

PA-106

밀 생산량 예측을 위한 머신러닝 기반 생육 데이터 중요도 분석 및 주요 조사항목 선정 Importance Analysis of Wheat Growth Data and Selection of Optimal Sampling Items to Predict Wheat Yield based on Machine Learning

강다영¹, 전해진¹, 김태곤^{1*}

Dayoung Kang¹, Hyejin Jeon¹, Taegon Kim^{1*}

¹전북대학교 스마트팜학과

¹Dep. of Smart Farm, Jeonbuk National Univ, Jeonju 54896, Korea

[서론]

밀은 1인당 연간 소비량이 연간 33kg로 59kg인 쌀 다음으로 많이 소비되는 식량작물이다. 우리밀의 자급률은 0.7%로 98%인 쌀에 비해 매우 낮아, 밀산업 육성에 대한 요구가 있어왔다. 2020년 밀산업육성법이 시행됨에 따라, 농식품부는 2025년까지의 밀 자급률 5% 달성을 목표로 재배면적을 확대 및 국산 밀 생산성 증대를 위해 노력하고 있다. 생산량 예측 모델은 생산성 향상 계획 및 효과를 정량화하기 위한 주요한 도구로, 밀산업육성법 시행과 함께 밀 생산량 예측모형에 대한 수요도 높아지고 있다. 생산량 예측을 효율적으로 수행하기 위하여, 생육데이터를 선별하여 생산량 예측에 필요한 주요한 생육 조사 항목을 찾아내고자 한다.

[재료 및 방법]

Amit 등(2022)은 주간 날씨 데이터, 토양 조건 및 작물 계절 변수(파종, 개화 및 수확 날짜)를 사용하여 밀의 수확량을 예측하였고, Keach 등(2021) 최소 및 최대 기온, 강수, 지구 태양 복사 조도, 눈 깊이의 5가지 기상 요소를 사용하여 밀의 수확량을 예측하였다. 기존 연구에서는 주로 환경요인을 활용하여 생산량을 예측하였지만, 필지 내 수분관리에 따른 생산량 변화 및 생육단계별 작물의 생육자료를 활용한 생산량 예측 연구를 찾아보기 어려웠다. 본 연구는 밀의 생육단계별 생육 데이터를 수집하여, 이를 바탕으로 생산량을 예측해 중요한 요인을 찾고자 한다. 분석에 사용된 생육데이터는 전라북도 남원시 운봉읍에 위치한 실증단지 4개 처리구(한밭처리구 A, 무처리구 B, 적정처리구 C, 과습처리구 D)에서 수집하였다. 각 처리구별 3개 블록으로 실험하고, 생육데이터는 각 블록별로 5반복 샘플링을 수행하였다. 생육 데이터는 잎, 줄기, 종자, 각각의 건물중, 생체중, 수분 함량과 잎의 너비와 길이를 측정하여 확보하였고, 동시에 생육지표로 식생군의 LAI, SPAD를 측정하였다. 개화기를 포함한 4월 28일부터 5월 31일까지 2주 간격으로 조사된 생육데이터를 활용하여 생산량을 예측해 분석하였다.

[결과 및 고찰]

생산량 예측모델 중 XGBoost모델(MAE 163.44)은 개화기 이후 4주차의 SPAD 및 잎의 너비, 개화기의 잎 너비, LAI, 개화기 이후 2주차의 LAI가 중요한 변수로 나타났다. Linear Regression모델(MAE 191.22)은 개화기 이후 2주차 수분함량과 LAI, 개화기의 LAI와 잎의 길이가 중요한 변수로 나타났다. 본 연구에서 생육조사를 통해 밀의 생육단계별 작황에 따라 주요한 조사항목을 선별한 의미가 있다. 하지만 대상지역이 한 지역으로 한정되어 기상, 토양 등 환경요인을 반영하지 못한 점이 한계로 생각되는 바, 향후 연구에서는 지속적인 모니터링과 다른 지역의 생육자료를 확보하여, 환경조건, 생육조건을 모두 고려할 계획이다.

[사사]

본 연구는 농촌진흥청 연구사업(사업번호: PJ017048), 한국연구재단 이공분야 기초연구사업(사업번호: NRF-2022R1G1A1011147)지원에 의해 이루어진 결과로 이에 감사드립니다.

*Corresponding author: E-mail, taegon@jbnu.ac.kr Tel. +82-63-270-2557