

# 기후변화에 따른 과수작물 재배지 변화 예측 연구: 한라봉을 중심으로\*

박혜은\*\* · 이종태\*\*\*

## <목 차>

I. 서론	4.2 2단계 연구결과
II. 선행연구고찰	V. 결 론
III. 연구방법	5.1 시사점
3.1 연구모형	5.2 향후연구방향
3.2 데이터	참고문헌
IV. 연구결과	<Abstract>
4.1 1단계 연구결과	

## I. 서론

지속적인 기후변화와 평균기온의 상승으로 인해 우리나라의 농작물 재배지역 및 재배 품목이 변화하고 있다. 대구지역의 대표적인 농작물인 사과와 감의 경우 2010년 이후 10년간 경작지가 절반 가까이 감소했고, 국내 감자 품종 중 대부분을 차지하고 있는 수미감자는 기후변화로 인한 고온과 집중호우 등에 취약한 적응력을 보였다(조정일 등, 2019). 기후변화시나리오 중 현재 추세대로 온실가스가 배출되는 경우를 가정한 SSP5-8.5 시나리오에 따르면, 한반도는

금세기 말(2081~2100)에는 연평균 기온이 7.0°C 증가하면서 아열대 기후로 변화될 것으로 예상하고 있다. 실제로, 2000년대 초반부터 우리나라에서도 망고, 용과와 같은 열대, 아열대 과수작물 재배가 증가하고 있다(박경원, 2022). 이러한 기후변화에 따른 재배 적합 작물의 변화에 대비하여 유망아열대 작물인 애플망고의 재배기술 책자를 배포하고, 국내산 바나나의 유산균 발효를 통한 음료를 개발(정현진 등, 2023)하면서 환경변화에 선제적으로 대응하기 위한 노력이 지속되고 있다. 특히 과수작물은 장시간 노지에서 노출되는 만큼 기상외존도가

\* 본 연구는 박혜은(2023)의 석사 학위 논문의 일부를 발전시킨 논문임.

\*\* 서울여자대학교 공학석사 [asphe@swu.ac.kr](mailto:asphe@swu.ac.kr)(주저자)

\*\*\* 서울여자대학교 경영학과 부교수, [light4u@swu.ac.kr](mailto:light4u@swu.ac.kr)(교신저자)

매우 높은 산업 분야로서 기후변화에 따른 다양한 작물과 수목에 대한 복합적인 반응을 예측하고, 그에 따른 대응책을 마련하기 위한 연구와 노력들을 요구하고 있다. 이러한 필요성을 반영하여 농촌진흥청은 우리나라 6대 과수 작물인 사과, 배, 복숭아, 단감, 감귤, 포도의 재배 한계선 변화를 기후변화 시나리오를 반영하여 2090년까지 10년 단위로 예측한 결과를 발표했다. 해당 보고서에 따르면 SSP5 시나리오를 적용한 결과 2090년대에는 국내에서 고품질 사과 재배 가능성이 없어지는 것으로 예측되었고, 감귤은 제주도에 집중되어 있던 재배면적이 2005년부터 지속적으로 감소하여, 2070년대엔 남해안 일대로 재배한계선이 상승하고 강원도 해안가와 제주도 중산간 지역으로 재배 가능성이 이동할 것으로 예측되었다. 이와같이, 기후변화에 대응하여 작물의 생산량 및 재배지 변화 예측 연구가 최근들어 더욱 활발하게 진행되고 있다. 반면, 감귤과 같이 제주도를 대표하는 특산물인 한라봉은 재배면적 및 생산량의 지속적인 증가와 더불어 러시아, 미국을 포함한 해외 수출 또한 본격화 되고있는 고부가가치 작물임에도 불구하고 기후변화에 대응하기 위한 관련 연구가 부족한 실정이다. 따라서 본 연구에서는 향후 한라봉 재배가 가능할 것으로 예측되는 지역인 충청북도 진천군, 경상북도 문경시, 강원도 홍천군 세 지역의 기후변화시나리오로 예측된 평균기온 데이터를 활용하여 한라봉 재배 가능 여부를 파악하고자 한다. 2017년도부터 전라남도 나주를 비롯한 전남권에서는 이미 한라봉 재배가 시작되었고, 현 시점엔 충청북도 영동군에서도 한라봉 재배가 이루어지고 있으나, 아직까진 난방비용이 많이 들고 상품성 또

한 제주산 한라봉에 비해 부족한 평가를 받고 있다. 그러나 온난화와 같은 기후변화가 지속된다면, 영동군을 비롯하여 영동군보다 위도가 높은 지역에서도 한라봉 재배가 활발하게 재배 될 것으로 예상된다. 따라서 영동군을 기준으로 위도가 높고, 농작물 재배에 관심이 많은 지역으로 선정하여 연구를 진행했다.

## II. 선행연구고찰

농업 분야는 생산 활동에 있어 기후변화의 직접적인 영향을 받는 산업으로, 기후변화로 인한 식량 생산성 변화, 축산물 생산성 변화, 작물 재배 적지 변화 등을 고려해야 한다(임영아, 2020). 기존 기후변화가 식량 생산에 미치는 영향에 관한 국내 선행연구는 대부분 미곡을 중심으로 진행되었으나(김대준 등, 2012, 이태석 등, 2012) 최근에는 머신러닝, 딥러닝 기법을 활용한 예측 연구가 활발하게 진행되면서 농산물 관련 예측연구에도 적용하는 사례들이 많아지고 있다. 머신러닝 알고리즘을 이용하여 딸기 생산량을 예측한 연구(김나은 등, 2022)나 작물 생산량 예측을 위한 머신러닝 기법 활용에 관한 연구(김세원 등, 2021)가 있고 LSTM과 같은 딥러닝 알고리즘을 활용하여 데이터 확보가 가능한 쌀, 양파, 대파, 애호박, 시금치의 가격을 예측한 연구(신성호 등, 2018)와 같은 사례에서는 높은 예측력을 보였다. 작물 생산량 예측을 위한 머신러닝 기법 활용연구에서는 작물의 품질 향상과 출하량 증대에 있어서 중요한 요인을 찾고, 가장 예측력이 높은 모델로 XGBoost모형을 선정한 사례가 있다(김세원

등, 2021).

한편 농업분야 연구에서 기후변화시나리오를 적용한 예측연구 사례가 증가하는 추세를 보이고있는데, 생물의 변화를 예측한 연구에는 서양금혼초 개화 생물계절 예측연구(김명현 등, 2021), 우리나라 미래 수종 분포 예측 연구(김찬우 등, 2023), 남한의 여름 배추 재배지 변동 예측(신민지 등, 2021) 연구가 있다. 2010년대 들어 매년 반복되는 이상기상 현상에 따른 채소 작물의 피해 양상 및 경감 기술 개발을 위한 연구에서는 RCP-8.5 기후변화 시나리오를 채택하여 기후변화가 배추와 고추의 생장 및 생산량에 미치는 영향을 평가했다(김성겸 등, 2020). 기후변화 시나리오를 반영하여, 우리나라의 주요 과수 기후학적 재배지 변동 예측연구 및 주요 과수작물 재배지 변동지도를 제작한 연구(문경환 등, 2022)도 진행되었다. 기후 데이터를 활용하여 XGBoost, Support Vector Machine, Random Forest와 같은 머신러닝 기법을 적용하여 높은 예측력을 보인 모델을 찾고 콩 재배지 변화까지 예측(Torsoni 등, 2022)한 연구가 있으며, 기후데이터에 토양조건까지 고려한 연구에서는 기존의 모델보다 더 우수한 예측력을 보였음을 증명한 연구(Sonal Agarwal 등, 2021)사례를 통해 농업분야에서의 머신러닝 적용 연구의 발전을 위해서 다양한 데이터 확보가 필요함을 알 수 있다.

이처럼 농업분야에서도 머신러닝, 딥러닝 방법론을 활용한 연구가 활발하게 진행되고 있으며 기후변화시나리오를 반영한 예측연구 또한 증가하는 추세다. LSTM 네트워크를 이용하여 높은 한라봉 가격 예측력을 보인 연구(정대호 등, 2022)가 있긴 하지만, 한라봉의 재배지 및

생산량 예측과 관련된 연구 사례는 드물었다. 따라서 본 연구에서는 기후변화시나리오 데이터에 머신러닝, 딥러닝기법을 적용하여 미래 한라봉 생산량을 예측했다는 점에서 차별성이 있다.

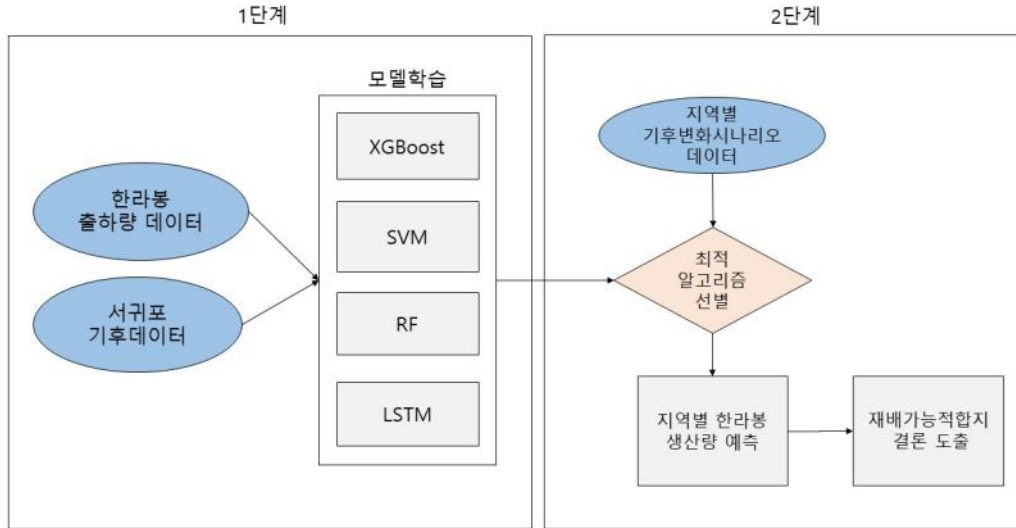
### III. 연구방법

#### 3.1 연구모형

본 연구는 총 2단계로 진행했다. 첫 번째로 제주지역의 한라봉 출하량 데이터와 서귀포지역의 기후 데이터를 활용하여 Random Forest, Support Vector Machine, XGBoost, LSTM과 같은 머신러닝, 딥러닝 기법을 사용하여 생산량 예측을 진행한 뒤 예측력이 가장 높은 알고리즘을 찾는다. 그다음 기상청에서 발표한 기후변화시나리오 데이터를 활용하여 2021년부터 2100년까지 예측된 경북 문경시, 충북 진천군, 강원도 홍천군의 평균기온 데이터 중 2024년부터 2050년까지의 데이터만 추린 뒤 첫 번째 단계에서 선택된 알고리즘을 적용하여 1단계부터 새로 학습하여 선정한 세 지역의 미래 한라봉 출하량을 예측한다.

#### 3.2 데이터

한라봉 출하량 데이터는 ‘제주 데이터 허브’의 ‘신고소별 감귤류 및 만감류 출하량 통계 데이터’를 활용했다. 해당 데이터는 신고소별 감귤류와 한라봉, 천혜향, 레드향과 같은 만감류의 출하량 통계 자료로 톤 단위로 2008년부터



<그림 1> 연구모형

2021년까지 일 데이터로 이루어져 있다. 본 연구에서는 만감류 중 한라봉 출하량 데이터만 추출했다. 제주특별자치도 농업기술원 자료에 따르면 만감류 전체 재배면적 중 한라봉이 68%를 차지하고 있고, 지역별로는 서귀포지역에서 78%가 재배되고 있다. 따라서 연구모형 첫 번째 단계에서 활용할 기후데이터는 서귀포지역의 데이터만 이용했다. 기후 데이터는 기상청 기상자료개방포털에서 중관규모의 날씨를 파악하기 위하여 정해진 시각에 모든 관측소에서 같은 시각에 실시하는 지상관측인 중관기상관측(ASOS)데이터를 활용했다. 기후데이터 필드는 기온(평균기온, 최고기온, 최저기온), 강수량(일강수량), 바람(최대풍속, 평균풍속), 평균이슬점온도, 습도(평균상대습도), 기압(평균증기압), 지면온도(평균지면온도)로 2008년 1월 1일부터 2021년 11월 30일까지 일 데이터로 총 5,083개의 row, 날씨를 포함한 12개의 column으로 이루어져 있다. 1단계에서는 한라봉 출하

량과 기후데이터로 표1과 같은 데이터 테이블을 구성했으며, 실제 활용한 데이터 갯수는 결측치를 제외한 4,468개다.

<표 1> 연구 변수 테이블

#	column	Unit	Non-Null Count	Dtype
0	일시	-	4468 non-null	object
1	출하량총합	ton	4468 non-null	float64
2	평균기온	°C	4468 non-null	float64
3	최저기온	°C	4468 non-null	float64
4	최고기온	°C	4468 non-null	float64
5	일강수량	mm	4468 non-null	float64
6	최대 풍속	m/s	4468 non-null	float64
7	평균 풍속	m/s	4468 non-null	float64
8	평균이슬점온도	°C	4468 non-null	float64
9	평균 상대습도	%	4468 non-null	float64
10	평균 증기압	hPa	4468 non-null	float64
11	평균지면온도	°C	4468 non-null	float64

2단계에서는 기상청에서 제공하는 기후변화 시나리오 SSP-8.5로 예측한 월 평균기온(°C) 데이터를 활용했다. 해당 데이터는 각 시, 군별 면, 읍 단위로 구분되어있어 진천군은 진천읍, 덕산읍, 초평면, 문백면, 백곡면, 이월면, 광혜원면 총 7개 단위 지역의 평균기온(°C), 문경시의 경우 문경읍, 가은읍, 영순면, 산양면, 호계면, 산북면, 동로면, 마성면, 농암면, 점촌1동~5동까지 14개의 단위 지역의 평균기온(°C), 홍천군의 경우 홍천읍, 화촌면, 두촌면, 내촌면, 서석면, 남면, 서면, 북방면, 내면, 동면까지 10개의 단위 지역의 평균기온(°C) 데이터를 평균내어 각 지역별 월 평균기온(°C) 데이터셋을 구축했다.

## IV. 연구결과

### 4.1. 1단계 연구결과

첫 번째 단계에서 학습 데이터에 사용한 전체 인스턴스 수는 4,468개로 그중 80%인 3,574개는 학습모델, 20%인 894개로 모델의 정확성을 테스트했다. 표1에 기술한 10개의 독립변수 중 target value인 ‘출하량총합’ 변수와의 Correlation 분석을 통해 가장 연관성이 높은 변수들만 추렸다.

<표 2> 학습/테스트 데이터셋 정보

구분	건수
Training	3,574
Test	894
합	4,468

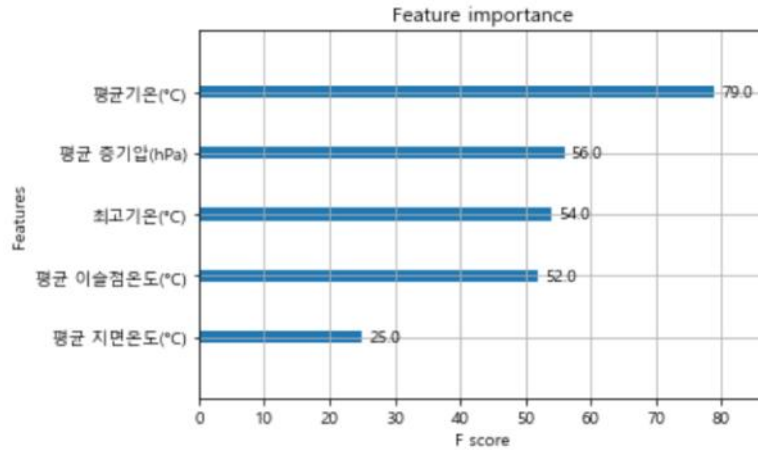
<표 3> Correlation 정보

출하량총합	1.000000
평균기온(°C)	-0.520948
최저기온(°C)	-0.510340
최고기온(°C)	-0.523131
일강수량(mm)	-0.043023
최대 풍속(m/s)	0.02224
평균 풍속(m/s)	-0.022760
평균이슬점온도(°C)	-0.468718
평균 상대습도(%)	-0.205300
평균 증기압(hPa)	-0.425144
평균지면온도(°C)	-0.504882

Correlation 값이 0.2 이하인 변수는 무의미한 변수라고 판단(최인호 등, 2022)했고, 종속 변수인 ‘출하량총합’을 제외한 변수 중 correlation value의 절댓값이 0.3 이상인 변수를 선별한 결과, ‘평균기온(°C)’, ‘최저기온(°C)’, ‘최고기온(°C)’, ‘평균 이슬점온도(°C)’, ‘평균 증기압(hPa)’, ‘평균 지면온도(°C)’ 까지 6개의 독립변수가 선택되었고 feature importance 함수로 변수의 중요도를 측정한 결과는 표4, 그림2와 같다.

<표 4> F Score 정보

변수명	F score
평균기온(°C)	79
평균 증기압(hPa)	56
최고기온(°C)	54
평균 이슬점온도(°C)	52
평균 지면온도(°C)	25
최저기온(°C)	0



<그림 2> feature importance 계산 결과 시각화

선별된 변수 중 평균기온(°C)이 가장 중요도가 높은 변수로 나타났고, 다음으로 중요한 변수로는 온도변수인 최고기온(°C)보다 평균증기압(hPa)으로 나타났다. 반면 최저기온(°C)변수의 중요도가 가장 낮게 나온 이유로는 한라봉의 생육온도 조건과 관련이 있다. 한라봉은 온도가 높은 지역으로 겨울이 따뜻하고 평균기온이 16.5°C 이상, 겨울 최저기온이 -3°C 이하로 내려가지 않는 곳이 적합하다. 1단계에서 구축한 데이터 셋을 살펴보면 최저기온(°C)이 대체로 0°C를 웃돌고 있어 적합, 부적합을 판단하기엔 변별력이 없기 때문인것으로 해석된다.

적용모델은 Random Forest, Support Vector Machine, LSTM 등 가격 및 생산량 예측에서 자주 사용하는 방법론과 함께 XGBoost를 반영하였다. 배성완·유정석(2018), 신인경 등(2021), 김선미·조두연(2022), 김선웅(2023) 등의 선행 연구에서 확인할 수 있는 바와 같이, 가격이나 생산 및 수요량 예측과 같은 정량적 지표의 예측에는 Random Forest, SVM, LSTM이 다양하

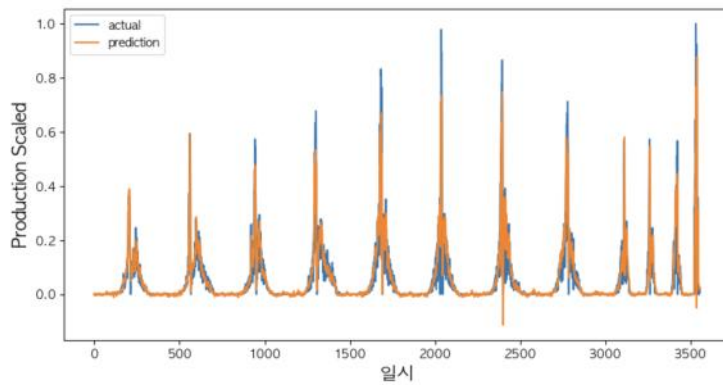
게 활용될 수 있다. 또한, XGBoost는 각종 시계열 데이터 예측연구에서 활발하게 사용되면서 높은 예측력을 보이고 있으며, Deshmukh et al. (2022)이 진행한 작물 수확량 연구에서도 XGBoost 알고리즘이 가장 좋은 예측력을 보인 바 있다. 본 연구에서도 모델 예측 성능 평가는 예측값의 오차를 측정하는 RMSE와 모델의 설명력을 측정하는  $R^2$ (R-squared) 값으로 비교했고, 1단계에서 각 방법론을 활용한 데이터 분석을 수행한 결과, XGBoost 모델의 예측력이 가장 안정적인 것으로 나타났다. 그 결과는 표5와 같다.

<표 5> RMSE 및  $R^2$  정보

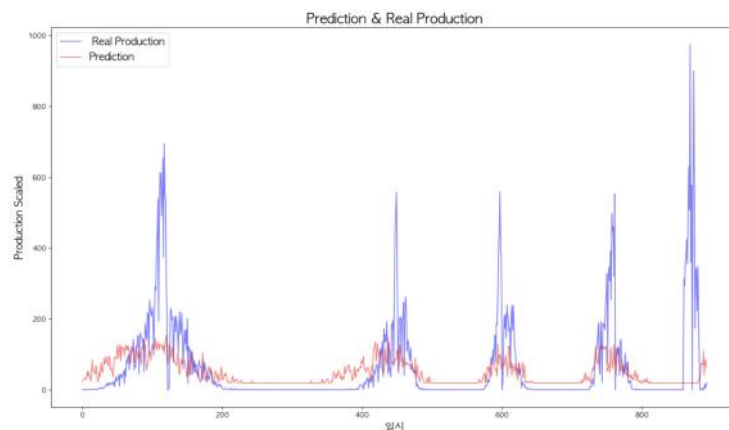
모델	RMSE	Train $R^2$	Test $R^2$
XGBoost	89	0.30	0.26
Random Forest	80	0.89	0.33
LSTM	54	0.012	0.006
SVM	98	0.033	0.044

예측값과의 오차를 보여주는 RMSE값은 LSTM이 54, Random Forest가 80으로 XGBoost보다 더 우수하게 나왔지만, Train  $R^2$  대비 Test  $R^2$ 이 절반 이하로 감소했다는 점에서 과적합(over-fitting) 문제가 있는 것으로 판단했다. 이에 따라, 10-fold 교차 검정을 시행했고(박종우 등, 2022, 이경윤 등, 2019) 각각 RMSE값이 소수점 단위로 소폭 감소하긴 했지만 유의미하게 개선되진 않았다. 이와 비슷한 연구 결과는 Alisson et al. (2023)이 진행한 머신러닝과 기후 데이터를 이용한 콩 재배지 예

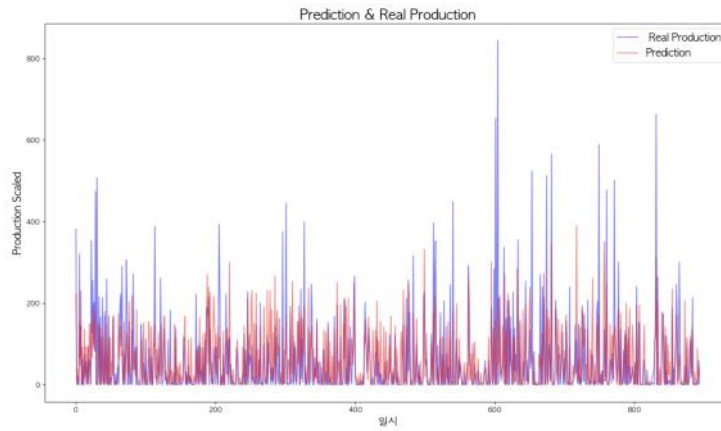
측연구에서도 찾아볼 수 있다. 이와 관련하여, 이홍석 등(2020)의 연구에서는 LSTM은 실시간 비선형적 데이터의 예측에서는 한계가 있을 수 있으며, 이를 해결하기 위한 대안으로 멀티층 LSTM과 같은 변형된 알고리즘이 고려되어야 함을 제시한 바 있다. 반면 SVM의 경우 RMSE값과  $R^2$  모두 가장 예측력이 낮게 측정되어 2단계에서 고려할 알고리즘 대상에서 제외하였다. SVM을 제외한 LSTM과 XGBoost, Random Forest 알고리즘을 적용한 모델의 1단계 예측 그래프는 다음과 같다.



<그림 3> LSTM모델 1단계 예측 결과 그래프



<그림 4> XGBoost모델 1단계 예측 그래프



<그림 5> Random Forest모델 1단계 예측 그래프

1단계 예측 결과는 RMSE값이 가장 낮게 나왔던 LSTM알고리즘을 적용한 모델이 실제값과 가장 비슷한 예측을 보여주고 있지만,  $R^2$ 값과 Train data와 Test data  $R^2$ 의 차이 정도를 종합적으로 고려한 결과 1단계에서 가장 안정적인 예측력을 보여준 알고리즘으로 XGBoost를 선정했다.

두 번째 단계에서는 1단계에서 선택된 XGBoost 알고리즘에 기후변화시나리오로 예측한 평균기온 데이터를 변수로 넣어 미래 한라봉 출하량을 예측했다. 1단계에서 선택했던 6개의 변수 중 feature의 중요도를 나타내는 F score 점수가 가장 높은 평균기온(°C)만으로 학습을 시켰고, 이러한 변화에서도 RMSE값은 89.25에서 89.17로 오히려 소폭 개선된 변화를 보였다. 이를 통해 한라봉 출하량 예측 및 재배 적합지를 예측할 때 가장 우선적으로 고려해야 할 요인은 평균기온(°C)임을 확인할 수 있다.

#### 4.2. 2단계 연구결과

첫 번째 단계에서 제주도지역 한라봉 출하량

데이터와 서귀포지역의 기후 데이터를 이용하여 가장 예측력이 높은 알고리즘이 XGBoost임을 확인하였으며, 이를 바탕으로 수행한 두 번째 단계에서는 기후변화시나리오 SSP5-8.5로 예측한 2024년부터 2050년까지의 평균기온(°C) 데이터를 XGBoost모델로 학습하여 미래 한라봉 출하량을 예측했다(표 6).

2단계 예측 결과에 따르면, 세 지역 모두 2020년대에 비해 2030년대에 출하량이 대체로 증가하지만 2040년대에는 소폭 감소하는 양상을 보일 것으로 예측되었다. 한라봉 생육 온도 조건으로는 겨울이 따뜻하고 연 평균기온이 16.5°C 이상, 겨울 최저기온이 -3°C 이하로 내려가지 않는 곳이 적합하기 때문에, 평균기온이 더 높은 2040년대에 더 높은 출하량을 보일 것으로 예상했던 것과는 달리 2030년대 수준의 평균 온도가 한라봉 재배에 가장 적합한 수준인 것으로 해석할 수 있다. 지역적으로는 가장 위도가 낮고 만감류 재배에 적극적인 관심을 가지고 있어 가장 높은 출하량을 보일 것으로 예상되었던 진천군이 의외로 제일 낮은 출하량



<표 6> 2단계 생산량 예측결과(단위: 톤)

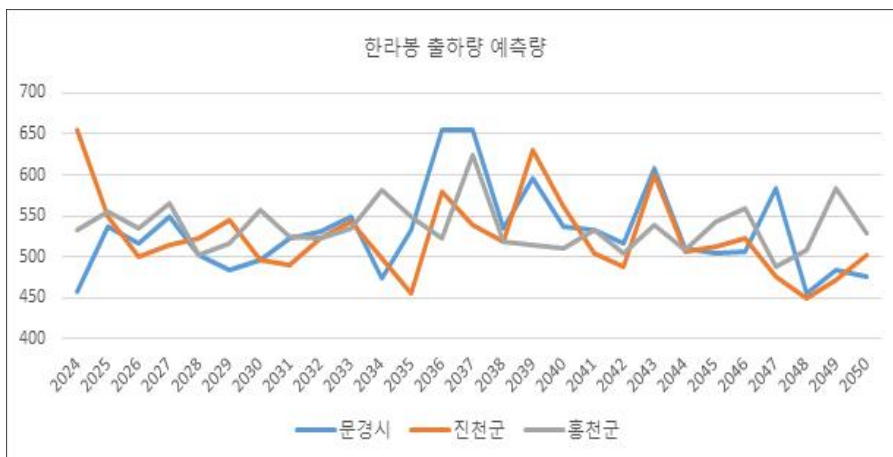
연도	문경시	진천군	홍천군
2024	257	654	532
2025	537	550	556
2026	517	501	534
2027	549	515	565
2028	503	522	503
2029	485	545	517
2030	496	497	558
2031	522	490	525
2032	531	523	522
2033	549	543	535
2034	473	498	582
2035	532	455	549
2036	655	579	523
2037	655	538	625
2038	534	519	519
2039	595	630	515
2040	536	561	511
2041	532	504	532
2042	517	488	505
2043	608	600	538
2044	511	507	509
2045	504	512	543
2046	507	523	560
2047	583	476	488
2048	456	449	509
2049	484	472	583
2050	476	503	529

예측 결과를 보였고, 위도가 제일 높은 홍천군에서 가장 높은 출하량을 보일 것으로 예측되었다. 이러한 결과는 미래 한라봉 재배선이 점진적으로 북상할 가능성이 있는 것으로 해석할 수 있다.

## V. 결 론

### 5.1. 시사점

본 연구는 머신러닝, 딥러닝 기법을 활용하여 한라봉 생산량 예측력이 가장 좋은 알고리즘을 찾고, 기후변화 시나리오를 반영하여 한라봉 생산 가능 예상 지역의 미래 생산량까지 예측했다. 연구는 총 2단계로 진행이 되었으며, 1 단계에서는 모델의 정확도를 높이기 위해 종속 변수와 상관관계가 높은 독립변수만 추린 후 Random Forest, Support Vector Machine, XGBoost, LSTM과 같은 머신러닝, 딥러닝 알고리즘에 적용했다. 각 알고리즘의 예측력은



<그림 6> 한라봉 출하량 예측량 그래프

RMSE와  $R^2$ 을 복합적으로 평가했다. LSTM 알고리즘을 적용한 모델의 경우 RMSE값이 54로 가장 좋았지만, 변수의 설명력을 보여주는  $R^2$ 값이 train  $R^2$  0.012, test  $R^2$  0.006으로 매우 낮았고, Random Forest의 경우 RMSE 80으로 두 번째로 좋은 결과가 나왔지만 train  $R^2$  0.89에서 test  $R^2$ 이 0.33으로 큰 차이를 보이는 것으로 보아 과적합의 문제가 있다고 판단했다. XGBoost의 경우 RMSE 89로 세 번째로 좋은 결과였지만, train  $R^2$  0.30, test  $R^2$  0.26으로 큰 차이 없이 적절하게 예측되었으므로 예측력이 가장 좋은 알고리즘으로 XGBoost를 선정할 수 있었다. 이러한 결과는 최근 김세원 등(2021), Allison et al.(2022) 등의 다양한 최근 연구에서 LSTM이나 SVM 등 그동안 다양하게 활용되었던 분석 방법론 이상으로 XGBoost와 같이 Ensemble화된 방법론들이 우수한 예측력을 보이고 있음을 설명한 것과 같은 맥락으로 이해할 수 있다. 실제 캐글 대회에서도 XGBoost를 포함한 Ensemble화된 방법론들이 예측력에서 높은 성과를 보이고 있다.

둘째, 본 연구에서는 다양한 기후 요인 중에서 만감류 과실을 포함한 농작물의 생산성 예측에서 가장 유의한 영향력을 보인 요인은 평균기온인 것으로 나타났다. 특히, 두 번째 연구 단계에서 평균기온을 제외한 나머지 요인을 예측을 위한 독립변수로 사용하지 않았음에도 예측력 면에서 큰 차이를 보이지 않았던 것을 주지할 필요가 있다. 이것은 평균 온도, 특히 태양 복사에 의해 전달된 열에서 영향을 받기 때문에, 평균기온이 식물 및 토양의 온도 변화와 이에 따른 성장 촉진에서 가장 중요한 요인임을 제시한 Eom et al.(2014)과 이병국 등(2015)의

선행연구와 같은 결론이라고 할 수 있다.

셋째, 선형적 구조가 명확하지 않거나 특정 시기에 특화된 비선형적 데이터를 활용한 예측에서는 LSTM 등의 선형적 알고리즘보다 XGBoost와 같이 Ensemble화된 알고리즘이 더욱 높은 예측력을 보여줄 수 있음을 확인할 수 있다. 이홍석 등(2020)의 연구에서 제시된 바와 같이 선형적 기법에 기초한 LSTM 등의 알고리즘은 다양한 분야에서의 활용성에도 불구하고, Training이 아닌 Testing 및 실무 적용에서는 급격한 예측력 하락 및 데이터의 신뢰성 저하가 발생할 수 있다.

기술적 측면에서의 시사점으로는 본 연구에서 진행한 2단계 모델로 한라봉을 비롯한 만감류 과실의 생산량을 예측할 수 있다는 점에서 활용도가 있으며, 생산량 예측을 통해 생산 가능지역 선별과 같은 영농 의사결정 및 정책 수립에 반영할 수 있다. 특히, 본 연구는 그동안 머신러닝 기법의 활용이 필요함이 강조되었으나 실무에서 적극적으로 활용되지 못했던 농산물 생산 분야에서 XGBoost를 포함한 머신러닝 알고리즘을 활용함으로써 보다 효율적인 농산물 생산 및 유통정책을 마련할 수 있음을 보여주고 있다.

## 5.2. 연구의 한계와 향후 연구방향

본 연구는 기후 데이터를 이용하여 한라봉 생산량을 예측하기 위한 데이터셋을 구축하고 가장 높은 예측력이 높은 알고리즘을 찾은 뒤 미래의 출하량을 예측했다는 점에서 한라봉을 포함한 만감류의 생산 지역 변화를 기계학습 방법론을 기반으로 예측한 초기 연구라는 점에

서 차별점을 가지고 있다. 그러나 본 연구 역시 다음과 같은 한계점들이 존재한다.

첫째. 확보가능한 데이터의 한계로 본 연구의 예측에서는 재배면적이 반영되지 않았다. 실제 농산물은 시범 생산보다 중요한 것이 전체 지역에서 본격적인 출하가 이루어지고, 상품화 단계에 진입하는 시점이 중요하지만, 본 연구에서는 가용 데이터의 범위와 특성 때문에 한계를 가지고 있다. 만감류 감귤의 시범 생산이 제주 지역 외에서 부분적으로 이루어지고 있지만, 아직 이러한 변화가 본격적으로 시장에 반영되지 않은 상태이며, 이에 따라 재배면적 데이터는 아직 관리 대상에 포함되지 않았기 때문이다. 향후 본 연구의 성과를 보다 구체화하기 위해서는 단위 면적 당 생산량에 따라 본격적인 생산 여부를 결정하는 농산물의 특성상 해당 요인에 대한 향후 연구의 수행이 필요하다.

둘째. 앞서 분석한 바와 같이 과수 작물의 생산량 예측에서 가장 의미있는 역할을 하는 변인은 월/일 평균기온이다. 그렇지만, 무조건 높은 온도보다는 지속적인 온도 유지, 상대적으로 크지 않은 온도 변화 폭, 습도 및 광량 등도 검토할 필요가 있다. 그럼에도 불구하고, 현재 우리나라의 농업 관련 공공/민간 기관에서 제공하는 기후 관련 정보는 온도 위주로 제공된다는 한계를 가지고 있다. 본 연구 역시 이러한 한계로 인하여 온도 중심으로 예측을 수행하였다는 한계를 가지고 있다.

셋째. 평균 온도 외 한라봉 생장에 영향을 미치는 생물학적 주요 요인에 대한 변수가 반영되지 않았다. 한라봉은 야간온도 기준 24~25°C가 최적 생육온도이며 백화기부터 낙과가 끝날 때까지 이 온도가 유지되는 것 역시 중요하다.

따라서 한라봉 생산량에 영향을 미치는 온도뿐만 아니라 온도의 유지 기간까지 반영한다면 더 높은 생산량 예측이 가능할 것으로 예상된다. 향후 출하량 및 과수 작물 생산량 예측에 주요한 영향을 미치는 요인들을 추가하고, 재배면적을 고려하여 면적대비 생산량을 계산한다면 생산량 예측에서 더 나아가 지역별로 한라봉 재배 적합 여부까지 판단할 수 있을 것으로 기대한다.

## 참고문헌

- 고성보, 김배성 “한라봉 수급전망 모형 개발 연구” 한국산학기술학회, Volume 13, Issue 11, 2012, pp. 5163-5168
- 김명준, 박영호, 김태규, 정재석 “오차항 러닝 기법을 활용한 예측진단 시스템 개선 방안 연구” 품질경영학회지, 47(4), 2019, pp. 783-793
- 김명현, 최순균, 김민경, 어진우, 엽소진, 방정환 “기후변화 시나리오에 따른 서양금혼초 개화 생물계절 예측”, 한국환경농학회, 학술발표논문집, 2021, pp. 287-288
- 김성겸, 이희주, 위승환 “기후변화 대응을 위한 채소 분야 적응 연구” 한국원예학회, 한국원예과학기술지 제38권 별호[초록집], 2020, pp. 51-51
- 김세원, 김영희 “작물 생산량 예측을 위한 머신러닝 기법 활용 연구” 한국산학기술학회, 제22권, 제7호, 2021, pp. 403-408
- 김찬우, 김지원, 홍민아, 송철호, 이우균 “RCP, SSP 기후변화 시나리오를 활용한 우

- 리나라 미래 수종 분포 예측 연구”. 한국기후변화학회지, Vol. 14, No.2, 2023, pp. 119-134
- 농정연구센터, 농정연구 73호, “농업부문의 기후변화 대응, 현황과 과제”, 임영아, 2020
- 농촌진흥청, “온난화로 미래 과일 재배 지도가 바뀐다 - 기후변화 시나리오(SSP5)” 문경환, 2022
- 민경태 “LSTM 기반 스마트 팜 데이터를 활용한 파프리카 생산량 예측 모델”, 강원대학교 석사학위논문, 2021
- 박해원, 오상현, 소남호, 박지인, 김근아, 정재원 “사과 주산지 변화 모형 추정”, 한국원예학회, 한국원예과학기술지 제 40권 별호 I, 2022, pp. 98-99
- 방경원 “기후변화로 인한 아열대 작물 연구 현황 및 전망”, 식품산업과 영양, 27(1), 2022, pp. 20-26
- 신민지, 이성은, 문경환 “신 기후변화 시나리오 SSP에 근거한 남한의 여름배추 재배지 변동 예측” 한국원예학회, 원예과학기술지 제39권 별호II, 2021, pp. 93-93
- 신성호, 이미경, 송사광 “ LSTM 네트워크를 활용한 농산물 가격 예측 모델”, 한국콘텐츠학회, Volume 18, Issue11, 2018, pp. 416-429
- 이병국, 정필균, 이우균, 임철희, 엄기철, “기온과 작물 잎 및 토양 표면온도의 변화양상 분석”, 한국기후변화학회, 제 6권, 제3호, 2015, pp. 209-221
- 이세연, 양현정, 김민영, 김준경, 손아영, 홍성훈 “스마트팜 활용을 위한 BI-LSTM 기반의 토마토 생산량 예측에 관한 연구”, 한국통신학회, 제48권, 제 4호, 2023, pp. 457-468
- 이태석, 최진용, 유승환, 이상현, 오윤경 “기후변화 시나리오에 따른 미래 논벼의 소비수량 및 생산량 변화 분석”, 한국농공학회, Volume 54 Issue 1, 2012, pp. 47-54
- 이홍석, 부이 각 남, 선충녕 “도심지 교통흐름 및 미세먼지 예측을 위한 딥러닝 LSTM프레임워크” 정보과학학회, 제 47권, 제 3호, 2020, pp. 292-297
- 임정주, 김태완, 임지섭, 김준호, 유태용, 이원주 “A Design and Implement of Efficient Agricultural Product Price Prediction Model”, 한국컴퓨터정보학회, Volume 27 Issue 5, 2022, pp. 29-36
- 정대호, 조영열, “LSTM 네트워크를 이용한 한라봉 가격 예측 모델”, 원예과학기술학회, Vol.40, Issue 5, 2022, pp. 571-577
- 정형진, 김도영, 오임경 “기후변화 대응 국내산 바나나 유산균 발효를 통한 음료 개발” 한국식품과학회, Vol 55, No. 2, 2023, pp. 165-172
- 조용장, 유제혁, 황인준 “미래 기후변화에 따른 양상불 모델 기반 침엽수 서식지 분포 예측”, 한국정보과학회, Vol. 36, No. 3, 2020, pp. 3-19
- 조정일, 이윤호, 상완규, 서명철, 김준환, 신평, 백제경 “지구온난화에 따른 기후변화가 감자의 생육과 수량이 미치는 영향 평가”, 한국작물학회, 추계학술대회, PA-78, 2019, pp. 116-116

최인호, 이원영, 은범진, 허정숙, 장광현 “Deep Neural Network를 활용한 초미세먼지 농도 예측에 관한 연구”, 한국환경영향평가학회, Vol. 22, No. 2, 2022, pp. 83-94

Alisson Gaspar Chiquitto, Jose Reinaldo da Silva Cabral Moraes, Glauco de Souza Rolim “Soybean yeild prediction by machine learning and climate”, *Theoretical and Applied Climatology*, 151 , 2023, pp. 1709-1725

Eom KC, Lee BK, Kim YS, Eom HY. “Comparison of the chane in daily air and surface temperature of redpepper leaf and soil”. *KJSSSF*, Volume 47, No. 5, 2014, pp, 1-6.

Femilda Josephin Joseph Shobana Bai, Kaliraj Shanmugaiah, Ankit Sonthalia, Yuvarajan Devarajan, Edwin Geo Varuvel “Application of machine learning algorithms for predicting the engine characteristics of a wheat germ oileHydrogen fuelled dual fuel engine” *International Journal of Hydrogen Energy*, Volume 48, Issue 60, 2023, pp. 23308-23322

Sonal Agarwall and Sandhya Tarar, “A hybrid appoach for crop yeild prediction using machine learning and deep learning algorithms” *Journal of Physics*, Conf. Ser. 1714, 2021, 012012

**박혜은 (Park, Hye Eun)**



서울여자대학교 공학석사  
관심분야: 머신러닝, 데이터분석, 빅데이터 활용연구

**이종태 (Lee, Jong Tae)**



서울대학교에서 학사와 석사, 한국과학기술원(KAIST)에서 박사 학위를 취득하였으며, 현재 서울여자대학교 경영학과 부교수로 재직 중이다. 주요 관심분야는 IT서비스개발, 정보화 전략, 데이터마케팅 기반 전략 수립 등이다.

<Abstract>

## **Research on predicting changes in crop cultivation areas due to climate change: Focusing on Hallabong**

Park, Hye Eun · Lee, Jong Tae

### **Purpose**

The purpose of this study is to use climate data to find the algorithm with the highest Hallabong production prediction ability and to predict future Hallabong production in areas where Hallabong cultivation is expected to be possible.

### **Design/methodology/approach**

The research is conducted in two stages. In the first step, find the algorithm with the highest predictive power among XGBoost, Random Forest, SVM, and LSTM methodologies. In the second stage, the algorithm found in the first stage is applied to predict future Hallabong production in three regions where Hallabong production is expected to be possible.

### **Findings**

As with many prediction studies, we found that XGBoost showed the highest prediction power. Even in areas where Hallabong production is expected to be possible, Hallabong production was predicted to be highest in Hongcheon, Gangwon-do, which has the highest latitude.

**Keyword:** Machine learning, Deep learning, XGBoost, LSTM, Random Forest

\* 이 논문은 2023년 11월 19일 접수, 2023년 12월 19일 1차 심사, 2024년 1월 11일 게재 확정되었습니다.