스마트팜 환경 데이터와 생육지표 간 상관관계 분석 및 머신러닝 기반 과중 예측

곽찬영, 박근호, 장훈석, 최주환

한국전자기술연구원

gcy0222@keti.re.kr, root@keti.re.kr, jhs0053@keti.re.kr, netside@keti.re.kr

Correlation Analysis of Smart Farm Environmental Data and Growth Indicators with Machine Learning-Based Fruit Weight Prediction

Chanyoung Gwak, Keunho Park, Hoon-Seok Jang, Juhwan Choi

Korea Electronics Technology Institute

요 약

본 논문은 전남 나주시 스마트팜 멜론 하우스에서 수집한 환경데이터와 멜론 생육지표 간의 상관관계를 분석하고, 머신러닝 기반 과중 예측 가능성을 검토하였다. 내부 온도, 내부 습도, 누적 일사와 같은 환경요인과 멜론 과중(수동 측정)을 중심으로 데이터를 전처리하고, 상관분석 및 선형회귀·랜덤 포레스트 모델을 비교하였다. 상관분석 결과, 과중(수동)은 내부습도, 과중(자동)과 양(+)의 상관($r \approx 0.69$), 누적 일사와 음(-)의 상관($r \approx -0.63$)을 보였다. 예측 모델에서는 랜덤 포레스트가 R^2=0.76, RMSE=195g으로 선형회귀(R^2=0.65, RMSE=238g)보다 우수한 성능을 나타냈다. 본 결과는 자동계측 데이터와 환경데이터의 상관관계를 보여주고, 머신러닝 모델로도 멜론 과중 예측이 가능함을 시사한다.

Ⅰ. 서론

스마트팜 기술의 확산과 함께, 환경데이터를 활용하여 작물의 생육과 수확량을 예측하려는 연구가 활발히 진행되고 있다. 특히 온도, 습도, 일사량 등 환경요인이 수확량에 미치는 영향은 다양한 작물에서 보고되어왔다[1-3]. 그러나 기존 연구는 기상·환경 데이터 중심의 분석에 치중되어 있으며, 실제 생육지표(예:과중)와의 상관관계를 정량적으로 규명하고 이를 예측 모델과 연결한 사례는 제한적이다.

본 연구는 두 가지로 구성된다. 첫째, 멜론 재배 과정에서 수집한 환경데이터와 생육지표 간의 상관관계를 분석하였다. 둘째, 머신러닝 기법을 적용하여 과중 예측 성능을 비교·평가하였다.

Ⅱ. 본론

2. 데이터 및 방법

본 연구에서는 2024년 5월부터 6월까지 전남 나주 스마트팜 멜론 하우스에서 수집된 멜론 생육 및 환경데이터를 활용하였다. 생육 데이터는 과중(수동·자동), 엽장 등을 포함하며, 환경데이터는 내부온도(2개 지점), 내부습도(2개지점), 누적 일사(외부)로 구성되었다. 내부온도와 습도는 각각 평균값으로 변환하여 분석에 사용하였다. 타깃 변수는 과중(수동)으로 설정하고, 입력 변수는 내부온도_avg, 내부습도_avg, 누적 일사, 과중(자동)으로 구성하였다.

멜론 과중(자동)은 수집된 RGB 이미지를 입력으로 하여 과실 영역을 탐지하고, 과실 크기 및 형태 정보를 활용하여 비전 기반 과중 추정치를 산출하였다. 추정된 자동 과중 데이터는 실제 계측 값과 상관분석을 수행하여 모델의 정확도를 평가하였다.

2.1 데이터 처리, 분석 및 시각화

수집된 기상 및 스마트팜 환경 센서 데이터는 Pandas 라이브러리를 활용하여 원시데이터를 불러오고, 전처리 과정을 수행하였다. 먼저 결측값(missing value)은 NaN으로 처리한 뒤 주요 변수에서 결측이 발생한 경우 행 단위로 제거하였다. 이상값(outlier)에 대해서는 사분위수 범위(IQR)와 같은 통계적 기준을 적용하여 제거하였다. 그리하여 원본데이터 891개에서 결측값 및 이상 값 제거 후 262개로 정제되었다.

환경데이터와 생육지표 간의 관계를 파악하기 위해 상관분석을 실시하였다. 상관계수를 계산하여 변수 간 관계의 정도를 정량적으로 파악하였다. 그 결과는 그림 1과 같이 상관행렬(correlation matrix)으로 시각화하였다.

또한, 과중(수동 측정값)을 예측하기 위한 회귀 기반 모델링을 수행하였다. 변수 간의 인과관계를 탐색하기 위해 선형회귀 분석을 수행하였으며, 회귀식과 추세선, 결정계수(R^2)를 도출하여 설명력을 평가하였다. 더불어, 비선형 관계를 반영하기 위해 랜덤포레스트(Random Forest) 모델을 추가적으로 적용하였다. 랜덤포레스트는 여러개의 의사결정나무를 무작위로 학습시킨 후 , 예측값의 평균을 구하는 방식으로 최종 결과를 도출하는 앙상블 기법이다. 이는 변수 간 상호작용과 비선형성을 효과적으로 포착할 수 있다는 장점이 있다.

이와 같은 데이터 처리 및 분석 절차를 통해 환경 요인의 기본 특성을 파악하였고, 이를 바탕으로 생육지표와의 상관 분석 및 예측 모델링에 활용하였다.

2.2 상관분석 결과

상관계수 분석 결과는 그림 1과 같이 과중(수동)은 내부습도_avg(r=0.69)와 과중(자동)(r=0.79) 강한 양의 상관, 외부 누적일사와 강한

음의 상관을 나타냈다. 내부온도_avg와는 중간 수준의 상관(r=0.34)이 확인되었으며, 엽장은 과중과 유의한 상관관계가 나타나지 않았다.

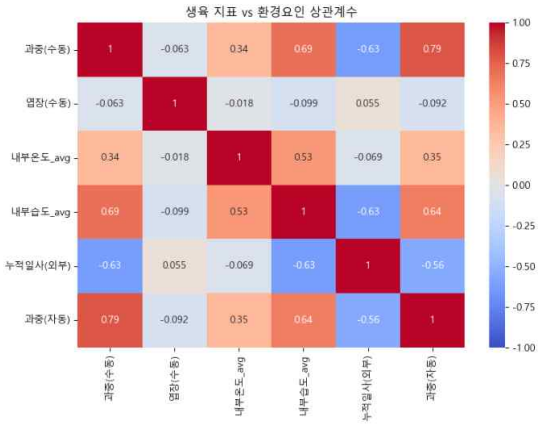

그림 1. 생육지표 vs 환경요인 상관계수

이는 자동 과중 계측이 실측값을 비교적 잘 반영하고 있고, 멜론의 과중 형성에 있어 습도와 온도가 긍정적 요인으로, 일사량은 부정적 요인으로 작용할 가능성을 보여준다.

3. 예측모델 성능 비교

머신러닝 기반 예측 모델의 성능 비교 결과, 표1과 같이 선형회귀 모델은 R^2=0.65, RMSE=238g을 보였으며, 랜덤포레스트 모델은 R^2=0.76, RMSE=195g으로 더 우수한 성능을 나타냈다. 이는 멜론 과중과 환경 요인 간 관계가 단순 선형이 아닌 비선형성을 포함하고 있음을 의미하며, 랜덤포레스트 모델이 이를 보다 효과적으로 포착했음을 보여준다.

표 1. 예측모델 성능비교

모델	R^2	RMSE (g)
선형회귀	0.649	237.7
랜덤포레스트	0.764	194.8

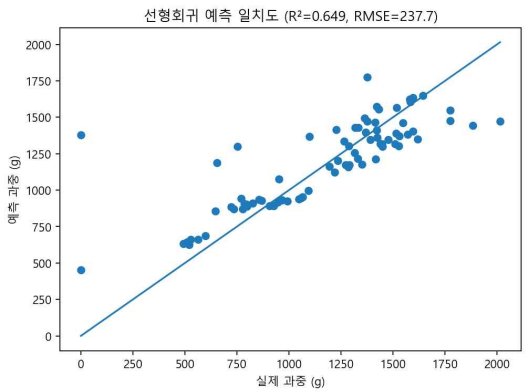

그림 2. 선형회귀 예측 일치도

3.1 변수 중요성 분석

랜덤포레스트의 변수 중요도 분석 결과, 과중(자동)이 74%로 가장 큰 기여도를 보였으며, 내부습도_avg(18%), 누적일사(7%), 내부온도_avg(1%) 순으로 나타났다. 이는 자동계측 데이터가 수동 계측값을 예측하는 데 있어 중요한 보조 변수임을 확인한 결과로, 스마트팜 환경에서 자동계측 센서 활용의 가능성을 뒷받침한다.

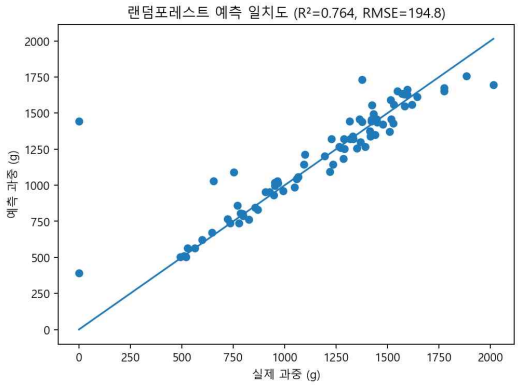

그림 3. 랜덤포레스트 예측 일치도

Ⅲ. 결론

본 연구는 스마트팜 환경에서 수집된 환경데이터와 멜론 생육지표 간의 관계를 규명하고, 머신러닝 기반 예측 모델의 성능을 평가하였다. 연구를 위해 수집된 내부온도, 내부습도, 누적일사, 자동·수동 과중 데이터를 전처리하여 상관분석을 수행하였으며, 멜론 과중 예측으로 선형회귀와 랜덤포레스트 모델을 비교하였다.

분석 결과, 멜론 과중(수동)은 내부 습도(r=0.69), 과중(자동)(r=0.69)과 양(+)의 상관, 누적 일사와 음(-)의 상관(r=-0.63)을 보여 습도와 일사가 주요 환경요인으로 작용함을 확인하였다. 예측 모델 성능 비교에서는 랜덤포레스트가 선형회귀보다 높은 설명력(R^2=0.76)과 낮은 예측 오차(RMSE=195g)를 확인하였다.

결론적으로, 수집한 환경데이터와 생육지표(과중)를 대상으로 상관분석을 통해 온도·습도·일사량과 같은 환경요인이 과중과 유의미한 관계를 가짐을 확인하였다. 멜론 과중 예측 머신러닝 모델들의 성능 비교하여, 랜덤포레스트모델이 상대적으로 우수한 예측 정확도를 보였다.

이러한 결과는 스마트팜 환경제어 및 과중 예측 서비스 개발에 기초 자료로 활용될 수 있다. 다만 본 연구는 단기간·소규모 데이터에 기반하므로 일반화에 제약이 있으며. 향후에는 장기적·다지역 데이터와 심층학습 기반 모델을 적용하여 예측 정확도와 실용성을 높이는 연구가 필요하다.

ACKNOWLEDGMENT

이 논문은 과학기술정보통신부의 재원으로 정보통신산업진흥원(NIPA)에서 지원한 AI융합 지능형 농업 생태계 구축 사업으로 수행된 연구임 (S0103-24-1001).

참 고 문 헌

[1] 박철우, 김재현, “환경데이터 기반의 작물 생육량 예측에 관한 연구,” 한국농업정보학회지, 2020.

[2] Chlingaryan, A., Sukkarieh, S., & Whelan, B., “Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review,” Computers and Electronics in Agriculture, vol. 151, pp. 61 - 69, 2018.

[3] Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D., “Machine learning in agriculture: A review,” Sensors, vol. 18, no. 8, pp. 2674, 2018.