

온실 환경 변화에 따른 작물 수분 스트레스 예측 알고리즘 분석 연구

김현우, 이명훈*

국립순천대학교, *국립순천대학교

kimhw51793@naver.com, *leemh777@scnu.ac.kr

Analysis of Crop Water Stress Prediction Algorithms in Response to Greenhouse Environmental Changes

Kim Hyeonwoo, MeongHun Lee*

Sunchon National Univ., *Sunchon National Univ.

요약

최근 기후변화로 인한 극단적 기상 현상 증가는 시설원에 환경에서도 작물 수분 관리의 어려움을 가중시키고 있다. 기존의 관수 방식은 환경 변화에 따른 작물의 실제 수분 상태를 충분히 반영하지 못해 수분 스트레스 발생 문제가 지속적으로 지적되고 있는 실정이다. 본 연구에서는 온실 환경에서 수집된 복합 센서 기반 시계열 데이터를 활용하여 작물 수분 스트레스 예측 알고리즘을 비교·분석하였다. 분류 문제에는 Random Forest, XGBoost, LSTM을, 회귀 문제에는 LightGBM과 DNN을 적용하였다. 실험 결과, LSTM 모델은 가장 높은 분류 성능을 보였으며, 트리 기반 모델은 낮은 연산 비용으로 안정적인 성능을 나타냈다. 본 연구는 예측 정확도와 실시간 적용성을 함께 고려한 작물 수분 스트레스 예측 알고리즘 선택 기준을 제시한다.

I. 서론

최근 기후변화의 여파로 이상고온 및 국지성 폭우 등 극단적 기상 현상이 빈번하게 발생하고 있으며, 이는 농업 생산 전반에 걸쳐 다양한 피해를 유발하고 있다.[1] 이러한 현상은 노지뿐만 아니라 시설원에 환경에서도 피해 사례가 나타나며, 외부 기상 변화에 따른 온실 내부 환경 변동성 증가로 인해 작물 생육 관리가 더욱 어려워지고 있는 실정이다.[2]

이로 인해 기존의 경험 기반 또는 고정된 기준에 의존한 관수 방식은 환경 변화에 적절히 대응하지 못하는 한계를 드러내고 있으며,[3] 그 결과 비효율적인 관수로 인한 물 자원 낭비와 작물 수분 스트레스 발생 문제가 지속적으로 지적되고 있다. 작물 수분 스트레스는 토양 수분 부족 또는 수분 흡수 불균형으로 인해 작물의 생리적 기능이 저하되는 상태로, 생육 지연 및 수량 감소로 이어질 수 있다.[4]

기존의 관수 시스템은 토양 수분 값이나 고정된 스케줄에 의존하는 경우가 많아, 환경 조건 변화에 따른 작물의 실제 수분 상태를 충분히 반영하지 못하는 한계를 지닌다.[5]

이에 본 연구에서는 열화상 영상과 온·습도, 토양 수분, 광량, 이산화탄소 농도 등 다양한 온실 환경 요소와 복합 센서 기반 시계열 데이터를 활용하여 작물 수분 스트레스를 예측하는 알고리즘을 분석하고자 한다. 머신러닝 및 딥러닝 기반의 다양한 예측 모델을 비교·평가하고, 예측 성능과 더불어 연산 효율성과 실시간 적용 가능성을 함께 고려함으로써 실제 스마트팜 환경에 적합한 수분 스트레스 예측 모델을 제시하는 것을 본 연구의 목표로 한다.

본 논문의 순서는 다음과 같다. 2장에서 실험 환경 및 연구 방법에 대해 설명하고, 도출된 분석 결과를 설명한다. 마지막으로 3장에서 결론을 서술한다.

II. 본론

본 장에서는 온실 환경에서 수집된 복합 센서 기반 데이터셋을 구성하고, 이를 바탕으로 작물 수분 스트레스 예측을 위한 다양한 알고리즘을 적용하여 성능을 분석하였다.

II.1 실험 환경 구성

본 연구에서는 실제 온실 환경을 반영한 테스트를 진행하기 위해 AI Hub에서 제공하는 농업 환경 데이터셋을 활용하였다. 해당 데이터는 온실 재배 환경에서 수집된 실측 데이터로, 작물 생육에 영향을 미치는 다양한 환경 변수를 포함한다. 데이터는 온실 내부에 설치된 복합 센서를 통해 일정 주기로 수집되었으며, 온도, 습도, 토양 수분, 광량, 이산화탄소 농도 등 주요 환경 변수를 포함한다. 수집된 데이터는 시계열 형태로 두달 간 수집하였으며, 구성은 다음 표와 같다.

측정 항목	단위	설명	수집 주기
온도	°C	온실 내부 온도	5분
습도	%	온실 내부 습도	
토양 수분 함량	%	작물 근권부 토양 수분	
광량	lux	생육 유효 광량	
CO ₂ 농도	ppm	광합성 환경 변수	
엽온	°C	작물 잎 표면 온도	
측정 시각	Timestamp	시계열 데이터 정렬 기준	

표 1 데이터 수집 조건 구성

수집된 데이터는 일부 결측치와 이상치를 포함하고 있어, 분석에 앞서 전처리 과정을 수행하였다. 전처리 과정에서는 결측 데이터 제거 및 보정, 이상치 필터링, 변수 간 스케일 차이를 고려한 정규화를 적용하였다. 또한 예측 모델 학습을 위해 데이터를 시계열 분석에 적합한 형태로 구성하였다.

단계	전처리 항목	설명
1	결측치 처리	결측 데이터 제거
2	이상치 제거	물리적 비현실적 값 및 급격한 변동 데이터 제거
3	시간 동기화	수집 주기 5분 단위 리샘플링
4	보간 처리	결측 구간 보완
5	정규화	각 변수 값 0-1 범위 변환
6	시계열 구성	연속된 시간 구간 기준 시계열 데이터셋 구성
7	라벨링	수분 스트레스 상태 Normal, Mild, Severe 단계 구분

표 2 데이터 전처리

II.2 예측 알고리즘 선정 및 학습

작물 수분 스트레스 예측을 위해 머신러닝 및 딥러닝 기반의 다양한 알고리즘을 적용하였다. 예측 문제는 수분 스트레스 상태를 단계별로 판단하는 분류 문제와 연속적인 수분 스트레스 지표를 예측하는 회귀 문제로 구분하였다. 분류 모델로는 Random Forest, XGBoost, LSTM을 적용하였으며, 회귀 분석에는 LightGBM과 DNN 모델을 사용하였다. 모든 모델은 동일한 데이터셋과 전처리된 시계열 데이터를 기반으로 학습 및 검증을 수행하여, 알고리즘 간 예측 성능을 객관적으로 비교하였다.

구분	알고리즘	출력
분류	Random Forest	Normal / Mild / Severe
분류	XGBoost	Normal / Mild / Severe
분류	LSTM	Normal / Mild / Severe
회귀	LightGBM	수분 스트레스 지표
회귀	DNN	수분 스트레스 지표

표 3 예측 알고리즘 구성

II.3 실험 결과 및 분석

분류 모델의 성능 평가는 Accuracy, Precision, Recall, F1-score 지표를 사용하였으며, 회귀 모델의 경우 RMSE와 MAE를 활용하여 예측 오차를 정량적으로 평가하였다. 이를 통해 알고리즘 간 예측 성능을 객관적으로 비교하였다.

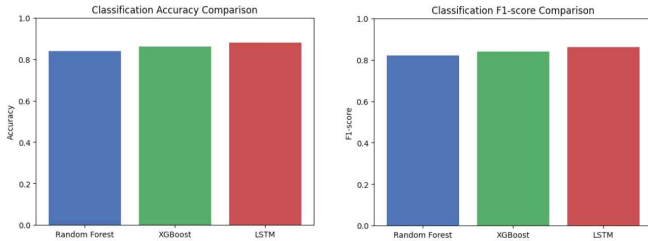


그림 1 분류 모델 성능 비교 결과

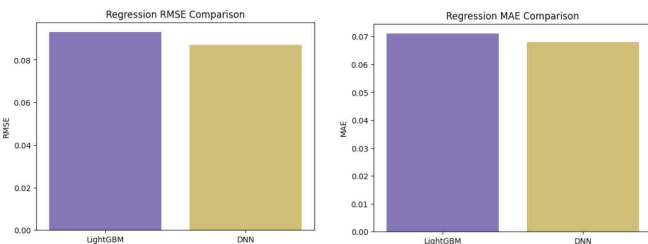


그림 2 회귀 모델 오차 비교 결과

실험 결과, 트리 기반 모델인 Random Forest와 XGBoost는 각각 Accuracy 0.84와 0.86 수준의 비교적 안정적인 분류 성능을 보였다. 반면 시계열 특성을 반영한 LSTM 모델은 Accuracy 0.88, F1-score 0.86으로 가장 우수한 예측 성능을 나타냈으며, 특히 환경 변화가 급격한 구간에서 분류 정확도가 상대적으로 향상 되는 경향을 보였다.

회귀 분석 결과에서도 LightGBM과 DNN 모델은 수분 스트레스 지표를 효과적으로 예측하였으며, DNN 모델은 RMSE 0.0087, MAE 0.0068로 LightGBM(RMSE 0.0093, MAE 0.0071) 대비 다소 낮은 예측 오차를 나타냈다. 다만 두 모델 간 오차 차이는 제한적인 수준으로 나타났다.

딥러닝 기반 모델은 상대적으로 높은 예측 성능을 보였으나 연산 비용이 증가하는 경향을 보였으며, 반면 트리 기반 모델은 비교적 낮은 연산 비용으로 안정적인 예측 성능을 유지하였다. 이를 통해 예측 정확도와

실시간 적용성을 동시에 고려한 모델 선택의 필요성을 확인하였다.

III. 결론

본 논문에서는 온실 환경의 복합 센서 기반 시계열 데이터를 활용하여 작물 수분 스트레스를 예측하고, 다양한 머신러닝 및 딥러닝 알고리즘의 성능을 비교·분석하였으며, 실제 온실 환경을 반영한 데이터셋을 기반으로 분류 및 회귀 문제를 구성하고, 예측 정확도와 실용성 측면에서 알고리즘별 특성을 분석하였다.

실험 결과, Random Forest와 XGBoost는 낮은 연산 비용으로 안정적인 예측 성능을 보였으며, 시계열 정보를 연속적으로 학습하는 LSTM 모델은 분류 정확도와 F1-score 측면에서 가장 우수한 성능을 나타냈다. 회귀 분석에서는 DNN 모델이 LightGBM 대비 다소 낮은 예측 오차를 보였으나, 모델 간 성능 차이는 제한적인 수준으로 확인되었다.

종합적으로, 정확도가 중요한 분석 환경에서는 LSTM 모델이 적합하며, 실시간성과 시스템 안정성이 요구되는 현장 환경에서는 트리 기반 모델이 실용적인 대안이 될 수 있다. 본 연구는 시계열 데이터 활용 방식에 따른 예측 성능과 실용성의 차이를 분석하여, 작물 수분 스트레스 예측 알고리즘 선택에 대한 기준을 제시한다는 점에서 의의가 있다.

향후 연구에서는 작물 및 재배 환경을 확대하여 예측 모델의 일반화 성능을 검증하려 한다. 또한 관수 의사결정 시스템과의 연계를 통해 예측 결과의 실제 현장 적용 가능성을 실증적으로 평가하는 연구를 진행할 예정이다.

ACKNOWLEDGMENT

“본 연구는 2025년도 전라남도 재원으로 전남인재평생교육진흥원의 지원을 받아 수행되었습니다.”

참 고 문 헌

- [1] Heeryong Ryu 외, “Damage index estimation by analysis of meteorological disasters on film plastic greenhouses.” International Journal of Agricultural and Biological Engineering, 2019, 12:58-63.
- [2] Vanthoor, B. H. E., Gázquez, J. C., Magán, J. J., Ruijs, M. N. A., Baeza, E., Stanghellini, C., van Henten, E. J., & de Visser, P. H. B. (2012). A methodology for model-based greenhouse design: Part 4, economic evaluation of different greenhouse designs: A Spanish case. Biosystems Engineering, 111(4), 336 - 349.
- [3] Hamlyn G. Jones, Irrigation scheduling: advantages and pitfalls of plant-based methods, Journal of Experimental Botany, Volume 55, Issue 407, November 2004, Pages 2427 - 2436,
- [4] Chaves MM, Maroco JP, Pereira JS. Understanding plant responses to drought - from genes to the whole plant. Funct Plant Biol. 2003 Mar;30(3):239-264. doi: 10.1071/FP02076. PMID: 32689007.
- [5] Conde, G., Guzmán, S. M., & Athelly, A. (2024). Adaptive and predictive decision support system for irrigation scheduling: An approach integrating humans in the control loop. Computers and Electronics in Agriculture, 217, 108640.