


머신러닝 기법을 활용한 주요 침엽수종의 수관급 분류와 간벌목 선정 연구

이용규 · 이정수 · 박진우 *

강원대학교 산림경영학전공

A Study on Classification of Crown Classes and Selection of Thinned Trees for Major Conifers Using Machine Learning Techniques

Yong-Kyu Lee, Jung-Soo Lee, and Jin-Woo Park *

Department of Forest Management, Kangwon National University, Chuncheon 24341, Korea

요약: 본 연구는 효율적인 산림시업계획 수립을 위하여 임목의 측정정보와 머신러닝 알고리즘을 이용하여 주요 침엽수종(소나무, 잣나무, 낙엽송)의 수관급 분류를 목적으로 하였다. 임목의 측정정보는 9년간 수집된 국유림 모니터링 정보를 활용하였으며, 머신러닝 알고리즘은 Random Forest (RF), XGBoost (XGB), Light GBM (LGBM)을 사용하였다. 알고리즘의 정확도, 정밀도, 재현율, F1 score를 이용한 성능평가를 통하여 알고리즘의 정확도를 비교·평가하였다. 분석결과, 소나무림, 잣나무림, 낙엽송 모두 RF 알고리즘이 성능평가 점수가 가장 높았으며, 수종별로는 소나무가 정확도 약 65%, 정밀도 약 72%, 재현율 약 60%, F1 score 약 66%로 성능평가 점수가 가장 높았다. 수관급은 우세목의 정확도가 약 80%이상으로 높았으나, 준우세목과 중간목, 피압목의 분류 정확도는 낮게 평가되었다. 본 연구결과는 산림시업의 간벌목 선정에 있어 의사결정을 위한 참조자료로 활용이 가능할 것으로 판단된다.

Abstract: Here we aimed to classify the major coniferous tree species (*Pinus densiflora*, *Pinus koraiensis*, and *Larix kaempferi*) by tree measurement information and machine learning algorithms to establish an efficient forest management plan. We used national forest monitoring information amassed over nine years for the measurement information of trees, and random forest (RF), XGBoost (XGB), and light GBM (LGBM) as machine learning algorithms. We compared and evaluated the accuracy of the algorithm through performance evaluation using the accuracy, precision, recall, and F1 score of the algorithm. The RF algorithm had the highest performance evaluation score for all tree species, and highest scores for *Pinus densiflora*, with an accuracy of about 65%, a precision of about 72%, a recall of about 60%, and an F1 score of about 66%. The classification accuracy for the dominant trees was higher than about 80% in the crown classes, but that of the co-dominant trees, the intermediate trees, and the overtopper trees was evaluated as low. We consider that the results of this study can be used as reference data for decision-making in the selection of thinning trees for forest management.

Key words: machine learning, crown classes, *Pinus densiflora*, *Pinus koraiensis*, *Larix kaempferi*, thinning planning

서론

산림경영계획은 산림의 경영목적에 따라 현재의 임분상태를 판단하고 향후 변화하게 된 미래의 산림을 계획하는

것이다. 경영계획을 수립하기 위한 방법으로 산림생장 및 수확모델을 활용한다(Choi et al., 2015). 우리나라는 1970년대 치산녹화사업을 시작으로 황폐해진 산림을 복구하여 2020년 기준 ha당 약 165.2 m²의 축적으로 높은 수준을 유지하고 있으나, 영급 구조는 IV영급 이상의 산림이 전체 면적의 약 77.3%를 차지하여 영급불균형을 이루고 있어 적극적인 산림경영활동이 필요한 시점이다(An and Shin, 2004; Korea Forest Service, 2020).

* Corresponding author

E-mail: pjw@kangwon.ac.kr

ORCID

Jin-Woo Park  https://orcid.org/0000-0001-6004-5866

산림경영에 있어 주요한 시업내역은 간벌을 통한 임분의 밀도조절이다(Kwon et al., 2007). 간벌은 보통 임분의 생육공간조절, 성장조절, 성장집중, 혼효조절, 임분수직구조개선, 임연부 보호관리 등의 목적으로 수행되며, 미성숙단계의 유령림에서부터 주벌이 이루어지기 전까지 임분의 목적에 맞도록 유도하는 별채적 조정행위이다(Korea Forest Service, 2007).

산림경영분야에서 임분상태를 파악하고 미래의 산림을 계획하기 위하여 성장모델 연구가 대부분을 이루고 있으며, 이것은 성장변수들을 이용하여 모델식을 조제하고 임분 및 개체목에 대한 성장량을 추정하여 산림경영 기준 및 목표량을 제시하는 연구이다. Kwon and Chung(2004)은 잣나무에 대한 성장모델을 구축하고 이를 통한 개체목단위의 성장모델을 제시하였으며, Kwon(2013)은 개체목의 성장예측 시뮬레이터를 개발하고 이를 기반으로 목재생산 최적화 모델을 개발하였다. 이 외에도 다수 성장모델을 이용한 산림경영 목표 수립을 위한 연구가 진행되었다.

효과적인 산림경영계획의 수립을 위해서는 임분의 성장 및 수확량에 대한 기준제시뿐만 아니라 간벌작업에 대한 기준제시가 필요하다. 일반적으로 간벌작업의 작업 순서에 있어 간벌률 및 간벌본수 결정과 간벌목의 선정(선목작업)이 핵심단계이다. 간벌율과 간벌본수는 임분수확표를 통하여 임령에 따른 임목본수 정보를 확보할 수 있다. 그러나 간벌목의 선정은 임분 내 개체목에 대한 수관급 또는 수형급을 통하여 등급을 정하고 존치목에 대한 형질불량목 및 경쟁목을 제거하는 방법으로 수행된다(Korea Forest Service, 2007). 개체목에 대한 등급화는 수관급 및 수형급에 대한 기준이 있으나, 정성적인 기준으로 작성되어 실질적으로 현장 작업자의 주관적 판단에 의하여 결정되고 있어 간벌목의 선정에 있어서도 정성적인 기준에 의하여 선정이 되고 있다.

효과적인 간벌목의 선정을 위하여 데이터 기반 객관적인 임목의 평가와 분류 방법이 필요하다. 일본에서는 이미 2000년대 초반 산림경영에 있어 의사결정트리 및 신경망 구조를 이용한 간벌목의 선정 알고리즘을 개발·적용하는 연구가 진행되었다(Minowa, 2001; Tasaka et al., 2001; Minowa, 2005). 최근 머신러닝 및 딥러닝 등 AI 모델들이 다양한 분야에 활용되면서 산림분야에서는 수종분류 및 층위구조 분석 등 이미지기반 연구들이 진행되고 있다(Kim et al., 2020; Kwon et al., 2019). AI 기술은 여러 가지 요인을 종합적으로 분석하여 대상을 판단하고 예측하는 방법이며, 그 중 머신러닝은 지도학습 방법으로 주어진 데이터를 클래스별로 분류하는 것이 대표적인 기술이다(Lee, 2019).

따라서 본 연구는 효과적인 간벌목의 선정을 위한 기준 자료가 되는 임목의 수관급 선정을 목적으로 하였다. 객관

적인 수관급의 등급분류를 위하여 임목의 성장요인 데이터를 기반한 머신러닝 알고리즘을 적용하여 임목의 수관급을 분류하고 정확도를 비교·분석을 통하여 최적의 알고리즘 선정을 목적으로 하였다.

재료 및 방법

1. 연구자료

본 연구에서 사용된 주요 침엽수종의 자료는 산림청에서 9년간 진행한 ‘국유림 내 주요수종 간벌효과 장기모니터링’ DB를 활용하였으며, 장기모니터링 자료는 우리나라 북부, 동부, 남부 지방산림청의 주요수종(소나무, 잣나무, 낙엽송)의 성장, 시업 등 간벌시업 모니터링에 의한 지속가능한 산림자원관리를 목표로 하고 있다. 장기모니터링은 총 142개 Site와 426개 Plot 정보를 가지고 있으며, 21,644개의 임목조사 자료를 수집하고 있다. 또한 임목의 기본적인 정보인 수종, 흉고직경 수고, 지하고, 수관폭 등의 정보를 가지고 있으며, 임목의 등급을 판단할 수 있는 수관급 정보가 수집되어 있다. 본 연구는 소나무, 잣나무, 낙엽송 전체 데이터 20,608본 중, 수고, 수관폭, 수관급 등 3,316본의 정보를 활용하였다.

2. 연구방법

본 연구는 주요 침엽수종의 간벌목 선정을 위하여 머신러닝 기법을 활용하여 임목의 수관급을 분류하는 것을 목적으로 하였다. 우리나라 주요 침엽수종에 대한 9년간의 모니터링 데이터를 통하여 소나무, 잣나무, 낙엽송의 데이터를 추출하였으며, 조사항목 중 간벌목 선정인자(흉고직경, 수고, 수관폭) 정보와 수관급 정보를 추출하였다. 추출된 정보는 세 가지 머신러닝 알고리즘(Random Forest, XGBoost, Light GBM)을 이용하여 분석을 실시하였다. 알고리즘에 적용한 학습자료와 검증자료는 7:3의 비율로 적용하였으며, 알고리즘에 의하여 구축된 모델의 성능 및 정확도 비교·분석을 통하여 간벌목 선정을 위한 최적의 모델을 선정하였다(Figure 1).

1) 간벌목 선정 요인과 수관급 분류

우리나라에서 수행되는 간벌방법은 정량간벌, 도태간벌, 열식간벌이 있으며, 보통은 정량간벌과 도태간벌을 적용하고 있다. 정량간벌은 수종별 직경급 잔존본수 기준표에 따라 간벌량을 결정하는 방법으로 임목의 흉고직경과 수고, 수관폭 등을 고려하여 간벌목을 선정하며, 도태간벌은 미래목에 대한 방해목을 선정하여 간벌하는 방법으로 임목의 흉고직경, 수고, 형질급, 개체목 간 거리 등을 고려하여 간벌목을 선정한다. 열식간벌은 열식인공조림지나 생장이

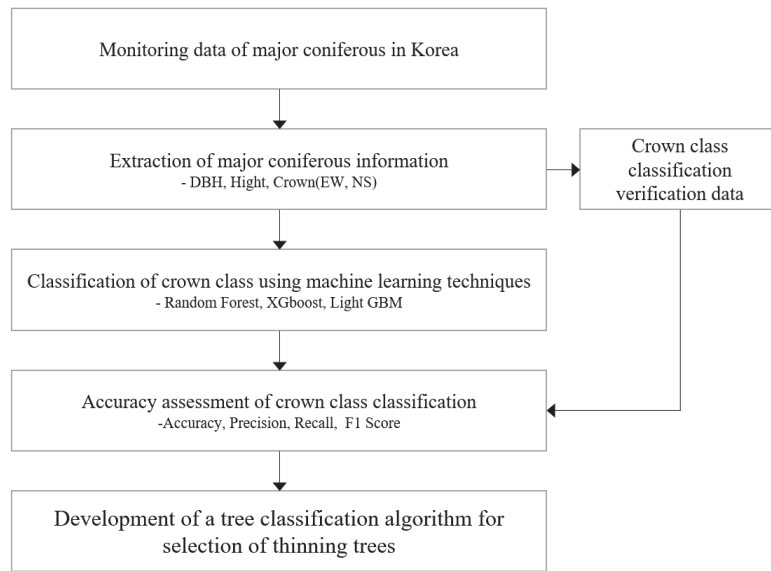


Figure 1. Flow chart of research.

균일한 입지에서 간벌열과 존치열을 선정하여 간벌을 시행되는 방법이다. 일반적으로 간벌방법에 있어 정량간벌을 기준으로 하지만 잔존목에 대한 대략적인 선정과 잔존목의 직경이 편중되는 것을 막기 위하여 도태간벌의 형태도 함께 적용하고 있다. 간벌목을 선정하는데 있어 기본적으로는 양적으로 얼마나 간벌을 할 것인가를 결정하게 되고 결정에 따라 고사목과 형질이 좋지 못한 나무를 위주로 간벌하게 된다. 일반적으로 모든 간벌방식에서 간벌량 및 간벌률을 결정 후, 미래목 또는 우량목에 대한 방해목을 간벌목으로 선정하고 있으며, 간벌목의 선정에 있어 고사목, 피해목, 피압목(열세목), 생장불량목, 형질열등목, 우량목의 순으로 입목의 발달상태가 낮은 입목부터 차례로 간벌목을 선정한다(Korea Forest Service, 2007). 본 연구는 입목의 발달상태를 평가하는 기준으로 수관급(우세목, 준우세목, 중층목, 피압목)을 적용하였다. 또한, 수관급을 구분하는 요인으로 입목의 측정요인인 흉고직경, 수고, 수관폭을 선정하였다. 수관급은 입목의 발달 상태에 따라 구분한 기준으로 국가산림자원조사의 산림의 건강·활력도 조사에 활용되고 있으며, 구분 기준은 Table 1과 같다.

2) 머신러닝을 이용한 수관급 분류 알고리즘 구축

(1) 데이터셋 구축

수관급에 의한 입목 분류를 위한 입력데이터는 ‘국유림 내 주요수종 간벌효과 장기모니터링’ DB를 이용하였으며, 머신러닝 알고리즘에 적용하기 위하여 데이터셋을 구축하였다. 모니터링 데이터는 Plot 단위의 조사 정보이며, 모든 정보를 취합하여 사용할 경우, 지역 및 지위에 따라 입목의 편차가 발생할 수 있다. 따라서 입력데이터의 Plot 단위 데이터의 Z-Score를 통하여 표준화된 데이터셋을 구축하였다.

(2) 머신러닝 분류 알고리즘 구축

의사결정트리에 의한 분류 알고리즘은 비모수 통계 방법으로 가정이 필요 없어 일반적인 머신러닝 데이터 분석에 가장 많이 활용된다(Kim et al., 2020). 그러나 의사결정트리는 분류정확도를 높이기 위하여 과적합의 문제를 가지고 있으며, 일반화 성능이 좋지 못한 단점이 있다. 이러한 의사결정트리 알고리즘을 보완하기 위하여 여러 의사결정트리 분류기를 하나로 묶어서 사용하는 앙상블을 도

Table 1. Classification standard of crown class.

Classification	Content
Dominant	As a standing tree that forms the top layer of the canopy, the canopy is well developed so that it receives sufficient vertical light and receives a part of the horizontal light from the side.
Co-dominant	As a standing tree forming the upper layer of the stand, it receives sufficient vertical light, but receives some horizontal light from the side.
Intermediate	The crown forms the lower layers of the dominant and semi-dominant trees.
Overtopper	A deciduous tree whose crown belongs to the lower tier and has not grown properly due to neighboring pressure.

입하였으며, 본 연구에서는 의사결정트리기반 앙상블 알고리즘의 대표적인 Random Forest, XGBoost, LightGBM을 사용하였다.

① Random Forest (RF)

Random Forest는 앙상블 모델의 대표적인 방법으로 데이터셋을 복원 추출하여 의사결정나무 모델의 구축하는 Bagging과 의사결정나무 모델의 구축 시 변수를 무작위로 선택하는 Random subspace 알고리즘의 조합으로 구성된다. Random Forest는 개별 의사결정나무의 노드 분할에서 일정한 개수의 변수를 랜덤으로 선택하여 최적의 분할 가능 변수를 선택하며, Random subspace 를 통하여 개별적 의사결정나무가 중복되는 것을 방지하여 데이터셋의 과적합을 방지한다. Random Forest는 다수의 의사결정나무 모델에 의하여 예측된 정보의 종합 분석으로 일반적으로 하나의 의사결정나무 보다 높은 예측 정확성을 보여준다(Breiman, 2001; Biau et al., 2016).

② XGboost (XGB)

XGboost는 Gradient boosting 알고리즘의 최적화 방법으로 의사결정나무의 작성에서 Error 값을 낮추는 부스팅 기반 모델이다. 회귀와 분류, 순위 등을 지원하는 유연한 모델이다(Chen et al., 2016). Gradient boosting은 각각의 의사결정나무를 수직으로 연결하여 개별적으로 데이터의 입력과 출력과정을 거치게 된다. 이 과정에서 도출된 잔여 오차를 새로운 의사결정나무에 학습시켜 분류정확도를 높이는 방법이다. Gradient boosting 모델은 뛰어난 예측력을 가지고 있으나, 과정에서 수행과정이 오래걸리는 단점을 가지고 있어 이러한 직렬 알고리즘을 병렬로 처리가 가능하도록 하여 예측 성능과 수행시간을 단축시킨 모델이 XGboost 이다(Jiang et al., 2019).

③ Light GBM (LGBM)

Light GBM은 XGboost와 마찬가지로 Gradient boosting 알고리즘을 기반으로 하고 있으며, 기존 알고리즘보다 가

볍게 구현함으로써 수행시간을 획기적으로 단축시킨 알고리즘이다. Light GBM은 기존의 균형 트리 분할(Level Wkse) 방식을 사용하지 않고, 의사결정나무의 최대 손실 값을 가지는 노드(Leaf node)를 지속적으로 분할하여 트리의 깊이를 고려하지 않는 방법이다. 따라서 기존의 균형 트리 분할에서 균형을 맞추기 위하여 소요된 시간을 단축시킬 수 있다. 반면, 데이터의 수가 적을 경우 과적합 가능성이 발생한다는 단점을 가지고 있다(Ke et al., 2017).

3) 머신러닝 분류 알고리즘의 성능 및 정확도 평가

알고리즘에 의하여 구축된 머신러닝 모델의 성능평가는 학습된 분류 데이터와 검증데이터 간의 오차행렬(Confusion Matrix)을 기반으로 산정하며, 오차행렬 값을 조합하여 정확도(Accuracy), 정밀도(Precision), 재현율(Recall), F1 Score 를 산출하고 모델을 평가하였다(Table 2, 3).

정확도는 실제값과 예측치가 일치한 정도를 계산한 값으로 정확도는 분류 문제에서 가장 간단하고 직관적인 평가 지표이다. 정밀도는 모델이 True라고 분류한 것(True Positive; TP + False Positive; FP) 중에서 실제로 True로 분류된 비율이며, 정답률이라고 불리기도 한다. 재현율은 실제 True인 것(True Positive; TP + False Negative; FN) 중 모델이 True로 예측한 비율을 뜻하며, 재현율과 정밀도 모두 TP를 높이는데 동일하게 초점을 맞추지만 재현율은 FN, 정밀도는 FP를 낮추는데 초점을 맞추기 때문에 서로 보완적인 지표이다. 가장 좋은 성능평가는 재현율과 정밀도 모두 높은 수치를 얻는 모델이며, F1 score는 정밀도와 재현율의 조화평균으로 데이터간 불균형정도를 평가해 볼 수 있다(Lee and Sun, 2020).

Table 2. Confusion matrix decision table.

	Actual	True	False
Predict			
True		True Positive (TP)	False Positive (FP)
False		False Negative (FN)	True Negative (TN)

Table 3. Performance metrics.

Classification	Performance metrics
Accuracy	$(TP+TN)/(TP+FN+FP+TN)$: Accuracy if how close a measured value is to the actual(true) value.
Precision	$TP/(TP+FP)$: Precision quantifies the number of positive class predictions that actually belong to the positive class.
Recall	$TP/(TP+FN)$: Recall quantifies the number of positive class predictions made out of all positive examples in the dataset.
F1 score	$2/(1/Precision + 1/Recall)$: F+Measure provides a single score that balances both the concerns of precision and recall in one number.

결과 및 고찰

1. 주요 침엽수의 현황

본 연구에 활용된 주요 침엽수 데이터를 보면, 평균흉고 직경은 27.7 cm, 수고 평균은 19.4 m, 수관폭(동서) 평균 4.9 m, 수관폭(남동) 평균 5.1 m이다. 소나무의 평균흉고 직경은 31.6 cm로 잣나무와 낙엽송보다 각각 약 1.1배, 약 1.2배 높았으며, 낙엽송의 평균 수고는 22.4 m로 소나무와 잣나무보다 약 1.3배 높았다. 소나무의 평균 수관폭은 동서로 5.6 m, 남북으로 5.7 m로 잣나무와 낙엽송보다 높았으며, 낙엽송은 수관이 좁고 겹가지가 가늘게 형성되는 특징이 있어 소나무와 잣나무보다 수관폭이 낮은 것으로 판단된다(Table 4). 한편, 본 연구에 사용된 자료를 “지속가

능한 산림자원관리 표준매뉴얼”의 시업체계와 비교한 결과, 모든 수종이 우량대경재 III, IV영급 기준에 부합하며, 수고 및 흉고직경의 생장이 좋은 것으로 나타났다. 수종별 수관급에 따른 흉고직경-수고, 수관폭 분포를 보면, 낙엽송이 흉고직경에 따른 수고생장이 가장 좋았으며, 분포 밀도도 가장 좋았다. 수관폭은 모든 수종이 대칭형을 이루고 있어 침엽수의 생육 특성인 원추형태를 보이고 있으며, 방위에 따른 큰 차이가 없음을 알 수 있다(Kim et al., 2013). 수관급에 따른 분포 특성을 보면, 수관급에 따라 분포 군집을 이루고 있지만, 우세목과 준우세목은 전반적으로 분포특징이 유사하며, 중간목과 피압목은 흉고직경과 수고가 낮은 원점에 가깝게 군집을 이루어 분포하고 있다 (Figure 2).

Table 4. Current status of measurement factors.

Species	Content	DBH (cm)	Height (m)	Crown-EW (m)	Crown-NS (m)
<i>Pinus densiflora</i>	Average	31.6	17.0	5.6	5.7
	Maximum	61.4	28.2	12.1	13.1
	Minimum	8.5	4.3	1.1	0.9
<i>Pinus koraiensis</i>	Average	27.8	16.9	5.0	5.0
	Maximum	63.5	29.4	12.6	12.8
	Minimum	6.8	3.8	0.3	0.6
<i>Larix kaempferi</i>	Average	26.4	22.4	4.7	4.9
	Maximum	60.8	36.4	13.2	11.5
	Minimum	9.5	7.7	0.8	0.9

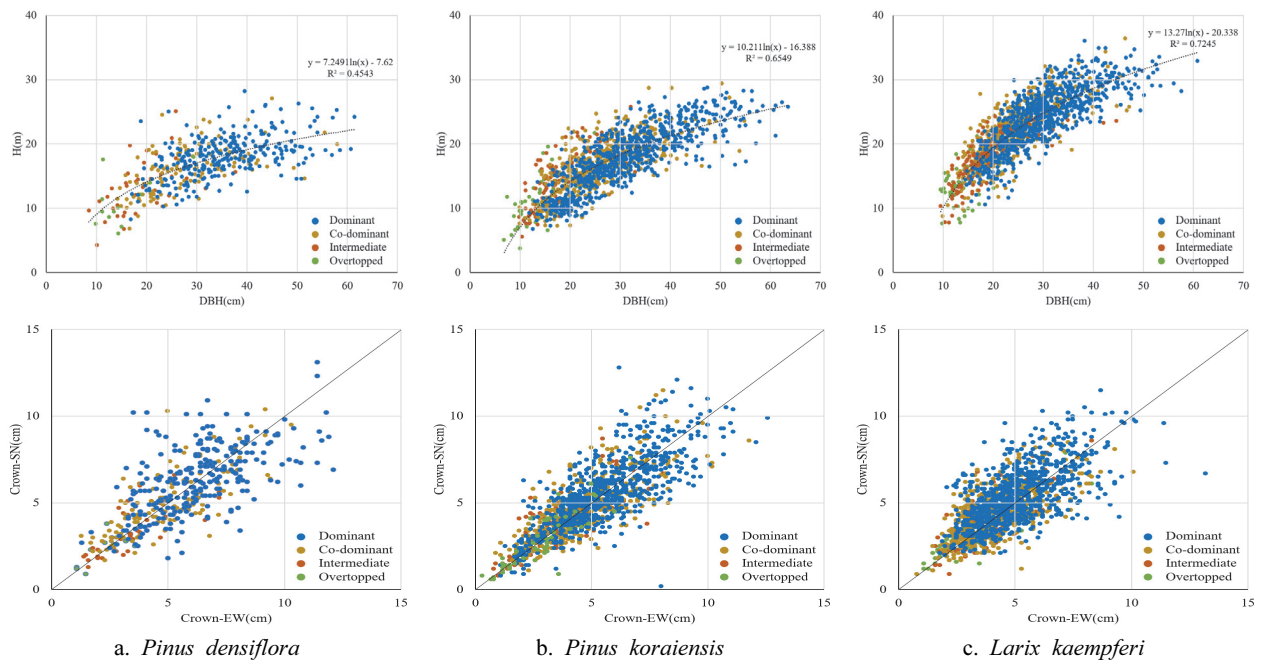


Figure 2. Current status of measurement factors of major conifers by crown class.

2. 수종별 수관급분류 알고리즘의 성능 및 정확도 평가

1) 소나무

소나무의 입목정보를 이용하여 수관급 분류에 대한 성능 및 정확도 평가 결과, 정확도는 Random Forest(이하 RF)가 64.7%로 가장 높았으며, Light GBM(이하 LGBM)(59.6%), XGboost(이하 XGB)(58.8%)의 순으로 분석되었다. 소나무는 알고리즘의 성능평가에 있어 정밀도와 재현율, F1-Score 모두 RF > LGBM > XGB의 순으로 분석되었으며, RF의 정확도는 XGB와 LGBM 비교하여 정확도가 5% 이상 높아 알고리즘 성능평가 점수가 가장 양호하였다. 그러나 우세목을 제외한 모든 항목에서 평가점수는 매우 낮았다. 소나무의 평가 데이터는 총 136본으로 검증데이터로 사용되어 학습 및 평가 데이터의 수가 적었으며, 이로 인한 누락 및 오차에 대한 평가점수 차이가 크게 발생하였다(Figure 3). 오차행렬을 통하여 수관급별 성능평가를 보면, 모든 알고리즘이 우세목에 대한 정밀도, 재현

율이 70%이상이었으며, RF에서 우세목이 정밀도는 83.3%로 우수한 분류 정확도를 보였다. 그러나 중간목과 피압목의 정확도는 30% 이하로 매우 낮아 상반된 경향을 보였다. 이것은 앞서 설명한 것처럼 분석에 사용된 학습데이터와 검증데이터의 수가 적어 오차에 대한 정확도 점수가 영향을 크게 영향을 받았다. 소나무에 대한 분류 알고리즘은 RF를 사용하며, 우세목의 분류를 통한 도태간벌에 적용이 가능할 것으로 판단된다(Table 5).

2) 잣나무

잣나무의 정확도는 RF가 60.2%로 가장 높았으며, LGBM(57.1%), XGB(56.6%)의 순으로 분석되었다. 잣나무는 알고리즘별 평가에서 모든 항목에서 RF의 평가점수가 높았으며, 정밀도를 제외한 모든 항목에서 RF가 3% 이상 높게 평가되어 소나무와 유사한 경향을 보였다. 한편 RF를 제외한 XGB와 LGBM은 모든 평가항목에서 1% 이하의 차

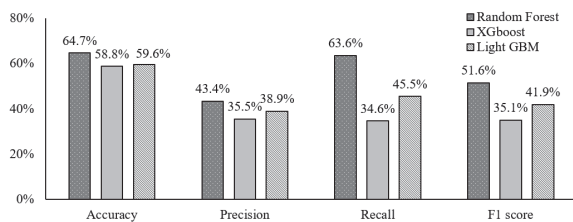


Figure 3. Performance evaluation of machine learning algorithms by *Pinus densiflora*.

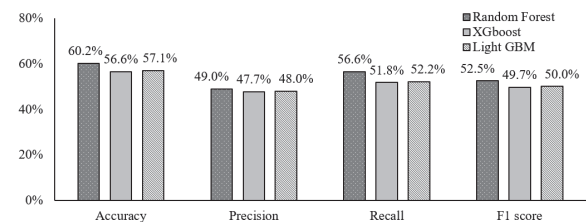


Figure 4. Performance evaluation of machine learning algorithms by *Pinus koraiensis*.

Table 5. Confusion matrix of *Pinus densiflora* by machine learning algorithms.

		Reference data				Total
		Dominant	Co-dominant	Intermediate	Overtopper	
Random Forest	Class					
	Dominant	70	20	4	-	94
	Co-dominant	14	14	4	1	33
	Intermediate	-	3	3	2	8
	Overtopper	-	-	-	1	1
	Total	84	37	11	4	136
XGboost	Class					
	Dominant	68	22	6	-	96
	Co-dominant	14	10	1	1	26
	Intermediate	1	4	1	2	8
	Overtopper	1	1	3	1	6
	Total	84	37	11	4	136
Light GBM	Class					
	Dominant	65	21	5	-	91
	Co-dominant	17	13	3	1	34
	Intermediate	2	3	2	2	9
	Overtopper	-	-	1	1	2
	Total	84	37	11	4	136

이로 매우 유사한 결과로 분석되었으며, 오차행렬에서도 XGB와 LGBM의 분류결과가 유사하였다. LGBM은 XGB의 개선 모델로 기본 알고리즘이 유사하게 설계되어 본 연구에서 성능평가 점수는 유사하게 산출된 것으로 판단된다(Figure 4). 오차행렬을 보면, 잣나무 분류 결과는 소나무 모델과 유사한 형태를 보이고 있으며, 모든 알고리즘

이 우세목에 대한 정밀도, 재현율이 매우 높은 반면, 준우세목과 중간목, 피압목에서 성능 및 정확도 점수가 낮았다. 특히, 잣나무의 검증데이터는 준우세목과 중간목으로 분류되었으나, 알고리즘에 의하여 분류는 우세목 또는 준우세목으로 분류된 입목이 많아 분류정확도가 낮아졌다. 잣나무림은 비교적 모든 평가항목이 안정적이고 유사하

Table 6. Confusion matrix of *Pinus koraiensis* by machine learning algorithms.

		Reference data				
	Class	Dominant	Co-dominant	Intermediate	Overtopper	Total
Random Forest	Dominant	186	74	6	1	267
	Co-dominant	40	44	21	3	108
	Intermediate	5	9	9	2	25
	Overtopper	-	-	2	8	10
	Total	231	127	38	14	410
XGboost	Dominant	167	67	9	2	245
	Co-dominant	60	48	18	1	127
	Intermediate	3	11	9	3	26
	Overtopper	1	1	2	8	12
	Total	231	127	38	14	410
Light GBM	Dominant	163	61	7	2	233
	Co-dominant	65	55	20	1	141
	Intermediate	3	10	8	3	24
	Overtopper	-	1	3	8	12
	Total	231	127	38	14	410

Table 7. Confusion matrix of *Larix kaempferi* by machine learning algorithms.

		Reference data				
	Class	Dominant	Co-dominant	Intermediate	Overtopper	Total
Random Forest	Dominant	191	66	3	-	260
	Co-dominant	56	60	25	3	144
	Intermediate	6	9	12	5	32
	Overtopper	-	4	1	9	14
	Total	253	139	41	17	450
XGboost	Dominant	195	63	5	-	263
	Co-dominant	54	59	23	5	141
	Intermediate	4	12	11	4	31
	Overtopper	-	5	2	8	15
	Total	253	139	41	17	450
Light GBM	Dominant	180	65	4	1	250
	Co-dominant	69	61	24	6	160
	Intermediate	4	11	12	4	31
	Overtopper	-	2	1	6	9
	Total	253	139	41	17	450

게 분류되어 머신러닝 알고리즘으로 인한 차이는 크지 않았다. 그러나, 준우세목과 중간목의 분류를 위한 개선이 필요할 것으로 판단된다(Table 6).

3) 낙엽송

낙엽송의 정확도는 XGB가 60.2%로 가장 높았으며, RF(60.4%), LGBM(57.6%)의 순으로 분석되었다. 낙엽송은 정확도에서 XGB가 높게 평가되었으나, 정밀도, 재현율, F1-Score는 모두 RF가 가장 높게 평가되었으며, 재현율에서는 XGB가 가장 낮게 평가되어 검증데이터에 대한 모델의 정확도는 낮게 평가된 것으로 나타났다(Figure 5). 오차행렬을 통하여 수관급별 성능평가를 보면, 앞선 두 수종의 분류와 같이 모든 알고리즘이 우세목에 대한 정밀도, 재현율이 매우 높은 반면, 준우세목과 중간목, 피압목에서 성능평가 점수가 낮았다. 낙엽송은 소나무와 잣나무 보다 준우세목의 성능평가 점수가 42%~43% 수준으로 분류정확도가 높았으나, 전체적인 평가점수가 50% 미만으로 일반적인 성능평가 점수는 낮게 평가되었다(Table 7).

결론

본 연구는 간벌목 선정을 위한 수관급 분류 방법의 개발을 목표로 머신러닝 알고리즘 성능평가를 통한 적용가능성을 검토하였다. 수종별 수관급 분류 결과는 수종 및 알고리즘의 차이와 관계없이 우세목은 70%~80%의 우수한 분류정확도를 보였으나, 준우세목, 중간목, 피압목은 50% 이하의 낮은 분류정확도를 보였다. 우리나라에서 많이 활용되고 있는 정량간벌과 도태간벌의 경우 간벌목의 선정을 위하여 우세목 또는 미래목을 먼저 선정하고 작업을 수행한다. 현재 알고리즘은 우세목의 분류 결과가 높게 평가되었으며, 간벌작업 시 존치목 선정을 위하여 활용가능할 것으로 판단된다. 한편, 현재 알고리즘은 수관급 분류를 위하여 흉고직경, 수고, 수관폭 정보를 활용하였으나, 수관급 분류는 임분내 입목간의 상대적 평가에 의하여 등급이 결정된다. 수관급의 결정 및 간벌목 선정에 있어 형질급 및 입목 위치에 대한 고려가 필요하다. 수관폭은 수종 및 주변 입목과의 관계에 따라 상대적인 수치를 보일 수 있으며, 흉고직경과 수고보다 측정에 대한 소요시간이 많이 증가하여 현장조사의 업무 증가를 발생시킨다. 향후, 형질급 및 입목간의 거리를 변수로 투입하여 알고리즘을 개선한다면, 수관급 분류를 위한 알고리즘의 정확도가 향상될 것으로 판단된다. 또한, 본 연구는 수관급 구분을 위하여 입목조사자료를 활용하였으나, 수관급 판정을 위하여 임분내 개체목의 사회적 등급을 임분별로 평가가 필요하여 향후 임분단위의 수관급을 객관적으로 평가할 수 있

는 지표 개발 및 적용이 필요할 것으로 판단된다. 머신러닝 알고리즘을 통한 간벌목의 선정은 기존의 주관적이고 정성적이었던 기준을 객관적이고 정량적으로 표현이 가능하였으며, 간벌목 선정에 있어 참조자료로 활용이 가능할 것이라 판단된다.

기존 간벌에서 선목작업은 사업시행자가 표준지 선정을 통하여 설계안을 마련하고 검토를 통하여 시행하도록 하고 있다. 전문적인 지식과 경험이 없이는 매우 어려운 작업이다. 본 연구방법에 적용된 머신러닝 기법은 간단한 입력의 수치적 데이터를 통하여 객관적인 결과를 도출함으로써 사유림의 산주들이 본인의 산에 대한 경영계획 수립과 가꾸는 방법의 의사결정에 있어 참고자료로 활용이 가능할 것이다. 이러한 시스템의 구축은 우리나라 60%이상 되는 사유림 경영자들에게 적극적인 산림경영 참여가 가능하도록 할 것이며, 부재산주들에게도 본인의 산을 경영·관리를 위한 의사결정에 있어 참고자료가 될 것이라 판단된다.

감사의 글

본 연구는 산림청(한국임업진흥원) 산림과학기술 연구개발사업(2021359D10-2123-BD01)의 지원에 의하여 이루어진 것입니다.

References

- An, K.H. and Shin, S.H. 2014. A study on the consciousness of proxy management in private forest. *Korean Journal of Forest Economics* 12(2): 14-24.
- Biau, G., Scornet, E., and Welbl, J. 2016. Neural random forests. *arXiv preprint arXiv:1604.07143*.
- Breiman, L. 2001. Random forests. *Machine Learning* 45(1): 5-32.
- Chen, T., and Guestrin, C. 2016. XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. pp. 785-794.
- Choi, J.Y., Lee, S.H. and Lee, S.H. 2015. Anticipation of the future suitable cultivation areas for Korean pines in Korean peninsula with climate change. *The Korea Society For Environmental Restoration And Revegetation Technology* 18(1): 103-113.
- Jiang, Y., Tong, G., Yin, H., and Xiong, N. 2019. A pedestrian detection method based on genetic algorithm for optimize XGBoost training parameters. *IEEE Access*. 7: 118310-118321.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W.,

- and Liu, T.Y. 2017. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems* 30: 3146-3154.
- Kim, B.M., Woo, H.S. and Park, J.W. 2020. Performance of deep learning-based automatic classification of forest plants: A comparison of data collection methods. *Journal of Korean Forest Society* 109(1): 23-30.
- Kim, D.Y., Im, S.J., Kwon, M.H. and Yim, J.H. 2010. Monitoring of vegetation coverage for selecting plants for beach revegetation. *Journal of the Korean Society of Civil Engineers* 30(5): 519-524.
- Kim, E.M., Kim, S.B., and Cho, E.S. 2020. Using mechanical learning analysis of determinants of housing sales and establishment of forecasting model. *Journal of Cadastre & Land InformatiX*, 50(1): 181-200.
- Korea Forest Service. 2007. Forest Management Standard Textbook. https://www.forest.go.kr/kfswweb/cop/bbs/selectBoardArticle.do?nttId=3104628&bbsId=BBSMS TR_1008&pageUnit=9&pageIndex=1&searchtitle=title&searchcont=&searchkey=&searchwriter=&searchWrd=%ec%88%b2%ea%b0%80%ea%be%b8%ea%b8%b0+%ed%91%9c%ec%a4%80&ctgryLrcls=&ctgryMdcls=&ctgrySmcls=&ntcStartDt=&ntcEndDt=&mn=NKFS_06_09_05 (2021.11.22.).
- Korea Forest Service. 2020. 2020 Forest Basic Statistics. <https://kfss.forest.go.kr/stat/ptl/fyb/frstyYrBookList.do?curMenu=9854> (2021.11.22.).
- Kwon, K.B. 2013. Development of a forest stand management model for carbon and timber management. Seoul University, Seoul, Korea.
- Kwon, S.D. and Chung, J.S. 2004. Development of individual-tree distance-independent simulation model for growth prediction of *Pinus koraiensis* stands. *Journal of Korean Forest Society* 93(1): 43-49.
- Kwon, S.D., Kim S.Y., Chung, J.S. and Kim H.H. Development of thinning effect analysis model (TEAM) using individual-tree distance-independent growth model of *Pinus koraiensis* stands. *Journal of Korean Forest Society* 96(6): 742-749.
- Kwon, S.K., Lee, Y.S., Kim, D.S., and Jung, H.S. 2019. Classification of forest vertical structure using machine learning analysis. *Korean Journal of Remote Sensing* 35(2): 229-239.
- Lee, H.W. 2019. Development of supervised machine learning based catalog entry classification and recommendation system. *Journal of Internet Computing and Services* 20(1): 57-65
- Lee, Y.J. and Sun, J.W. 2020. Predicting highway concrete pavement damage using XGBoost. *Korean Journal of Construction Engineering and Management* 21(6): 46-55.
- Minowa, Y. 2001. Analyzing a combination of factors for thinning trees with a neural network. *Journal of Forest Research* 6(2): 95-100.
- Minowa, Y. 2005. Classification rules discovery from selected trees for thinning with the C4. 5 machine learning system. *Journal of Forest Research* 10(3): 221-231.
- Tasaka, T., Kumakura, Y., Uchiage, S., Yano, Y., Saito, K., Shinoda, T., Ueki, S. and Saito, N. 2001. Evaluation of a method for the estimation of logging productivity using data-miningprogram C4. 5. *Bulletin of the Utsunomiya University Forests, Japan*.

Manuscript Received : November 23, 2021

First Revision : April 17, 2022

Second Revision : May 27, 2022

Accepted : May 30, 2022