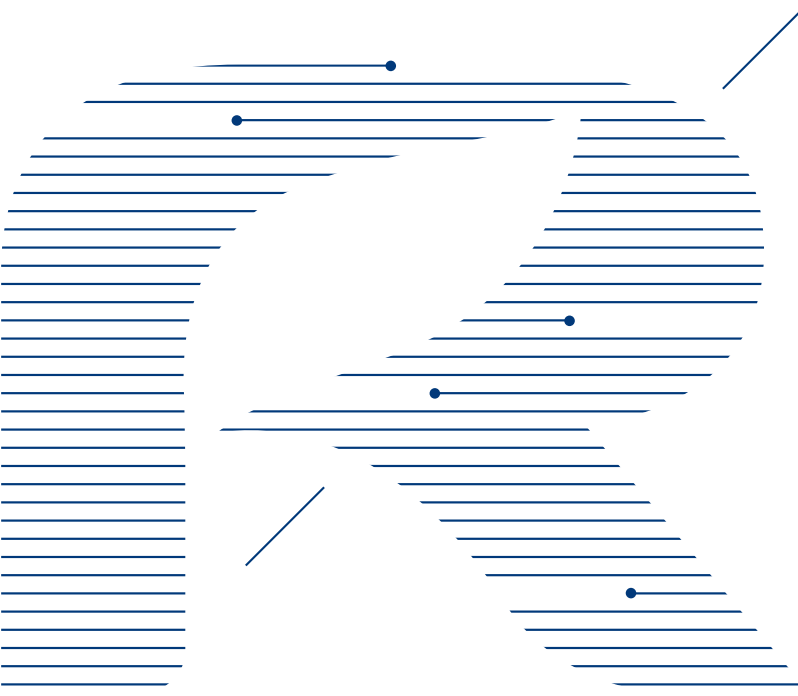


인공지능(AI)을 활용한 농업관측 고도화 방안 연구

- 항공촬영 및 실측조사를 중심으로

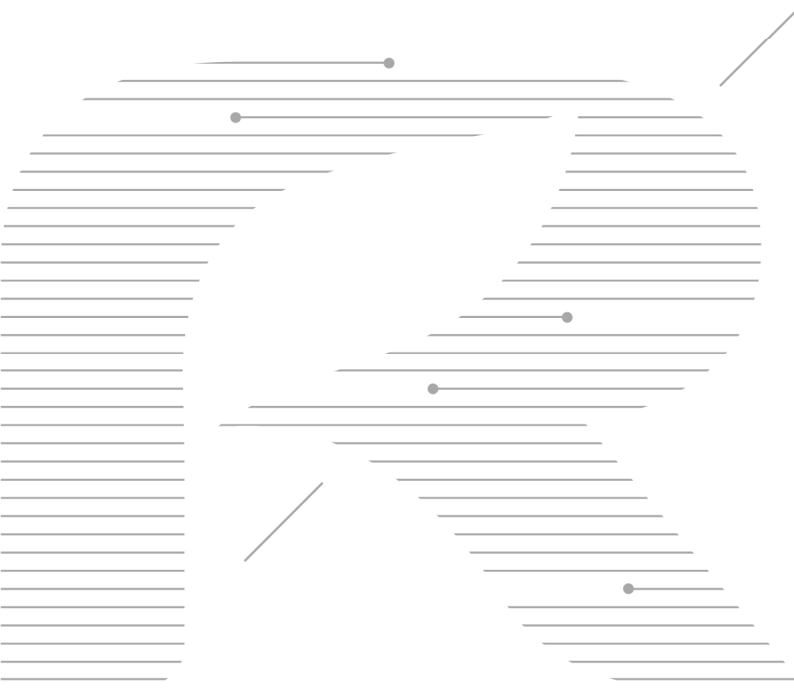
김원태 · 김라이 · 김연중 · 문홍성 · 윤성주 · 김가을



인공지능(AI)을 활용한 농업관측 고도화 방안 연구

- 항공촬영 및 실측조사를 중심으로

김원태 · 김라이 · 김연중 · 문홍성 · 윤성주 · 김가을



연구 담당

김원태 | 전문위원 | 연구 총괄, 제1~5장 집필

김라이 | 전문연구원 | 제2장 집필

김연중 | 명예선임연구위원 | 제1, 5장 집필

문홍성 | 전문연구원 | 제3~4장, 부록 집필

윤성주 | 전문연구원 | 제3장, 부록 집필

김가을 | 연구원 | 제3장 집필

특별연구 E2024-04

인공지능(AI)을 활용한 농업관측 고도화 방안 연구

- 항공촬영 및 실측조사를 중심으로

등 록 | 제6-0007호(1979. 5. 25.)

발 행 | 2024. 12.

발 행 인 | 한두봉

발 행 처 | 한국농촌경제연구원

우) 58321 전라남도 나주시 빛가람로 601

대표전화 1833-5500

인 쇄 처 | 세일포커스(주)

I S B N | 979-11-6149-758-7 95520

※ 이 책에 실린 내용은 한국농촌경제연구원의 공식 견해와 반드시 일치하는 것은 아닙니다.

※ 이 책에 실린 내용은 출처를 명시하면 자유롭게 인용할 수 있습니다.

무단 전재하거나 복사하면 법에 저촉됩니다.

최근 기후변화와 농업 환경의 급속한 변화는 농업 생산성과 품질에 중대한 영향을 미치고 있다. 이에 따라 농업 분야에서도 디지털 전환이 활발히 이루어지고 있으며, 특히 인공지능(AI)을 활용한 정밀 농업의 중요성이 대두되고 있다. AI 기술은 복잡한 데이터를 신속히 분석하고 정확한 예측을 제공함으로써 농업 생산 효율성을 극대화할 수 있는 잠재력을 가지고 있다.

본 연구에서는 농업관측센터에서 생성하고 있는 생육 실측 및 항공촬영 데이터를 대상으로 인공지능 분석 방법을 활용한 농업관측 고도화 방안을 모색하고자 한다. 생육 실측은 작물의 상태를 정밀하게 파악하여 생산성과 품질을 예측하는데 도움을 주며, 항공촬영은 넓은 필지의 재배 및 출하 상태를 파악하거나 좁은 필지의 작물 생육상황을 수집하는 데 효과적이다. 이러한 기술을 인공지능과 접목함으로써 농업관측의 정확성을 높이고, 이를 통해 농산물 수급관리 및 의사결정에 기여할 수 있다.

본 연구에서는 이러한 접근의 필요성과 가능성을 살펴보고, 실질적인 적용 사례와 문제점 및 개선 방향을 제시함으로써 농업관측 고도화에 대한 새로운 방향성을 제시하고자 한다.

그동안 데이터 전처리는 물론 코드 작성 및 분석 등 이 연구에 매진한 연구진의 노고에 감사하며, 아무쪼록 이 연구가 인공지능을 활용한 농업관측 고도화에 기여하는 기초자료로 활용되길 기대한다.

2024. 12.

한국농촌경제연구원장 **한두봉**

연구 목적

- 이 연구의 목적은 생육실측조사와 드론을 활용한 항공촬영에 의해 생성된 농업관측 정형·비정형 데이터와 기상변수를 접목하기 위하여 인공지능(AI) 활용기법을 제안하고, 일부 품목에 대해 주산지 시군의 구증을 예측하고 이를 실제치와 비교함으로써 농업관측 데이터 분석 고도화에 반영할 수 있는 방안을 도출하는 것임.
- 연구의 범위는 2021~2024년 농업관측센터에서 수집한 대서종 마늘과 준고랭지 배추의 데이터 수집 및 활용 실태를 정리하고 마늘과 배추의 정형·비정형 데이터를 활용하여 인공지능으로 구증을 예측한 후 실제치와 비교하는 것임. 지역은 마늘 주산지인 창녕군과 준고랭지 배추 주산지인 평창군으로 한정하며, 2021~2023년 데이터로 학습하였고, 2024년 데이터를 테스트하였음.

연구 방법

- 생육실측조사 및 영상촬영 데이터를 활용하여 전처리 과정을 수행한 후, 인공지능(AI)으로 마늘과 배추의 구증을 추정하였음. 생육실측조사 데이터를 이용하여 정형 데이터 분석, 영상촬영 데이터를 이용한 비정형 데이터 분석, 정형·비정형 데이터를 결합하여 추정한 멀티모달 분석을 진행하였음. 정형 데이터와 멀티모달 분석은 랜덤포레스트, XGBoost, ConvLSTM을 동시에 추정하여 예측력을 비교하였으며, 비정형 데이터 분석은 합성곱 신경망(Convolutional Neural Network: CNN)을 사용하였음.

연구 결과

- 생육실측 자료를 이용한 구증 예측 모델은 마늘의 경우 ConvLSTM, 배추 1·2기작은 XGBoost가 상대적으로 우수하였으나 전반적으로 예측력은 좋지 않았음. 멀티모달 분석 결과와 생육실측 자료 결과가 동일하게 나타났으나, 생육실측 자료만을 활용한 모델보다 예측력이 소폭 개선되었음. 이미지 자료를 이용한 CNN 모델은 배추 1기작에서 어느 정도 좋은 예측력을 보였으나 마늘과 배추 2기작의 예측력은 좋지 않았음.
- 전반적인 예측력이 좋지 않은 이유는 생육 실측 데이터와 이미지 데이터의 생성 목적이 본 연구의 AI 분석 목적과 달라 이를 매칭할 때 많은 정보들이 소실되었고 데이터의 축적 기간도 짧았기 때문임. 생육 실측 데이터는 비슷한 시기의 데이터 간 편차가 심한 경우도 많아 전처리 과정이 요구됨. 이미지 데이터는 생육실측의 보조 역할로 생성되었기 때문에 예측 모형의 데이터로 이용하는 데 한계가 있음.

시사점

- 생육실측조사 및 영상촬영 품목의 작기 또는 수확 종료 후 각 데이터의 조사 일자, 필지 등을 비교하여 사전에 매칭하는 인공지능 학습용 데이터셋 구축이 필요함. 이를 위해 생육실측조사와 항공촬영 가이드라인이 연동할 수 있도록 조정하는 작업도 요구됨. 또한, 농촌진흥청에서 발사 예정인 차세대중형위성 4호(농림위성)를 통해 제공되는 재배면적, 출하면적, 생육상황 등의 데이터를 농업관측에 활용하기 위한 기관 간 협력체계 구축도 필요함.

Research on Enhancing Agricultural Outlook Using Artificial Intelligence (AI)

- Focusing on Aerial Imaging and Field Surveys

Purpose of Research

- The objective of this study is to propose AI-based methodologies that integrate structured and unstructured agricultural outlook data generated from crop growth surveys and drone-based aerial imaging with meteorological variables. Additionally, the study aims to predict the final weight (yield) of specific crops in major producing regions and compare these predictions with actual data, thereby identifying ways to enhance agricultural observation data analysis.
- The scope of the research includes organizing the data status and utilization practices of garlic and semi-highland cabbage collected by the Agricultural Outlook Center from 2021 to 2024. Using structured and unstructured data for garlic and cabbage, the study employs AI techniques to predict final weights and compares these predictions to actual values. The study focuses on Changnyeong County (the main production area for garlic) and Pyeongchang County (the main production area for semi-highland cabbage). Due to limited data, the prediction is restricted to the final stage weight.

Research Method

- Using crop growth survey data and imaging data, preprocessing was performed, followed by estimating the weight of garlic and cabbage using Artificial Intelligence (AI). Structured data analysis was conducted using the growth survey data, unstructured data analysis was carried out using imaging data, and a multimodal analysis was performed by combining structured and unstructured data. For structured data and multimodal analysis, prediction accuracy was compared using Random Forest, XGBoost, and ConvLSTM simultaneously, while unstructured data analysis utilized a Convolutional Neural Network (CNN).

Main Findings

- For the crop weight prediction model using crop growth survey data, ConvLSTM showed relatively better performance for garlic, while XGBoost performed better for the first and second cabbage growing cycles. However, the overall prediction accuracy was not satisfactory. The multimodal data analysis showed similar results to the growth survey data models but demonstrated a slight improvement in prediction accuracy compared to models using only growth survey data. The CNN model using image data showed relatively good prediction accuracy for the first cabbage growing cycle but performed poorly for garlic and the second cabbage growing cycle.

- The primary reasons for the overall low prediction accuracy were the differing purposes for which growth survey data and image data were generated, leading to significant information loss during data matching, and the short duration of data accumulation. Growth survey data often exhibited high variance within similar time periods, necessitating preprocessing. Image data, being originally intended to supplement growth survey data, faced limitations in being used as primary data for prediction models.

Implications

- It is necessary to develop AI training datasets by pre-matching the survey dates, plots, and other data for each crop growth survey and imaging crop or post-harvest period. To achieve this, adjustments are also required to align the guidelines for crop growth surveys and aerial imaging. Additionally, it is essential to establish an inter-agency collaboration system to utilize agricultural outlook data provided by the upcoming Next-Generation Medium-Satellite No. 4 (Agricultural Satellite) scheduled for launch by the Rural Development Administration. This data includes variables such as cultivation area, harvest area, and crop growth prediction models.

Researchers: KIM Wontae, KIM Rayi, KIM Yeanjung, MUN Hongsung, YUN Sungju,
KIM Gaeul

Research period: 2024. 3. - 2024. 12.

E-mail address: wtkim@krei.re.kr

제1장 서론

1. 연구의 필요성과 목적 1
 2. 연구내용 및 방법 4
 3. 선행연구 검토 및 차별성 7

제2장 농업관측 관련 데이터 현황

1. 농업관측센터 데이터 수집 현황 11
 2. 유관기관 데이터 수집 현황 23
 3. 시사점 28

제3장 생육·영상 자료 수집 및 전처리

1. 생육실측 자료 수집 및 전처리 31
 2. 영상 자료 수집 및 전처리 37
 3. 기상 자료 수집 43
 4. 시사점 44

제4장 인공지능(AI) 활용 생산량 추정

1. 인공지능(AI) 활용 분석 개요 47
 2. 정형 데이터 분석 49
 3. 이미지 데이터 분석 65
 4. 멀티모달 분석 74
 5. 시사점 83

제5장 농업관측 데이터 활용 고도화 방안

- 1. 인공지능 학습용 데이터 구축 87
- 2. 농림위성 데이터 활용 협력 방안 구축 89
- 3. 인공지능 예측 모델 고도화 90

부록

- 1. 생육 및 수량 실측조사 평균값 95
- 2. 코드 예시: 마늘 모형 분석 코드 97

참고문헌 107

제2장

〈표 2-1〉 농업관측 품목 및 제공 정보 12

〈표 2-2〉 농업관측 실측조사 사업 품목별 대상자 현황(2024년) 14

〈표 2-3〉 품목별 조사 횟수 및 시기 16

〈표 2-4〉 품목별 생육실측조사 항목 17

〈표 2-5〉 품목별 생산량 실측조사 항목 18

〈표 2-6〉 농업관측센터 항공촬영 사업 현황 21

〈표 2-7〉 농촌진흥청 농작물 생육조사 현황 23

〈표 2-8〉 농촌진흥청 원격탐사 및 빅데이터 활용 주요 사업 내용 24

〈표 2-9〉 차세대중형위성 4호(농림위성) 주요 특징 25

〈표 2-10〉 농림위성 활용 계획 26

〈표 2-11〉 한국지능정보사회진흥원 인공지능 학습용 데이터 수집 현황 28

제3장

〈표 3-1〉 마늘 실측조사 회차별 조사항목 상세 내용 32

〈표 3-2〉 창녕지역 마늘 실측조사 시기 및 횟수 33

〈표 3-3〉 창녕지역 조사 회차별 마늘 생구 무게 실측조사 평균값과 표본 수 ... 34

〈표 3-4〉 평창지역 고랭지배추 정식시기별 실측조사 시기 35

〈표 3-5〉 평창지역 조사 회차별 고랭지배추 구중 실측조사 평균값과 표본 수 ... 36

〈표 3-6〉 고도별 촬영 면적 37

〈표 3-7〉 생육실측조사 영상 자료 현황: 마늘, 차수별 39

〈표 3-8〉 생육실측조사 영상 자료 현황: 마늘, 지역별 40

〈표 3-9〉 생육실측조사 영상 자료 현황: 여름배추, 차수별 41

〈표 3-10〉 생육실측조사 영상 자료 현황: 여름배추, 지역별 41

〈표 3-11〉 분석에 활용된 영상 자료: 마늘	42
〈표 3-12〉 분석에 활용된 영상 자료: 배추	43

제4장

〈표 4-1〉 마늘 정형 데이터 구중 모형 하이퍼파라미터 최적화	59
〈표 4-2〉 마늘 정형 데이터 구중 모형 성능지표	60
〈표 4-3〉 배추 1기작 정형 데이터 구중 모형 하이퍼파라미터 최적화	62
〈표 4-4〉 배추 2기작 정형 데이터 구중 모형 하이퍼파라미터 최적화	62
〈표 4-5〉 배추 1기작 정형 데이터 구중 모형 성능지표	65
〈표 4-6〉 배추 2기작 정형 데이터 구중 모형 성능지표	65
〈표 4-7〉 마늘 CNN 모델 하이퍼파라미터 최적화 결과	70
〈표 4-8〉 CNN 모델 요약	71
〈표 4-9〉 마늘 CNN 모델 성능지표 비교	71
〈표 4-10〉 배추 CNN 모델 하이퍼파라미터 최적화 결과	72
〈표 4-11〉 CNN 모델 요약	73
〈표 4-12〉 CNN 모델 성능지표 비교	73
〈표 4-13〉 마늘 멀티모달 구중 모형 하이퍼파라미터 최적화	77
〈표 4-14〉 마늘 멀티모달 구중 모형 성능지표	78
〈표 4-15〉 배추 1기작 멀티모달 구중 모형 하이퍼파라미터 최적화	80
〈표 4-16〉 배추 2기작 멀티모달 구중 모형 하이퍼파라미터 최적화	80
〈표 4-17〉 배추 1기작 멀티모달 구중 모형 성능지표	83
〈표 4-18〉 배추 2기작 멀티모달 구중 모형 성능지표	83

부록

〈부표 1-1〉 창녕지역 마늘 생육 및 수량 실측조사 평균값	95
〈부표 1-2〉 평창지역 고랭지배추 생육 및 수량 실측조사 평균값	96

제2장

<그림 2-1> 농업관측정보 수집 및 분석 체계 13
 <그림 2-2> 실측조사 조사구역 선정 방식 15
 <그림 2-3> 배추 생육실측조사 방법(예시) 17
 <그림 2-4> 품목별 생산량 실측조사 방법(예시) 18
 <그림 2-5> 농업관측센터 항공촬영 사업 추진 체계 21
 <그림 2-6> 항공촬영 영상(예시) 22
 <그림 2-7> 품목 재배분포도(고흥군 금산면 마늘, 양파 예시) 22

제4장

<그림 4-1> Random Forest 기본구조 53
 <그림 4-2> XGBoost 기본구조 55
 <그림 4-3> ConvLSTM 기본구조 57
 <그림 4-4> 마늘 정형 데이터 구중 모형 특성 중요도 60
 <그림 4-5> 배추 1기작 정형 데이터 구중 모형 특성 중요도 63
 <그림 4-6> 배추 2기작 정형 데이터 구중 모형 특성 중요도 64
 <그림 4-7> CNN 기본구조 69
 <그림 4-8> 마늘 이미지 데이터 유사도 70
 <그림 4-9> 배추 이미지 데이터 유사도 72
 <그림 4-10> 마늘 멀티모달 구중 모형 특성 중요도 78
 <그림 4-11> 배추 1기작 멀티모달 구중 모형 특성 중요도 81
 <그림 4-12> 배추 2기작 멀티모달 구중 모형 특성 중요도 82

1

서론

1. 연구의 필요성과 목적

1.1. 연구 배경과 필요성

- 농업관측센터에서 조사·분석하여 제공하고 있는 농업관측정보는 정부의 수급안정정책 수립에 기초자료로 활용되고 있으며, 이에 따라 관측정보의 정확도 제고를 위해 다양하고 객관적인 산지 정보수집 방법이 사용되고 있음.
 - 농업관측센터는 표본농가 및 모니터 전화조사, 주산지 방문조사, 인력을 활용한 실측조사, 드론을 이용한 항공촬영 등의 방법으로 품목별 재배면적, 출하면적, 생육상황, 단수, 생산량 등의 정보를 제공하고 있음.

- 농업관측센터는 통계청의 농업 관련 통계와 전화조사를 활용한 농업관측정보의 불일치 문제가 지속적으로 제기됨에 따라 2020년부터 5대 채소류인 배추, 무, 마늘, 양파, 건고추를 대상으로 실측조사를 도입하였음.

- 실측조사는 원예조사원 또는 외부 전문업체의 조사원들이 농업관측 품목별 표본농가의 필지를 직접 방문하여 재배면적과 생육상황, 생산량을 실제로 측정하는 방법임.
- 농업관측 실측조사 결과가 통계청에 비해 생산 현장을 비교적 잘 반영하고 있다는 농업인단체의 의견이 높아짐에 따라 농식품부에서는 5대 채소류의 수급안정대책 수립 시 통계청 농업통계를 대신하여 농업관측정보를 활용하고 있음.

○ 한편, 농업관측센터는 인력에 의존하는 재배면적 실측조사의 한계를 보완하고자 드론을 활용한 항공촬영 사업도 도입하여 조사범위 확대 및 효율성 제고를 도모하였음.

- 2019년까지는 농업관측센터 인력을 활용하여 강원지역 여름배추와 전남 지역 겨울배추에 대한 영상정보를 수집하여 재배면적을 산출하였음.
- 항공촬영을 통한 재배면적 산출 요구가 주요 5대 품목으로 확대되면서 2020년부터 시범사업 형태로 외부전문업체 위탁을 통한 항공촬영을 적극 활용하고 있음.

○ 농업관측 항공촬영 사업은 주산지가 집중되어 있는 품목의 소규모 단위 재배면적 전수조사와 특정 시기의 출하면적 및 잔여면적 산출이라는 이전과는 차별적인 관측정보를 제공할 수 있게 되었음.

- 제주도 월동작물 항공촬영 사업으로 생산 집중도가 높은 지역에 분포하는 품목의 전수조사 가능성을 확인하였으며, 촬영지역 주제도 등 영상정보를 관측월보에 포함함으로써 단기관측정보의 시각화 자료를 제공하는 다각적인 성과가 있었음.

○ 다양한 기관에서 생성되는 생산 관련 방대한 정보의 복합 활용은 수급 관련 정보 고도화에 기여할 수 있지만, 축적된 데이터 활용에 대한 연구는 미흡한 상황임.

- 농업부문에서는 주로 드론을 활용한 항공촬영이 진행되고 있는데, 농촌진흥청은 항공촬영으로 취득한 화상자료, 초분광자료, 식생지수 등을 활용하여 작물의 작황을 파악하고 있으며, 농산물품질관리원은 직불제 이행 점검 수단으로 드론을 활용하고 있음. 또한 통계청은 농업재배면적조사의 보조 수단으로 드론을 활용하고 있음.

○ 이러한 농업부문 항공촬영은 대부분 작물 식별과 재배면적 산출을 목적으로 진행되고 있으며, 농촌진흥청에서 진행하는 항공촬영이 그나마 작물의 작황 파악에 상대적으로 큰 비중을 두고 있음. 그러나 각 기관에서 생성된 항공촬영 데이터가 서로 공유되지 못하고 있어 축적된 데이터를 통합적으로 활용하지 못하고 있음.

○ 따라서 관측 품목의 정확한 생산량 예측을 위해 농업관측센터의 실측조사와 항공촬영으로 생성된 데이터뿐만 아니라 증장기적으로 유관기관에서 보유하고 있는 병해충이미지 등 비정형 데이터를 접목하여 인공지능으로 학습시켜 품목분류 및 재배면적 산출, 작황 및 단수 예측이 가능한 시스템을 구축하는 것이 필요함.

1.2. 연구 범위 및 목적

○ 이 연구의 범위는 첫째, 관측 정보화 사업의 일환으로 2021~2024년 농업관측센터에서 수집한 창녕군 대서종 마늘과 평창군 준고랭지 배추의 데이터 현

황 및 활용 실태를 정리하고 개선방안을 도출함. 둘째, 마늘과 배추의 정형·비정형 데이터를 활용하여 인공지능 분석 방법으로 구증을 추정한 후 실제치와 비교하여 예측력을 검증함. 셋째, 수집된 데이터가 4년으로 짧아 차수별 구증 예측을 하지 않고 최종 차수의 구증만을 예측하는 것으로 한정함.

- 이 연구의 목적은 농업관측센터의 생육실측조사 결과, 드론을 활용한 항공촬영 영상 데이터 등 정형·비정형 데이터와 유관기관의 기상 데이터 등을 접목하여 인공지능 분석 방법으로 구증을 예측한 후 이를 실제치와 비교함으로써 농업관측 데이터 분석 고도화 방안을 도출하는 것임.

2. 연구내용 및 방법

2.1. 연구내용

- 제2장에서는 농업관측사업 관련 데이터 생성 현황을 파악함. 먼저 농업관측 전담기관인 농업관측센터에서 제공하고 있는 농업관측 관련 데이터의 종류와 수집 및 작성 방법을 정리함.
 - 다음으로 지방자치단체, 농업관련 유관기관 등에서 작성하고 있는 데이터의 수집 및 이용 현황을 검토하고 시사점을 도출함.
- 제3장에서는 본 연구의 분석 데이터인 창녕군 대서 마늘과 평창군 준고랭지 배추에 대한 생육 실측 자료와 드론 촬영 영상 데이터 수집 현황을 설명함.
 - 먼저 마늘과 배추의 생육실측조사 회차별 조사항목, 실측조사 시기 및 횟수, 회차별 실측조사 표본 수 및 평균값 등 실측조사 현황을 정리함.

- 다음으로 마늘과 배추 실측포전에 대한 드론 촬영 시기 및 횟수, 회차별 영상촬영 포전 수, 영상촬영 불가 횟수 및 이유 등 드론을 활용한 영상 데이터 수집 현황을 정리함.
 - 이어서 생육실측조사 데이터와 드론 촬영 데이터의 조사시기, 장소 등을 비교하여 조사 시기와 장소가 일치하는 자료를 추출하여 분석 데이터로 활용함.
- 제4장에서는 전처리된 데이터셋으로 인공지능을 이용하여 마늘과 배추의 구중을 추정하고 그 예측력을 평가함.
- 먼저 다양한 인공지능(AI) 모델을 활용하여 마늘과 배추의 구중 추정이 가능하도록 데이터를 전처리하여 데이터셋을 구축함.
 - 생육실측 자료, 기상 자료 등 정형 데이터를 활용하여 마늘과 배추의 구중을 추정함. 다음으로 마늘과 배추의 영상 데이터 등 비정형 데이터를 이용하여 마늘과 배추의 구중을 추정함. 마지막으로 생육실측 자료, 기상 자료, 영상 데이터 등을 모두 활용하는 멀티모달 예측모형으로 마늘과 배추의 구중을 추정함.
 - 인공지능을 활용한 마늘과 배추의 구중 추정은 2021~2023년 데이터를 학습데이터로 이용하고, 2024년 데이터는 테스트데이터로 활용하여 모형의 예측력을 평가함.
- 제5장에서는 생육실측 자료와 영상 자료, 기상 자료 등 정형·비정형 데이터를 대상으로 인공지능(AI)을 활용한 농업관측 고도화 방안을 제시함.
- 농업관측센터에서 수집하고 있는 정형·비정형 데이터의 효율적 관리 및 활용방안을 제시함. 그리고 인공지능(AI) 모델을 이용하여 추정된 마늘과 배추의 구중 예측 결과와 그 의미, 예측력 개선방안 등을 제시함.

2.2. 연구방법

- 농업부문 인공지능(AI) 활용 실태, 정형·비정형 데이터를 활용한 농업부문 생산량 예측방법론 등에 관한 선행연구를 국내외 문헌조사를 통해 수행함.
- 마늘과 배추의 생육실측조사 데이터, 드론 영상촬영 데이터를 수집하여 인공지능 분석에 사용 가능한 데이터를 추출하고, 부족한 영상 데이터 확충을 위해 회전, 좌우반전, 밝기조절, 대비조절 등의 이미지 증강 전처리 과정을 수행함.
- 전처리된 정형·비정형 데이터를 사용하여 인공지능(AI)으로 마늘과 배추의 구증을 추정하였음. 생육실측조사 데이터를 이용하여 정형 데이터 분석, 영상 촬영 데이터를 이용한 비정형 데이터 분석, 정형·비정형 데이터를 모두 결합하여 추정한 멀티모달 분석을 진행하였음.
 - 정형 데이터 분석은 랜덤포레스트, XGBoost, ConvLSTM으로 동시에 추정하여 예측력을 비교하였음.
 - 비정형 데이터 분석은 예측력 제고를 위해 이미지 증강을 진행한 후, 선행 연구에서 예측력이 우수한 것으로 나타난 합성곱 신경망(Convolutional Neural Network: CNN)을 사용하였음.
 - 정형·비정형 데이터를 결합한 멀티모달 분석은 랜덤포레스트, XGBoost, ConvLSTM으로 추정하였으며, 정형 데이터만을 이용한 추정 결과와 예측력을 비교하였음.
- 마늘과 배추의 구증을 다양한 인공지능 방법으로 추정하기 위한 데이터 수집 과정에서 도출한 시사점과 구증 추정 결과를 바탕으로 농업관측 고도화를 위한 개선방안을 도출함.

3. 선행연구 검토 및 차별성

3.1. 선행연구 검토

- 국승용 외(2022)는 ‘드론을 활용한 농업생산조사 방법 연구’에서 농업생산조사에 드론 활용 시 발생하는 사항들을 검토하여 신뢰성을 높일 수 있도록 효율적인 기준을 수립하고, 기관 간 논의 및 협력 강화, 표준 제정 등과 같은 활용방안을 제시하였음. 연구 결과 원예농산물의 작목 식별 및 재배면적 확인 방식에 대한 후속 딥러닝 모형 개발 및 양파, 마늘 등의 개체 식별 모형 개발에 관한 추가적인 연구 필요성을 언급함.
- 윤병삼 외(2023)는 ‘농업관측 항공촬영 사업 평가’에서 농업관측센터에서 수행 중인 항공촬영 사업의 현황조사 및 사업 운영 자료 검토를 통해 농업관측 사업의 성과평가 및 한계점을 도출하고, 사업 개선방안과 중장기 계획을 제안함.
- 문태원 외(2020)는 ‘합성곱 신경망을 이용한 온실 파프리카의 작물 생체중 추정’에서 온실 파프리카의 생육 이미지를 이용하여 CNN 기법으로 파프리카의 생체중을 추정하였음. 그 결과 그동안 품목 분류, 면적 계산, 질병 판독 등에 활용되던 CNN 기법이 품목의 생체중 추정에도 이용될 수 있음을 확인하였음.
- 박준 외(2022)는 ‘ResNet 기반 작물 생육단계 추정 모델 개발’에서 ResNet의 Pooling Layer를 수정한 CNN(Convolution Neural Network) 모델을 사용하여 사과와 복숭아의 생육단계를 추정한 결과, ResNet 모델과 DenseNet 모델의 생육단계 추정보다 우수한 추정 결과를 확인하였음.

- 채철주 외(2022)는 ‘농업 공공 빅데이터를 이용한 머신러닝 기반 생산량 및 판매 수익금 예측’에서 시설 원예 스마트팜에서 수집된 농업 디지털 데이터를 활용하여 다중선형회귀모형, 서포트벡터머신(Support Vector Machine: SVM), 인공신경망(Artificial Neural Network: ANN), 순환신경망(Recurrent Neural Network: RNN), LSTM, ConvLSTM(Convolutional LSTM) 등으로 농작물의 생산량 및 판매 수익금을 예측한 결과, LSTM에 Convolutional Layer를 추가한 ConvLSTM이 가장 우수한 성능을 보였음.
- 김원성(2022)은 ‘실측정보를 이용한 딥러닝 기반 마늘·양파 생구무게 예측 Neural Prophet 시차모델’에서 농촌경제연구원의 난지형 마늘과 중만생종 양파 실측정보를 Neural Prophet의 시차모델에 적용하여 최종 생산량을 예측함.
- 이여진 외(2022)는 ‘다변수 LSTM 딥러닝 네트워크를 이용한 육계 시세 예측 모델 연구’에서 육계 시세의 형성에 영향을 주는 다양한 시계열 데이터를 다변수 LSTM 네트워크에 적용하여 가격 예측 모형을 설계함.
- 박혜빈 외(2023)는 ‘딥러닝을 활용한 위성영상 기반의 강원도 지역의 배추와 무 수확량 예측’에서 위성영상과 통계청 생산량 자료를 활용하여 다양한 딥러닝 기법으로 강원도 고랭지 배추와 무의 수확량을 예측하였음. 연구 결과 무의 수확량 예측은 CNN 모델이 가장 우수하였으며, 배추의 수확량 예측은 RF(Random Forest) 모델이 가장 우수한 예측력을 보였음.
- 장일식·박구만(2023)은 ‘딥러닝 기반 시설재배 생산량 예측 연구’에서 시설 원예 스마트팜 작물인 오이와 파프리카의 환경 및 생육 정보를 바탕으로 딥러닝 모델인 MPL, RNN, LSTM, GRU, TCN 등을 적용하여 예측 성능의 결과를 분석하였음.

- 김재휘 외(2023)는 ‘대파 도매가격 예측모형 고도화 방안’에서 대파 품목을 대상으로 ARIMA, Holt-윈터스 평활법, 랜덤포레스트를 활용하여 대파 가격 예측력을 높일 수 있는 시계열 모형을 구축함.
- Alhnaity et al.(2019)은 ‘Using deep learning to predict plant growth and yield in greenhouse environments’에서 온실에서 측정된 토마토 줄기의 직경과 온실에 장착된 센서로 내부 요인을 이용하여 SVR, 랜덤포레스트, LSTM으로 예측력을 측정한 결과, LSTM이 가장 우수한 성능을 보였음.

3.2. 선행연구와의 차별성

- 기존의 연구는 실측조사 자료를 활용하여 생산량을 예측하거나 항공촬영 데이터를 이용하여 품목 식별과 재배면적을 산출하는 데 국한되어 있음. 본 연구에서는 생육실측 자료, 항공촬영 데이터, 기상 자료 등 정형·비정형 데이터를 활용하여 다양한 인공지능 분석 방법으로 구중을 예측한다는 데 선행연구와 차별성이 있음.

2

농업관측 관련 데이터 현황

1. 농업관측센터 데이터 수집 현황

1.1. 농업관측사업 추진 현황

○ 농업관측사업은 주요 농축산물 36개 품목을 대상으로 수급 및 가격 전망을 제공하여 농업인은 물론 정책담당자와 관련 종사자의 합리적인 의사결정을 지원하고 있음. 이를 위해 농업관측센터는 정기적으로 다양한 정보를 수집하고 분석하여 발표하고 있음.

- 농업관측센터는 경제주체의 다양한 의사결정을 지원하기 위해 품목별 특성을 반영하여 정보수집 및 분석 체계를 구축하였으며, 생육 및 수급 전 주기의 주요 지표 산출을 위해 다양한 조사 및 분석 방법을 활용하고 있음.
- 농업관측사업을 통해 제공되는 정보는 아래 <표 2-1>과 같음. 품목별로 재배(입식) 등 의사결정에 앞서 재배 의향을 조사하며, 행위가 완료된 이후 실제 재배(입식)면적 등을 조사함. 품목별 수급량 예측을 통해 시장으로 출하

될 물량을 산출하고 저장 동향, 가격 동향 등의 정보를 활용하여 시기별 가격을 예측하고 있음.

- 품목 관측정보 외 주요 채소류 5개 품목을 대상으로 실시하고 있는 생육조사 결과와 해외 주요 축산물의 수급 동향, 증장기 수급 전망을 위한 농업전망 보고서 등 다양한 정보를 제공하고 있음.

〈표 2-1〉 농업관측 품목 및 제공 정보

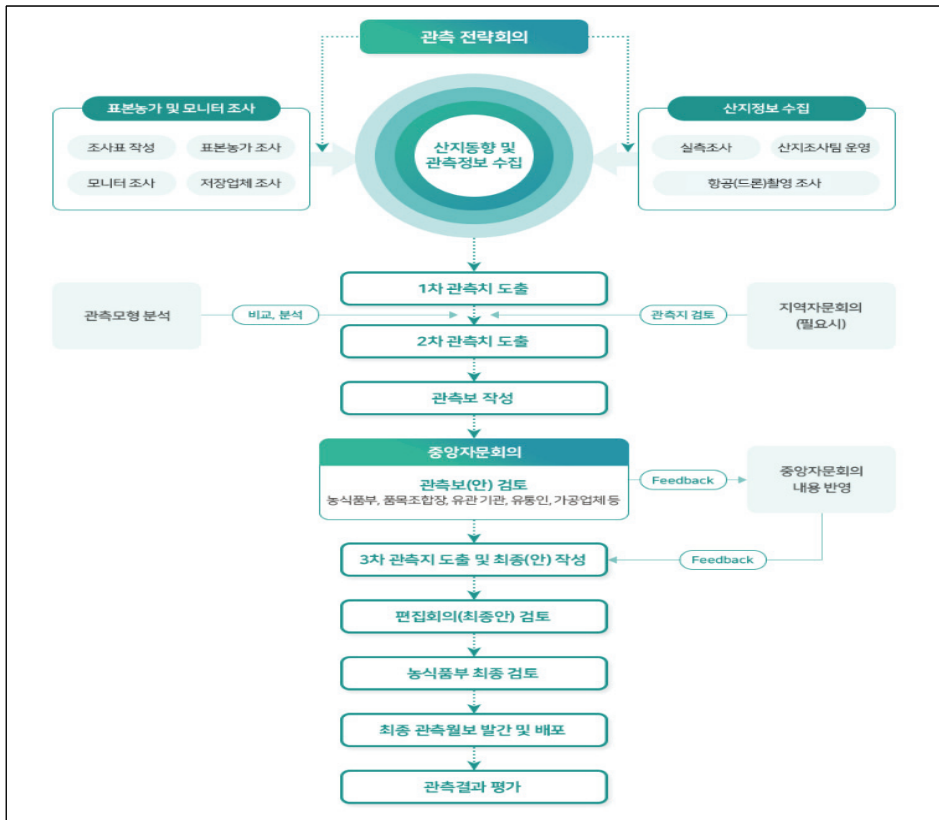
구분		공표 주기	공표일	내용
채소 (8)	엽근채소	배추, 무, 당근, 양배추	3~12월	가격 및 출하동향, 재배(의향)면적, 작황, 출하량 전망, 수입동향 및 해외정보, 기상 정보, 주간농사정보 등
	양념채소	건고추, 마늘, 양파, 대파	3~12월	
버섯(1)		3, 6, 9, 12월	1일	
과일 (6)	사과, 배, 포도, 감귤		4~12월	
	복숭아		4~9월, 12월	
	단감		6, 8~12월	
과채 (8)	오이, 애호박, 토마토, 풋고추, 파프리카		3~12월	
	수박		3~10월	
	참외		3~9, 12월	
식량 작물 (3)	쌀		3, 5, 9, 10, 11월	
	콩		3, 5, 10, 11월	
	감자		2~11월	
축산 (6)	육계		3~12월	사육동향, 도축동향, 수입량, 가격동향 및 전망 등
	한육우, 젓소, 돼지, 산란계, 오리		4~9월, 12월	
국제 곡물 (4)	밀, 옥수수, 대두, 국제쌀		3~12월	가격동향, 국내외 시장동향 및 전망 등

자료: 농업관측센터 홈페이지(aglook.krei.re.kr.), 검색일: 2024. 11. 20.

- 농업관측센터는 관측정보의 생산 및 확산을 위해 정보수집→정보분석→정보확산의 3단계를 거치고 있으며 분석을 위한 기초데이터 수집을 위해 자원을 집중하고 있음. 관측사업 추진 체계는 〈그림 2-1〉과 같음.

- 과거의 품목별 수급 동향 및 전망 자료 생산은 농업경영체 중심의 표본농가 조사와 지역의 전반적인 수급 상황을 광범위하게 파악할 수 있는 모니터 조사, 저장성이 있는 품목의 입출고 동향 및 전망 등 정보를 수집하기 위한 저장업체 조사의 의존도가 높았음.
- 기존 관측정보가 분야별 표본 응답을 통해 생산된 데이터의 의존도가 높다는 의견에 따라 농업관측사업 고도화를 추진하였으며, 주요 채소류를 중심으로 시기별로 작황 및 생산량을 측정하는 실측조사 사업과 드론을 활용한 광범위한 지역의 재배면적 확인을 위한 항공촬영 사업을 추진함.

〈그림 2-1〉 농업관측정보 수집 및 분석 체계



자료: 농업관측센터 홈페이지(aglook.krei.re.kr.), 검색일: 2024. 11. 20.

1.2. 농업관측 실측조사 사업

○ 농업관측센터는 기존 표본조사 방식의 비표본오차 발생 가능성을 줄이고 정량화된 실제 측정 데이터를 기반으로 한 분석체계 도입을 위해 주요 채소류 5개 품목(마늘, 양파, 건고추, 배추, 무)을 대상으로 2020년부터 실측조사 사업을 시행하고 있음.

- 2024년 기준 정기 실측조사는 5대 주요 채소류 표본농가 4,893호를 대상으로 시행하고 있음.
- 최근 사과·배 등 주요 과실류의 수급 불안정이 상시화되고 '과수산업 경쟁력 제고 대책(2024~2030)'의 일환으로 2024년부터 2개 품목(사과 100호, 배 3호 대상)을 대상으로 생육상황을 시범적으로 조사하고 있음.

〈표 2-2〉 농업관측 실측조사 사업 품목별 대상자 현황(2024년)

단위: 호

품목명	재배면적 조사	생육 조사	생산량 조사	비고
양파	928	100	928	-
마늘	621	100	621	-
건고추	1,125	100	1,125	-
배추	1,721	330	1,721	-
무	498	160	498	-
사과	100	100	100	시범
배	3	3	3	시범

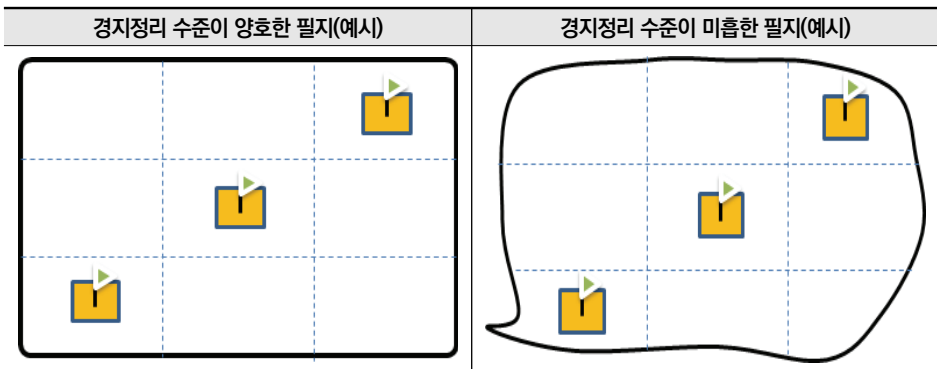
자료: 농업관측센터 내부자료.

- 실측조사 사업은 조사 과정에서 조사원에 의한 비표본오차 발생 가능성이 있음. 이를 해결하기 위해 품목별로 재배면적, 생육, 생산량 실측조사 매뉴얼을 개발하였고 조사원은 매뉴얼에 기재된 정해진 절차에 따라 표준화된 조사를 수행하고 데이터를 수집하고 있음.

○ 재배면적 실측조사는 표본농가를 대상으로 해당 품목의 정식(파종) 이후 농가 별 실제 경작이 이루어진 모든 필지를 방문하여 경작 여부 및 재배면적을 확인하는 조사임.

- 표본농가 전화조사 시 경영체가 응답한 면적과 필지 개수를 확인하고 조사 일정이 확정되면 표본농가와 동행하여 재배된 필지를 모두 조사함. 현장조사에서는 필지별 정식(파종)면적을 확인하고 정확한 면적 측정을 위해 GPS 면적측정단말기를 활용하여 실제 면적을 측정하고 있음. 1개의 필지에 2개 이상의 품목이 경작되고 있는 경우 각 품목의 재배면적을 산출하여 면적측정 정확도를 높이고 있음.
- 현장방문을 통해 실측된 재배면적은 전화조사에서 응답한 면적과 비교하여 응답 정확성을 검토하고 상이한 결과가 도출된 경우 재배면적을 보정하는 근거로 활용하고 있음.
- 재배면적 실측조사 시 향후 정기적인 생육조사와 최종 생산량 확인을 위한 실측조사구역(필지당 3개소)을 선정하여 관리·운영하고 있음.

〈그림 2-2〉 실측조사 조사구역 선정 방식



자료: 저자 작성.

○ 생육실측조사는 해당 조사 품목의 시기별 생육상황을 확인하고 최종 생산량을 예측하기 위한 기초자료로 활용하기 위해 생산량 조사구역 중 일부를 선정하여 실시하는 정기조사임.

- 생육실측조사는 품목별 생육 주기와 특성을 고려하여 아래 <표 2-3>과 같이 생육 단계별 개체의 크기를 직접 측정하는 조사로 조사 지침서에 근거한 시기별 표준화된 조사방식으로 데이터를 수집하고 있음.

<표 2-3> 품목별 조사 횟수 및 시기

구분	양파, 마늘	건고추	배추, 무
횟수	11회	9회	4회
조사 주기	10~15일 (구비대기 전) 이전 회차 기준 15일 (구비대기 후) 이전 회차 기준 10일	15일	(1차) 정식(파종) 후 20~30일 (2차) 1차 조사 후 20~30일 (3차) 2차 조사 후 15~20일 (4차) 3차 조사 후 5일(혹은 수확예정일 5일 전)

자료: 농업관측센터 내부자료를 바탕으로 저자 작성.

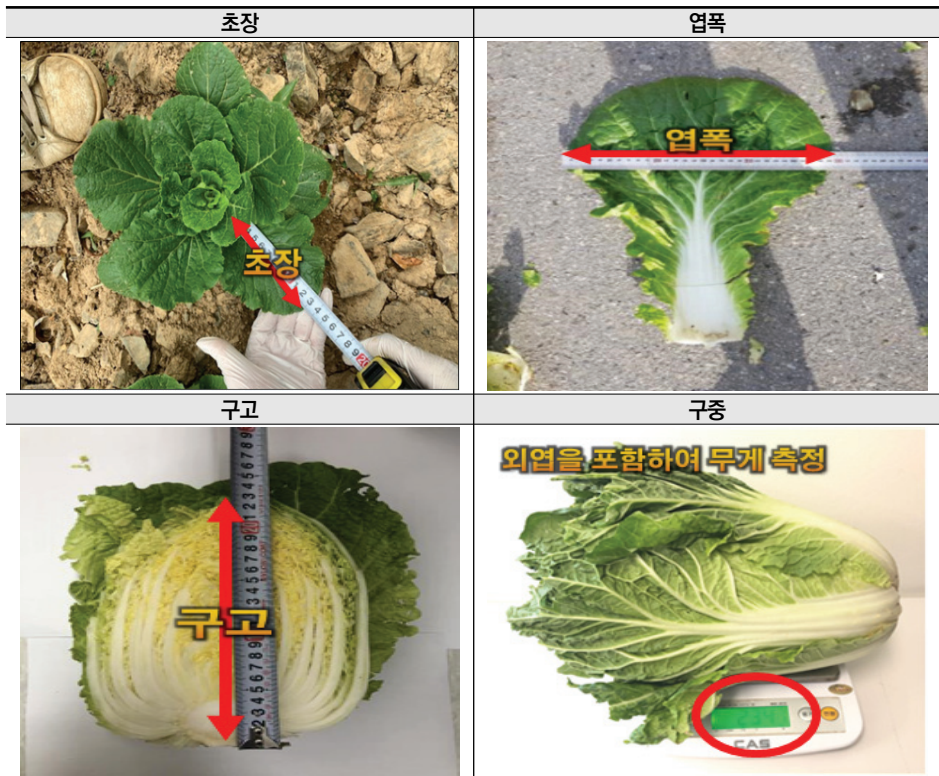
- 품목별 생육상황을 실측조사하기 위해 생육 조사구역의 경영체 정보, 품종, 종자, 관수시설 유무, 정식(파종)일, 수확(예정)일, 필지 정보(이랑폭, 두둑폭, 재식주수, 결주수) 등을 수집함.
- 품목별로 생육 상태를 파악하기 위해 품목 생육 단계를 고려한 지표를 개발하고 표준화된 방법으로 조사하고 있음. 생육 단계를 고려하기 위해 생육 초기/중·후기로 구분한 후 조사항목을 추가하여 실측이 이루어지고 있음.
- 생육 초기에는 작물의 키, 폭 등 재식 상태에서 측정할 수 있는 지표를 중심으로 정보를 수집하고 일정 수준 생육이 진행된 중·후기에는 개체의 크기, 무게 등을 추가로 조사하고 있음.
- 생육 지표 조사 시 표준화된 촬영 방식에 따라 작물을 촬영하고 있으며, 조사구역 내 병해충 발병 여부(빈도)와 드론을 활용한 시기별 영상, 토양 수분 정보 등을 추가로 수집하여 모형개발을 위한 기초자료를 수집하고 있음.

〈표 2-4〉 품목별 생육실측조사 항목

품목명	공통	생육 초기	생육 후기(추가)	기타 정보
양파	경영체 정보, 품목, 품종, 종자명, 관수유무, 조사시작일, 정식(파종)일, 수확(예정)일, 필지정보(이랑폭, 두둑폭, 재식주수, 결주수 등)	초장, 엽수, 엽초장, 엽초경	구직경, 1주 무게, 경영 무게, 생구 무게 등	작물 사진, 병해충 발병여부, 드론영상 (5, 7, 10m), 토양 수분 정보 등
마늘			주당 수확과중	
건고추		초장, 주경장, 경경, 분지수, 착과수, 미숙과수, 낙과수	구고, 구폭, 구중, 주중	
배추		초장, 엽수, 엽폭	근장, 근중	
무		초장, 엽수	근장, 근중	

자료: 농업관측센터 내부자료를 바탕으로 저자 작성.

〈그림 2-3〉 배추 생육실측조사 방법(예시)



자료: 농업관측센터 내부자료.

○ 실측조사 대상 품목의 최종 생산량 예측을 위해 건고추를 제외한 양파, 마늘, 배추, 무 4개 품목을 대상으로 수확예정일 5일 전 생산량 실측조사를 수행하고 있음.

- 재배면적 실측조사 시 표본농가의 대표필지 중 1개의 필지 내 3개의 생산량 조사구역을 설정하고 수확이 임박한 시점에 현장을 방문하여 조사구역 (3.3㎡) 내 생육 중인 해당 품목 전체를 수확하여 실제로 생산량을 조사함.
- 양파, 마늘, 배추, 무와 달리 건고추는 생육 기간 동안 여러 번 수확하는 품목의 특성을 반영하여 최종 생산량을 조사하지 않음. 생산량 조사는 정식 (파종) 이후 1회 수확하는 품목을 중심으로 수행하고 있음.

〈표 2-5〉 품목별 생산량 실측조사 항목

품목명	공통	조사 항목	기타 정보
양파	경영체 정보, 조사일시, 품목, 품종, 이랑폭, 두둑폭, 재식주 수(정상, 비정상, 결주) 등	구직경, 생구중, 특이사항	피해 상황, 조사 사진 등
마늘		생구중, 개체별 특이사항	
배추			
무			

자료: 농업관측센터 내부자료를 바탕으로 저자 작성.

〈그림 2-4〉 품목별 생산량 실측조사 방법(예시)



자료: 농업관측센터 내부자료.

1.3. 항공촬영(드론) 방식 활용 재배 및 출하면적 조사 사업

○ 농업관측센터는 관측고도화 사업의 일환으로 실측조사와 더불어 항공촬영(드론) 방식을 활용한 광범위한 지역의 재배 및 출하면적 조사 사업을 본격적으로 도입하며 원격탐사의 기반을 마련하였음.

- 그동안 관측사업을 위한 조사방식은 농업경영체 중심의 표본조사 중심으로 데이터를 수집하였음. 이러한 방식은 짧은 시간에 전국의 표본을 대상으로 요구하는 데이터를 수집하는 데 장점이 있으나, 표본 설계 및 조사 응답 결과의 오류 시 비현실적 데이터를 생산할 가능성이 있음.
- 인력을 활용한 특정 주산지의 필지 전수조사를 실시할 경우 다수의 인력과 자원이 소요되어 비효율적인 방식으로 평가받고 있으며, 필지 진입 등이 어려워 전수조사에는 한계가 있는 것으로 평가됨. 이러한 한계점을 극복하기 위해 원격탐사 방식의 조사 체계 도입이 필요하였고 이동의 제한이 없고 신속한 조사가 가능한 장점이 있는 드론을 활용하기 시작함.

○ 농업관측센터는 2015년부터 드론을 활용하여 재배 및 출하면적을 조사하기 위해 가능성을 검토하기 시작함. 시범사업 후 2020년부터 농업관측사업의 특수성을 반영한 항공촬영 방식을 도입하여 운영하고 있음.

- 농업관측센터는 2015년 회전익 드론 1기를 처음 도입하였고 정량적인 조사보다는 필지 작황의 정성적인 판단을 위해 필지별 동영상 촬영하여 정보를 수집하였음. 이러한 방식은 주관적인 판단이 개입될 가능성이 높아 체계적이고 과학적인 조사방식의 도입 필요성이 제기됨.
- 2017년부터 고정익 항공기를 활용한 정사영상을 촬영하고 이를 토대로 배추 재배면적을 분석하는 방식을 도입함. 해당 분야의 전문적인 지식이 부족함에 따라 농촌진흥청과 협업하여 사업을 추진하였으며, 고랭지배추 3

개 봉우리(안반데기, 매봉산, 귀네미)를 대상으로 촬영을 진행하였음.

- 해당 조사방식은 특정 지역의 전체 필지를 촬영하고 그 영상을 분석하여 필지 중심의 전수조사 방식 데이터를 수집한다는 데 의의가 있음. 그러나 고고도에서 촬영이 이루어지는 만큼 영상 해상도가 낮고 영상을 가공하는 데 많은 자원이 필요하였음. 또한 영상을 기준으로 필지를 구획하고 분류하는 데 있어 낮은 해상도로 인해 정확도가 낮다는 단점이 있음.

○ 드론 기체와 센서(카메라) 사양의 발달 및 촬영 기술의 발달, 농림수산물교육문화진흥원에서 서비스하고 있는 농경지 전자지도(팜맵) 활용을 검토하면서 농업관측사업의 특수성을 반영한 촬영 방식을 도입할 수 있었음.

- 농업관측사업의 특성상 짧은 시간 내 광범위한 지역의 자료를 수집하여야 하며, 경지정리가 잘 되어 있지 않은 소규모 필지 위주로 구성된 채소류 주산지의 특성을 고려하였을 때 고해상도 항공영상을 활용한 품목 분류가 필요함. 또한 1개의 필지 내 다양한 품목을 혼작하는 경우가 있어 정확한 면적 산출을 위한 광각 영상이 필요함.

○ 과거 사업을 추진하며 도출된 한계점을 극복하고 드론을 활용한 효율적인 조사 방식을 설계 후 주요 채소류 주산지를 중심으로 재배 및 출하면적 조사사업을 본격적으로 추진하였음.

- 농업관측센터는 주요 채소류 주산지 53개 읍면 전체 농경지를 대상으로 약 10만 5천 ha(70만 개 필지)를 5개 권역으로 구분하여 촬영하고 있음.
- 농업관측센터는 드론을 활용한 사업 운영을 위해 기준지도로 농림수산물교육문화정보원에서 공개하는 농경지 전자지도(팜맵)를 활용하고 있음. 해당 지도는 전국 농경지를 대상으로 실제 경계를 표시하고 면적을 산출하고 있으며, 위성영상을 활용하여 매년 갱신하고 있음. 해당 지도에는 필지

별 관리를 위한 ID, 주소, 지목, 면적 등이 수록되어있어 과거 필지별 농경지를 구획하여야 했던 업무체계의 개선 및 사전 촬영 계획 수립이 가능하게 되었음. 또한 고해상도 낱장 사진으로 필지별 품목을 분류할 수 있는 체계 도입에 크게 기여한 것으로 판단하고 있음.

- 고해상도 영상을 기초로 배추, 무, 당근, 양배추, 마늘, 양파, 대파, 감자, 사과, 감귤, 기타 등 관측 품목 중심으로 재배면적을 분류하고 있으며, 효율적인 사업 운영을 위해 자가 소비용으로 판단되는 소규모 필지(200㎡ 미만)는 분류에서 제외하였음.

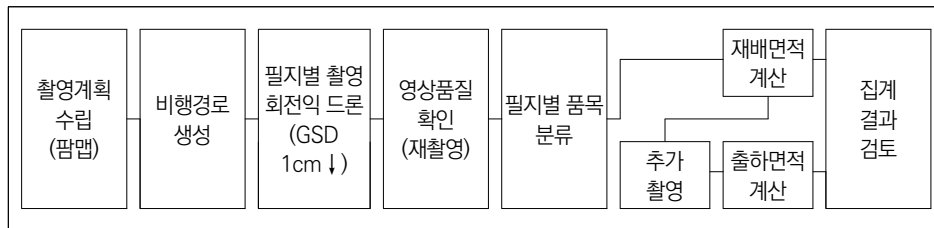
〈표 2-6〉 농업관측센터 항공촬영 사업 현황

구분	촬영 면적(ha)			읍면(개)		주요 품목
	2023년	2024년	증감률	2023년	2024년	
여름채소(1권역)	8,081	8,081	-	5	5	배추, 무, 대파
여름채소(2권역)	6,428	6,428	-	12	12	
겨울채소(1권역)	28,945	28,945	-	11	11	배추, 무, 마늘, 양파, 대파, 당근
겨울채소(2권역)	25,748	25,748	-	17	17	
겨울채소(제주)	55,323	35,907	-35.1%	12	8	
합계	124,525	105,109	-15.6%	57	53	-

자료: 저자 작성.

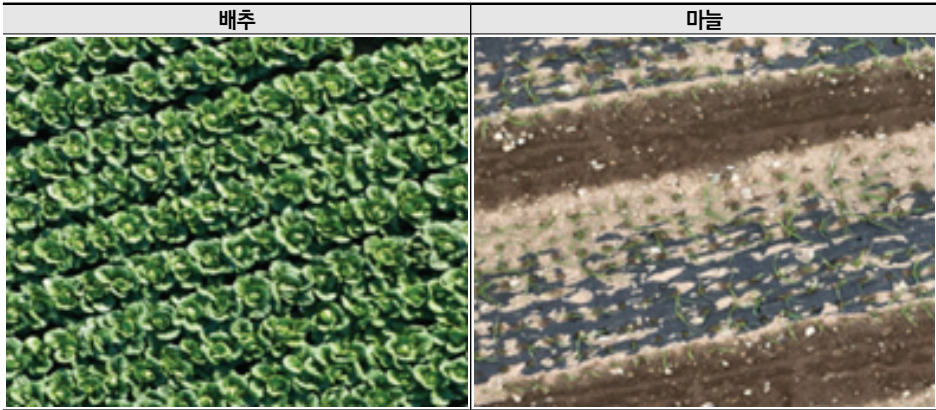
- 농업관측센터는 사업 특성상 생육 초기 개체 크기가 작은 작물을 정확하게 분류할 필요가 있음. 이를 위해 회전익 드론으로 고해상도 영상을 촬영하고 인력을 활용하여 품목을 분류하고 면적을 집계하고 있음.

〈그림 2-5〉 농업관측센터 항공촬영 사업 추진 체계



자료: 저자 작성.

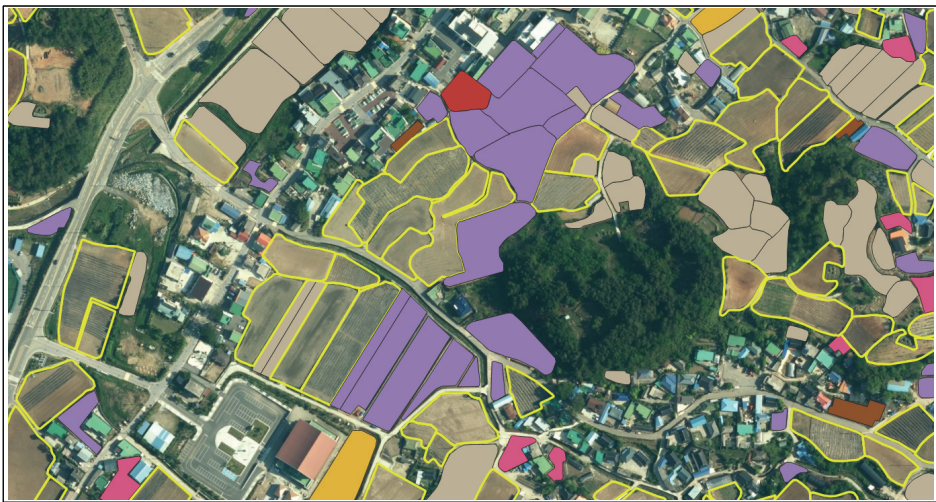
<그림 2-6> 항공촬영 영상(예시)



자료: 농업관측센터 내부자료.

- 항공촬영 사업을 통해 수집된 데이터는 필지별 항공촬영 영상, 필지별 품목명, 품목별 면적(재배, 출하), 혼작 여부 등 영상을 통해 확인할 수 있는 다양한 정보를 수록하고 있으며, 이를 GIS 기반의 shp 파일에 메타데이터를 저장하고 있음.

<그림 2-7> 품목 재배분포도(고흥군 금산면 마늘, 양파 예시)



자료: 저자 작성.

○ 항공촬영 사업을 통해 수집된 데이터는 팜맵의 ID를 기준으로 관리되고 있으며, 주요 채소류 실측조사 사업 시 팜맵 ID를 추가로 수집하고 있음.

2. 유관기관 데이터 수집 현황

2.1. 농촌진흥청

○ 농촌진흥청은 작물의 생육 파악 및 예측모델 개발을 위해 실측조사를 실시하고 그 결과를 발표하고 있으며, 원격탐사 연구를 수행하고 있음.

〈표 2-7〉 농촌진흥청 농작물 생육조사 현황

품목	필지수(개)	조사횟수(회)	조사항목
벼	690	8	초장, 경수, 출수기, 수수, 총입수, 완전입 수, 완전입 비율
보리	204	8	초장, 경수, 주간엽 수, 고사주율, 고엽률, 고사경율, 출수기, 간장, 수수, 성숙기, 1수립 수, 1,000립중, 10a당 수량 등
콩		6	초장, 잎 수, 개체 수, 가지 수, 개화시, 꽃 핏 때, 꼬투리 수, 알 수
마늘, 양파	671	8	3.3㎡당 주수, 초장, 엽 수, 생육재생기, 결주율, 구직경, 10a 당 예상 수량
배추		4(가을배추)	3.3㎡당 주 수, 초장, 엽 수, 병충해, 10a당 수량 등
무		4(가을배추)	3.3㎡당 주 수, 초장, 엽 수, 병충해, 10a당 수량 등
고추		7	3.3㎡당 주수, 초장, 착과수, 주당 착과수, 수확량, 10a당 예상 수량 등
과수	555	3	개화일, 만개일, 꽃수, 착과량, 10a당 착과수, 예상 생산량 등

자료: 농촌진흥청(2014), 농작물 생육조사 표준매뉴얼.

○ 벼, 마늘, 양파, 고추 등 주요 품목을 대상으로 생육상황을 조사하고 있으며 그 결과를 공표하고 있음.

- 농촌진흥청(2014)에 따르면 농작물 생육조사가 가능한 품목은 벼, 보리, 콩, 마늘, 양파, 배추, 무, 고추, 과수 등이며 2,120개 조사구역을 대상으로 전국 시군농업기술센터 작목별 담당자가 생육 시기별로 조사하고 있음.

〈표 2-8〉 농촌진흥청 원격탐사 및 빅데이터 활용 주요 사업 내용


연도	주요 사업 내용
2014	<ul style="list-style-type: none"> ○ 항공영상을 활용한 고랭지배추 면적 및 생육 분석기술 개발 - 고랭지배추 주산지 재배면적 산정 및 정식시기 분포지도 작성: 3개소 286ha(안반데기, 매봉산, 귀네미) - 분광 특성을 이용한 재배 작물 분류 및 면적 산정: 배추, 감자, 양배추 등 - 센서 기반의 고랭지배추 생육 추정 및 영상 기반의 병 발생 분석
2015	<ul style="list-style-type: none"> ○ 위성·항공 영상 기반 고랭지배추 작황 평가 기술 개발 - 배추 정식상황 파악을 위한 항공영상 제공: 3개소 286ha(안반데기, 매봉산, 귀네미) - 고랭지배추 주산지(안반덕 등 3개 지역) 재배현황 맵 제공(농식품부, KREI, 농진청) - 영상분석을 통한 정식시기, 생육 이상 판별 모델 구축, 정식시기, 생육 이상 발생 현황별 지도 제작 - 시범 연구: 원격탐사 활용 5대 채소 주산지 작황 변동 추정 기술 개발
2016	<ul style="list-style-type: none"> ○ 농업환경정보 통합관리 및 서비스 체계 구축 - 원격탐사 기술 활용 재배현황 지도 작성 및 작황 추정 모형 고도화: 고랭지배추(안반덕 등 286ha, 가을배추무(고창 65.7ha), 마늘양파(합천 등 300ha) - 항공·위성 영상 기반 채소 주산지 작황 정보 공유 및 협력체계 구축 및 5대 채소 재배현황맵 배포 - 고랭지배추 주산지(안반덕 등 3개 지역) 재배현황 맵 제공(농식품부, KREI, 농진청) ○ 국내외 농업 생산 환경 모니터링 및 예측 고도화 - 위성영상 활용 국내외 농경지 이용 현황 및 작황 분석: 주요 곡물 시대(중국, 미국) 작물 구분 및 수량 추정, 북한 벼 생산량 추정 - 원격탐사를 기반으로 한 동계 맥류, 조사료 재배현황 및 생육진단 추정 연구: 위성, 항공영상을 활용한 동계작물의 주산지 작목 구분 및 수량 추정 기술 개발
2017	<ul style="list-style-type: none"> ○ 마늘 생육 및 단수 예측 모형 개발 - 원격탐사 자료와 기상 자료 활용 마늘 생육인자 추정: 드론 활용 주산지별 생육 추정 모형 개발 - 필지 단위 생육 변화 적용 이상 판별 기준 설정: 식생지수 변화 적용 연차별 필지 단위 생육 이상 현황도 작성 - 마늘 생육단계별 작황 정보 생산: GIS 기법 활용 마늘 주산단지 생육인자 분포도 작성
2018	<ul style="list-style-type: none"> ○ 원격탐사 기반 동계작물 판별 및 생육 평가 기술 개발 - 작물 분류를 위한 계층 분류 체계 구축 및 시군단위 동계작물 분포도 작성: 위성영상 기반 분포도 작성 기술 개발 - 무인기를 활용한 동계작물 생육 평가 기술 개발: 맥류 생육 단계별 작황 정보 생산, 사료가치 평가 모델 개발 등
2019	<ul style="list-style-type: none"> ○ 채소 수급 안정 목적의 드론 활용 맞춤형 정보 생산 및 협업체계 구축 - 채소 재배현황 변동 추정 및 의사결정 지원 기술 개발: 주산지 맞춤형 작황 정보 생산 및 제공 - 맞춤형 실시간 재배현황 정보 제공 시스템 운영: 웹 GIS 기반 정보제공 시스템 구축, 관측월보 수록
2021	<ul style="list-style-type: none"> ○ 인공지능 기반 실시간 작물 병해 진단 기술 고도화 - 인공지능 기반 토마토, 딸기 병해 진단 고도화 연구 추진
2022	<ul style="list-style-type: none"> ○ 농업환경정보 생산을 위한 위성영상 공동활용 기술 개발 - 위성영상 대기보정 및 범용활용 알고리즘 적용 다중분광 위성 공동활용 분광밴드조정계수 평가 - 위성·드론 공동활용을 통한 농경지 생육 모니터링(위성영상 활용 보 도열병 탐지 및 분석)
2023	<ul style="list-style-type: none"> ○ 농경지 모니터링을 위한 광학·레이더 위성 영상 융합 활용기반 구축 - 레이더 영상(후방산란계수) 기반 농경지 식생지수 추정 - 식생지수 기반 생육추정 활용 체계 구축(직물제 이행점검) - 위성영상 기반 시·군단위 벼, 콩 생육 모니터링(김제시 부량면)

주: 국승용 외(2022: 39), 드론을 활용한 농업생산조사 방법 연구의 내용을 추가 보완함.

자료: 농촌진흥청 국립농업과학원(각 연도), 농업과학연구사업연보.

- 원격탐사와 관련하여 재배현황 및 식생지수 산출을 위한 정사영상을 제작하는 사업을 실시하였으며, 작물 식별을 위한 방법론적 연구 중심으로 사업이 이루어지고 있음. 주요 품목의 생육 시기와 이상 여부, 작황 분석, 단수예측모형 등 예측 중심의 연구가 활발하게 이루어지고 있음.
- 농촌진흥청은 2025년 하반기 차세대중형위성 4호(농림위성) 발사 후 농업분야에 위성영상을 활용한 데이터 수집 및 분석체계를 구축하는 사업을 수행 중에 있으며 2026년 현업화를 목표로 사업을 수행 중에 있음.
 - 농림위성은 발사 준비단계로 발사 이후 정보 활용을 위해 외국 위성영상을 활용한 재배면적, 생육 추정 등의 연구를 수행하고 있음.
 - 위성영상에서 추출할 수 있는 NDVI값을 활용, 식생지수를 산출하는 방식이 주를 이루고 있으며, 벼, 동계 맥류, 마늘, 양파, 배추 재배면적, 벼 수량 예측, 북한 벼 작황 모니터링 등 시범사업을 실시함.
 - 차세대중형위성 4호(농림위성)는 공간해상도 5m급으로 기존 위성 대비 고해상도 영상촬영이 가능하며, 전국을 3일 내 촬영할 수 있음.
 - 농림위성 발사 후 농산물 수급 안정, 식량정책 자급률 제고, 농업수자원 관리, 기후변화 감시/대응, 북한농업 통계, 해외농업(식량안보), 국제개발협력(ODA), 디지털 농업 분야에서 활용 예정임.

〈표 2-9〉 차세대중형위성 4호(농림위성) 주요 특징

구분	주요 내용	
위성 유형	중형위성(500kg) ※설계수명 5년	
운영 고도	상공 900km	
촬영 사양	<ul style="list-style-type: none"> ○ 공간해상도(GSD) 5m급(외국 위성 6.5~30m) ○ 관측폭 120km(전국 3일 촬영) ※ 한반도 전체 5회 촬영 시 영상 취득(한반도 가로길이 540km) ○ 분광밴드 5개(청/녹/적/적색경계/근적외) 	

자료: 농촌진흥청(2024), “하늘에서 농업을 보다”, 농업위성센터 홍보자료.

〈표 2-10〉 농림위성 활용 계획

농산물 수급안정	식량정책 자금률	농업 수자원 관리	기후변화 감시/대응
채소 등 수급 민감 작물 재배면적, 생육이상 관측, 수량파악 등	벼, 콩 등 식량작물 재배면적, 생육 모니터링, 생산량 관측 등	농경지 밭 토양 수분 관측, 농업용 물자원 현황, 저수지 수표면적 변화 등	농경지 토지이용, 피복변화, 주요 작물 재배지, 작황 변동, 농업재해 평가 등
북한농업 통계	해외농업(식량안보)	국제개발협력(ODA)	디지털 농업
북한 농경지 재배면적, 관개 현황, 주요 발작물 생육 현황 등 파악	북/남반구 주요 곡물 생산국 작황정보 파악	아시아, 아프리카, 중남미 등 주요 협력국 작황 정보	작물구분도, 수량분포, 재배시기, 생육현황 분석, 순일차 생산량 등

자료: 농촌진흥청(2024), “하늘에서 농업을 보다”, 농업위성센터 홍보자료.

2.2. 제주특별자치도

○ 지방자치단체 중 제주특별자치도는 주요 동계작물 중심으로 재배면적 조사를 위해 한국농촌경제연구원 농업관측센터와 공동으로 드론을 활용한 조사 사업을 실시함.

- 농림수산식품교육문화진흥원에서 배포하고 있는 농경지 전자지도(팜맵) 대상 28만 개 필지를 대상으로 전수조사를 실시였으며, 주요 동계작물의 정식 및 파종이 완료된 11월 상순부터 중순까지 촬영을 수행하고 있음.
- 2021~2024년까지 총 4회에 걸쳐 사업이 진행되었으며, 2021~2023년까지는 한국농촌경제연구원 농업관측센터와 협업으로 공동사업을 추진하였으나, 2024년부터는 각 기관의 목적에 맞는 사업 추진을 위해 별도의 사업을 추진하고 있음.
- 제주특별자치도에서 생산되는 주요 동계작물인 마늘, 양파, 무, 당근, 양배추, 감귤, 브로콜리 등 15개 품목 분류기준에 따른 재배면적을 조사하였으며, 필지별 품목 분류를 위한 고해상도 영상과 필지 운영 현황을 파악하기 위한 광각영상을 수집하였음.
- 수집된 데이터 결과는 품목별 재배면적 행정자료와 수급 조절을 위한 기초 자료로 활용하고 있음.

2.3. 한국지능정보사회진흥원(NIA)

- 한국지능정보사회진흥원(NIA)에서는 2020년부터 인공지능 학습용 데이터 구축, 빅데이터 수집 사업 등을 통해 농업분야의 다양한 데이터를 수집하는 사업을 지원하고 있음.

- 해당 사업을 통해 생산된 농업관련 데이터는 AI Hub(aihub.or.kr) 사이트를 통해 국민을 대상으로 서비스하고 있으며, 농축수산 관련 데이터는 총 83건을 제공하고 있음.
 - 이 중 농업관련 데이터는 28건, 축산물 관련 데이터는 12건이 구축되어있으며, 이미지, 텍스트, 비디오, 오디오, 센서값 등 다양한 자료를 제공하고 있음. 구축된 모든 데이터는 이미지 정보를 담고 있는 것으로 파악됨.
 - 본 연구와 관련된 인공지능 학습데이터는 7건이며 데이터의 주요 특징은 아래 <표 2-11>과 같음.
 - 인공지능 학습용 데이터 수집 사업 초기에는 작물의 이미지 수집을 위해 정사영상을 활용한 방식을 활용하였음. 인공지능학습을 위해 작물의 생육 정보가 필요함을 인지하여 2022년부터 일부 품목의 경우 지상조사를 동반하여 생육 이상 지역을 방문하여 조사하고 있으며, 수확량 데이터를 동반하여 수집하고 있음.

〈표 2-11〉 한국지능정보사회진흥원 인공지능 학습용 데이터 수집 현황

데이터명	구축연도	데이터 수집 현황	형태
노지작물(배추 등) 착상 데이터	2022	○ 수급 민감 품목(배추, 무)의 신속한 재배면적 산출 등을 위한 드론 영상 기반 주요 채소 작물 구분용 학습용 데이터 ○ 구축된 데이터 활용 수급 조절을 위한 생산량 예측 인공지능 개발용 공공데이터	이미지
벼 생육이상 인식 데이터	2022	○ 벼 경작지에서 드론으로 촬영한 다분광 이미지 상에서 벼의 생육 이상 범위를 탐지하기 위한 학습용 데이터	이미지
제주 주요작물 자동탐지 데이터	2021	○ 제주특별자치도 주요 농작물 15종(식량작물 5종, 월동작물 10종)에 대한 재배지역 학습데이터 구축 및 재배현황 파악, 다양한 농업정책 활용성 제공	이미지
정밀농업 노지작물 통합 데이터	2021	○ 정밀농업 필수 노지작물 생육 데이터 확보를 목적으로, 작물 10종에 대해 지상 촬영 2D RGB 이미지 데이터, 다분광 적외선 촬영 온도값, 드론 촬영 데이터, 토양 및 환경 센서 데이터 등을 융합한 학습 데이터셋	이미지
제주 월동작물 자동탐지 드론 영상	2020	○ 제주 월동작물 자동 탐지를 위한 드론촬영 이미지 데이터	이미지
주요 농작물 생육 이미지	2020	○ 10종의 주요 농작물 품질 및 수확량 예측용 생육 단계별 드론 초분광 이미지 데이터	이미지
드론 농경작지 촬영 영상	2020	○ 5종의 농지 경작 주요 작물에 대한 드론 초분광 카메라 농지 정보 이미지 데이터	이미지

자료: 한국지능정보사회진흥원 AI Hub(aihub.or.kr), 검색일: 2024. 12. 8.

3. 시사점

○ 농업관측센터에서 생성하고 있는 재배면적 및 생육실측조사 데이터와 항공 촬영을 통한 재배면적 및 출하면적 이미지 데이터는 보다 정확한 농업관측정보를 제공하기 위한 분석용 데이터로 활용되고 있음.

- 재배면적 및 생육실측조사는 그동안 표본농가 전화조사와 모니터 설문조사에 의존하던 농업관측정보 생성의 비표본오차를 개선하는 데 크게 기여하였으며, 항공촬영 사업은 수치로만 제공하던 품목별 주산지 재배면적 및 출하면적 정보를 시각화하여 제공하게 되었다는 데 의미가 있음.
- 그러나 각 사업이 서로 연관되어 진행되지 못하고 있고 데이터도 통일된 형식 없이 서로 다른 저장공간에 보관되어 있어 이를 개선할 필요가 있음.

- 한편 한국지능정보사회진흥원에서 제공하고 있는 인공지능 학습용 데이터는 과제별로 인공지능 데이터 구축·활용 가이드라인을 제공하고 있음. 이 가이드라인은 데이터 명세 정보, 데이터 구축 가이드로 구분하여 작성되었음.
 - 데이터 명세 정보는 데이터명, 활용 분야, 데이터 요약, 데이터 출처, 데이터 포맷, 어노테이션(Annotation) 포맷, 데이터 구성, 데이터 통계, 원시 데이터 특징, 기타 정보 등이 포함되어 있음.
 - 데이터 구축 가이드에는 데이터 구축 개요, 문제 정의, 획득·정제, 어노테이션/라벨링, 검수, 활용 등이 포함되어 있음.

- 농업관측센터의 생육실측조사와 재배면적 항공촬영 사업은 사업별 특성을 반영한 매뉴얼을 구축하여 데이터를 수집하고 있으나, 드론을 활용한 생육상황 항공촬영은 필지의 생육상황 변화를 파악하는 참고자료에 불과하여 촬영 대상이 구체적으로 명시되어 있지 않음.
 - 따라서 농업관측센터에서 생성하는 정보를 활용하여 인공지능을 활용하기 위해서라도 학습용 데이터 구축을 위한 생육실측조사와 생육상황 항공촬영의 연계 작업이 필요함.

3

생육·영상 자료 수집 및 전처리

1. 생육실측 자료 수집 및 전처리

1.1. 생육실측 자료의 생성

○ KREI의 마늘, 배추 생육실측조사는 조사자가 포전을 직접 방문해 농촌진흥청 조사 방법을 준용하여 개체별 실측을 진행하고 있으며, 주요 조사 항목으로는 조사구역 농가 및 필지 현황(필지정보), 생육 및 수량 등이 있음.

- 생육 및 병해충 조사는 모든 조사 회차(마늘의 경우 1~11차, 배추의 경우 1~4차)에서 고정 개체 5개, 수확 개체 5개를 정기적으로 조사하며, 특정 회차(마늘 5차, 배추 3차) 이후 직접 개체를 수확하여 조사가 이루어짐.
- 필지 내에 3.3m²(1평)씩 3개의 구역을 선정하여 '반복'이라 칭하며, 각 필지에 3개의 반복, 각 반복 구역 내에 개체(마늘의 경우 최대 10개, 배추의 경우 최대 6개)에 대한 조사가 이루어짐.

- 필지정보에는 품목, 품종, 재배면적, 관수시설 유무, 재식주수, 정식, 수확 예정일 등이 포함됨.
- 생육 및 수량 조사는 품목별 특성을 반영하여 조사 내용이 다르게 구성되어 있음. 마늘은 생엽장, 엽초장, 엽초경, 엽수, 구 직경, 1주 무게, 경엽 무게 등을 조사하고, 배추는 초장, 엽수, 엽폭, 외엽수, 내엽수, 구폭, 구고, 주중, 구중 등을 조사함.

〈표 3-1〉 마늘 실측조사 회차별 조사항목 상세 내용

구분		조사 회차	조사 항목
재식상태	마늘	1차	이랑폭, 3.3㎡당 주수, 결주율 등
	배추		
생육상태	마늘	1~11차	초장, 엽수, 엽초장, 엽초경
	배추	1~4차	초장, 엽수, 엽폭
병해충	마늘	1~11차	잎마름병, 노균병, 흑색썩음균핵병, 고자리파리 등
	배추	1~4차	무름병, 뿌리혹병, 검은무늬병, 미결구 등
수확조사	마늘	5차~	구 비대기, 추대기, 구 직경, 1주 무게, 경엽 무게, 생구 무게
	배추	3차~	외엽수, 내엽수, 구폭, 구고, 주중, 구중
경영정보	마늘	수시	정식일, 수확 예정일, 관수시설 유무, 총 수확량 등
	배추		

자료: 이형용 외(2023: 32), 마늘·양파 구중 예측모형 개발 연구의 내용을 토대로 재구성.

○ 창녕군 마늘 실측 생육 자료는 KREI 농업관측센터의 마늘 전체 생육 실측 표본농가(100호) 중 8호를 대상으로 조사가 진행되고 있음.

- 창녕지역 마늘 생육 실측 표본농가는 2020~2022년까지 6호였으나, 표본을 재구축하면서 2023년 이후 8호를 대상으로 조사가 이뤄지고 있음. 조사 시기는 2월 중순, 3월 상·중순, 4월 상·중·하순, 5월 상·중·하순이며, 조사 횟수는 해마다 다르지만 평균 9회 정도로 진행되고 있음.

〈표 3-2〉 창녕지역 마늘 실측조사 시기 및 횟수

연도	2월	3월	4월	5월	6월	조사 횟수	표본농가 수
2020	19~20일	9일 23일	6~7일 16~20일 27일	6일 13일 22일	2일	10회	6
2021	17~18일	4일 17~18일	5~6일 14~15일 22~23일	3일 12~13일 25~26일	1일	10회	6
2022	17일	3일 21일	5~6일 18일 25~26일	9~10일 16~17일 25~26일	-	9회	6
2023	14일	2~3일 14~15일 28~30일	11~12일 18~20일	2~3일 9~10일 23~24일	-	9회	8
2024	13일 27일	12~13일 26~27일	8일 17~18일 26~30일	8~9일 22~23일	-	9회	8
평균	중순	상순, 중순	상순, 중순, 하순	상순, 중순, 하순	-	9회	7

자료: 저자 작성.

1.2. 생육실측 자료의 현황

○ 본 연구에 활용된 창녕군 대서종 마늘 실측 생육 항목별 평균값은 〈부표 1-1〉에 정리하였으며, 이 중 종속변수인 생구 무게 실측조사 평균값은 〈표 3-3〉과 같음.

- 2020~2021년에는 수확이 6월 초까지 진행되어 10차까지 조사 결과 값이 있는 반면, 2022~2024년에는 수확이 5월 하순에 마무리되어 9차까지 조사 결과 값이 있음.
- 생구 무게는 5차부터 조사가 시작되고 조사 회차가 진행될수록 평균값이 증가하는 모습을 보이고 있으나, 일부 연도에서는 조사 회차가 진행될수록 실측값이 하락한 모습을 보이고 있음. 이는 수확 후반기가 될수록 표본농가 중 수확이 완료된 포전은 실측조사에서 제외되기 때문임.

〈표 3-3〉 창녕지역 조사 회차별 마늘 생구 무게 실측조사 평균값과 표본 수

구분		조사 회차					
		5차	6차	7차	8차	9차	10차
2020년	생구 무게	44.3	56.9	74.6	91.9	95.1	92.7
	표본 수	6	6	6	6	4	1
2021년	생구 무게	35.8	43.0	65.0	84.9	101.0	80.4
	표본 수	6	6	6	6	6	1
2022년	생구 무게	27.1	40.3	63.8	66.7	85.3	-
	표본 수	6	6	6	6	3	-
2023년	생구 무게	24.7	35.7	57.7	67.6	63.0	-
	표본 수	8	8	8	8	6	-
2024년	생구 무게	15.7	19.4	33.0	50.9	68.6	-
	표본 수	8	8	8	8	4	-

자료: 저자 작성.

○ 평창군 배추 실측 생육 표본 수는 해마다 차이가 있으나, KREI 농업관측센터의 고랭지배추 표본농가(5년 평균 132호) 중 40호 내외를 대상으로 조사가 진행되고 있음.

- 평창지역 고랭지배추 생육 실측 표본농가는 2021년(26호)을 제외하고 40호 내외에서 유지되고 있음.
- 고랭지배추 생육실측조사는 정식시기에 따라 최대 5번(4월 정식~8월 정식)으로 나뉘어지며, 각 정식 시기별 1~4회차 조사가 진행됨. 1기작 준고랭지 배추를 중심으로 5월 중순에 실측조사를 시작하여 2기작 준고랭지 배추가 수확되는 10월 하순까지 조사가 이루어짐.

○ 본 연구에 활용된 평창군 고랭지배추 실측 생육 항목별 평균값은 〈부표 1-2〉에 정리하였으며, 종속변수인 구중 실측조사 평균값은 〈표 3-4〉와 같음.

- 2020~2022년까지 구중은 3차에만 조사가 이루어졌으나, 여름철 기상여건이 극심하여 2023년 이후 4차까지 구중 조사가 진행되었음.

〈표 3-4〉 평창지역 고랭지배추 정식시기별 실측조사 시기

연도		4월 정식	5월 정식	6월 정식	7월 정식	8월 정식	표본 수
2020	1차	-	6. 9., 6. 17.	6. 22.~24., 6. 29., 7. 2., 7. 7., 7. 15.~16., 7. 20.~21.	7. 27.~30., 8. 5., 8. 11., 8. 19.~20.	-	43
	2차	-	6. 9.~10., 6. 22.~23., 6. 29.	7. 15.~16., 7. 20.~23., 7. 28.~30., 8. 4.~5., 8. 10.~12.	8. 10.~13., 8. 18.~20., 8. 24.~26., 8. 31., 9. 2., 9. 7.~9.	-	
	3차	-	6. 29.~7. 1., 7. 14.~15., 7. 21.	7. 28.~30., 8. 4.~5., 8. 11.~13., 8. 18.~20., 8. 24.~26., 9. 1.~2., 9. 9., 9. 15., 9. 21.	8. 31.~9. 1., 9. 15.~17., 9. 21.~22., 10. 6.~7., 10. 28.	-	
2021	1차	5. 17.~18.	5. 31.~6. 1., 6. 8.	6. 22., 7. 13., 7. 22.	-	-	26
	2차	6. 8.~9.	6. 16., 6. 22., 6. 30.	7. 14., 7. 27., 8. 11.	-	-	
	3차	6. 22.~23., 6. 29.	7. 1., 7. 6.~7., 7. 13.~14.	8. 11., 8. 18., 8. 30.	-	10. 12.	
2022	1차	5. 17.~18.	5. 23., 6. 2., 6. 13., 6. 20.	6. 20., 6. 29., 7. 6., 7. 13.	7. 26.~27., 8. 3.	8. 31.~9. 1.	44
	2차	5. 30., 6. 7.	6. 13., 6. 20., 7. 5., 7. 13.	7. 14., 7. 19.~20., 7. 26.~27., 8. 3.~4., 8. 9.	8. 16.~17., 8. 23.	9. 20.	
	3차	6. 20., 6. 29.	7. 5., 7. 12., 7. 26., 8. 2.~3.	8. 3., 8. 8.~10., 8. 16., 8. 23.~24.	8. 30.~31., 9. 15.	10. 13.	
2023	1차	5. 16.	5. 23., 6. 1., 6. 13.	6. 29., 7. 4., 7. 11.~13., 7. 19.	7. 19., 7. 25.	8. 22.~23., 8. 31.	39
	2차	6. 8.	6. 12.~13., 6. 20.~21., 7. 5.	7. 19.~20., 7. 25.~26., 8. 1.~2., 8. 8.	8. 9., 8. 16., 8. 23.	9. 11.~12., 9. 19.	
	3차	6. 20.	6. 27.~28., 7. 4.~5., 7. 17.~19.	8. 1.~2., 8. 8.~10., 8. 16.~17., 8. 22.	8. 23., 8. 30., 9. 5.	9. 25.~26., 10. 5.	
	4차	-	7. 4.~6., 7. 12., 7. 24.	8. 9., 8. 17., 8. 22.~23., 8. 31.	8. 31., 9. 5., 9. 11., 9. 19.	10. 10.	
2024	1차	5. 28.~29.	5. 29., 6. 12.~13.	6. 26., 7. 3.~4., 7. 9., 7. 17.	7. 29.~31., 8. 12.	8. 27., 9. 4., 9. 10.	39
	2차	6. 4., 6. 19.	6. 19., 7. 1., 7. 9.	7. 17., 7. 23., 7. 29.~30.	8. 19., 9. 3.	9. 17.~19., 9. 23., 9. 30.~10. 1.	
	3차	6. 19., 6. 26.	7. 2.~4., 7. 15.~17., 7. 23.	7. 30.~8. 1., 8. 5.~7., 8. 12.~14.	9. 2., 9. 19.	9. 30., 10. 7.~8., 10. 15.	
	4차	6. 27.	7. 4., 7. 21.~22., 7. 30.	8. 6.~9., 8. 14., 8. 19.	9. 9.~10., 9. 23.	9. 30., 10. 7.~8., 10. 15.~16., 10. 23.	

자료: 저자 작성.

〈표 3-5〉 평창지역 조사 회차별 고랭지배추 구중 실측조사 평균값과 표본 수

구분		조사 회차	
		3차	4차
2020년	구중	1,829.7	-
	표본 수	21	-
2021년	구중	2,194.9	-
	표본 수	24	-
2022년	구중	2,032.6	-
	표본 수	29	-
2023년	구중	1,756.8	1,748.5
	표본 수	31	19
2024년	구중	1,483.7	1,528.2
	표본 수	22	17

자료: 저자 작성.

1.3. 생육실측 자료의 전처리

○ 필지정보에는 관수시설 유무, 수확 예정일 등 분석에 불필요한 정보가 포함되고 조사 항목은 품목별, 차수별로 각기 다르게 구성되어 있어, 통일성을 갖추도록 데이터 전처리의 과정이 필요함.

○ 마늘의 경우, 1~4차는 고정 개체 5개, 5~11차는 수확 개체 5개를 추가로 조사함. 따라서, 1~4차는 각 반복 당 5개, 5~11차는 10개의 개체에 대한 조사값이 존재함.

- 영상 자료가 각 필지의 반복별 이미지이므로, 생육실측 자료 역시 존재하는 개체 조사값을 평균하여 각 필지의 반복별 조사값을 이용함.
- 조사 불가, 입력 누락 등의 결측치를 제외하고 타깃변수인 생구 무게가 존재하는 필지의 반복만을 분석에 최종적으로 이용함.

- 배추의 경우, 연도별로 각 필지의 반복별 개체 수가 다르며, 3~6개의 개체를 대상으로 조사가 진행되었음. 또한, 타깃변수인 구중은 마지막 차수에만 조사되어, 2022년까지는 3차, 2023년부터는 4차의 구중 값을 이용하였음.
- 배추 역시, 영상 자료와의 결합을 위해 개체 조사값을 평균하여 각 필지의 반복별 조사값을 이용하였으며, 조사 불가, 입력 누락 등의 결측치를 제외한 뒤 최종적으로 분석에 이용함.

2. 영상 자료 수집 및 전처리

2.1. 영상 자료의 생성

- 농업관측센터에서는 정량적 수치인 생육실측조사와 더불어 작물의 생육상황을 시각적으로 확인할 수 있도록 2021년부터 영상 자료를 추가로 수집하고 있음.
- 영상 자료는 회전익 드론을 이용한 항공촬영을 통해 수집되며 필지 변화, 기상여건 등 현장 상황에 따라 5, 7, 10m 고도에서 촬영을 진행함.
- 고도별로 촬영되는 필지의 면적은 <표 3-6>과 같으며, 약 6.9~51.4평의 면적이 촬영됨.

<표 3-6> 고도별 촬영 면적

고도(m)	가로(m)	세로(m)	면적(㎡)	면적(평)
5	5.45	4.17	22.73	6.89
7	10.25	7.04	72.16	21.87
10	15.98	10.62	169.71	51.43

자료: 농업관측센터 내부자료.

- 생육실측조사구역에 대한 영상 자료 수집의 목적은 이미지 분석을 통한 생육 상황 자료 수집이 아니라 생육 실측치에 대한 교차검증과 시각적 파악을 위한 참고용 자료로 활용하기 위함이며, 촬영 대상 선정, 고도 등 구체적인 매뉴얼이 통일되어 있지는 않음.
 - 즉, 모든 조사 대상을 촬영하는 것이 아닌 생육실측조사 대상 필지 중 일부 필지의 일부 차수에 대하여 가능한 경우에 영상 자료를 수집함.
 - 또한, 영상 자료에는 농가의 작업 상황, 조사원의 조사 상황 등이 포함되므로 담당 부서 및 관련 품목 담당자만이 열람할 수 있는 대외비로 관리되고 있음.

- 특히, 영상 자료는 참고용으로 활용되는 만큼 비교적 체계적으로 관리되지는 못하고 있음. 예를 들면 기상 상황에 따라 촬영이 불가능하여 일부 영상 자료가 누락되어 있거나 다른 품목과 섞여 있고, 파일명의 형식이 통일되지 않았거나 오탈자가 포함되어 있는 등의 사례가 있음.

2.2. 영상 자료의 현황

2.2.1. 마늘

- 마늘의 생육실측조사 대상지 중 항공촬영 대상 필지는 2021년 82개, 2022년 89개, 2023년 87개, 2024년 70개이며, 강풍, 우천, 드론 파손 등의 이유로 실제 촬영이 이루어진 필지는 각각 81개, 89개, 87개, 69개임.
 - 2021~2024년의 기간 동안 항공촬영을 통해 취득한 영상 자료는 총 3,220개(각각 770개, 983개, 850개, 620개)이며, 2021년 7m 고도에서 촬영된 18개 영상 자료 이외에는 모두 5m 고도에서 촬영되었음.

○ 차수별로는 마늘의 주 생육기인 3~8차의 시기에 영상 자료도 많이 분포하고 있으며, 수확이 늦게 진행된 일부 필지에 대하여 9~11차까지도 촬영이 진행되었음을 확인할 수 있음.

〈표 3-7〉 생육실측조사 영상 자료 현황: 마늘, 차수별

단위: 개

구분	1차	2차	3차	4차	5차	6차	7차	8차	9차	10차	11차	합계
2021년	0	0	0	180	111	120	140	111	57	42	9	770
2022년	36	87	165	162	114	72	126	96	116	0	9	983
2023년	75	147	105	117	66	88	102	87	42	12	9	850
2024년	0	27	57	53	77	111	101	114	60	14	6	620
합계	111	261	327	512	368	391	469	408	275	68	33	3,220

자료: 저자 작성.

○ 항공촬영 대상 필지의 지역적 특성은 〈표 3-8〉과 같으며, 충청북도 제천시, 단양군, 전라남도 함평군, 경상북도 군위군은 2022년까지의 영상 자료만 존재하고, 경상남도 남해군, 제주특별자치도는 2024년의 영상 자료가 존재하지 않음.

- 전라남도 무안군, 신안군, 함평군, 해남군, 경상남도 남해군, 제주특별자치도의 영상 자료는 모두 난지형 남도종이며, 충청남도 서산시, 태안군, 전라북도 완주군, 경상북도 영천시, 경상남도 창원군, 합천군은 난지형 대서종, 충청북도 단양군, 제천시는 한지형 마늘임.
- 전라북도 부안군과 전라남도 고흥군은 난지형 남도종과 난지형 대서종이 모두 포함되어 있으며, 경상북도 군위군과 의성군은 난지형 대서종과 한지형 마늘이 포함되어 있음.

〈표 3-8〉 생육실측조사 영상 자료 현황: 마늘, 지역별

단위: 개

구분		2021년	2022년	2023년	2024년
충청북도	제천시	3	33	0	0
	단양군	12	39	0	0
충청남도	서산시	48	39	39	33
	태안군	42	36	6	45
전라북도	부안군	78	80	78	60
	완주군	39	48	69	15
전라남도	고흥군	111	60	159	81
	무안군	9	12	21	20
	신안군	27	66	60	48
	함평군	30	48	0	0
	해남군	24	78	121	71
경상북도	영천시	36	75	63	30
	군위군	66	57	0	0
	의성군	54	60	96	85
경상남도	남해군	45	69	15	0
	창녕군	72	78	66	117
	합천군	56	84	33	15
제주특별자치도	서귀포시	3	12	21	0
	제주시	15	9	3	0
합계		770	983	850	620

자료: 저자 작성.

2.2.2. 여름배추

○ 여름배추의 생육실측조사 대상지 중 항공촬영 대상 필지는 2021년 122개, 2022년 85개, 2023년 73개, 2024년 51개이며, 실제 촬영이 이루어진 필지는 각각 119개, 73개, 69개, 47개임.

○ 2021~2024년의 기간 동안 항공촬영을 통해 취득한 영상 자료는 총 1,511개로, 연도별로 각각 602개, 349개, 317개, 243개임.

- 차수별로 보면, 3차 이후의 영상 자료 개수가 1, 2차에 비해 현저히 적은 것을 확인할 수 있는데, 이는 예정보다 일찍 수확하여 싹배추로 출하하는 경우가 있기 때문으로 판단됨.

〈표 3-9〉 생육실측조사 영상 자료 현황: 여름배추, 차수별

단위: 개

구분	1차	2차	3차	4차	합계
2021년	257	225	120	0	602
2022년	104	165	80	0	349
2023년	81	102	78	56	317
2024년	114	54	51	24	243
합계	556	546	329	80	1,511

자료: 저자 작성.

○ 항공촬영 대상 필지의 지역적 특성은 다음 표와 같은데, 대부분의 지역이 2021~2024년 기간 동안의 영상 자료가 모두 존재함.

- 정선군은 2022년, 영월군은 2023년에만 영상 자료가 존재하며, 춘천시는 생육실측조사를 진행하는 필지가 존재하나 해당 필지의 영상 자료는 취득하지 않은 것으로 나타남.

〈표 3-10〉 생육실측조사 영상 자료 현황: 여름배추, 지역별

단위: 개

구분	2021년	2022년	2023년	2024년
강릉시	87	33	27	9
삼척시	182	30	50	45
태백시	120	66	36	36
영월군	0	0	3	0
정선군	69	48	39	87
평창군	129	154	147	57
홍천군	9	6	6	3
횡성군	6	12	9	6
합계	602	349	317	243

자료: 저자 작성.

2.3. 영상 자료의 전처리

○ 이미지 파일명은 생육실측 자료와 다른 방식으로 구성되어 있어 생육실측 자료와 결합하여 사용할 수 있도록 이미지 파일명 재구성이 필요함.

- 이미지 파일명은 필지 ID, 연도, 차수, 반복, 촬영날짜의 내용을 담도록 재구성한 뒤, 생육실측 자료와 일치하는 이미지만을 분석에 이용함.

○ 본 연구의 대상 지역인 경상남도 창원군의 영상 자료는 총 333개이며, 차수 별로도 역시 3~8차의 시기에 영상 자료가 많이 분포하고 있었음.

- 창원군 마늘 영상 자료 중, 생육실측조사 날짜와 항공촬영 날짜가 상이한 4개 필지를 제외한 뒤, 생구 무게 값이 존재하는 5~9차인 219개의 영상 자료만을 분석에 활용하였음.

〈표 3-11〉 분석에 활용된 영상 자료: 마늘

단위: 개

구분	5차	6차	7차	8차	9차	합계
2021년	18	0	15	15	6	69
2022년	12	15	12	0	3	78
2023년	6	12	0	18	6	66
2024년	9	24	21	18	9	108
합계	45	51	48	51	24	219

자료: 저자 작성.

○ 본 연구의 대상 지역인 강원특별자치도 평창군의 배추 영상 자료는 총 487개 임. 2024년의 영상 자료 개수가 매우 적은 것은 고온 및 병해 발생, 싹배추 출하 증가 등의 영향인 것으로 판단됨.

- 평창군 배추 영상 자료 중, 생육실측조사 날짜와 항공촬영 날짜가 상이하거나 항공촬영 날짜를 알 수 없는 자료와 완전고령지 작기에 해당하는 자료를 제외한 뒤, 425개의 영상 자료를 검토하여 분석에 활용하였음.

〈표 3-12〉 분석에 활용된 영상 자료: 배추

단위: 개

구분	1차	2차	3차	4차	합계
2021년	36	39	51	0	126
2022년	38	45	36	0	119
2023년	36	54	24	24	138
2024년	24	9	6	3	42
합계	134	147	117	27	425

자료: 저자 작성.

3. 기상 자료 수집

- 생육실측 자료와 이미지 자료는 모두 필지 단위로 데이터가 구축되어있음에 따라 필지 단위의 기상정보 자료를 활용할 필요가 있음.
- 기상관측자료로는 기상청의 종관기상관측(ASOS), 방재기상관측(AWS), 농업기상관측(AAOS) 등이 있음. 이들 자료는 시간 단위별, 기상 요소별로 세부적으로 제공하고 있지만, 제공지점을 통해서 관측되는 기상정보이기 때문에 농가 필지의 기상정보로 활용하기에는 제한적임.
- 국립농업과학원은 국내 지역 특성상 좁은 지역 내 기상분포 변동이 크게 나타나는 상황에 따라 정밀한 기후 정보를 표현하기 위해 ‘소기후 모형’을 개발·운영하고 있음.
- 소기후 모형은 우리나라 전역의 농장 필지 단위까지 공간적으로 농업 기상 정보를 산출할 수 있음(김대준 외, 2019).
 - 소기후 모형은 지속적인 개선을 통해 일 최저 및 최고기온과 강수량, 일사량, 일조시간 등 상세 기상 실황 및 예보정보를 제공하고 있으며, 농장규모

기상정보를 토대로 각종 농작물의 생육 정보와 기상재해 예측정보를 생산하고 있음(김대준 외, 2019).

- 현재, 생육실측 자료는 생육조사 시점의 조사일자와 해당 조사농가의 주소와 같은 필지 정보를 갖고 있음.
 - 생육실측 자료의 해당 품목과 필지 정보 바탕으로 소기후 모형을 통해서 해당 농가 필지의 기상정보를 생육 기간에 맞게 일자별로 산출할 수 있음.
 - 따라서, 기상 자료는 각 해당 농가 필지별로 소기후 모형으로 산출된 기상정보를 활용하기로 하였음. 또한, 해당 데이터는 일자별로 산출된 자료이므로 시차를 적용하여 이용하기도 용이함.

- 소기후 모형으로부터 제공받은 기상정보 자료는 농가 필지단위 평균기온, 최저기온, 최고기온, 강수량, 일사량, 일조시간, 상대습도, 평균풍속 등이며, 2020년부터 현재 산출 가능한 시점까지로 정리하였음.

4. 시사점

- 본 연구에 사용된 생육실측 자료와 드론 촬영 이미지 자료는 서로 활용 방법이 달라 데이터의 수집일자, 수집장소, 표본농가 ID 등이 통일된 형태로 기재되어 있지 않음.
 - 또한 기상여건 등에 의해 정해진 일자에 실측조사나 촬영이 이루어지지 못하여 데이터가 누락되거나 다른 품목의 저장공간에 혼재되어 있는 등 데이터 관리가 제대로 이루어지지 않는 문제가 나타났음.

- 이러한 문제들로 인해 분석용 데이터를 구축할 때 생육실측 자료와 이미지 자료의 조사일자를 확인하여 매칭시키는 추가 작업이 필요하였으며, 각각의 파일명을 통일시키는 작업에 많은 시간이 소요되었음.
- 본 연구에 활용된 이미지 데이터의 수가 인공지능으로 구증을 예측하기에 부족하여 1장을 9개로 나누는 이미지 분할을 시도하였음.
- 그러나 일정 간격으로 이미지를 분할한 결과 포전의 이량이 너무 큰 비중을 차지하거나 개체가 잘려나가는 등의 문제가 발생하였음.
 - 따라서 이미지 분할 보다는 좌우반전, 밝기조절, 회전, 리사이즈, 대비조절 등의 이미지 증강을 시도하여 분석에 필요한 이미지 데이터를 확보하였음.
- 농작물의 생육에 기상 조건이 미치는 영향이 중요하여 국립농업과학원의 소기후 모형에서 제공하는 필지별 기상 자료를 분석에 활용하였음.
- 본 연구에서는 대서종 마늘 주산지인 창녕군의 생육실측조사 필지별 기상 자료를 분석에 사용하였으며, 고랭지배추에 대해서는 기상 자료 수집 시간이 부족하여 분석에서 제외하였음.
 - 향후 추가적인 연구 및 관측 고도화를 위해 국립농업과학원의 소기후모형을 활용하여 품목별 주산지의 생육실측조사 필지에 대한 기상 자료를 사전에 확보하는 것이 필요함.
- 농업관측센터는 전화조사, 생육실측조사, 드론을 활용한 작물 이미지 촬영 등 관측 대상 품목의 생육과 관련한 많은 정형·비정형 데이터를 생성하고 보유하고 있음. 따라서 관측 고도화를 위해 사전에 품목별 인공지능 학습용 데이터셋을 구축하는 것이 요구됨.

4

인공지능(AI) 활용 생산량 추정

1. 인공지능(AI) 활용 분석 개요

- 농업관측센터는 관측정보 수집 및 예측력 제고를 위해 실측조사 및 항공촬영 등의 다양한 방법으로 정보를 수집하고 있음.
- 정보수집 조사는 배추, 무, 건고추, 마늘, 양파를 대상으로 진행하고 있으며, 본 연구에서는 데이터 수집이 비교적 용이한 배추와 마늘을 대상으로 분석하고자 함.
- 생산량 추정을 위해서 구중 추정 모형을 구축하고자 하였고, 이용할 수 있는 자료는 생육실측 자료와 항공촬영 이미지 자료임.
 - 생육실측 자료는 생육 단계별 작황을 조사한 자료로 구중, 초장, 엽수 등의 생육 정보를 갖고 있는 형태임.
 - 항공촬영 이미지 자료는 해당 실측 구역을 촬영한 자료로 조사시기에 따라

해당 구역을 촬영하여 시간이 지남에 따라 생산 동향을 파악할 수 있음.

- 생육실측조사 자료는 수치 정보를 갖는 정형 데이터이고, 항공촬영 자료는 이미지 정보를 갖고 있는 비정형 데이터임. 따라서, 정형 데이터와 비정형 데이터를 이용한 분석이 필요함.

- 또한, 분석에 이용한 데이터는 모형 간 비교가 용이하도록 실측 데이터와 이미지 데이터를 결합하여 전처리를 진행하여, 같은 데이터로 각 구중 모델을 추정하였음.
 - 실측 데이터와 항공촬영 데이터는 각각 2020년, 2021년부터 축적되기 시작하여 시계열의 길이가 매우 짧은 편임.
 - 또한, 실측조사 데이터와 항공촬영 데이터는 동일한 시점에서 같이 축적된 것이 아니라 실측조사 중 일부 시점들에서만 항공촬영이 이루어져 결합하게 되면 항공촬영 데이터와 결합되지 못한 실측조사 데이터가 소실되는 한계점이 있음.

- 본 연구는 해당 데이터를 바탕으로 다양한 시도를 하여 예측모형 구축이 필요함에 따라 정형 데이터를 이용한 구중 모델, 비정형 데이터를 이용한 구중 모델, 정형 데이터와 비정형 데이터를 결합한 멀티모달(multi modal)¹⁾ 구중 모델 분석을 진행하였음.
 - 일반적으로 멀티모달은 텍스트, 이미지, 음성, 비디오 등을 결합하여 학습 및 처리하는 방식을 의미하는데 정형 데이터와 비정형 데이터의 결합도 멀티모달의 일종으로 볼 수 있음.

1) 멀티모달은 주변 환경을 이해하기 위해 다양한 정보 채널을 동시에 결합하는 것을 의미함(Akkus et al., 2023).

2. 정형 데이터 분석

2.1. 분석자료

2.1.1. 마늘

○ 마늘 생육 실측 데이터는 주산지들을 대상으로 조사가 이루어지지만 난지형 대서종 마늘 최대 주산지인 창녕군 지역으로 한정하였고, 2021~2024년 데이터를 이용하였음.

- 마늘 생육조사는 각 포전마다 개체별로 1~9차(혹은 10차)까지 진행되며 그 데이터가 축적되어있음.
- 1~4차 조사는 지하부 생구가 형성되지 않아 지상부 초장, 엽수, 엽초장, 엽초경, 잎끝고사장 등의 생육 정보가 조사되며, 5~9차에는 1~4차에 조사되는 정보와 더불어 일부 개체에 대해 구직경, 생구 무게 등이 조사됨.

○ 생산량 추정을 위해서는 구중 무게 예측모형을 구축하는 것이 필요하고 그에 따라 타깃변수는 구중 무게가 되며, 피처변수로는 초장, 엽초장, 엽초경 및 기상정보인 평균온도, 강수량, 일사량을 설정하였음.

- 엽수는 생육 특성 및 상관관계 등을 고려했을 때 피처변수로 활용이 제한적이라 판단하여 제외하였음.
- 기상정보는 국립농업과학원의 소기후모형을 통해 각 필지의 기상정보를 추정하고, 해당 시점의 과거 15일 동안의 기상정보의 평균값과 누적값을 분석에 이용하였음.

$$\text{구중 무게} = f(\text{초장, 엽초장, 엽초경, 평균온도, 강수량, 일사량}) \quad (1)$$

○ 5~9차 조사는 조사항목이 고정되어 있어 비교적 시계열 데이터로 가공이 가능함에 따라 5~9차의 데이터를 시계열 형태로 이용하여 분석을 진행하고자 하였음.

- 그러나 각 데이터 특징에 따른 분석을 위해 동일한 데이터프레임으로 분석이 필요함에 따라 이미지 데이터를 결합하는 과정을 진행하였으며, 이에 따라 시계열을 이용한 분석이 제한됨.

○ 따라서, 마늘 데이터는 시계열 형태이지만 횡단면 형태로 조정하여 분석자료로 이용하였고, 2021~2023년 데이터를 훈련셋으로 하고 2024년 데이터는 테스트셋으로 구분하여 분석하였음.

- K-폴드 교차검증을 진행함에 따라 별도의 검증셋은 구성하지 않음.

2.1.2. 여름 배추

○ 여름 배추는 마늘과 다르게 1~3차(혹은 4차)로 생육실측조사가 이루어짐에 따라 시계열 형태로 구축하는 것이 더욱 제한적임.

- 자료 수집 지역은 평창군으로 한정하였고, 그 중 완전고랭지 주산지인 대관령면을 제외하고 준고랭지 1, 2기작 데이터를 분석 대상으로 설정함.

○ 여름 배추는 회차별 동일한 조사항목으로 생육정보가 축적되는 마늘과 달리 1차 조사(초장, 엽수, 엽폭), 2차 조사(초장, 엽폭), 3차 조사(초장, 엽수, 엽폭, 내엽수, 외엽수, 구고, 구폭) 등 회차별 조사항목이 상이하여 연속성을 갖는 변수는 초장과 엽폭 뿐임.

○ 따라서, 3차의 구중 무게를 타깃변수로 두고, 피쳐변수로는 1차의 초장, 엽

수, 엽폭, 2차의 초장, 엽폭을 설정하였음.

- 4이 예측모형의 한계는 데이터가 적어 시계열이 아닌 횡단면으로 구성된다는 것임.
- 3차의 구중 무게는 데이터값의 범위가 너무 넓어 이상치 탐지(anomaly detection)가 필요하다고 판단됨에 따라 Isolation Forest 기법으로 이상치를 제거하고 분석에 이용함.²⁾
- 또한, 기상 데이터의 확보가 용이하지 않고, 해당 데이터 구성에서 기상 설정은 제한적이기 때문에 기상정보는 고려하지 않았음.

$$\text{구중 무게} = f(\text{1차 초장, 1차 엽수, 1차 엽폭, 2차 초장, 2차 엽폭}) \quad (2)$$

○ 마늘과 마찬가지로 여름 배추의 훈련셋 2021~2023년 데이터로 이용하였고, 테스트셋은 2024년 데이터로 진행함.

- 1기작 데이터의 관측치는 241개이며, 그중 훈련셋은 232개, 테스트셋은 9개임. 2기작의 총관측치는 81개이고, 훈련셋은 67개, 테스트셋은 14개임.
- 배추도 K-폴드 교차검증을 진행함에 따라 검증셋은 설정하지 않음.

²⁾ Isolation Forest는 Liu et al.(2008)이 고안한 이상치 탐지 알고리즘으로 이상치는 저밀도 영역에 존재하기 때문에 격리(isolation)하기 쉽다는 점에 초점 맞춘 기법임. 각 관측치의 격리 경로 길이(path length)를 이용하여 이상치 판별 점수(anomaly score)를 산출해서 1에 가까우면 이상치, 0.5 이하면 정상치로 판별하는 형태임. 이때, 이상치 판별 점수는 0과 1 사이의 범위를 가짐.

2.2. 분석 방법

- 인공지능을 기반으로 한 마늘과 배추의 구중 예측력 비교를 위해서는 다양한 알고리즘 기반의 분석 방법이 필요하다고 판단되었고, 그중에서 결정트리 기반의 랜덤포레스트(Random Forest), 그래디언트 부스팅 기법의 발전된 형태인 XGBoost(Extreme Gradient Boosting), 합성곱 신경망을 이용한 딥러닝 기법인 ConvLSTM(Convolutional Long Short-Term Memory)으로 모형을 구축하였음.

2.2.1. Random Forest

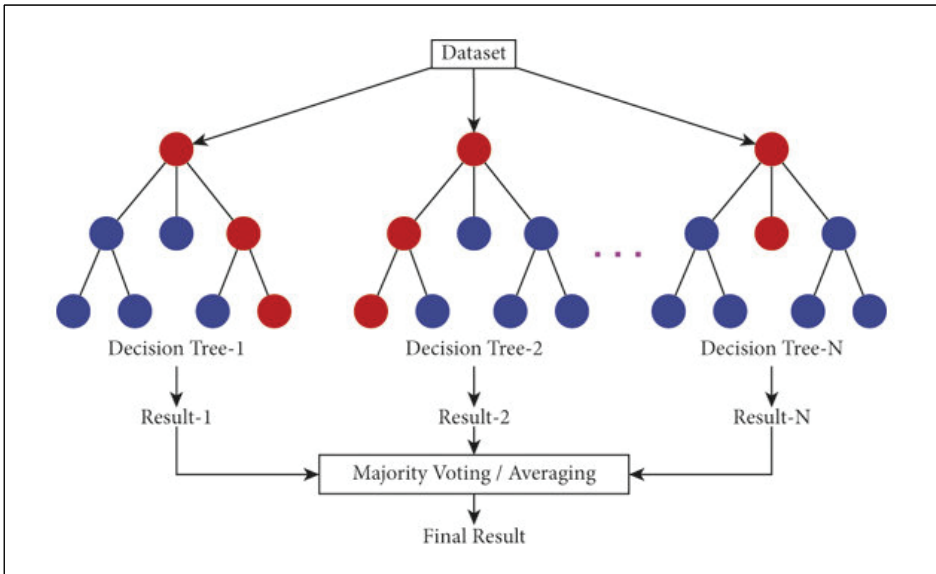
- 랜덤포레스트는 B개의 결정트리 기반으로 타깃변수를 예측하는데 각 트리는 독립적으로 학습을 진행함. 여기서 피쳐변수는 벡터 $X = [x_1, x_2, x_3, \dots, x_n]$ 이고, 각 트리 노드에서 기준값을 기준으로 특정 피쳐 x_i 로 서브셋으로 나누고 노드의 불순도(impurity)를 최소화하는 방향으로 수행함.
- 결정트리에서 학습된 새로운 입력 벡터 X 는 리프 노드에서 예측값 $h_b(X)$ 로 도출되고 최종 예측값은 아래와 같은 식으로 산출됨.

$$\hat{y} = \frac{1}{B} \sum_{b=1}^B h_b(X) \quad (3)$$

- 예측력 제고를 위해 주요 하이퍼파라미터인 n_estimators, max_depth, min_samples_split, min_samples_leaf, max_features 등의 범위를 설정하고, Optuna를 통해 하이퍼파라미터 범위 내에서 최적의 조합을 찾으려 하였음.

- n_estimators: 생성할 트리 개수로 50~500 사이로 설정하였으며, 많을수록 성능이 개선되지만 계산 비용은 증가함.
- max_depth: 각 결정트리의 최대 깊이로 5~30 범위로 지정하였고, 깊을수록 세부 학습이 가능하지만 과적합(overfitting)의 가능성이 존재함.
- min_samples_split: 노드 분할의 최소한 샘플 수로 2~20 사이로 지정하였음. 이 파라미터는 작으면 세부 분할이 이루어져 복잡한 모델을 형성하며 크면 반대로 단순한 모델로 만들어짐.
- min_samples_leaf: 리프 노드를 구성하는 최소 샘플 개수로 값이 작을수록 과적합 위험이 증가할 수 있으므로 1~10의 사이로 선택함.
- max_features: 각 노드에서 분할할 때 사용할 최대 피쳐 수로 'sqrt'나 log2를 이용하면 피쳐 수가 제한되어 과적합을 방지할 수 있음.

〈그림 4-1〉 Random Forest 기본구조



자료: Khan et al.(2021).

2.2.2. XGBoost

○ XGBoost는 입력 피쳐 벡터 X 에 대해 타깃변수 y 의 예측값을 생성하는데 여기에 새로운 모델인 t 번째 학습된 결정트리 $f_t(x)$ 가 학습률(learning rate)인 η 를 통해서 타깃변수 y 와 오차를 줄이도록 동작함.

$$\hat{y}^{(t)} = \hat{y}^{(t-1)} + \eta f_t(X) \quad (4)$$

○ 이를 바탕으로 손실함수를 최소화하는 것이 목표이며, 아래의 손실함수를 최소화함. $\iota(y_i, \hat{y}_i)$ 는 타깃변수 실측치와 예측치 간의 오차를 측정하는 함수이고, $\Omega(f_t)$ 는 트리 복잡도 정규화 항임.

- γ 은 트리 리프 노드 추가에 따른 페널티 계수, T 는 트리 리프 노드 개수, ω_j 는 리프 j 의 가중치, λ 는 리프 노드 가중치를 제어하는 정규화 계수임.

$$L = \sum_{i=1}^n \iota(y_i, \hat{y}_i) + \sum_{t=1}^T \Omega(f_t) \quad (5)$$

$$\Omega(f_t) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T \omega_j^2 \quad (6)$$

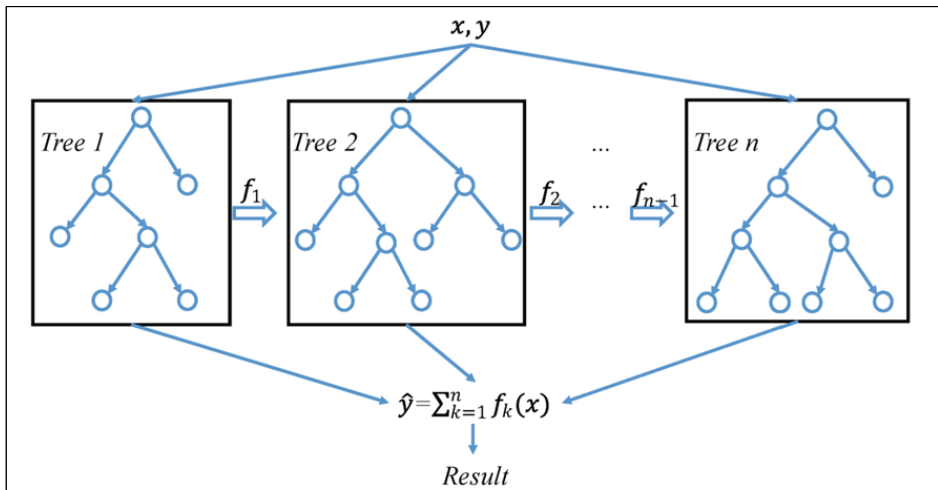
○ XGBoost에서 하이퍼파라미터 최적화는 max_depth, learning_rate, n_estimators, subsample, colsample_bytree, gamma로 최적의 조합을 찾도록 하였음.

- n_estimators: 50~500 사이로 설정하여 최적 트리 개수를 검색함.
- max_depth: 3~30 범위로 지정해 모델의 복잡성을 고려함.
- learning_rate: 학습률로 작으면 학습이 느리지만 세밀하며 반대로 크면

학습이 빠르지만 최적화가 제한적임. 0.005~0.5 사이로 탐색함.

- subsample: 각 부스팅 단계에서 사용할 샘플링 비율을 의미하며, 0.5~1.0 범위로 설정함.
- colsamples_bytree: 비율이 적정하면 트리 간 독립성으로 과적합을 방지할 수 있으므로 0.5~1.0 사이로 진행함.
- gamma: 트리 분할에 따른 최소 손실 감소량으로 작으면 더 분할이 세부적으로 진행되어 복잡한 모델이 되며 그에 따라 과적합 가능성이 존재함. 1e-8~1.0 사이의 로그스케일 값으로 설정함.

〈그림 4-2〉 XGBoost 기본구조



자료: Wang et al.(2019).

2.2.3. ConvLSTM(Convolutional LSTM)

- ConvLSTM은 합성곱 신경망(CNN)과 LSTM을 결합한 모델로 시간적 의존성과 공간적 패턴을 동시에 학습하기 유리함.

○ ConvLSTM은 LSTM의 게이트 구조인 입력 게이트(i_t), 망각 게이트(f_t), 출력 게이트(o_t)와 셀 상태(C_t)에 합성곱 연산(\otimes)으로 진행하여 학습하는 형태임. 따라서 아래와 같이 나타낼 수 있음.

- X_t 는 t 기 피쳐변수 벡터, H_{t-1} 은 $t-1$ 기의 은닉 상태, b 는 바이어스, σ 는 시그모이드 활성화 함수, \tanh 는 하이퍼볼릭 탄젠트 함수임.

$$i_t = \sigma(W_{xi} * X_t + W_{hi} * H_{t-1} + W_{ci} \odot C_{t-1} + b_i) \quad (7)$$

$$f_t = \sigma(W_{xf} * X_t + W_{hf} * H_{t-1} + W_{cf} \odot C_{t-1} + b_f) \quad (8)$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tanh(W_{xc} * X_t + W_{hc} * H_{t-1} + b_c) \quad (9)$$

$$o_t = \sigma(W_{xo} * X_t + W_{ho} * H_{t-1} + W_{co} \odot C_t + b_o) \quad (10)$$

$$H_t = o_t \odot \tanh(C_t) \quad (11)$$

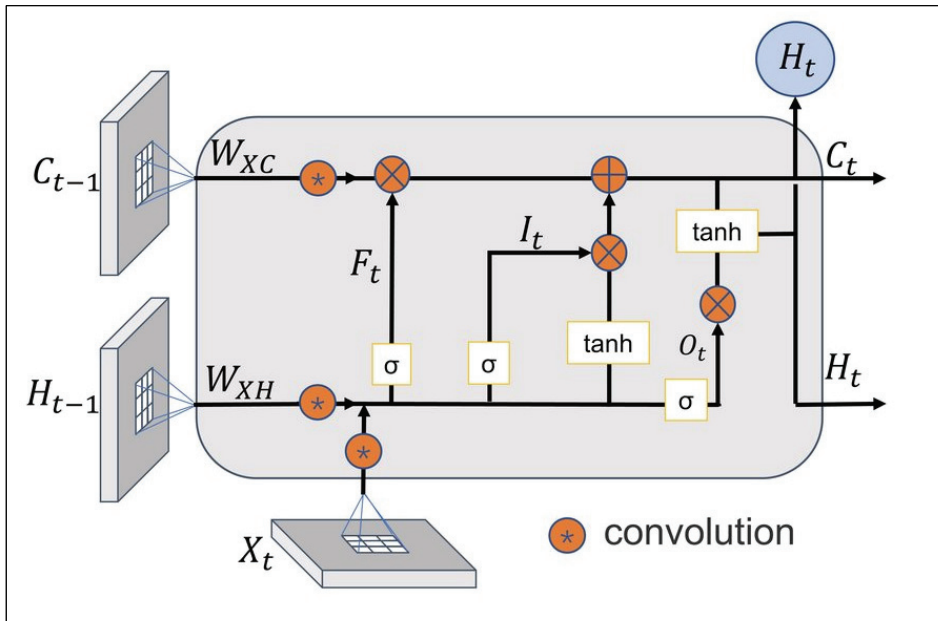
○ 식(11)의 은닉 상태의 H_t 가 Dense layer에 연결되어 최종 예측값인 타깃 변수 y 를 산출하는 형태임.

○ ConvLSTM의 하이퍼파라미터는 filters, kernel_size, learning_rate, batch_size, epochs, dropout_rate의 조합으로 최적화를 진행하였음.

- filters: 레이어의 필터 개수로 많으면 표현력이 증가하지만 그에 따라 계산 비용이 증가함. 8~128 사이로 설정하였음.
- kernel_size: 레이어의 필터 크기로 클수록 넓은 부분의 정보를 학습할 수 있음. 1×1 , 3×3 중에 선택하도록 함.

- learning_rate: 1e-5~1e-2 사이로 설정하여 적정 학습률을 탐색하고자 하였음.
- batch_size: 모델의 가중치 학습 과정에서 처리되는 데이터 샘플 수로 병렬 연산을 위해 2의 제곱 형태로 8, 16, 32, 64 중에서 탐색하도록 함.
- epochs: 모델이 학습하기 위한 반복 횟수로 많으면 충분히 학습하지만 과적합의 가능성이 발생하므로 10~50 사이로 검색함.
- dropout_rate: 학습 도중 뉴런 일부를 비활성화시켜 과적합을 방지하는 지표로 높을수록 뉴런의 비활성화 빈도가 높아짐. 0.0~0.5 사이로 설정함.

<그림 4-3> ConvLSTM 기본구조



자료: Shi et al.(2022).

2.3. 분석결과

2.3.1. 마늘

- 마늘 구중 모형 분석은 랜덤포레스트, XGBoost, ConvLSTM으로 동시 진행하여 예측력을 비교해보고자 하였음.
- 주요 하이퍼파라미터 최적화와 교차검증은 Optuna 라이브러리를 이용하여 진행하였고, 최적화 방법은 베이지안 최적화의 일종인 트리 기반 파젠 추정(Tree-structured Parzen Estimator: TPE)³⁾으로 진행함.
- 랜덤포레스트 모델의 결정트리($n_estimators$)는 142, 결정트리층(max_depth)은 10, 내부 노드분할을 위한 샘플 수($min_samples_split$)는 4, 리프 노드에 있어야 할 최소 샘플 수(min_sample_leaf)는 1, 트리 분할 시 최대 개수($max_features$)는 전체 특징 수의 제곱근만큼 사용하는 것이 최적으로 나타났음.
- XGBoost 모델의 하이퍼파라미터는 부스팅 단계 수($n_estimators$) 450, 트리 최대 깊이(max_depth) 14, 학습 속도($learning_rate$) 0.010, 부스팅 단계에서 데이터 샘플 비율($subsample$) 0.701, 트리 생성 시 사용할 특징 비율($colsample$) 0.541, 트리 복잡도 제어 값($gamma$) $1.330e-08$ 에서 최적임.

³⁾ TPE는 하이퍼파라미터 최적화 기법 중 베이지안 최적화(Bayesian Optimization) 기반의 알고리즘임. 이 기법은 목적함수의 정보를 이용하여 하이퍼파라미터의 확률 분포를 추정하는 방법으로 목적함수 값을 이용하여 성능이 낮은 영역의 분포와 성능이 좋은 영역의 분포를 구분하고 성능이 좋은 영역의 분포 확률이 높은 하이퍼파라미터 값을 계속해서 찾는 과정을 거침. 다수의 시행으로 하이퍼파라미터 값이 수렴하는 값을 최종 선택하게 됨(Bergstra et al., 2011, 2013; Watanabe, 2023).

○ ConvLSTM은 합성곱 층 필터 개수(filters)는 100개, 커널 사이즈(kernel size)는 1, 학습률(learning_rate)은 0.008, 배치 사이즈(batch size)는 8, 에포크(epochs)는 41, 드롭아웃 비율(dropout rate)은 0.248로 결정됨.

〈표 4-1〉 마늘 정형 데이터 구중 모형 하이퍼파라미터 최적화

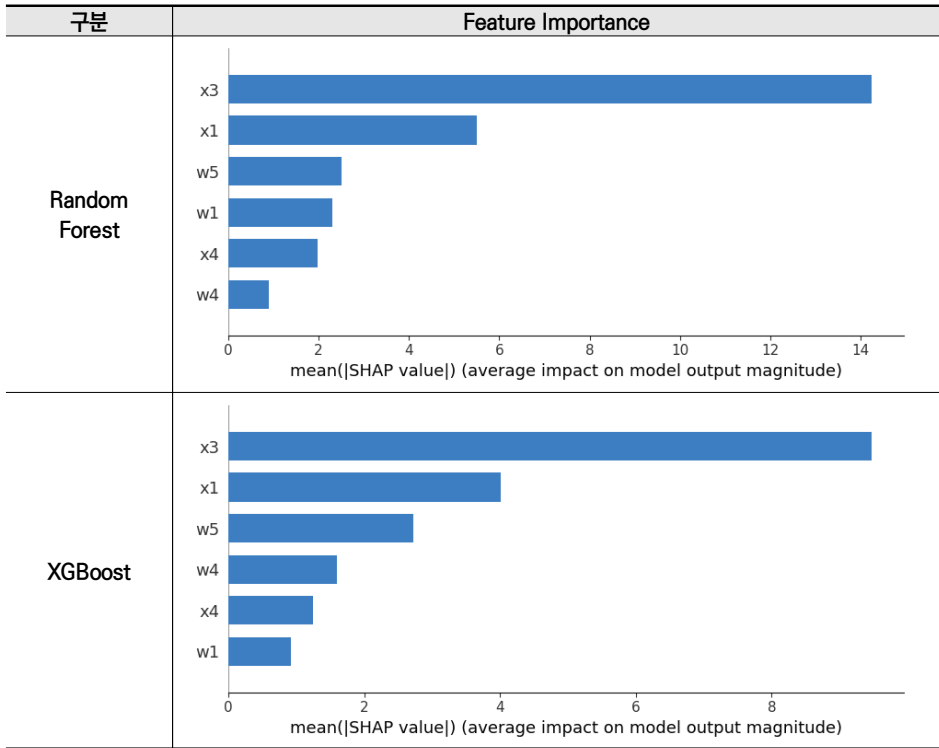
구분	Random Forest	XGBoost	ConvLSTM
n_estimators	142	450	-
max_depth	10	14	-
min_samples_split	4	-	-
min_samples_leaf	1	-	-
max_features	sqrt	-	-
learning_rate	-	0.010	0.008
subsample	-	0.701	-
colsample	-	0.541	-
gamma	-	0.000	-
filters	-	-	100
kernel size	-	-	1
batch size	-	-	8
epochs	-	-	41
dropout rate	-	-	0.248

주: 하이퍼파라미터 값은 소수점 셋째 자리로 제시함.

자료: 저자 작성.

- 마늘 구중 모형 구축에 이용된 피처변수들이 타깃변수인 구중 무게 예측에 얼마나 기여했는지를 확인하기 위해 피처 중요도를 산출하였음.
- 랜덤포레스트와 XGBoost 모두 엽초장(x3)이 구중 무게 예측에 가장 기여하는 변수로 나타났으며, 그 뒤로 초장(x1), 일사량(w5) 순으로 중요한 변수로 확인되었음.
 - 엽초경(x4)과 평균기온(w1), 강수량(w4)도 구중 무게에 기여하는 변수이지만 그 중요도는 낮은 편이고 모형에 따라 다르게 나타남.

〈그림 4-4〉 마늘 정형 데이터 구중 모형 특성 중요도



자료: 저자 작성.

○ 2021~2023년의 데이터로 학습하여 구축된 마늘 구중 모형에 2024년 테스트 데이터를 이용하여 산출된 예측치를 실제치와 비교한 성능지표를 보면 ConvLSTM이 두 모형에 비해 예측력이 좋은 모델로 나타났음.

〈표 4-2〉 마늘 정형 데이터 구중 모형 성능지표

구분	Random Forest	XGBoost	ConvLSTM
RMSE	22.830	22.406	21.147
MAE	18.619	19.909	17.988
MSE	521.209	502.025	447.178
MAPE	77.980	89.160	74.960

주: 성능지표 값은 소수점 셋째 자리까지 제시함.

자료: 저자 작성.

2.3.2. 배추

- 배추 1기작과 2기작 랜덤포레스트 모델의 하이퍼파라미터는 결정트리 ($n_estimators$) 52, 153, 결정트리층(max_depth) 10, 16, 내부 노드분할을 위한 샘플 수($min_samples_split$) 16, 12, 리프 노드에 있어야 할 최소 샘플 수(min_sample_leaf) 3, 1이며, 트리 분할 시 최대 개수($max_features$)는 1기작의 경우 전체 특징 수의 제곱근만큼 사용하는 것이 최적으로 나타났다으며, 2기작은 모든 특징을 그대로 이용하는 것이 가장 최적으로 계산됨.
- 1기작과 2기작 각각 XGBoost 하이퍼파라미터는 부스팅 단계 수($n_estimators$) 189, 77, 트리 최대깊이(max_depth) 8, 28, 학습 속도는($learning_rate$) 0.010, 0.035, 부스팅 단계에서 데이터 샘플 비율($subsample$) 0.600, 0.578, 트리 생성 시 사용할 특징 비율($colsample$) 0.934, 0.833, 트리 복잡도 제어 값 ($gamma$) $8.101e-05$, $6.305e-05$ 로 설정됨.
- ConvLSTM의 하이퍼파라미터를 보면 합성곱 층 필터 개수($filters$)는 1기작 101, 2기작 92, 커널 사이즈($kernel\ size$)는 모두 1, 학습률($learning_rate$)은 0.010, 0.009, 배치 사이즈($batch\ size$)는 모두 8, 에포크($epochs$)는 44, 38, 드롭아웃 비율($dropout\ rate$)은 0.291, 0.245로 결정됨.

〈표 4-3〉 배추 1기작 정형 데이터 구중 모형 하이퍼파라미터 최적화

구분	Random Forest	XGBoost	ConvLSTM
n_estimators	52	189	-
max_depth	10	8	-
min_samples_split	16	-	-
min_samples_leaf	3	-	-
max_features	sqrt	-	-
learning_rate	-	0.010	0.010
subsample	-	0.600	-
colsample	-	0.934	-
gamma	-	0.000	-
filters	-	-	101
kernel size	-	-	1
batch size	-	-	8
epochs	-	-	44
dropout rate	-	-	0.291

주: 하이퍼파라미터 값은 소수점 셋째 자리로 제시함.
 자료: 저자 작성.

〈표 4-4〉 배추 2기작 정형 데이터 구중 모형 하이퍼파라미터 최적화

구분	Random Forest	XGBoost	ConvLSTM
n_estimators	153	77	-
max_depth	16	28	-
min_samples_split	12	-	-
min_samples_leaf	1	-	-
max_features	None	-	-
learning_rate	-	0.035	0.009
subsample	-	0.578	-
colsample	-	0.833	-
gamma	-	0.000	-
filters	-	-	92
kernel size	-	-	1
batch size	-	-	8
epochs	-	-	38
dropout rate	-	-	0.245

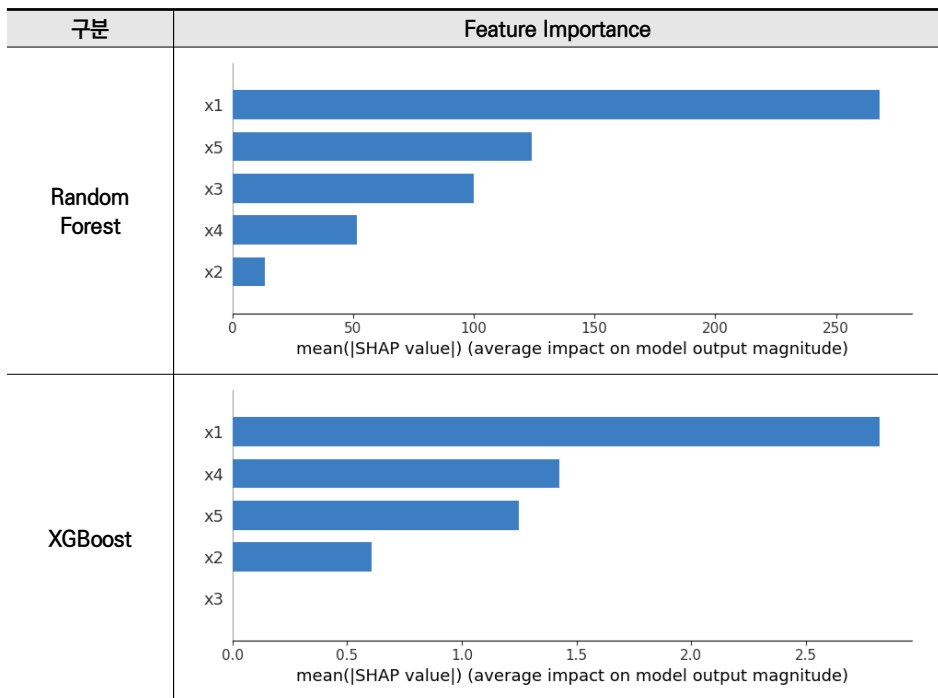
주: 하이퍼파라미터 값은 소수점 셋째 자리로 제시함.
 자료: 저자 작성.

○ 배추 1기작 구중 모형은 1차 초장(x1)이 구중 무게 예측하는데 가장 크게 기여하는 변수로 나타났고, 그 뒤로는 모형에 따라 다르게 나타났음. 2기작 구

중 모형은 1차 업수(x1)가 가장 중요도가 높은 변수로 산출되었으며, 2차 업폭(x5)이 두 번째로 중요도가 높은 것으로 확인됨.

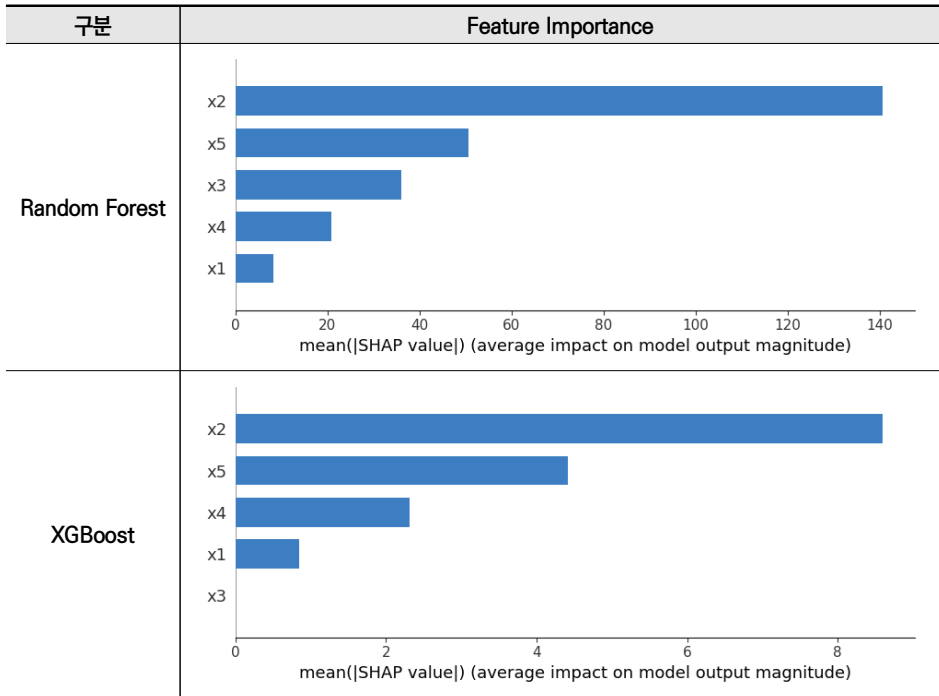
○ 1기작 모형과 2기작 모형은 각각 구중 무게 예측에 중요도가 높은 변수는 다른 양상을 보이는 것으로 나타났으며, 추후 이러한 특징을 반영하여 피처 엔지니어링을 할 필요가 있음.

〈그림 4-5〉 배추 1기작 정형 데이터 구중 모형 특성 중요도



자료: 저자 작성.

〈그림 4-6〉 배추 2기작 정형 데이터 구중 모형 특성 중요도



자료: 저자 작성.

- 배추 1기작 모형과 2기작 모형의 성능지표를 보면 1기작 모형이 2기작 모형에 비해서 매우 예측력이 높은 것으로 나타났는데 이는 데이터 수 차이 및 생육 특성에 따른 것으로 파악됨.
- 1기작 모형은 XGBoost가 예측력이 가장 좋은 것으로 나타났으며, 랜덤포레스트, ConvLSTM 순으로 나타났고, 배추 2기작 모형은 성능 개선 시도가 필요한 것으로 고려됨. 특히, 데이터의 적은 관측치와 품질적 요인이 합쳐져서 다른 모형에 비해 성능 저하를 보이는 것으로 판단됨에 따라 데이터 수 확충 및 품질 개선이 필요해 보임.

〈표 4-5〉 배추 1기작 정형 데이터 구중 모형 성능지표

구분	Random Forest	XGBoost	ConvLSTM
RMSE	1,082.279	614.085	1,151.449
MAE	996.387	502.825	1,098.435
MSE	1,171,327.038	377,100.642	1,325,833.958
MAPE	49.620	26.220	53.520

주: 성능지표 값은 소수점 셋째 자리까지 제시함.
자료: 저자 작성.

〈표 4-6〉 배추 2기작 정형 데이터 구중 모형 성능지표

구분	Random Forest	XGBoost	ConvLSTM
RMSE	854.816	754.000	775.443
MAE	806.675	728.825	719.364
MSE	730,709.813	568,516.061	601,311.934
MAPE	150.580	137.900	143.340

주: 성능지표 값은 소수점 셋째 자리까지 제시함.
자료: 저자 작성.

3. 이미지 데이터 분석

3.1. 분석자료

3.1.1. 마늘

○ 마늘 이미지는 2021년부터 2024년까지 창녕군에서 촬영된 데이터를 이용하였음. 각 이미지 파일은 JPG로 되어있고, 총 219개의 이미지 파일을 이용하였으며, 예측력을 높이기 위해 이미지 증강을 진행하였음.

- 좌우반전, 회전, 밝기조절, 대비조절, 리사이즈 등을 진행함. 다만, 상하반전은 이미지의 의미 왜곡이 발생할 가능성이 높아 이용하지 않음(Shorten & Khoshgoftaar, 2019; Yang et al., 2022).

○ 2021~2023년 이미지로 학습 및 모델을 구축하고 2024년을 테스트셋으로 하여 성능지표를 산출함.

3.1.2. 배추

○ 배추 이미지 데이터는 2021년부터 생산하였고, 매년 데이터 축적을 진행하고 있음. 본 이미지 데이터 모형은 실측 데이터와 마찬가지로 평균군을 한정하여 진행하였고, 그중에서도 준고랭지 여름배추를 대상으로 하였음.

○ 여름배추 생육실측 자료는 1기작, 2기작으로 구분할 수 있는데 각 작형은 특성이 다를 수 있기 때문에 1기작과 2기작으로 데이터를 나눠서 전처리하였음.

- 여름배추 자료에서 대관령면 지역과 출하 시점 기준으로 고랭지배추는 제외하였음.

- 또한, 출하 시점이 6, 7월이면 1기작, 9, 10월이면 2기작으로 구분하여 각 데이터프레임을 구축함.

○ 1기작 모델과 2기작 모델에 이용한 이미지는 JPG 파일로 각각 64개, 29개이며, 관측치가 적은 것으로 판단되어 예측력을 높이기 위해 이미지 증강기법을 이용하였음.

○ 마늘과 동일하게 2021~2023년 이미지로 학습 및 모델을 구축하고 2024년을 테스트셋으로 하여 성능지표를 산출함.

3.2. 분석 방법

- 이미지 예측모형 구축 전에 이미지 파일들 간의 이미지 유사도를 검증하기 위해 이미지 간 사인 코사인 유사도(Cosine Similarity)를 확인함. 이는 이미지들의 동일성을 파악하는 것이 목적으로 예측력을 판단하는 데 도움이 될 것임.
- 이미지를 이용하여 예측모형을 구축하는 방법은 일반적으로 이미지에서 특징 추출(feature extraction)을 진행하고 타깃변수를 예측하는 방법과 이미지 자체를 학습하는 End-to-End 방식으로 타깃변수를 예측하는 방법으로 나눌 수 있음.
- 본 연구는 이미지 특징 추출보다는 세부적 특징까지 학습하기 위해 이미지를 그대로 이용하여 타깃변수를 예측하는 방법을 시도하였음.
 - End-to-End 방법은 이미지 자체를 별도의 전처리나 피쳐 엔지니어링 없이 그대로를 받아서 이미지의 중요한 특징인 텍스처나 색상, 패턴 등을 자동으로 학습하는 형태임.
 - 이러한 방법은 합성곱 신경망(Convolutional Neural Network: CNN)을 이용하여 이미지의 데이터 패턴을 그대로 학습하도록 하는 것임.
- 먼저, 이미지 데이터의 크기는 $224 \times 224 \times 3$ 으로 설정하였음.
 - 224×224 는 이미지 크기를 의미하며, 이미지의 세부적인 특징을 학습하기 위해 큰 이미지로 설정하였음. 또한 3은 RGB를 의미함.
- CNN 설계는 아래와 같이 진행하였음.

$$Y[i, j, k] = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \sum_{c=0}^{C-1} X[i+m, j+n, c] \cdot K[m, n, c, k] + b_k \quad (1)$$

$Y[i, j, k]$: k 번째 출력 특징 맵의 (k, k) 위치값

$X[i+m, j+n, c]$: 입력 이미지 픽셀값

$K[m, n, c, k]$: k 번째 필터 가중치

b_k : k 번째 필터의 바이어스

M, N : 필터 크기

C : 입력채널수

- 활성화함수(Activation)는 비선형성을 추가하여 복잡한 패턴을 학습할 수 있도록 하는데 본 연구에서는 ReLU 함수를 이용하였음. 식(2)의 ReLU 함수를 통해 모든 음수값은 0으로 변환되고 양수값은 그대로 진행함.

$$f(x) = \max(0, x) \quad (2)$$

- 풀링(Pooling)은 이미지 크기를 줄여 계산량을 감소시킴과 동시에 이미지의 중요한 특징은 유지하는 역할을 하며 식(3)과 같음. s 는 스트라이드(strides)로 합성곱이 이동하는 수준을 나타냄.

$$Y[i, j, k] = \max(X[si+m, sj+n, k]) \quad (3)$$

- 위 과정들을 거쳐 1차원 벡터로 변환된 특징 맵들을 이용하여 완전 연결 레이어(Fully Connected Layer)에서 최종 예측값을 출력함. y 는 최종 예측값, x 는 1차원 변환 입력 벡터, W 는 가중치, b 는 바이어스임.

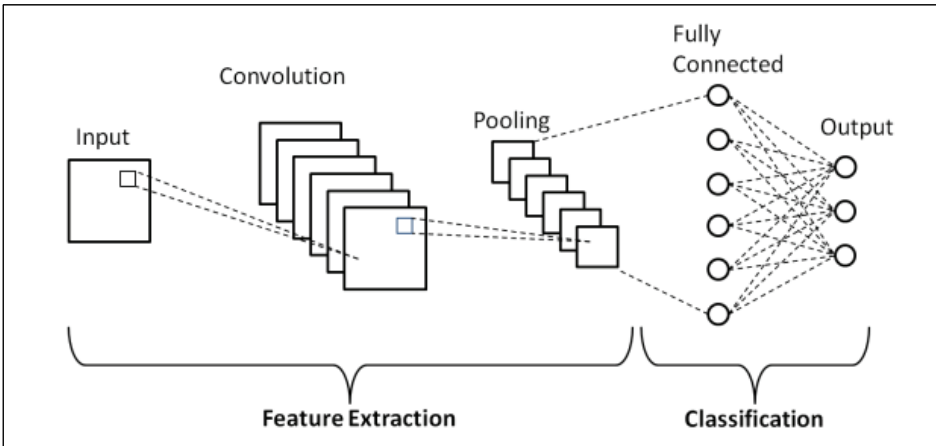
$$y = W \cdot x + b \quad (4)$$

- 데이터의 관측치가 소규모이기 때문에 예측력 제고를 위해 이미지 증강을 진행하였으며, 좌우반전, 밝기조절, 대비조절, 각도 조절(90, 180, 270도), 리

사이즈 등을 진행하였음.

- 하이퍼파라미터 최적화(Hyperparameter Optimization: HPO)와 K -폴드 교차검증은 Optuna 패키지를 이용하여 최적모형 학습을 진행하였고, MSE 값이 최소가 되도록 하였음.

<그림 4-7> CNN 기본구조



자료: Phung & Rhee(2019).

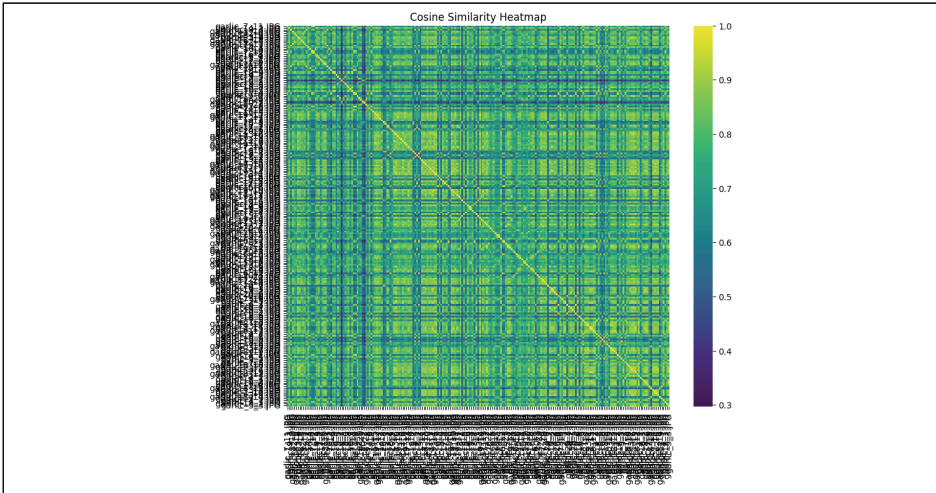
- 테스트셋에 대한 성능지표는 SE, MAE , MSE , $MAPE$ 로 진행함.

3.3. 분석 결과

3.3.1. 마늘

- 마늘 이미지 데이터의 전체 평균 유사도는 73.18%로 나타났으며, 유사도가 80%가 넘는 이미지 쌍은 35.84% 수준으로 나타나 마늘 이미지들이 유사할 가능성이 높음. 따라서 예측력 문제가 발생할 수 있음.

〈그림 4-8〉 마늘 이미지 데이터 유사도



자료: 저자 작성.

○ 하이퍼파라미터 튜닝은 주요 하이퍼파라미터인 필터 수, 커널 사이즈, 드롭아웃 비율, 학습률, 배치 사이즈를 최적화하여 모형을 구축하였음.

〈표 4-7〉 마늘 CNN 모델 하이퍼파라미터 최적화 결과

하이퍼파라미터	미늘
필터 수	32
커널 사이즈	(5, 5)
드롭아웃 비율	0.440
학습률	0.000
배치 사이즈	16

주: 하이퍼파라미터 값은 소수점 셋째 자리로 제시함.

자료: 저자 작성.

○ Conv2D는 합성곱레이어로 횡수가 늘어날수록 더 높은 수준의 특징을 추출하는 것을 의미하며 2번 진행하였음. Dense는 완전연결층으로 학습된 특징을 바탕으로 예측을 수행하며 마지막 횡수에는 최종출력층으로 예측값을 생성하는 파라미터임.

○ 마늘 CNN 모델은 23,065,281개의 파라미터를 이용하여 학습하여 모형이 구축되었음.

〈표 4-8〉 CNN 모델 요약

구분	마늘	
	출력 크기	파라미터수
Conv2D_1	(None, 220, 220, 32)	2,432
MaxPooling2D_1	(None, 110, 110, 32)	0
Conv2D_2	(None, 106, 106, 64)	51,264
MaxPooling2D_2	(None, 53, 53, 64)	0
Flatten	(None, 179776)	0
Dense_1	(None, 128)	23,011,456
Dropout	(None, 128)	-
Dense_2	(None, 1)	129

자료: 저자 작성.

○ 마늘 CNN 모델의 성능은 매우 좋지 않은 것으로 나타났으며 MAPE 기준으로 102.57로 지속적인 개선이 필요함.

〈표 4-9〉 마늘 CNN 모델 성능지표 비교

성능지표	마늘
RMSE	24.740
MAE	21.687
MSE	612.076
MAPE	102.570

주: 성능지표 값은 소수점 셋째 자리까지 제시함.

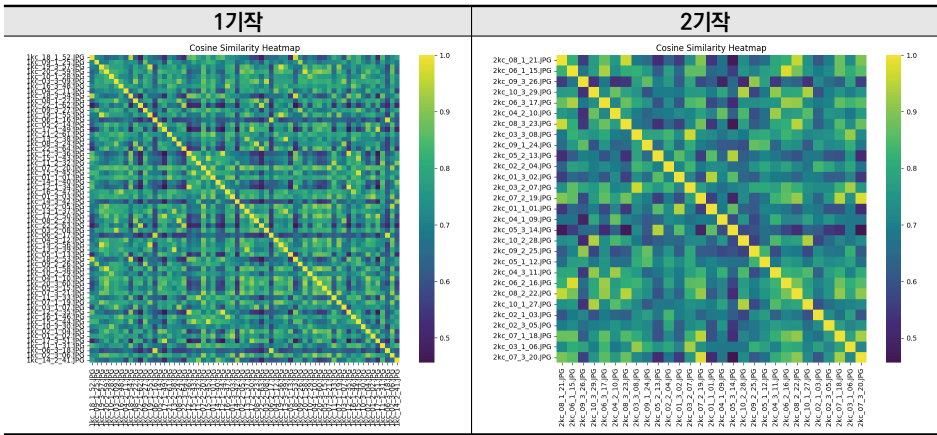
자료: 저자 작성.

3.3.2. 배추

○ 배추 이미지의 1기작과 2기작 이미지들의 유사도를 확인한 결과 1기작 이미지들의 평균 유사도는 73.16%로 나타났으며, 80% 이상 비율은 28.17%로 나타났음.

○ 2기작의 경우 평균 유사도 72.22%, 80% 이상 비율은 21.18%로 나타나 이미지의 유사도 높을 가능성이 있으며 그에 따라 예측력 제고에 문제가 있을 수 있음.

〈그림 4-9〉 배추 이미지 데이터 유사도



자료: 저자 작성.

○ 마늘과 동일하게 하이퍼파라미터 튜닝은 주요하이퍼파라미터인 필터수, 커널 사이즈, 드롭아웃 비율, 학습률, 배치 사이즈를 최적화하였음.

〈표 4-10〉 배추 CNN 모델 하이퍼파라미터 최적화 결과

하이퍼파라미터	1기작	2기작
필터 수	32	128
커널 사이즈	(5, 5)	(3, 3)
드롭아웃 비율	0.240	0.404
학습률	0.001	0.003
배치 사이즈	16	16

주: 하이퍼파라미터 값은 소수점 셋째 자리로 제시함.

자료: 저자 작성.

○ 배추의 1기작 모델의 학습가능한 파라미터는 23,065,281개이고, 2기작은 95,850,497개임.

〈표 4-11〉 CNN 모델 요약

구분	1기작		2기작	
	출력 크기	파라미터수	출력 크기	파라미터수
Conv2D_1	(None, 220, 220, 32)	2,432	(None, 222, 222, 128)	3,584
MaxPooling2D_1	(None, 110, 110, 32)	0	(None, 111, 111, 128)	0
Conv2D_2	(None, 106, 106, 64)	51,264	(None, 109, 109, 256)	295,168
MaxPooling2D_2	(None, 53, 53, 64)	0	(None, 54, 54, 256)	0
Flatten	(None, 179776)	0	(None, 746496)	0
Dense_1	(None, 128)	23,011,456	(None, 128)	95,551,616
Dropout	(None, 128)	-	(None, 128)	-
Dense_2	(None, 1)	129	(None, 1)	129

자료: 저자 작성.

○ 배추 1기작 모델과 2기작 모델의 성능지표를 보면 1기작 모델이 2기작 모델에 비해 MAPE 기준으로 더 예측력이 좋은 것으로 나타남.

〈표 4-12〉 CNN 모델 성능지표 비교

성능지표	1기작	2기작
RMSE	682.421	676.808
MAE	561.452	650.409
MSE	465,698.169	458,069.614
MAPE	26.790	121.370

주: 성능지표 값은 소수점 셋째 자리까지 제시함.

자료: 저자 작성.

4. 멀티모달 분석

4.1. 분석자료

4.1.1. 마늘

- 마늘 실측 데이터와 이미지 데이터는 각 조사 시점과 그 시점에서의 데이터 개수로 된 시계열 형태로 구성하였지만, 그 관측치 수가 짧아 횡단면 형태로 분석을 진행하였음.
- 분석은 앞서 정형 데이터에서 분석한 대로 타깃변수는 마늘 구중으로 설정하였고, 피처변수는 초장, 엽초장, 엽초경, 평균기온, 강수량, 일사량으로 지정하였음.
 - 엽수는 마늘 생육이 일정 시점 지남에 따라 잎이 떨어지게 되는데 모형 설정에 있어 예측력을 떨어뜨리는 요인으로 작용할 수 있기에 제외하였음.
 - 평균기온, 강수량, 일사량은 국립농업과학원의 소기후모형을 통해 각 필지별 추정된 기상 자료를 이용하였으며, 해당 구중 무게 조사 시점의 과거 15일 동안의 평균값과 누적값을 적용하였음.
- 전처리된 마늘 생육 데이터에 각 개체특성(ID, 시간정보, 시점 개수)에 따라 이미지 파일명을 매칭하여 데이터프레임을 구성하였고, 이를 바탕으로 분석을 진행하였음.
- 이미지 데이터는 CNN으로 다차원벡터로 구성된 특징을 1차원 벡터로 평탄화(Flatten) 연산을 진행하여 단일 값으로 출력되고, 이 이미지 특징 추출값을 생육 데이터와 매칭하였음.

- 훈련셋은 2021~2023년 데이터로 설정하였고, 테스트셋은 2024년의 데이터로 진행하여 성능을 확인하였음.
- 검증 데이터는 따로 확보하지 않고 훈련셋 내에서 K-폴드 교차검증을 진행하여 관측치 확보와 교차검증력을 높이고자 하였음.

4.1.2. 배추

- 배추는 조사차수가 1~3, 4회차로 마늘보다 짧게 이루어져 시계열 확보에 더욱 제한적인 상황이라 마늘과 다르게 데이터를 구축하였음.
- 앞서 정형 데이터에서 설정한 함수 형태에서 CNN으로 이미지 특성을 추출하여 그 추출값을 실측 데이터에 합쳐 분석을 진행함.
- 정형 데이터 분석과 동일한 자료를 이용하되 이미지 특징 추출값만 추가하여 동일한 모델 기법으로 진행하였음.
- 시계열 데이터는 아니지만 훈련셋은 2021~2023년 데이터로 테스트셋은 2024년 데이터로 임의구분하여 진행하였으며, 검증셋 구성 없이 교차검증으로 진행함.

4.2. 분석 방법

- 멀티모달은 복수의 데이터 형식을 입력으로 하여 학습하는 인공지능을 의미하는데, 본 연구에서는 이미지 데이터를 CNN으로 특징을 추출하여 이미지의 고차원적 정보를 단일 수치값으로 변환하여 수치 데이터셋에 결합하여 분

석에 이용하였음.

- 복수 데이터를 융합하는 방식은 early fusion, intermediate fusion, late fusion으로 구분되는데, 본 연구에서 이용하는 방식은 intermediate fusion 방법에 가까움.

○ CNN을 이용하여 특징을 추출하는 과정은 합성곱 연산, 풀링 연산, 평탄화 연산 과정으로 진행되며 평탄화된 벡터가 완전 연결 계층(Dense layer)을 통해 단일 수치로 변환되어 정형 데이터 프레임에 매칭되어 결합할 수 있음.

○ 이러한 과정들을 하나의 수식으로 나타내면 아래와 같음.

- $f_{Dense}()$ 는 Dense Layer 연산이고, $Flatten$ 은 평탄화 연산, $g_{Conv}()$ 는 합성곱 계층과 활성화함수의 조합임. b 는 바이어스임.

$$\hat{y} = f_{Dense}(Flatten(g_{Conv}(W_{Conv} * X + b_{Conv}))) \quad (15)$$

○ 이미지 특징 추출값은 단일 수치로 수치 데이터 프레임과 병합하여 피처변수의 하나로 예측모형에 활용함.

4.3. 분석결과

4.3.1. 마늘

○ 멀티모달 랜덤포레스트 모델의 결정트리(n_estimators)는 239, 결정트리층(max_depth)은 7, 내부 노드분할을 위한 샘플 수(min_samples_split)는 6, 리프 노드에 있어야 할 최소 샘플 수(min_sample_leaf)는 2, 트리 분할 시 최대 개수(max_features)는 \log_2 (전체 특징수)만큼이 최적임.

○ 멀티모달 XGBoost 하이퍼파라미터는 부스팅 단계수(n_estimators)를 262, 트리 최대깊이(max_depth)는 14, 학습속도는(learning_rate) 0.016, 부스팅 단계에서 데이터 샘플 비율(subsample)은 0.711, 트리 생성 시 사용할 특징 비율(colsample)은 0.611, 트리 복잡도 제어 값(gamma)은 8.861e-06로 나타남.

○ ConvLSTM은 합성곱 층 필터 개수(filters)는 92, 커널 사이즈(kernel size)는 1, 학습률(learning_rate)은 0.009, 배치 사이즈(batch size)는 8, 에포크(epochs)는 38, 드롭아웃 비율(dropout rate)은 0.245로 확인됨.

〈표 4-13〉 마늘 멀티모달 구중 모형 하이퍼파라미터 최적화

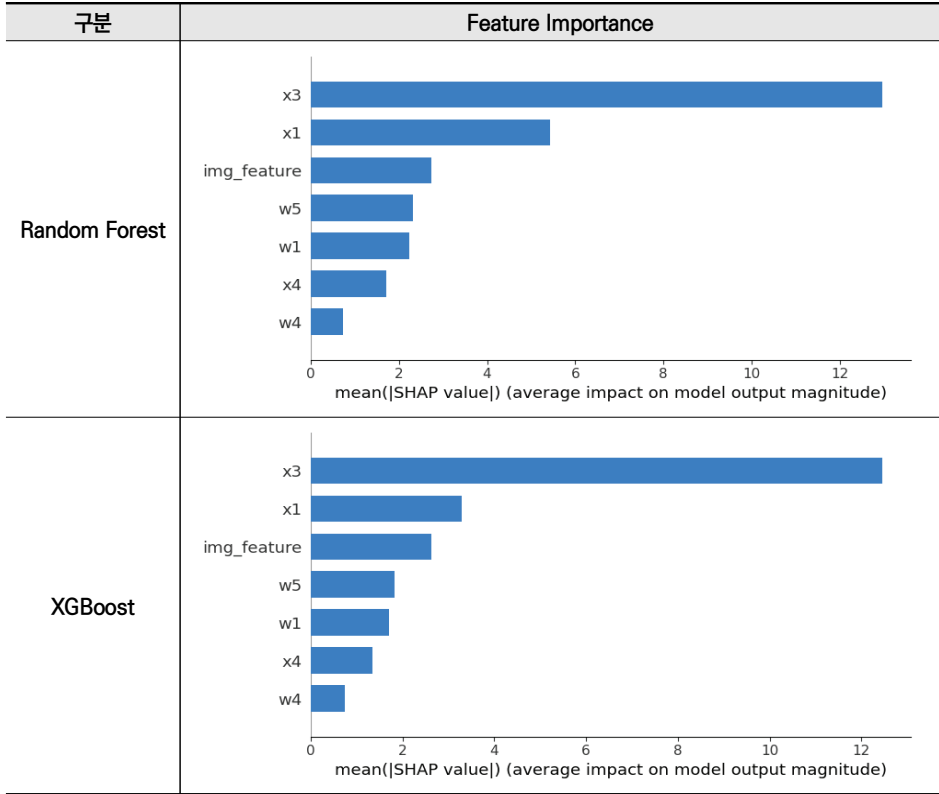
구분	Random Forest	XGBoost	ConvLSTM
n_estimators	239	262	-
max_depth	7	14	-
min_samples_split	6	-	-
min_samples_leaf	2	-	-
max_features	log ₂	-	-
learning_rate	-	0.016	0.009
subsample	-	0.711	-
colsample	-	0.611	-
gamma	-	0.000	-
filters	-	-	92
kernel size	-	-	1
batch size	-	-	8
epochs	-	-	38
dropout rate	-	-	0.245

주: 하이퍼파라미터 값은 소수점 셋째 자리로 제시함.

자료: 저자 작성.

○ 정형 데이터 피쳐중요도와 다르게 랜덤포레스트와 XGBoost 모두 동일하게 구중(y)에 엽초장(x3), 초장(x1), 이미지(img_feature), 일사량(w5), 평균기온(w1), 엽초경(x4), 강수량(w4) 순으로 나타남.

〈그림 4-10〉 마늘 멀티모달 구중 모형 특성 중요도



자료: 저자 작성.

○ 예측력은 정형 데이터만을 이용한 구중 모형보다 소폭 개선된 것으로 나타남. MAPE 기준으로 랜덤포레스트는 77.98이었으나 76.73으로, XGBoost는 89.16에서 76.16으로 ConvLSTM은 74.96에서 73.94로 개선됨.

〈표 4-14〉 마늘 멀티모달 구중 모형 성능지표

구분	Random Forest	XGBoost	ConvLSTM
RMSE	22.139	20.299	21.142
MAE	18.412	17.456	18.270
MSE	490.120	412.065	446.977
MAPE	76.730	76.160	73.940

주: 성능지표 값은 소수점 셋째 자리까지 제시함.

자료: 저자 작성.

4.3.2. 배추

- 배추 1기작과 2기작의 랜덤포레스트 모델의 결정트리($n_estimators$)는 각각 375, 357, 결정트리층(max_depth)은 25, 7, 내부 노드분할을 위한 샘플 수($min_samples_split$)는 6, 13, 리프 노드에 있어야 할 최소 샘플 수(min_sample_leaf)는 8, 2, 트리 분할 시 최대 개수($max_features$)는 1기작, 2기작 모두 전체 특징을 사용하는 것이 가장 최적으로 나타남.
- XGBoost 하이퍼파라미터는 부스팅 단계 수($n_estimators$)를 각각 482, 368, 트리 최대깊이(max_depth)는 8, 28, 학습 속도는($learning_rate$) 0.007, 0.053, 부스팅 단계에서 데이터 샘플 비율($subsample$)은 0.585, 0.594, 트리 생성 시 사용할 특징 비율($colsample$)은 0.942, 0.800, 트리 복잡도 제어 값($gamma$)은 $1.07e-08$, $3.445e-04$ 로 최적화가 이루어졌음.
- ConvLSTM은 합성곱 층 필터 개수($filters$)는 각각 100, 96, 커널 사이즈($kernel\ size$)는 모두 1, 학습률($learning_rate$)은 0.008, 0.009, 배치 사이즈($batch\ size$)는 모두 8, 에포크($epochs$)는 41, 44, 드롭아웃 비율($dropout\ rate$)은 0.248, 0.392로 결정됨.

〈표 4-15〉 배추 1기작 멀티모달 구중 모형 하이퍼파라미터 최적화

구분	Random Forest	XGBoost	ConvLSTM
n_estimators	375	482	-
max_depth	25	3	-
min_samples_split	6	-	-
min_samples_leaf	8	-	-
max_features	None	-	-
learning_rate	-	0.007	0.008
subsample	-	0.585	-
colsample	-	0.942	-
gamma	-	0.000	-
filters	-	-	100
kernel size	-	-	1
batch size	-	-	8
epochs	-	-	41
dropout rate	-	-	0.248

주: 하이퍼파라미터 값은 소수점 셋째 자리로 제시함.
자료: 저자 작성.

〈표 4-16〉 배추 2기작 멀티모달 구중 모형 하이퍼파라미터 최적화

구분	Random Forest	XGBoost	ConvLSTM
n_estimators	357	368	-
max_depth	7	15	-
min_samples_split	13	-	-
min_samples_leaf	2	-	-
max_features	None	-	-
learning_rate	-	0.053	0.009
subsample	-	0.594	-
colsample	-	0.800	-
gamma	-	0.000	-
filters	-	-	96
kernel size	-	-	1
batch size	-	-	8
epochs	-	-	44
dropout rate	-	-	0.392

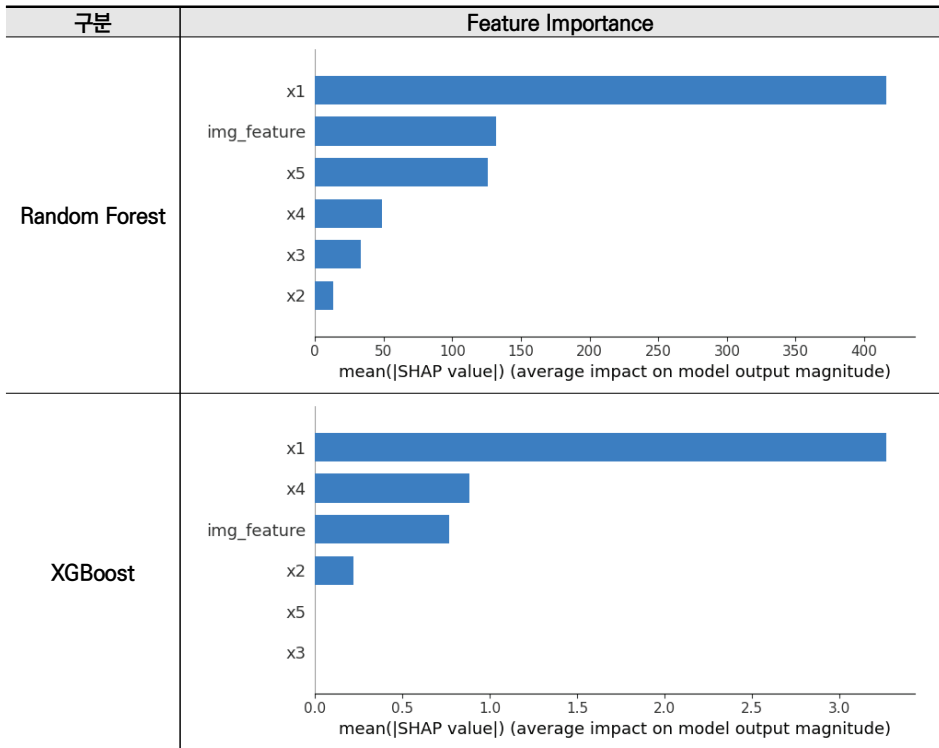
주: 하이퍼파라미터 값은 소수점 셋째 자리로 제시함.
자료: 저자 작성.

○ 배추 1기작 구중 모형의 SHAP 피쳐 중요도를 보면 1기작 모델에서는 1차 초장(x1)이 가장 구중 무게 예측에 중요도가 높게 나타났으며 이미지 변수의 경우 랜덤포레스트에서는 2순위로 중요, XGBoost에서는 3순위로 중요하게

나타나 의미 있는 변수로 확인됨.

- 2기작 구중 모형에서 랜덤포레스트 모델은 1차 엽수(x2)가 가장 높은 중요도를 나타냈으며 이미지(img_feature)가 2순위로 중요도가 높은 변수로 확인되었으며, XGBoost에서는 이미지 데이터(img_feature)가 구중 무게 예측에 가장 중요도가 높은 변수로 확인되었고 그 뒤로 1차 엽수(x2) 순으로 나타났음.

<그림 4-11> 배추 1기작 멀티모달 구중 모형 특성 중요도



자료: 저자 작성.

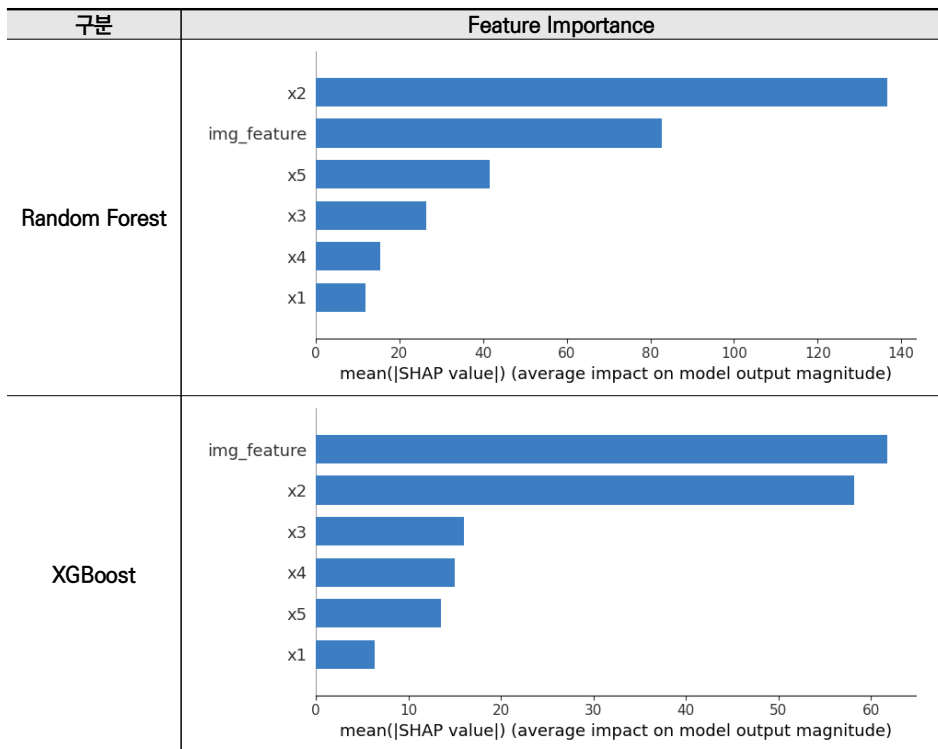
- 배추 1기작과 2기작 모형도 마늘 모형과 마찬가지로 멀티모달 데이터를 활용함으로써 예측력이 제고된 것으로 나타났음.

○ 1기작 모델은 MAPE 기준으로 랜덤포레스트는 49.62에서 49.50, XGBoost는 26.22에서 26.13, ConvLSTM은 53.52에서 50.44로 소폭이나마 개선 여력을 보임.

○ 2기작 모델은 랜덤포레스트 150.58에서 137.35, XGBoost는 137.90에서 127.43으로 개선 가능성을 확인하였음.

- 다만, ConvLSTM은 오히려 예측 성능이 떨어져 검토가 필요함.

〈그림 4-12〉 배추 2기작 멀티모달 구중 모형 특성 중요도



자료: 저자 작성.

〈표 4-17〉 배추 1기작 멀티모달 구중 모형 성능지표

구분	Random Forest	XGBoost	ConvLSTM
RMSE	1,082.148	612.144	1,098.685
MAE	990.347	501.206	1,017.219
MSE	1,171,044.198	374,719.776	1,207,109.358
MAPE	49.500	26.130	50.440

주: 성능지표 값은 소수점 셋째 자리까지 제시함.
자료: 저자 작성.

〈표 4-18〉 배추 2기작 멀티모달 구중 모형 성능지표

구분	Random Forest	XGBoost	ConvLSTM
RMSE	791.571	723.705	812.948
MAE	753.596	686.061	754.233
MSE	626,583.859	523,748.221	660,884.749
MAPE	137.250	127.430	150.360

주: 성능지표 값은 소수점 셋째 자리까지 제시함.
자료: 저자 작성.

5. 시사점

- 본 연구는 생육 실측 정보 데이터와 이미지 데이터를 이용하여 마늘과 배추의 구중 예측모형 구축을 시도하였음.
 - 마늘은 창녕군, 배추는 평창군 준고랭지 1, 2기작으로 구분하여 데이터를 한정하였음.
 - 데이터는 생육 실측 정보 데이터, 이미지 데이터, 생육 실측 정보 데이터와 이미지 데이터를 결합한 멀티모달 형태인 3가지 형태로 분석에 이용함.
- 생육 실측 정보 데이터와 멀티모달 데이터는 머신러닝 기법인 랜덤포레스트와 XGBoost, 딥러닝 기법으로 ConvLSTM을 이용하여 분석을 진행하였고, 이미지 데이터는 딥러닝 기법인 CNN을 이용하여 End-to-End 방식으로 분석함.

- 하이퍼파라미터와 교차검증은 적절한 범위를 검색하여 지정하고 범위 내에서 최적의 조합을 찾아 예측 모델을 구축하는 방법으로 진행함.
 - 하이퍼파라미터는 모델 기본값 혹은 지정하던 기존 방식에서 모델 최적화를 위해 하이퍼파라미터 최적화(HPO)를 Optuna의 TPE로 최적 조합을 산출하도록 함.
 - 교차검증은 데이터의 적은 관측치 수로 별도의 검증 데이터를 확보하는 대신 훈련데이터 내 교차검증을 진행하여 과적합을 방지하는 최적화로 시도함.

- 분석 결과를 보면 생육실측 자료를 이용한 마늘 모델은 ConvLSTM, 배추 1·2기작 모델은 XGBoost가 상대적으로 좋은 모델이었으나 전체적인 예측력은 좋지 못했음. 멀티모달 데이터도 생육 실측 자료 결과와 동일하게 나타났음.

- 그중에서 멀티모달은 생육실측 자료만을 활용한 예측력보다 소폭 개선되는 것으로 보아 이미지 데이터를 추가로 활용한 예측 모델을 구축하는 것이 예측력을 개선하는 하나의 방법이 될 수 있음을 확인하였음.

- 이미지 자료를 이용한 결과는 배추 1기작 모델은 어느 정도 예측력이 좋게 나타났으나 마늘과 배추 2기작 모델은 예측력이 좋지 않았음.

- 전체적으로 보면 예측력이 좋지 않게 나타났는데 먼저, 생육 실측 데이터와 이미지 데이터 구축 목적이 다름에도 이를 결합함으로써 많은 데이터 정보들이 소실되었기 때문으로 판단하였음. 또한, 데이터의 축적 기간이 짧았음.

- 두 번째로 데이터의 품질 문제를 들 수 있음. 생육 실측 데이터는 조사자마다 데이터를 생산하는 방법이 조금씩 다를 가능성이 있고, 비슷한 기간의 생육

정보 값들의 편차가 심하게 나타나는 경우도 많이 있어서 이에 대한 많은 전처리 과정이 요구됨.

○ 그리고, 이미지 데이터의 경우 예측을 위한 뚜렷한 목표를 바탕으로 구축되는 것이 아니기 때문에 예측 모형의 데이터로 이용하는 데 제한적임.

- 이미지 유사도 검증에서 마늘과 배추 이미지 데이터들은 각각 70%를 상회하는 유사도를 보여, 예측력 개선에 문제가 있을 수 있음.

○ 마지막으로 농업관측의 기능 중 하나는 수확기 이전에 품목별 최종 생산량과 가격 예측 정보를 제공하는 것임. 그러나 본 연구에서는 분석에 사용된 데이터의 축적 기간이 짧아 조사 회차별 최종 구중을 예측하기에는 아직 한계가 있어 향후 정형, 비정형 데이터를 지속적으로 수집하고 인공지능 분석을 할 수 있는 데이터셋 구축이 필요함.

5

농업관측 데이터 활용 고도화 방안

1. 인공지능 학습용 데이터 구축

○ 농업관측센터에서 생성하고 있는 품목별 생육실측 데이터, 드론을 이용한 항공촬영 영상 데이터 등은 서로 연계되지 못하고 따로 분석되어 농업관측월보에 정보를 제공하고 있음. 또한 생성된 데이터는 서로 다른 저장공간에 각기 보관하고 있어 체계적인 관리가 이루어지지 못하고 있음.

- 생육실측 데이터와 영상 데이터의 관리 프로세스가 명확하지 못하여 본 연구 수행을 위한 자료 수집단계에서 어느 저장공간에 어떠한 형태로 데이터가 보관·관리되고 있는지 혼돈이 발생하였음.
- 또한 생육실측 데이터와 영상 데이터의 수집일자, 수집공간, 표본농가 ID 등이 다른 형식으로 기재되어 있어, 인공지능 분석용 데이터로 전처리하는 과정에서 이를 서로 비교하여 통일시키는 업무의 비효율이 발생하고 있음.

○ 이러한 점을 개선하기 위해 생육실측 및 영상촬영을 시행하는 품목의 작기 또는 수확이 종료된 이후, 생육실측 데이터와 영상 데이터의 생성일자, 필지 등을 비교하여 동일한 일자에 동일 필지에서 생성된 자료를 서로 매칭하여 보관하는 인공지능 학습용 데이터 구축 작업이 필요함.

- 이는 농업관측센터 관측디지털팀이 주체가 되어 관측모형팀의 지원을 받아 수행하는 것이 바람직하며, 담당자를 지정하여 인공지능 학습용 데이터의 구축 및 관리업무를 담당하도록 할 필요가 있음.

○ 인공지능 학습용 데이터 구축을 위해 항공영상촬영 방법을 개선할 필요가 있음. 현재는 생육실측조사 시 드론으로 영상촬영을 병행하고 있으며 분석용 목적이 아닌 필지의 생육상황에 대한 시각적 변화를 파악하는 참고용 자료로 사용하고 있음. 따라서 항공촬영은 필지 변화, 기상 여건 등에 따라 촬영고도가 각기 다르며 촬영 대상 선정도 명시되어 있지 않음.

- 본 연구에서는 5~10m 고도에서 촬영한 이미지를 사용하여 비정형 데이터 분석을 수행하였으며, 그 결과 이미지의 유사도가 높아 예측력이 높지 않았음. 인공지능 학습용 이미지 데이터 구축을 위해서는 다양한 각도와 높이에서 촬영된 이미지가 필요하며, 향후 분석용 이미지 생성을 위한 촬영 방안을 고민할 필요가 있음.

○ 한편 유관기관의 경우 인공지능 학습용 데이터 구축·활용 가이드라인을 작성하여 데이터 수집 계획, 구축 및 활용방안을 체계적으로 관리하고 있음. 데이터 수집 계획에서는 데이터의 정보·포맷·구성·통계·특성 등을 정의하고, 데이터 구축 및 활용에서는 데이터 구축 목적·획득 방법·정제 및 가공·학습 모델·활용 방안 등을 제시하고 있음.

- 현재 농업관측센터 내에도 생육실측조사와 항공촬영 관련 가이드라인이 존재하지만, 각 가이드라인이 서로 연동되어 작동하지 않아 추가적인 작업이 필요한 상황임. 따라서 농업관측센터에서도 인공지능 학습용 데이터 구축을 위한 가이드라인을 작성하여 체계적인 자료 수집 및 관리가 필요할 것으로 판단됨.

2. 농림위성 데이터 활용 협력 방안 구축

- 농촌진흥청은 위성과 드론을 활용한 원격탐사 기술을 적용하여 농산물의 재배면적과 작황을 관측하고 있음. 농촌진흥청은 원격탐사 기술을 활용하여 벼의 재배면적과 생육상황을 관측하여 수량 추정 모형을 구축 운영하고 있으며, 동계 맥류 재배면적, 마늘·양파 재배면적 및 파종(정식)·출하 시기, 고랭지 배추 재배현황, 해남지역 가을배추 재배현황 등의 정보를 제공하고 있음.
 - 이러한 원격탐사 모형을 통한 관측은 접근이 어려운 지역을 포함한 넓은 지역에 대한 주기적이고 객관적인 자료 수집이 가능하다는 장점이 있음. 반면에 위성영상은 현황 파악이 가능하나 그 원인을 알기 어렵고 기상의 영향을 많이 받아 영상취득의 적시성이 떨어지는 단점이 있음.
- 농촌진흥청은 농업·산림 부문 관측 및 공공서비스를 목적으로 2025년 하반기에 차세대중형위성 4호(농림위성)를 발사하여 농산물 수급안정과 디지털 농업 실현 등 8개 부문에서 활용할 계획임.
 - 농림위성의 촬영폭은 120km 이상으로 전국을 3일에 촬영 가능하며, 동일 지역을 매일 촬영하여 작물의 생육변화 파악이 가능하도록 공간해상도 5m 간격으로 전국 데이터의 취득이 가능한 것으로 알려져 있음.

- 여기에 필지 단위의 드론 촬영을 융복합하여 주산지별 작물 재배면적, 생육상황 및 수량 정보, 필지별 작물 생육 정보, 병해충 정보 등을 제공하여 농업정책 및 영농현장의 의사결정을 지원할 계획임.
- 농림위성 자료는 재배면적, 출하면적, 생육예측 모형 변수 등을 농업관측 정보 생산에 활용 가능할 것으로 판단되며, 농림위성 자료의 농업관측 활용을 위한 기관 간 협력체계 구축이 요구됨. 농림위성센터는 위성사진을 이용하여 생성된 품목별/지역별 재배면적을 제공하고, 농업관측센터는 개인정보보호를 이유로 제공하지 못했던 생육실측·항공촬영 표본필지 정보를 제공하는 방안을 검토할 필요가 있음.
- 농업관측센터는 농림위성에서 제공된 영상의 분석 방법과 활용 계획을 수립하고 데이터 활용을 위한 기초연구를 추진하는 등 농림위성 자료의 활용도 제고를 위한 연구협력체계를 구축할 필요가 있음.

3. 인공지능 예측 모델 고도화

- 본 연구에서는 인공지능을 이용하여 마늘과 배추의 구중을 추정함에 있어 정형 데이터는 랜덤포레스트, XGBoost, ConvLSTM 모형을 사용하였고, 비정형 데이터는 합성곱 신경망(Convolutional Neural Network: CNN)을 사용하였으며, 정형·비정형 데이터를 결합하여 추정한 멀티모달 모형은 이미지 데이터의 특징을 CNN으로 추출하여 단일 수치값으로 변환한 랜덤포레스트, XGBoost, ConvLSTM 모형을 사용하였음.

- 정형 데이터를 이용한 마늘 구중 추정 모형은 ConvLSTM의 예측력이 상대적으로 우수하였으며, 준고랭지 배추 1기작의 구중 추정 모형은 XGBoost 모형의 예측력이 가장 좋았음.
- 비정형 데이터를 이용하여 CNN으로 추정된 마늘 구중 모형의 MAPE는 102.57로 매우 좋지 않았으며, 배추 구중 모형의 1기작 MAPE는 26.79로 마늘에 비해 우수하였으나 2기작의 MAPE는 121.37로 예측력이 매우 낮았음.
- 멀티모달 모형으로 추정된 마늘 구중 모형의 MAPE는 ConvLSTM 모형, 배추 1기작의 구중 모형의 MAPE는 XGBoost 모형의 예측력이 상대적으로 우수하였으나, 정형 데이터를 사용하여 추정된 예측력에 비해 소폭 개선된 것에 불과하였음. 다만, 배추 2기작 모형은 정형 데이터 활용 모형의 예측력에 비해 개선 가능성을 확인하였음.

○ 이러한 결과는 ① 본 연구에 사용된 정형·비정형 데이터의 생성 목적이 구중 예측보다는 농업관측월보 작성의 부가적 자료로 활용되기 위함이기 때문이며 ② 비대기까지는 외형부의 측정값이 증가하는 경향을 보이다 수확기를 앞두고 잎의 영양분이 구중으로 이동하는 마늘의 생육 특성으로 인해 발생하였으며, ③ 생육실측조사나 항공촬영사업 기간이 짧아 누적 데이터의 양이 부족하여 발생한 것으로 판단됨.

- 이와 같은 요인들로 인해 인공지능 모형의 최적 하이퍼파라미터를 찾는 데 많은 시간이 소요되었으며, 이러한 하이퍼파라미터 최적화 작업은 새로운 데이터가 입력될 때마다 모형의 성능과 효율성을 극대화하기 위해 반드시 지속되어야 함.

○ 이를 해결하기 위해 향후 도입할 수 있는 방안으로는 데이터 증강기법, 식별

모델, 자동화 또는 최신 모델 기법을 활용해서 검토하는 것이 필요함.

- 먼저, 이미지 데이터 구축에서는 기존 RGB 형태의 품질을 높이는 방법이나 초분광, 다분광 이미지 데이터 형태로 구축을 검토할 필요가 있음.
- 또한, 적은 데이터 수를 보완하기 위해 본 연구에서는 전통적인 방식(회전, 이동, 확대 등)을 이용하였으나 최근에는 자동화, 생성 모델, 혼합 등으로 다양하게 발전하고 있음에 따라 이를 적용할 수 있는 방안을 검토할 필요가 있음.
 - 자동화 기법은 단순한 증강을 넘어서 증강한 데이터를 이용하여 최적 조합을 찾는 과정도 추가되는 형태임. Auto-Augment 기반의 기법으로 볼 수 있음.
 - 생성모델은 기존 이미지를 조정하는 형태가 아닌 이미지를 학습하여 새로운 형태의 이미지를 구성하는 방식으로 GAN(Generative Adversarial Network), diffusion model 등이 있음.
 - 혼합은 기존 방식에서 이미지 데이터 간 혼합을 통해 다양한 패턴을 학습하도록 유도하는 것으로 이미지 데이터 간 특정 비율로 섞어서 새로운 이미지 증강을 시도하는 것으로 Mixup, Cutmix 등이 있음.
- 모형 측면에서는 현행 모델들의 하이퍼파라미터 최적화를 확대 또는 최신 최적화 기법을 이용하거나 사전학습 모델을 적용, 앙상블(Ensemble) 기법으로 예측력을 향상시키는 방법들이 있음.
 - 본 연구에서 이용한 하이퍼파라미터 최적화는 최근 사용되는 기법으로 비교적 사용이 용이하므로 이 기법을 기반으로 발전시킬 필요가 있음.
 - 이미지 예측모델은 이미지 데이터를 바탕으로 학습모델을 자체적으로 만들었으나 대규모 데이터를 이용하여 만들어진 사전학습 모델을 이용하면

예측력을 향상시킬 가능성이 있음. CNN 계열(ResNet, EfficientNet, ConvNeXt)과 Vision Transformer 계열(Vit, DeiT) 등을 중심으로 지속적으로 연구되고 있으므로 이를 검토할 필요가 있음.

- 앙상블 기법은 다수의 모델을 결합하여 예측력을 높이는 기법으로 배깅(Bagging), 부스팅(Boosting), 스택킹(Stacking) 등으로 나뉘어지며, 단일 예측모델을 이용하기보다는 여러 모델을 결합하는 방식도 활용해 볼 필요가 있음.

○ 영상 자료 등 비정형 데이터를 이용한 인공지능 분석은 보다 높은 수준의 컴퓨터 사양을 요구하고 있음. 특히 멀티모달 분석 시에는 이미지에서 특징을 추출하는 단계와 학습 단계에서 많은 계산이 필요함에 따라 인공지능 분석용 고사양 컴퓨터 확충이 필요함. 또한 인공지능을 활용한 농업관측 고도화를 위해 인공지능을 전공한 전담인력의 확충도 필요함.

○ 본 연구에서는 시계열 데이터가 짧아 조사 차수별로 마늘의 최종 구중 예측치 변화를 살펴볼 수 없었음. 따라서 실측조사 및 영상 데이터를 충분히 확보하여 차수별 최종 구중을 예측할 필요가 있음. 또한 이번 연구에서는 품목별로 1지역의 주산지를 선정하여 최종 구중을 예측하였으나, 분석 모델을 안정화시킨 이후 분석지역을 전국으로 확대할 필요가 있음. 그리고 분석 품목·품종·작형을 확대하는 등 추가적인 보완 연구가 계속 이어질 필요가 있음.

부 록 1

생육 및 수량 실측조사 평균값

〈부표 1-1〉 창녕지역 마늘 생육 및 수량 실측조사 평균값

구분	1차	2차	3차	4차	5차	6차	7차	8차	9차	10차		
2020년	초장(cm)	34.0	42.5	58.7	80.9	88.1	93.2	93.1	83.6	78.7	59.8	
	엽수(매)	6.9	8.5	8.8	9.0	9.2	8.5	7.5	6.5	5.1	3.3	
	엽초장(cm)	3.4	10.5	18.9	22.8	29.7	34.2	37.6	37.4	40.4	37.5	
	엽초경(mm)	13.7	17.3	18.6	20.6	20.2	19.9	18.4	16.7	14.4	11.3	
	구 직경(mm)	-	-	-	-	39.1	47.5	54.0	59.3	63.3	62.5	
	1주 무게(g)	-	-	-	-	-	-	-	-	-	-	
	생구 무게(g)	-	-	-	-	44.3	56.9	74.6	91.9	95.1	92.7	
	경엽 무게(g)	-	-	-	-	-	-	-	-	-	-	
	2021년	초장(cm)	21.5	31.9	41.9	68.4	80.9	87.9	88.4	86.1	76.1	-
		엽수(매)	5.8	6.6	7.5	7.9	8.0	7.8	7.1	6.4	4.5	2.6
엽초장(cm)		5.4	6.8	10.3	19.1	27.7	30.6	37.5	37.7	39.5	36.6	
엽초경(mm)		8.5	11.7	14.3	16.2	18.3	18.3	17.3	16.2	14.9	10.9	
구 직경(mm)		-	-	-	-	35.7	36.3	48.7	57.5	64.7	60.6	
1주 무게(g)		-	-	-	-	129.7	129.6	158.7	163.5	162.6	-	
생구 무게(g)		-	-	-	-	35.8	43.0	65.0	84.9	101.0	80.4	
경엽 무게(g)		-	-	-	-	93.9	86.6	93.7	78.6	61.6	-	
2022년		초장(cm)	14.2	27.9	37.0	55.1	75.1	83.4	79.8	71.9	66.7	-
		엽수(매)	6.0	6.0	6.7	7.4	8.1	8.1	8.3	6.6	4.6	-
	엽초장(cm)	2.5	4.5	8.2	14.3	25.5	31.1	33.9	35.2	40.1	-	
	엽초경(mm)	5.1	10.0	10.3	12.1	14.6	14.8	12.8	12.3	9.8	-	
	구 직경(mm)	-	-	-	-	34.3	37.7	49.0	53.4	56.4	-	
	1주 무게(g)	-	-	-	-	74.0	106.7	129.9	119.1	122.4	-	
	생구 무게(g)	-	-	-	-	27.1	40.3	63.8	66.7	85.3	-	
	경엽 무게(g)	-	-	-	-	46.9	66.4	66.1	52.5	37.2	-	
	2023년	초장(cm)	28.5	29.4	37.8	54.5	73.5	83.0	84.9	81.8	64.3	-
		엽수(매)	6.1	6.5	7.6	8.2	8.3	8.3	7.2	7.9	4.4	-
엽초장(cm)		5.2	4.7	8.4	14.2	21.7	27.3	35.8	38.0	37.1	-	
엽초경(mm)		10.4	10.1	11.5	13.5	15.2	15.2	15.3	14.9	10.3	-	
구 직경(mm)		-	-	-	-	30.6	36.6	48.0	54.6	54.8	-	
1주 무게(g)		-	-	-	-	95.0	127.5	142.7	143.3	103.1	-	
생구 무게(g)		-	-	-	-	24.7	35.7	57.7	67.6	63.0	-	
경엽 무게(g)		-	-	-	-	70.3	91.8	85.0	75.7	40.1	-	
2024년		초장(cm)	23.0	30.3	35.2	46.6	60.2	74.8	82.1	74.9	75.0	-
		엽수(매)	5.6	6.5	6.6	7.2	7.4	7.7	7.1	6.1	4.3	-
	엽초장(cm)	3.2	5.2	6.9	10.1	15.6	22.5	29.8	31.6	37.8	-	
	엽초경(mm)	7.3	9.4	10.3	12.3	13.9	14.5	15.1	12.8	11.9	-	
	구 직경(mm)	-	-	-	-	23.5	29.7	37.5	47.4	57.1	-	
	1주 무게(g)	-	-	-	-	64.7	83.8	108.3	110.5	122.4	-	
	생구 무게(g)	-	-	-	-	15.7	19.4	33.0	50.9	68.6	-	
	경엽 무게(g)	-	-	-	-	49.1	64.4	75.3	59.6	53.8	-	

자료: 저자 작성.

〈부표 1-2〉 평창지역 고랭지배추 생육 및 수량 실측조사 평균값

구분	1차	2차	3차	4차	
2020년	초장(cm)	18.2	27.3	33.2	-
	엽수(매)	18.3	27.7	96.3	-
	엽폭(cm)	13.3	19.6	29.1	-
	외엽수(매)	-	-	17.2	-
	내엽수(매)	-	-	72.5	-
	구고(cm)	-	-	18.4	-
	구폭(cm)	-	-	25.7	-
	구중(g)	-	-	1,829.7	-
2021년	초장(cm)	16.9	28.5	32.9	-
	엽수(매)	14.0	46.0	87.1	-
	엽폭(cm)	12.1	21.2	22.7	-
	외엽수(매)	-	-	77.2	-
	내엽수(매)	-	-	10.0	-
	구고(cm)	-	-	25.4	-
	구폭(cm)	-	-	20.1	-
	구중(g)	-	-	2,194.9	-
2022년	초장(cm)	17.0	29.7	33.8	-
	엽수(매)	13.5	-	73.1	-
	엽폭(cm)	11.9	20.8	22.9	-
	외엽수(매)	-	-	8.9	-
	내엽수(매)	-	-	64.1	-
	구고(cm)	-	-	26.2	-
	구폭(cm)	-	-	19.0	-
	구중(g)	-	-	2,032.6	-
2023년	초장(cm)	16.3	29.2	33.5	34.7
	엽수(매)	29.2	-	69.8	72.3
	엽폭(cm)	11.4	20.4	22.5	23.1
	외엽수(매)	-	-	8.9	9.1
	내엽수(매)	-	-	61.0	63.1
	구고(cm)	-	-	24.9	25.4
	구폭(cm)	-	-	18.6	18.0
	구중(g)	-	-	1,756.8	1,748.5
2024년	초장(cm)	19.7	28.7	34.2	36.1
	엽수(매)	15.1	-	66.8	68.6
	엽폭(cm)	13.7	21.1	22.9	23.0
	외엽수(매)	-	-	11.4	12.1
	내엽수(매)	-	-	55.4	56.5
	구고(cm)	-	-	26.2	28.1
	구폭(cm)	-	-	17.6	17.5
	구중(g)	-	-	1,483.7	1,528.2

자료: 저자 작성.

부 록 2

코드 예시: 마늘 모형 분석 코드

○ 해당 코드는 python 3.9, tensorflow 2.10.0, Compute Capability 8.9, CUDA v11.2, cuDNN 8.1.1의 환경에서 작성되었음.

```
# install required packages
!pip install pandas scikit-learn xgboost matplotlib tensorflow-addons
torch optuna shap --quiet

# import libraries
import os
import random
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.model_selection import TimeSeriesSplit, KFold
from sklearn.ensemble import RandomForestRegressor
from sklearn.preprocessing import MinMaxScaler
from sklearn.metrics import mean_absolute_error, mean_squared_error,
r2_score
import tensorflow as tf
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import (
    ConvLSTM2D, Conv2D, MaxPooling2D, Flatten, Dense, Dropout
)from tensorflow.keras.preprocessing.image import load_img, img_to_ar-
ray
```

```

import torch
import xgboost as xgb
import optuna
import shap

# reproducibility
RANDOM_SEED = 42
np.random.seed(RANDOM_SEED)
random.seed(RANDOM_SEED)
tf.random.set_seed(RANDOM_SEED)
torch.manual_seed(RANDOM_SEED)

# load dataset
excel_path = r'C:\Users\user\Desktop\mun\ml\garlic\df.xlsx'
image_folder = r'C:\Users\user\Desktop\mun\ml\image'
sheet_name = 'data'
img_size = (224, 224)
data = pd.read_excel(excel_path, sheet_name=sheet_name)

# define image processing function
def extract_image_features(image_folder, data, img_column, img_size):
    image_features = []
    for _, row in data.iterrows():
        img_filename = row[img_column]
        img_path = os.path.join(image_folder, f"{img_filename}.jpg")
        if os.path.exists(img_path):
            img = load_img(img_path, target_size=img_size)
            img_array = img_to_array(img) / 255.0
            image_features.append(img_array)
        else:

```

```

        image_features.append(np.zeros((*img_size, 3)))
    return np.array(image_features)

# define CNN
def build_custom_cnn(input_shape):
    model = Sequential([
        Conv2D(32, (3, 3), activation='relu', input_shape=input_shape),
        MaxPooling2D((2, 2)),
        Conv2D(64, (3, 3), activation='relu'),
        MaxPooling2D((2, 2)),
        Conv2D(128, (3, 3), activation='relu'),
        MaxPooling2D((2, 2)),
        Flatten(),
        Dense(128, activation='relu'),
        Dropout(0.5),
        Dense(1, activation='linear')
    ])
    model.compile(optimizer='adam', loss='mse')
    return model

# extract image features
images = extract_image_features(image_folder, data, 'i', img_size)

# train CNN & extract image features
cnn_model = build_custom_cnn(input_shape=(224, 224, 3))
cnn_model.fit(images, data['y'], epochs=10, batch_size=16, validation_split=0.1, verbose=1)
image_features = cnn_model.predict(images).flatten()

# add extracted image features to dataframe

```

```

data['img_feature'] = image_features

# define target and features variable
features = ['x1', 'x3', 'x4', 'w1', 'w4', 'w5', 'img_feature']
target = 'y'

# normalize
scaler = MinMaxScaler()
data[features] = scaler.fit_transform(data[features])

# split data into training data and test data
train_data = data[data['n'] <= 138]
test_data = data[(data['n'] >= 139) & (data['n'] <= 219)]
# separate target and features
X_train, y_train = train_data[features], train_data[target]
X_test, y_test = test_data[features], test_data[target]

# k-fold cross-validation
kf = KFold(n_splits=5, shuffle=True, random_state=RANDOM_SEED)

# 1. RandomForest using Optuna
def objective_rf(trial):
    n_estimators = trial.suggest_int('n_estimators', 50, 500)
    max_depth = trial.suggest_int('max_depth', 5, 30)
    min_samples_split = trial.suggest_int('min_samples_split', 2, 20)
    min_samples_leaf = trial.suggest_int('min_samples_leaf', 1, 10)
    max_features = trial.suggest_categorical('max_features', ['sqrt', 'log2',
None])
    model = RandomForestRegressor(n_estimators=n_estimators, max_depth=
max_depth, min_samples_split=min_samples_split, min_samples_leaf=

```

```

min_samples_leaf, max_features=max_features, random_state=RANDOM_
SEED)
    mae_list = []
    for train_index, val_index in kf.split(X_train):
        X_tr, X_val = X_train.iloc[train_index], X_train.iloc[val_index]
        y_tr, y_val = y_train.iloc[train_index], y_train.iloc[val_index]
        model.fit(X_tr, y_tr)
        y_pred = model.predict(X_val)
        mae_list.append(mean_absolute_error(y_val, y_pred))
    return np.mean(mae_list)
study_rf = optuna.create_study(direction='minimize', sampler=optuna.sa
mplers.TPESampler(seed=RANDOM_SEED))
study_rf.optimize(objective_rf, n_trials=50)
# predictions using RF
best_params_rf = study_rf.best_params
rf_model = RandomForestRegressor(**best_params_rf, random_state=RAN
DOM_SEED)
rf_model.fit(X_train, y_train)
y_pred_rf = rf_model.predict(X_test)

# output the optimal hyperparameters
for key, value in study_rf.best_params.items():
    print(f"{key}: {value}")

# compute SHAP importance for RF
explainer_rf = shap.TreeExplainer(rf_model)
shap_values_rf = explainer_rf.shap_values(X_test)
shap.summary_plot(shap_values_rf, X_test, plot_type="bar")
shap.summary_plot(shap_values_rf, X_test)

```

```

# 2. XGBoost using Optuna
def objective_xgb(trial):
    param = {
        'objective': 'reg:squarederror',
        'max_depth': trial.suggest_int('max_depth', 3, 30),
        'learning_rate': trial.suggest_loguniform('learning_rate', 0.005, 0.5),
        'n_estimators': trial.suggest_int('n_estimators', 50, 500),
        'subsample': trial.suggest_uniform('subsample', 0.5, 1.0),
        'colsample_bytree': trial.suggest_uniform('colsample_bytree', 0.5, 1.0),
        'gamma': trial.suggest_loguniform('gamma', 1e-8, 1.0),
    }
    model = xgb.XGBRegressor(**param, random_state=RANDOM_SEED)
    mae_list = []
    for train_index, val_index in kf.split(X_train):
        X_tr, X_val = X_train.iloc[train_index], X_train.iloc[val_index]
        y_tr, y_val = y_train.iloc[train_index], y_train.iloc[val_index]
        model.fit(X_tr, y_tr, eval_set=[(X_val, y_val)], verbose=False)
        y_pred = model.predict(X_val)
        mae_list.append(mean_absolute_error(y_val, y_pred))
    return np.mean(mae_list)

study_xgb = optuna.create_study(direction='minimize', sampler=optuna.samplers.TPESampler(seed=RANDOM_SEED))
study_xgb.optimize(objective_xgb, n_trials=50)

# prediction using XGBoost
best_params_xgb = study_xgb.best_params
xgb_model = xgb.XGBRegressor(**best_params_xgb, random_state=RANDOM_SEED, early_stopping_rounds=10)
xgb_model.fit(X_train, y_train, eval_set=[(X_test, y_test)], verbose=False)
y_pred_xgb = xgb_model.predict(X_test)

```

```

# output the optimal hyperparameters
for key, value in study_xgb.best_params.items():
    print(f"{key}: {value}")

# compute SHAP importance for XGBoost
explainer_xgb = shap.TreeExplainer(xgb_model)
shap_values_xgb = explainer_xgb.shap_values(X_test)
shap.summary_plot(shap_values_xgb, X_test, plot_type="bar")
shap.summary_plot(shap_values_xgb, X_test)

# 3. ConvLSTM using Optuna
def objective_convlstm(trial):
    filters = trial.suggest_int('filters', 8, 128)
    kernel_size = trial.suggest_categorical('kernel_size', [1, 3])
    learning_rate = trial.suggest_loguniform('learning_rate', 1e-5, 1e-2)
    batch_size = trial.suggest_categorical('batch_size', [8, 16, 32, 64])
    epochs = trial.suggest_int('epochs', 10, 50)
    dropout_rate = trial.suggest_uniform('dropout_rate', 0.0, 0.5)
    model = Sequential()
    model.add(ConvLSTM2D(filters=filters, kernel_size=(kernel_size, ker-
nel_size), activation='relu', padding='same', input_shape=(1, 1, len(fea-
tures), 1)))
    model.add(tf.keras.layers.Dropout(dropout_rate))
    model.add(Flatten())
    model.add(Dense(1))
    model.compile(optimizer=tf.keras.optimizers.Adam(learn-
ing_rate=learning_rate), loss='mse')
    # restructure data for ConvLSTM
    X_train_lstm = X_train.values.reshape((X_train.shape[0], 1, 1, len(fea-
tures), 1))

```



```

mae_list = []
for train_index, val_index in kf.split(X_train_lstm):
    X_tr, X_val = X_train_lstm[train_index], X_train_lstm[val_index]
    y_tr, y_val = y_train.iloc[train_index], y_train.iloc[val_index]
    model.fit(X_tr, y_tr, epochs=epochs, batch_size=batch_size, ver-
bose=0)
    y_pred = model.predict(X_val).flatten()
    mae_list.append(mean_absolute_error(y_val, y_pred))
return np.mean(mae_list)

study_conv_lstm = optuna.create_study(direction='minimize', sampler=opt
una.samplers.TPESampler(seed=RANDOM_SEED))
study_conv_lstm.optimize(objective_conv_lstm, n_trials=20)
best_params_conv_lstm = study_conv_lstm.best_params
conv_lstm_model = Sequential()
conv_lstm_model.add(ConvLSTM2D(filters=best_params_conv_lstm['filters
'], kernel_size=(1, 1), activation='relu', padding='same', input_shape=(1, 1,
len(features), 1)))
conv_lstm_model.add(Flatten())
conv_lstm_model.add(Dense(1))
conv_lstm_model.compile(optimizer=tf.keras.optimizers.Adam(learn-
ing_rate=best_params_conv_lstm['learning_rate']), loss='mse')

# prediction using ConvLSTM
X_train_lstm = X_train.values.reshape((X_train.shape[0], 1, 1, len(fea-
tures), 1))
X_test_lstm = X_test.values.reshape((X_test.shape[0], 1, 1, len(features), 1))
conv_lstm_model.fit(X_train_lstm, y_train, epochs=best_params_conv_lstm
['epochs'], batch_size=best_params_conv_lstm['batch_size'], verbose=1)
y_pred_lstm = conv_lstm_model.predict(X_test_lstm).flatten()

```

```
# output the optimal hyperparameters
for key, value in study_convlstm.best_params.items():
    print(f"{key}: {value}")

# define evaluation metrics
def evaluate_model(y_true, y_pred, model_name):
    mae = mean_absolute_error(y_true, y_pred)
    mse = mean_squared_error(y_true, y_pred)
    rmse = np.sqrt(mse)
    mape = mean_absolute_percentage_error(y_true, y_pred)
    print(f"{model_name} - RMSE: {rmse:.4f}, MAE: {mae:.4f}, MSE: {mse:.4f},
MAPE: {mape:.2f}%")

# calculate evaluation metrics
evaluate_model(y_test, y_pred_rf, "RandomForest")
evaluate_model(y_test, y_pred_xgb, "XGBoost")
evaluate_model(y_test, y_pred_lstm, "ConvLSTM")
```


참고문헌

- 국승용·김라이·박영구·장익훈(2022), 드론을 활용한 농업생산조사 방법 연구, 한국농촌경제연구원.
- 김대준·김수옥·김진희·윤은정(2019), “농장맞춤형 농업기상정보 생산을 위한 소기후 모형 구축”, 한국농림기상학회지, 21(3): 146-157, 한국농림기상학회.
- 김원성(2022), 실측정보를 이용한 딥러닝 기반 마늘·양파 생구무게 예측 Neural Prophet 시차모델, 제2회 농업관측 논문경진대회 자료집: 191-208, 한국농촌경제연구원.
- 김재휘·김치운·노호영(2023), “대파 도매가격 예측모형 고도화 방안”, 농업생명과학연구, 57(4): 143-150, 경상국립대학교 농업생명과학연구원.
- 농촌진흥청(2014), 농작물 생육조사 표준매뉴얼, 농촌진흥청 농촌지원국.
- _____ (2024), “하늘에서 농업을 보다”, 농업위성센터 홍보자료, 농촌진흥청 국립농업과학원 농업위성센터.
- 농촌진흥청 국립농업과학원(각 연도), 농업과학연구사업연보.
- 문태원·박준영·손정익(2020), “합성곱 신경망을 이용한 온실 파프리카의 작물 생체중 추정”, 생물환경조절학회지, 29(4): 381-387, 한국생물환경조절학회.
- 박준 외(2022), “ResNet 기반 작물 생육단계 추정 모델 개발”, 스마트미디어저널, 11(2): 53-62, 한국스마트미디어학회.
- 박혜빈·이예진·박선영(2023), “딥러닝을 활용한 위성영상 기반의 강원도 지역의 배추와 무 수확량 예측”, 대한원격탐사학회지, 39(5): 1031-1042, 대한원격탐사학회.
- 윤병삼·서홍석·진영현(2023), 농업관측 항공촬영 사업 평가, 한국농촌경제연구원.
- 이여진 외(2022), “다변수 LSTM 딥러닝 네트워크를 이용한 육계 시세 예측 모델 연구”, 한국통신학회논문지, 47(12): 2058-2064, 한국통신학회.
- 채철주·이현조·김용기·구현정(2022), “농업 공공 빅데이터를 이용한 머신러닝 기반 생산량 및 판매 수익금 예측”, 스마트미디어저널, 11(4): 19-29, 한국스마트미디어학회.
- 이형용·김준기·최성천·서동주(2023), 마늘·양파 구중 예측모형 개발 연구 -변수 선택법을 중심으로, 한국농촌경제연구원.
- 장일식·박구만(2023), “딥러닝 기반 시설재배 생산량 예측 연구”, 방송공학회논문지, 28(4): 448-546, 서울과학기술대학교 나노IT디자인융합대학원.

- Akkus, C. et al.(2023), Multimodal deep learning, arXiv.
- Alhnaity. B., S. Pearson, G. Leontidis & S. Kollias(2019), Using deep learning to predict plant growth and yield in greenhouse environments, arXiv.
- Bergstra, J., R. Bardenet, Y. Bengio & B. Kégl(2011), “Algorithms for hyper-parameter optimization”, Advances in neural information processing systems, 24, NIPS.
- Bergstra, J., D. Yamins & D. Cox(2013), “Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures”, Proceedings of the 30th International conference on machine learning, 28(1): 115-123, PMLR.
- Khan, M. Y. et al.(2021), “Automated Prediction of Good Dictionary EXamples (GDEX): A Comprehensive Experiment with Distant Supervision, Machine Learning, and Word Embedding-Based Deep Learning Techniques”, Complexity, 2021(1), Wiley.
- Liu, F. T., K. M. Ting & Z. H. Zhou(2008), “Isolation forest”, 2008 eighth IEEE international conference on data mining, IEEE.
- Phung, V. H. & E. J. Rhee(2019), “A high-accuracy model average ensemble of convolutional neural networks for classification of cloud image patches on small datasets”, Applied Sciences, 9(21): 4500, MDPI.
- Shi, C. et al.(2022), “Learning Multiscale Temporal-Spatial-Spectral Features via a Multipath Convolutional LSTM Neural Network for Change Detection With Hyperspectral Images”, IEEE Transactions on Geoscience and Remote Sensing, 60, IEEE.
- Shorten, C. & T. M. Khoshgoftaar(2019), “A survey on image data augmentation for deep learning”, Journal of big data, 6(60): 1-48, Springer.
- Wang, Y. et al.(2019), “A hybrid ensemble method for pulsar candidate classification”, Astrophysics and Space Science, 364(139): 1-13, Springer.
- Watanabe, S.(2023), Tree-structured parzen estimator: Understanding its algorithm components and their roles for better empirical performance, arXiv.

〈온라인 자료〉

농업관측센터 홈페이지(aglook.krei.re.kr), 검색일: 2024. 11. 20.

한국지능정보사회진흥원 AI Hub(<https://aihub.or.kr>), 검색일: 2024. 12. 8.

〈관련 자료 목록〉

농민신문(2024. 10. 31.), ““우주농업시대 활짝”...농진청 농업위성센터 출범”.

Chen, T. & C. Guestrin(2016), “Xgboost: A scalable tree boosting system”,
Proceedings of the 22nd ACM SIGKDD International Conference on
Knowledge Discovery and Data Mining: 785-794, ACM.

Lundberg, S. M., G. G. Erion & S. I. Lee(2018), Consistent individualized
feature attribution for tree ensembles, arXiv.

Kaggle(<https://www.kaggle.com/>), 검색일: 2024. 11. 29.



인공지능(AI)을 활용한
농업관측 고도화 방안 연구
- 항공촬영 및 실측조사를 중심으로

한국농촌경제연구원

전라남도 나주시 빛가람로 601 T.1833-5500 F.061) 820-2211

