

## 파프리카 환경 데이터를 활용한 AI 기반 생육 조건 예측 모델 연구

정광훈, 여현\*

\*국립순천대학교

gwanghoon5035@gmail.com, \*yhyun@scnu.ac.kr

## A Study on AI-Based Growth Condition Prediction Model Using Paprika Environmental Data

Jung Gwang Hoon, Yoe Hyun\*

\*Sunchon National Univ.

## 요 약

본 연구는 노지에서 재배되는 파프리카의 생육 환경 예측을 위한 인공지능(AI) 기반 모델을 개발하고, 그 성능을 비교 분석하는 데 목적이 있다. 2024년 한 해 동안 5분 간격으로 수집된 파프리카 노지 환경 센서 데이터를 활용하여, 기온(°C), 습도(%), 풍속(m/s), 풍향(°), 일사량(W/m²) 등의 변수를 입력으로 하고 30분 후 기온을 예측하는 머신러닝 모델을 구축하였다. 예측 모델로는 선형 회귀, 랜덤 포레스트, 서포트 벡터 머신(SVM) 회귀를 적용하였으며, 성능 평가는 평균제곱근오차(RMSE), 평균절대오차(MAE), 결정계수(R²)를 기준으로 수행하였다. 실험 결과, 랜덤 포레스트 모델이 RMSE 0.6°C, R² 0.99로 가장 우수한 예측 성능을 보였고, SVM은 이보다 약간 낮은 성능을 나타냈다. 반면, 선형 회귀는 단순 구조로 인해 비선형 환경 변화를 충분히 반영하지 못해 예측 정확도가 낮았다. 본 연구는 노지 환경에서도 머신러닝을 통한 단기 기온 예측이 효과적으로 수행될 수 있음을 입증하였으며, 향후 생육 적합성 예측, 중기 기상 예측, 딥러닝 모델 적용 등으로의 확장이 기대된다.

## I. 서론

최근 농업 분야에서는 ICT 기술을 접목한 스마트 농업 시스템이 확산되고 있으며, 다양한 환경 데이터를 기반으로 한 작물 생육 예측 연구도 활발히 이루어지고 있다[1].

특히 기계학습(Machine Learning)이나 딥러닝(Deep Learning) 알고리즘을 이용한 생육 조건 예측 모델이 제안되고 있지만, 대부분의 연구는 토마토, 상추 등 온실 재배 작물을 중심으로 진행되어 왔다[2].

반면, 노지 환경에서 수집된 실제 데이터를 활용한 생육 조건 예측 연구는 상대적으로 부족하며, 특히 파프리카를 대상으로 한 연구는 거의 전무한 실정이다.

최근 농업 분야에서는 ICT 기술을 접목한 스마트 농업 시스템이 확산되고 있으며, 다양한 환경 데이터를 기반으로 한 작물 생육 예측 연구도 활발히 이루어지고 있다[3].

특히 기계학습(Machine Learning)이나 딥러닝(Deep Learning) 알고리즘을 이용한 생육 조건 예측 모델이 제안되고 있지만, 대부분의 연구는 토마토, 상추 등 온실 재배 작물을 중심으로 진행되어 왔다[4].

반면, 노지 환경에서 수집된 실제 데이터를 활용한 생육 조건 예측 연구는 상대적으로 부족하며, 특히 파프리카를 대상으로 한 연구는 거의 전무한 실정이다[5]. 본 연구는 노지에서 수집된 파프리카 재배 환경 데이터를 기반으로, AI 기법을 활용한 생육 조건 예측 모델을 설계하는 것을 목적으로 한다. 이를 위해 실제 노지에서 5분 간격으로 수집된 온도, 습도, 일사량, 풍속 등의 환경 데이터를 활용하여, 기계학습 기반 예측 모델을 구축하고, 파프리카 생육에 적합한 조건을 도출하고자 한다.

본 연구는 노지 환경의 불확실성과 변동성을 고려한 데이터 기반 예측 모델을 제시함으로써, 파프리카 노지 재배의 생산성과 품질 향상에 기여할 수 있는 기반을 마련하고자 한다.

## II. 본론

본 연구에는 2024년 한 해 동안 국내 한 파프리카 노지 재배지에 설치된 환경 센서로부터 수집된 데이터가 활용되었다.

센서는 5분 간격으로 대기 정보를 기록하였으며, 주요 수집 변수는 기온(°C), 상대습도(%), 풍향(°에서의 방향각), 풍속(m/s), 일사량(W/m²)의 5가지이다.

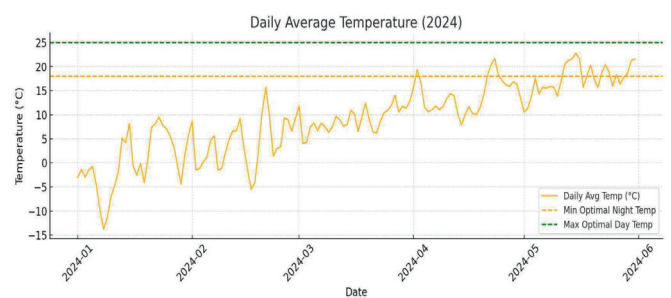


그림 1. 파프리카 일별 평균 온도

fig 1. Paprika Daily Average Temperature

<그림 1>은 수집된 파프리카 일별 평균 기온을 나타낸 그래프이다. 수집된 원시 데이터에 대해 이상치 제거 및 결측 보정 등의 전처리를 수행하여 분석에 적합한 상태로 정제하였다.

특히 일사량의 경우 야간 시간대에는 0에 수렴하는 값만 나타나며 주간에만 유의한 변화를 보이는데, 이러한 일주기 변동 특성이 온도 변화에 미치는 영향을 모델이 학습할 수 있도록 다른 기상 변수와 함께 입력 변수로 사용하였다.

단기 온도 예측을 위한 학습 모델을 구성하기 위해, 각 시점의 센서 관측 값을 입력으로 하고 30분 후의 기온을 출력으로 하는 데이터셋을 구축하였다. 전체 데이터 중 80% 가량을 훈련용으로 사용하고 나머지 20%를 테스트용으로 하여 모델의 예측 성능을 평가하였다.

머신러닝 기법으로는 (1) 선형 회귀(Linear Regression), (2) 랜덤 포레스트(Random Forest) 회귀, (3) 서포트 벡터 머신(Support Vector Machine, 회귀용 SVR)을 선정하였다.

선형 회귀는 입력 변수와 출력 사이의 선형 관계를 가정한 가장 단순한 모델로서 기준선 역할을 한다. 랜덤 포레스트는 다수의 의사결정나무 모델을 앙상블하여 비선형 패턴까지 포착하는 강력한 방법으로, 본 연구에서는 100개의 트리를 갖는 RF 모델을 구축하였다. SVM 회귀모델은 RBF 커널을 사용하여 입력-출력 간의 비선형 관계를 학습하도록 하였으며, 기타 하이퍼파라미터는 기본값을 적용하였다.

세 모델 모두 훈련 데이터에 대해 학습을 완료한 후, 미사용된 테스트 데이터에 대한 예측을 수행하였다.

모델 성능 비교를 위해 평균제곱근오차(RMSE), 평균절대오차(MAE), 결정계수( $R^2$ )를 산출하였다. RMSE는 예측값과 실제값의 차이의 제곱 평균에 대한 제곱근으로 오류의 크기를 나타내며, MAE는 오류 절대값의 평균으로 직관적인 평균 오차 크기를 의미한다.  $R^2$ 는 모델이 실제 변동성을 얼마나 설명하는지를 0~1 범위에서 표시하는 지표로, 1에 가까울수록 설명력이 높음을 의미한다. RMSE와 MAE는 낮을수록,  $R^2$ 는 높을수록 예측 정확도가 우수한 모델로 판단된다.

표 1. 모델별 온도 예측 비교

Table 1. Comparison of Temperature Prediction by Model

모델	RMSE	MAE	$R^2$
선형 회귀	0.9	0.6	0.97
랜덤 포레스트	0.6	0.4	0.99
서포트 벡터 머신	0.7	0.5	0.98

<표 1>은 모델별 성능 지표를 정량적으로 비교한 것이다. 선형 회귀의 경우 RMSE가 약 0.9°C 수준으로 가장 높았으며, MAE 역시 0.6°C 내외로 나타나 세 모델 중 정확도가 가장 낮았다. 반면 랜덤 포레스트 모델은 RMSE 0.6°C, MAE 0.4°C 수준으로 오류를 크게 줄였고 결정계수  $R^2$ 도 0.99에 달해 가장 우수한 성능을 기록하였다. SVM 모델의 성능은 RF와 선형 회귀의 중간 정도로, RMSE 약 0.7°C, MAE 0.5°C,  $R^2$  약 0.98을 보였다. 세 모델 모두  $R^2$  값이 0.95 이상으로 비교적 높아 단기(30분) 온도 변동의 대부분을 설명하고 있었지만, 미세한 오차 측면에서는 비선형 머신러닝 기법(RF, SVM)이 선형 모델에 비해 예측 정밀도가 향상됨을 확인할 수 있다.

랜덤 포레스트 모델의 예측값이 실제 관측된 온도와 가장 근접하였으며, SVM도 비교적 정확하게 예측 온도를 맞추는 경향을 보였다. 반면 선형 회귀의 예측값은 급격한 기온 상승/하강 구간에서 다소 편차를 나타냈는데, 이는 일사량 변화 등에 따른 비선형 응답을 선형 모델이 충분히 따라가지 못한 결과로 보인다. 예를 들어 일출 후 햇빛이 강해지면서 기온이 급격히 상승하는 구간에서, 선형 회귀는 상승 폭을 과소예측하는 경향이 있었으나 랜덤 포레스트와 SVM은 이러한 변화를 보다 잘 포착하여 실제 값에 근접한 예측선을 그렸다. 이처럼 비선형 모델들은 주변 환경 변수와 온도 간의 복잡한 상관관계를 학습함으로써 단기간의 기온 변화를 정밀하게 예측하는 데 강점을 나타내었다. 본 연구 결과는 기온 예측에 있어 머신러닝 기법의 효용을 확인한 것이며, 특히 랜덤 포레스트와 같은 앙상블

기법이 뛰어난 성능을 보였다.

### III. 결론

본 연구에서는 파프리카 노지 재배지의 센서 데이터를 활용하여 30분 후의 기온을 예측하는 머신러닝 모델을 개발하고, 선형 회귀, 랜덤 포레스트, SVM 세 기법의 성능을 비교하였다. 실험 결과 랜덤 포레스트 모델이 가장 높은 예측 정확도를 보였으며, SVM 역시 선형 회귀에 비해 낮은 오차로 우수한 성능을 나타냈다. 이는 노지 환경에서 기온 변화의 비선형성을 포착하기 위해서는 복잡한 모델 구조가 효과적임을 시사한다. 반면, 가장 단순한 선형 회귀 모델도 일정 수준의 정확도를 보여주었는데( $R^2 \approx 0.97$ ), 이는 예측에서는 현재 기온 자체가 강한 예측 인자로 작용하기 때문인 것으로 판단된다. 그럼에도 불구하고 수 °C 이내의 오차 감소가 작물 생육 환경 관리에는 큰 의미를 가질 수 있다. 또한 본 연구에서 제시한 방법론은 노지 재배 뿐 아니라 시설원예, 축산 등 다양한 농업 분야의 미기상 예측에도 적용 가능하며, 향후 연구에서는 예측 리드 타임을 1시간 이상으로 늘려 중기 예측에 도전하거나, 딥러닝(LSTM 등) 모델을 포함한 추가적인 기법들과의 성능 비교, 그리고 기상 예보 데이터와의 융합을 통한 예측 정확도 향상 등에 대해서도 지속적인 연구가 필요하다.

### ACKNOWLEDGMENT

“본 연구는 과학기술정보통신부 및 정보통신기획평가원의 지역지능화 혁신인재양성(Grand ICT연구센터) 사업의 연구결과로 수행되었음”(IITP-2025-2020-0-01489)

### 참 고 문 헌

- [1] Aekyung Moon, Eunryung Lee, and Seunghan Kim, "Development of Microclimate-based Smart farm Predictive Platform for Intelligent Agricultural Services," Journal of the Korea Industrial Information Systems Research, vol. 26, no. 1, pp. 21-29, 2021.
- [2] Hoon-Seok Jang and Tai-Gil Kwon, "GRU-based temperature prediction model using environmental data from a paprika greenhouse," 대한전자공학회 학술대회, pp. 2518-2520, 제주, 2024-06-26.
- [3] Chang Il-Sik, Yang Ji Hee, Kang Hyun Seok, Jang Jun Young, and Park Goo Man, "A study on predicting paprika production using environmental information and growth information in smart farms," Proceedings of KIIT Conference, pp. 11-11, 제주, 2023-06-01.
- [4] Chang Il-Sik, Yang Ji Hee, Kang Hyun Seok, Jang Jun Young, and Park Goo Man, "A study on predicting paprika production using environmental information and growth information in smart farms," Proceedings of KIIT Conference, pp. 11-11, 제주, 2023-06-01.
- [5] 황인철, 노희선, 양동일, and 김만배, "다중 회귀 분석을 이용한 파프리카 생산량 예측," 한국통신학회논문지, vol. 46, no. 11, pp. 2048-2055, 2021. (10.7840/kics.2021.46.11.2048)