

Comparison of Machine Learning-Based Greenhouse VPD Prediction Models

Jang Kyeong Min[†] · Lee Myeong Bae^{††} · Lim Jong Hyun^{†††} · Oh Han Byeol^{††††} ·
Shin Chang Sun^{†††††} · Park Jang Woo^{†††††}

ABSTRACT

In this study, we compared the performance of machine learning models for predicting Vapor Pressure Deficits (VPD) in greenhouses that affect pore function and photosynthesis as well as plant growth due to nutrient absorption of plants. For VPD prediction, the correlation between the environmental elements in and outside the greenhouse and the temporal elements of the time series data was confirmed, and how the highly correlated elements affect VPD was confirmed. Before analyzing the performance of the prediction model, the amount and interval of analysis time series data (1 day, 3 days, 7 days) and interval (20 minutes, 1 hour) were checked to adjust the amount and interval of data. Finally, four machine learning prediction models (XGB Regressor, LGBM Regressor, Random Forest Regressor, etc.) were applied to compare the prediction performance by model. As a result of the prediction of the model, when data of 1 day at 20 minute intervals were used, the highest prediction performance was 0.008 for MAE and 0.011 for RMSE in LGBM. In addition, it was confirmed that the factor that most influences VPD prediction after 20 minutes was VPD (VPD_{y_71}) from the past 20 minutes rather than environmental factors. Using the results of this study, it is possible to increase crop productivity through VPD prediction, condensation of greenhouses, and prevention of disease occurrence. In the future, it can be used not only in predicting environmental data of greenhouses, but also in various fields such as production prediction and smart farm control models.

Keywords : Machine Learning, Vapor Pressure Deficit(VPD), Prediction model, Smart Farm, Light Gradient Boosting Machine(LGBM), RandomForest

머신러닝 기반의 온실 VPD 예측 모델 비교

장 경 민[†] · 이 명 배^{††} · 임 종 현^{†††} · 오 한 별^{††††} · 신 창 선^{†††††} · 박 장 우^{†††††}

요 약

본 연구에서는 식물의 영양분 흡수에 따른 식물 성장뿐만 아니라 기공 기능 및 광합성에도 영향을 끼치는 온실의 수증기압차(VPD, Vapor Pressure Deficit)예측을 위한 머신러닝 모델들의 성능을 비교해보았다. VPD 예측을 위해 온실 내·외부 환경요소 및 시계열 데이터의 시간적 요소들 간의 상관관계를 확인하고 상관관계가 높은 요소들이 VPD에 어떤 영향을 미치는지 확인하였다. 예측 모델의 성능을 분석하기 전 분석 시계열 데이터의 양(1일, 3일, 7일), 간격(20분, 1시간)이 예측 성능에 미치는 영향을 확인하여 데이터의 양과 간격을 조절하였다. 마지막으로 4개의 머신러닝 예측 모델(XGB Regressor, LGBM Regressor, Random Forest Regressor 등)을 적용하여 모델별 예측 성능을 비교했다. 모델의 예측 결과로 20분 간격의 1일의 데이터를 사용했을 때 LGBM에서 MAE는 0.008, RMSE는 0.011의 가장 높은 예측 성능을 보였다. 또한 20분 후 VPD 예측에 가장 큰 영향을 미치는 요소는 환경적 요인보다는 과거 20분 전의 VPD(VPD_{y_71})임을 확인하였다. 본 연구의 결과를 활용하여 VPD 예측을 통해 작물의 생산성을 높이고, 온실의 결로, 병 발생 예방 등이 가능하다. 향후 온실의 환경 데이터 예측뿐만 아니라 더 나아가 생산량 예측, 스마트팜 제어 모델 등 다양한 분야에 활용할 수 있을 것이다.

키워드 : 기계학습, 수증기압차(VPD), 예측모델, 스마트팜, Light Gradient Boosting Machine(LGBM), RandomForest

1. 서 론

시설 재배의 경우 외부의 온·습도, 바람 등 외부 날씨의 영향을 줄이고 온도, 습도, 양액 등 작물 생장에 필요한 환경적 요소를 최적의 조건으로 제어할 수 있어 노지 재배에 비해 작물의 수확량과 품질 향상이 가능하다[1]. 이러한 장점으로 실제 우리나라의 농업지역 경지 면적이 지속적으로 축소되는 반면(1988년 2,138.0×103ha → 2018년 1,595.0×103ha, 약 25.4% 감소) 시설재배의 수와 면적이 꾸준히 증가(2018

※ This research was supported by "Regional Innovation Strategy (RIS)" through the National Research Foundation of Korea funded by the Ministry of Education(MOE)(2021RIS-002).

† 정 회 원 : 순천대학교 정보통신공학전공 석사과정

†† 준 회 원 : 순천대학교 정보통신공학전공 박사수료

††† 준 회 원 : 순천대학교 정보통신공학전공 박사과정

†††† 비 회 원 : 순천대학교 정보통신공학전공 박사과정

††††† 정 회 원 : 순천대학교 인공지능공학부 교수

Manuscript Received : November 30, 2022

First Revision : January 13, 2023

Accepted : January 25, 2023

* Corresponding Author : Park Jang Woo(jwpark@snu.ac.kr)

년 53,274.2ha, 약 135.7% 증가)하고 있다[2].

작물의 성장에 관여하며 온실에서 제어 가능한 요소 중 수증기압차(VPD, Vapor Pressure Deficit)는 온실에서 수집되는 온도와 상대습도 등을 통해 계산되는 값으로 식물 내부의 물 수송의 주요 원동력이며, 식물의 영양분 흡수에 영향을 미치는 요소이다. 식물의 영양분 흡수에 따른 식물 성장뿐만 아니라 기공 기능 및 광합성에도 영향을 끼친다[3, 4]. 예를 들어 VPD 수치가 높은 경우(2.2kPa 이상, 습도가 높을 경우) 식물은 기공을 닫게 되고, 기공 폐쇄로 인해 광합성은 감소하며, 효소 활동을 포함한 대사활동이 제한되어 영양분의 흡수와 축적을 감소시키는 경향이 있다[5, 6]. 실제로 VPD를 조절하여 토마토의 총 수확량을 개선한 연구 사례도 있다[7]. 또한 VPD는 결로나 증산에 있어서 수증기 흐름을 표시할 수 있는 유용한 수치로, 병 발생 위험, 결로 여부, 관개 시기 등을 평가하는데 사용되고 있다[8].

최근 수십 년 동안 전 세계적으로 온도가 상승하여 평균 VPD가 기하급수적으로 증가하였고 노지 재배의 경우 온도와 습도의 조절이 불가능하여 식물 성장 최적의 VPD를 유지하기 어려워졌다. 하지만 시설 재배의 경우의 내부 환경 요소를 조절하여 표준 식물 최적의 VPD인 0.4kPa~0.8kPa를 유지해줄 수 있다[9, 10]. 특히 파프리카 온실의 온도는 주간 22~25°C, 야간 18~20°C, 습도는 70~75%로 관리해주어야 한다. 이러한 환경 조건에 적합한 VPD는 주간 0.66kPa~0.95kPa, 야간 0.52kPa~0.70kPa이다[11, 12]. 네덜란드의 대표적인 스마트팜 솔루션 기업인 '프리카바(privava)'의 온실 자동 관리 시스템에서는 VPD 데이터를 계산하여 수집하고 있으며, 미국의 상업용 온실 회사인 Argus Control Systems에서는 VPD의 제어를 위해 1.25kPa 이상이 되면 온실 내부에 가습 시스템을 가동하여 VPD를 조절하여 온실의 환경을 유지한다[13].

본 연구에서는 온실 작물의 성장에 영향을 끼치는 VPD의 예측을 위해 온실 내부 데이터를 활용하여 VPD의 값을 구하고, 환경적·시간적 요소들이 VPD에 어떠한 영향을 미치는지 상관관계를 분석해보았다. 그 후 분석에 사용되는 데이터의 양과 간격을 변화시켜 예측 모델에 적절한 데이터를 선정하고, 과거 데이터를 포함하여 다양한 예측 모델을 적용하여 최종적으로 최적의 VPD 예측 모델을 찾았다.

본 연구의 결과를 활용하여 온실의 VPD뿐만 아니라 습도, CO2농도 등의 예측에 적용 가능하다. 또한 본 연구의 향후 주제인 스마트팜 통합 관리 플랫폼에 적용하여 VPD 데이터를 기반으로 한 온실 제어 시스템을 구축해보고자 한다.

2. 데이터 수집 및 전처리

2.1 데이터의 수집

온실 환경 데이터는 전남 고흥에 위치한 벤로형 온실(Venlo type greenhouse)에서 수집했다. 벤로형 온실은 기상재해에 강해 농업인의 선호도가 높아져 많이 사용되고 있

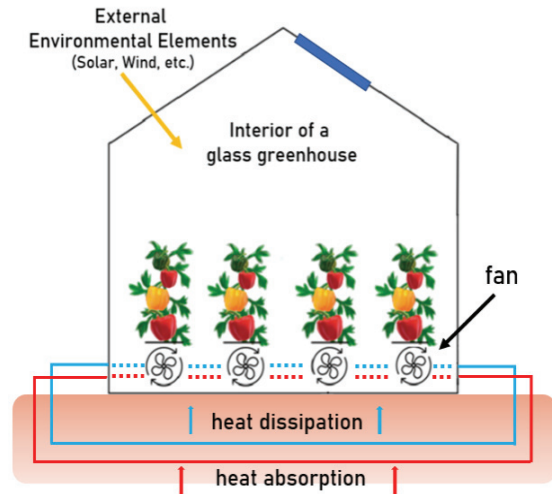


Fig. 1. Structure of Goheung Semi-enclosed Greenhouse

는 유형의 온실이다[14]. 재배하고 있는 작물의 종류는 파프리카이고, 온실은 반밀폐형이며 구조는 Fig. 1과 같다.

천창의 사이즈를 최소화하여 온실의 온·습도를 유지하고 내부 공기의 순환을 위해 외부와 연결된 환기팬이 작동된다. 여름과 겨울철 등 외부와 온실의 온도 차이가 큰 경우에는 지열 히트펌프를 이용하여 온·습도를 조절하여 파프리카 재배에 적합한 환경을 유지하고 있다.

이 온실은 한 개의 동에 총 4개의 블록에서 파프리카를 재배하고 있으며, 블록별로 총 4개의 내부 센서가 설치되어 있다. 본 연구에서는 4개의 블록 중 1번 블록의 데이터를 사용하였다. 이 온실에 설치되어 있는 센서를 통해 수집한 내부 환경 데이터와 외부 환경 데이터는 Table 1과 같이 총 13개로 구성되어 있다.

데이터는 20분 간격으로 수집되었으며 약 14개월(2020-12-30~2022-02-22) 동안 수집된 데이터를 사용하여 분석하였다.

Table 1. Environment Data Collection Items

Sortation	name	Description	Unit
Internal Environment data	in_temp	Air temperature	°C
	Wtemp	Wet bulb temperature	°C
	Dtemp	Dew point temperature	°C
	in_RH	Relative Humidity	%
	Enthalpy	Enthalpy	kJ/m ³
	CO2	CO2 concentration	ppm
External Environment Data	out_temp	Outside temperature	°C
	out_RH	Humidity	%
	wind_d1	wind direction	°
	wind_s	wind speed	m/s
	RI	Radiation intensity	W/m ²
	RI[ave]	Standard radiation intensity	W/m ²
	RI[sum]	Radiation sum	J/cm ²

2.2 데이터의 전처리

온실의 내부 데이터와 외부 데이터가 분리되어 있어 time 을 기준으로 merge를 사용하여 데이터를 취합하였다. 취합 후 데이터를 확인해 본 결과 중복 및 누락은 없었다.

수증기압차(VPD, Vapor Pressure Deficit)는 공기가 포화되었을 때 포함할 수 있는 수분량(SVP, Saturated Vapor Pressure)과 공기중의 수분량(AVP, Active Vapor Pressure) 차이로 계산한다. 각각의 공식은 다음 Equation (1)~(3)과 같다. 이 공식에서 T는 섭씨온도(°C), exp는 Euler's Number라고 불리는 수학 상수로 약 2.71828을 의미한다.

$$SVP = 0.6108 \times \exp\left(\frac{17.27T}{T+237.3}\right) [kPa] \quad (1)$$

$$AVP = SVP \times \frac{RH}{100} \quad (2)$$

$$VPD = SVP - AVP \quad (3)$$

수집된 온실 데이터는 시계열 데이터로(Time Series data) 균등한 시간 간격에서 연속적으로 측정된 값들의 시퀀스이다. 시계열 데이터는 년, 월, 일, 시간, 분 등으로 기록된다[15]. 환경적 요소뿐만 아니라 시간적 요소와의 상관관계를 확인하기 위해 시간별, 일별, 월별, 계절별 데이터로 분리했다. 시간 데이터는 주기적인 속성을 갖지만 순차적인 값으로 표시되지 않는다. 예를 들면 1월 31일과 2월 1일은 인접하지만 시퀀스 형식에서는 31의 차이가 생긴다. 이러한 문제를 해결하기 위해 시간별, 일별, 월별, 계절별로 분리한 일차원 공간에서의 순차적인 시간 데이터를 sin, cos함수를 사용하여 2차원 공간의 연속적인 데이터로 변환하였다[16].

$$hour_x = \sin\left(\left(\frac{360}{24}\right) \times hourmin\right) \quad (4)$$

$$hour_y = \cos\left(\left(\frac{360}{24}\right) \times hourmin\right) \quad (5)$$

Equation (4),(5) 공식을 적용하면 Fig. 2와 같이 시간을 2차원 공간에서의 연속된 표현이 가능해진다.

VPD 데이터를 박스플롯을 사용하여 월별로 분석해보았다. Fig. 3을 보면 6~8월의 데이터에서 다른 달과 다르게 너무 높거나 낮은 수치의 분포를 보였다. 파프리카 농장의 작기가 반영된 결과이다. 농장은 작기를 운영하지 않을 때 온실 내부의 온도, 습도 등의 환경적 제어를 진행하지 않는다.

농장의 작기를 운영(1, 1~5월, 9~12월), 미운영(0, 6~8월)으로 구분하여 분석해보았다. Fig. 4와 같이 운영 기간과 미운영 기간의 VPD 값의 차이가 확인된다. 특히 8시~17시에 확연한 차이를 보인다. 작기 운영 중 VPD는 평균 0.26 kPa, 최소 0kPa, 최대 1.64kPa의 값을 보였으며, 작기 미운영 중 VPD는 평균 0.52kPa, 최소 0kPa, 최대 4.98kPa의 값을 보였다.

본 연구에서는 시간적 주기적인 속성을 반영하기 위해 시간 데이터를 2차원 공간의 연속적 데이터로 변환했기 때문에

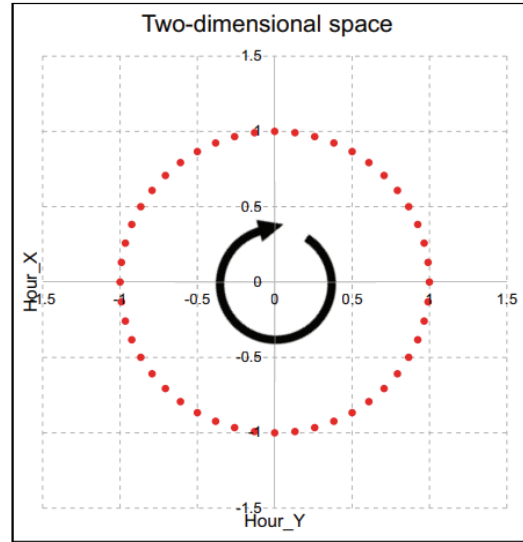


Fig. 2. Two-dimensional Representation of a Day based on 30-min Interval[16]

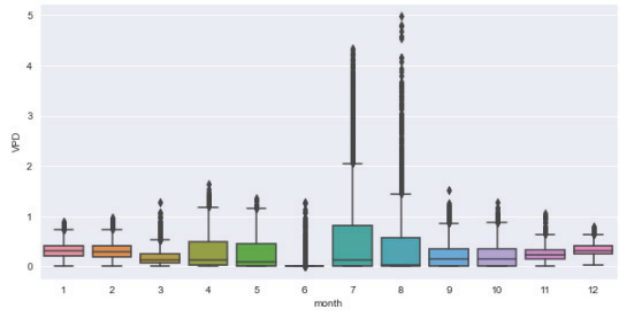


Fig. 3. Monthly VPD Distribution Chart

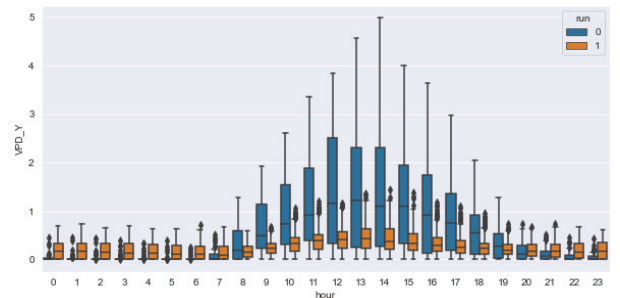


Fig. 4. Hourly VPD Distribution Chart

작기 운영 중(1) 데이터와 미운영(0) 데이터를 모두 사용하여 분석을 진행하였다.

3. 데이터의 분석

3.1 상관관계 분석

VPD와 내·외부 환경적 특성, 시간적 특성 사이의 상관관계를 분석해보았다.

VPD와 내부 환경적 요소 사이의 상관관계를 확인해본 결

과 Fig. 5와 같이 in_temp와 0.58의 양의 상관관계를 보였고 in_RH와 -0.84의 강한 음의 상관관계를 보였다. VPD를 구하는 공식이 반영된 결과로 판단된다. VPD는 온도(in_temp) 보다 습도(in_RH)에 큰 영향을 받는 것을 확인할 수 있다.

VPD와 외부 환경적 요소 사이의 상관관계를 확인해본 결과 Fig. 6과 같이 RI가 0.54, RI[ave]가 0.51의 양의 상관관계를 보였으며 그 외에 외부 환경적 요소는 VPD에 큰 영향을 미치지 않았다. 반밀폐형 온실의 특성상 외부 환경의 영향을 크게 받지 않는 것으로 판단된다. 일사량의 경우 온실의 천창이 암막 형태가 아니므로 온실의 내부 온도(in_temp)와 유사한 정도의 영향을 미치는 것을 확인할 수 있다.



Fig. 5. Correlation between Internal Environmental Data and VPD



Fig. 6. Correlation between External Environmental Data and VPD

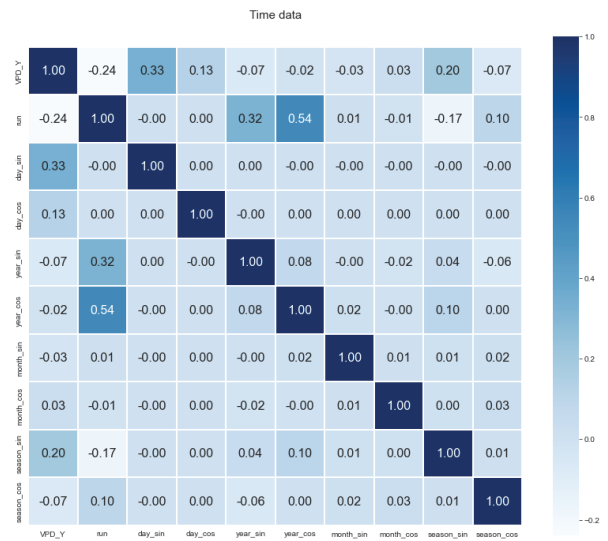


Fig. 7. Correlation between Time Data and VPD

VPD와 시간적 요소 사이의 상관관계를 확인해본 결과 Fig. 7과 같이 작기 운영여부(run)와 약한 음의 상관관계를 보였다. 작기 운영(1)일 때 VPD가 낮게 유지되고, 작기 미운영(0)일 때 온실 환경을 제어하지 않아 값이 높아지는 상황이 반영된 것으로 판단된다. 또한 작기 운영여부(run)와 year_sin, year_cos도 약한 음의 상관관계를 보였는데 작기의 1년의 주기성이 반영된 것으로 판단된다.

3.2 데이터 분석

상관관계 분석 결과 양의 상관관계를 보인 온실의 내부온도와 강한 음의 상관관계를 보인 내부 습도에 따라 VPD가 어떻게 변하는지 확인하기 위해 VPD_Y(온·습도 모두 제어), VPD_N_TEM(온도제어를 하지 않을 때, 외부온도와 같다고 가정), VPD_N_RH(습도 제어를 하지 않을 때, 외부습도와 같다고 가정), VPD_N(온·습도를 모두 제어하지 않을 때, 바깥의 온·습도와 동일하다고 가정)의 VPD 값을 비교해보았다.

Fig. 8과 같이 온실의 온도를 제어하지 않는 경우 VPD의 값이 낮아져 평균 0.17kPa의 최적의 VPD보다 0.5배 낮은 VPD 값을 가지게 된다. 온실의 습도를 제어하지 않는 경우에는 반대로 평균 0.81kPa의 최적의 VPD보다 2배 높은 값을 가지게 되며, 온도와 습도를 모두 제어하지 않는 경우 VPD_N_RH와 VPD_N_TEM의 사이인 평균 0.43kPa의 값이 나온다. 평균 0.43kPa인 경우 작물에 적절한 VPD 값이지만 VPD_N의 경우 외부의 환경 조건과 동일한 상태로 작물의 최적 온·습도를 벗어나므로, 최적의 VPD라고 볼 수 없다.

상관관계 분석 결과와 환경 제어를 통한 분석 결과를 통해 VPD는 온실 외부의 환경적 요소, 시간적 환경 요소보다는 내부의 환경적 요소에 따라 결정되며, 온도보다는 습도에 큰 영향을 받는 것을 확인할 수 있다.

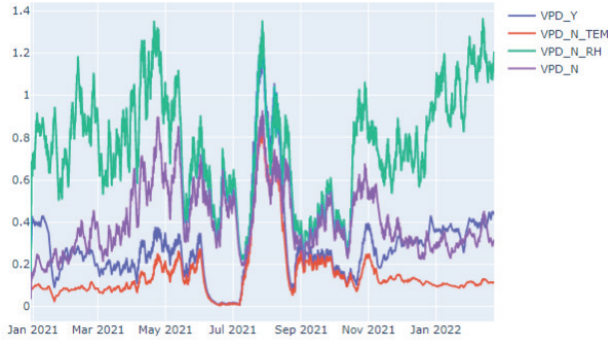


Fig. 8. VPD Data Changes According to Changes in Internal Temperature and Humidity

3.3 데이터 분석 방법론

선형회귀(LinearRegression)는 회귀 모델 중 가장 많이 사용되는 모델로 실제 값과 예측값의 차이(RSS, 오류의 제곱값)를 최소화하는 직선형 회귀선을 최적화하는 방식이다. 선형회귀에는 규제를 적용하지 않는 일반 선형 회귀, 상대적으로 큰 회귀 계수 값의 예측 영향도를 감소시키기 위해서 회귀 계수 값을 더 작게 만드는 방식의 Ridge, 예측 영향력이 작은 피처의 회귀 계수를 0으로 만들어 회귀 예측 시 피처가 선택되지 않도록 하는 Lasso 등이 있으며 그 식은 아래 Equation (6)-(8)과 같다.

$$J(\beta) = \frac{1}{n} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 \beta_{1i} - \dots - \beta_k \beta_{ki})^2 \quad (6)$$

$$J(\beta) = \frac{1}{n} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 \beta_{1i} - \dots - \beta_k \beta_{ki})^2 + \lambda \sum_{j=1}^k \beta_j^2 \quad (7)$$

$$J(\beta) = \frac{1}{n} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 \beta_{1i} - \dots - \beta_k \beta_{ki})^2 + \lambda \sum_{j=1}^k |\beta_j| \quad (8)$$

GradientBoost는 Gradient(또는 잔차(Residual))를 이용하여 이전 모형의 약점을 보완하는 새로운 모형을 순차적으로 적합한 뒤 이들을 선형 결합하여 얻어진 모형을 생성하는 지도 학습 알고리즘이다.

XGB(eXtra Gradient Boost)는 2016년 Chen과 Guestrin이 개발한 알고리즘으로이다[17]. 트리 기반의 앙상블 학습에서 가장 각광받고 있는 알고리즘 중 하나로 분류와 회귀 영역에서 예측 성능이 뛰어나고 자체 과적합 규제(Regularization) 기능을 가졌다는 장점이 있다. XGBRegressor의 하이퍼 파라미터로는 일반 파라미터, 부스터 파라미터, 학습 태스크 파라미터가 있다.

Light GBM(Light GBM boost)는 XGB 알고리즘의 속도를 보완하는 모델이다. XGB보다 가볍고 더 나은 성능을 보이며 메모리를 더 적게 사용하는 장점이 있고, 데이터 양이 적을 경우 과적합 될 수 있다는 단점이 있다. LGBM 모델은 알고리즘 구조 측면에서 수직적 성장(leaf-wise)을 적극적으로 활용하여[18], 비대칭적 트리를 형성하므로 균형을 잡기

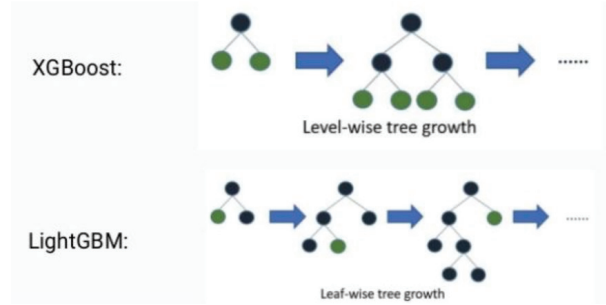


Fig. 9. Comparison of XGB and LGBM Models [19]

위한 연산을 하지 않는다. Fig. 9를 통해 균형 트리 분할 방식(level-wise)을 사용하는 XGB 모델과의 차이점을 확인할 수 있다.

랜덤포레스트(RandomForestRegressor)는 의사결정나무와 배깅 모델을 소개한 Breiman이 2001년 소개한 방법이다[20]. 여러개의 의사결정 트리를 조합해서 사용하는 방식인 앙상블 학습의 한 방법으로 Bagging(배깅)을 사용하는 대표적인 알고리즘이다. RandomForest는 빠른 속도와 이상치에 민감하지 않다는 장점과 메모리 사용이 많다는 단점을 가지고 있다. 많은 연구들이 랜덤포레스트를 이용하여 연구를 수행하고 있으며 다른 분류기법들과 비교하여 우수성을 보여주고 있다[21].

VPD 값 예측을 위해 MinMaxScaler를 사용하여 데이터들의 값을 0과 1사이의 범위값으로 변환하였다. 시계열 데이터므로 과거의 순서가 의미가 중요하기 때문에 데이터를 셔플하지 않고 test data와 train data로 분리하였다. 또한 과거의 데이터와 밀접한 관계를 갖는 시계열 데이터의 특성을 반영하기 위해 과거 VPD 데이터를 변수로 추가하여 환경적, 시간적 요소들과 함께 분석하였다. test 데이터는 전체 데이터의 약 25%, train data는 75%로 분리하였다.

4. 데이터의 예측

4.1 데이터의 양 변화

예측에 사용되는 데이터의 양 변화에 따른 예측 정확도를 확인해보았다. LinearRegression의 Ridge와 Lasso를 사용하여 데이터를 예측 전 1일(24시간), 3일(72시간), 7일(168시간)로 변화시켜 MAE와 RMSE를 확인하였다. MAE(Mean Absolute Error)와 RMSE(Root Mean Squared Error)는 회귀 모델의 성능을 평가하기 위한 지표로 0에 가까울수록 높은 정확도를 가진다.

데이터의 양이 많아질수록 Ridge의 정확도는 낮아졌고 Lasso의 정확도는 유사한 값을 보였다. 데이터 양의 증가와 예측 정확도 사이에 양의 상관관계가 성립되지 않는 것을 확인하였다. 예측 정확도가 동일할 때 데이터의 양이 증가할수록 예측에 많은 시간이 소요되므로 1일과 3일의 데이터 중 1일의 데이터를 사용했다.

Table 2. Accuracy Comparison based on Amount of Data

Sortation	1Day(24Hours)		3Day(72Hours)		7Day(168Hours)	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
Ridge	0.012	0.008	0.012	0.009	0.014	0.010
Lasso	0.020	0.014	0.020	0.014	0.019	0.014

Table 3. Accuracy Comparison According to Data Spacing

Sortation	1hour interval		20 minutes interval	
	RMSE	MAE	RMSE	MAE
Ridge	0.018	0.013	0.012	0.008
Lasso	0.022	0.017	0.020	0.014

4.2 데이터의 시간 간격 변화

데이터의 간격 변화에 정확도를 비교하기 위해 데이터 간격을 20분, 1시간 간격으로 변화시켜 MAE와 RMSE를 측정하였다.

MAE와 RMSE값을 구한 결과 데이터의 간격이 1시간일 때 보다 20분일 때 약 35% 더 높은 정확도를 보였다. 또한 Ridge보다 Lasso 모델을 사용했을 때 약 79% 더 높은 정확도를 보였다.

온실은 작물의 특성에 맞게 온도와 습도를 제어하고 있고, 외부 기상의 변화는 급격하게 이루어지지 않는 특징이 있다. 시간 간격이 짧을수록 VPD의 변화량도 적기 때문에 높은 예측 정확도 보이는 것으로 판단된다.

4.3 모델 변화

분석 결과를 반영하여 20분 간격의 1일의 데이터를 다양한 모델을 적용하여 최적의 예측값을 갖는 모델을 확인해보았다. 모든 회귀 모델은 파라미터를 적용하지 않고 n_estimators=100으로 통일하여 분석하였다. 정확도 분석 결과 LGBM을 사용했을 때 MAE는 0.011, RMSE는 0.008로 제일 높은 정확도를 보였다.

4.4 결과

온실의 특징으로 내부 환경적 요소들을 파프리카 재배에 적절하게 유지하고 있고, 반밀폐 온실로 외부 환경적 영향을 거의 받지 않는다. 또한 데이터의 예측 간격을 20분으로 설정하여 VPD 변동의 폭이 좁기 때문에 모델간 정확도에 큰 변화가 보이지는 않았다. 그 중 LGBM을 사용했을 때 MAE는 0.008, RMSE는 0.011로 제일 높은 정확도를 보였으며 생성한 예측데이터와 실제 데이터의 차이는 Fig 10의 그래프를 통해 확인할 수 있다.

feature_importances를 사용하여 예측 변수의 중요도를 확인해보았다. Table 5를 보면 VPD 값 예측에 가장 중요한 변수는 VPD_Y_71(20분 전의 VPD)이며 중요도는 약 95%임을 확인할 수 있다.

Table 4. Accuracy Comparison According to Model Change

Sortation	20 minutes interval	
	RMSE	MAE
Ridge	0.012	0.008
Lasso	0.020	0.014
GradientBoostingRegressor	0.012	0.008
XGBRegressor	0.012	0.008
LGBMRegressor	0.011	0.008
RandomForestRegressor	0.012	0.008

Real/prediction

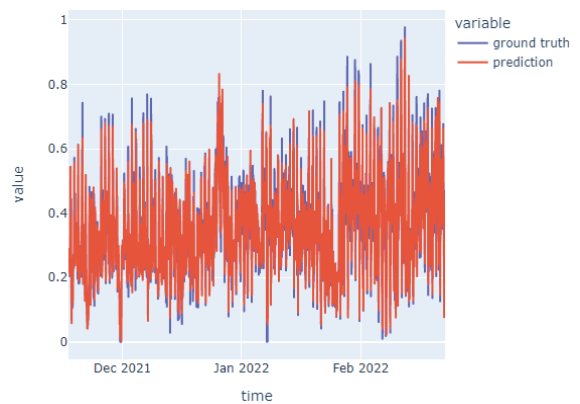


Fig. 10. Comparison of Predicted and Actual Values

Table 5. feature_importances

Colom	Important
YPD_Y_71	0.947741
in_RH_71	0.006444
RI[lave]_71	0.000755
VPD_Y_70	0.000488
in_temp_71	0.000486
VPD_Y_4	0.000417
VPD_Y_9	0.000396
VPD_Y_8	0.000353
VPD_11	0.000301
wind_d1_10	0.000291

5. 결 론

식물의 영양분 흡수에 따른 식물 성장뿐만 아니라 기공 기능 및 광합성에도 영향을 끼치는 VPD를 예측하기 위해 온실의 내·외부 환경적 요소, 시간적 요소들과의 상관관계를 분석해본 결과 온실 내부 환경적 요소로는 in_temp(0.56), in_RH(-0.85)와 상관관계를 보였으며 온실 외부 환경적 요소는 RI(0.54), RI[lave](0.51)와 상관관계를 보였다. 시간적 요소는 큰 상관관계를 보이지 않았으나 그 중 run(-0.24)과 약한 상관관계를 보였다.

다양한 조건(데이터의 양, 간격, 모델 등)을 예측 모델에 적용해본 결과 정확도는 데이터 양에 영향을 받지 않았다. 데이터의 간격에는 큰 영향을 받았는데, 온실의 VPD가 급격하게 변하지 않기 때문에 시간 간격이 짧을수록 정확도가 높아지는 것으로 해석된다. 마지막으로 온실 내·외부 환경적 요소, 시간적 요소, 과거 데이터를 변수로 하여 예측 정확도를 비교한 결과 LGBM을 사용했을 때 MAE는 0.008, RMSE는 0.011로 제일 높은 예측 성능을 보였다.

현재 많은 시설원에 농가에서는 스마트팜을 도입하고 있으며 빅데이터를 활용하여 농가 시스템의 의사결정에 도움을 받고자 하는 농가의 수요가 높아지고 있다[22]. 본 연구 결과를 활용하여 VPD 예측을 통해 작물의 생산성을 높이고, 온실의 결로, 병 발생 등 예방에 활용 가능하며 VPD뿐만 아니라 온실의 습도, CO₂농도 등의 예측에 사용 가능하다. 더 나아가 에너지 사용을 효율화 및 비용 절감을 통해 수익을 증대시킬 수 있는 다양한 예측 모델을 개발에 사용 가능하며, 다양한 스마트팜 시스템에 접목하여 활용이 가능할 것이라고 기대해본다. 또한 본 연구의 예측 모델을 추후 연구하고자 하는 스마트팜 통합 관리 플랫폼에 적용하여 VPD 데이터를 기반으로 한 온실 제어 시스템 구축에 활용 가능할 것이라고 생각한다.

References

- [1] S. J. Oh, "A design of intelligent information system for greenhouse cultivation," *The Society of Digital Policy and Management*, Vol.15, Iss.2, pp.183-190, 2017.
- [2] M. B. Son and D. Y. Han, "Assessment of feature agricultural land use and climate change impacts on irrigation water requirement considering greenhouse cultivation," *Journal of the Korean Association of Geographic Information Studies*, Vol.23, No.4, pp.120-139, 2020.
- [3] C. Amitrano, C. Arena, Y. Rouphael, S. Pascale, and V. De Micco, "Vapour pressure deficit: The hidden driver behind plant morphofunctional traits in controlled environments," *Annals of Applied Biology*, Vol.175, Iss.3, pp.313-325, 2019.
- [4] J. Ding, X. Jiao, P. Bai, Y. Hu, J. Zhang, and J. Li, "Effect of vapor pressure deficit on the photosynthesis, growth, and nutrient absorption of tomato seedlings," *Scientia Horticulturae*, Vol.293, pp.110736, 2022.
- [5] T. Inoue, M. Sunaga, M. Ito, Q. Yuchen, Y. Matsushima, K. Sakoda, and W. Yamori, "Minimizing VPD fluctuations maintains higher stomatal conductance and photosynthesis, resulting in improvement of plant growth in lettuce," *Frontiers in Plant Science*, Vol.12, 2021.
- [6] J. Ding, X. Jiao, P. Bai, Y. Hu, J. Zhang, and J. Li, "Effect of vapor pressure deficit on the photosynthesis, growth, and nutrient absorption of tomato seedlings," *Scientia Horticulturae*, Vol.293, 2022.
- [7] H. W. Lee, Y. S. Kim, S. Y. Shim, and J. W. Lee, "Variation of vapor pressure deficit and condensation flux of air heating plastic greenhouse installed with two layers thermal curtain in winter," *The Korean Society for Bio-Environment Control*, Vol.22, No.1, pp.35-41, 2013.
- [8] H. M. Noh and J. H. Lee, "The effect of vapor pressure deficit regulation on the growth of tomato plants grown in different planting environments," *Applied Sciences*, Vol.12, No.7, pp.3667, 2022.
- [9] C. Grossiord et al., "Plant responses to rising vapor pressure deficit," *The New Phytologist*, Vol.226, No.6, pp.1550-1566, 2020.
- [10] Heuvelink Ep, "Tomatoes," Cambridge MA: CABI Pub, 2005.
- [11] H. C. Kim, S. G. Jung, J. H. Lee, and H. J. Bae, "Effects of greenhouse covering material on environment factors and fruit yield in protected cultivation of sweet pepper," *Journal of Bio-Environment Control*, Vol.18, No.3, pp.253-257, 2009.
- [12] Perfect Grower Vapor Pressure Deficit Recommendations (kPa), Google Chrome [Internet], <https://www.perfectgrower.com/knowledge/knowledge-base/vpd-chart-vapor-pressure-deficit/>
- [13] S. Redmond, C. M. Hasfalina, Z. Abdjamil, V. B. Peter, A. Desa, W. Ismail, and W. Ishak, "Membership function model for defining optimality of vapor pressure deficit in closed-field cultivation of tomato," *Acta Horticulturae*, 10.17660/ActaHortic.2017.1152.38, 2016.
- [14] benro-type vinyl greenhouse, Google Chrome [Internet], <http://www.jjn.co.kr/news/articleView.html?idxno=788176>
- [15] J. H. kim, C. H. Lee, and K. S. Shim, "Time series prediction using clustering algorithm," *KIISE Transactions on Computing Practices*, Vol.20, No.3, 191-195, 2014.
- [16] J. H. Moon, S. W. Park, S. M. Rho, and E. J. Hwang, "A comparative analysis of artificial neural network architectures for building energy consumption forecasting," *International Journal of Distributed Sensor Networks*, Vol.15, No.9, 2019.
- [17] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," *arXiv:1603.02754v3 [cs.LG]*, 2016.
- [18] S. Cheon, J. Yu, J. G. Kim, J. S. Oh, T.-H. Nam, and T. Lee, "Predicting deformation behavior of additively manufactured Ti-6Al-4V based on XGB and LGBM," *Transactions of Materials Processing*, Vol.31, No.4, 2022.
- [19] XGBoost & LightGBM, slidshare [Internet], <https://www.sli deshare.net/GabrielCyprianoSaca/xgboost-lightgbm>
- [20] L. Breiman, "Random Forests," *Machine Learning*, Vol.45, pp.5-32, 2001.

- [21] J. E. Ha, H. C. Shin, and J. G. Lee, "Korean text classification using randomforest and XGBoost focusing on Seoul metropolitan civil complaint data," *The Journal of Bigdata*, Vol.2, No.2, pp.95-104, 2017.
- [22] S. W. Rye and D. Y. Park, "Economic analysis based on smart farm colling technology," *KIEAE Journal*, Vol.21, No.5, pp.55-65, 2021.



오 한 별

<https://orcid.org/0000-0003-3146-9090>
 e-mail : ohb8095@scnu.ac.kr
 2017년 순천대학교 화학공학전공(학사)
 2022년 순천대학교 정보통신공학전공(석사)
 2022년~현 재 순천대학교
 정보통신공학전공 박사과정

관심분야: 스마트팜, 농업 IoT, 빅데이터, 데이터분석



장 경 민

<https://orcid.org/0000-0001-5774-6303>
 e-mail : jangkm0625@scnu.ac.kr
 2013년 순천대학교 정보통신공학전공
 (학사)
 2021년~현 재 순천대학교
 정보통신공학전공 석사과정

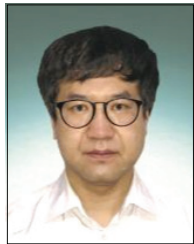
관심분야: 기계학습, 농업 IoT, 빅데이터



신 창 선

<https://orcid.org/0000-0002-5494-4395>
 e-mail : css@scnu.ac.kr
 1996년 우석대학교 전산학과(학사)
 1999년 한양대학교 컴퓨터교육과(석사)
 2004년 원광대학교 컴퓨터교육과(박사)
 2005년~현 재 순천대학교
 인공지능공학부 교수

관심분야: 분산컴퓨팅, 실시간 객체모델, 농업 IoT



이 명 배

<https://orcid.org/0000-0002-7160-2637>
 e-mail : lmb@scnu.ac.kr
 2010년 순천대학교 컴퓨터공학과(학사)
 2012년 순천대학교 정보통신공학전공(석사)
 2014년 순천대학교 정보통신공학전공
 (박사수료)

2021년~현 재 순천대학교 지능기술연구소 부소장
관심분야: 첨단기술농업, 농업 IoT, 클라우드, 기계학습



박 장 우

<https://orcid.org/0000-0001-8201-8949>
 e-mail : jwpark@scnu.ac.kr
 1987년 한양대학교 전자공학과(학사)
 1989년 한양대학교 전자공학과(석사)
 1993년 한양대학교 전자공학과(박사)
 1995년~현 재 순천대학교
 인공지능공학부 교수

관심분야: SoC, 기계학습, RFID/USN



임 종 현

<https://orcid.org/0000-0001-6832-4077>
 e-mail : sshb56@s.scnu.ac.kr
 2019년 순천대학교 정보통신공학전공(학사)
 2021년 순천대학교 정보통신공학전공(석사)
 2021년~현 재 순천대학교
 정보통신공학전공 박사과정

관심분야: 소프트웨어 시스템, 유비쿼터스, 빅데이터, 스마트팜