


라이다 플랫폼과 딥러닝 모델에 따른 잣나무와 낙엽송의 분류정확도 비교 및 평가

이용규 · 이상진 · 이정수 *

강원대학교 산림경영학과

Comparison and Evaluation of Classification Accuracy for *Pinus koraiensis* and *Larix kaempferi* based on LiDAR Platforms and Deep Learning Models

Yong-Kyu Lee, Sang-Jin Lee, and Jung-Soo Lee *

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

요약: 본 연구는 잣나무와 낙엽송을 대상으로 라이다로부터 취득된 3차원의 Point cloud data (PCD)를 이용하여 딥러닝 기반의 수종 분류 모델을 구축하고 분류정확도를 비교·평가하였다. 수종 분류 모델은 라이다 플랫폼(고정식과 이동식), Farthest point sampling (FPS) 기반의 다운샘플링 강도(1024개, 2048개, 4096개, 8192개), 딥러닝 모델(PointNet, PointNet++) 3가지 조건에 따라 총 16개의 모델을 구축하였다. 분류 정확도 평가 결과, 고정식 라이다는 다운샘플링 강도가 8192개인 PCD 자료에 PointNet++ 모델을 적용하였을 때 카파계수가 93.7%로 가장 높았으며, 이동식 라이다는 다운샘플링 강도가 2048개에 PointNet++을 적용하였을 때 카파계수가 96.9%로 가장 높았다. 또한, 플랫폼과 다운샘플링 강도가 동일한 경우 PointNet++이 PointNet보다 정확도가 높았다. 구축된 16개 모델의 오분류 사례는 첫 번째, 수종 간의 구조적인 특징이 유사한 개체목 두 번째, 경사지 또는 임도 주변에 위치하여 편심성장한 개체목 세 번째, 개체목 분할 시 수관부가 수직으로 분할된 개체목에 대해 발생하였다.

Abstract: This study aimed to use three-dimensional point cloud data (PCD) obtained from Terrestrial Laser Scanning (TLS) and Mobile Laser Scanning (MLS) to evaluate a deep learning-based species classification model for two tree species: *Pinus koraiensis* and *Larix kaempferi*. Sixteen models were constructed based on the three conditions: LiDAR platform (TLS and MLS), down-sampling intensity (1024, 2048, 4096, 8192), and deep learning model (PointNet, PointNet++). According to the classification accuracy evaluation, the highest kappa coefficients were 93.7% for TLS and 96.9% for MLS when applied to PCD data from the PointNet++ model, with down-sampling intensities of 8192 and 2048, respectively. Furthermore, PointNet++ was consistently more accurate than PointNet in all scenarios sharing the same platform and down-sampling intensity. Misclassification occurred among individuals of different species with structurally similar characteristics, among individual trees that exhibited eccentric growth due to their location on slopes or around trails, and among some individual trees in which the crown was vertically divided during tree segmentation.

Key words: tree species classification, terrestrial laser scanning, mobile laser scanning, point cloud data, PointNet, PointNet++

서론

정확한 산림자원조사는 산림정책 수립, 생물다양성 보

니터링, 생태계 유지 등에서 매우 중요하다. 특히, 수종 분류는 산림 중 구조, 산림자원의 지속가능한 관리를 위해 산림자원조사에서 필수적인 항목이다(Jones et al., 2010; Ballanti et al., 2016). 하지만, 기존의 산림자원조사는 현장 조사자에 의한 산림자원조사가 수행되고 있으며 특히, 수종 정보는 주관적인 판단이 적용된다. 또한, 수종을 분류

* Corresponding author

E-mail: jslee72@kangwon.ac.kr

ORCID

Jung-Soo Lee  https://orcid.org/0000-0003-2770-0528

하기 위해서는 많은 교육과 경험이 필요하다(Pu and Landry, 2012; Yan et al., 2021). 한편, 원격탐사 분야에서는 다양한 플랫폼으로 취득한 데이터에 딥러닝 모델을 이용하여 수종을 분류하는 연구가 진행되고 있다. 특히, 국외의 경우 최근 라이다(Light Detection And Ranging, LiDAR) 기술의 발전으로 산림 분야에서 라이다를 활용하여 흉고직경, 수고, 재적, 수종 등 정보를 추출하기 위한 연구가 다양하게 진행되고 있다(Wessel et al., 2018; Axelsson et al., 2018; Guan et al., 2015). 라이다는 레이저를 주변으로 방출하여 사물의 표면으로부터 반사되어 되돌아오는 신호를 분석하여 대상 물체에 대한 3차원 정보인 포인트 클라우드 데이터(Point cloud data, PCD) 형태로 취득할 수 있다(Ko et al., 2022). 라이다는 장착된 플랫폼에 따라 항공 라이다(Airborne Laser scanning, ALS), 지상 라이다(Terrestrial Laser scanning, TLS), 이동식 라이다(Mobile Laser scanning, MLS)로 구분된다. 일반적으로 ALS는 항공기 또는 드론에 장착되어 대규모 면적의 데이터 취득이 가능하지만, 포인트 클라우드의 밀도가 낮고 분포가 고르지 않다. TLS는 라이다를 지상에 고정하여 데이터를 취득하는 방법으로 고밀도의 PCD를 취득할 수 있으나, 데이터 수집 과정이 노동 집약적이며 비용이 많이 든다. MLS는 지상에서 이동하면서 PCD를 취득하며, 빠르고 편리하게 획득할 수 있으며 다른 플랫폼에 비해 비용이 크게 절감되지만, TLS에 비해 정밀도가 낮은 단점이 있다(Liang et al., 2018; Liu et al., 2022).

최근, PCD를 분류하기 위한 딥러닝 기술에 관한 연구가 다양하게 이루어지고 있다(Tao et al., 2015; Zeybek and Şanlıoğlu, 2019). 기존의 PCD를 처리하기 위한 딥러닝은 3D 모델을 2D 데이터로 변환하거나, 3D 복셀 그리드 형태로 변환하여 입력데이터를 구축하고 합성곱층(Convolution layer)에 입력하였다(Wu et al., 2015; Su et al., 2015). 이러한 기존의 합성곱 신경망(Convolution Neural Network,

CNN)의 적용을 위한 데이터 변형은 데이터에 대한 정보와 공간적 특징이 손실되기 때문에 3D 객체 정보를 그대로 반영하기에 한계가 있다. 따라서, PCD의 특성을 그대로 적용할 수 있도록 PCD가 모델에 입력되어 학습이 진행되는 PCD 딥러닝 모델이 개발되었다(Xi et al., 2020). 대표적인 PCD 딥러닝 모델은 PointNet과 PointNet++가 있으며, 두 모델을 활용하여 수종을 분류하는 다양한 연구가 진행되고 있다(Qi et al., 2017a; Qi et al., 2017b; Liu et al., 2021; Briechele et al., 2020; Liu et al., 2022).

따라서, 본 연구는 우리나라의 산림자원조사에 라이다를 적용하기 위해 PCD와 딥러닝 모델을 이용하여 수종 정보를 추출하는 것을 목적으로 하였다. 이를 위해 본 연구에서는 우리나라 주요 침엽수종인 잣나무와 낙엽송에 대한 데이터를 수집하고, 플랫폼 별 다운샘플링 강도, 딥러닝 모델에 따른 수종 분류 결과를 도출함으로써 라이다를 이용한 산림자원조사에 수종 분류 모델을 적용하고자 하였다.

재료 및 방법

1. 연구대상지 및 장비

1) 연구대상지

연구대상지는 강원도 춘천시 동산면과 홍천군 북방면에 위치한 강원대학교 학술림의 잣나무림과 낙엽송림 지역에 각각 5개의 표본지를 선정하였다. 표본지는 국가산림자원조사(NFI) 표본지 설정에 따라 기본 조사구 반경이 11.3 m, 면적이 0.04 ha인 원형 Plot으로 설정하여 조사하였다(Korea Forest Service, 2021)(Figure 1). 잣나무림과 낙엽송림 표본지는 모두 5영급 인공림으로, 평균 경사는 잣나무림에서 약 20°, 낙엽송림에서 약 26°로 분포하였다. 수종별 5개 표본지의 개체목 수는 잣나무 205본, 낙엽송 86본이 분포하였으며, 수종별 평균 흉고직경과 수고는 잣

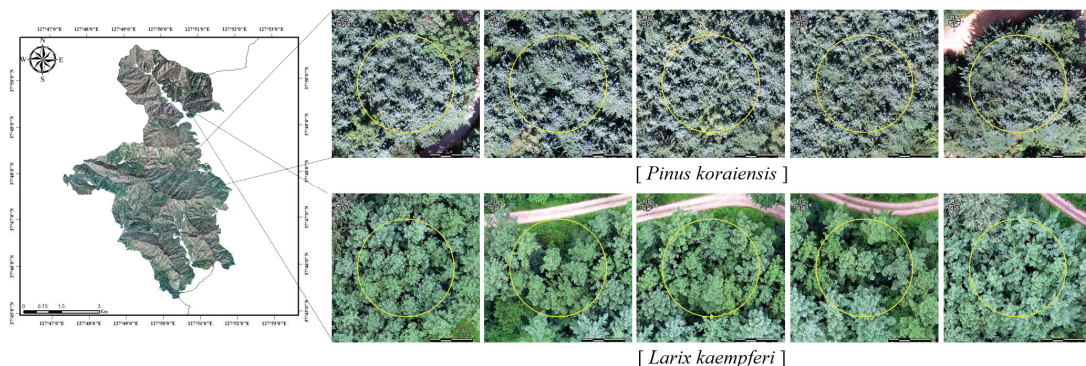


Figure 1. The location of study area.

Table 1. Distribution of diameter at breast height (DBH) and tree height in the study area.

Species	Age class	DBH(cm)				Height(m)			
		Min	Max	Mean	STD	Min	Max	Mean	STD
<i>Pinus koraiensis</i>	V	12.0	50.3	27.6	7.3	10.3	27.4	19.0	2.4
<i>Larix kaempferi</i>	V	16.2	47.5	32.7	7.2	10.2	34.1	26.1	4.9

Table 2. Specifications and performance of the BLK360 and Hovermap LiDAR.

Parameters	BLK 360	Hovermap LiDAR
Measurement speed	360,000 points/s	300,000 points/s
Range accuracy	4mm @ 10m/ 7mm @ 20m	±15mm
Horizontal angle	360°	360°
Vertical angle	300°	360°
Maximum measurement distance	60 m	100 m
Laser wavelength	830 nm	905 nm
Average scanning time per station [mm:ss]	05:30	-
Point density	~300,000 per/m ²	-

나무림에서 27.6 cm와 19.0 m, 낙엽송림에서 32.7 cm와 26.1 m로 분포하였다(Table 1).

2) 연구장비

TLS는 BLK360 (Leica Geosystem, Heerbrug, Switzerland)을 이용하였다. BLK360은 최대 60 m의 스캔 범위, 위치 정확도는 10 m 거리에서 4 mm, 360° × 300° 시야각도, 초당 360,000 포인트를 취득하는 동시에 RGB 정보에 대한 정보를 취득할 수 있다. 또한, BLK360은 포인트 밀도 설정을 통해 포인트 수를 설정하여 데이터의 수집 시간과 크기를 설정할 수 있으며, 본 연구에서는 단일 스캔 시간을 약 5분 30초로 설정하였다. MLS는 Hovermap (Emesent Pty Ltd, Queensland, Australia)을 이용하였다. Hovermap은 최대 100 m의 스캔 범위, 위치 정확도는 30 mm, 360° × 360° 시야각도, 초당 300,000 포인트를 취득할 수 있다 (Table 2).

2. 연구방법

본 연구에서는 TLS와 MLS를 이용하여 PCD를 수집하였으며, 수집된 표본지 단위의 PCD를 노이즈 제거 및 지면점 분리, 높이 정규화, 개체목 분할 과정의 데이터 전처리를 통해 개체목 단위의 PCD로 구축하였다. 각 개체목 단위의 PCD는 데이터 형태에 따른 딥러닝 모델의 비교를 위해 1024개(S1), 2048개(S2), 4096개(S3), 8192(S4) 포인트로 다운샘플링을 수행하였다. 데이터세트는 7:3 비율로 학습데이터와 검증데이터로 구분하였으며, 라이다 플랫폼과 다운샘플링 강도에 따라 8종류의 데이터세트를 구축

하였다. 딥러닝 모델은 PointNet과 PointNet++를 이용하였으며, 검증데이터를 이용하여 혼동 행렬을 작성하여 수종 분류 모델의 정확도를 비교·평가하였다(Figure 2).

1) 라이다를 이용한 PCD 취득

TLS를 이용하여 수집된 PCD는 표본지당 9회 스캐닝을 진행하였으며, 스캐닝 위치는 중심점과 중심점으로부터 11.3 m를 기준으로 8방위 지점으로 설정하였다(Bauwens et al., 2018; Chen et al., 2021). 또한, 각 스캐닝 위치에는 PCD 정확도와 지상기준점(Ground Control Point, GCP)을 이용한 기하 보정을 위해 말뚝을 설치하고 Real-Time Kinematic (RTK)을 이용하여 위치 정보를 취득하였다. MLS를 이용하여 수집된 PCD는 사전에 표준지 내에 수직으로 24 m, 수평으로 3 m의 간격으로 경로를 설정하였으며, 배낭에 MLS를 부착하여 경로를 따라 스캐닝을 수행하였다 (Hartley et al., 2022; Del Perugia et al., 2019). 마지막으로 표준지 내 폐색 효과를 최소화하기 위해 경로의 대각선 방향으로 데이터를 취득하였다.

2) PCD 전처리 및 개체목 단위의 PCD 구축

개체목 단위의 수종 분류 딥러닝 모델을 구축하기 위해서는 수집된 표준지 단위의 PCD를 개체목 단위의 PCD로 분할하는 과정이 필요하다. 개체목 단위의 PCD 구축은 수집된 PCD에서 노이즈 제거, 개체목 분할, 육안검토 순으로 수행되었다. 노이즈 제거는 LiDAR 데이터 취득 시 발생하는 대상지 외의 불필요한 포인트를 대상으로 제거하였다. 노이즈 제거 방법은 임분 상단부 또는 지면 하단부

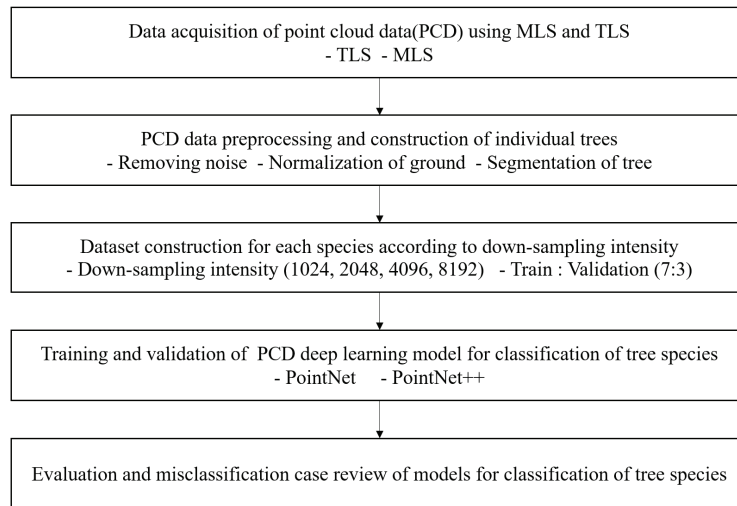


Figure 2. Flowchart of overall research for tree species classification using LiDAR point cloud data and deep learning.

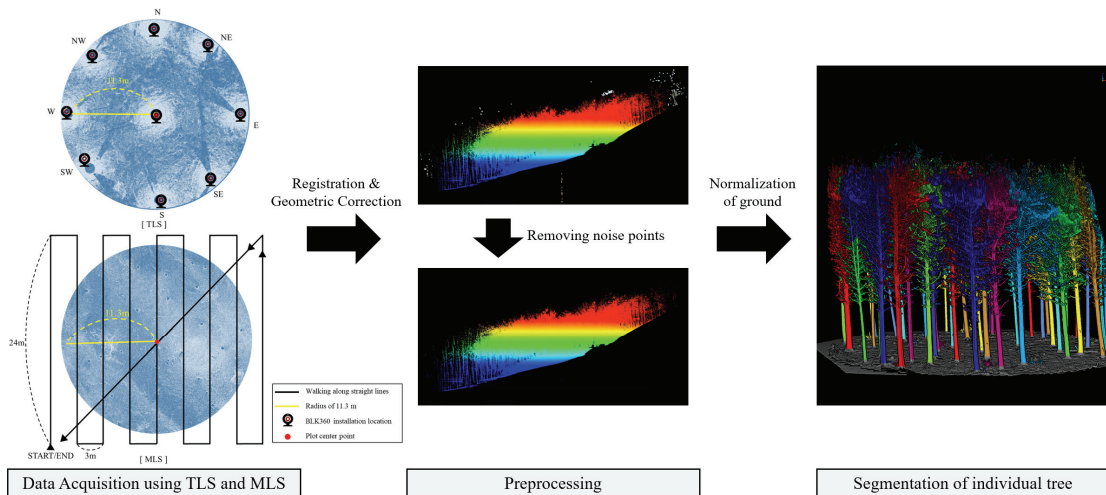


Figure 3. Data Acquisition and Preprocessing in the tree species dataset.

에 분포한 노이즈 포인트를 육안판독을 통해 수동으로 제거한 후에 Statistical Outlier Removal (SOR) 방법을 적용하였다. SOR은 각 이웃 점과의 거리의 평균과 표준편차 (std)를 산출하고, 표준편차가 기준값보다 크면 노이즈 포인트로 간주하여 제거한다. 본 연구에서는 이웃점의 수는 10, 기준값은 $5 \times \text{std}$ 로 설정하였다.

개체목 단위의 PCD 분할을 위해서는 포인트 클라우드 정규화(Point cloud normalization) 과정이 선행되어야 한다. 포인트 클라우드 정규화는 DEM의 z값을 포인트 클라우드의 z값에 차하여 지형 기록의 영향을 제거하는 과정을 의미한다. DEM은 지면의 포인트를 대상으로 추출한 후에 Kriging 기법을 이용하여 구축되었으며, 해상도는 0.5 m로 구축하였다(Zhao et al., 2016). 포인트 클라우드 정규화가 수행된 표본지 단위의 PCD는 DBSCAN과 Comparative

shortest-path (CSP)을 이용하여 수간 부분을 탐지하고, 개체목 단위의 PCD로 분할되었다. 노이즈제거와 개체목 단위의 PCD 분할 과정은 LiDAR360 (GreenValley International, California, USA) 프로그램을 이용하였다(Tao et al., 2015). 마지막으로, 모든 개체목 단위의 PCD에 대해 육안판독을 수행하였으며, 분류되지 않은 포인트, 노이즈, 지면점, 하층 식생 등을 제거하여 최종적으로 개체목 단위의 PCD를 검토했다(Figure 3).

3) 다운샘플링 강도에 따른 잣나무와 낙엽송 데이터셋 구축

잣나무와 낙엽송 데이터셋은 플롯 내에 위치한 모든 개체목을 대상으로 구축되었으며, MLS와 TLS 데이터에서 동일한 개체목을 식별하고, 분석에 용이하도록 번호를

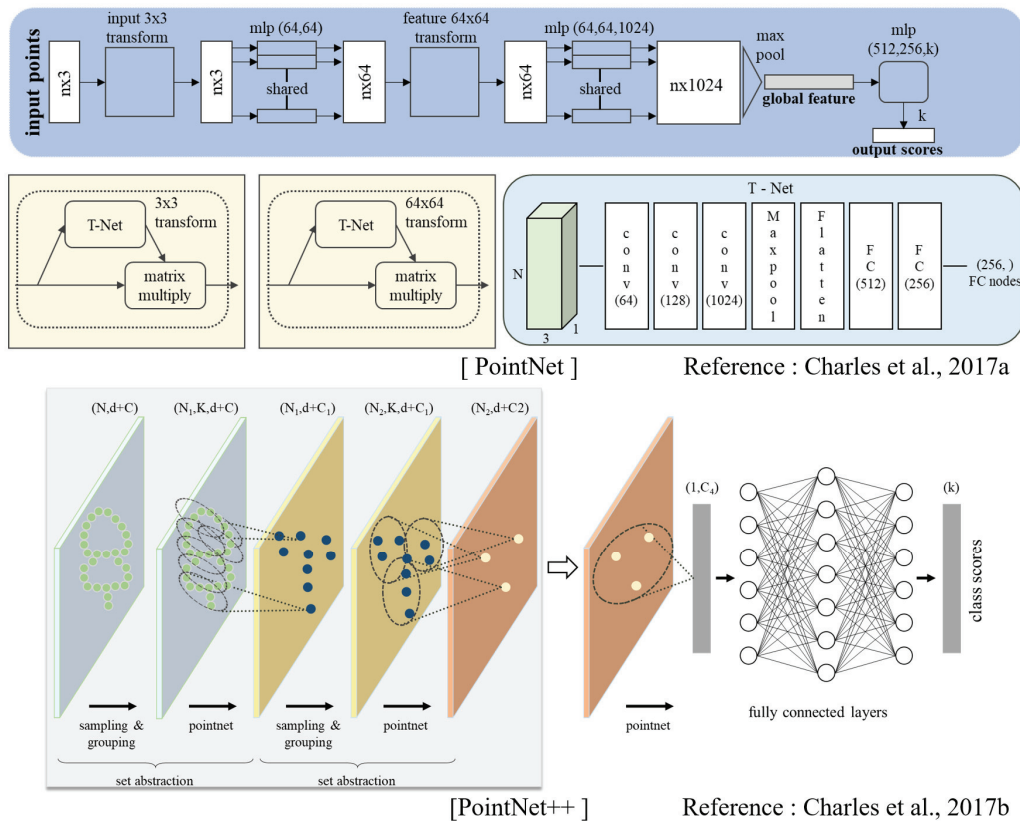


Figure 4. Architecture visualization of PCD deep learning models.

지정하였다. 구축된 데이터세트는 무작위 추출 방법으로 7:3 비율을 적용하여 학습데이터와 검증데이터로 구분하였다. PCD 딥러닝 모델을 구축하기 위해서는 입력데이터의 포인트 개수가 동일해야 하며, 입력데이터가 커질수록 입력 및 출력에 시간이 오래 소요되는 단점이 있다. 따라서, 데이터세트의 포인트 수를 통일하고, 간소화하기 위해 다운샘플링을 적용하였다. 이 과정에서 각 데이터의 공간적인 위치 정보를 제거하기 위해 각 개체목의 xyz좌표의 평균값이 원점에 위치하도록 하였으며, 각각의 xyz 좌표 값들에 대해 max값으로 나누어 주는 정규화 과정이 포함되었다. 다운샘플링은 선행연구에서 모델의 정확도가 가장 높게 분포했던 Farthest point sampling (FPS) 방법을 이용하였다(Liu et al., 2022). FPS 방법을 이용한 다운샘플링은 Python 프로그램을 이용하였다. 이에 따라, 본 연구에서는 라이다 플랫폼(TLS, MLS), 다운샘플링 강도(S1, S2, S3, S4)에 따라 8종류의 데이터세트를 구축하였다(Chen et al., 2021).

4) 수종 분류를 위한 딥러닝 모델의 학습 및 검증
수종 분류를 위한 PCD 딥러닝 모델은 최근 PCD를 이용한 분류, 객체분할 문제에서 많이 활용되고 있으며, 입력

자료가 PCD로 구성된 PointNet과 PointNet++를 이용하였다. PointNet 모델은 구조에서 T-Net과 matrix multiply 단계로 구성된 2개의 Transform 단계에서 PCD의 방향과 스케일에 대한 보정이 적용되며, PointNet은 모델의 구조 특성상 PCD 내에서 전역적인 특성을 이용하여 분류가 수행된다(Qi et al., 2017a; Sakharova et al., 2022). 한편, PointNet++ 모델은 샘플링 계층, 그룹화 계층, PointNet 계층으로 구성되어 있으며, 첫 번째로 수행되는 샘플링 및 그룹화 계층의 경우에는 입력된 데이터를 다운샘플링하여 512개의 대표 포인트를 선정하고, 선정된 포인트를 기준으로 입력된 데이터의 포인트가 가까운 점들을 그룹화하는 작업이 수행된다. 그리고, 각 그룹의 포인트들에 PointNet을 이용하여 특징을 추출한다. 이러한 과정의 반복을 통해 PointNet++는 지역적인 특성을 모델 학습 과정에서 반영할 수 있다(Qi et al., 2017b; Briechele et al., 2020; Shin et al., 2022; Chen et al., 2021)(Figure 4).

PointNet과 PointNet++는 분류 모델과 객체분할 모델로 구분되며, 본 연구에서는 개체목 단위의 PCD에 대해 수종을 분류하는 문제에 적합한 분류모델을 이용하였다. 두 모델의 하이퍼 파라미터는 학습횟수 200회, 배치사이즈 16, 최적화 함수 Adaptive-moment estimation (Adam), 학습률

0.001로 설정하여 학습과 검증을 수행하였다. 최적화 함수로 사용한 Adam은 이전 학습 단계의 조정 방향을 활용하고, 학습이 진행됨에 따라 학습률을 조정하여 안정적인 최적화가 가능한 장점을 가지고 있다(Zhang et al., 2022). 또한, PointNet++ 모델의 샘플링 계층과 그룹과 계층에서는 각각 FPS방법과 Multi-scale grouping 방법을 이용하였다(Liu et al., 2022). 이에 따라, 수종 분류 모델은 8가지 종류의 데이터세트와 PointNet과 PointNet++ 모델을 조합하여 총 16개 모델을 구축하였으며, 딥러닝 모델의 학습 및 검증은 Python 프로그램의 PyTorch 라이브러리를 이용하였다.

5) 수종 분류모델의 평가

수종 분류모델의 평가는 검증데이터를 이용하였으며, 검증데이터의 참값과 예측값을 이용하여 혼동 행렬(confusion matrix)을 작성하였다. 혼동 행렬을 기반으로 전체정확도(Overall accuracy; OA), 생산자정확도(Producer's accuracy; PA), 사용자정확도(User's accuracy; UA), F-score와 카파계수(kappa)를 산출하여 모델을 평가하였다(Ramadhani et al., 2020)(Table 3). OA는 올바르게 분류된 샘플의 수를 모든 샘플의 수로 나눈 값이며, 간단하게 성능을 평가할 수 있는 장점이 있다. PA는 참값과 분류결과가 정확하게 일치하는 비율을 의미하며, UA는 분류된 결과 중에서 참값과 얼마나 일치하는지를 의미한다. F-score는 PA와 UA의 가중 평균이며,

데이터 라벨이 불균형할 때, 모델의 성능을 정확하게 파악할 수 있는 장점이 있다. 카파계수는 혼동 행렬을 기반으로 계산되며 일관성 검증에 사용된다(Landis and Koch, 1977). 또한, 분석시간에 대한 비교를 위해 다운샘플링과 모델의 학습 과정에 대한 소요 시간을 측정하고 비교하였다.

마지막으로, 수종 분류 모델의 오분류 원인을 도출하기 위해 검증데이터에 해당하는 개체목들을 모든 모델에서 바르게 분류한 그룹(True)과 하나의 모델 이상에서 오분류된 그룹(False)으로 구분하였다. True 그룹과 False 그룹의 흉고직경과 수고에 차이가 있는지 검정하기 위해 t-test를 시행하였다. 또한, 오분류된 개체목의 PCD에 대한 검토를 수행하고 수종 분류 모델의 개선방안을 제시하였다.

결과 및 고찰

1. 잣나무와 낙엽송의 PCD 데이터세트 구축

본 연구에서 구축한 데이터세트는 총 263개의 개체목 단위의 PCD로 구성되었으며, 무작위 추출법을 이용하여 학습데이터(185개)와 검증데이터(78개)로 구분되었다. 학습데이터는 잣나무 130개, 낙엽송 55개, 검증데이터는 잣나무 55개, 낙엽송 23개로 구성되었다. 구축된 각 개체목 단위의 PCD는 포인트 수가 TLS의 경우 최소 992천개에서 최대 10,292천개까지 분포한 반면, MLS는 최소 182천

Table 3. Evaluation of confusion matrix for tree species classification models.

Class	Validation data				
	P	L	Σ	UA(%)	
Classified by deep learning model	P	N_{PP}	N_{PL}	ΣPR	$N_{PP}/\Sigma PR \times 100\%$
	L	N_{LP}	N_{LL}	ΣLR	$N_{LL}/\Sigma LR \times 100\%$
	Σ	ΣPP	ΣPL	N	
	PA(%)	$N_{PP}/\Sigma PP \times 100\%$	$N_{LL}/\Sigma PL \times 100\%$		

$$F\text{-score} = 2 \times \frac{PA \times UA}{PA + UA}$$

$$OA = \frac{N_{PP} + N_{LL}}{N} \times 100\%$$

$$kappa = \frac{N \times (N_{PP} + N_{LL}) - (\Sigma PP \times \Sigma PR + \Sigma PL \times \Sigma LR)}{N^2 - (\Sigma PP \times \Sigma PR + \Sigma PL \times \Sigma LR)}$$

Note: N is total points, R is test classes, and P is classified class.

※ P: *Pinus koraiensis*, L: *Larix kaempferi*

Table 4. Distribution of tree species dataset with number of point cloud data.

Tree Species	Number		Number of points(unit: k)						
	Total	Train	Test	TLS			MLS		
				Min	Max	Mean	Min	Max	Mean
<i>Pinus koraiensis</i>	185	130	55	568	10,292	2,832	198	2,915	771
<i>Larix kaempferi</i>	78	55	23	992	9,309	4,156	182	2,264	1,072

개에서 최대 2,915천개로 평균 포인트 수가 TLS가 MLS에 비해 약 1.4배 많았다(Table 4).

개체목 단위의 PCD에서 잣나무의 구조적인 특징은 수고가 높아짐에 따라 수관이 점점 넓어지고 초두부에 가까워질수록 다시 좁아지는 반면, 낙엽송은 잣나무보다 수고에 따른 수관 폭이 일정하게 발달하였다. 이러한 수종별 특징들은 항공사진을 이용한 수종판독 시 잣나무의 경우 우산형, 낙엽송의 경우 원추형의 구조를 갖는다는 점과 유사하였으며, 다운샘플링을 적용하였을 때도 이러한 특징들이 유지되었다(Lim et al., 2019)(Figure 5).

2. PCD 딥러닝 모델의 학습 및 검증

플랫폼에 따른 최적 모델은 TLS는 다운샘플링 S4, PointNet++을 적용한 모델이 검증정확도가 97.4%로 가장

높았으며, MLS는 다운샘플링 S2, PointNet++을 적용한 모델이 검증정확도가 98.7%로 플랫폼별 최적의 모델 기준, MLS가 TLS보다 약 1.3% 우수하였다.

PCD 딥러닝 모델의 학습 및 검증 그래프에 따르면, PointNet은 학습이 진행됨에 따라 학습 손실은 지속적으로 감소하였지만 검증 손실 값은 감소하지 않았다. 이에 따라, PointNet에서는 검증데이터에 대한 성능 개선이 이루어지지 않아 학습데이터에 대한 과적합이 발생한 것으로 판단된다. 또한, 학습횟수 100회~200회에서의 검증정확도의 추이를 보면, PointNet++의 변동계수는 평균 약 33%로 PointNet에 비하여 약 24% 높게 분포하였다. 하지만, 200회까지의 학습 중 검증 정확도가 가장 높았던 두 모델을 비교해보면, PointNet++가 PointNet보다 약 8% 우수하였다(Figure 6). 이는 모델의 구조적인 특징에 따라

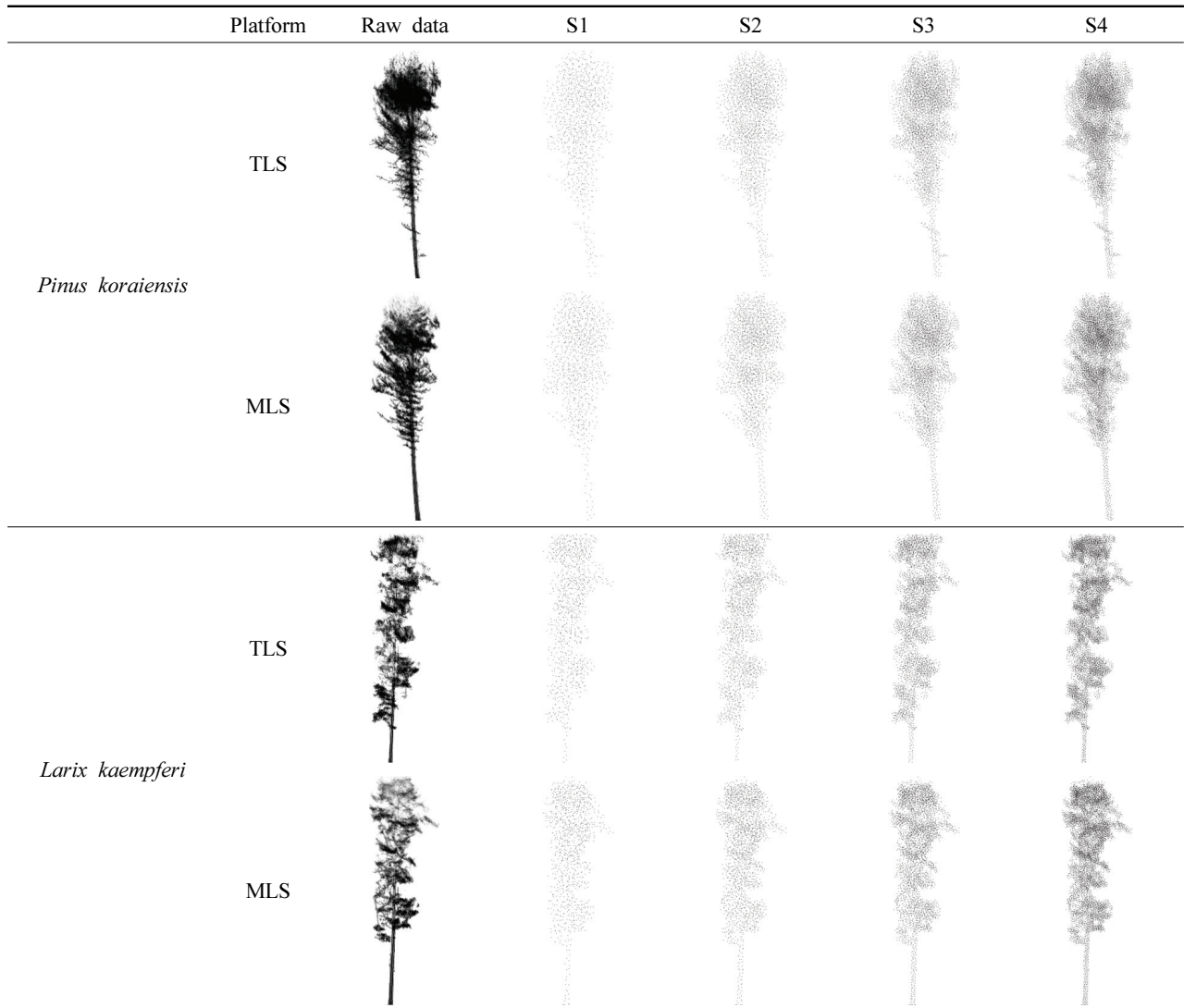


Figure 5. The samples of individual tree point clouds obtained by the farthest point sampling method down-sampling on TLS and MLS data.

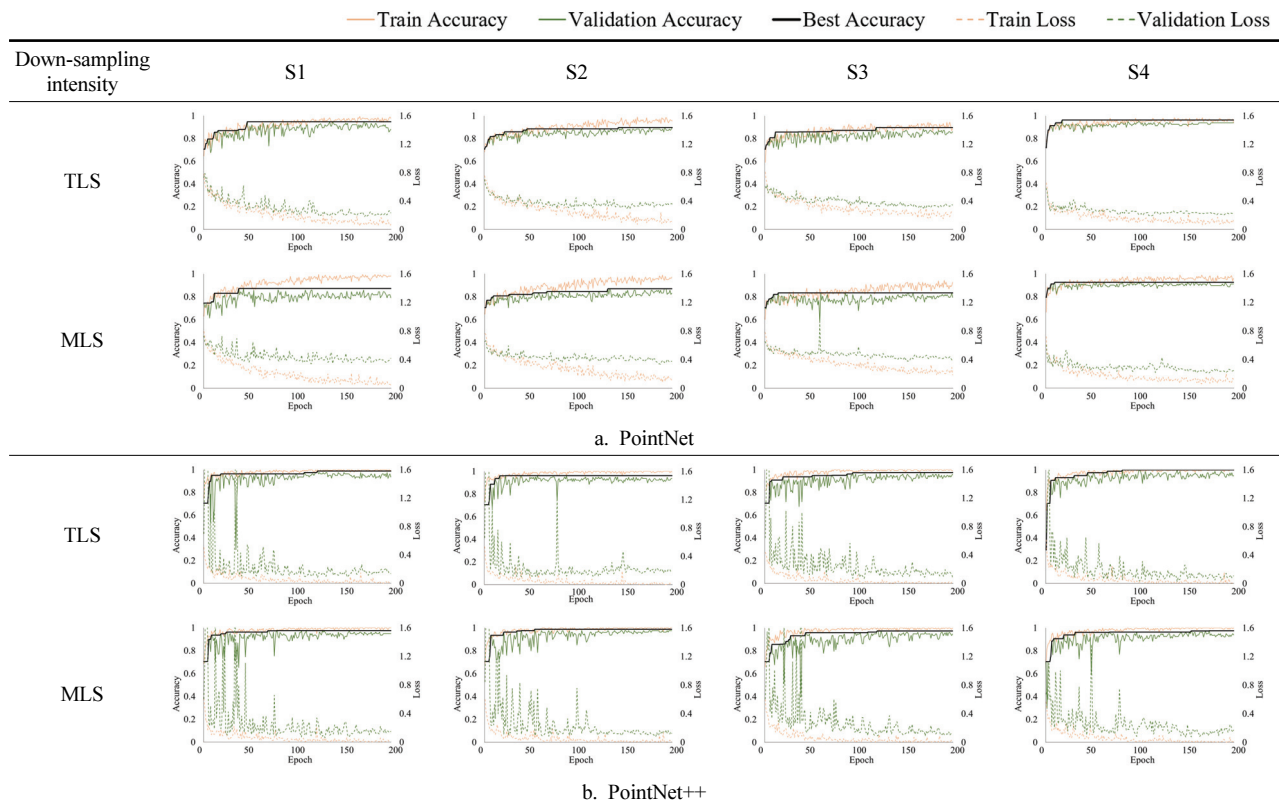


Figure 6. Accuracy and loss curves of tree species classification using two deeplearning models with LiDAR platform and down-sampling method.

PointNet은 개체목에 대해 전역적인 특성을 이용하여 수종을 분류하지만, PointNet++는 학습이 진행되면서 개체목의 특징을 잘 나타내는 지역적인 특성을 찾아내는 과정이 포함된 것이 원인으로 판단된다. 이에 따라, 모델의 특성을 고려하여 검증정확도의 변동을 최소화하고 최적의 모델을 구축하기 위해서는 최적화 함수와 학습률에 대한 하이퍼파라미터 최적화 작업이 필요하다고 판단된다 (Soydaner, 2019). 선행연구에 따르면, 수종 분류에 있어 학습 과정에서 PointNet보다 PointNet++가 약 10% 정도 더 높은 성능을 보였다. 본 연구에서는 두 모델 간의 정확도 차이가 약 7%로 분포하였다. 이는 분류 대상 수종이 2가지로 선행 연구에 비해 분류 범주가 작기 때문인 것으로 판단된다(Liu et al., 2022; Xi et al., 2020; Briechle et al., 2020; Shin et al., 2022; Chen et al., 2021).

3. 수종 분류모델의 정확도 평가

TLS를 이용하여 구축된 수종 분류 모델 중 다운샘플링 S4, PointNet++을 이용한 모델의 카파 계수가 93.7%로 가장 높았으며, MLS는 다운샘플링 S2, PointNet++을 이용한 모델이 96.9%로 가장 높았다. 특히, 구축된 수종 분류 모델에서 다운샘플링 강도와 플랫폼이 동일한 모든 경우에

서 PointNet++가 PointNet보다 더 높은 정확도로 산출되었다. 따라서, 잣나무와 낙엽송의 분류 모델은 PointNet++이 적절한 것으로 판단된다.

한편, PointNet++을 이용한 모델을 기준으로 다운샘플링 강도에 따른 검증정확도의 변화는 TLS와 MLS의 패턴이 상이하였다. TLS는 다운샘플링이 S1에서 S2로 증가할 때 검증정확도가 감소하였으나 이후에는 증가하여 S4일 때 검증정확도가 가장 높았다. 하지만, MLS는 일정한 패턴 없이 다운샘플링 강도에 따른 정확도가 변화하여 S2의 검증정확도가 높았다. 이와 같이 최적의 다운샘플링 강도에 따른 정확도의 경향은 수종 분류 모델을 구축하는 환경에 따라 달라질 수 있다. PointNet++는 포인트의 개수가 증가할수록 그룹화 계층에서 선정되는 그룹 내의 포인트 간격이 좁아지는 특성이 있다. 이에 따라, 데이터가 정밀하게 수집되는 TLS는 지역적인 특징이 반영되어 다운샘플링 S4에서 정확도가 향상되었다. 하지만, MLS는 TLS에 비해 PCD의 밀도와 정확도가 낮기 때문에 정확도의 향상이 이루어지지 않은 것으로 판단된다(Chen et al., 2021). 한편, 잣나무와 낙엽송의 F-Score은 모든 모델에서 잣나무가 낙엽송보다 높게 분포하였으며, 이는 수종 간의 데이터셋 수의 불균형으로 인한 것으로 사료된다(Table 5, Table 6).

Table 5. Confusion matrix of PCD deep learning models according to platform and down-sampling intensity.

Model & Platform	Ground truth																
	S1			S2			S3			S4							
Down-sampling intensity	<i>Pinus koraiensis</i>	<i>Larix kaempferi</i>	Total	PA	<i>Pinus koraiensis</i>	<i>Larix kaempferi</i>	Total	PA	<i>Pinus koraiensis</i>	<i>Larix kaempferi</i>	Total	PA	<i>Pinus koraiensis</i>	<i>Larix kaempferi</i>	Total	PA	
Pointnet	<i>Pinus koraiensis</i>	54	3	57	94.7%	53	6	59	89.8%	53	6	59	89.8%	54	2	56	96.4%
	<i>Larix kaempferi</i>	1	20	21	95.2%	2	17	19	89.5%	2	17	19	89.5%	1	21	22	95.5%
	Total	55	23	78		55	23	78		55	23	78		55	23	78	
	UA	98.2%	87.0%		96.4%	73.9%		96.4%	73.9%	98.2%	91.3%		98.2%	91.3%			
Pointnet++	<i>Pinus koraiensis</i>	53	8	61	86.9%	52	7	59	88.1%	50	8	58	86.2%	55	2	57	96.5%
	<i>Larix kaempferi</i>	2	15	17	88.2%	3	16	19	84.2%	5	15	20	75.0%	0	21	21	100.0%
	Total	55	23	78		55	23	78		55	23	78		55	23	78	
	UA	96.4%	65.2%		94.5%	69.6%		90.9%	65.2%	100.0%	91.3%		100.0%	91.3%			
Pointnet++	<i>Pinus koraiensis</i>	54	2	56	96.4%	51	1	52	98.1%	53	2	55	96.4%	52	3	55	94.5%
	<i>Larix kaempferi</i>	1	21	22	95.5%	4	22	26	84.6%	2	21	23	91.3%	3	20	23	87.0%
	Total	55	23	78		55	23	78		55	23	78		55	23	78	
	UA	98.2%	91.3%		92.7%	95.7%		96.4%	91.3%	94.5%	87.0%		94.5%	87.0%			
Pointnet++	<i>Pinus koraiensis</i>	54	2	56	96.4%	55	1	56	98.2%	52	3	55	94.5%	54	4	58	93.1%
	<i>Larix kaempferi</i>	1	21	22	95.5%	0	22	22	100.0%	3	20	23	87.0%	1	19	20	95.0%
	Total	55	23	78		55	23	78		55	23	78		55	23	78	
	UA	98.2%	91.3%		100.0%	95.7%		94.5%	87.0%	94.5%	87.0%		98.2%	82.6%			

Table 6. Comparison of classification accuracy by PCD deep learning models.

Unit: %

Model	Platform	Down-sampling intensity	OA	Kappa	F-Score	
					<i>Pinus koraiensis</i>	<i>Larix kaempferi</i>
PointNet	TLS	1024	94.9	87.3	96.4	90.9
		2048	89.7	74.0	93.0	81.0
		4096	89.7	74.0	93.0	81.0
		8192	96.2	90.6	97.3	93.3
	MLS	1024	87.2	66.6	91.4	75.0
		2048	87.2	67.5	91.2	76.2
		4096	83.3	58.3	88.5	69.8
		8192	92.3	81.5	94.5	87.0
PointNet++	TLS	1024	96.2	90.6	97.3	93.3
		2048	93.6	85.1	95.3	89.8
		4096	94.9	87.7	96.4	91.3
		8192	97.4	93.7	98.2	95.5
	MLS	1024	96.2	90.6	97.3	93.3
		2048	98.7	96.9	99.1	97.8
		4096	92.3	81.5	94.5	87.0
		8192	93.6	84.0	95.6	88.4

Table 7. Time required to down-sampling and training and validation for PointNet and PointNet++.

Unit: minute

Content of analysis	Down-sampling by FPS				Training and validation							
					PointNet				PointNet++			
Down-sampling intensity	S1	S2	S3	S4	S1	S2	S3	S4	S1	S2	S3	S4
TLS	83	159	312	617	42	43	44	49	59	61	60	69
MLS	21	41	81	161	42	42	44	49	60	61	61	69

수종 분류 분석과정의 소요시간은 다운샘플링 단계에서 평균 약 184분, 모델 학습 및 검증 단계에서 평균 약 83분으로 다운샘플링 단계가 평균 약 2.2배 더 오래 소요되었다. 다운샘플링의 강도에 따른 총 분석 소요시간은 TLS가 MLS보다 3.9배 오래 소요되었으며, 다운샘플링의 포인트 수가 S4가 S1에 비해 TLS에서 7.4배, MLS에서 7.6배 오래 소요되었다. 이처럼 FPS를 이용한 다운샘플링은 원본데이터의 포인트 수와 다운샘플링의 포인트 수가 많을수록 시간이 증가하였다. 모델의 학습 및 검증 시간은 포인트의 수가 증가할수록 소요시간이 증가하는 패턴을 보였으나, 포인트 수보다 모델의 구조 차이에 따라 학습시간에 미치는 영향이 컸다. 모델별 평균 소요시간을 측정하였을 때, PointNet++는 PointNet보다 약 1.4배 더 오래 소요되었다. 이러한 점은 PointNet++이 모델의 크기를 의미하는 파라미터 수가 PointNet보다 더 큰 것이 원인으로 판단된다. 다운샘플링 과정과 학습 및 검증의 소요시간을 비교한 결과, 효율적인 수종 분류모델의 구축 및 적용을 위해서는

데이터의 구축과정에 대한 검토가 필요하며, 다양한 샘플링기법의 적용 및 평가를 통한 효율화가 필요할 것으로 판단된다(Table 7).

4. 데이터세트의 유형에 따른 오분류 검토

True 그룹과 False 그룹의 t-test 결과, 유의수준 0.1에서 두 그룹 간의 모집단 평균이 유의한 차이가 없었다(Table 8) (Figure 7). 이에 따라, 개체목의 흉고직경과 수고는 오분류와 관계가 없는 것으로 판단된다. 이는 선행연구에서 수고가 개체목 분류 정확도에 영향을 끼치지 않는다는 결과와 일치 하였으며, 추가적으로 흉고직경도 동일하게 분류 정확도에 영향을 끼치지 않는 결과를 도출하였다(Liu et al., 2022). 이처럼 개체목의 크기와 관련된 인자가 정확도에 영향을 끼치지 않는 원인은 크게 두 가지로 판단된다. 첫 번째는 개체목의 다운샘플링 과정에서 수행된 정규화 과정을 통해 개체목 크기의 스케일이 조정되었다. 이 과정에서 평균과 비교하여 흉고직경과 수고의 차이가 있는 개

Table 8. Comparative t-test analysis of DBH and tree height for true and false classifications in *Pinus koraiensis* and *Larix kaempferi*

Categories		Observations	Mean	Variance	t-value	p-value(two-tail)	
<i>Pinus koraiensis</i>	DBH	True	40	19.0	0.10	0.92	
		False	15	18.9			11.8
	Tree Height	True	40	26.9	45.9	-1.59	0.13
		False	15	31.0	78.9		
<i>Larix kaempferi</i>	DBH	True	6	25.5	10.2	0.65	0.53
		False	17	24.1	52.3		
	Tree Height	True	6	30.6	37.7	0.33	0.75
		False	17	29.5	64.3		

Downsampling intensity ○ 1024(S1) × 2048(S2) □ 4096(S3) △ 8192(S4)

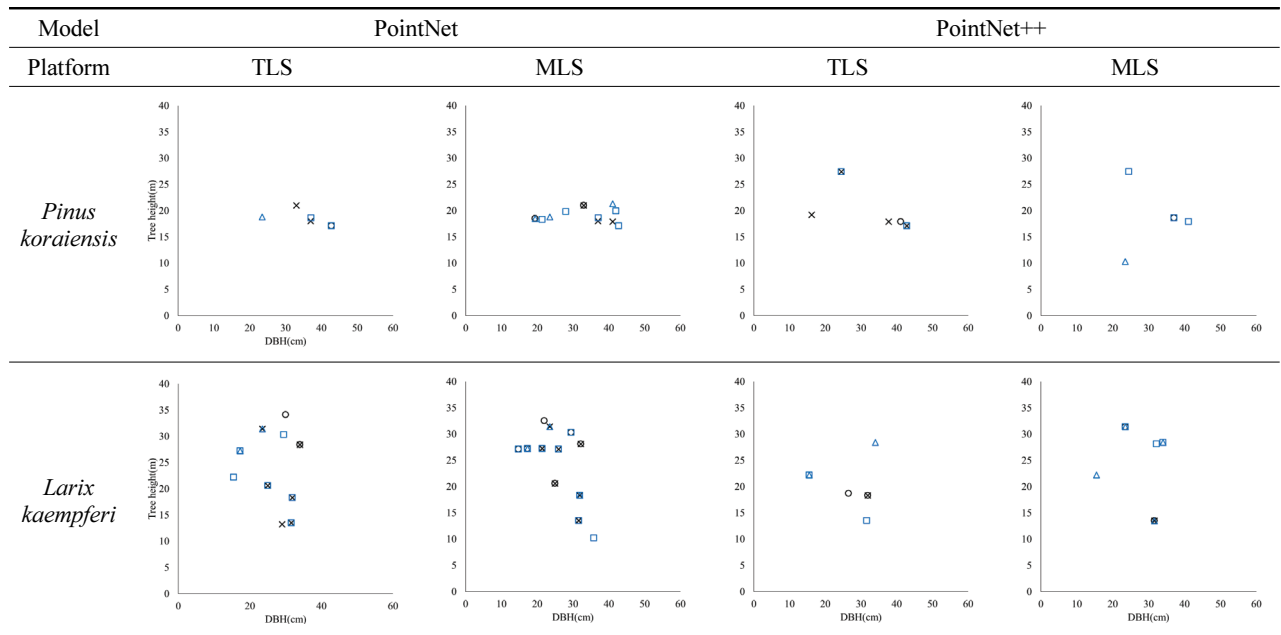


Figure 7. The distribution of DBH and Tree height in misclassified tree species.

체목들은 정규화 후에 유사하게 나타나기 때문에 흉고직경과 수고의 정보는 일부 손실된다. 두 번째는 PointNet과 PointNet++ 모델에서 회전과 기하학적 변환의 영향을 최소화하기 위한 Transform 단계에서 스케일이 다른 개체목에 대한 보정이 수행되기 때문이다.

오분류된 개체목의 PCD를 검토한 결과, 주요 오분류 사례로 3가지 유형(A, B, C)을 선정하였다. 유형 A는 각 수종의 구조적인 특징이 나타나지 않고, 개체목 단위의 PCD가 잣나무와 낙엽송이 유사한 개체목에 관한 오분류 사례이다. PointNet 모델에서 유형 A의 오분류가 주로 발생하였으며, 모델의 구조가 개선된 PointNet++에서는 유형 A 개체목에 대해 바르게 분류하였다. 또한, 유형 B는 편심성장한 개체목을 오분류한 사례로서, 주로 경사지 또는 입도

부근에 위치한 개체목으로 편심생장의 영향으로 인해 수관이 한 방향으로 치우쳐진 구조적인 특징이 있었다. 유형 C는 개체목 분할 시 수관부가 수직 방향으로 분할된 개체목 중에서 발생하였다. 수관부가 수직으로 분할된 TLS 자료는 8개 모델 중 3개 모델에서 오분류가 발생했지만, 정상적으로 분할된 MLS 자료는 8개 모델 중 1개 모델에서 오분류가 발생하였으며, 입분 밀도가 높은 잣나무림에서 주로 발생하였다(Chen et al., 2019, Ko, 2021)(Figure 8).

오분류 검토 결과를 종합하면, 두 수종 간의 구조적인 특징이 유사하거나 주변 생육 환경에 의해 특징이 나타나지 않는 개체목들이 오분류 되었다. 따라서, 위와 같은 오분류 사례를 개선하기 위해서는 다양한 환경에서 생육하고 있는 개체목들에 대한 데이터 수집이 필요하다고 판단된다. 또한,













Type	A		B		B	
Platform	TLS	MLS	TLS	MLS	TLS	MLS
<i>Pinus koraiensis</i>						
<i>Larix kaempferi</i>						

Figure 8. The Samples of individual tree point clouds in misclassification types.

입력데이터를 포인트들의 위치 정보만 입력하는 기존 모델에서 라이다의 반사강도, 색상 값 등 다양한 입력자료를 활용하는 수종 분류 모델의 개발이 필요하다고 생각된다.

결론

본 연구는 TLS와 MLS를 이용하여 잣나무와 낙엽송을 대상으로 개체목 단위의 PCD를 구축하고 라이다 플랫폼, 다운샘플링 강도, PCD 딥러닝 모델에 따라 수종분류 모델을 구축 및 평가하였다. 수종분류 모델의 구축 결과, TLS를 이용하였을 때는 S4 포인트로 다운샘플링된 자료에 PointNet++을 적용한 모델, MLS는 S2 포인트에 PointNet++을 적용한 모델의 정확도가 가장 우수하였다. 정확도가 가장 높은 모델을 기준으로 MLS를 이용하여 구축된 수종분류 모델이 TLS보다 카파계수가 약 1.3% 높았으며, MLS 데이터 구축과정에서 정합하는 과정이 생략되어 데이터셋 구축에 용이하였다. 따라서, 잣나무와 낙엽송의 분류에 있어서는 TLS 보다 MLS의 활용 가능성이 높다고 판단된다. 하지만, 다양한 수종의 데이터셋이 구축된다면 고해상도의 TLS 자료의 정확도가 더 높게 산출될 가능성이 있기 때문에 TLS와 MLS 자료에 대한 지속적인 수집과 연구가 필요하다고 사료 된다.

다운샘플링 강도는 모델을 구축하는 환경에 따라 최적의 값이 달라지기 때문에 향후, 수종분류 모델 구축 시 다운샘플링 강도를 최적화하는 단계가 필요하다고 판단

된다. 또한, 다운샘플링의 최적화 단계에서는 다운샘플링 강도에 따라 소요 시간이 급격하게 변화하는 점을 고려해야 한다. 마지막으로, PCD 딥러닝 모델은 동일한 플랫폼과 다운샘플링 강도를 이용하였을 때, 모든 모델에서 PointNet++의 정확도가 PointNet보다 높았기 때문에 수종분류 모델에는 PointNet++가 적합하다고 판단된다.

본 연구에서 구축된 16개의 모델을 검증데이터에 적용하여 도출된 오분류 사례에 대해 분석한 결과, 데이터셋에 대한 보완이 필요하다고 판단된다. 본 연구의 데이터셋은 5영급 잣나무림과 낙엽송림에 한정되었으며, 환경사이에 위치하였다. 또한, 임분단위로 수집된 PCD 자료를 개체목 단위의 PCD로 구축하는 과정에서 수관 부분을 명확하게 구분하는데 어려움이 있었다. 이에 따라 향후에는 다양한 수종과 환경조건의 라이다 자료 수집을 통해 다양한 수종 및 혼효림을 대상으로 한 수종분류 모델의 적용가능성에 대해 검토해야 한다. 또한, 우리나라 수종의 특징을 추출할 수 있는 개체목 분할 알고리즘과 PCD 딥러닝 모델이 개발된다면, 산림자원조사 수종 분류에도 라이다를 활용할 수 있을 것으로 기대된다.

감사의 글

본 연구는 산림청(한국임업진흥원) 산림과학기술 연구개발사업(‘2021359B10-2323-BD01, 2019151D10-2323-0301’)의 지원에 의하여 이루어진 것입니다.

References

- Axelsson, A., Lindberg, E. and Olsson, H. 2018. Exploring multispectral ALS data for tree species classification. *Remote Sensing* 10(2): 183.
- Ballanti, L., Blesius, L., Hines, E. and Kruse, B. 2016. Tree species classification using hyperspectral imagery: a comparison of two classifiers. *Remote Sensing* 8(6): 445.
- Briechle, S., Krzystek, P. and Vosselman, G. 2020. Classification of tree species and standing dead trees by fusing UAV-based lidar data and multispectral imagery in the 3D deep neural network PointNet++. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 2: 203-210.
- Chen, J., Chen, Y. and Liu, Z. 2021. Classification of typical tree species in laser point cloud based on deep learning. *Remote Sensing* 13(23): 4750.
- Chen, S., Liu, H., Feng, Z., Shen, C. and Chen, P. 2019. Applicability of personal laser scanning in forestry inventory. *PLoS One* 14(2): e0211392.
- Chen, Y., Liu, G., Xu, Y., Pan, P. and Xing, Y. 2021. PointNet++ network architecture with individual point level and global features on centroid for ALS point cloud classification. *Remote Sensing* 13(3): 472.
- Del Perugia, B., Giannetti, F., Chirici, G. and Travaglini, D. 2019. Influence of scan density on the estimation of single-tree attributes by hand-held mobile laser scanning. *Forests* 10(3): 277.
- Guan, H., Yu, Y., Ji, Z., Li, J. and Zhang, Q. 2015. Deep learning-based tree classification using mobile LiDAR data. *Remote Sensing Letters* 6(11): 864-873.
- Hartley, R.J., Jayathunga, S., Massam, P.D., de Silva, D., Estarija, H.J., Davidson, S.J., Wuraola, A. and Pearse, G.D. 2022. Assessing the potential of backpack-mounted mobile laser scanning systems for tree phenotyping. *Remote Sensing* 14(14): 3344.
- Jones, T.G., Coops, N.C. and Sharma, T. 2010. Assessing the utility of airborne hyperspectral and LiDAR data for species distribution mapping in the coastal Pacific Northwest, Canada. *Remote Sensing of Environment* 114(12): 2841-2852.
- Ko, B.J., Park, S.I., Park, H.J. and Lee, S.H. 2022. Measurement of tree height and diameter using terrestrial laser scanner in coniferous forests. *Journal of Environmental Science International* 31(6): 479-490.
- Ko, C.U., Lee, J.W., Kim, D. and Kang, J.T. 2022. The application of terrestrial light detection and ranging to forest resource inventories for timber yield and carbon sink estimation. *Forests* 13(12): 2087.
- Korea Forest Service and Korea Forestry Promotion Institute. 2021. The 8th national forest inventory and forest health monitoring. -Field manual-. Seoul: Korea Forestry Promotion Institute. pp. 7-8.
- Landis, J.R. and Koch, G.G. 1977. The measurement of observer agreement for categorical data. *biometrics*. pp. 159-174.
- Liang, X. et al. 2018. International benchmarking of terrestrial laser scanning approaches for forest inventories. *ISPRS Journal of Photogrammetry and Remote Sensing* 144: 137-179.
- Lim, J.B., Kim, K.M., and Kim, M.K. 2019. The development of major tree species classification model using different satellite images and machine learning in Gwangneung area. *Korean Journal of Remote Sensing* 35(6): 1037-1052
- Liu, B., Chen, S., Huang, H. and Tian, X. 2022. Tree Species Classification of backpack laser scanning data using the PointNet++ point cloud deep learning method. *Remote Sensing* 14(15): 3809.
- Liu, B., Huang, H., Su, Y., Chen, S., Li, Z., Chen, E. and Tian, X. 2022. Tree species classification using ground-based LiDAR data by various point cloud deep learning methods. *Remote Sensing* 14(22): 5733.
- Liu, B., Huang, H., Tian, X. and Ren, M. 2022. Individual tree species classification using the pointwise mlp-based point cloud deep learning method. *Environmental Sciences Proceedings* 22(1): 19.
- Liu, M., Han, Z., Chen, Y., Liu, Z. and Han, Y. 2021. Tree species classification of LiDAR data based on 3D deep learning. *Measurement* 177: 109301.
- Loudermilk, E.L., Pokswinski, S., Hawley, C.M., Maxwell, A., Gallagher, M.R., Skowronski, N.S., Hudak, A.T., Hoffman, C. and Hiers, J.K. 2023. Terrestrial laser scan metrics predict surface vegetation biomass and consumption in a frequently burned southeastern US ecosystem. *bioRxiv*, 2023-01.
- Pu, R. and Landry, S. 2012. A comparative analysis of high spatial resolution IKONOS and WorldView-2 imagery for mapping urban tree species. *Remote Sensing of Environment* 124: 516-533.
- Qi, C.R., Su, H., Mo, K. and Guibas, L.J. 2017. Pointnet: deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 652-660.
- Qi, C.R., Yi, L., Su, H. and Guibas, L.J. 2017. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in Neural Information Processing*

- Systems 30: 5105-5114.
- Ramadhani, F., Pullanagari, R., Kereszturi, G. and Procter, J. 2020. Mapping of rice growth phases and bare land using Landsat-8 OLI with machine learning. *International Journal of Remote Sensing* 41(21): 8428-8452.
- Sakharova, E.K., Nurlyeva, D.D., Fedorova, A.A., Yakubov, A.R. and Kanev, A.I. 2022. Issues of tree species classification from LiDAR data using deep learning model. In *advances in neural computation, machine learning, and cognitive research V: Selected papers from the XXIII international conference on neuroinformatics*, Moscow, Russia. Springer International Publishing. pp. 319-324.
- Shin, Y. H., Son, K. W. and Lee, D. C. 2022. Semantic segmentation and building extraction from airborne LiDAR data with multiple return using PointNet++. *Applied Sciences* 12(4): 1975.
- Soydaner, D. 2020. A comparison of optimization algorithms for deep learning. *International Journal of Pattern Recognition and Artificial Intelligence* 34(13): 2052013.
- Su, H., Maji, S., Kalogerakis, E. and Learned-Miller, E. 2015. Multi-view convolutional neural networks for 3d shape recognition. In *Proceedings of the IEEE international conference on computer vision*. pp. 945-953.
- Tao, S. et al. 2015. Segmenting tree crowns from terrestrial and mobile LiDAR data by exploring ecological theories. *ISPRS Journal of Photogrammetry and Remote Sensing* 110: 66-76.
- Wessel, M., Brandmeier, M. and Tiede, D. 2018. Evaluation of different machine learning algorithms for scalable classification of tree types and tree species based on Sentinel-2 data. *Remote Sensing* 10(9): 1419.
- Wu, Z., Song, S., Khosla, A., Yu, F., Zhang, L., Tang, X. and Xiao, J. 2015. 3d shapenets: A deep representation for volumetric shapes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 1912-1920.
- Xi, Z., Hopkinson, C., Rood, S.B. and Peddle, D.R. 2020. See the forest and the trees: effective machine and deep learning algorithms for wood filtering and tree species classification from terrestrial laser scanning. *ISPRS Journal of Photogrammetry and Remote Sensing* 168: 1-16.
- Yan, S., Jing, L. and Wang, H. 2021. A new individual tree species recognition method based on a convolutional neural network and high-spatial resolution remote sensing imagery. *Remote Sensing* 13(3): 479.
- Zeybek, M. and Şanlıoğlu, I. 2019. Point cloud filtering on UAV based point cloud. *Measurement* 133: 99-111.
- Zhang, F., Tian, X., Zhang, H., and Jiang, M. 2022. Estimation of Aboveground carbon density of forests using deep learning and multisource remote sensing. *Remote Sensing* 14(13): 3022.
- Zhao, X., Guo, Q., Su, Y. and Xue, B. 2016. Improved progressive TIN densification filtering algorithm for airborne LiDAR data in forested areas. *ISPRS Journal of Photogrammetry and Remote Sensing* 117: 79-91.

Manuscript Received : February 24, 2023

First Revision : April 25, 2023

Second Revision : June 1, 2023

Accepted : June 2, 2023