



머신러닝을 활용한 아까시나무의 개화 및 낙화일 예측모델 개발

오성준, 심근섭*, 박장현

(주)스마트팜 에이전트

Predicting Blooming and Fall Dates for *Robinia pseudoacacia* Using Machine Learning

Sungjun OH, Keunseob Shim* and Janghyun Park

Co. SmartFarmAgent, Anyang 14093, Republic of Korea

Abstract

In recent years, due to the effects of rapid climate change, the blooming and falling seasons of acacia trees distributed on the Korean Peninsula have changed significantly, causing great damage to the beekeeping industry. Therefore, this study aimed to find a prediction model that predicts the actual blooming and falling season of acacia trees with a small margin of error by using a total of 27 variables, including weather and location data, and machine learning to predict the blooming and falling season of acacia trees observed at five locations nationwide over the past 10 years (2015~2024). Three models were used: RandomForest, Support Vector Machine (SVM), and CatBoost, and various hyperparameter combinations were created and applied to the models using GridSearchCV. The prediction results showed that RandomForest and Support Vector Machine (SVM) performed well, with an error range of 0.1~2.66 days.

Keywords

Robinia pseudoacacia, Blooming, Petals falling

서 론

최근 급격한 기후 변화와 이상 기후 패턴의 영향으로 식물의 개화 및 낙화 시기가 점차 불확실해지고 있다. 특히 아까시나무(*Robinia pseudoacacia*)는 생태계와 경제적 가치가 높은 밀원수 종으로, 우리나라 꿀 생산량의 70% 내외를 차지하는 등 여러 면에서 중요한 역할을 한다(Kim *et al.*, 2021a). 따라서 아까시나무의 개화 및 낙화 시기 예측은 농업 및 환경 관리 측면에서 중요한 연구 주제로 부각되고 있다. 지금까지 아까시나무 개화 시기에 대한 연구는 기온 기반의 통계 모델(Kim *et al.*, 2021b)이나 기후 시계열 분석을 활용한 예측 연구(Son *et al.*, 1999) 등이 주를 이루어 왔다. 그러나 기존 연구들은 주로 기온, 강수량 등 주요 기후 요소에만 초점을 맞추고 있어, 예측 정밀

도를 높이기 위해서는 보다 다양한 변수의 고려가 요구된다. 실제로 개화 시기는 적산온도, 일조량, 강수량의 시기적 분포, 계절별 기후 패턴 등 복합적인 요인의 영향을 받는다. 이러한 파생 변수들은 식물의 성장과 생리 반응에 핵심적인 역할을 하므로, 보다 정교한 분석과 예측을 위해 이들 요소를 통합적으로 반영하는 접근이 필요하다. 최근에는 이러한 복합 요인의 중요성을 반영한 다변량 예측모델에 대한 연구가 증가하고 있으며, 이는 지역적 기후 특성과 지형적 요인까지 고려하는 방향으로 확장되고 있다(Lee *et al.*, 2020; Cho *et al.*, 2021). 특히, 지역 간 기후 편차가 큰 우리나라의 경우, 지역 맞춤형 모델이 개화 예측의 정확도를 높이는 데 효과적일 수 있다. 머신러닝은 다양한 기후 및 지리적 특성 데이터를 학습하여 패턴을 도출하고 정밀한 예측을 가능하게 하는 강력한 도구로, 개화

시기 예측에서도 활발히 활용되고 있다(Taylor and White, 2020). 위도, 경도, 고도, 기온, 강수량, 일조량 등의 기초 변수뿐만 아니라, 적산온도와 같은 파생 변수까지 통합적으로 분석함으로써 아까시나무 개화 시기의 예측 정확도를 높일 수 있다(Yun *et al.*, 2001). 한편, 기존 연구에서는 낙화일에 대한 직접적인 관측값이 부족하여, 만개일 기준 일정 일수 후(예: 4일 또는 6일 후)를 낙화일로 간주하는 방식이 일반적으로 사용되었다(Kim *et al.*, 2021b). 그러나 이러한 방식은 낙화일 예측의 독립성을 저하시켜, 예측모델이 만개일의 단순한 선형 결합에 머무는 한계를 지닌다. 이에 따라 본 연구에서는 낙화일 예측을 주요 목표로 설정하기보다는 보조 지표로 활용하고, 개화 시기 예측에 보다 초점을 맞추는 전략을 채택하였다. 이러한 배경을 바탕으로, 본 연구는 기후 및 지리적 특성을 반영한 지역별 아까시나무 개화 예측모델을 구축하고자 하였다. 이를 위해 국내 5개 지역을 대상으로, 다양한 파생 변수와 기후 데이터를 포함한 데이터셋을 기반으로 머신러닝 기법을 적용하였다. 예측모델로는 머신러닝 기반 회귀 기법 중 랜덤포레스트(RandomForestRegressor, RFR), 서포트 벡터 회귀(Support Vector Regressor, SVR), 그리고 캣부스트(CatBoostRegressor) 세 가지를 선정하였다. 각 모델은 서로 다른 학습 방식과 해석 가능성을 지니며, 예측 특성에도 차이를 보인다. 랜덤포레스트는 다수의 결정 트리를 기반으로 예측을 수행하는 앙상블 학습 기법으로, 과적합(overfitting)을 방지하면서도 안정적인 예측 성능을 제공한다(Breiman, 2001). 또한 변수 중요도(feature importance)를 제공함으로써, 개화 및 낙화 시기에 영향을 미치는 주요 기후 및 지리적 요인을 식별할 수 있는 해석적 장점이 있다. 단, 많은 수의 트리를 학습해야 하므로 연산 자원이 많이 소요될 수 있다. 서포트 벡터 회귀는 커널 함수를 이용해 입력 변수 간의 복잡한 비선형 관계를 효과적으로 모델링할 수 있는 알고리즘이다(Vapnik, 1995). 고차원 공간에서의 마진 최적화를 통해 예측 성능을 향상시킬 수 있으나, 커널 및 하이퍼파라미터의 선택에 따라 성능이 크게 좌우되며, 대규모 데이터셋에서는 계산 효율성의 한계가 존재한다. 캣부스트는 부스팅(Boosting) 알고리즘 기반 모델로, 범주형 변수 처리에 특화되어 있으며 변수 간의 상호작용을 효과적으로 학습할 수 있다(Prokhorenkova *et al.*, 2018). 다양한 기후 및 지리 정보를 포함한 본 연구의 데이터셋에 적합하며, 결측치와 노이즈에 대한 견고한 성능을 보이는 장점이 있다. 다만, 학습 속

도가 비교적 느릴 수 있는 점은 고려가 필요하다. 따라서 본 연구는 랜덤포레스트, 서포트 벡터 회귀, 캣부스트 등 세 가지 머신러닝 회귀 모델을 활용하여, 국내 5개 지역을 대상으로 아까시나무의 개화 및 낙화 시기를 예측하고, 기후 및 지리적 요인을 반영한 최적의 예측모델을 도출하는 것을 주요 목적으로 한다. 이를 통해 기후 변화에 따른 개화 시기 변화를 보다 정밀하게 파악하고, 예측 정보를 양봉 농가에 실질적으로 활용할 수 있는 기반 자료로 제공하고자 한다. 또한, 다양한 기후 및 지형 변수를 반영한 지역 맞춤형 예측 접근은 향후 다른 주요 수종의 생육 예측모델 개발에도 적용 가능한 확장성을 지닌다.

재료 및 방법

1. 예측모델에 적용 자료

본 연구에서는 2015년부터 2024년까지 10년간 아까시나무 개화 및 낙화 시기 예측 최적 모델을 찾기 위해 연도별로 총 10개의 데이터셋을 활용하였으며, 연구 대상 지역은 광주, 대구, 전주, 청주, 포항으로 총 5개 지역을 선정하였다. 아까시나무 개화 데이터는 기상청 기상자료개방포털의 계절 관측 자료를 활용하였고, 예측모델 학습에 활용한 27개의 기상 관측 자료 및 관측소 관련 데이터는 기상청 기상자료개방포털의 종관기상관측(Automated Synoptic Observing System, ASOS)과 방재기상관측(Automatic Weather System, AWS)의 데이터를 융합하여 생성하였다(Table 1). 계절 관측 자료에는 개화 시작일과 만화일 정보는 존재하였으나, 낙화 시작일과 종료일 데이터는 제공되지 않았다. 이에 따라, 기존 연구(Kim *et al.*, 2021b)를 참고하여 만화일 기준 4일 후를 낙화 시작일로, 낙화 시작일 기준 2일 후를 낙화 종료일로 계산하였다. 본 연구에서 활용한 주요 변수로는 각 지역의 월별 평균 기온, 월별 일사량, 월별 평균 강수량, 적산온도

Table 1. Latitude longitude altitude information for five regions

Region	Longitude (°E)	Latitude (°N)	Elevation (m)
Gwangju	126.89	35.17	70.28
Daegu	128.65	35.88	54.27
Jeonju	127.11	35.84	60.44
Cheongju	127.44	36.64	58.7
Pohang	129.38	36.03	3.94

Table 2. Presents the average seasonal temperature and precipitation for the five selected regions (Gwangju, Daegu, Jeonju, Cheongju, and Pohang) over the period from 2015 to 2024

Region	Wintertime (December, January, February)		Springtime (March, April)	
	Average precipitation (mm)	Average temperature (°C)	Average precipitation (mm)	Average temperature (°C)
Gwangju	1.205	2.861	2.54	11.827
Daegu	0.741	2.328	2.006	12.1
Jeonju	1.237	1.882	2.146	11.1
Cheongju	0.909	0.654	1.752	11.358
Pohang	1.059	3.757	2.341	12.268

The seasonal data is categorized into wintertime (December, January, and February) and springtime (March and April), with values representing the mean temperature (°C) and total precipitation (mm) for each season.

(Winkler scale), 계절별 평균 기온 및 강수량 등이 포함되었다(Walkovszky, 1998). 이 중 적산온도는 하루 평균 기온이 5°C 이상일 때의 기온을 누적하여 계산하였으며 계절별 평균 기온 및 강수량에서 계절은 겨울철(12~2월), 봄철(3~4월)로 나뉘서 계산하였다. 겨울철 평균강수량은 전주 1.237 mm, 광주 1.205 mm, 포항 1.059 mm 등의 순으로 많았으며, 겨울철 평균온도는 포항, 광주, 대전 등의 순으로 높게 나타났다. 봄철의 평균온도는 지역 간 큰 차이가 없는 것으로 나타났다(Table 2).

2. 데이터 전처리 및 모델 선정

수집된 기상 데이터의 결측값은 인접한 관측값의 평균으로 대체하였으며, 정상 범주를 벗어난 이상치는 제거하였다. 모델은 아까시나무의 개화 및 낙화일 예측에 적합하다고 판단되는 랜덤포레스트(RandomForestRegressor, RFR), 서포트 벡터 회귀(SupportVectorRegressor, SVR), 캣부스트(CatBoostRegressor) 총 세 가지로 선정하였다. 각 지역별로 데이터의 특성이 다르기 때문에, 각 지역에 대해 별도의 하이퍼파라미터 최적화를 수행하였다. 하이퍼파라미터 최적화는 교차 검증(Cross Validation)과 그리드 서치(GridSearch)를 통해 진행하였으며, 이 과정에서 테스트 데이터(2024년 데이터)는 교차 검증에 포함되지 않고 별도로 분리하여 모델의 일반화 성능을 평가하는 데 사용하였다. 특히, 랜덤포레스트 모델의 경우 트리의 개수(n_estimators), 최대 깊이(max_depth), 노드 분할 최소 샘플 수(min_samples_split), 리프 노드 최소 샘플 수(min_samples_leaf)를 최적화하였으며, 서포트 벡터 머신 모델

은 규제 파라미터(C), 커널(kernel), 감마(gamma)를, 캣부스트 모델은 트리 깊이(depth), 학습률(learning_rate), 반복 횟수(iterations)를 최적화하였다.

3. 모델 학습 및 평가

학습은 2015년부터 2023년까지의 데이터를 학습 데이터로, 2024년 데이터를 테스트 데이터로 사용하여 모델의 예측 성능을 검증하였다. 모델 성능 평가는 실제 개화 및 낙화일과의 차이를 측정하는 평균제곱오차(Mean Squared Error, MSE)와 설명계수(R^2 , Coefficient of Determination)를 사용하였다. 데이터 수집, 전처리, 모델링 및 평가는 전부 파이썬(Python) 프로그래밍 언어를 사용하여 진행하였다.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (\text{Eq. 1})$$

n : 데이터 셋의 샘플 개수로 평가에 포함된 관측값의 총 개수

y_i : 실제 관측 값으로 i 번째 샘플에 해당하는 실제 개화일 또는 낙화일 데이터

$\hat{y}_i$: 모델이 예측한 개화일 또는 낙화일의 예측값으로 i 번째 샘플에 대해 모델이 예측한 결과를 의미

$\bar{y}$: 실제 관측 값들의 평균

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (\text{Eq. 2})$$

n : 데이터 셋의 샘플 개수로 평가에 포함된 관측값의 총 개수
 y_i : 실제 관측 값으로 i 번째 샘플에 해당하는 실제 개화일 또는 낙화일 데이터
 $\hat{y}_i$: 모델이 예측한 개화일 또는 낙화일의 예측값으로 i 번째 샘플에 대해 모델이 예측한 결과를 의미
 $\bar{y}$: 실제 관측 값들의 평균

결 과

총 5개 지역에 대해서 적산온도, 기온, 강수량, 계절별 기상 데이터 등 기후 및 위도, 경도, 고도를 나타내는 지리적 특성 데이터 총 27개의 변수를 투입하여 상관관계 분석과

개화 시작일, 만화일, 낙화 시작일 및 낙화 종료일을 예측 진행하였다 (Fig. 1, Table 3). 예측 결과는 절대오차가 0.1~2.66일 정도로 매우 좁은 오차 범위 내의 예측 정확도를 보였으며, 특히 개화 시작일에 대해 매우 높은 예측 정확도를 보인 것을 확인할 수 있었다. 또한 Random ForestRegressor, SVR, CatBoostRegressor 모델을 비교한 결과, CatBoostRegressor의 경우 예측 오차 범위가 3~4일로 성능이 다소 떨어진 반면, RandomForestRegressor와 SVR 모델은 상대적으로 높은 예측률을 보였다. 특히 지역별로 분석한 결과, 광주와 대구, 포항에서는 SVR 모델이 우수한 예측 성능을 보였으며, 전주와 청주 지역에서는 RandomForestRegressor가 더 정확한 예측 결과를 나타냈다. 이를 통해, 하나의 통합된 모델 대신 지역별 특화된 모

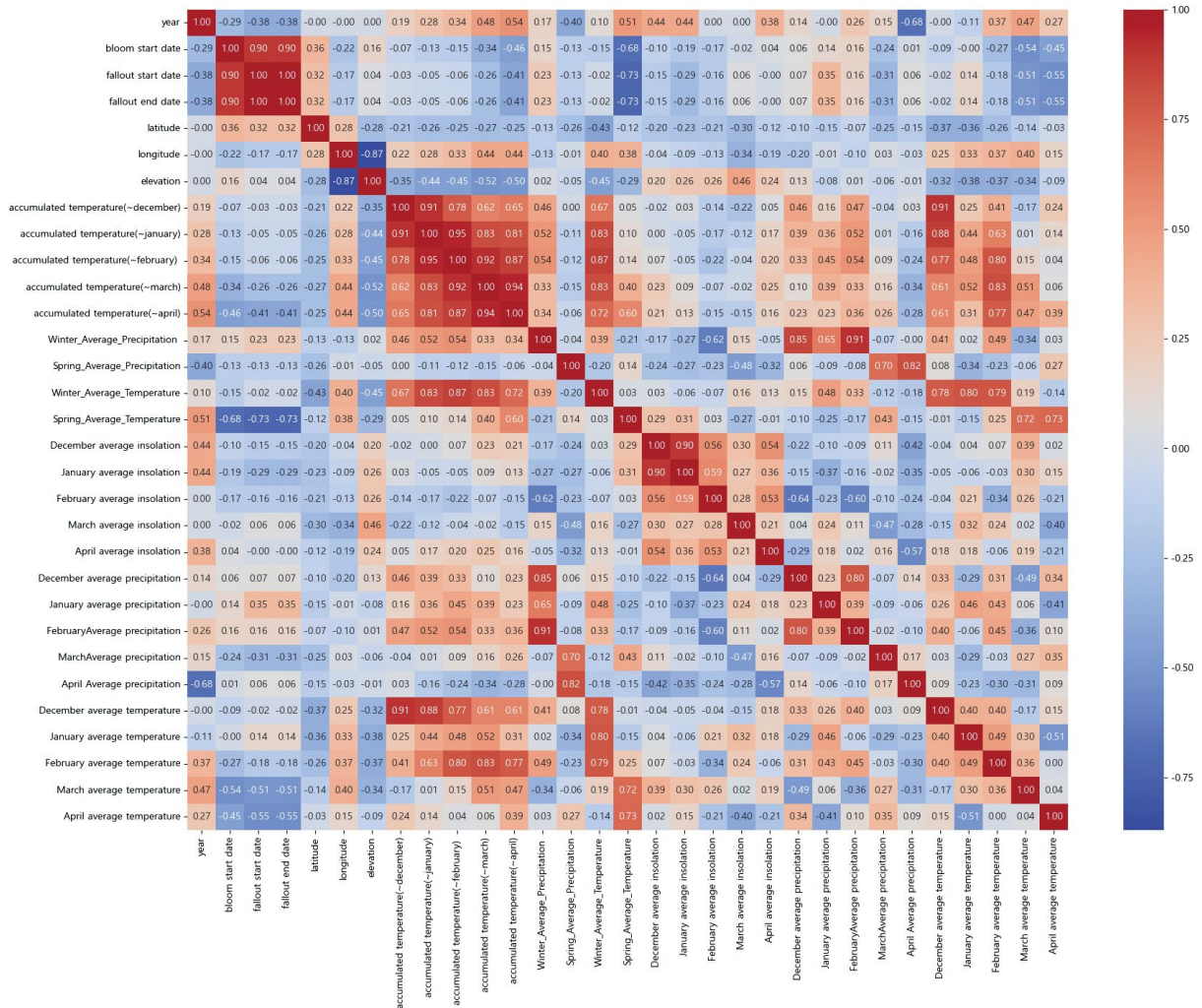


Fig. 1. Visualize correlations in data with heatmaps (A correlation heatmap is a graphical representation that uses color to indicate the strength and direction of relationships between multiple variables. Darker colors typically signify stronger correlations).

Table 3. Each model (RandomForestRegressor, SupportVectorRegressor, CatBoostRegressor) with their predicted values, actual values, absolute errors, and best performing parameters, and R^2 for each of the blooming start, full bloom, fall start, and fall end dates

Region	Target	Model	Predicted value	Actual value	Absolute error	Best params	R^2
Gwangju	start	SVR	112.87	114	1.13	C: 0.1, gamma: auto, kernel: poly	0.655
	full	SVR	120.24	118	2.24	C: 1.0, gamma: scale, kernel: rbf	0.733
	fall	SVR	124.24	122	2.24	C: 1.0, gamma: scale, kernel: rbf	0.793
	end	SVR	126.24	124	2.24	C: 1.0, gamma: scale, kernel: rbf	0.802
Daegu	start	SVR	116.86	118	1.14	C: 1.0, gamma: scale, kernel: rbf	0.681
	full	SVR	121.05	123	1.95	C: 100, gamma: scale, kernel: rbf	0.574
	fall	SVR	125.05	127	1.95	C: 100, gamma: scale, kernel: rbf	0.608
	end	SVR	127.05	129	1.95	C: 100, gamma: scale, kernel: rbf	0.514
Jeonju	start	RFR	119.1	119	0.1	max_depth: None, min_samples_leaf: 1, min_samples_split: 2, n_estimators: 100	0.667
	full	RFR	126.66	124	2.66	max_depth: None, min_samples_leaf: 1, min_samples_split: 2, n_estimators: 100	0.869
	fall	RFR	130.66	128	2.66	max_depth: None, min_samples_leaf: 1, min_samples_split: 2, n_estimators: 100	0.898
	end	RFR	132.66	130	2.66	max_depth: None, min_samples_leaf: 1, min_samples_split: 2, n_estimators: 100	0.886
Cheongju	start	RFR	120.84	119	1.84	max_depth: None, min_samples_leaf: 1, min_samples_split: 2, n_estimators: 100	0.876
	full	RFR	126.39	124	2.39	max_depth: 30, min_samples_leaf: 1, min_samples_split: 5, n_estimators: 100	0.922
	fall	RFR	130.39	128	2.39	max_depth: None, min_samples_leaf: 1, min_samples_split: 2, n_estimators: 50	0.882
	end	RFR	132.36	130	2.36	max_depth: None, min_samples_leaf: 1, min_samples_split: 2, n_estimators: 100	0.774
Pohang	start	RFR	117.1	117	0.1	max_depth: None, min_samples_leaf: 1, min_samples_split: 2, n_estimators: 100	0.833
	full	SVR	122.05	121	1.05	C: 1.0, gamma: scale, kernel: rbf	0.753
	fall	SVR	126.05	125	1.05	C: 1.0, gamma: scale, kernel: rbf	0.851
	end	SVR	128.05	127	1.05	C: 1.0, gamma: scale, kernel: rbf	0.739

The predicted values and actual values are based on daily predictions, starting from January 1st, with each subsequent value increasing by 1 for each day that passes

텔을 구축하는 것이 각 지역의 특성을 반영하여 더욱 정확한 예측을 위해 필요하다는 점을 확인할 수 있었다. 또한 모델의 성능을 평가한 결과 R^2 값은 0.655에서 0.922 사이에 분포하며, 전반적으로 모델이 데이터의 변동성을 잘 설명하는 것으로 나타났다. 이러한 높은 R^2 값은 예측모델이 실제 관측값과 잘 일치하고 있음을 나타내며, 예측 정확도가 높은 수준임을 확인할 수 있었다.

랜덤포레스트 모델의 변수 중요도 (Feature Importance)를 분석한 결과 광주 지역의 경우 2월 평균 일사량, 1월 평균 강수량, 12월부터 3월까지의 적산온도가 예측모델에 큰 영향을 준 것을 확인하였고, 대구의 경우 1월 평균 기온, 2월

평균 강수량, 봄철 평균 기온 등이 예측모델에 큰 영향을 준 것을 확인할 수 있었다. 이와 같이 지역별 기상정보에 따라 예측모델에서 변수의 영향력이 다르게 나타난다는 것을 확인할 수 있었다 (Fig. 2).

예측모델을 이용하여 도출한 아까시나무 개화 예측 결과는 광주의 경우 개화 예측일이 4월 22일, 실제 개화일이 4월 23일로 나타났으며, 대전의 경우 개화 예측일이 4월 28일, 실제 개화일이 4월 27일로 1일의 차이가 나타난 반면에 전주와 포항지역의 개화 예측은 실제 개화일과 동일하게 예측되었다 (Fig. 3, Table 4).

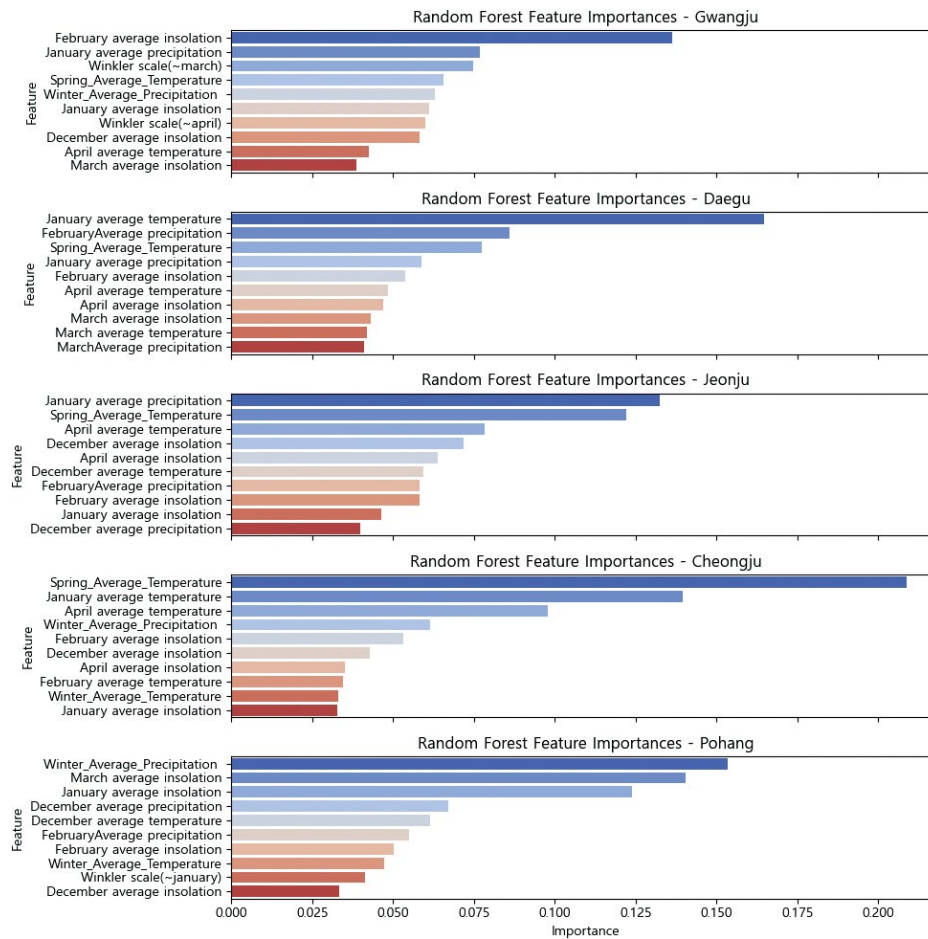


Fig. 2. Random Forest feature importance by region (A higher feature importance score indicates that the corresponding feature plays a more significant role in the model's predictions).

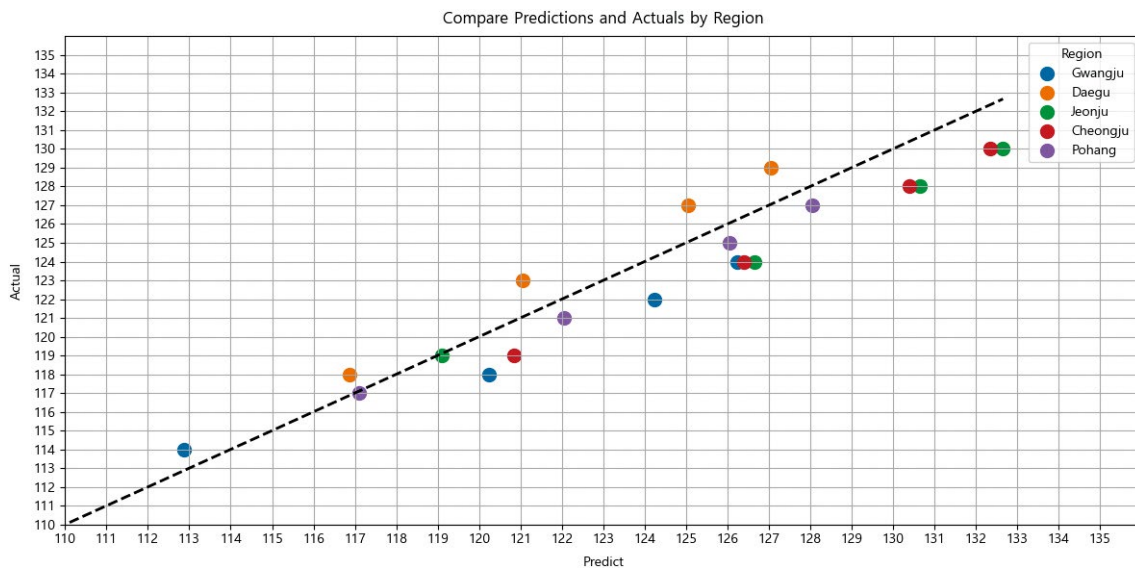


Fig. 3. Scatter plot visualization of actual vs. predicted values.

Table 4. Actual and predicted bloom and fall times by region (2024)

Region	Target	Predicted date	Actual date
Gwangju	start	4.22	4.23
	full	4.29	4.27
	fall	5.3	5.1
	end	5.5	5.3
Daegu	start	4.28	4.27
	full	4.30	5.2
	fall	5.4	5.6
	end	5.6	5.8
Jeonju	start	4.28	4.28
	full	5.5	5.3
	fall	5.9	5.7
	end	5.11	5.9
Cheongju	start	4.30	4.28
	full	5.5	5.3
	fall	5.9	5.7
	end	5.11	5.9
Pohang	start	4.26	4.26
	full	5.1	4.30
	fall	5.5	5.4
	end	5.7	5.6

고 찰

본 연구에서는 우리나라의 기후 및 지리적 특성 데이터를 활용하여 지난 10년간(2015~2024년)의 지역별 아까시나무 개화 및 낙화 시기의 변화를 분석하고, 최적의 예측모델을 도출하였다. 모델 성능 분석 결과, 평균제곱오차(Mean Squared Error, MSE)와 설명계수(R^2)를 주요 성능 지표로 활용하였으며, 전반적으로 최적의 하이퍼파라미터를 적용한 모델이 실제 관측값과 유사한 예측 결과를 보였다. 특히, 개화 및 낙화 시기 예측에서 R^2 값이 높은 모델일수록 입력된 기후 및 환경 요인이 예측 성능에 중요한 역할을 하는 것으로 나타났다. 이러한 연구 결과는 기존 연구와 비교했을 때 몇 가지 중요한 시사점을 제공한다. 기존의 단순 기후 요소 기반 예측 연구(Yun *et al.*, 2001)에서는 주로 기온과 강수량만을 고려하여 개화 시기를 예측했으나, 본 연구에서는 적산온도, 일조량, 지역별 위도·경도 및 고도 등 다양한 기후 및 지리적 요인을 포함함으로써 보다 정교한 예측이 가능하였음을 확인하였다. 특히,

기존 연구에서는 단일 통합 모델을 적용하여 전국 단위로 개화 시기를 예측하는 경우가 많았으나, 단일 모델이 예측 가능성 범위 내에서 충분한 정밀도를 확보할 경우 장기적으로 신뢰도 높은 접근법이 될 수 있다는 점도 고려되어야 한다. 이에 반해, 본 연구는 지역별 기후 특성을 반영한 머신러닝 모델을 도입함으로써, 각 지역의 기후 차이를 보다 정확하게 반영할 수 있음을 보였다. 이는 기존 연구에서 간과되었던 지역별 기후 차이에 따른 개화·낙화 시기의 변동성을 보다 체계적으로 분석할 수 있도록 기여한 점에서 중요한 의미를 가진다. 그러나 본 연구에는 몇 가지 한계점이 존재한다. 첫째, 본 연구에서 사용한 기후 데이터의 기간이 10년(2015~2024년)으로 한정되어 있어, 장기적인 기후 변화에 따른 개화 및 낙화일의 변동성을 충분히 반영하지 못할 가능성이 있다. 기존 연구 중 일부(Kim and Han, 2019)는 30년 이상의 장기 기후 데이터를 활용하여 예측의 신뢰도를 높였으며, 이에 비해 본 연구는 최근 10년간의 변화에 초점을 맞추었다는 점에서 차이가 있다. 따라서 향후 연구에서는 보다 장기간의 데이터를 포함하여 모델의 일반화 성능을 검증할 필요가 있다. 둘째, 개화 및 낙화일에 영향을 미치는 다양한 요인 중 기후 요소 이외의 변수(예: 토양의 영양 상태, 수분 함량, 주변 식생 등)를 충분히 고려하지 못했다는 한계가 있다. 기존 연구에서 일부 생태학적 요소가 개화 시기에 미치는 영향을 강조한 바 있으며, 이를 반영할 경우 더욱 정밀한 예측이 가능할 것으로 예상된다. 향후 연구에서는 이러한 추가 변수들을 포함하여 모델을 개선하는 것이 필요하다. 셋째, 본 연구에서는 미기상(microclimate) 및 미기후(micrometeorology) 데이터를 추가하려고 하였으나, 적절한 데이터를 확보하지 못하여 분석에 포함하지 못했다. 미기상 데이터는 지역 내 미세한 기온·습도 변화까지 반영할 수 있어 더욱 정밀한 예측이 가능하다는 점에서 의미가 크다(Gardner *et al.*, 2021). 향후 연구에서는 이러한 데이터를 확보하여, 보다 상세한 기후 변화를 반영한 모델링을 수행하는 것이 중요하다. 본 연구의 결과는 기후 변화에 따른 개화 및 낙화 시기 변동을 예측함으로써, 양봉업의 생산성 향상뿐만 아니라 생태계 변화에 대한 대응 전략을 마련하는 데 유용한 자료로 활용될 수 있다. 특히, 지역별 기후 특성을 고려한 모델링이 필수적임을 시사하며, 향후 연구에서는 이를 더욱 정교하게 개선할 필요가 있다.

감사의 글

본 연구는 농촌진흥청의 주요 밀원 개화기 예측 이용 벌꿀 생산 최적 모델 개발(과제번호: RS-2023-00230940)에 의해 수행되었습니다.

인 용 문 헌

- Breiman, L. 2001. Random forests. *Mach. Learn.* 45(1): 5-32.
- Cho, J. G., S. Kumar, S. H. Kim, J.-H. Han, C. S. Durso and P. H. Martin. 2021. Apple phenology occurs earlier across South Korea with higher temperatures and increased precipitation. *Int. J. Biometeorol.* 65: 265-276.
- Gardner, A. S., I. M. D. Maclean, K. J. Gaston and L. Bütikofer. 2021. Forecasting future crop suitability with microclimate data. *Agric. Syst.* 190: 103084.
- Kim, D.-I. and D. Han. 2019. Comparative study on long term climate data sources over South Korea. *Journal of Water and Climate Change* 10(3): 504-523.
- Kim, M. J., S. J. Na, H. Y. Kwon and Y. K. Kim. 2021a. The main source of Honey Plants in Korea I (Tree Species). pp. 28. National Institute of Forest Science.
- Kim, S. K., T. K. Kim, S. K. Yoon, K. C. Jang, H. M. Lim, W. Y. Lee, M. S. Won, J. H. Lim and H. S. Kim. 2021b. Recent Changes in Bloom Dates of *Robinia pseudoacacia* and Bloom Date Predictions Using a Process-Based Model in South Korea. *J. Korean Soc. For. Sci.* 3: 322-340.
- Lee, H. K., S. J. Lee, M. K. Kim and S. D. Lee. 2020. Prediction of Plant Phenological Shift under Climate Change in South Korea. *Sustainability* 12(21): 9276.
- Prokhorenkova, L., G. Gusev, A. Vorobev, A. V. Dorogush and A. Gulin. 2018. CatBoost: unbiased boosting with categorical features. *Advances in Neural Information Processing Systems (NeurIPS)* 31.
- Son, J. H., J. G. Kim, K. G. Oh, I. H. Ji and S. C. Park. 1999. Prediction of a Flowering Period and Selection of Trees with High Nectar Production in Black Locust. pp. 95. Ministry of Agriculture and Forestry.
- Taylor, S. D. and E. P. White. 2020. Automated data-intensive forecasting of plant phenology throughout the United States. *Ecol. Appl.* 30(1): e02025.
- Vapnik, V. 1995. The nature of statistical learning theory. Springer Science & Business Media.
- Walkovszky, A. 1998. Changes in phenology of the locust tree (*Robinia pseudoacacia* L.) in Hungary. *Int. J. Biometeorol.* 41: 155-160.
- Yun, J. I., J. Y. Choi and J. H. Ahn. 2001. Seasonal Trend of Elevation Effect on Daily Air Temperature in Korea. *Korean J. Agric. For. Meteorol.* 3(2): 96-104.