.8786 | 0.9880 | 0.9748 | 935 | 0.9858 | 0.9766 | 0.7870 | 1.5804 |
| 23 | 343 | 0.9954 | 0.9825 | 0.3226 | 0.9924 | 0.9767 | 980 | 0.9866 | 0.9762 | 0.6853 | 1.6415 |
| 24 | 358 | 0.9918 | 0.9822 | 0.5264 | 0.9853 | 0.9734 | 1025 | 0.9892 | 0.9703 | 0.7782 | 1.8193 |
| 25 | 374 | 0.9457 | 0.9381 | 1.1543 | 0.9772 | 0.9642 | 1070 | 0.9914 | 0.9736 | 0.5841 | 1.7670 |
| Average | 184 | 0.9458 | 0.9532 | 1.1521 | 0.9673 | 0.9512 | 530 | 0.9669 | 0.9564 | 0.9298 | 2.0688 |

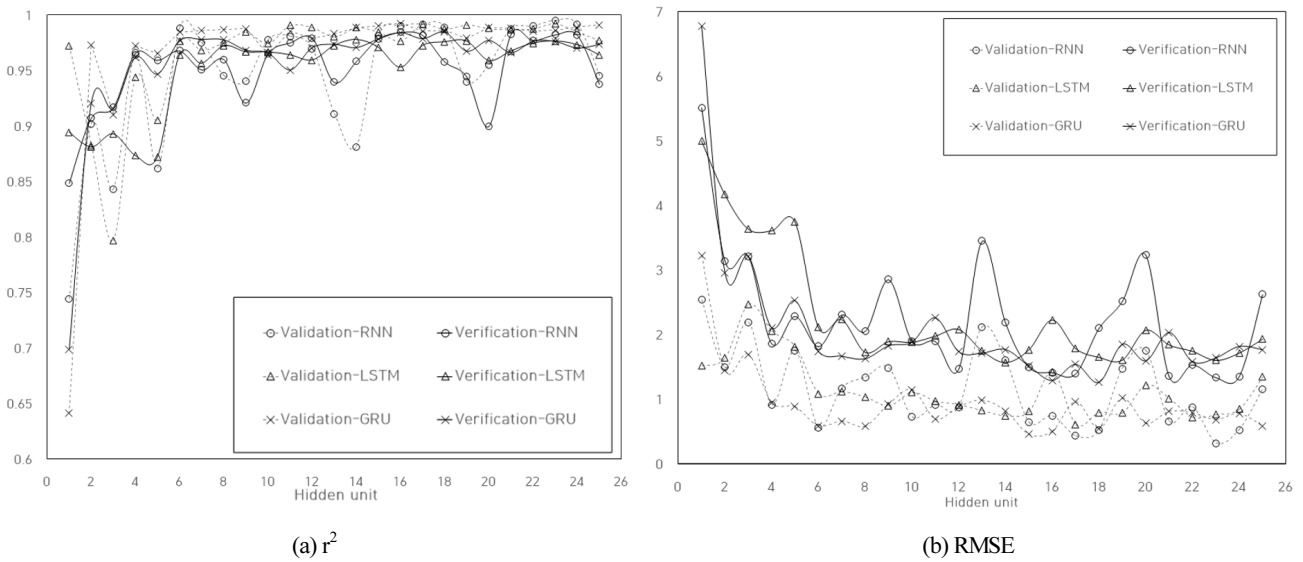


Fig. 7. Model validation and verification results according to the algorithm.

RNN의 정확도는 훈련모델 0.9458로 분석 알고리즘 중 가장 낮게 도출되었다. 또한, 다른 두 알고리즘의 훈련모델 정확도는 각각 LSTM 0.9673, GRU 0.9669로 비슷하게 나타났는데 알고리즘에 추가된 게이트 특성으로 인한 결과로 판단된다. 각 알고리즘의 테스트모델 평균 정밀도는 RNN 0.9532, LSTM 0.9512, GRU 0.9564로 모두 0.95 이상으로 모두 시계열 데이터 시뮬레이션 모델링에 적절한 알고리즘으로 판단된다.

Fig. 7의 히든 유닛 증가에 따른 모델 특성 분석 결과 두 가지 모델의 척도를 살펴보면 숨은층 6 이상에서 안정적인 결과가 나타났다. 숨은층 증가에 따른 모델 정확도(Fig. 5(a))는 대부분 비슷한 경향을 나타내었으며, 모델 정밀도(Fig. 5(b))는 테스트 결과보다 훈련결과가 높게 나타나는 기계학습 특성이 나타났다. 이렇게 파라미터에 따라 다양한 모델 성능 차이가 발생하기 때문에 최적화를 위한 분석이 이루어져야 한다.

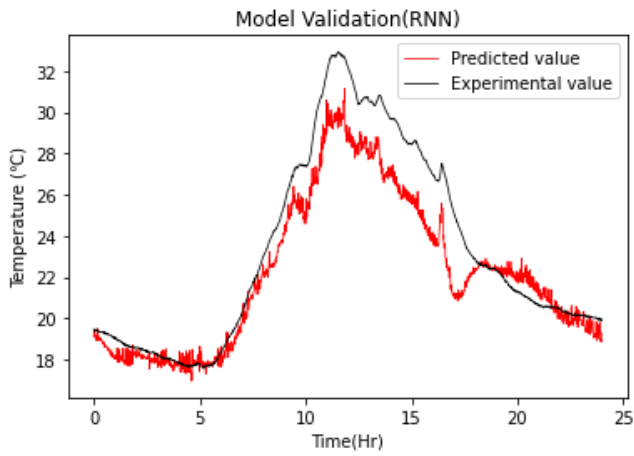
RNN 알고리즘의 경우 6 이하의 숨은층에서 모델 정확도 및 정밀도가 낮게 나타나는데 이는 비교적 적은 매개변수로 인하여 과소 적합(under fitting)이 발생한 것으로 사료된다. 따라서 기계학습모델의 학습 시 숨은층의 개수는 최소한 6개 이상으로 선정하는 것이 적절하다고 판단된다. 또한, RNN 알고리즘의 경우 숨은층이 증가함에도 모델 안정화가 이루어지지 않는 것으로 나타나는데 이는 기울기 폭발 및 소실로 인하여 모델 오차가 증가하는 것으로 판단된다. 또한, LSTM과 GRU 알고리즘은 비슷한 수준의 모델 정확도와 정밀도 결과가 도출되었다. 하지만 LSTM 알고리즘은 GRU 알고리즘 보다 약 30% 이상의 많은 숨은층을 가지고 있으며 모델 개발 시 파라미터 개수가 늘어남에 따라 높은 계산부하량이 요구되기 때문에 모델

학습 시간이 증가하는 문제점이 발생한다. 따라서 최적의 정확도와 정밀도를 가진 모델을 도출하기 위하여 GRU 알고리즘이 대상 시스템에 가장 적합한 알고리즘이라 판단된다. 따라서 최종적으로 본 연구의 대상 시스템의 데이터에 대한 최적의 숨은층은 6개, GRU 알고리즘으로 선정되었다.

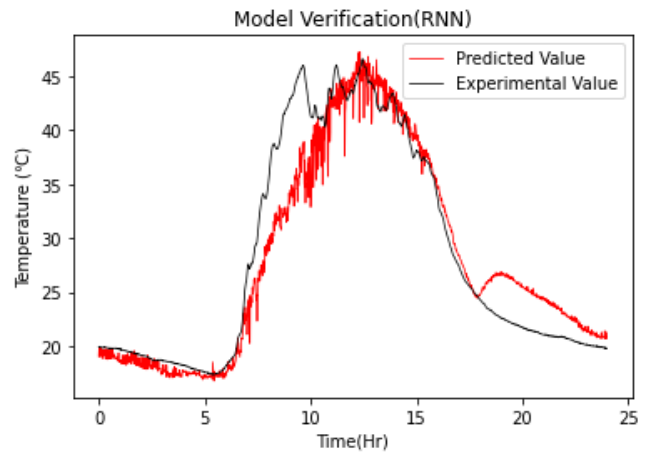
3. 실험 및 시뮬레이션 결과 비교

숨은층 25개에서 알고리즘에 따른 온실 내부온도 예측 결과는 Fig. 8과 같다. Fig. 8의 (a), (c), (e)의 모델 훈련 결과를 살펴보면 RNN 알고리즘의 경우 너무 많은 학습으로 인하여 과대 적합이 발생하여 약 4°C 이상의 오차를 나타내고 있으며, GRU와 LSTM 알고리즘의 경우 약 2°C 이내의 상대적으로 낮은 오차가 나타났다. 또한, Fig. 8의 (b), (d), (f)의 모델 테스트 결과 RNN 알고리즘은 오전 8시에서 10시 사이의 온도 상승 부분을 잘 예측하지 못하는 한계점이 나타났으며, GRU와 LSTM 알고리즘은 대부분의 시간대에서 모델이 잘 예측하는 결과를 도출하였지만 16시에서 20시 구간에서 서로 다른 예측 특성이 나타났다. 이는 두 모델의 서로 다른 게이트 숫자와 메모리 블록 특성에서 나타난 결과라 판단된다.

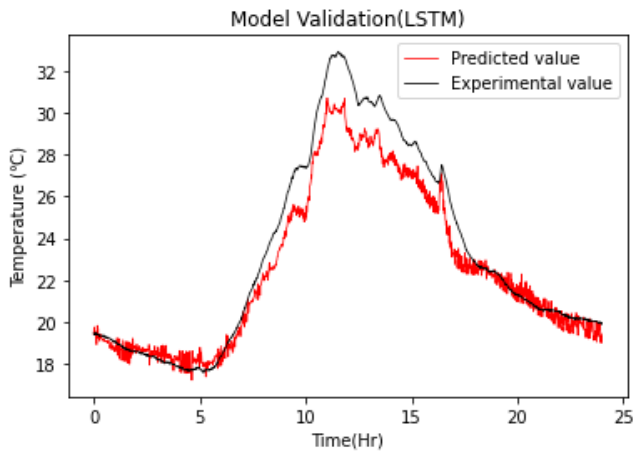
본 연구에서는 기본적인 시계열 알고리즘을 활용하여 온실 내부온도 예측 모델이 개발되었으며, 예측 결과 일정 파라미터 개수 이상의 학습 구간에서 약 2°C 이내의 오차의 성능을 가지는 모델 정확도를 나타냈다. 더 나아가 본 연구에서 다루지 못한 추가적인 파라미터의 및 학습방법 개선을 통하여 더욱 높은 성능의 모델 개발이 필요할 것으로 판단된다.



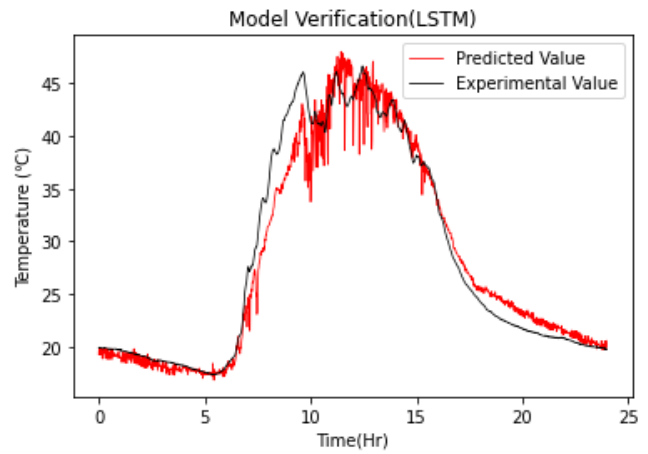
(a) Training result of RNN algorithm



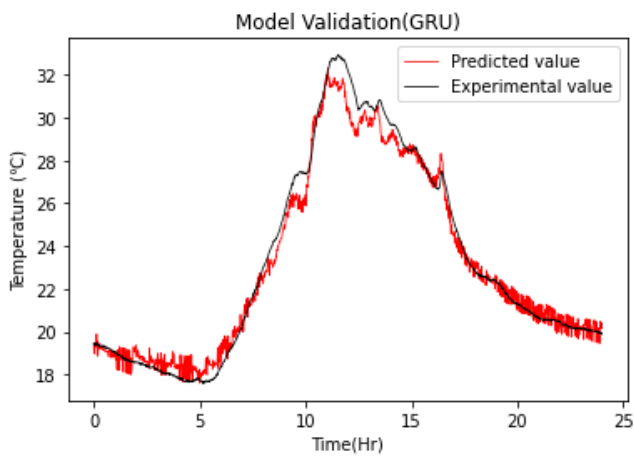
(b) Test result of RNN algorithm



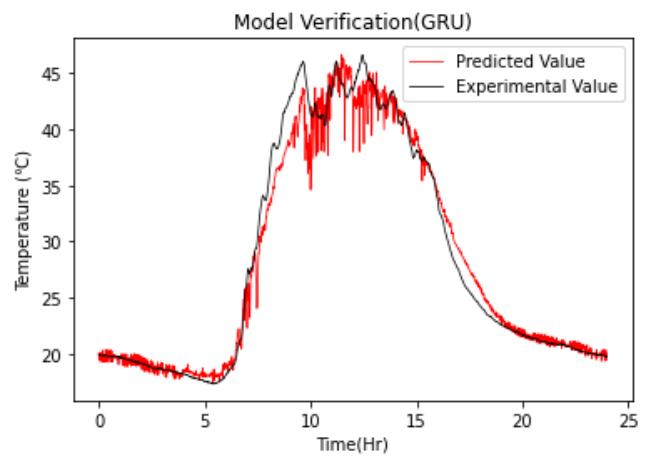
(c) Training result of LSTM algorithm



(d) Test result of LSTM algorithm



(e) Training result of GRU algorithm



(f) Test result of GRU algorithm

Fig. 8. Comparison of experimental and simulation results according to the algorithm (at hidden unit 6).

적 요

본 연구는 데이터를 기반으로 한 인공지능 기계학습 기법을 활용하여 온실 내부온도 예측 시뮬레이션 모델을 개발을 수행하였다. 온실 시스템의 내부온도 예측을 위해서 다양한 방법이 연구됐지만, 가외 변인으로 인하여 기존 시뮬레이션 분석 방법은 낮은 정밀도의 문제점을 지니고 있다. 이러한 한계점을 극복하기 위하여 최근 개발되고 있는 데이터 기반의 기계학습을 활용하여 온실 내부온도 예측 모델 개발을 수행하였다. 기계학습모델은 데이터 수집, 특성 분석, 학습을 통하여 개발되며 매개변수와 학습방법에 따라 모델의 정확도가 크게 변화된다. 따라서 데이터 특성에 따른 최적의 모델 도출방법이 필요하다. 모델 개발 결과 습은층 증가에 따라 모델 정확도가 상승하였으며 최종적으로 GRU 알고리즘과 습은층 6에서 r^2 0.9848과 RMSE 0.5857°C로 최적 모델이 도출되었다. 본 연구를 통하여 온실 외부 데이터를 활용하여 온실 내부온도 예측 모델 개발이 가능함을 검증하였으며, 추후 다양한 온실데이터에 적용 및 비교분석이 수행되어야 한다. 이후 한 단계 더 나아가 기계학습모델 예측(predicted) 결과를 예보(forecasting) 단계로 개선하기 위해서 데이터 시간 길이(sequence length)에 따른 특성 분석 및 계절별 기후변화와 작물에 따른 사례별로 개발 모델을 관리하는 등의 다양한 추가 연구가 수행되어야 한다.

추가주제어: 기계학습, 시계열 데이터, 시뮬레이션 모델, 인공지능, 자동화 온실

사 사

이 논문은 2022년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(2021R1A6A1A0304424211 and 2022R1C1C2009821).

Literature Cited

- Ahn W.Y., and H.C. Lee 2016, Automatic control system for cultivation environment of crops. *J Korea Inst Inf Commun Eng* 20:2167-2171. (in Korean) doi:10.6109/jkiice.2016.20.11.2167
- Ariga M., S. Nakayama, and D. Nishibayasi 2018, *Machine learning at work*. O'Reilly Media, CA, USA, pp 137-140.
- Chang Z., Y. Zhang, and W. Chen, 2019, Electricity price prediction based on hybrid model of adam optimized LSTM neural network and wavelet transform. *Energy* 187:115804. doi:10.1016/j.energy.2019.07.134
- Charu C.A. 2018, *Neural networks and deep learning: a Textbook*. Springer, Heidelberg, Germany, pp 123-127.
- Cho K., B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio 2014, Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*. doi: 10.48550/arXiv.1406.1078
- Cho K.J., K.Y. Kim, and W.M. Yang 2015, Survey of ICT apply to plastic greenhouse, rack·pinion adaption to single span and CFD analysis. *Protected Hort Plant Fac* 24:308-316. (in Korean) doi:10.12791/KSBEC.2015.24.4.308
- Choi H.Y., T.W. Moon, D.H. Jung, and J.E. Son 2019, Prediction of air temperature and relative humidity in greenhouse via a multilayer perceptron using environmental factors. *Protected Hort Plant Fac* 28:95-103. (in Korean) doi:10.12791/KSBEC.2019.28.2.95
- Choi Y.S., H.J. Lee, and S.T. Joung 2012, A design and implementation of web-based green house automation system. *J Korea Inst Electron Commun Sci* 1519-1527. (in Korean) doi:10.13067/JKIECS.2012.7.6.1519
- Fatnassi H., C. Poncet, M.M. Bazzano, R. Brun, and N. Bertin 2015, A numerical simulation of the photovoltaic greenhouse microclimate. *Solar Energy* 120:575-584. doi:10.1016/j.solener.2015.07.019
- François C. 2017, *Deep learning with Python*. Manning Publications Company, NY, USA, pp 79-92.
- Gabriel K.R. 1971, The biplot graphic display of matrices with application to principal component analysis. *Biometrika* 58:453-467. doi:10.1093/biomet/58.3.453
- Hong S.W., and I.B. Lee 2014, Predictive model of micro-environment in a naturally ventilated greenhouse for a model-based control approach. *Protected Hort Plant Fac* 23:181-191. (in Korean) doi:10.12791/KSBEC.2014.23.3.181
- Hope T., Y.S. Resheff, and I. Lieder 2017, *Learning tensorflow: A guide to building deep learning systems*. O'Reilly Media, CA, USA, pp 84-91.
- Huh M.H. 2017, Representing variables in the latent space. *Korean J Appl Stat* 30:555-566. (in Korean) doi:10.5351/KJAS.2017.30.4.555
- Hwang I.C., H. Noh, D. Yang, and M. Kim 2021, Prediction of paprika yield using multiple linear regression. *J Korean Inst Commun* 46:21-11. (in Korean) doi:10.7840/kics.2021.46.11.2048
- Kim D.H., B.M. Jenkins, T.R. Rumsey, M.W. Yore, and N.J. Kim 2007, Simulation and model validation of a horizontal shallow basin solar concentrator. *Solar Energy* 81(4):463-475. doi:10.1016/j.solener.2006.08.007
- Kim H.S. 2001, Prediction of cooling effect for fog cooling system in greenhouse by CFD simulation. Master Diss., Seoul National University, Seoul, Korea, pp 1-45. (in Korean)
- Kim S.Y., and Y.J. Jung 2017, *First learning machine learning*. Hanbit Media, Seoul, Korea, pp 137-140. (in Korean)

- Kwon H.W., K.C. Oh, Y. Choi, Y.G. Chung, and J. Kim 2021, Development and application of machine learning-based prediction model for distillation column. *Int J Intell* 36: 1970-1997. doi:10.1002/int.22368
- Lee I.B., and T.H. Short 1999, Analysis of the efficiency of natural ventilation in multi-span greenhouse using CFD simulation. *J Bio-Env Con* 8:9-18. (in Korean)
- Lee I.B., N.K. Yun, T. Boulard, J.C. Roy, S.H. Lee, G.W. Kim, S.K. Lee, and S.H. Kwon 2006a, Development of an aerodynamic simulation for studying microclimate of plant canopy in greenhouse: (1) Study on aerodynamic resistance of tomato canopy through wind tunnel experiment. *J Bio-Env Con* 15:289-295. (in Korean)
- Lee I.B., N.K. Yun, T. Boulard, J.C. Roy, S.H. Lee, G.W. Kim, S.W. Hong, and S.H. Sung 2006b, Development of an aerodynamic simulation for studying microclimate of plant canopy in greenhouse: (2) Development of CFD model to study the effect of tomato plants on internal climate of greenhouse. *J Bio-Env Con* 15:296-305. (in Korean)
- Lee J.K., J.W. Oh, Y.J. Cho, and D.H. Lee 2020, A research about time domain estimation method for greenhouse environmental factors based on artificial intelligence. *Protected Hort Plant Fac* 3:277-284. (in Korean) doi:10.12791/KSBEC.2020.29.3.277
- Muller A.C., and S. Guido 2017, Introduction to machine learning with Python: a guide for data scientists. O'Reilly Media, CA, USA, pp 27-31.
- Na M.H., Y. Pack, and W. Cho 2017, A study on optimal environmental factors of tomato using smart farm data. *J Korean Data Inf Sci Soc* 28:1427-1435 doi:10.7465/jkdi.2017.28.6.1427
- Oh K.C., H.W. Kwon, J.W. Roh, Y.Y. Cho, H.D. Park, H.T. Cho, and J.H. Kim 2020, Development of machine learning-based platform for distillation column. *Korean Chem Eng Res* 58:565-572. (in Korean) doi:10.9713/kcer.2020.58.4.565
- Park Y.M., S.M. Gang, J.H. Chae, and J.J. Lee 2018, Classification method of plant leaf using DenseNet. *J Korea Multimedia Soc* 21:571-582. doi:10.9717/kmms.2018.21.5.571
- Song Y.E., A.Y. Moon, S.Y. An, and H.Y. Jung 2019, Prediction of smart greenhouse temperature-humidity based on multi-dimensional LSTMs. *J Korean Soc Precis Eng* 36(3):239-246. (in Korean) doi:10.7736/KSPE.2019.36.3.239
- Tadj N., T. Bartzanas, D. Fidaros, B. Draoui, and C. Kittas 2010, Influence of heating system on greenhouse microclimate distribution. *Trans ASABE* 53:225-238. doi:10.13031/2013.29498
- Yu I.H., N.K. Yun, M.W. Cho, H.R. Ryu, and D.G. Moon 2014, Development of CFD model for analyzing the air flow and temperature distribution in greenhouse with air-circulation fans. *Korean J Agric Sci* 41:461-472. (in Korean) doi:10.7744/cnujas.2014.41.4.461