Master's Thesis of Landscape Architecture

Identification of tree species and the estimation of aboveground biomass in an urban forest using multi-period airborne LiDAR with hyperspectral datasets

다중시기 항공 LiDAR와 초분광 영상을 활용한 도시림 수종 분류 및 바이오매스 추정

February **2023**

Seoul National University
Graduate School of Environment Studies
Landscape Architecture
Dae Yeol Kim

Identification of tree species and the estimation of aboveground biomass in an urban forest using multi-period airborne LiDAR with hyperspectral datasets

Prof. Song, Youngkeun

Submitting a master's thesis of Public Administration
October 2022

Seoul National University
Graduate School of Environment Studies
Landscape Architecture
Dae Yeol Kim

Confirming the master's thesis written by
Dae Yeol Kim
December 2022

Chair ______________________ (Seal)

Vice Chair ______________________ (Seal)

Examiner ______________________ (Seal)

Identification of tree species and the estimation of aboveground biomass in an urban forest using multi-period airborne LiDAR with hyperspectral datasets

서울대학교 환경대학원 환경조경학과
Dae Yeol Kim

위 논문은 서울대학교 및 환경대학원 환경조경학과 학위논문
관련 규정에 의거하여 심사위원의 지도과정을 충실히
이수하였음을 확인합니다.

2023년 2월

위 원 장 ______________ (서울대학교 환경대학원 교수)

부위원장 ______________ (서울대학교 농업생명과학대학 교수)

위 원 ______________ (서울대학교 환경대학원 교수)

Abstract

Identification of tree species and the estimation of aboveground biomass in an urban forest using multi-period airborne LiDAR with hyperspectral datasets

Dae Yeol Kim

Dept. of Landscape Architecture

Graduate School of Environmental Studies

Seoul National University

As climate change has emerged as a global concern and the proportion of urban residents has increased, the importance of urban forest as a space that relieves air pollution and urban heat islands and provides various benefits such as biomass generation, biodiversity conservation, and carbon storage has increased. Given that the amount of carbon absorption and accumulation based on the biomass calculation differs by tree species that comprise a forest, accurate tree species classification is required to quantitatively calculate urban forest benefits and manage endangered species. Regarding conventional forest monitoring, the Korea Forest Service produces and manages forest type maps by aerial image analysis and field surveys, a labor-intensive and time-consuming approach. In addition, because aerial

imaging cannot identify the vertical structure of urban forest vegetation, a method for classifying the species of the research sites and accurately distinguishing boundaries is required. Notably, many effective forest monitoring studies are being conducted using forest structure characteristics derived from airborne LiDAR and the spectral reflectance of hyperspectral images. With survey technology advancements, LiDAR point data with high density (10 point/m^2) can be obtained, point data can be easily used in open-source software, and hyperspectral images can be developed with expanded vegetation index lists and preprocessing and correction algorithms. In this research, the traditional forest survey method was improved by combining airborne hyperspectral images (AHI) with airborne LiDAR data from two periods, thereby leveraging the characteristics of each data set and seasonal characteristics of vegetation. The goal was to increase the accuracy and efficiency of tree classification, understand species distribution in urban forests by creating an environmental planning map, and calculate the research site Aboveground biomass (AGB) based on the classification results. The research site is an urban forest in Gwacheon, Gyeonggi-do, located at 37° 23'–37° 27' north latitude and 126° 57'–127° 02' east longitude, with an area of 2,034 ha and 10 major species. Forest surveys were conducted from August to October to gather field survey data for classification; the airborne LiDAR dataset was acquired during the leaf-on (November) and leaf-off (April) periods, whereas the AHI dataset was acquired during the leaf-on (September) and leaf-off (November) periods. The airborne LiDAR and AHI datasets were calibrated through preprocessing, and 29 independent variables for tree classification were extracted by calculating the PC1 band of AHI, the vegetation index related to the pigment and photosynthesis

properties of leaves, and the height of airborne LiDAR images. In addition, 165,216 points were obtained by generating 16,522 random points for 10 major species of the research site, excluding missing values. Classification and verification were performed by learning five machine learning classification models of logistic regression (LR), support vector machine (SVM), decision tree (DT), random forest (RF), and light gradient boosting machine (LGBM) based on tree species information obtained from field surveys. The tree classification results indicated that the average accuracy of the five classifiers for the multitemporal multidataset was 71%, exceeding those of the single temporal multidataset (leaf-on: 57%; leaf-off: 61%) and multitemporal single dataset (AHI: 64%; airborne LiDAR: 55%). Comparing the accuracy of each machine learning classifier on five datasets revealed that RF had the highest average accuracy (76%), followed by LGBM (70%), DT (61%), SVM (60%), and LR (39%). Consequently, the classification accuracy of multitemporal multidatasets using RF techniques was also highest (83.3%; Kappa: 0.80). The main independent variables contributing to tree classification were the CRI (Importance: 0.064) extracted from the AHI dataset acquired in November and the leaf area index (Importance: 0.062) of the airborne LiDAR images acquired in April. The diameter at breast height (DBH) of the independent tree was derived using the modified logistic tree height (TH) – DBH relational expression of 928,015 tree crown areas, which were extracted using the independent tree segmentation algorithm. By substituting TH and DBH into the allometric equations for each part of the tree volume, biomass, and stand yield table, AGB was derived for trees with at least 2 m height, and a total biomass of 45,351 t was calculated. Tree classification using airborne LiDAR and AHI could result in over 80%

accuracy when employed in urban forests, and the forest's actual seasonal characteristics were clearly increased by combining images acquired based on leaf growth, fall foliage season, and leaf fall season compared to single temporally acquired images. Overall, research into the estimation of carbon absorption and storage through the management of climate change-vulnerable species and AGB computation can benefit from tree species maps visualized in the classification results.

Keywords: : AIRBORNE LiDAR, AIRBORNE HYPERSPECTRAL IMAGING, URBAN FOREST, TREE SPECIES CLASSIFICATION, ABOVEGROUND-BIOMASS (AGB), MULTI-TEMPORAL

Student Number : 2021-22194

Contents

[List of Tables, Figures and Equations]

▌ List of Tables

▌ List of Figures

▌List of Equations

Chapter 1. Introduction

1.1. Research Background and Purpose

1.1.1. Background of the Research

1) Climate Change and Carbon Neutrality

Climate change has become a global concern, with increasing atmospheric carbon dioxide (CO_2) concentration being identified as the main cause. Carbon-based organic matter, a material that is the basis of life, accounts for approximately half of the mass of all living organisms and is primarily found as fossil fuel in sedimentary rocks and as CO_2 in the atmosphere. Before the Industrial Revolution, the CO_2 concentration averaged 280 parts per million (ppm) for approximately 6,000 years. However, in 2022, it reached 412.5 ppm, an increase of more than 50%. In 2017, the average global temperature reached 1° C above pre-industrial levels.

Owing to their global occurrence, climate change-induced problems, such as heat waves, droughts, floods, natural disasters, food shortages, spread of pests, species extinction, and ecosystem changes, have emerged as global concerns. The Intergovernmental Panel on Climate Change (IPCC) is an agency under the UN, and countries around the world are implementing policies aimed at realizing carbon neutrality by 2050. In addition, to comply with international regulations, methods for quantifying carbon absorption and storage are being actively discussed.

2) Urban Forest and Forest Monitoring Using Remote Sensing

The Creation And Management Of Urban Forest Act (National Assembly of the Republic of Korea, 2020) defines urban forests as forests and trees created and managed in cities to promote public health and recreation, cultivation of emotions, and experiential activities. Urban forest is a concept that includes urban forests, parks, and street trees. According to the UN 2011 report, the share of the global population residing in cities would be estimated at 56.15% by 2020 owing to rising migration to cities. In Korea, 91.4% of the population based on administrative districts lives in the city (Ministry of Land, Infrastructure and Transport, 2021). Notably, urban forests are becoming increasingly important as a space that provides various benefits such as carbon storage, biomass generation, air pollution mitigation, heat island reduction, and biodiversity conservation (Escobedo et al., 2011).

Since the emergence of the global COVID-19 pandemic, the demand for green spaces in cities, which had been previously neglected, has increased. Therefore, the management and restoration of natural spaces that play a multifunctional role in the city are becoming increasingly crucial. The time has come to integrate smart technology into quality management of green spaces. With the developed countries government's smart green city policy, discussions on environmental management and urban ecosystem preservation are ongoing, which are aimed at using various data collection sensors to address urban issues. Big data, Global Positioning System (GPS), and sensor technologies are now integrated as sustainable development goals are achieved at the city level and the spatial scale is reduced from the country to the city level. Natural environment management technology that monitors, evaluates, and preserves the natural ecological environment is becoming crucial for addressing urban issues and reducing carbon emissions. Because the amount of carbon absorbed and accumulated in the atmosphere varies by tree species, precise classification of tree species is necessary to quantify and manage urban forest benefits.

Traditional forest monitoring employed by national agencies such as the Korea Forest Service entails producing and managing forest type maps by interpreting aerial images and conducting field surveys; however, this strategy is time-consuming, labor-intensive, and can result in a poor understanding of the vertical structure of vegetation and inaccurate classification of species boundaries. Therefore, an accurate tree species classification method that improves existing methodologies is required for quantifying urban forest benefits, conserving biodiversity, and managing climate change-vulnerable species.

3) Remote Sensing Using Multiple Datasets

While previous studies have focused on improving object classification accuracy using satellite images or orthophoto processing, recent forest monitoring studies have focused on data convergence techniques that utilize the advantages of various types of datasets rather than a single remote sensing dataset. The price of remote sensing sensors is decreasing owing to technological advancements, and the use of convergence analysis, which targets a large region, is expanding. When used for forest monitoring, airborne Light Detection and Ranging (LiDAR) can effectively estimate the tree height (TH), diameter at breast height (DBH), tree crown area, and volume because it can create a three-dimensional model of the target site and extract it for the crown area after recognizing individual trees (Simonson et al., 2014). Recently, the efficiency of surveying techniques and sensors has increased to the point that data on the target site can be obtained at a resolution of at least 10 points/㎡, thereby enabling detailed reproduction. Improved accuracy of classification of species and biomass calculation may be obtained using high-resolution airborne LiDAR data. In addition, airborne hyperspectral imaging (AHI) can leverage the advantage of high spectral resolution via more than 100 spectral bands to identify vitality and health status beyond the classification of species. In urban forests in temperate climates, where artificial forests, natural forests, and various species of trees coexist, the classification of tree units requires hyperspectral images with high-resolution spectral wavelengths and various vegetation indices suitable for leaf and reflective characteristics.

4) Classification of Tree Species through Machine Learning

Due to their structural nature, the size and complexity of raw data are large, and the researcher's expertise and research setting ability are required to remove unnecessary data and classify the desired target species. Supervised machine learning for tree classification is widely used in forests and ecology. Here various machine learning classification techniques are employed to create a model that describes the relationship between the dependent and independent variables after extracting the independent variables used for tree classification from remote sensing datasets.

Although logistic regression (LR) is mainly used for binary classification, it can also be used for multinomial classification when it contains at least three categories in the case of multinomial logistic regression (Kwak and Clayton-Matthew, 2002). Even when independent and dependent variables are not linearly correlated, LR can classify them. Furthermore, a support vector machine (SVM) performs classification using a hyperplane that optimizes margins to classify high-dimensional data, and the classification criteria are evenly distributed on both sides of the data to avoid bias (Georganas et al., 2001). Its advantages over neural network methods include its applicability to classification and prediction tasks and its reduced overfitting and impact on error data. However, SVM requires multiple combination tests to control kernel and model parameters in order to create an optimized model; this slows down model construction when many input datasets exist, and it is difficult to interpret and is a complex black box. Decision tree (DT), a machine learning algorithm for supervised learning, can be classified and regressed as a decision rule. DT is a classification algorithm that comprises node segmentation and pruning trees and can evaluate the validity of the optimal tree using data for verification (Kotsiantis, 2013). Owing to its intuitive structure, DT is easy to interpret and can identify valid input variables, but its poor accuracy when overfit and unstable prediction of new data are drawbacks. Random forest (RF) is a machine learning technique that learns a model by randomly selecting some variables and forming multiple DTs (Breiman, 2001). It uses an ensemble method that combines multiple predictions to classify the most votes received as the final prediction (Hastie et al., 2001). RF generates and learns several DT classifiers using a bagging method that randomly selects features from extracted samples using a bootstrap method that permits redundancy. RF accuracy surpasses that of DT, and it can retain high accuracy even as the percentage of missing values increases by reducing predictive variability and preventing overfitting. However, as the number of data increases, RF speed decreases relative to DT, and tree separation becomes complicated, making it difficult to analyze individual trees and interpret the results.

Because the light gradient boosting machine (LGBM) grows trees vertically using the leaf-wise algorithm, it can minimize the loss of predictive errors, unlike the level-wise algorithm method (Ke, 2017). Despite its speed, less memory usage, extremely accurate results, and graphics processing unit (GPU) utilization, LGBM is more prone to over-aggregation on small datasets; hence, datasets with at least 10,000 rows are recommended for its effective use.

5) Biomass Calculation Using Growth Allometry

Because carbon emission rights in countries around the world enable transactions between countries and companies in a lifelike concept, the calculation of green carbon, which is carbon absorbed and stored by the land ecosystem, has become important. Forest biomass is part of green carbon and is an important indicator of forest productivity and carbon circulation (Lim, 2009); the carbon storage and carbon intake of target sites can be calculated by applying biomass expansion and carbon coefficients. Therefore, after classifying the target site species for the preservation value of urban forests, the ground biomass is estimated using the carbon emission coefficient and biomass relative bio-decorations for each species provided by the National Institute of Forest Science. While Cho Hyun-gil (1999) computed the carbon storage and annual carbon absorption of domestic trees using allometric equations per species, Lim Jong-soo (2009) evaluated forest biomass statistics using a regression model and k-nearest neighbor (k-NN) with Landsat TM-5. Integrating the data from LiDAR and hyperspectral sensors can improve terrestrial biomass accuracy from a single dataset (Koch, 2010) while quantifying biomass and visualizing it with tree classification results.

1.1.2. Research Purpose and Significance

1) Research Purpose

Therefore, this research classifies major species of forests in Gwacheon, a temperate climate area, and calculates aboveground biomass by combining AHI and airborne LiDAR during the leaf-on and leaf-off periods, assuming that changes in forest structure and leaf growth cycle in different urban forests can be identified by combining remote sensing data.

The main variables for tree classification are extracted from a forest, which is a natural urban forest excluding street trees and park vegetation, and the accuracy of five classifiers (LR, SVM DT, RF, and LGBM) is compared. Field survey data are used to answer and validate the following research questions: (1) Can the classification between species be explained by employing different datasets from airborne LiDAR and AHI collected during two distinct seasonal periods? (2) Which machine learning classifier learns high-resolution remote sensing data, reflects the situation in the field, and records high accuracy? (3) Is it possible to calculate biomass and review accuracy in indivisible tree units based on tree classification results?

2) Significance of Research

Although comparative studies that integrate high-resolution hyperspectral and LiDAR datasets for domestic tree classification and carbon accumulation calculation are insufficient, some studies on urban forests at the city level, as opposed to small experimental sites, exist. In addition, high-resolution airborne LiDAR with an average point density of 42.7 points/㎡ was used as a variable for classifying forests and trees during the leaf-on and leaf-off periods, and 1 m spatial resolution and 127 spectral spectra were used compared to Sentienl-2 satellite images that provide 10 m spatial resolution and 13 bands. Because most existing studies have focused on pixel-based tree classification, pixel outliers are often reflected in the tree classification results, although it is meaningful as it calculates the polygon-based independent tree area and uses the pixel's median value through zone statistics.

1.1.3. Research Scope

1) Spatial and Temporal Extent

The research site is a natural urban forest, including Umyeonsan Mountain (293 m above sea level [ASL]), Gwanaksan Mountain (632 m ASL), and Cheonggye Mountain (618 m ASL), adjacent to Gwanak-gu and Seocho-gu in Seoul in the north, Seongnam-si in the east, Uiwang-si in the south, and Anyang-si in the west. Since the city area is located in the center of Gwacheon-si, small forest patches exist near the city area, surrounding the forest areas on the left and right sides of the center. The temporal range of the research was November 03, 2021, April 02, 2022, and September 01, 2022, and the airborne LiDAR data were acquired in November 2021 and April 2022, and the AHI data were acquired in November 2021 and September 2022. As for the field survey data, minor class attribute data among the representative biotop types surveyed by the Gyeonggi Research Institute from August to October 2022 were used.

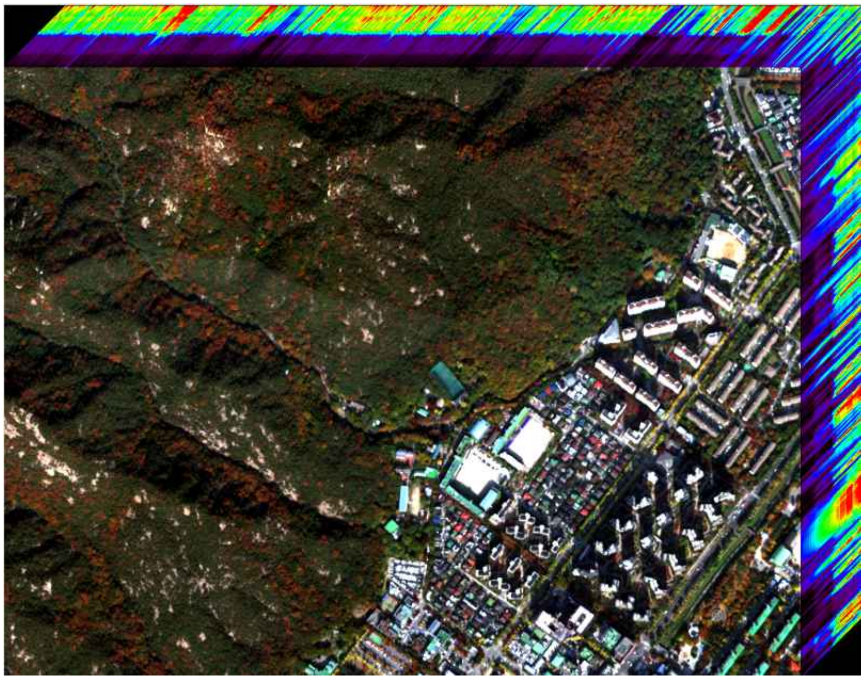


Fig. 1.1 An example of a hyperspectral imaging dataset of the research area

2) Content Scope

Using airborne LiDAR and AHI, which are different types of remote sensing data from the two periods, this research classified 10 representative species of urban forest using 5 machine learning techniques: LR, SVM, DT, RF, and LGBM. Based on the results of random sampling and classification of 29 features and forest field survey data obtained through the interband calculation and structure metrics algorithm from remote sensing data, the above ground biomass (AGB) of urban forests was estimated by substituting parameters for the TH-DBH relational expression and ground biomass calculation.

Chapter 2. Literature Review

2.1. Tree Species Classification and Biomass Estimation

2.1.1. Tree Species Classification and Biomass Estimation

Forest monitoring and biomass calculation by integrating remote sensing data and environmental attribute information have been actively practiced. Recently, many effective forest monitoring studies have utilized the canopy structure derived from airborne LiDAR and vegetation spectral characteristics of hyperspectral images (De Almeida et al., 2021). As survey technology and measurement sensors advance, LiDAR point data with a high density (10 point/m^2) can be acquired, and point data can be used in easily accessible open-source software. Table 2.1 lists the contents and results of previous studies related to vegetation monitoring, such as tree species classification and biomass calculation through data fusion.

Table 2.1 LIterature review of tree species classification and biomass estimation.

Year	Author	Explanation
2008	An-Jin chang, Hyung-Tae Kim.	Extract tree crown area and calculate biomass using aerial images and airborne LiDAR. The regression equation of tree height and diameter at breast height is calculated through field surveys, and biomass is estimated by application of allometric equation.
2009	Cho, H. K., Shin, M. Y.	Using Landsat TM-5 satellite images and field survey sample points, biomass statistics and biomass distribution plots in uninvestigated area are estimated using regression models and k-Nearest Neighbor.

2012	Lee, Hyun Jik, Ru, Ji Ho.	Classified tree species and calculated forest biomass and carbon absorption using airborne LiDAR data and KOMPSAT-2 satellite images. Classify tree species by 90% or more accuracy on average.
2013	Englhart et al	Developed regression modelbusing aerial images and RapidEye multispectral images from 2007 and 2011. Moniotored changes in average canopy height and ground biomass for peatland area in rainforests.
2014	Hyunggab Cho, Kyu-Sung Lee.	Comparison of classification results for coniferous species of hyperspectral imaging and multispectral imaging. The maximum likelihood method was applied to the hyperspectral image converted by dimensionality reduction, and it was calculated with a classification accuracy of 90% or more.
2014	Alonzo et al	Using airborne hyperspectral imaging and airborne LiDAR tree crown area of 29 tree species in the city was extracted, and an overall accuracy of 93.5% was recorded using a total of 28 independent variables including 7 lidar metrics. 4.2% improvement in accuracy for combined datasets compared to single datasets.
2016	Park Jeong-seo	A representative spectroscopic library of hyperspectral images was created for land cover classification of seven classes. Tree classification technique, overall accuracy is improved by 85% or more.
2017	Luxia Liu et al	Classification of 15 species of urban forest trees using the attribute values of airborne LiDAR and hyperspectral images as variables. Analyze fused datasets with 70% or more of total accuracy when classified as random forest classifier.
2019	De Almeida et al	Hyperspectral and lidar variables that affect biomass calculation are reviewed through various regression equations. Compared to a single dataset, the model of the fusion dataset was found to have a higher correlation with the actual biomass calculation.

2021	Janne et al	8 species classification using 3D-CNN techniques by fusing airborne LiDAR and airborne hyperspectral images. Analyzing the entire fusion dataset with 87% classification accuracy when analyzed using 3D-CNN techniques.
2021	Li, Q et al	Species classification using multiple classifiers and comparison of fused datasets in mangrove areas using airborne LiDAR and airborne hyperspectral images. The accuracy of CNN classification was the highest, and it was analyzed that the correlation of the amount of biomass was high.
2021	Kwon, S. K., Kim, K. M., Lim, J	Classified 9 types of forest into random forest classifier using RapidEye sensor. Using spectral information and texture information, tree species classification accuracy was 69.29%.
2021	De Almeida et al	Biomass estimation using 3 lidar matrices and 18 hyperspectral image matrices related to canopy height and leaf area index index as variables.
2022	Chen et al	Convergence of drone LiDAR images of leaf-off and leaf-on periods for temperate forests with high vegetation density. The tree segmentation accuracy of the fusion image was 16.7% more accurate in broad-leaved forests and 4.4% more accurate in coniferous forests.

In international research cases, many studies are being conducted to maximize the advantages of each sensor and utilize seasonal differences by fusing hyperspectral images and airborne LiDAR datasets or by fusing the same dataset with images acquired at different times. Enghart (2013) emphasized the need to monitor forest conservation and sustainable management by calculating the change in canopy height and AGB for petland in rainforest forests using airborne LiDAR data and multispectral images acquired in different years.

Alonzo (2014) classified 29 species of trees in a city and found that the dataset with 7 LiDAR variables outperformed the single hyperspectral dataset by 4.2% in terms of classification accuracy, resulting in 93.5% of tree species classification accuracy. While Janne Mayra (2021) classified 8 species using 3D-CNN techniques by fusing airborne LiDAR and hyperspectral images, Luxia Liu (2017) classified a fusion dataset comprising variables extracted from AHI and airborne LiDAR as RF analysis. Chen (2022) classified the independent tree area 12.3% more accurately than the single temporal leaf-on data by fusing drone LiDAR images from the leaf-on and leaf-off periods in order to increase the reproduction of tree stems and trees. However, the aforementioned studies focused mainly on classifying coniferous species in urban street trees or park green areas, which are relatively easy to distinguish between independent trees, many were conducted on small target sites rather than city level space scales, and few used AHI and airborne LiDAR images obtained at the same time. Therefore, applying these methods to urban forest species in Korea, where deciduous and coniferous trees coexist, will be difficult, and the raster image-based classification results make it difficult to distinguish species boundaries and do not obtain TH for biomass calculation.

Since the 2000s, various studies have been conducted in Korea to classify species using satellite images, hyperspectral images, and airborne LiDAR and to calculate biomass, and methods to increase accuracy have been developed. Im Jong-soo (2009) conducted regression model analysis and k-NN analysis with 16 independent variables from satellite imaging bands, and real measurements of forest type biomass obtained from 58 outdoor sample points in Muju-gun, Jeollabuk-do. As a result, the accuracy of statistical verification using the k-NN technique was higher, and about 8.39 million tons of biomass and 149 t/ha per unit area were calculated at the target site. The research emphasized the need to calculate forest biomass for each species using high-resolution satellite images to compensate for the 30 m spatial resolution of Landsat TM-5 and to expand biomass estimation from small areas to large areas. While Kwon Soo-kyung (2021) used the RapidEye sensor to classify nine forest types using RF techniques, spectral information, and texture information, realizing 69.29% accuracy, Park Jung-seo (2016) selected and applied a representative spectral library for land cover classification of hyperspectral images to improve overall accuracy by more than 85%. While Cho Hyung-gap (2014) employed dimensional reduction to apply hyperspectral images, realizing over 90% classification accuracy via maximum likelihood classification, Lee Hyun-jik (2012) and Jangan-jin (2008) integrated orthophoto, satellite image data, and airborne LiDAR data to obtain biomass estimate results.

Although studies on tree species classification or biomass estimation using satellite images have become common in Korea, studies integrating airborne LiDAR and hyperspectral images remain insufficient, with only a few studies on the independent variable setting for tree classification in temperate forests. In addition, airborne LiDAR has a point density of less than 10 points/㎡, resulting in errors due to low-resolution data, and its application to large-scale target sites is limited when targeting artificial forests with a clear distinction between small data samples and trees. Due to the large size and complexity of raw data when handling the 3D structure of airborne LiDAR images and the AHI dataset comprising a large number of bands, researcher expertise and an analysis method for extracting independent variables that remove unnecessary data and improve tree species classification accuracy are required.

Chapter 3. Methods

3.1. Research Progress

3.1.1. Research Progress

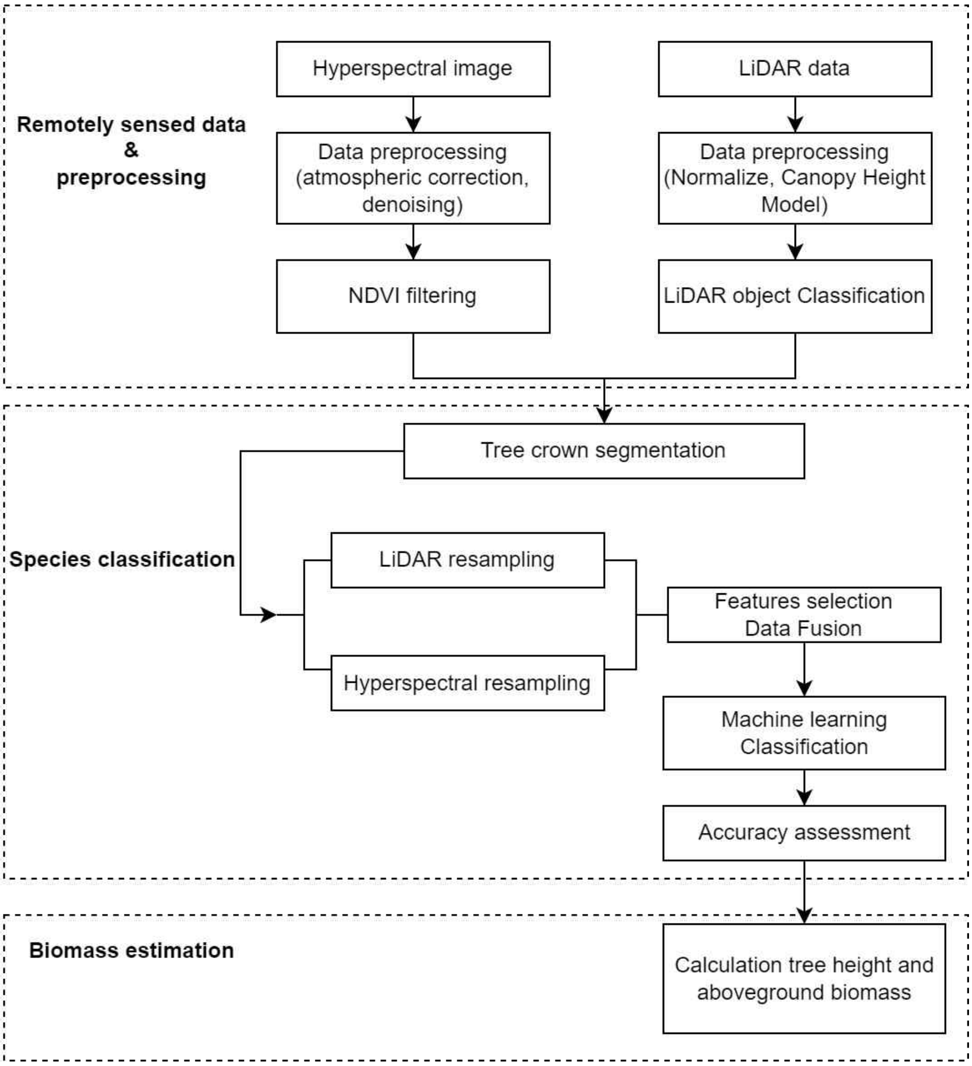


Fig. 3.1 The workflow proposed for tree species classification and AGB calculation

Fig. 3.1 shows this research's order of data processing and analysis, as follows: 1) calibration of remote sensing dataset, 2) individual tree segmentation, 3) calculation of major independent variables for tree classification, 4) training of machine learning classification model and tree classification and accuracy evaluation based on field survey data, and 5) calculation of individual tree heights and aboveground biomass.

3.2. Research Area

3.2.1. Overview of Research Area

Research area is an urban forest located at 37° 23'–37° 27' north latitude and 126° 57'–127° 02' east longitude, with an area of 2,034 ha out of 3,587 ha in Gwacheon Fig. 3.2. Among the forest areas, grassland, cultivated land, and orchards, which are tree species group codes No. 80–99 (Korea Forest Service: forest type map), were excluded from the research's scope, leaving a research area of 2,327 ha containing only the tree species of the research subject. According to the 2020 Forest Basic Statistics (Korea Forest Service, 2020), Gwacheon-si's forest-growing stock was 392,645 ㎥, with a forest growth rate of 64.87% and an average forest-growing stock of 168.73 ㎥/ha. In 2018, As stated in 2020~2024 Climate Change Adaptation Second Action Plan (2019), Gwacheon City had a precipitation of 1,170.5 mm, with an average annual temperature of 12.1° C (ranging from 39.9° C to 17.9° C). Some of the area's forests are adjacent to Seoul, the most populous city in Korea, with a population of 77,818 (Ministry of Security and Public Administration, 2022), so these forests play a role in absorbing carbon dioxide, improving air quality, alleviating heat island, and reducing fine dust.

In addition, as they are adjacent to Seoul City, where it is difficult to obtain remote sensing data due to aviation security, calculating biomass and carbon intake using tree classification seems reasonable. Every five years, the Korea Forest Service produces a forest type map to monitor the forests of the target site. According to the accurate forest type map in 2020, 14 major tree species exist and are occupied in order of area, as shown in Fig. 3.3. Most of the composition was The others quercus forest 29%, Mixed forest 18%, The others deciduous forest 17%, Robinia pseudoacacia 11%, Pinus densiflora 10%. The remaining small percentages of area were Pinus rigida, Quercus acutissima, Quercus variabilis, Pinus koraiensis, Quercus mongolica, Larix Kaempferi, Populus canadensis, Pinus thunbergii, and Abies holophylla. The analysis results of the tree species area ratio reveal that quercus and deciduous forests dominate the area and create a representative landscape.

Notably, the trees in the target area do not form a large colony but are irregularly scattered and distributed throughout the entire area (Fig. 3.2). Owing to its geographical location, Gwacheon-si is surrounded by forests, and the city area is located in the center. Gwacheon-si is largely divided into Gwanaksan Mountain in the west and Cheonggyesan and Umyeonsan mountains in the east. Table 3.1 shows the area and area ratio of the target land species calculated based on the precision forest type map.

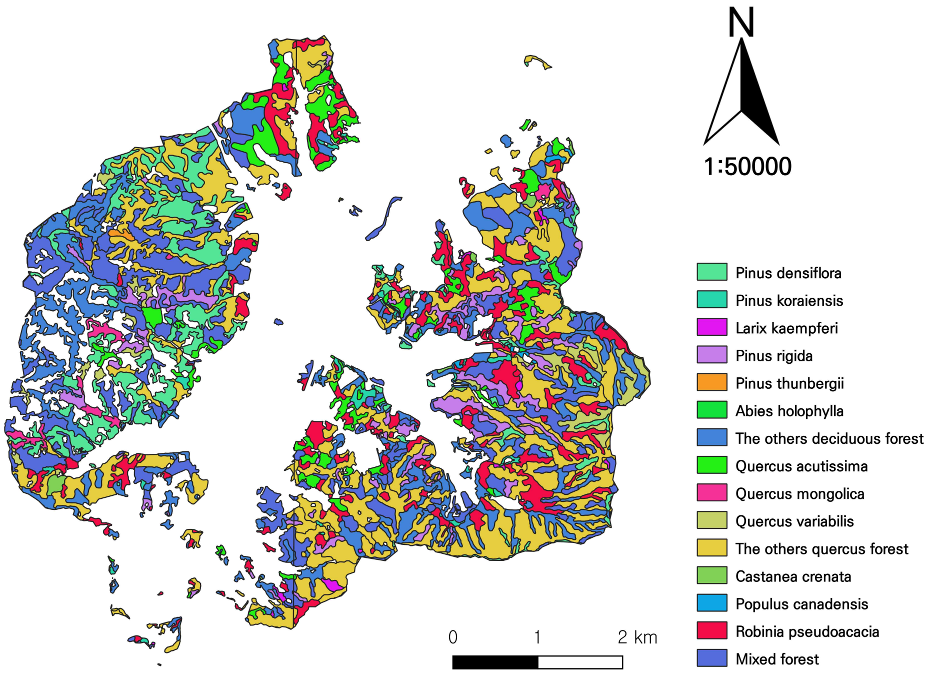


Fig. 3.2 Distribution of forest species in Gwacheon (Korea Forest Service: forest type map)

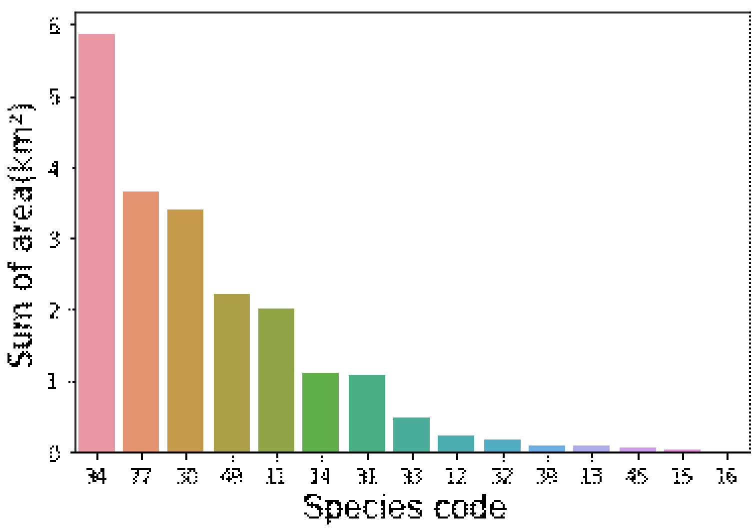


Fig. 3.3 Area ranking of research area species (forest type map).

Table 3.1 Area and area ratio by species.

Code no.	Species	Area (ha)	Area ratio (%)
11	*Pinus densiflora*	200.06	9.83
12	*Pinus koraiensis*	21.83	1.07
13	*Larix Kaempferi*	7.45	0.37
14	*Pinus rigida*	109.96	5.41
15	*Pinus thunbergii*	2.24	0.11
16	*Abies holophylla*	0.36	0.02
30	*The others deciduous forest*	339.25	16.68
31	*Quercus acutissima*	108.21	5.32
32	*Quercus mongolica*	17.81	0.88
33	*Quercus variabilis*	48.45	2.38
34	*The others quercus forest*	587.56	28.88
45	*Populus canadensis*	3.92	0.19
49	*Robinia pseudoacacia*	222.13	10.92
77	*Mixed forest*	365.16	17.95

The analysis of data from the Korea Forest Service's 1:5000 forest type map and 2020 Forest Basic Statistics revealed that the average age of the target area was 4.17, indicating that most forests have a crown occupancy of more than 50% between 31 and 40 years old. In the forest area, Age 4 and Age 5 accounted for 51.4% and 38.7%, whereas Age 4 and Age 5 accounted for 51.5% and 41.8% of the forest-growing stocks.

Table 3.2 Statistic of age class (2020 Forest Basic Statistics).

Age class	Explanation	Forest area (ha)	forest-growing stocks (㎥)
Age 2	Above 50% occupancy of 11~20 year-old trees	10	357
Age 3	Above 50% occupancy of 21~30 year-old trees	149	13,832
Age 4	Above 50% occupancy of 31~40 year-old trees	1,136	202,568
Age 5	Above 50% occupancy of 41~50 year-old trees	856	164,155
Age 6	Above 50% occupancy of 51~60 year-old trees	57	11,733

Notably, 965 polygons were used in the Korea Forest Service's 1:5000 precision forest type map data to analyze the destination's forest canopy height. The frequency of a forest canopy height of 15 m or more and less than 17 m was the highest, and the ratios of 13 to 21 m and 5 to 7 m were relatively high. The forest basin average was 13.86 m, indicating that 8–30 m of trees had grown.

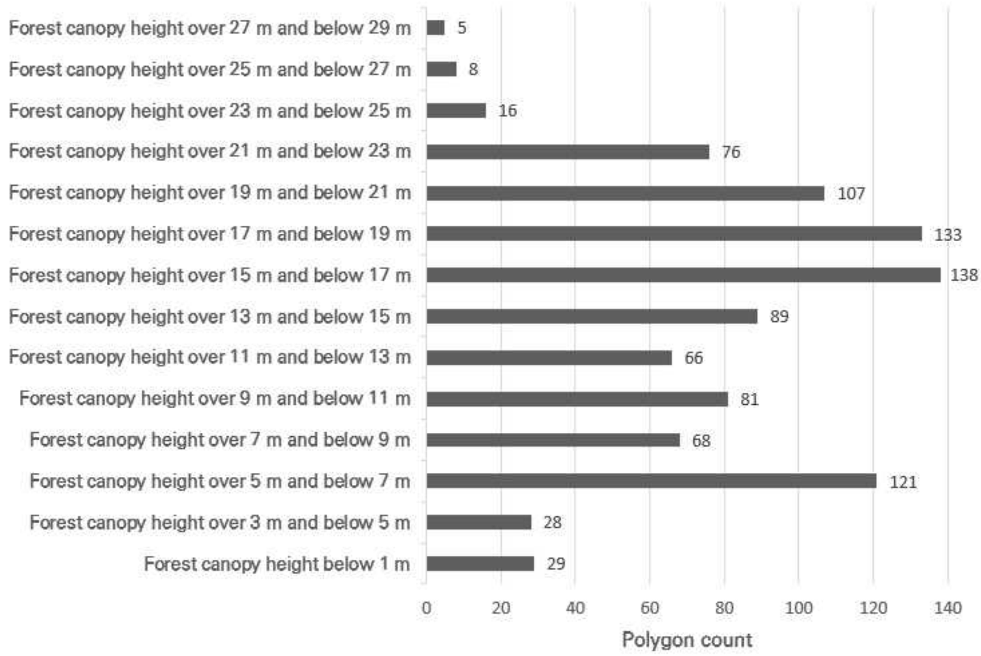


Fig. 3.4 Polygon count by forest canopy height (Korea Forest Service: forest type map).

3.3. Data Acquisition

3.3.1. Airborne LiDAR and Hyperspectral Imaging

For data acquisition, considering the seasonal environmental characteristics of the research area, the airborne LiDAR dataset was considered leaf-on in November before leaf fall and leaf-off in April after leaf fall in an urban forest. While the September AHI dataset was designated leaf-on because it was acquired during the post-monsoon leaf growth period, the November dataset was considered leaf-off because it was acquired during the fall season. Airborne LiDAR data[1)] were acquired on November 03, 2021 and

1) Provided by ASIA Aero Survey co., Ltd.

April 02, 2022 and were scanned with Leica's TerrainMapper 1 at 6000 ft ASL, with an average point density of 42.7 points per square meter (pp㎡), a maximum reflection number of 5, and a scan angle of ± 19.998°. Reflection intensity, GPS time, and number of returns were also acquired as attributes. In the case of AHI[2], the target areas were filmed on November 03, 2021 and September 01, 2022. AHI is a SPECIM AISA Eagle sensor comprising 127 units with a spatial resolution of 1 m and a 404–996 nm wavelength range. While the hyperspectral images in November comprised 12 individual strips, those in September comprised 13 individual strip images. Table 3.3 lists the details of the collected remote sensing data.

Table 3.3 Remote sensing data information.

Sensor	Attribute	Value
Airborne Hyperspectral Imaging	Date of Acquisition	2021.11.03., 2022.09.01
	Sensor Model	AISA Eagle
	Spatial Resolution	1 m
	Spectral Range	404–996 nm
	Full Width at Half Maximum (FWHM)	0.44–0.48 nm
	Radiometric Resolution	12 bit

2) Provided by ASIA Aero Survey co., Ltd.

Airborne LiDAR	Date of Acquisition	2021.11.03., 2022.04.02
	Flying Height	6000 ft
	Sensor model	TerrainMapper 1
	Point Density	42.7 points per ㎡(pp㎡)
	Maximum Number of Return	5
	Scan Angle	± 19.998°

3.3.2. Ground Truth Data

Ground truth data, which provides information on specifications and x and y coordinates, was obtained from August to October 2022 as a field survey. Data from several types of minor class attribute subclasses of forest survey result[3] data were used, and the minimum area of sampled field survey data exceeds 100 ㎡ (10 m × 10 m). Forest surveys in polygon area units were conducted for these data, and the polygon ID and field data at the survey point were cross-referenced. The minor class attribute information on these

3) Provided by Gyeonggi Research Institute.

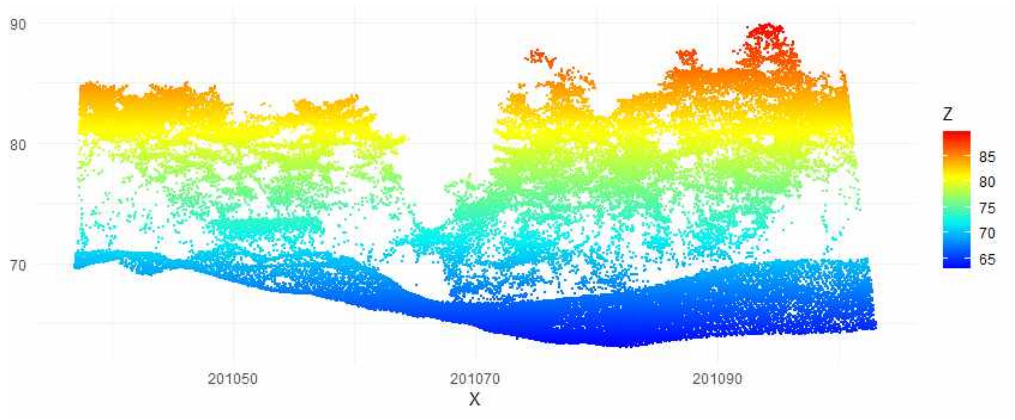
data comprised a polygon containing a single species and a polygon containing multiple mixed forests; however, only a single species from the polygon was used to improve the training accuracy of the machine learning classifier. The forest survey data served as the ground truth for tree classification training and verification.

3.4. Data Processing

3.4.1. Pre-Processing of Airborne LiDAR and Hyperspectral Imaging

Concerning airborne LiDAR images, surface and non-surface points were classified using automatic ground classification and manual classification of the acquired raw LAS data to produce ground classification data (DTD). The standard height was corrected by applying KNGeoid18, a korean national geoid model developed by the Korea National Geographic Information Institute, and system errors were erased through the IMU Calibration. The acquired airborne LiDAR images were mooted into a single LAS file using GreenValley International LiDAR 360 software, and only forest areas were extracted at the urban forest boundary using QGIS' LAStools (https://rapidlasso.com/lastools/). Through R software's cloth simulation filtering, the point where the height value (z) was located at the lowest point was selected and the ground and non-ground areas of the research site points were automatically classified.

Raw data(Leaf−off)



CSF ground classification(Leaf−off)

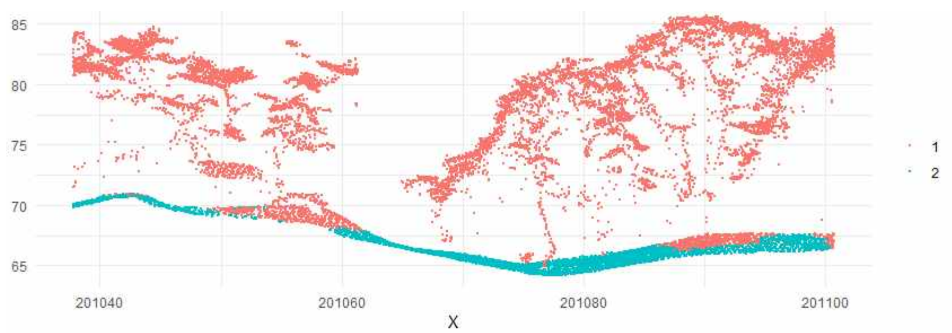


Fig. 3.5 Ground classification using the CSF algorithm.
(1:non-ground; 2:ground; X: x coordinate, Z: Altitude)

Because noise points away from the surface or object may cause errors in LiDAR analysis and classification, noise caused by flying currents or frequency interference was removed through noise filtering in LiDAR 360, and the remaining abnormal points were manually removed through visual inspection. Finally, a normalized three-dimensional LiDAR point cloud was created by subtracting the altitude of the indicator from that of each point. The number of point data for airborne LiDAR acquired in November 2021 was 1,193,108,897, with an average elevation of 6.30 m (standard deviation 5.48) and an average reflection intensity of 730.2 (standard deviation 758.0). The number of point data for airborne LiDAR acquired in April 2022 was 875,226,444, with an average elevation of 5.62 m (standard deviation 5.08) and an average reflectance of 538.89 (standard deviation 792.25). Table 3.4 lists detailed attribute information of the point cloud of the airborne LiDAR image.

Table 3.4 Point cloud attributes of Leaf-on and Leaf-off season.

	Point number	Mean height (m)	Mean intensity
Leaf-on	1,193,108,897	6.30 ± 5.48	730.2 ± 758.0
Leaf-off	875,226,444	5.62 ± 5.08	538.89 ± 792.25

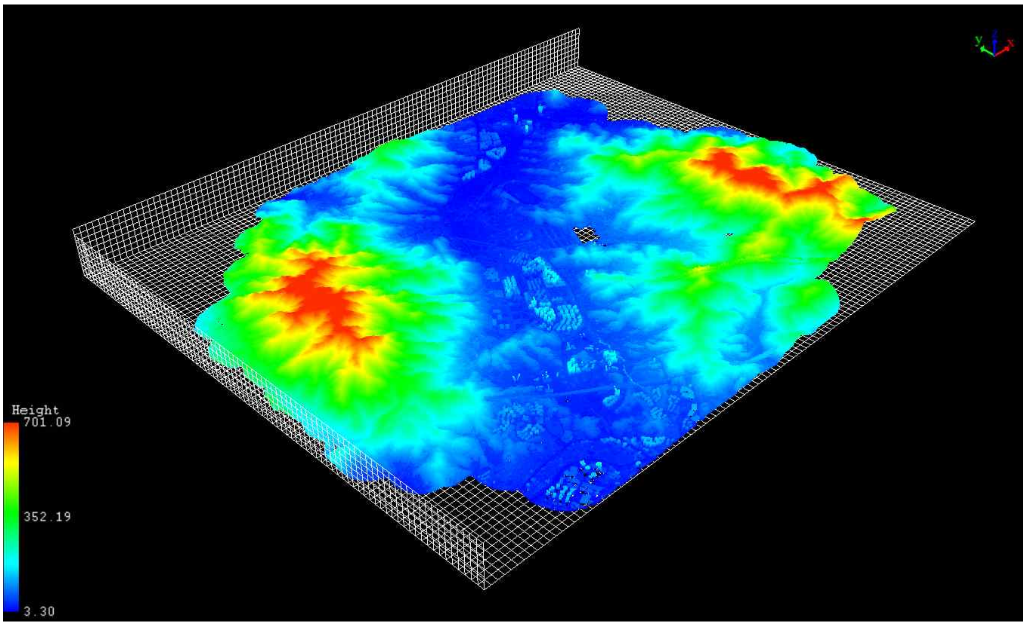


Fig. 3.6 Airborne LiDAR imaging of leaf-on season

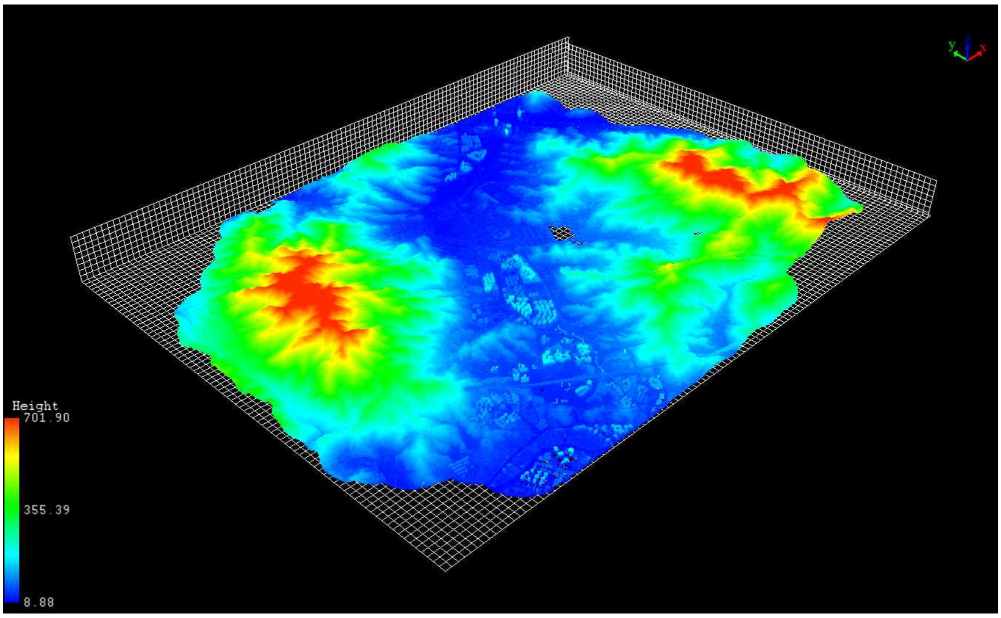


Fig. 3.7 Airborne LiDAR imaging of leaf-off season

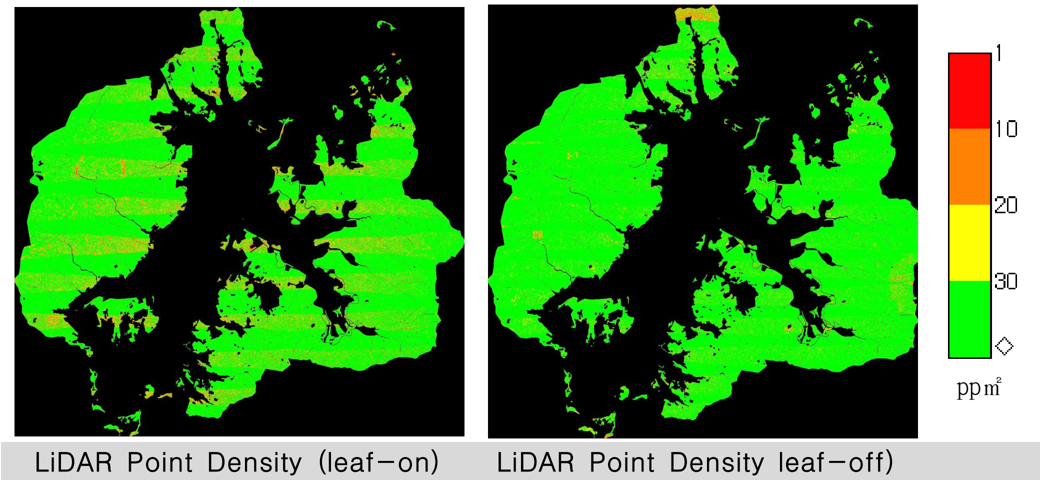


Fig. 3.8 LiDAR point density of leaf-on and leaf-off condition

AHI requires atmospheric, radiation, and geometric correction processes to eliminate errors caused by differences in brightness values from aerosols in the ground atmosphere. While radiation and geometrical corrections were performed using Spectir's SHIPS module, atmospheric correction was performed using Res' ATCOR-4 module. The image coordinates were geo-referenced based on the National Geographic Information Institute's aerial orthophoto photograph (UTM 52N) acquired on May 15, 2020. To compensate for differences in reflectance values caused by differences in shooting time along aircraft flight paths, additional Quick Atmospheric Correction (QUAC) were performed using ENVI 5.6.2 software from L3Harris Geospatial. This software is effective in urban ecosystems where artificial structures, water, vegetation, and soil coexist. The 12 strips acquired in November and 13 strips acquired in September were orthogonally calibrated with each single mosaic image using the Seamless Mosaic algorithm in ENVI 5.6.3.

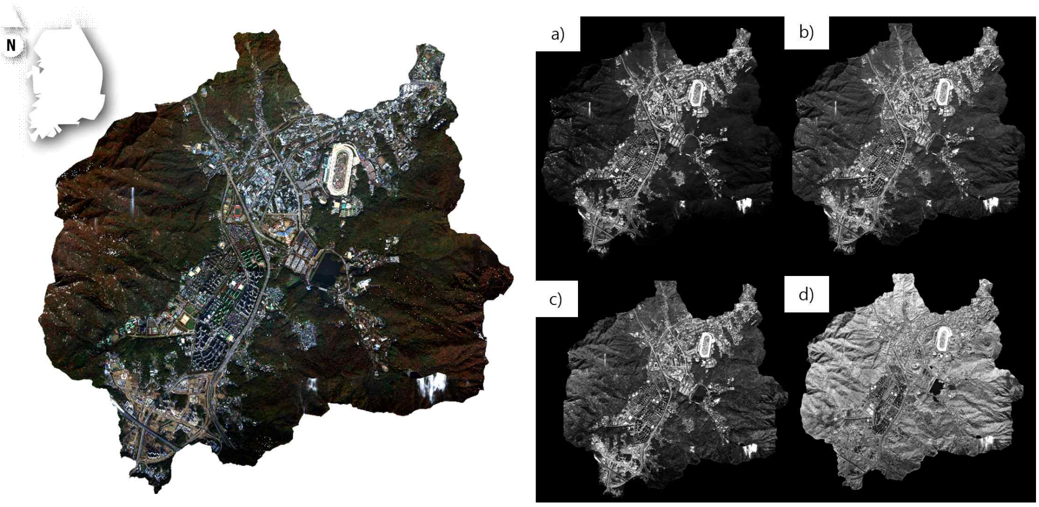


Fig. 3.9 Mosaic image of airborne hyperspectral Imaging (November)
a)Blue, b)Green, c)Red, d)Near infrared

3.4.2. Dataset Composition

In this research, while the classification accuracy of the combined dataset of AHI and airborne LiDAR was compared with that of a single dataset, the classification accuracy of a single temporal dataset was compared with that of a multitemporal dataset in order to determine whether the classification accuracy of the fusion dataset was improved. The front part of the string was set with Hyperspectral Imaging (H) and LiDAR (L), and the back part of the string was set with alphabetical capitalization at the time of data acquisition.

Regarding HS-LN and HN-LA, the acquisition timing differed, but at the time of acquisition, these datasets were considered a single temporal multidataset, considering the environmental characteristics of urban forests. As a result, five datasets were generated: multitemporal single datasets (HN-HS, LN-LA), single temporal multidataset (HS-LN, HN-LA), and multitemporal multidataset (HN-HS-LN-LA). Table 3.5 lists the details of the divided datasets.

Table 3.5 Dataset composition by sensors and date condition.

ID	Dataset	Number of bands
HN-HS	Hyperspectral November, Hyperspectral September	16
LN-LA	LiDAR November, LiDAR April	13
HS-LN	Hyperspectral September, LiDAR November	14
HN-LA	Hyperspectral November, LiDAR April	14
HN-HS-LN-LA	Hyperspectral November, Hyperspectral September, LiDAR November, LiDAR April	29

3.4.3. Selection of Major Tree Species

To eliminate errors caused by the nonexclusive characteristics of dependent variables during machine learning classification, similar tree species should be unified. In the minor class field containing tree species classification of field forest survey data, the top 10 species with a high area ratio were selected for classification based on a single tree area, excluding mixed forests and rock vegetation. Quercus mongolica, Pinus densiflora, Robinia pseudoacacia, Quercus acutissima, Pinus rigida, Quercus variabilis, Castanea crenata, Pinus koraiensis, Quercus serrata, and Larix Kaempferi were representative species, and the classification codes were assigned in order of area. Species investigated with an area ratio of less than 0.2 were Fraxinus lanuginosa, Populus tomentiglandulosa, Quercus aliena, Prunus sargentii, Abies holophylla, and Prunus subg. Cerasus, Liriodendron tulipifera, Zelkova serrata, Cercidiphyllum japonicum, Metasequoia glyptostroboides, and Ziziphus jujuba Mill were excluded from the classification model learning because the area occupied on a wide-area spatial scale was small.

Table 3.6 Area of target tree species.

Species	Class number	Survey area (km^2)	Area ratio (%)
Quercus mongolica	1	3.49	33.3
Pinus densiflora	2	1.96	18.7
Robinia pseudoacacia	3	1.61	15.4
Quercus acutissima	4	1.15	10.9
Pinus rigida	5	0.77	7.3
Quercus variabilis	6	0.53	5.1
Castanea crenata	7	0.33	3.1
Pinus koraiensis	8	0.15	1.4
Quercus serrata	9	0.06	0.6
Larix kaempferi	10	0.02	0.2

3.4.4. Tree Crown Segmentation

The merging of hyperspectral and airborne LiDAR imagery was utilized to extract only the vegetation area of the research area. Using the threshold value of the NDVI index (range: 0.2–0.8) extracted from hyperspectral images, areas below 0.2 were identified as non-vegetation areas, and raster pixels with vegetation reflectance were extracted as masking layers to distinguish between vegetation and non-vegetation areas. Regarding airborne LiDAR data, the University of Dayton DALES dataset was subjected to deep learning training using PointNet++, a hierarchical neural network of MATLAB, and labeled into eight classes: ground, vegetation, cars, trucks, powerlines, fences, poles, and buildings. Point clouds such as power transmission towers, buildings, and fences mixed in urban forests were removed, and AHI was used as a masking layer to extract only the areas filtered as vegetation areas (TH of 2 m or more; tree radius of 2 m or more).

Airborne LiDAR data collected during the leaf-on and leaf-off periods were merged to generate data that expressed both the upper and lower forest layers. The digital elevation model (DEM) and digital surface model (DSM) with 0.5 m spatial resolution were generated using triangulated irregular network (TIN) interpolation by calculating the altitude values (Z) of the ground and non-ground of the data. The canopy height model (CHM), which is a normalized height value from the ground, was generated through the difference between DEM and DSM. Local maximum filtering analysis was performed on CHM to calculate the tree top by selecting the maximum height value of the point cloud, and through watershed analysis, the tree's crown area pixels were selected based on the tree apex and distance of the local minima, and the boundaries between the trees were distinguished. Each tree was assigned a unique tree ID, and the attribute value of each feature was determined by calculating the median value of the raster pixels included in the tree crown area.

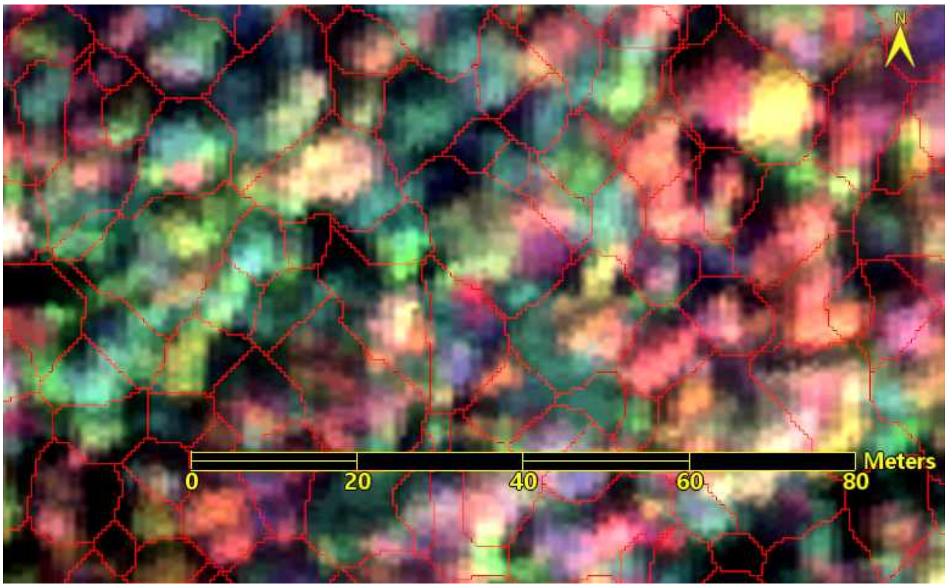


Fig. 3.10 Tree crown area extraction image using local maxima filtering

3.5. Feature Extraction

3.5.1. LiDAR Feature Extraction

The spatial resolution of the AHI and airborne LiDAR image metrics was set to 1 m resolution because of the noise impact and geometrical placement error caused by spatial resolution mismatch (Frair et al., 2010). Airborne LiDAR metrics convert a 3D LiDAR dataset into a 2D raster and incorporate it into a classification model as a variable (Davies & Asner, 2014; Simonson et al., 2014). R Software's stdmetrics and GreenValley International LiDAR 360's Forest Metrics were used to generate independent variables using the distribution and intensity characteristics of the point group data. Notably, the LiDAR metrics variable includes elevation metrics relating to height and density and intensity metrics variables capable of using different surface reflectance and properties for classification by utilizing return intensity—the amount of energy returned to the LiDAR sensor by laser pulses reflected on the target object. The standard deviation of the intensity and height of the LiDAR point were used, and the 9th quartile of 75th percentile reflectance and point cloud density were used. In addition, the canopy cover and leaf area index, which are mainly related to the vegetation area, were used as independent variables. Using the CloudCompare software's distance computation algorithm, an absolute distance variable representing the point change between leaf-on and leaf-off was derived by calculating the difference between the vertical and horizontal point changes.

3.5.2. Hyperspectral Imaging Feature Extraction

Several vegetation indices are based on the calculation of the near-infrared and red bands because vegetation reflectance differences in these bands occur mainly in the spectral reflectance of general leaves. In this research, the hyperspectral vegetation index was selected for vegetation classification by referencing previous studies. Using the photochemical reflectance index (PRI), which evaluates photosynthetic efficiency, carotenoid pigments in living leaves can be detected and plant photosynthetic function can be ascertained (Gamon, 1997). The red green ratio index (RGRI), like PRI, measures photosynthetic efficiency and is useful for estimating leaf stress and developmental processes in response to leaf red light caused by anthocyanins in chlorophyll (Gamon, 1999). The structure-insensitive pigment index is sensitive to detecting the ratio of carotenoid pigments to chlorophyll and can also detect increases in carotenoid-rich canopy stress (Penuelas, 1995). Furthermore, the carotenoid reflectance index 1 (CRI1), representing carotenoids in plant leaves, can detect stress-related carotenoid concentration. The plant senescence reflectance index is associated with plant senescence and detects carotenoid increases in chlorophyll (Merzlyak et al., 1999). The modified red edge simple ratio index is used for leaf reflection and stress detection using bands in the red edge area (Sims, 2002). The anthocyanin reflectance index 1 (ARI1) utilizes anthocyanin, which is abundant in leaves during the open and deciduous periods, and hyperspectral imaging can identify vegetation stress because it is expressed as a change in the anthocyanin pigment concentration (Gitelson, 2001). The modified chlorophyll absorption ratio index is used to indicate the relative degree of chlorophyll content (Daughtry et al., 2000). The Vogelmann red edge index 1 (VREI1) is sensitive to chlorophyll concentration, leaf area, and water content and utilizes the red region wavelength (Vogelmann, 1993).

3.5.3. Principal Component Analysis

In most cases, the entire wavelength band of the AHI data is not used due to limitations associated with large data dimensions. Misclassification is possible and significant variable discrimination is made more challenging when all bands are used in the species classification process. Therefore, limiting the dimension is essential to prevent data loss when removing redundant data. To reduce the high-dimensional dataset to highly correlated variables, a new variable creation process was performed through principal component analysis (PCA). PCA extracts orthogonal principal components by designating the axis with the largest variance as the first principal component and the axis with the second-largest variance as the second principal component. PCA results can be visualized and combined with other remote sensing data. To select independent variables for species classification, PCA was performed on the two periods' hyperspectral images using the PCA tool of ENVI 5.6.3. The PCA results of the 127 multidimensional bands of the September AHI dataset indicated that PC1 (99.5%), PC2 (0.3%), and PC3 (0.1%) had explanatory power for the variables of the entire dataset, and PC1 was used. For the November AHI dataset, the PCA results indicated PC1 (99.6%), PC2 (0.21%), and PC3 (0.08%). As a result, each of the 127 hyperspectral bands of the two periods was dimensionally reduced to the PC1 band through PCA and used as an independent variable for species classification.

3.5.4. Feature Selection

Table 3.7 lists the composition of the independent variables used in vegetation classification. A total of 29 variables, including AHI PC1 (2 variables), hyperspectral vegetation index (14 variables), and airborne LiDAR variables (13 variables), were used as independent variables for classification. In order to prevent the multicollinearity problem caused by a high correlation between independent variables, variables with a correlation coefficient of at least 0.8 between independent variables were considered to have a strong correlation and were excluded from being overlapped (Kim, 2019).

Table 3.7 Independent variable used for tree species classification.

Datasets	Independent variable	Source
Airborne LiDAR	Canopy Cover(CC) Nov, Apr	Jennings et al., 1999
	Leaf Area Index(LAI) Nov, Apr	Nilson, 1971
	75th percentile of return intensity(IP75) Nov, Apr	Liu, et al., 2017
	Standard deviation of return intensity(IS) Nov, Apr	Liu, et al., 2017
	Standard deviation of elevation(ES) Nov, Apr	Liu, et al., 2017
	Ninth decile of elevation density(ED9) Nov, Apr	Korpela et al., 2010
	Absolute distance(AD) Difference between Nov and Apr	Esposito et al., 2017

Airborne Hyper-spectral Imaging	First Principal Component (PC1)Sep, Nov	-
	Anthocyanin Reflectance Index 1(ARI1) Nov	Gitelson et al., 2001
	Carotenoid Reflectance Index1 Sep, Nov	Gitelson et al., 2002
	Modified Chlorophyll Absorption Ratio Index(MCARI) Nov	Daughtry et al., 2000
	Modified Red Edge Simple Ratio(MRESR) Sep	Sims et al., 2002
	Photochemical Reflectance Index(PRI) Sep, Nov	Penuelas et al., 2005
	Plant Senescence Reflectance Index(PSRI) Sep	Merzlyak et al., 1999
	Red Green Ratio Index(RGRI) Sep, Nov	Gamon et al., 1999
	Structure Insensitive Pigment Index(SIPI) Sep, Nov	Penuelas et al., 1995
	Vogelmann Red Edge Index 1(VREI1) Nov, Sep	Vogelmann et al., 1993

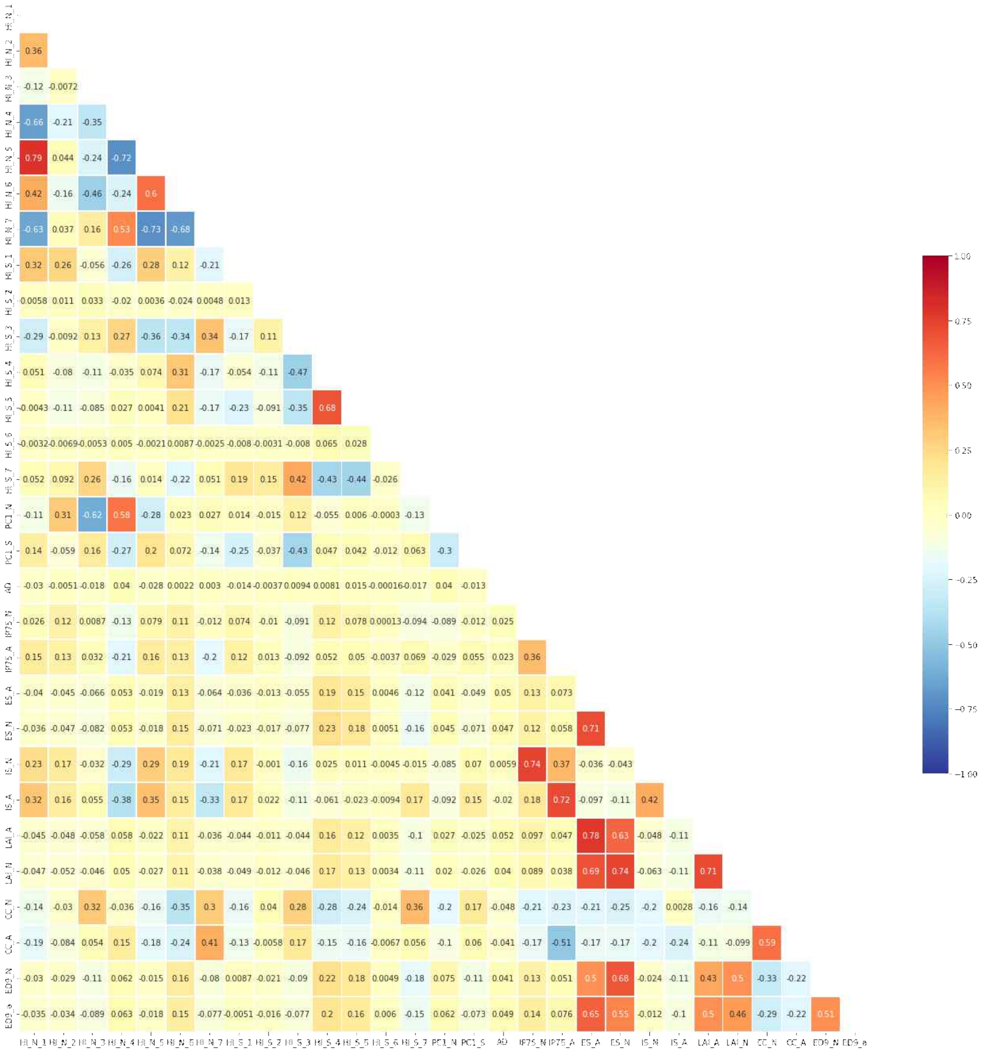


Fig. 3.11 The results of correlation analysis between independent variables

3.6. Classification Technique

3.6.1. Classification Technique

A 2.5 m buffer was set internally from the tree species boundary to prevent data collection errors due to the size of the raster pixel during random point sampling at this boundary. In addition, to prevent sampling in areas such as the forest's lower part and the gap in the forest, areas less than 2 m in height were excluded using the height layer extracted from the LiDAR data to conduct sampling in the canopy's upper part. To prevent overfitting and outlier data generation caused by the difference in the range and size of each variable in continuous data, a robust scaler was used and is shown in Equation (1). The robust scaler sets the median value (x_{med}) to 0 and makes the interquartile range (IQR, x_{75}-x_{25}), which is the difference between the 3rd quartile and the 1st quartile, to be 1 for data (x_i). This minimizes the influence of extreme values in continuous independent variables.

$$x' = \frac{x_i - x_{med}}{x_{75} - x_{25}} \tag{1}$$

Noting that 16,522 random points were sampled for each type to remove missing data, 165,216 points were generated for 10 types. For the training and evaluation of machine learning classifiers, 165,216 points were divided into a ratio of 80 to 20, and stratified sampling was performed to maintain the population data ratio and prevent training bias toward a specific tree species. For machine learning analysis, 5 classifiers, LR, SVM, DT, RF, and LGBM, were used, and each classifier was cross-validated 5 times in 10 layers through an optimized hyperparameter tuning process using Python's

GridSearchCV (Table 3.8).

Table 3.8 List of hyperparameter values of LR, SVM, DT, RF, and LGBM classifier.

Classifier	Hyperparameter	Tuned value	Defalut value
LR	solver	newton-cg	lbfgs
	C	0.01	1.0
	max_iter	1000	100
SVM	C	1.0	1.0
	gamma	0.1	scale
DT	max_depth	None	None
	min_samples_split	2	2
	criterion	gini	gini
RF	n_estimators	500	100
	max_depth	50	None
	oob_score	True	False
	random_state	42	None
LGBM	eval_metric	multi_logloss	None
	n_estimators	400	100

In K-fold cross-validation, after dividing the total dataset into k partitions, k−1 partitions are used for training, the remaining 1 partition is used as test data, and the value of the verified average is calculated as accuracy by k iterations (A. Ramzan et al., 2021).

The confusion matrix and Cohen's kappa index were used to evaluate the analysis results. The confusion matrix entails evaluating a trained classification model by comparing the actual class value to the predicted class value and determining the classified class accuracy by representing it as a matrix.

- True Positives (TP): Actual positive values are classified as positive
- False Positives (FP): Actual negative values are classified as positive
- False negative (FN): Actual positive values are classified as negative
- True Negative (TN): Actual negative values are classified as negative

		Actual Class	
		Positive	Negative
Predicted Class	Positive	TP (True Positive)	FP (False Positive)
	Negative	FN (False Negatvie)	TN (True Negatvie)

Fig. 3.12 Verification of classification accuracy using confusion matrix

Precision refers to the percentage of values calculated using Equation (2) and classified by the classification model as positive that are actually positive. Recall is the ratio of the value predicted by the model as positive among the actual positive values calculated using Equation (3). These two items were evaluated for each model type generated by the F1 score, which is Equation (4) for harmonic means, and the accuracy of each dataset and classifier was compared using Equation (5). Regarding the Cohen's kappa index, the tree classification accuracy can be evaluated by comparing the predicted and actual values and measuring the degree of agreement.

$$(Precision) = \frac{TP}{TP+FP} \tag{2}$$

$$(\mathrm{Recall}) = \frac{TP}{TP+FN} \tag{3}$$

$$(\text{F1-score}) = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

$$(\text{Accuracy}) = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

3.7. Estimating Aboveground Biomass

3.7.1. Estimating Aboveground Biomass

Because forest trees are determined by the forest-growing stock of the clinical classification unit of the forest type map, urban foresters use allometric equations to calculate biomass (Park, 2009). The biomass calculation method using allometric equations estimates the biomass and carbon storage amount by substituting the DBH obtained from sample surveys for each species into allometric equations.

In this research, the unit of analysis was the independent tree, and biomass was calculated using TH-DBH relational expression verified in previous studies utilizing tree species classification results and airborne LiDAR TH information. AGB was calculated by summing the biomass of stems, branches, and leaves.

The modified logistic (6) model explained the relationship between TH and DBH in forests with strong explanatory power (Ratkowsky and Reedy 1986, Huang et al., 2000) and was applied to TH after being converted to Equation (7).

$$H = 1.2 + \frac{a}{1 + b^{-D^{-c}}} \quad (6)$$

$$D = \left(b\left(\frac{a}{H - 1.2} - 1\right)\right)^{-1/c} \quad (7)$$

$$y = aD^{b}H^{c} \quad (8)$$

H = Tree Height; D= Diameter at breast height; a~c= Model parameters

DBHs of Quercus mongolica, Pinus densiflora, Quercus acutissima, Pinus rigida, Quercus variabilis, Quercus serrata, and Larix kaempferi were derived using the parameter values of a–c of 7 major species of Seo Yeon-ok (2011). Robinia pseudoacacia, Castanea crenata, and Pinus koraiensis referred to the application criteria for tree volume, biomass, and stand yield table 2021 (KFS and NIFoS, 2021), and the parameters of Quercus acutissima, Quercus mongolica, and Pinus densiflora were set to be used (Robinia pseudoacacia → Quercus acutissima, Castanea crenata → Quercus mongolica, Pinus koraiensis → Pinus densiflora). Table 3.9 displays the values of a, b, and c used to calculate DBH based on the modified logistic growth model.

Table 3.9 Parameters used in TH-DBH relational expression by species.

Species	a	b	c
Quercus mongolica	15.3958	0.03	1.3391
Pinus densiflora	14.9639	0.0174	1.5132
Robinia pseudoacacia	14.263	0.0068	2.0675
Quercus acutissima	14.263	0.0068	2.0675
Pinus rigida	13.7491	0.0302	1.5015
Quercus variabilis	19.9118	0.029	1.2394
Castanea crenata	15.3958	0.03	1.3391
Pinus koraiensis	14.9639	0.0174	1.5132
Quercus serrata	15.9803	0.0295	1.3246
Larix kaempferi	20.5193	0.0016	2.4893

AGB was calculated by adding the biomass of stems, branches, and leaves by substituting the TH value of the individual tree derived by CHM segmentation and the DBH value estimated according to the tree volume, biomass, and stand yield table's allometric equations (8).

While DBHs smaller than the range of use of the equation were excluded in the calculation process, those exceeding the range of the equation were deemed maximum DBH. Trees not included in the standard yield table were replaced based on the current criteria for table application. The parameter values of Quercus acutissima for Robinia pseudoacacia and Quercus mongolica for Castanea crenata and Quercus serrata were used (Robinia pseudoacacia → Quercus acutissima, Castanea crenata → Quercus mongolica, Quercus serrata → Quercus mongolica). Concerning Pinus densiflora, allometric equations for each part of Pinus densiflora in the central region were used. Table 3.10 shows the values of a, b, and c substituted for allometric requirements by region for each tree type to calculate AGB.

Table 3.10 Parameters used to estimate aboveground biomass
(Tree volume, biomass, and stand yield table 2021. Seoul, Republic of Korea: KFS and NIFoS).

Species	Stems (a)	Stems (b)	Stems (c)	branches (a)	branches (b)	branches (c)	leaves (a)	leaves (b)	leaves (c)
Quercus mongolica	0.098	1.406	1.135	0.018	3.083	-0.493	0.023	2.609	-0.833
Pinus densiflora	0.034	1.734	1.025	0.008	3.586	-1.158	0.077	1.931	-0.566
*Robinia pseudoacacia**	0.008	2.334	1.069	0.012	2.853	0.006	0.008	2.518	-0.151
Quercus acutissima	0.008	2.334	1.069	0.012	2.853	0.006	0.008	2.518	-0.151
Pinus rigida	0.029	1.824	1.036	0.00002	2.632	2.058	0.053	1.82	-0.22
Quercus variabilis	0.053	1.81	0.881	0.082	2.553	-0.608	0.108	1.63	0.406
*Castanea crenata**	0.098	1.406	1.135	0.018	3.083	-0.493	0.023	2.609	-0.833
Pinus koraiensis	0.034	1.734	1.025	0.008	3.586	-1.158	0.077	1.931	-0.566
*Quercus serrata**	0.098	1.406	1.135	0.018	3.083	-0.493	0.023	2.609	-0.833
Larix kaempferi	0.005	2.458	0.904	0.143	4.482	-2.9	0.022	1.877	-0.023

*Parameters are replaced and applied according to the current inventory application standard

Chapter 4. Results

4.1. Tree Species Classification

4.1.1. Tree Species Classification

The tree classification results of the trained classifiers indicated that the average accuracy of multitemporal multidatasets was 71%, exceeding those of single temporal multidatasets (leaf-on:57%; leaf-off:61%) and multitemporal single datasets (AHI:64%; airborne LiDAR 55%). The combination of AHI and airborne LiDAR images during the leaf-off period showed about 4% higher accuracy than the combination of AHI and airborne LiDAR images during the leaf-on period. The comparison of the multitemporal single datasets revealed that the AHI dataset was approximately 9% more accurate than the airborne LiDAR dataset.

The comparison results of the five machine learning classifiers indicated that the RF algorithm had the highest average accuracy (76%), followed by LGBM (70%), DT (61%), SVM (60%), and LR (39%). The accuracy of classification using multitemporal multidatasets as RF algorithms was 83.5% (Kappa: 0.82), indicating the highest accuracy in all classifiers and combinations of datasets. Following RF, accuracy of multitemporal multidatasets was shown in the order of LGBM (81%), SVM (76%), DT (68%), and LR (47%). Fig. 4.1 depicts the classification result map of the research area based on the trained RF model.

Table 4.1 Comparison of classification accuracy by dataset and classifier.

Classifier	Dataset					Mean
	HN-HS	LN-LA	HS-LN	HN-LA	HN-HS-LN-LA	
Lr	0.43	0.30	0.34	0.40	0.47	0.39
DT	0.61	0.62	0.55	0.61	0.68	0.61
SVM	0.66	0.45	0.56	0.59	0.76	0.60
LGBM	0.72	0.64	0.66	0.68	0.81	0.70
RF	0.78	0.72	0.72	0.75	0.83	0.76
Mean	0.64	0.55	0.57	0.61	0.71	

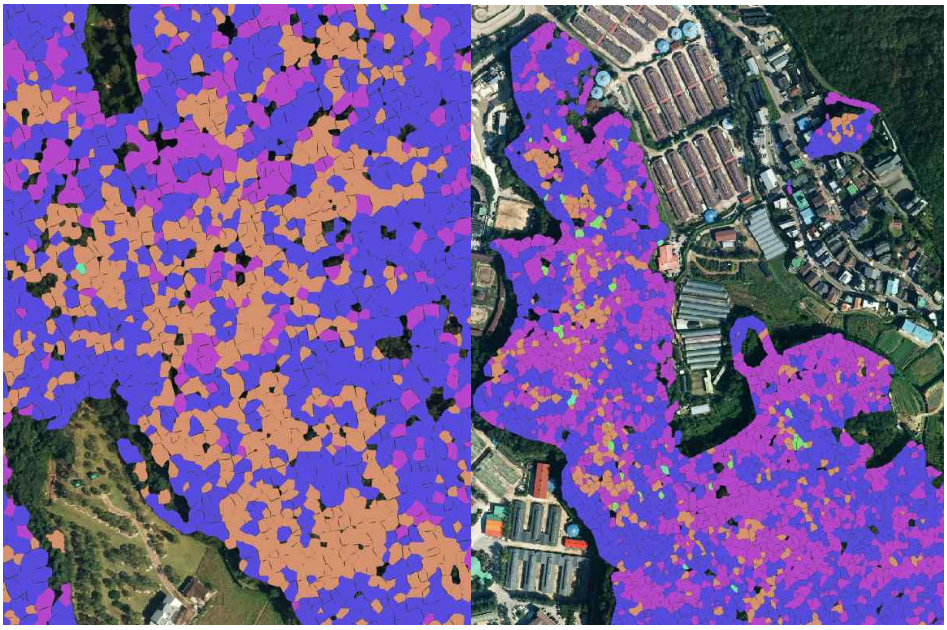


Fig. 4.1 The results of individual tree segmentation and species classification

The F1-score evaluation of classification accuracy by species revealed that Larix kaempferi (95%), Pinus koraiensis (88%), Quercus serrata (88%), Pinus densiflora (85%), Robinia (84%), Castanea crenata (82%), and Quercarius (82%) had high scores in that order. As shown in Table 4.2, when multitemporal multidatasets were classified as RF classifiers, both precision (min: 78%; max. 93%) and recall (min: 74%; max. 98%) showed over 70% classification accuracy for the 10 major species.

Table 4.2 Random forest classification confusion matrix (multitemporal multidataset).

True label / Predicted label	a	b	c	d	e	f	g	h	i	j	Recall
a: Quercus mongolica	2529	54	90	83	72	212	50	8	157	48	77%
b: Pinus densiflora	53	2827	34	20	130	46	25	73	83	13	86%
c: Robinia pseudoacacia	60	29	2886	53	70	26	65	21	55	39	87%
d: Quercus acutissima	136	81	175	2459	109	68	68	84	50	74	74%
e: Pinus rigida	132	146	46	38	2521	95	51	201	24	51	76%
f: Quercus variabilis	162	55	104	87	129	2495	28	7	184	54	75%
g: Castanea crenata	100	26	139	89	67	137	2578	55	62	52	78%
h: Pinus koraiensis	14	26	41	25	130	11	68	2971	1	17	90%
I: Quercus serrata	27	102	22	5	5	26	7	1	3103	6	94%
j: Larix kaempferi	17	6	18	11	7	4	5	4	5	3228	98%
Precision	78%	84%	81%	86%	78%	80%	88%	87%	93%	90%	83.5%
Overall Accuracy: 83.5%, Kappa coefficient: 0.82											

Table 4.3 Random forest classification F1-score (multitemporal multidataset).

Species	F1-score	support
Quercus mongolica	0.78	3304
Pinus densiflora	0.85	3304
Robinia pseudoacacia	0.84	3304
Quercus acutissima	0.80	3304
Pinus rigida	0.77	3305
Quercus variabilis	0.78	3305
Castanea crenata	0.82	3305
Pinus koraiensis	0.88	3304
Quercus serrata	0.88	3304
Larix kaempferi	0.94	3305
Mean	0.83	33044

4.2. Important Independent Variables

4.2.1. Important Independent Variables

Analyzing the main variables affecting tree classification results using the Feature Importance function of the Python library showed that the carotenoid reflection index derived from AHI was the highest variable in the fall topology reason (Importance: 0.064). CRI during the full-leaf period in September was also identified as a highly significant variable among the hyperspectral image variables (Importance: 0.048). VREI, which utilizes the near-infrared wavelength bands of the September hyperspectral images, was also a highly significant variable (Importance: 0.045). In the hyperspectral image obtained in November, ARI (Importance: 0.044), photosynthesis-related index (Importance: 0.042), and RGRI (Importance: 0.042) had importance, as indicated. For airborne LiDAR images, the leaf area index (Importance: 0.062) and standard deviation of electricity (Importance: 0.049) obtained in April were also high. Notably, the above findings confirmed that, among the data extracted through interband calculations from the AHI dataset, the index associated with the pigment and photosynthesis performance of leaves and the properties associated with the vertical structure of airborne LiDAR generated by metrics calculations appeared to be major factors in determining overall accuracy. In addition, the 75th percentile intensity (leaf-on, leaf-off, Importance: 0.023), standard deviation of reflected intensity (leaf-on, leaf-off: Importance: 0.022, 0.221), canopy cover (leaf-off, Importance: 0.022), and the difference in distance between the points of the two periods (Importance: 0.017) were independent variables of low importance in species classification.

Table 4.4 Feature Importance (N: November; S: September; A: April).

Feature	Explanation	Feature Importance
CRI1_N	Carotenoid Reflectance Index 1_N	0.064
LAI_A	Leaf Area Index_A	0.062
ES_A	Standard deviation of elevation_A	0.049
CRI1_S	Carotenoid Reflectance Index 1_S	0.048
VREI1_S	Vogelmann Red Edge Index 1_S	0.045
ARI1_N	Anthocyanin Reflectance Index 1_N	0.044
ES_N	Standard deviation of elevation_N	0.043
PRI_N	Photochemical Reflectance Index_N	0.042
RGRI_N	Red Green Ratio Index_N	0.042
VREI1_N	Vogelmann Red Edge Index 1_N	0.041
SIPI_N	Structure Insensitive Pigment Index_N	0.039

LAI_N	Leaf Area Index__N	0.038
MRESR_S	Modified Red Edge Simple Ratio_S	0.036
PC1_S	First Principal Component_S	0.034
PC1_N	First Principal Component_N	0.032
MCARI_N	Modified Chlorophyll Absorption Ratio Index_N	0.031
CC_A	Canopy Cover_A	0.028
ED9_A	Ninth decile of elevation density_A	0.027
PSRI_S	Plant Senescence Reflectance Index_S	0.027
RGRI_S	Red Green Ratio Index_N_S	0.026
ED9_N	Ninth decile of elevation density_N	0.025
SIPI_S	Structure Insensitive Pigment Index_S	0.025
PRI_S	Photochemical Reflectance Index_S	0.024
IP75_A	75th percentile of return intensity_A	0.023
IP75_N	75th percentile of return intensity_N	0.023
IS_N	Standard deviation of return intensity_N	0.022
CC_N	Canopy cover_N	0.022
IS_A	Standard deviation of return intensity_A	0.021
AD	Absolute distance	0.017

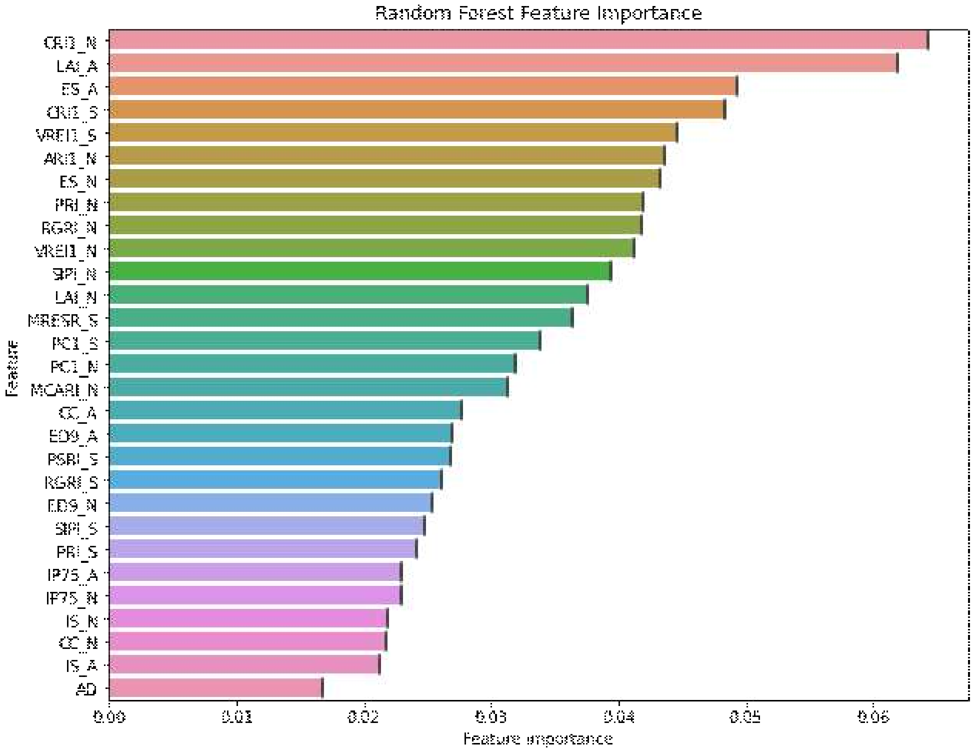


Fig. 4.2 Ranking of important independent variables

4.3. Estimating Aboveground Biomass

4.3.1. Estimating Individual Tree Height and DBH

The estimated individual tree through local maxima segmentation of the CHM model created by fusing LiDAR images of leaf-on and leaf-off was 928,015 trees. Individual TH values calculated by dividing the research area's central point into left and right halves revealed a significant height variation. Comparing 250 K geological maps (Korea Institute of Geoscience and Mineral Resources), most of the Gwanak Mountain area west of the research site comprised Granite, and trees with an average height of 5.14 m were distributed due to shallow soil depth and many cliffs and bedrock. In addition, trees with an average TH of 7.33 m were distributed in the eastern Cheonggyesan Mountain area, which comprises gneiss and is relatively deep in the weathered earth core. The difference in TH of the research area vegetation due to differences in topography and bedrock was revealed, and the average height of all research area trees was approximately 6.4 m.

Table 4.5 The statistic of individual tree height

	Mean	Std. dev.	Freq.
Left	5.14	2.24	393,211
Right	7.33	2.37	534,804
Total	6.40	2.56	928,015

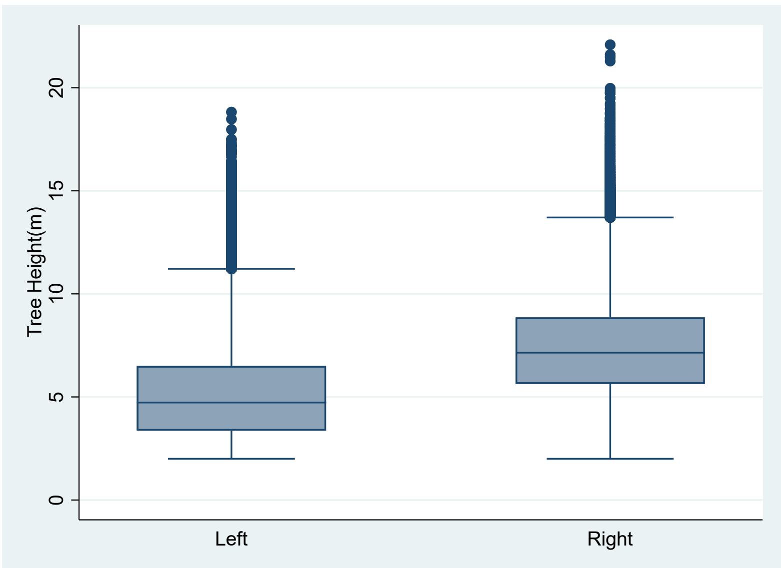


Fig. 4.3 Comparison of tree height in the research area

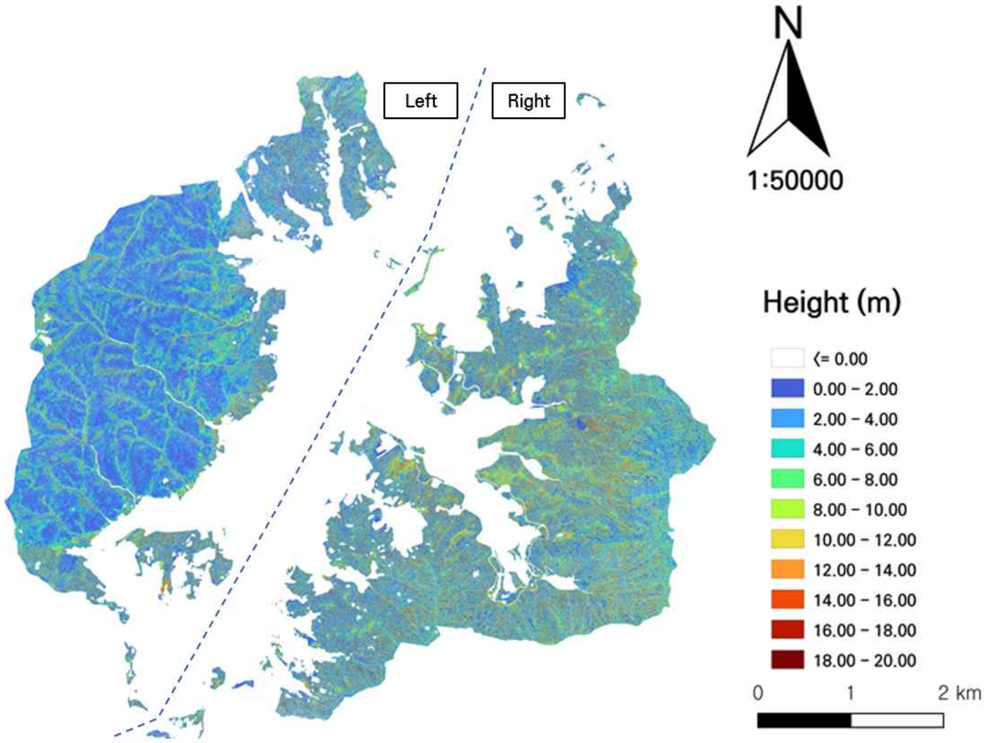


Fig. 4.4 Distribution of individual tree heights in the research area

4.3.2. Estimating Aboveground Biomass

Tree species values were assigned to individual tree crown area polygons based on the results of classifying multitemporal multidatasets using the RF algorithm. The DBH of the independent tree was derived using Logistic TH-DBH relational expression. Different a–c parameters were applied according to TH, DBH, and tree species to calculate the biomass of stems, branches, and leaves and AGB by adding them to the allometric equations of tree volume, biomass, and stand yield table. Stem biomass had an average of 29.6 kg, a standard deviation of 44.9 kg, a maximum of 587.9 kg, and a minimum of 2.7 kg, with a total of 27,504,963.4 kg (27,505 t). Branch biomass had an average of 16.7 kg, a standard deviation of 43.4 kg, a maximum of 431.6 kg, a minimum of 2.0 kg, and a total of 15,454,894.9 kg (15,455 t). Leaf biomass was calculated with an average of 2.6 kg, a standard deviation of 4.2 kg, a maximum of 39.5 kg, and a minimum of 0.6 kg, with a total of 2,390,777.8 kg (2391 t). The average AGB of stem, branch, and leaf biomass was 48.9 kg, with a standard deviation of 91.5 kg, a maximum of 954.5 kg, and a minimum of 7.3 kg. A total of 45,350,636 kg (45,351 t) of AGB was calculated for the research area. Fig. 4.5 shows the outcomes of mapping the research area biomass calculation result (Table 4.6) by combining the biomass result and attribute value of the indivisible tree crown area polygon. According to the AGB distribution at the research area, DBH calculated based on TH of trees was used as input values, so AGB was calculated to be higher in forests near Seoul Grand Park and Cheonggye Mountain in the east than forests in Gwanak Mountain in the west.

Table 4.6 Biomass calculation result for tree part

Biomass (kg)	Mean	Stdev	Max	Min	Sum
Stems	29.6	44.9	587.9	2.7	27,504,963.4
branch	16.7	43.4	431.6	2.0	15,454,894.9
leaves	2.6	4.2	39.5	0.6	2,390,777.8
Above ground	48.9	91.5	954.5	7.3	45,350,636.0

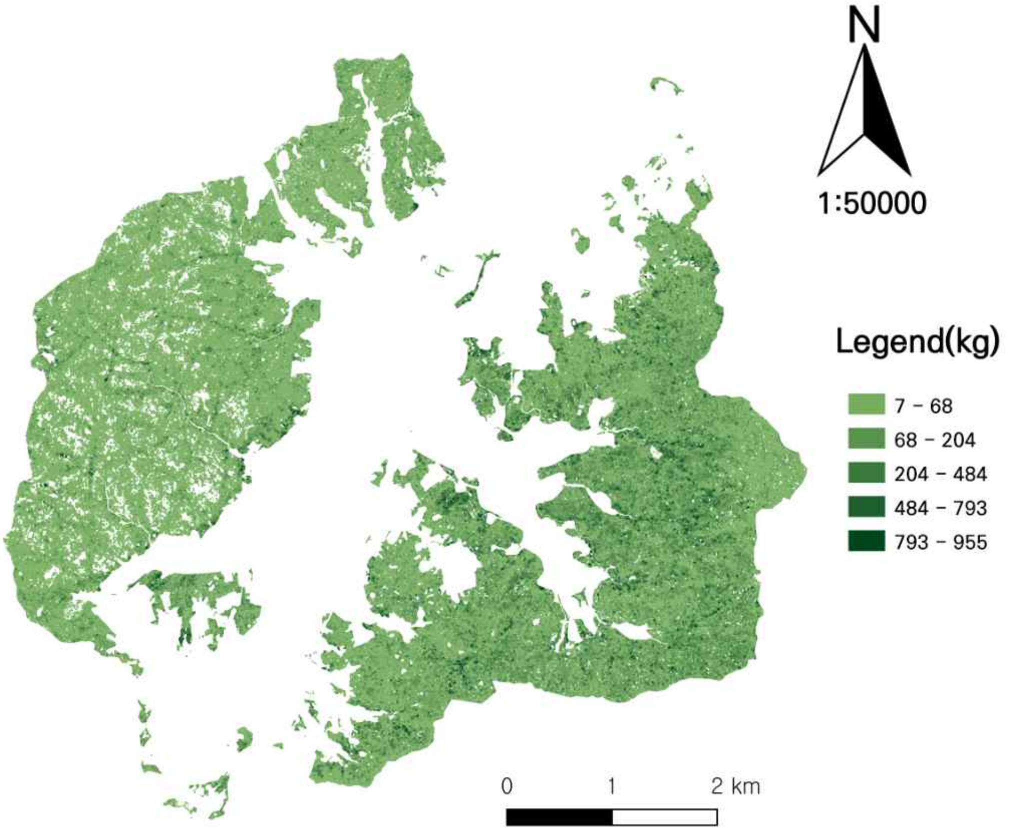


Fig. 4.5 The map of aboveground biomass estimation in the city of Gwacheon

Chapter 5. Discussion

5.1. Discussion

5.1.1. Tree Species Classification

This research suggests that classifying urban forest species by fusing AHI and airborne LiDAR data acquired over two different periods could be more accurate than traditional aerial image-based research methods. The average accuracy of the five machine learning classifiers in multitemporal multidatasets was 71%, exceeding those of single temporal multidatasets (leaf-on:57%; leaf-off:61%) and multitemporal single datasets (AHI:64%; airborne LiDAR 55%). Notably, employing the seasonal characteristics of forests can improve classification accuracy because the spectral reflectance of leaves and degree of leaf development in forests vary in seasons. In addition, combining data can improve classification results because two distinct sensors can collect a variety of attribute information for the target species. The canopy leaf surface reflectance is strongly prominent in AHI, whereas airborne LiDAR can confirm the tree's vertical structure and height-based volume. Datasets that fuse different periods and sensors may have high classification accuracy because they use the overall features of the tree as variables for tree classification. Therefore, compared to existing aerial or satellite images that classify coniferous and deciduous trees, the combination of different sensors in the studied two periods resulted in more than 80% classification accuracy for tree species.

The RF classification algorithm had the highest average accuracy (76%) because it uses ensemble methods that synthesize the classification results of multiple DTs to prevent overfitting, leading to high classification accuracy. In addition, RF makes non-relational predictions between trees through randomness and bagging during the selection of variables, hence reducing their sensitivity to the noise within the dataset. When segmenting nodes in the tree, 29 features derived from AHI and airborne LiDAR are randomly selected and optimal features are found, thus reducing bias while creating a suitable classification model. For this reason, high-dimensional datasets with 29 variables are thought to be classified with distinctly higher classification accuracy compared to other classifiers. The LGBM algorithm, which showed the second-highest classification accuracy, had an average classification accuracy of 70% because it finds the optimal classifier via error reduction of the previous model residuals by weighting the error-containing data while learning several weak learners sequentially. However, DT, SVM, and LR exhibited relatively low accuracy in classifying urban forest species using many independent variables owing to their limitations in reducing errors and calculating optimal models when learning and classifying high-dimensional datasets using various features of AHI and airborne LiDAR.

5.1.2. Important Independent Variables

The classification of the multitemporal multidataset using the RF classifier indicated that variables with important effects on tree classification did not appear to have a bias toward one sensor at a time, suggesting that some important independent variables may contribute to improving the accuracy of tree classification by period and sensor. Based on the analysis that the CRI of the hyperspectral image acquired in the fall foliage season was analyzed as the most important independent variable, the temporal difference and spectral reflectance difference of fall foliage increased by tree species in the corresponding period and can be detected through AHI. Fig. 5.1, which was captured from the ground in November when the AHI data were acquired, shows that the fall foliage season differs for each tree species and that it is relatively easier to distinguish the boundaries of deciduous trees during the leaf-off period. In the fall foliage season, the ARI also recorded a high importance of 0.044, and the classification result of the PRI showed a high contribution of 0.042, suggesting that species classification accuracy can be improved by utilizing the timing of chlorophyll destruction and anthocyanin synthesis. The CRI in the leaf-on period recorded a high importance of 0.045, suggesting that carotenoid concentration differs between species in both the fall foliage season and leaf growth period.

Fig. 5.1 Ground field image at the time of acquisition of Airborne Hyperspectral Imaging (2021.11.07.) (Left: Munwon Children's Park, Right: Airdrie Park)

Notably, VREI using the wavelength band of the red edge area could improve classification accuracy in all periods of November and September, suggesting that this can be used as an independent variable for classification by utilizing the difference in spectral reflectance for each species in the red edge area, which is a section where the spectral reflectance of vegetation increases rapidly in the infrared area.

The leaf area index obtained from an airborne LiDAR image showed a difference between April's leaf-off period (0.062) and November's leaf-on period (0.038). During the leaf-on period, which is the full-leaf period, the vegetation density increased, resulting in many overlaps at the boundary and making it difficult to distinguish between tree species boundaries. In addition, the high importance of the standard deviation variable in the height metrics of LiDAR points (leaf-off: 0.049; leaf-on: 0.043) may have affected the classification because the height of the point data for each species differed. Meanwhile, intensity is expected to act as an important variable in the classification process because it contains information on the difference in light scattering according to leaf surface conditions. However, in this research, the difference in intensity of laser pulses by season and aircraft course was reflected in the data acquisition process, indicating different attribute information for each data sampling point and low importance in the classification process.

Chapter 6. Conclusion

6.1. Overall Summary

6.1.1. Overall Summary

This research targeted the temperate urban forest and generated five datasets by dividing the AHI and airborne LiDAR images obtained during the two periods by period and sensor. Through interband and metrics calculations, an independent variable containing 29 species for tree classification was derived, and machine learning classification was performed for 10 representative species using independent and dependent variable values for 165,216 points by sampling 16,522 random points for each species. As a result, the average accuracy of the five classifiers of the multitemporal multidataset was the highest (71%), and the RF classifier had the highest average accuracy of the datasets for the five classifiers (76%). When classifying multitemporal multidatasets with RF, accuracy was highest (accuracy: 83.3%; Kappa: 0.82). Using the model trained with the RF classifier, the multitemporal multidataset with the highest classification accuracy was classified, and the TH-DBH relational equation and biomass calculation formula were applied to calculate 45,351 t of AGB in 928,015 tree crown areas at the research

6.2. Implications and Limitations

6.2.1. Implications of Research

This research is significant in that it quantified the classification of species and biomass for calculating carbon storage to cope with climate change policies and classified species based on machine learning analysis using the advantages of high-resolution hyperspectral and LiDAR images as an advanced method of traditional investigation. Although the spatial resolution of previous spectral satellite images was limited to 10–30 m and the number of available bands was limited to 10, this research used high-resolution image data with 1 m spatial resolution and 127 hyperspectral images to reflect the three-dimensional forest characteristics of the research area by combining airborne LiDAR and two-dimensional plane data. This research is expected to help preserve urban forests and manage climate change-vulnerable species, as it classifies urban forest species with over 80% accuracy. In addition, sampling was conducted based on forest field survey data to enable a more detailed tree boundary classification than the existing 1:5000 scale forest type map, and classification was performed in urban forests as opposed to test bed research area. The results of the area tree species classification are unique because the crown region was derived and analyzed as a vector polygon region, as opposed to the raster pixel format used in previous studies. In the individual tree crown area, the independent variable value was input as a median value by zone statistics to reduce noise and misclassification caused by fine pixels, thereby improving classification accuracy, and the highest point height of the TH was input as an attribute value to derive DBH. Carbon absorption in urban forests can be quantified in response to climate change policies using AGB, making it possible to establish and apply urban ecological status for sustainable urban development and urban management. The importance of the independent variables contributing to the classification can be calculated by extracting vegetation indices and

PCA derived from high-resolution hyperspectral images along with independent variables that reflect the situation at the area calculated by height and intensity metrics.

In addition, establishing DBH using hyperspectral and LiDAR datasets in the future and obtaining the corresponding time series data from various periods would contribute to quantifying the urban forest carbon cycle by measuring the growth and growth rate of major urban forest species and calculating their biomass.

6.2.2. Limitations of Research

Certain species exhibited lower categorization accuracy, which is a limitation of this research. Owing to vegetation production at various layers in actual forests, object classification accuracy at the boundary was deemed to be quite low due to overlap with adjacent trees. Although the vegetation area with at least 2 m TH was selected for tree species classification, this has a limitation because the case of vegetation that does not grow high due to bedrock and soil characteristics is not reflected. Furthermore, because data classification accuracy verification was conducted as a confusion matrix, comparing the tree heights derived from airborne LiDAR, the location of the tree crown area, and the classification results to the actual site is crucial. CHM was extracted from the airborne LiDAR dataset, and local maximum filtering was performed based on the height point to set the crown area because various vegetations are mixed in actual temperate forests. For horizontally growing trees, accurately calculating the tree crown area was difficult; hence, the research may result in an underestimation or overestimation of the actual one. DBH was derived from TH-DBH relational expression, and the results were obtained using the DBH-biomass calculation formula used in previous studies; hence, it may differ from the actual field when calculating DBH through actual tree height. Although a verified model was used, a re-verification should be undertaken because the verification comparing the actual biomass amount to the calculated biomass amount at this research area was omitted. To accurately calculate forest biomass, additional research is required to establish a regression equation between two variables that can estimate DBH for each species in the forest using tree heights derived from airborne LiDAR for various species. In addition, when defining the remote sensing image acquired in November, the timing was not unified in that the November image was regarded as leaf-on in AHI and leaf-off in airborne LiDAR. However, the November dataset was advantageous since it could be defined based on specific situations because it was acquired when both leaves and fall foliage existed. Regarding domestic forests, obtaining AHI and airborne LiDAR images in August–September was difficult due to the influence of the continuous rainy season and typhoons, suggesting that the November datasets can be used more for classification. In the future, improved accuracy for tree species classification can be realized by developing parameters that classify domestic forests with a large number of deciduous and coniferous trees based on their structural forms and by applying classification algorithms. Notably, if a vegetation index is developed using airborne LiDAR–hyperspectral images, forest monitoring research employing remote sensing data will be actively conducted in urban forest research.

References

▌ Domestic Literature

(1) Report

Gwacheon city, (2019), 2020~2024 Climate Change Adaptation Second Action Plan

Korea Forest Service, (2021), 2020 Forest Basic Statistics

Ministry of Land, Infrastructure and Transport (South Korea). (2021). Statistics of Urban Planning

Ministry of Security and Public Administration. (2022.10) Resident Registry Population. https://jumin.mois.go.kr/

Park Eun-jin, & Jwa Seung-hee. (2009). Quantification of CO2 Uptake by Urban Trees and Greenspace Management for C Sequestration. Gyeonggi Research Institute, 1-1.

Son, Y. M., Kim, R. H., Lee, K. H., Pyo, J. K., Kim, S. W., Hwang, J. S., ... & Park, H. (2014). Carbon emission factors and biomass allometric equations by species in Korea. Korea Forest Research Institute.

(2) Research Paper

An-Jin chang, & Hyung-Tae Kim. (2008). Study of Biomass Estimation in Forest by Aerial Photograph and LiDAR Data. Journal of the Korean Association of Geographic Information Studies, 11(3), 166-173

Cho, H. K., & Shin, M. Y. (2009). Estimation of Forest Biomass based upon Satellite Data and National Forest Inventory Data. Korean Journal of Remore Sensing, 25(4), 311-320.

Eun Jin Park, & Kyu Yi Kang. (2010). Estimation of C Storage and Annual CO2 Uptake by Street Trees in Gyeonggi-do. Korean Journal of Environment and Ecology, 24(5), 591-600.

Hyunggab Cho, & Kyu-Sung Lee. (2014). Comparison between Hyperspectral and Multispectral Images for the Classification of Coniferous Species. Korean Journal of Remote Sensing, 30(1), 25-36.

Jo, Hyun-Kil, & Ahn, Tae-Won. (2012). Carbon Storage and Uptake by Deciduous Tree Species for Urban Landscape. Journal of the Korean Institute of Landscape Architecture, 40(5), 160-168.

Jo, Hyun-Kil. (1999). Carbon Uptake and Emissions in Urban Landscape, and the Role of Urban Greenspace for several Cities in Kangwon Province. Journal of Korean institute of landscape architecture, 27(1), 39-53.

Joon Hyung Park, Su Young Jung, Kwang Soo Lee, Chang Hwan Kim, Yong Bae Park, & Byung Oh Yoo. (2016). Classification of Regional Types for Pinus densiflora stands Using Height-DBH Growth in Korea. Journal of Korean Society of Forest Science, 105(3), 336-341.

Lee, D. G., Yun, E. J., Kim, E. Y., & Park, E. J. (2006). Gwacheon City Street Tree Status Analysis and Maintenance Plan. In Proceedings of the Korean Institute of Landscape Architecture Conference (pp. 66-69). The Korean Institute of Landscape Architecture.

Lee, Hyun Jik, & Ru, Ji Ho. (2012). Application of LiDAR Data &

High-Resolution Satellite Image for Calculate Forest Biomass. Journal of Korean Society for Geospatial Information Science, 20(1), 53-63.

Lim, Ye Seul, Eo, Yang Dam, Jeon, Min Cheol, Lee, Mi Hee, & Pyeon, Mu Wook. (2016). Experiments of Individual Tree and Crown Width Extraction by Band Combination Using Monthly Drone Images. Journal of Korean Society for Geospatial Information Science, 24(4), 67-74.

Park, Jung-Seo., Seo, Jin-Jae & Go, Je-Woong. (2016). The Study on Improving Accuracy of Land Cover Classification using Spectral Library of Hyperspectral Image. Journal of Cadastre & Land InformatiX, 46(2), 239-251.

Seo Yeon Ok, Young Jin Lee, Dae-Kyun Nor, Kim Sung Ho, Choi Jung Kee, & LEE， Woo-Kyun. (2011). Height-DBH Growth Models of Major Tree Species in Chungcheong Province. Journal of Korean Society of Forest Science, 100(1), 62-69.

Soo-Kyung Kwon, ·Kyoung-Min Kim, & Joongbin Lim. (2021). A Study on Pre-evaluation of Tree Species Classification Possibility of CAS500-4 Using RapidEye Satellite Imageries. Korean Journal of Remote Sensing, 37(2), 291-304.

(3) Thesis

Choi, Heejoon. (2018). Estimation of the Urban Forest Growth Using Multi-temporal LiDAR Dataset (Doctoral dissertation, Department of Landscape Architecture Graduate School of Environmental Studies Seoul National University)

(4) Other Materials

KFS (Korea Forest Service), NIFoS (National Institute of Forest Science). 2021. Tree volume, biomass, and stand yield table 2021. Seoul, Republic of Korea: KFS and NIFoS (in Korean)

Korea Forest Service: forest type map(1:5000)

Korea Institute of Geoscience and Mineral Resources, Geo Big Data Open Platform, 250K geological maps

National Assembly of the Republic of Korea, 2020, Creation And Management Of Urban Forest Act(Act No. 17420, jun. 9, 2020)

▌ International Literature

(1) Research Paper

A. Ramezan, C., A. Warner, T., & E. Maxwell, A. (2019). Evaluation of sampling and cross-validation tuning strategies for regional-scale machine learning classification. Remote Sensing, 11(2), 185.

Alonzo, M., Bookhagen, B., & Roberts, D. A. (2014). Urban tree species mapping using hyperspectral and lidar data fusion. Remote Sensing of Environment, 148, 70-83.

Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.

Cao, Q. V. (2004). Predicting parameters of a Weibull function for modeling diameter distribution. Forest science, 50(5), 682-685.

Daughtry, C. S., Walthall, C. L., Kim, M. S., De Colstoun, E. B., & McMurtrey

Iii, J. E. (2000). Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. Remote sensing of Environment, 74(2), 229-239.

De Almeida, C. T., Galvao, L. S., Ometto, J. P. H. B., Jacon, A. D., de Souza Pereira, F. R., Sato, L. Y., ... & Longo, M. (2019). Combining LiDAR and hyperspectral data for aboveground biomass modeling in the Brazilian Amazon using different regression algorithms. Remote Sensing of Environment, 232, 111323.

De Almeida, D. R. A., Broadbent, E. N., Ferreira, M. P., Meli, P., Zambrano, A. M. A., Gorgens, E. B., ... & Brancalion, P. H. (2021). Monitoring restored tropical forest diversity and structure through UAV-borne hyperspectral and lidar fusion. Remote Sensing of Environment, 264, 112582.

Degerickx, J., Roberts, D. A., McFadden, J. P., Hermy, M., & Somers, B. (2018). Urban tree health assessment using airborne hyperspectral and LiDAR imagery. International journal of applied earth observation and geoinformation, 73, 26-38.

Din, M., Zheng, W., Rashid, M., Wang, S., & Shi, Z. (2017). Evaluating hyperspectral vegetation indices for leaf area index estimation of Oryza sativa L. at diverse phenological stages. Frontiers in plant science, 8, 820.

Englhart, S., Jubanski, J., & Siegert, F. (2013). Quantifying dynamics in tropical peat swamp forest biomass with multi-temporal LiDAR datasets. Remote sensing, 5(5), 2368-2388.

Esposito, G., Salvini, R., Matano, F., Sacchi, M., Danzi, M., Somma, R., & Troise, C. (2017). Multitemporal monitoring of a coastal landslide through SfM-derived point cloud comparison. The Photogrammetric Record, 32(160), 459-479.

Gamon, J. A., & Surfus, J. S. (1999). Assessing leaf pigment content and activity with a reflectometer. The New Phytologist, 143(1), 105-117.

Gamon, J., Serrano, L., & Surfus, J. S. (1997). The photochemical reflectance index: an optical indicator of photosynthetic radiation use efficiency across species, functional types, and nutrient levels. Oecologia, 112(4), 492-501.

Gitelson, A. A., Merzlyak, M. N., & Chivkunova, O. B. (2001). Optical properties and nondestructive estimation of anthocyanin content in plant leaves¶. Photochemistry and photobiology, 74(1), 38-45.

Gitelson, A. A., Zur, Y., Chivkunova, O. B., & Merzlyak, M. N. (2002). Assessing carotenoid content in plant leaves with reflectance spectroscopy¶. Photochemistry and photobiology, 75(3), 272-281.

Huang, S., Price, D., & Titus, S. J. (2000). Development of ecoregion-based height–diameter models for white spruce in boreal forests. Forest ecology and management, 129(1-3), 125-141.

Jennings, S. B., Brown, N. D., & Sheil, D. (1999). Assessing forest canopies and understorey illumination: canopy closure, canopy cover and other measures. Forestry: An International Journal of Forest Research, 72(1), 59-74.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. Advances in neural information processing systems, 30.

Kim, J. H. (2019). Multicollinearity and misleading statistical results. Korean journal of anesthesiology, 72(6), 558-569.

Knapp, N., Fischer, R., Cazcarra-Bes, V., & Huth, A. (2020). Structure metrics to generalize biomass estimation from lidar across forest types from different continents. Remote Sensing of Environment, 237, 111597.

Koch, B. (2010). Status and future of laser scanning, synthetic aperture radar and hyperspectral remote sensing data for forest biomass assessment. ISPRS Journal of Photogrammetry and Remote sensing, 65(6), 581-590.

Korpela, I., Ørka, H. O., Maltamo, M., Tokola, T., & Hyyppä, J. (2010). Tree species classification using airborne LiDAR–effects of stand and tree parameters, downsizing of training set, intensity normalization, and sensor type. Silva Fennica, 44(2), 319-339.

Kotsiantis, S. B. (2013). Decision trees: a recent overview. Artificial Intelligence Review, 39(4), 261-283.

Kwak, C., & Clayton-Matthews, A. (2002). Multinomial logistic regression. Nursing research, 51(6), 404-410.

L., Coops, N. C., Aven, N. W., & Pang, Y. (2017). Mapping urban tree species using integrated airborne hyperspectral and LiDAR remote sensing data. Remote Sensing of Environment, 200, 170-182.

Li, Q., Wong, F. K. K., & Fung, T. (2021). Mapping multi-layered mangroves from multispectral, hyperspectral, and LiDAR data. Remote Sensing of Environment, 258, 112403.

Mayra, J., Keski-Saari, S., Kivinen, S., Tanhuanpaa, T., Hurskainen, P., Kullberg, P., ... & Vihervaara, P. (2021). Tree species classification from airborne hyperspectral and LiDAR data using 3D convolutional neural networks. Remote Sensing of Environment, 256, 112322.

Merzlyak, J., et al. “Non-destructive Optical Detection of Pigment Changes During Leaf Senescence and Fruit Ripening.“ Physiologia Plantarum 106 (1999): 135-141.

Merzlyak, M. N., Gitelson, A. A., Chivkunova, O. B., & Rakitin, V. Y. (1999). Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening. Physiologia plantarum, 106(1), 135-141.

N. M. Singer and V. K. Asari, “DALES Objects: A Large Scale Benchmark Dataset for Instance Segmentation in Aerial Lidar,“ in IEEE Access, vol. 9, pp. 97495-97504, 2021, doi: 10.1109/ACCESS.2021.3094127.

Nilson, T. (1971). A theoretical analysis of the frequency of gaps in plant stands. Agricultural meteorology, 8, 25-38.

Penuelas, J., Filella, I., & Gamon, J. A. (1995). Assessment of photosynthetic radiation?use efficiency with spectral reflectance. New Phytologist, 131(3), 291-296.

Prins, A. J., & Van Niekerk, A. (2021). Crop type mapping using LiDAR, Sentinel-2 and aerial imagery with machine learning algorithms. Geo-spatial Information Science, 24(2), 215-227.

Ratkowsky, D. A., & Reedy, T. J. (1986). Choosing near-linear parameters in the four-parameter logistic model for radioligand and related assays. Biometrics, 575-582.

Sims, D. A., & Gamon, J. A. (2002). Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. Remote sensing of environment, 81(2-3), 337-354.

Song, Y., & Ryu, Y. (2015). Seasonal changes in vertical canopy structure in a temperate broadleaved forest in Korea. Ecological research, 30(5), 821-831.

Tong, S., & Chang, E. (2001, October). Support vector machine active learning for image retrieval. In Proceedings of the ninth ACM international conference on Multimedia (pp. 107-118).

Vogelmann, J. E., Rock, B. N., & Moss, D. M. (1993). Red edge spectral measurements from sugar maple leaves. TitleREMOTE SENSING, 14(8), 1563-1575.

Weinstein, B. G., Marconi, S., Aubry-Kientz, M., Vincent, G., Senyondo, H., & White, E. P. (2020). DeepForest: A Python package for RGB deep learning tree crown delineation. Methods in Ecology and Evolution, 11(12), 1743-1751.

Yan, P., Han, Q., Feng, Y., & Kang, S. (2022). Estimating LAI for Cotton Using Multisource UAV Data and a Modified Universal Model. Remote Sensing, 14(17), 4272.

Zhu, X., Skidmore, A. K., Wang, T., Liu, J., Darvishzadeh, R., Shi, Y., ... & Heurich, M. (2018). Improving leaf area index (LAI) estimation by correcting for clumping and woody effects using terrestrial laser scanning. Agricultural and forest meteorology, 263, 276-286.

(2) Book

Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). The elements of statistical learning: data mining, inference, and prediction (Vol. 2, pp. 1-758). New York: springer.

초록

다중시기 항공 LiDAR와 초분광 영상을 활용한 도시림 수종 분류 및 바이오매스 추정

김 대 열
서울대학교 환경대학원 환경조경학과

기후변화가 전 세계적 관심사로 부각되고 도시에 거주하는 인구의 비율이 증가함에 따라 대기오염과 열섬 현상을 완화하고 바이오매스 생성, 생물다양성 보존, 탄소 저장 등 다양한 편익을 제공하는 공간으로서 도시림의 중요성이 증가하고 있다. 산림을 구성하는 수종에 따라서 바이오매스 산정량에 따른 탄소 흡수량과 축적량이 다르기 때문에 도시림이 제공하는 편익을 정량적으로 계산하고 기후변화 취약종을 관리하기 위해서는 정확한 수종 분류가 필요하다. 전통적인 산림 모니터링의 경우 산림청에서 항공 영상을 이용한 판독과 현장 조사를 통해 임상도를 제작하여 관리하고 있지만 많은 노동력과 시간이 필요하고 항공 사진으로는 도시림 식생의 수직구조를 파악할 수 없기 때문에 대상지에서 생장하는 수종을 분류하고 경계를 정확하게 구분하는 방법이 요구되고 있다. 선행연구에 따르면 항공 LiDAR에서 파생된 산림구조 특성과 초분광영상의 분광 반사율을 이용하는 효과적인 산림 모니터링 연구가 많이 진행되고 있다. 최근에는 측량 기술의 발달로 인해 고밀도의(10 point/m^2) LiDAR 점군 데이터를 획득이 가능하게 되었고 오픈소스 소프트웨어서도 점군 데이터의 활용이 용이하게 되었으며 초분광 영상의 경우 다중 분광 영상에 비해 확대된 식생지수 목록과 전처리 및 보정 알고리즘 등이 개발되었다.

본 연구에서는 전통적인 산림조사 방법을 개선하기 위해 두시기의 초분광 영상과 항공 LiDAR 영상을 결합하여 각 자료가 가지는 특징과 식생의 계절적 특성 변화를 활용하여 수종 분류의 정확도와 효율성을 높이고 환경 계획에 활용 가능한 지도를 제작하여 도시림의 수종 분포를 파악하고자 하였으며 최종 분류 결과를 기반으로 대상지의 지상부 바이오매스를 산정하는 것을 목표로 하였다. 대상지는 북위 37° 23’ ~ 37° 27’, 동경 126° 57’ ~ 127° 02’의 경기도 과천시 도시림으로 면적은 2,034 ha이고 10종의 주요 수종이 성립하고 있다. 분류를 위한 현장 조사 자료는 8월부터 10월에 취득된 산림조사결과 데이터를 사용하였고, 항공 LiDAR는 Leaf-on (11월), Leaf-off (4월) 시기에 취득되었고 항공 초분광의 경우 Leaf-on (9월), Leaf-off (11월)에 취득된 데이터셋을 사용하였다.

항공 LiDAR와 항공 초분광 데이터셋은 전처리 과정을 통해 보정되었으며 도시림의 식생 영역을 대상으로 항공 초분광 영상의 PC1 밴드, 잎의 색소 및 광합성 특성과 관련된 식생지수와 항공 LiDAR 영상의 높이, 반사강도 메트릭스 계산을 통해 수종 분류를 위한 독립변수 29개를 추출하였다. 대상지의 대표 수종 10종을 대상으로 16,522개의 랜덤 포인트를 생성하여 결측치를 제외한 총 165,216개의 데이터셋을 생성하였고 현장 조사에서 획득된 수종 정보를 기반으로 로지스틱 회귀 (LR), 서포트 벡터 머신 (SVM), 의사결정나무 (DT), 랜덤포레스트 (RF), Light Gradient Boosting Machine (LGBM)의 5개의 머신러닝 분류 모델을 학습하여 분류와 검증을 수행하였다.

머신러닝 학습을 통한 분류 결과 다중시기 다중 데이터셋의 5개 분류기 평균 정확도는 71%으로 단일시기 다중 데이터셋 (leaf-on: 57%; leaf-off: 61%)와 다중시기 단일 데이터셋 (항공 초분광: 64%; 항공 LiDAR: 55%)에 비해 높게 나타났다. 5개의 데이터셋의 머신러닝 분류기별 정확도 비교 결과 RF의 평균 정확도는 76%으로 LGBM (70%), DT (61%), SVM (60%), LR (39%)에 비해 높게 나타났다. 결과적으로 다중시기 다중 데이터셋을 RF 기법을 이용한 분류의 정

확도가 83.3%로 (Kappa: 0.80) 가장 높은 것으로 나타났다. 수종 분류에 기여하는 주요 독립변수는 11월에 취득된 항공 초분광 데이터셋에서 추출된 Carotenoid 반사 지수 (Importance: 0.064)와 4월에 취득된 항공 LiDAR 영상의 엽면적 지수 (Importance: 0.062)로 추정되었다.

개별 수목 추출 알고리즘을 통해 추출한 928,015개의 수관 영역을 Modified Logistic 수고-흉고직경 관계식을 사용하여 개체목의 흉고직경을 도출하고 수고 및 흉고직경을 입목재적·바이오매스 및 임분수확표의 부위별 상대생장식에 대입하여 2 m 이상의 교목을 대상으로 지상부 바이오매스를 도출한 결과 총 45,351 t의 바이오매스를 산정하였다.

항공 LiDAR와 항공 초분광을 활용한 수종 분류는 시 단위의 도시림에서 80% 이상의 정확도로 수종 분류를 수행할 수 있음을 시사하였으며 단일 시기에 촬영된 영상에 비해 잎의 생장 시기, 갈변 시기, 낙엽 시기에 따라 촬영된 영상을 결합하여 실제 산림의 계절적 특징을 반영했을 때 분류 정확도가 증가하는 것이 뚜렷하게 나타났다. 분류 결과를 시각화한 수종 지도를 토대로 기후변화 취약종을 관리하고 지상부 바이오매스를 산정하여 탄소 흡수량과 저장량을 추정하는 연구에 기여할 수 있을 것으로 사료된다.

주요어 : 항공 LiDAR, 항공 초분광, 도시림, 수종 분류, 지상부 바이오매스, 다중시기

학 번 : 2021-22194

Master’s Thesis of Landscape Architecture

Identification of tree species and the estimation of aboveground biomass in an urban forest using multi-period airborne LiDAR with hyperspectral datasets

다중시기 항공 LiDAR와 초분광 영상을 활용한
도시림 수종 분류 및 바이오매스 추정

February **2023**

Seoul National University
Graduate School of Environment Studies
Landscape Architecture
Dae Yeol Kim

Identification of tree species and the estimation of aboveground biomass in an urban forest using multi-period airborne LiDAR with hyperspectral datasets

Prof. Song, Youngkeun

Submitting a master's thesis of Public Administration
October 2022

Seoul National University
Graduate School of Environment Studies
Landscape Architecture
Dae Yeol Kim

Confirming the master's thesis written by
Dae Yeol Kim
December 2022

Chair ______________________ (Seal)

Vice Chair ______________________ (Seal)

Examiner ______________________ (Seal)

Identification of tree species and the estimation of aboveground biomass in an urban forest using multi-period airborne LiDAR with hyperspectral datasets

서울대학교 환경대학원 환경조경학과
Dae Yeol Kim

위 논문은 서울대학교 및 환경대학원 환경조경학과 학위논문
관련 규정에 의거하여 심사위원의 지도과정을 충실히
이수하였음을 확인합니다.

2023년 2월

위 원 장 ______________ (서울대학교 환경대학원 교수)

부위원장 ______________ (서울대학교 농업생명과학대학 교수)

위 원 ______________ (서울대학교 환경대학원 교수)

Abstract

Identification of tree species and the estimation of aboveground biomass in an urban forest using multi-period airborne LiDAR with hyperspectral datasets

Dae Yeol Kim

Dept. of Landscape Architecture

Graduate School of Environmental Studies

Seoul National University

As climate change has emerged as a global concern and the proportion of urban residents has increased, the importance of urban forest as a space that relieves air pollution and urban heat islands and provides various benefits such as biomass generation, biodiversity conservation, and carbon storage has increased. Given that the amount of carbon absorption and accumulation based on the biomass calculation differs by tree species that comprise a forest, accurate tree species classification is required to quantitatively calculate urban forest benefits and manage endangered species. Regarding conventional forest monitoring, the Korea Forest Service produces and manages forest type maps by aerial image analysis and field surveys, a labor-intensive and time-consuming approach. In addition, because aerial

imaging cannot identify the vertical structure of urban forest vegetation, a method for classifying the species of the research sites and accurately distinguishing boundaries is required. Notably, many effective forest monitoring studies are being conducted using forest structure characteristics derived from airborne LiDAR and the spectral reflectance of hyperspectral images. With survey technology advancements, LiDAR point data with high density (10 point/m^2) can be obtained, point data can be easily used in open-source software, and hyperspectral images can be developed with expanded vegetation index lists and preprocessing and correction algorithms. In this research, the traditional forest survey method was improved by combining airborne hyperspectral images (AHI) with airborne LiDAR data from two periods, thereby leveraging the characteristics of each data set and seasonal characteristics of vegetation. The goal was to increase the accuracy and efficiency of tree classification, understand species distribution in urban forests by creating an environmental planning map, and calculate the research site Aboveground biomass (AGB) based on the classification results. The research site is an urban forest in Gwacheon, Gyeonggi-do, located at 37° 23'–37° 27' north latitude and 126° 57'–127° 02' east longitude, with an area of 2,034 ha and 10 major species. Forest surveys were conducted from August to October to gather field survey data for classification; the airborne LiDAR dataset was acquired during the leaf-on (November) and leaf-off (April) periods, whereas the AHI dataset was acquired during the leaf-on (September) and leaf-off (November) periods. The airborne LiDAR and AHI datasets were calibrated through preprocessing, and 29 independent variables for tree classification were extracted by calculating the PC1 band of AHI, the vegetation index related to the pigment and photosynthesis properties of leaves, and the height of airborne LiDAR images. In addition, 165,216 points were obtained by generating 16,522 random points for 10 major species of the research site, excluding missing values. Classification and verification were performed by learning five machine learning classification models of logistic regression (LR), support vector machine (SVM), decision tree (DT), random forest (RF), and light gradient boosting machine (LGBM) based on tree species information obtained from field surveys. The tree classification results indicated that the average accuracy of the five classifiers for the multitemporal multidataset was 71%, exceeding those of the single temporal multidataset (leaf-on: 57%; leaf-off: 61%) and multitemporal single dataset (AHI: 64%; airborne LiDAR: 55%). Comparing the accuracy of each machine learning classifier on five datasets revealed that RF had the highest average accuracy (76%), followed by LGBM (70%), DT (61%), SVM (60%), and LR (39%). Consequently, the classification accuracy of multitemporal multidatasets using RF techniques was also highest (83.3%; Kappa: 0.80). The main independent variables contributing to tree classification were the CRI (Importance: 0.064) extracted from the AHI dataset acquired in November and the leaf area index (Importance: 0.062) of the airborne LiDAR images acquired in April. The diameter at breast height (DBH) of the independent tree was derived using the modified logistic tree height (TH) - DBH relational expression of 928,015 tree crown areas, which were extracted using the independent tree segmentation algorithm. By substituting TH and DBH into the allometric equations for each part of the tree volume, biomass, and stand yield table, AGB was derived for trees with at least 2 m height, and a total biomass of 45,351 t was calculated. Tree classification using airborne LiDAR and AHI could result in over 80%

accuracy when employed in urban forests, and the forest's actual seasonal characteristics were clearly increased by combining images acquired based on leaf growth, fall foliage season, and leaf fall season compared to single temporally acquired images. Overall, research into the estimation of carbon absorption and storage through the management of climate change-vulnerable species and AGB computation can benefit from tree species maps visualized in the classification results.

Keywords: : AIRBORNE LiDAR, AIRBORNE HYPERSPECTRAL IMAGING, URBAN FOREST, TREE SPECIES CLASSIFICATION, ABOVEGROUND-BIOMASS (AGB), MULTI-TEMPORAL
Student Number : 2021-22194

Contents

[List of Tables, Figures and Equations]

▌ List of Tables

▌ List of Figures

▌ List of Equations

Chapter 1. Introduction

1.1. Research Background and Purpose

1.1.1. Background of the Research

1) Climate Change and Carbon Neutrality

Climate change has become a global concern, with increasing atmospheric carbon dioxide (CO_2) concentration being identified as the main cause. Carbon-based organic matter, a material that is the basis of life, accounts for approximately half of the mass of all living organisms and is primarily found as fossil fuel in sedimentary rocks and as CO_2 in the atmosphere. Before the Industrial Revolution, the CO_2 concentration averaged 280 parts per million (ppm) for approximately 6,000 years. However, in 2022, it reached 412.5 ppm, an increase of more than 50%. In 2017, the average global temperature reached 1° C above pre-industrial levels.

Owing to their global occurrence, climate change–induced problems, such as heat waves, droughts, floods, natural disasters, food shortages, spread of pests, species extinction, and ecosystem changes, have emerged as global concerns. The Intergovernmental Panel on Climate Change (IPCC) is an agency under the UN, and countries around the world are implementing policies aimed at realizing carbon neutrality by 2050. In addition, to comply with international regulations, methods for quantifying carbon absorption and storage are being actively discussed.

2) Urban Forest and Forest Monitoring Using Remote Sensing

The Creation And Management Of Urban Forest Act (National Assembly of the Republic of Korea, 2020) defines urban forests as forests and trees created and managed in cities to promote public health and recreation, cultivation of emotions, and experiential activities. Urban forest is a concept that includes urban forests, parks, and street trees. According to the UN 2011 report, the share of the global population residing in cities would be estimated at 56.15% by 2020 owing to rising migration to cities. In Korea, 91.4% of the population based on administrative districts lives in the city (Ministry of Land, Infrastructure and Transport, 2021). Notably, urban forests are becoming increasingly important as a space that provides various benefits such as carbon storage, biomass generation, air pollution mitigation, heat island reduction, and biodiversity conservation (Escobedo et al., 2011).

Since the emergence of the global COVID-19 pandemic, the demand for green spaces in cities, which had been previously neglected, has increased. Therefore, the management and restoration of natural spaces that play a multifunctional role in the city are becoming increasingly crucial. The time has come to integrate smart technology into quality management of green spaces. With the developed countries government's smart green city policy, discussions on environmental management and urban ecosystem preservation are ongoing, which are aimed at using various data collection sensors to address urban issues. Big data, Global Positioning System (GPS), and sensor technologies are now integrated as sustainable development goals are achieved at the city level and the spatial scale is reduced from the country to the city level. Natural environment management technology that monitors, evaluates, and preserves the natural ecological environment is becoming crucial for addressing urban issues and reducing carbon emissions. Because the amount of carbon absorbed and accumulated in the atmosphere varies by tree species, precise classification of tree species is necessary to quantify and manage urban forest benefits.

Traditional forest monitoring employed by national agencies such as the Korea Forest Service entails producing and managing forest type maps by interpreting aerial images and conducting field surveys; however, this strategy is time-consuming, labor-intensive, and can result in a poor understanding of the vertical structure of vegetation and inaccurate classification of species boundaries. Therefore, an accurate tree species classification method that improves existing methodologies is required for quantifying urban forest benefits, conserving biodiversity, and managing climate change-vulnerable species.

3) Remote Sensing Using Multiple Datasets

While previous studies have focused on improving object classification accuracy using satellite images or orthophoto processing, recent forest monitoring studies have focused on data convergence techniques that utilize the advantages of various types of datasets rather than a single remote sensing dataset. The price of remote sensing sensors is decreasing owing to technological advancements, and the use of convergence analysis, which targets a large region, is expanding. When used for forest monitoring, airborne Light Detection and Ranging (LiDAR) can effectively estimate the tree height (TH), diameter at breast height (DBH), tree crown area, and volume because it can create a three-dimensional model of the target site and extract it for the crown area after recognizing individual trees (Simonson et al., 2014). Recently, the efficiency of surveying techniques and sensors has increased to the point that data on the target site can be obtained at a resolution of at least 10 points/㎡, thereby enabling detailed reproduction. Improved accuracy of classification of species and biomass calculation may be obtained using high-resolution airborne LiDAR data. In addition, airborne hyperspectral imaging (AHI) can leverage the advantage of high spectral resolution via more than 100 spectral bands to identify vitality and health status beyond the classification of species. In urban forests in temperate climates, where artificial forests, natural forests, and various species of trees coexist, the classification of tree units requires hyperspectral images with high-resolution spectral wavelengths and various vegetation indices suitable for leaf and reflective characteristics.

4) Classification of Tree Species through Machine Learning

Due to their structural nature, the size and complexity of raw data are large, and the researcher's expertise and research setting ability are required to remove unnecessary data and classify the desired target species. Supervised machine learning for tree classification is widely used in forests and ecology. Here various machine learning classification techniques are employed to create a model that describes the relationship between the dependent and independent variables after extracting the independent variables used for tree classification from remote sensing datasets.

Although logistic regression (LR) is mainly used for binary classification, it can also be used for multinomial classification when it contains at least three categories in the case of multinomial logistic regression (Kwak and Clayton-Matthew, 2002). Even when independent and dependent variables are not linearly correlated, LR can classify them. Furthermore, a support vector machine (SVM) performs classification using a hyperplane that optimizes margins to classify high-dimensional data, and the classification criteria are evenly distributed on both sides of the data to avoid bias (Georganas et al., 2001). Its advantages over neural network methods include its applicability to classification and prediction tasks and its reduced overfitting and impact on error data. However, SVM requires multiple combination tests to control kernel and model parameters in order to create an optimized model; this slows down model construction when many input datasets exist, and it is difficult to interpret and is a complex black box. Decision tree (DT), a machine learning algorithm for supervised learning, can be classified and regressed as a decision rule. DT is a classification algorithm that comprises node segmentation and pruning trees and can evaluate the validity of the optimal tree using data for verification (Kotsiantis, 2013). Owing to its intuitive structure, DT is easy to interpret and can identify valid input variables, but its poor accuracy when overfit and unstable prediction of new data are drawbacks. Random forest (RF) is a machine learning technique that learns a model by randomly selecting some variables and forming multiple DTs (Breiman, 2001). It uses an ensemble method that combines multiple predictions to classify the most votes received as the final prediction (Hastie et al., 2001). RF generates and learns several DT classifiers using a bagging method that randomly selects features from extracted samples using a bootstrap method that permits redundancy. RF accuracy surpasses that of DT, and it can retain high accuracy even as the percentage of missing values increases by reducing predictive variability and preventing overfitting. However, as the number of data increases, RF speed decreases relative to DT, and tree separation becomes complicated, making it difficult to analyze individual trees and interpret the results.

Because the light gradient boosting machine (LGBM) grows trees vertically using the leaf-wise algorithm, it can minimize the loss of predictive errors, unlike the level-wise algorithm method (Ke, 2017). Despite its speed, less memory usage, extremely accurate results, and graphics processing unit (GPU) utilization, LGBM is more prone to over-aggregation on small datasets; hence, datasets with at least 10,000 rows are recommended for its effective use.

5) Biomass Calculation Using Growth Allometry

Because carbon emission rights in countries around the world enable transactions between countries and companies in a lifelike concept, the calculation of green carbon, which is carbon absorbed and stored by the land ecosystem, has become important. Forest biomass is part of green carbon and is an important indicator of forest productivity and carbon circulation (Lim, 2009); the carbon storage and carbon intake of target sites can be calculated by applying biomass expansion and carbon coefficients. Therefore, after classifying the target site species for the preservation value of urban forests, the ground biomass is estimated using the carbon emission coefficient and biomass relative bio-decorations for each species provided by the National Institute of Forest Science. While Cho Hyun-gil (1999) computed the carbon storage and annual carbon absorption of domestic trees using allometric equations per species, Lim Jong-soo (2009) evaluated forest biomass statistics using a regression model and k-nearest neighbor (k-NN) with Landsat TM-5. Integrating the data from LiDAR and hyperspectral sensors can improve terrestrial biomass accuracy from a single dataset (Koch, 2010) while quantifying biomass and visualizing it with tree classification results.

1.1.2. Research Purpose and Significance

1) Research Purpose

Therefore, this research classifies major species of forests in Gwacheon, a temperate climate area, and calculates aboveground biomass by combining AHI and airborne LiDAR during the leaf-on and leaf-off periods, assuming that changes in forest structure and leaf growth cycle in different urban forests can be identified by combining remote sensing data.

The main variables for tree classification are extracted from a forest, which is a natural urban forest excluding street trees and park vegetation, and the accuracy of five classifiers (LR, SVM DT, RF, and LGBM) is compared. Field survey data are used to answer and validate the following research questions: (1) Can the classification between species be explained by employing different datasets from airborne LiDAR and AHI collected during two distinct seasonal periods? (2) Which machine learning classifier learns high-resolution remote sensing data, reflects the situation in the field, and records high accuracy? (3) Is it possible to calculate biomass and review accuracy in indivisible tree units based on tree classification results?

2) Significance of Research

Although comparative studies that integrate high-resolution hyperspectral and LiDAR datasets for domestic tree classification and carbon accumulation calculation are insufficient, some studies on urban forests at the city level, as opposed to small experimental sites, exist. In addition, high-resolution airborne LiDAR with an average point density of 42.7 points/㎡ was used as a variable for classifying forests and trees during the leaf-on and leaf-off periods, and 1 m spatial resolution and 127 spectral spectra were used compared to Sentienl-2 satellite images that provide 10 m spatial resolution and 13 bands. Because most existing studies have focused on pixel-based tree classification, pixel outliers are often reflected in the tree classification results, although it is meaningful as it calculates the polygon-based independent tree area and uses the pixel's median value through zone statistics.

1.1.3. Research Scope

1) Spatial and Temporal Extent

The research site is a natural urban forest, including Umyeonsan Mountain (293 m above sea level [ASL]), Gwanaksan Mountain (632 m ASL), and Cheonggye Mountain (618 m ASL), adjacent to Gwanak-gu and Seocho-gu in Seoul in the north, Seongnam-si in the east, Uiwang-si in the south, and Anyang-si in the west. Since the city area is located in the center of Gwacheon-si, small forest patches exist near the city area, surrounding the forest areas on the left and right sides of the center. The temporal range of the research was November 03, 2021, April 02, 2022, and September 01, 2022, and the airborne LiDAR data were acquired in November 2021 and April 2022, and the AHI data were acquired in November 2021 and September 2022. As for the field survey data, minor class attribute data among the representative biotop types surveyed by the Gyeonggi Research Institute from August to October 2022 were used.

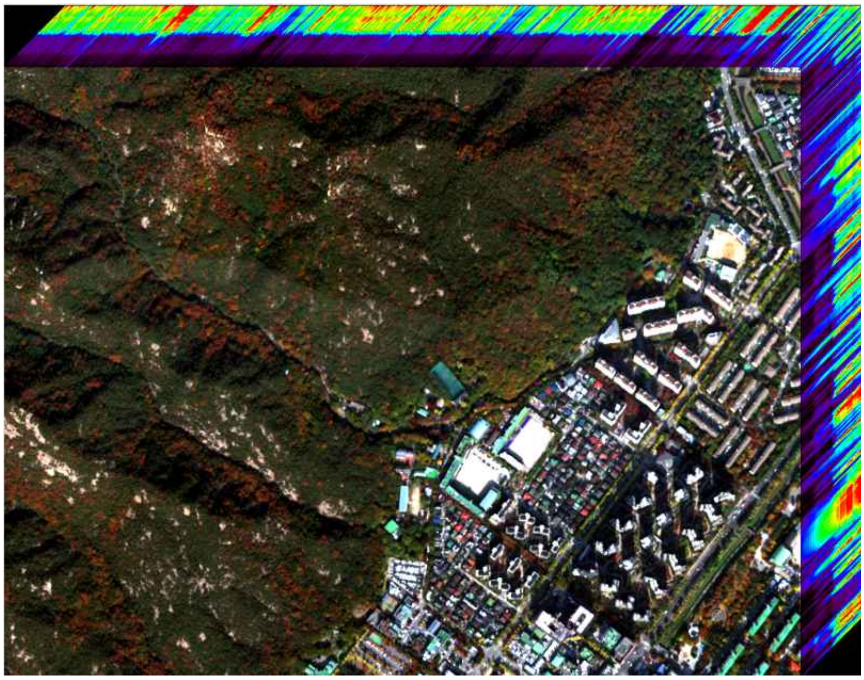


Fig. 1.1 An example of a hyperspectral imaging dataset of the research area

2) Content Scope

Using airborne LiDAR and AHI, which are different types of remote sensing data from the two periods, this research classified 10 representative species of urban forest using 5 machine learning techniques: LR, SVM, DT, RF, and LGBM. Based on the results of random sampling and classification of 29 features and forest field survey data obtained through the interband calculation and structure metrics algorithm from remote sensing data, the above ground biomass (AGB) of urban forests was estimated by substituting parameters for the TH-DBH relational expression and ground biomass calculation.

Chapter 2. Literature Review

2.1. Tree Species Classification and Biomass Estimation

2.1.1. Tree Species Classification and Biomass Estimation

Forest monitoring and biomass calculation by integrating remote sensing data and environmental attribute information have been actively practiced. Recently, many effective forest monitoring studies have utilized the canopy structure derived from airborne LiDAR and vegetation spectral characteristics of hyperspectral images (De Almeida et al., 2021). As survey technology and measurement sensors advance, LiDAR point data with a high density (10 point/m^2) can be acquired, and point data can be used in easily accessible open-source software. Table 2.1 lists the contents and results of previous studies related to vegetation monitoring, such as tree species classification and biomass calculation through data fusion.

Table 2.1 LIterature review of tree species classification and biomass estimation.

Year	Author	Explanation
2008	An-Jin chang, Hyung-Tae Kim.	Extract tree crown area and calculate biomass using aerial images and airborne LiDAR. The regression equation of tree height and diameter at breast height is calculated through field surveys, and biomass is estimated by application of allometric equation.
2009	Cho, H. K., Shin, M. Y.	Using Landsat TM-5 satellite images and field survey sample points, biomass statistics and biomass distribution plots in uninvestigated area are estimated using regression models and k-Nearest Neighbor.

2012	Lee, Hyun Jik, Ru, Ji Ho.	Classified tree species and calculated forest biomass and carbon absorption using airborne LiDAR data and KOMPSAT-2 satellite images. Classify tree species by 90% or more accuracy on average.
2013	Englhart et al	Developed regression modelbusing aerial images and RapidEye multispectral images from 2007 and 2011. Moniotored changes in average canopy height and ground biomass for peatland area in rainforests.
2014	Hyunggab Cho, Kyu-Sung Lee.	Comparison of classification results for coniferous species of hyperspectral imaging and multispectral imaging. The maximum likelihood method was applied to the hyperspectral image converted by dimensionality reduction, and it was calculated with a classification accuracy of 90% or more.
2014	Alonzo et al	Using airborne hyperspectral imaging and airborne LiDAR tree crown area of 29 tree species in the city was extracted, and an overall accuracy of 93.5% was recorded using a total of 28 independent variables including 7 lidar metrics. 4.2% improvement in accuracy for combined datasets compared to single datasets.
2016	Park Jeong-seo	A representative spectroscopic library of hyperspectral images was created for land cover classification of seven classes. Tree classification technique, overall accuracy is improved by 85% or more.
2017	Luxia Liu et al	Classification of 15 species of urban forest trees using the attribute values of airborne LiDAR and hyperspectral images as variables. Analyze fused datasets with 70% or more of total accuracy when classified as random forest classifier.
2019	De Almeida et al	Hyperspectral and lidar variables that affect biomass calculation are reviewed through various regression equations. Compared to a single dataset, the model of the fusion dataset was found to have a higher correlation with the actual biomass calculation.

2021	Janne et al	8 species classification using 3D-CNN techniques by fusing airborne LiDAR and airborne hyperspectral images. Analyzing the entire fusion dataset with 87% classification accuracy when analyzed using 3D-CNN techniques.
2021	Li, Q et al	Species classification using multiple classifiers and comparison of fused datasets in mangrove areas using airborne LiDAR and airborne hyperspectral images. The accuracy of CNN classification was the highest, and it was analyzed that the correlation of the amount of biomass was high.
2021	Kwon, S. K., Kim, K. M., Lim, J	Classified 9 types of forest into random forest classifier using RapidEye sensor. Using spectral information and texture information, tree species classification accuracy was 69.29%.
2021	De Almeida et al	Biomass estimation using 3 lidar matrices and 18 hyperspectral image matrices related to canopy height and leaf area index index as variables.
2022	Chen et al	Convergence of drone LiDAR images of leaf-off and leaf-on periods for temperate forests with high vegetation density. The tree segmentation accuracy of the fusion image was 16.7% more accurate in broad-leaved forests and 4.4% more accurate in coniferous forests.

In international research cases, many studies are being conducted to maximize the advantages of each sensor and utilize seasonal differences by fusing hyperspectral images and airborne LiDAR datasets or by fusing the same dataset with images acquired at different times. Enghart (2013) emphasized the need to monitor forest conservation and sustainable management by calculating the change in canopy height and AGB for petland in rainforest forests using airborne LiDAR data and multispectral images acquired in different years.

Alonzo (2014) classified 29 species of trees in a city and found that the dataset with 7 LiDAR variables outperformed the single hyperspectral dataset by 4.2% in terms of classification accuracy, resulting in 93.5% of tree species classification accuracy. While Janne Mayra (2021) classified 8 species using 3D-CNN techniques by fusing airborne LiDAR and hyperspectral images, Luxia Liu (2017) classified a fusion dataset comprising variables extracted from AHI and airborne LiDAR as RF analysis. Chen (2022) classified the independent tree area 12.3% more accurately than the single temporal leaf-on data by fusing drone LiDAR images from the leaf-on and leaf-off periods in order to increase the reproduction of tree stems and trees. However, the aforementioned studies focused mainly on classifying coniferous species in urban street trees or park green areas, which are relatively easy to distinguish between independent trees, many were conducted on small target sites rather than city level space scales, and few used AHI and airborne LiDAR images obtained at the same time. Therefore, applying these methods to urban forest species in Korea, where deciduous and coniferous trees coexist, will be difficult, and the raster image-based classification results make it difficult to distinguish species boundaries and do not obtain TH for biomass calculation.

Since the 2000s, various studies have been conducted in Korea to classify species using satellite images, hyperspectral images, and airborne LiDAR and to calculate biomass, and methods to increase accuracy have been developed. Im Jong-soo (2009) conducted regression model analysis and k-NN analysis with 16 independent variables from satellite imaging bands, and real measurements of forest type biomass obtained from 58 outdoor sample points in Muju-gun, Jeollabuk-do. As a result, the accuracy of statistical verification using the k-NN technique was higher, and about 8.39 million tons of biomass and 149 t/ha per unit area were calculated at the target site. The research emphasized the need to calculate forest biomass for each species using high-resolution satellite images to compensate for the 30 m spatial resolution of Landsat TM-5 and to expand biomass estimation from small areas to large areas. While Kwon Soo-kyung (2021) used the RapidEye sensor to classify nine forest types using RF techniques, spectral information, and texture information, realizing 69.29% accuracy, Park Jung-seo (2016) selected and applied a representative spectral library for land cover classification of hyperspectral images to improve overall accuracy by more than 85%. While Cho Hyung-gap (2014) employed dimensional reduction to apply hyperspectral images, realizing over 90% classification accuracy via maximum likelihood classification, Lee Hyun-jik (2012) and Jangan-jin (2008) integrated orthophoto, satellite image data, and airborne LiDAR data to obtain biomass estimate results.

Although studies on tree species classification or biomass estimation using satellite images have become common in Korea, studies integrating airborne LiDAR and hyperspectral images remain insufficient, with only a few studies on the independent variable setting for tree classification in temperate forests. In addition, airborne LiDAR has a point density of less than 10 points/㎡, resulting in errors due to low-resolution data, and its application to large-scale target sites is limited when targeting artificial forests with a clear distinction between small data samples and trees. Due to the large size and complexity of raw data when handling the 3D structure of airborne LiDAR images and the AHI dataset comprising a large number of bands, researcher expertise and an analysis method for extracting independent variables that remove unnecessary data and improve tree species classification accuracy are required.

Chapter 3. Methods

3.1. Research Progress

3.1.1. Research Progress

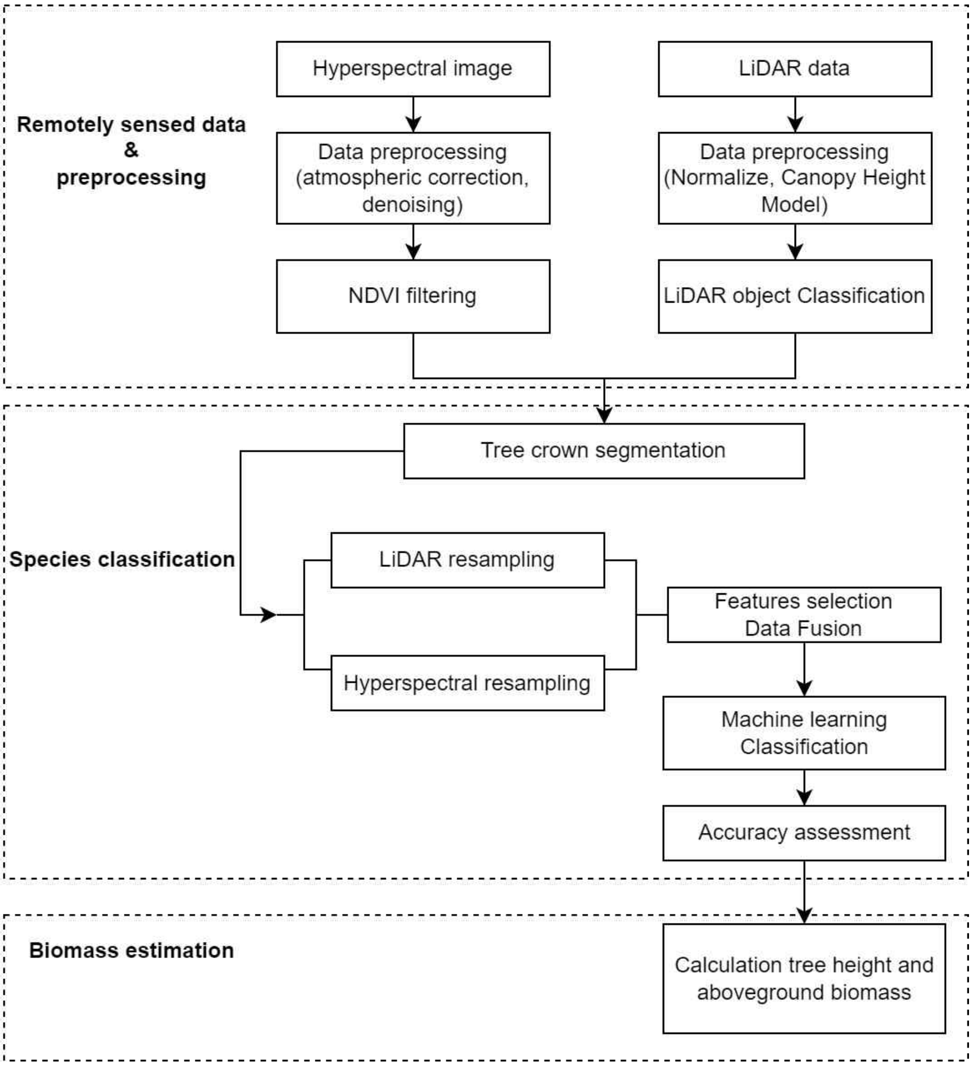


Fig. 3.1 The workflow proposed for tree species classification and AGB calculation

Fig. 3.1 shows this research's order of data processing and analysis, as follows: 1) calibration of remote sensing dataset, 2) individual tree segmentation, 3) calculation of major independent variables for tree classification, 4) training of machine learning classification model and tree classification and accuracy evaluation based on field survey data, and 5) calculation of individual tree heights and aboveground biomass.

3.2. Research Area

3.2.1. Overview of Research Area

Research area is an urban forest located at 37° 23'–37° 27' north latitude and 126° 57'–127° 02' east longitude, with an area of 2,034 ha out of 3,587 ha in Gwacheon Fig. 3.2. Among the forest areas, grassland, cultivated land, and orchards, which are tree species group codes No. 80–99 (Korea Forest Service: forest type map), were excluded from the research's scope, leaving a research area of 2,327 ha containing only the tree species of the research subject. According to the 2020 Forest Basic Statistics (Korea Forest Service, 2020), Gwacheon-si's forest-growing stock was 392,645 ㎥, with a forest growth rate of 64.87% and an average forest-growing stock of 168.73 ㎥/ha. In 2018, As stated in 2020~2024 Climate Change Adaptation Second Action Plan (2019), Gwacheon City had a precipitation of 1,170.5 mm, with an average annual temperature of 12.1° C (ranging from 39.9° C to 17.9° C). Some of the area's forests are adjacent to Seoul, the most populous city in Korea, with a population of 77,818 (Ministry of Security and Public Administration, 2022), so these forests play a role in absorbing carbon dioxide, improving air quality, alleviating heat island, and reducing fine dust.

In addition, as they are adjacent to Seoul City, where it is difficult to obtain remote sensing data due to aviation security, calculating biomass and carbon intake using tree classification seems reasonable. Every five years, the Korea Forest Service produces a forest type map to monitor the forests of the target site. According to the accurate forest type map in 2020, 14 major tree species exist and are occupied in order of area, as shown in Fig. 3.3. Most of the composition was The others quercus forest 29%, Mixed forest 18%, The others deciduous forest 17%, Robinia pseudoacacia 11%, Pinus densiflora 10%. The remaining small percentages of area were Pinus rigida, Quercus acutissima, Quercus variabilis, Pinus koraiensis, Quercus mongolica, Larix Kaempferi, Populus canadensis, Pinus thunbergii, and Abies holophylla. The analysis results of the tree species area ratio reveal that quercus and deciduous forests dominate the area and create a representative landscape.

Notably, the trees in the target area do not form a large colony but are irregularly scattered and distributed throughout the entire area (Fig. 3.2). Owing to its geographical location, Gwacheon-si is surrounded by forests, and the city area is located in the center. Gwacheon-si is largely divided into Gwanaksan Mountain in the west and Cheonggyesan and Umyeonsan mountains in the east. Table 3.1 shows the area and area ratio of the target land species calculated based on the precision forest type map.

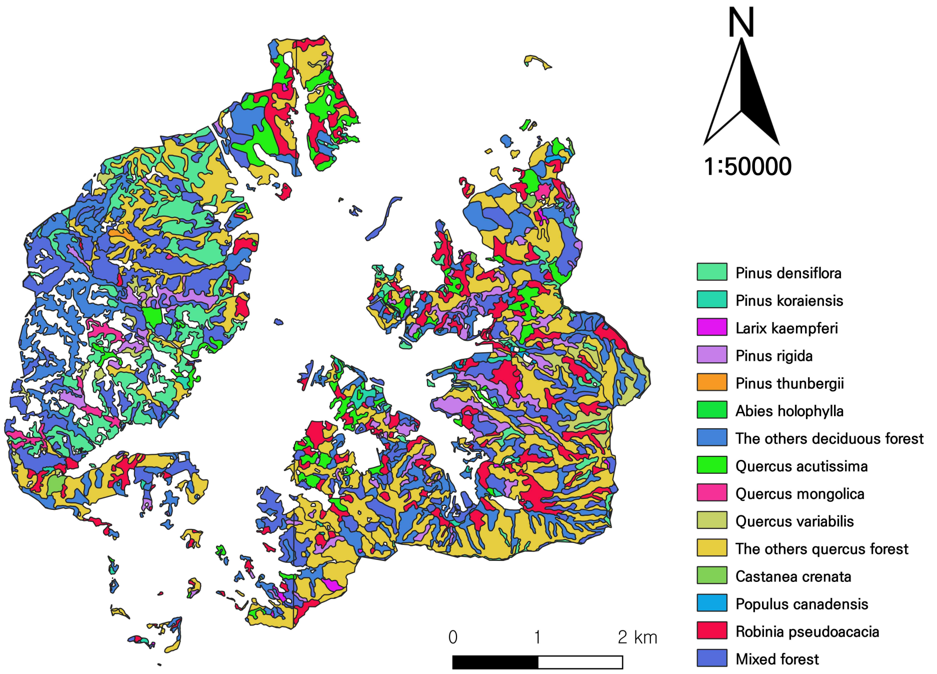


Fig. 3.2 Distribution of forest species in Gwacheon (Korea Forest Service: forest type map)

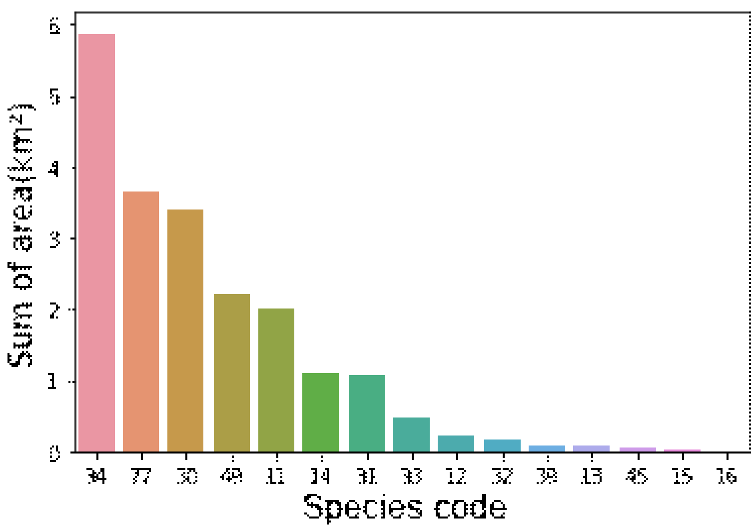


Fig. 3.3 Area ranking of research area species (forest type map).

Table 3.1 Area and area ratio by species.

Code no.	Species	Area (ha)	Area ratio (%)
11	*Pinus densiflora*	200.06	9.83
12	*Pinus koraiensis*	21.83	1.07
13	*Larix Kaempferi*	7.45	0.37
14	*Pinus rigida*	109.96	5.41
15	*Pinus thunbergii*	2.24	0.11
16	*Abies holophylla*	0.36	0.02
30	*The others deciduous forest*	339.25	16.68
31	*Quercus acutissima*	108.21	5.32
32	*Quercus mongolica*	17.81	0.88
33	*Quercus variabilis*	48.45	2.38
34	*The others quercus forest*	587.56	28.88
45	*Populus canadensis*	3.92	0.19
49	*Robinia pseudoacacia*	222.13	10.92
77	*Mixed forest*	365.16	17.95

The analysis of data from the Korea Forest Service's 1:5000 forest type map and 2020 Forest Basic Statistics revealed that the average age of the target area was 4.17, indicating that most forests have a crown occupancy of more than 50% between 31 and 40 years old. In the forest area, Age 4 and Age 5 accounted for 51.4% and 38.7%, whereas Age 4 and Age 5 accounted for 51.5% and 41.8% of the forest-growing stocks.

Table 3.2 Statistic of age class (2020 Forest Basic Statistics).

Age class	Explanation	Forest area (ha)	forest-growing stocks (㎥)
Age 2	Above 50% occupancy of 11~20 year-old trees	10	357
Age 3	Above 50% occupancy of 21~30 year-old trees	149	13,832
Age 4	Above 50% occupancy of 31~40 year-old trees	1,136	202,568
Age 5	Above 50% occupancy of 41~50 year-old trees	856	164,155
Age 6	Above 50% occupancy of 51~60 year-old trees	57	11,733

Notably, 965 polygons were used in the Korea Forest Service's 1:5000 precision forest type map data to analyze the destination's forest canopy height. The frequency of a forest canopy height of 15 m or more and less than 17 m was the highest, and the ratios of 13 to 21 m and 5 to 7 m were relatively high. The forest basin average was 13.86 m, indicating that 8–30 m of trees had grown.

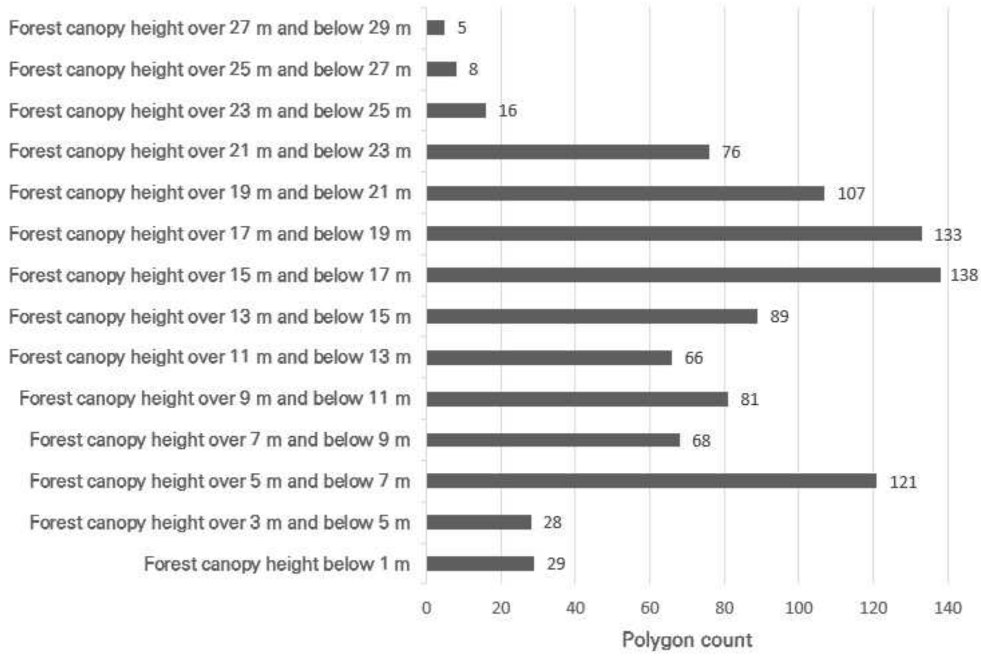


Fig. 3.4 Polygon count by forest canopy height (Korea Forest Service: forest type map).

3.3. Data Acquisition

3.3.1. Airborne LiDAR and Hyperspectral Imaging

For data acquisition, considering the seasonal environmental characteristics of the research area, the airborne LiDAR dataset was considered leaf-on in November before leaf fall and leaf-off in April after leaf fall in an urban forest. While the September AHI dataset was designated leaf-on because it was acquired during the post-monsoon leaf growth period, the November dataset was considered leaf-off because it was acquired during the fall season. Airborne LiDAR data[1] were acquired on November 03, 2021 and

1) Provided by ASIA Aero Survey co., Ltd.

April 02, 2022 and were scanned with Leica's TerrainMapper 1 at 6000 ft ASL, with an average point density of 42.7 points per square meter (pp㎡), a maximum reflection number of 5, and a scan angle of ± 19.998°. Reflection intensity, GPS time, and number of returns were also acquired as attributes. In the case of AHI[2], the target areas were filmed on November 03, 2021 and September 01, 2022. AHI is a SPECIM AISA Eagle sensor comprising 127 units with a spatial resolution of 1 m and a 404–996 nm wavelength range. While the hyperspectral images in November comprised 12 individual strips, those in September comprised 13 individual strip images. Table 3.3 lists the details of the collected remote sensing data.

Table 3.3 Remote sensing data information.

Sensor	Attribute	Value
Airborne Hyperspectral Imaging	Date of Acquisition	2021.11.03., 2022.09.01
	Sensor Model	AISA Eagle
	Spatial Resolution	1 m
	Spectral Range	404–996 nm
	Full Width at Half Maximum (FWHM)	0.44–0.48 nm
	Radiometric Resolution	12 bit

2) Provided by ASIA Aero Survey co., Ltd.

Airborne LiDAR	Date of Acquisition	2021.11.03., 2022.04.02
	Flying Height	6000 ft
	Sensor model	TerrainMapper 1
	Point Density	42.7 points per ㎡(pp㎡)
	Maximum Number of Return	5
	Scan Angle	± 19.998°

3.3.2. Ground Truth Data

Ground truth data, which provides information on specifications and x and y coordinates, was obtained from August to October 2022 as a field survey. Data from several types of minor class attribute subclasses of forest survey result[3] data were used, and the minimum area of sampled field survey data exceeds 100 ㎡ (10 m × 10 m). Forest surveys in polygon area units were conducted for these data, and the polygon ID and field data at the survey point were cross-referenced. The minor class attribute information on these

3) Provided by Gyeonggi Research Institute.

data comprised a polygon containing a single species and a polygon containing multiple mixed forests; however, only a single species from the polygon was used to improve the training accuracy of the machine learning classifier. The forest survey data served as the ground truth for tree classification training and verification.

3.4. Data Processing

3.4.1. Pre-Processing of Airborne LiDAR and Hyperspectral Imaging

Concerning airborne LiDAR images, surface and non-surface points were classified using automatic ground classification and manual classification of the acquired raw LAS data to produce ground classification data (DTD). The standard height was corrected by applying KNGeoid18, a korean national geoid model developed by the Korea National Geographic Information Institute, and system errors were erased through the IMU Calibration. The acquired airborne LiDAR images were mooted into a single LAS file using GreenValley International LiDAR 360 software, and only forest areas were extracted at the urban forest boundary using QGIS' LAStools (https://rapidlasso.com/lastools/). Through R software's cloth simulation filtering, the point where the height value (z) was located at the lowest point was selected and the ground and non-ground areas of the research site points were automatically classified.

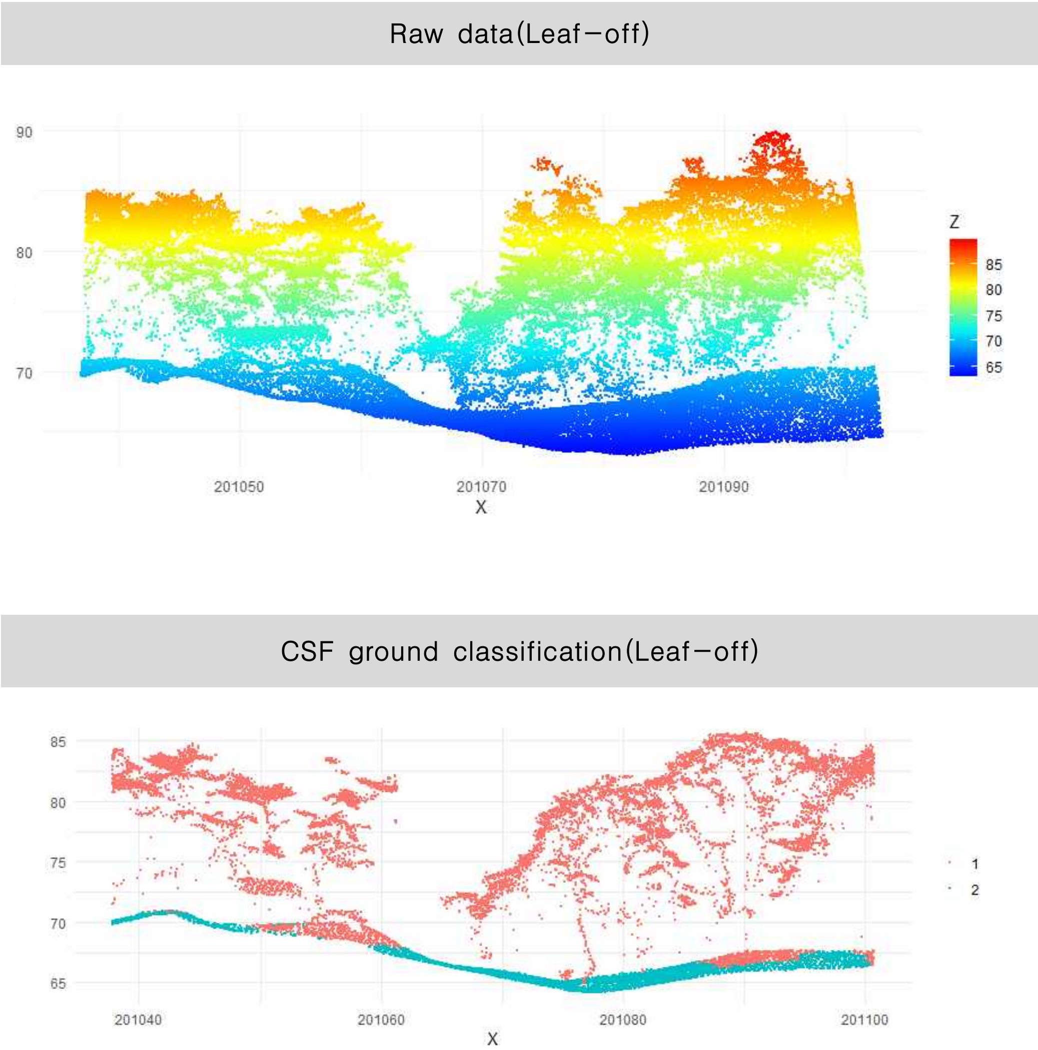


Fig. 3.5 Ground classification using the CSF algorithm.
(1:non-ground; 2:ground; X: x coordinate, Z: Altitude)

Because noise points away from the surface or object may cause errors in LiDAR analysis and classification, noise caused by flying currents or frequency interference was removed through noise filtering in LiDAR 360, and the remaining abnormal points were manually removed through visual inspection. Finally, a normalized three-dimensional LiDAR point cloud was created by subtracting the altitude of the indicator from that of each point. The number of point data for airborne LiDAR acquired in November 2021 was 1,193,108,897, with an average elevation of 6.30 m (standard deviation 5.48) and an average reflection intensity of 730.2 (standard deviation 758.0). The number of point data for airborne LiDAR acquired in April 2022 was 875,226,444, with an average elevation of 5.62 m (standard deviation 5.08) and an average reflectance of 538.89 (standard deviation 792.25). Table 3.4 lists detailed attribute information of the point cloud of the airborne LiDAR image.

Table 3.4 Point cloud attributes of Leaf-on and Leaf-off season.

	Point number	Mean height (m)	Mean intensity
Leaf-on	1,193,108,897	6.30 ± 5.48	730.2 ± 758.0
Leaf-off	875,226,444	5.62 ± 5.08	538.89 ± 792.25

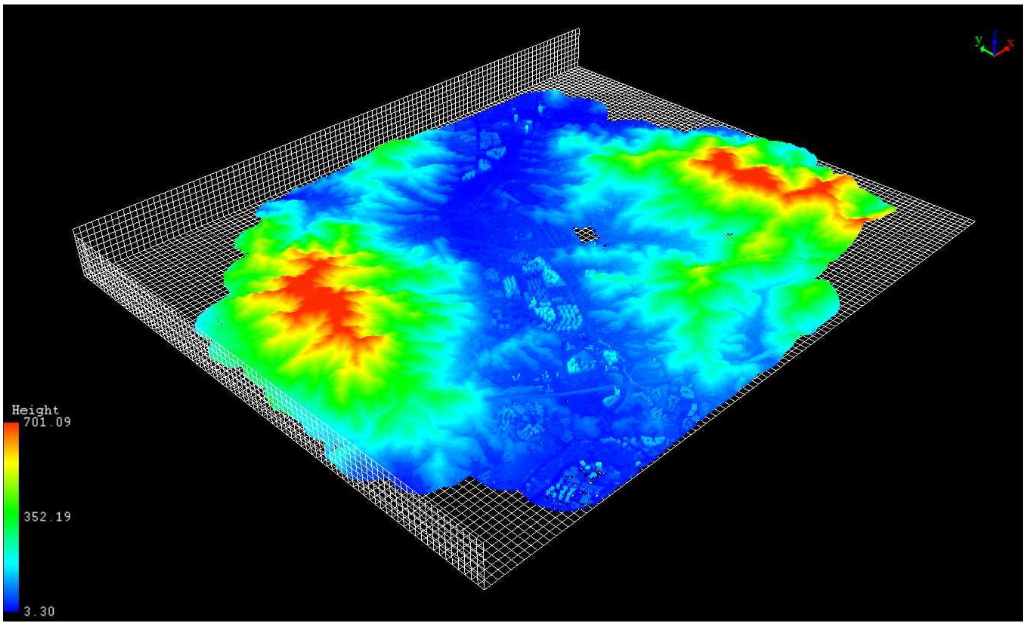


Fig. 3.6 Airborne LiDAR imaging of leaf-on season

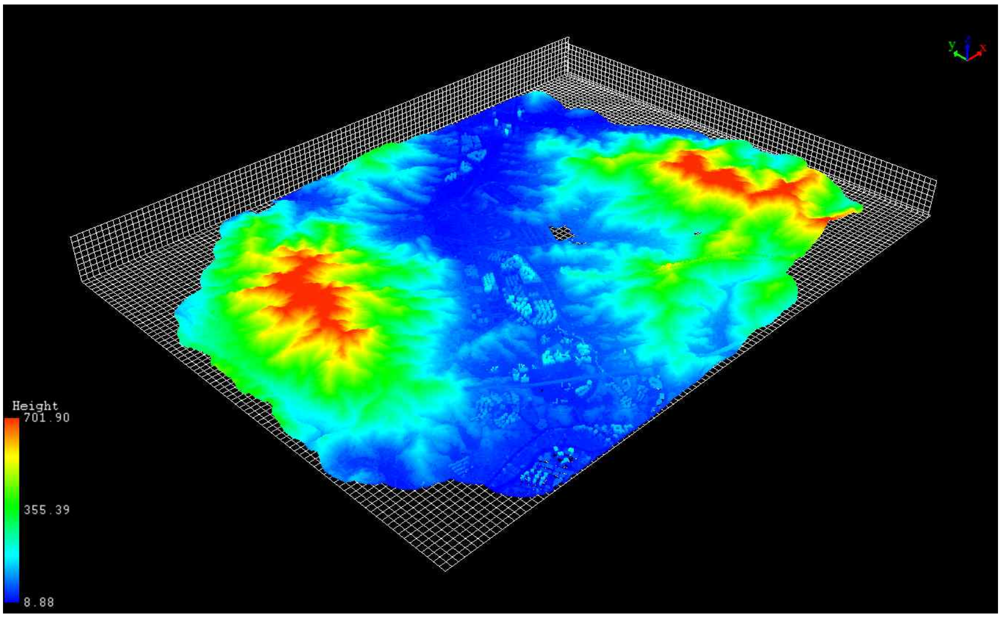


Fig. 3.7 Airborne LiDAR imaging of leaf-off season

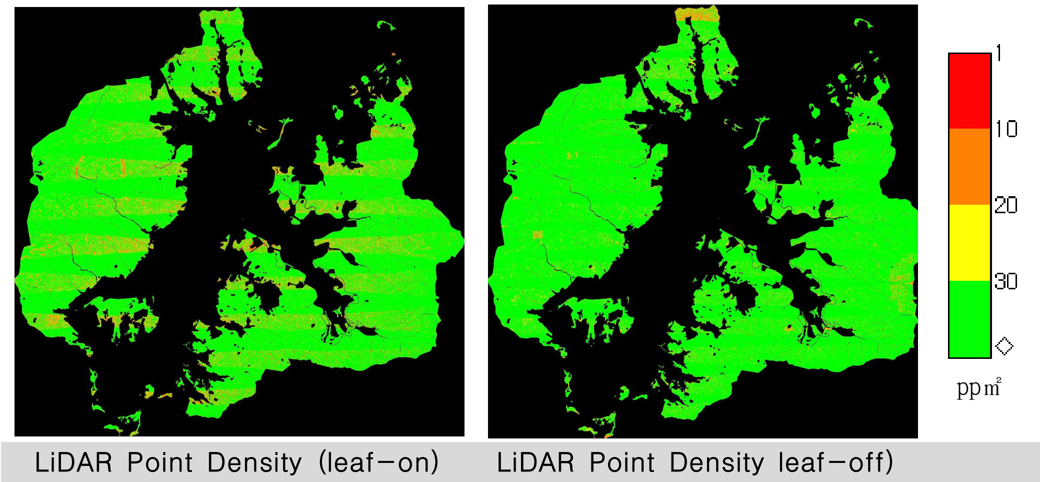


Fig. 3.8 LiDAR point density of leaf-on and leaf-off condition

AHI requires atmospheric, radiation, and geometric correction processes to eliminate errors caused by differences in brightness values from aerosols in the ground atmosphere. While radiation and geometrical corrections were performed using Spectir's SHIPS module, atmospheric correction was performed using Res' ATCOR-4 module. The image coordinates were geo-referenced based on the National Geographic Information Institute's aerial orthophoto photograph (UTM 52N) acquired on May 15, 2020. To compensate for differences in reflectance values caused by differences in shooting time along aircraft flight paths, additional Quick Atmospheric Correction (QUAC) were performed using ENVI 5.6.2 software from L3Harris Geospatial. This software is effective in urban ecosystems where artificial structures, water, vegetation, and soil coexist. The 12 strips acquired in November and 13 strips acquired in September were orthogonally calibrated with each single mosaic image using the Seamless Mosaic algorithm in ENVI 5.6.3.

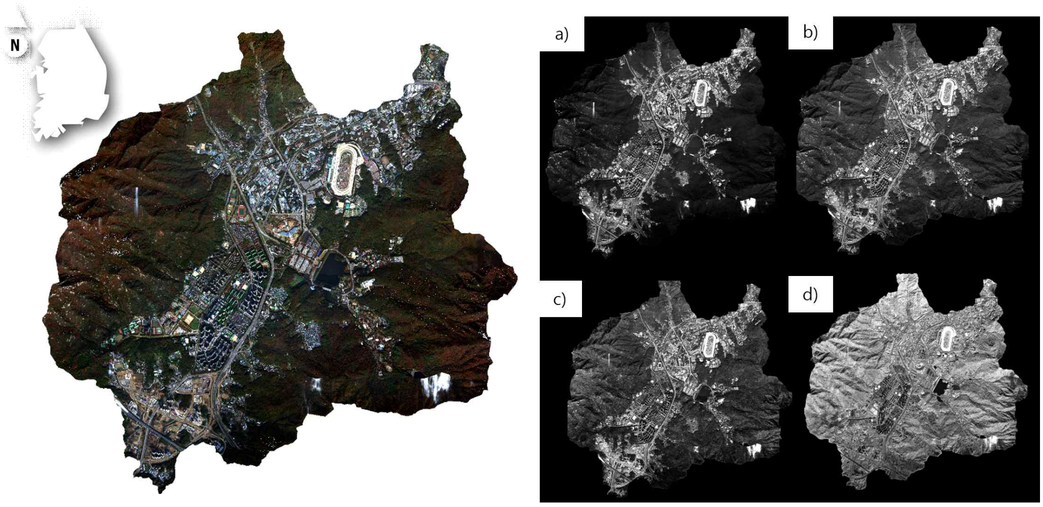


Fig. 3.9 Mosaic image of airborne hyperspectral Imaging (November)
a)Blue, b)Green, c)Red, d)Near infrared

3.4.2. Dataset Composition

In this research, while the classification accuracy of the combined dataset of AHI and airborne LiDAR was compared with that of a single dataset, the classification accuracy of a single temporal dataset was compared with that of a multitemporal dataset in order to determine whether the classification accuracy of the fusion dataset was improved. The front part of the string was set with Hyperspectral Imaging (H) and LiDAR (L), and the back part of the string was set with alphabetical capitalization at the time of data acquisition.

Regarding HS-LN and HN-LA, the acquisition timing differed, but at the time of acquisition, these datasets were considered a single temporal multidataset, considering the environmental characteristics of urban forests. As a result, five datasets were generated: multitemporal single datasets (HN-HS, LN-LA), single temporal multidataset (HS-LN, HN-LA), and multitemporal multidataset (HN-HS-LN-LA). Table 3.5 lists the details of the divided datasets.

Table 3.5 Dataset composition by sensors and date condition.

ID	Dataset	Number of bands
HN-HS	Hyperspectral November, Hyperspectral September	16
LN-LA	LiDAR November, LiDAR April	13
HS-LN	Hyperspectral September, LiDAR November	14
HN-LA	Hyperspectral November, LiDAR April	14
HN-HS-LN-LA	Hyperspectral November, Hyperspectral September, LiDAR November, LiDAR April	29

3.4.3. Selection of Major Tree Species

To eliminate errors caused by the nonexclusive characteristics of dependent variables during machine learning classification, similar tree species should be unified. In the minor class field containing tree species classification of field forest survey data, the top 10 species with a high area ratio were selected for classification based on a single tree area, excluding mixed forests and rock vegetation. Quercus mongolica, Pinus densiflora, Robinia pseudoacacia, Quercus acutissima, Pinus rigida, Quercus variabilis, Castanea crenata, Pinus koraiensis, Quercus serrata, and Larix Kaempferi were representative species, and the classification codes were assigned in order of area. Species investigated with an area ratio of less than 0.2 were Fraxinus lanuginosa, Populus tomentiglandulosa, Quercus aliena, Prunus sargentii, Abies holophylla, and Prunus subg. Cerasus, Liriodendron tulipifera, Zelkova serrata, Cercidiphyllum japonicum, Metasequoia glyptostroboides, and Ziziphus jujuba Mill were excluded from the classification model learning because the area occupied on a wide-area spatial scale was small.

Table 3.6 Area of target tree species.

Species	Class number	Survey area (km^2)	Area ratio (%)
Quercus mongolica	1	3.49	33.3
Pinus densiflora	2	1.96	18.7
Robinia pseudoacacia	3	1.61	15.4
Quercus acutissima	4	1.15	10.9
Pinus rigida	5	0.77	7.3
Quercus variabilis	6	0.53	5.1
Castanea crenata	7	0.33	3.1
Pinus koraiensis	8	0.15	1.4
Quercus serrata	9	0.06	0.6
Larix kaempferi	10	0.02	0.2

3.4.4. Tree Crown Segmentation

The merging of hyperspectral and airborne LiDAR imagery was utilized to extract only the vegetation area of the research area. Using the threshold value of the NDVI index (range: 0.2–0.8) extracted from hyperspectral images, areas below 0.2 were identified as non-vegetation areas, and raster pixels with vegetation reflectance were extracted as masking layers to distinguish between vegetation and non-vegetation areas. Regarding airborne LiDAR data, the University of Dayton DALES dataset was subjected to deep learning training using PointNet++, a hierarchical neural network of MATLAB, and labeled into eight classes: ground, vegetation, cars, trucks, powerlines, fences, poles, and buildings. Point clouds such as power transmission towers, buildings, and fences mixed in urban forests were removed, and AHI was used as a masking layer to extract only the areas filtered as vegetation areas (TH of 2 m or more; tree radius of 2 m or more).

Airborne LiDAR data collected during the leaf-on and leaf-off periods were merged to generate data that expressed both the upper and lower forest layers. The digital elevation model (DEM) and digital surface model (DSM) with 0.5 m spatial resolution were generated using triangulated irregular network (TIN) interpolation by calculating the altitude values (Z) of the ground and non-ground of the data. The canopy height model (CHM), which is a normalized height value from the ground, was generated through the difference between DEM and DSM. Local maximum filtering analysis was performed on CHM to calculate the tree top by selecting the maximum height value of the point cloud, and through watershed analysis, the tree's crown area pixels were selected based on the tree apex and distance of the local minima, and the boundaries between the trees were distinguished. Each tree was assigned a unique tree ID, and the attribute value of each feature was determined by calculating the median value of the raster pixels included in the tree crown area.

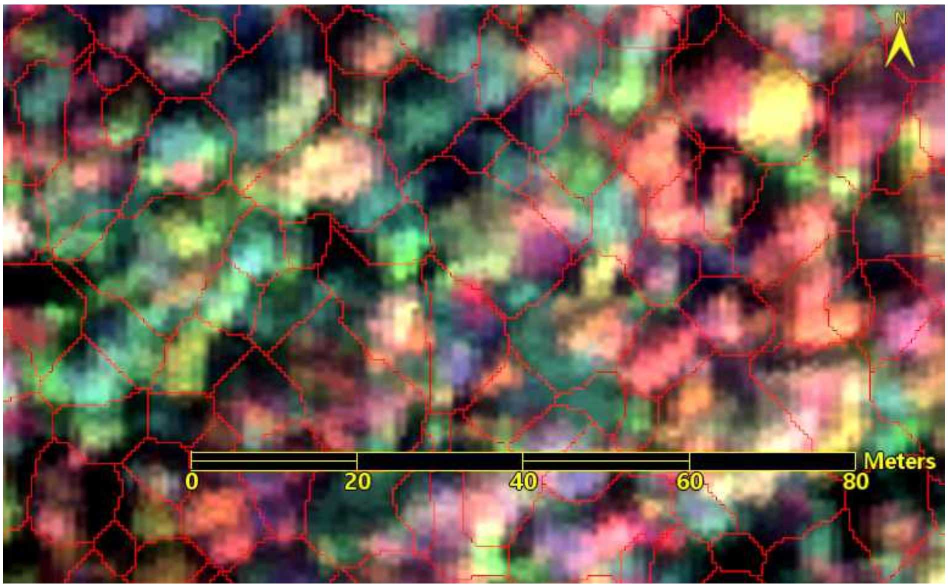


Fig. 3.10 Tree crown area extraction image using local maxima filtering

3.5. Feature Extraction

3.5.1. LiDAR Feature Extraction

The spatial resolution of the AHI and airborne LiDAR image metrics was set to 1 m resolution because of the noise impact and geometrical placement error caused by spatial resolution mismatch (Frair et al., 2010). Airborne LiDAR metrics convert a 3D LiDAR dataset into a 2D raster and incorporate it into a classification model as a variable (Davies & Asner, 2014; Simonson et al., 2014). R Software's stdmetrics and GreenValley International LiDAR 360's Forest Metrics were used to generate independent variables using the distribution and intensity characteristics of the point group data. Notably, the LiDAR metrics variable includes elevation metrics relating to height and density and intensity metrics variables capable of using different surface reflectance and properties for classification by utilizing return intensity—the amount of energy returned to the LiDAR sensor by laser pulses reflected on the target object. The standard deviation of the intensity and height of the LiDAR point were used, and the 9th quartile of 75th percentile reflectance and point cloud density were used. In addition, the canopy cover and leaf area index, which are mainly related to the vegetation area, were used as independent variables. Using the CloudCompare software's distance computation algorithm, an absolute distance variable representing the point change between leaf-on and leaf-off was derived by calculating the difference between the vertical and horizontal point changes.

3.5.2. Hyperspectral Imaging Feature Extraction

Several vegetation indices are based on the calculation of the near-infrared and red bands because vegetation reflectance differences in these bands occur mainly in the spectral reflectance of general leaves. In this research, the hyperspectral vegetation index was selected for vegetation classification by referencing previous studies. Using the photochemical reflectance index (PRI), which evaluates photosynthetic efficiency, carotenoid pigments in living leaves can be detected and plant photosynthetic function can be ascertained (Gamon, 1997). The red green ratio index (RGRI), like PRI, measures photosynthetic efficiency and is useful for estimating leaf stress and developmental processes in response to leaf red light caused by anthocyanins in chlorophyll (Gamon, 1999). The structure-insensitive pigment index is sensitive to detecting the ratio of carotenoid pigments to chlorophyll and can also detect increases in carotenoid-rich canopy stress (Penuelas, 1995). Furthermore, the carotenoid reflectance index 1 (CRI1), representing carotenoids in plant leaves, can detect stress-related carotenoid concentration. The plant senescence reflectance index is associated with plant senescence and detects carotenoid increases in chlorophyll (Merzlyak et al., 1999). The modified red edge simple ratio index is used for leaf reflection and stress detection using bands in the red edge area (Sims, 2002). The anthocyanin reflectance index 1 (ARI1) utilizes anthocyanin, which is abundant in leaves during the open and deciduous periods, and hyperspectral imaging can identify vegetation stress because it is expressed as a change in the anthocyanin pigment concentration (Gitelson, 2001). The modified chlorophyll absorption ratio index is used to indicate the relative degree of chlorophyll content (Daughtry et al., 2000). The Vogelmann red edge index 1 (VREI1) is sensitive to chlorophyll concentration, leaf area, and water content and utilizes the red region wavelength (Vogelmann, 1993).

3.5.3. Principal Component Analysis

In most cases, the entire wavelength band of the AHI data is not used due to limitations associated with large data dimensions. Misclassification is possible and significant variable discrimination is made more challenging when all bands are used in the species classification process. Therefore, limiting the dimension is essential to prevent data loss when removing redundant data. To reduce the high-dimensional dataset to highly correlated variables, a new variable creation process was performed through principal component analysis (PCA). PCA extracts orthogonal principal components by designating the axis with the largest variance as the first principal component and the axis with the second-largest variance as the second principal component. PCA results can be visualized and combined with other remote sensing data. To select independent variables for species classification, PCA was performed on the two periods' hyperspectral images using the PCA tool of ENVI 5.6.3. The PCA results of the 127 multidimensional bands of the September AHI dataset indicated that PC1 (99.5%), PC2 (0.3%), and PC3 (0.1%) had explanatory power for the variables of the entire dataset, and PC1 was used. For the November AHI dataset, the PCA results indicated PC1 (99.6%), PC2 (0.21%), and PC3 (0.08%). As a result, each of the 127 hyperspectral bands of the two periods was dimensionally reduced to the PC1 band through PCA and used as an independent variable for species classification.

3.5.4. Feature Selection

Table 3.7 lists the composition of the independent variables used in vegetation classification. A total of 29 variables, including AHI PC1 (2 variables), hyperspectral vegetation index (14 variables), and airborne LiDAR variables (13 variables), were used as independent variables for classification. In order to prevent the multicollinearity problem caused by a high correlation between independent variables, variables with a correlation coefficient of at least 0.8 between independent variables were considered to have a strong correlation and were excluded from being overlapped (Kim, 2019).

Table 3.7 Independent variable used for tree species classification.

Datasets	Independent variable	Source
Airborne LiDAR	Canopy Cover(CC) Nov, Apr	Jennings et al., 1999
	Leaf Area Index(LAI) Nov, Apr	Nilson, 1971
	75th percentile of return intensity(IP75) Nov, Apr	Liu, et al., 2017
	Standard deviation of return intensity(IS) Nov, Apr	Liu, et al., 2017
	Standard deviation of elevation(ES) Nov, Apr	Liu, et al., 2017
	Ninth decile of elevation density(ED9) Nov, Apr	Korpela et al., 2010
	Absolute distance(AD) Difference between Nov and Apr	Esposito et al., 2017

Airborne Hyper-spectral Imaging	First Principal Component (PC1)Sep, Nov	-
	Anthocyanin Reflectance Index 1(ARI1) Nov	Gitelson et al., 2001
	Carotenoid Reflectance Index1 Sep, Nov	Gitelson et al., 2002
	Modified Chlorophyll Absorption Ratio Index(MCARI) Nov	Daughtry et al., 2000
	Modified Red Edge Simple Ratio(MRESR) Sep	Sims et al., 2002
	Photochemical Reflectance Index(PRI) Sep, Nov	Penuelas et al., 2005
	Plant Senescence Reflectance Index(PSRI) Sep	Merzlyak et al., 1999
	Red Green Ratio Index(RGRI) Sep, Nov	Gamon et al., 1999
	Structure Insensitive Pigment Index(SIPI) Sep, Nov	Penuelas et al., 1995
	Vogelmann Red Edge Index 1(VREI1) Nov, Sep	Vogelmann et al., 1993

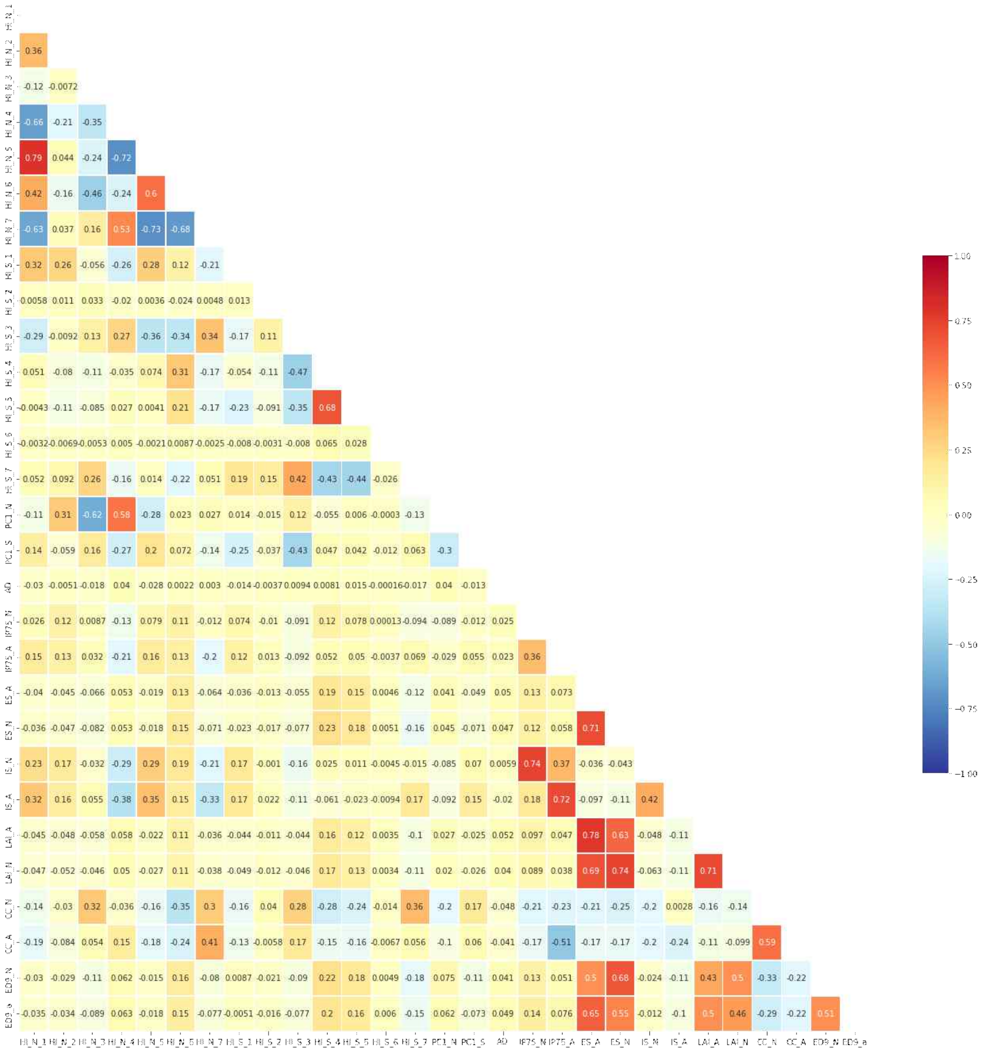


Fig. 3.11 The results of correlation analysis between independent variables

3.6. Classification Technique

3.6.1. Classification Technique

A 2.5 m buffer was set internally from the tree species boundary to prevent data collection errors due to the size of the raster pixel during random point sampling at this boundary. In addition, to prevent sampling in areas such as the forest's lower part and the gap in the forest, areas less than 2 m in height were excluded using the height layer extracted from the LiDAR data to conduct sampling in the canopy's upper part. To prevent overfitting and outlier data generation caused by the difference in the range and size of each variable in continuous data, a robust scaler was used and is shown in Equation (1). The robust scaler sets the median value (x_{med}) to 0 and makes the interquartile range (IQR, x_{75}-x_{25}), which is the difference between the 3rd quartile and the 1st quartile, to be 1 for data (x_i). This minimizes the influence of extreme values in continuous independent variables.

$$x' = \frac{x_i - x_{med}}{x_{75} - x_{25}} \tag{1}$$

Noting that 16,522 random points were sampled for each type to remove missing data, 165,216 points were generated for 10 types. For the training and evaluation of machine learning classifiers, 165,216 points were divided into a ratio of 80 to 20, and stratified sampling was performed to maintain the population data ratio and prevent training bias toward a specific tree species. For machine learning analysis, 5 classifiers, LR, SVM, DT, RF, and LGBM, were used, and each classifier was cross-validated 5 times in 10 layers through an optimized hyperparameter tuning process using Python's

GridSearchCV (Table 3.8).

Table 3.8 List of hyperparameter values of LR, SVM, DT, RF, and LGBM classifier.

Classifier	Hyperparameter	Tuned value	Defalut value
LR	solver	newton-cg	lbfgs
	C	0.01	1.0
	max_iter	1000	100
SVM	C	1.0	1.0
	gamma	0.1	scale
DT	max_depth	None	None
	min_samples_split	2	2
	criterion	gini	gini
RF	n_estimators	500	100
	max_depth	50	None
	oob_score	True	False
	random_state	42	None
LGBM	eval_metric	multi_logloss	None
	n_estimators	400	100

In K-fold cross-validation, after dividing the total dataset into k partitions, k−1 partitions are used for training, the remaining 1 partition is used as test data, and the value of the verified average is calculated as accuracy by k iterations (A. Ramzan et al., 2021).

The confusion matrix and Cohen's kappa index were used to evaluate the analysis results. The confusion matrix entails evaluating a trained classification model by comparing the actual class value to the predicted class value and determining the classified class accuracy by representing it as a matrix.

- True Positives (TP): Actual positive values are classified as positive
- False Positives (FP): Actual negative values are classified as positive
- False negative (FN): Actual positive values are classified as negative
- True Negative (TN): Actual negative values are classified as negative

		Actual Class	
		Positive	Negative
Predicted Class	Positive	TP (True Positive)	FP (False Positive)
	Negative	FN (False Negatvie)	TN (True Negatvie)

Fig. 3.12 Verification of classification accuracy using confusion matrix

Precision refers to the percentage of values calculated using Equation (2) and classified by the classification model as positive that are actually positive. Recall is the ratio of the value predicted by the model as positive among the actual positive values calculated using Equation (3). These two items were evaluated for each model type generated by the F1 score, which is Equation (4) for harmonic means, and the accuracy of each dataset and classifier was compared using Equation (5). Regarding the Cohen's kappa index, the tree classification accuracy can be evaluated by comparing the predicted and actual values and measuring the degree of agreement.

$$(Precision) = \frac{TP}{TP+FP} \tag{2}$$

$$(\mathrm{Recall}) = \frac{TP}{TP+FN} \tag{3}$$

$$(\text{F1-score}) = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

$$(\text{Accuracy}) = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

3.7. Estimating Aboveground Biomass

3.7.1. Estimating Aboveground Biomass

Because forest trees are determined by the forest-growing stock of the clinical classification unit of the forest type map, urban foresters use allometric equations to calculate biomass (Park, 2009). The biomass calculation method using allometric equations estimates the biomass and carbon storage amount by substituting the DBH obtained from sample surveys for each species into allometric equations.

In this research, the unit of analysis was the independent tree, and biomass was calculated using TH-DBH relational expression verified in previous studies utilizing tree species classification results and airborne LiDAR TH information. AGB was calculated by summing the biomass of stems, branches, and leaves.

The modified logistic (6) model explained the relationship between TH and DBH in forests with strong explanatory power (Ratkowsky and Reedy 1986, Huang et al., 2000) and was applied to TH after being converted to Equation (7).

$$H = 1.2 + \frac{a}{1 + b^{-D^{-c}}} \tag{6}$$

$$D = \left(b \left(\frac{a}{H - 1.2} - 1 \right) \right)^{-1/c} \tag{7}$$

$$y = aD^b H^c \tag{8}$$

H = Tree Height; D= Diameter at breast height; a~c= Model parameters

DBHs of Quercus mongolica, Pinus densiflora, Quercus acutissima, Pinus rigida, Quercus variabilis, Quercus serrata, and Larix kaempferi were derived using the parameter values of a–c of 7 major species of Seo Yeon-ok (2011). Robinia pseudoacacia, Castanea crenata, and Pinus koraiensis referred to the application criteria for tree volume, biomass, and stand yield table 2021 (KFS and NIFoS, 2021), and the parameters of Quercus acutissima, Quercus mongolica, and Pinus densiflora were set to be used (Robinia pseudoacacia → Quercus acutissima, Castanea crenata → Quercus mongolica, Pinus koraiensis → Pinus densiflora). Table 3.9 displays the values of a, b, and c used to calculate DBH based on the modified logistic growth model.

Table 3.9 Parameters used in TH-DBH relational expression by species.

Species	a	b	c
Quercus mongolica	15.3958	0.03	1.3391
Pinus densiflora	14.9639	0.0174	1.5132
Robinia pseudoacacia	14.263	0.0068	2.0675
Quercus acutissima	14.263	0.0068	2.0675
Pinus rigida	13.7491	0.0302	1.5015
Quercus variabilis	19.9118	0.029	1.2394
Castanea crenata	15.3958	0.03	1.3391
Pinus koraiensis	14.9639	0.0174	1.5132
Quercus serrata	15.9803	0.0295	1.3246
Larix kaempferi	20.5193	0.0016	2.4893

AGB was calculated by adding the biomass of stems, branches, and leaves by substituting the TH value of the individual tree derived by CHM segmentation and the DBH value estimated according to the tree volume, biomass, and stand yield table's allometric equations (8).

While DBHs smaller than the range of use of the equation were excluded in the calculation process, those exceeding the range of the equation were deemed maximum DBH. Trees not included in the standard yield table were replaced based on the current criteria for table application. The parameter values of Quercus acutissima for Robinia pseudoacacia and Quercus mongolica for Castanea crenata and Quercus serrata were used (Robinia pseudoacacia → Quercus acutissima, Castanea crenata → Quercus mongolica, Quercus serrata → Quercus mongolica). Concerning Pinus densiflora, allometric equations for each part of Pinus densiflora in the central region were used. Table 3.10 shows the values of a, b, and c substituted for allometric requirements by region for each tree type to calculate AGB.

Table 3.10 Parameters used to estimate aboveground biomass
(Tree volume, biomass, and stand yield table 2021. Seoul, Republic of Korea: KFS and NIFoS).

Species	Stems (a)	Stems (b)	Stems (c)	branches (a)	branches (b)	branches (c)	leaves (a)	leaves (b)	leaves (c)
Quercus mongolica	0.098	1.406	1.135	0.018	3.083	−0.493	0.023	2.609	−0.833
Pinus densiflora	0.034	1.734	1.025	0.008	3.586	−1.158	0.077	1.931	−0.566
*Robinia pseudoacacia**	0.008	2.334	1.069	0.012	2.853	0.006	0.008	2.518	−0.151
Quercus acutissima	0.008	2.334	1.069	0.012	2.853	0.006	0.008	2.518	−0.151
Pinus rigida	0.029	1.824	1.036	0.00002	2.632	2.058	0.053	1.82	−0.22
Quercus variabilis	0.053	1.81	0.881	0.082	2.553	−0.608	0.108	1.63	0.406
*Castanea crenata**	0.098	1.406	1.135	0.018	3.083	−0.493	0.023	2.609	−0.833
Pinus koraiensis	0.034	1.734	1.025	0.008	3.586	−1.158	0.077	1.931	−0.566
*Quercus serrata**	0.098	1.406	1.135	0.018	3.083	−0.493	0.023	2.609	−0.833
Larix kaempferi	0.005	2.458	0.904	0.143	4.482	−2.9	0.022	1.877	−0.023

*Parameters are replaced and applied according to the current inventory application standard

Chapter 4. Results

4.1. Tree Species Classification

4.1.1. Tree Species Classification

The tree classification results of the trained classifiers indicated that the average accuracy of multitemporal multidatasets was 71%, exceeding those of single temporal multidatasets (leaf-on:57%; leaf-off:61%) and multitemporal single datasets (AHI:64%; airborne LiDAR 55%). The combination of AHI and airborne LiDAR images during the leaf-off period showed about 4% higher accuracy than the combination of AHI and airborne LiDAR images during the leaf-on period. The comparison of the multitemporal single datasets revealed that the AHI dataset was approximately 9% more accurate than the airborne LiDAR dataset.

The comparison results of the five machine learning classifiers indicated that the RF algorithm had the highest average accuracy (76%), followed by LGBM (70%), DT (61%), SVM (60%), and LR (39%). The accuracy of classification using multitemporal multidatasets as RF algorithms was 83.5% (Kappa: 0.82), indicating the highest accuracy in all classifiers and combinations of datasets. Following RF, accuracy of multitemporal multidatasets was shown in the order of LGBM (81%), SVM (76%), DT (68%), and LR (47%). Fig. 4.1 depicts the classification result map of the research area based on the trained RF model.

Table 4.1 Comparison of classification accuracy by dataset and classifier.

Classifier	Dataset					Mean
	HN-HS	LN-LA	HS-LN	HN-LA	HN-HS-LN-LA	
Lr	0.43	0.30	0.34	0.40	0.47	0.39
DT	0.61	0.62	0.55	0.61	0.68	0.61
SVM	0.66	0.45	0.56	0.59	0.76	0.60
LGBM	0.72	0.64	0.66	0.68	0.81	0.70
RF	0.78	0.72	0.72	0.75	0.83	0.76
Mean	0.64	0.55	0.57	0.61	0.71	

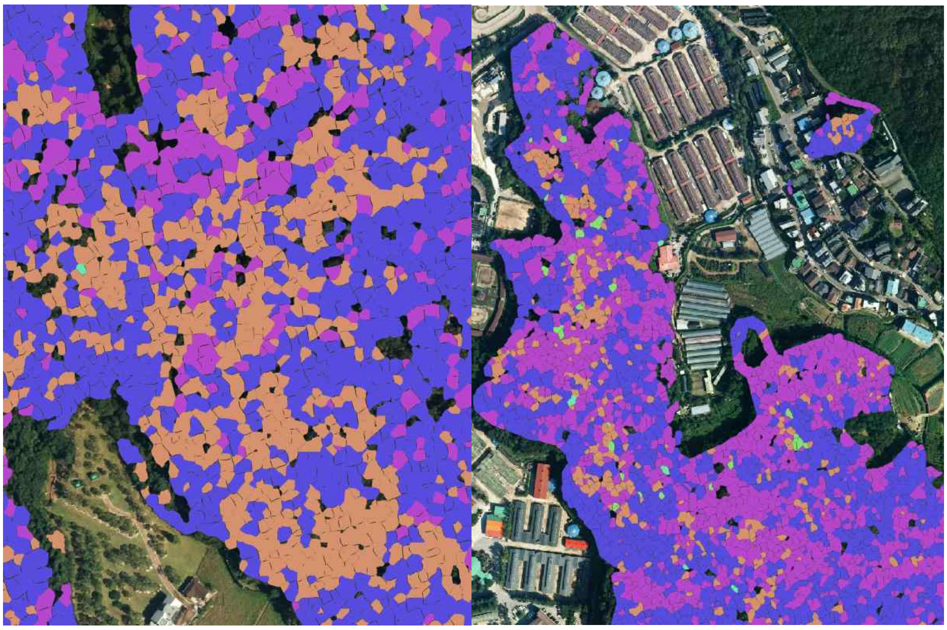


Fig. 4.1 The results of individual tree segmentation and species classification

The F1-score evaluation of classification accuracy by species revealed that Larix kaempferi (95%), Pinus koraiensis (88%), Quercus serrata (88%), Pinus densiflora (85%), Robinia (84%), Castanea crenata (82%), and Quercarius (82%) had high scores in that order. As shown in Table 4.2, when multitemporal multidatasets were classified as RF classifiers, both precision (min: 78%; max. 93%) and recall (min: 74%; max. 98%) showed over 70% classification accuracy for the 10 major species.

Table 4.2 Random forest classification confusion matrix (multitemporal multidataset).

Predicted label \ True label	a	b	c	d	e	f	g	h	i	j	Recall
a: Quercus mongolica	2529	54	90	83	72	212	50	8	157	48	77%
b: Pinus densiflora	53	2827	34	20	130	46	25	73	83	13	86%
c: Robinia pseudoacacia	60	29	2886	53	70	26	65	21	55	39	87%
d: Quercus acutissima	136	81	175	2459	109	68	68	84	50	74	74%
e: Pinus rigida	132	146	46	38	2521	95	51	201	24	51	76%
f: Quercus variabilis	162	55	104	87	129	2495	28	7	184	54	75%
g: Castanea crenata	100	26	139	89	67	137	2578	55	62	52	78%
h: Pinus koraiensis	14	26	41	25	130	11	68	2971	1	17	90%
I: Quercus serrata	27	102	22	5	5	26	7	1	3103	6	94%
j: Larix kaempferi	17	6	18	11	7	4	5	4	5	3228	98%
Precision	78%	84%	81%	86%	78%	80%	88%	87%	93%	90%	83.5%
Overall Accuracy: 83.5%, Kappa coefficient: 0.82											

Table 4.3 Random forest classification F1-score (multitemporal multidataset).

Species	F1-score	support
Quercus mongolica	0.78	3304
Pinus densiflora	0.85	3304
Robinia pseudoacacia	0.84	3304
Quercus acutissima	0.80	3304
Pinus rigida	0.77	3305
Quercus variabilis	0.78	3305
Castanea crenata	0.82	3305
Pinus koraiensis	0.88	3304
Quercus serrata	0.88	3304
Larix kaempferi	0.94	3305
Mean	0.83	33044

4.2. Important Independent Variables

4.2.1. Important Independent Variables

Analyzing the main variables affecting tree classification results using the Feature Importance function of the Python library showed that the carotenoid reflection index derived from AHI was the highest variable in the fall topology reason (Importance: 0.064). CRI during the full-leaf period in September was also identified as a highly significant variable among the hyperspectral image variables (Importance: 0.048). VREI, which utilizes the near-infrared wavelength bands of the September hyperspectral images, was also a highly significant variable (Importance: 0.045). In the hyperspectral image obtained in November, ARI (Importance: 0.044), photosynthesis-related index (Importance: 0.042), and RGRI (Importance: 0.042) had importance, as indicated. For airborne LiDAR images, the leaf area index (Importance: 0.062) and standard deviation of electricity (Importance: 0.049) obtained in April were also high. Notably, the above findings confirmed that, among the data extracted through interband calculations from the AHI dataset, the index associated with the pigment and photosynthesis performance of leaves and the properties associated with the vertical structure of airborne LiDAR generated by metrics calculations appeared to be major factors in determining overall accuracy. In addition, the 75th percentile intensity (leaf-on, leaf-off, Importance: 0.023), standard deviation of reflected intensity (leaf-on, leaf-off: Importance: 0.022, 0.221), canopy cover (leaf-off, Importance: 0.022), and the difference in distance between the points of the two periods (Importance: 0.017) were independent variables of low importance in species classification.

Table 4.4 Feature Importance (N: November; S: September; A: April).

Feature	Explanation	Feature Importance
CRI1_N	Carotenoid Reflectance Index 1_N	0.064
LAI_A	Leaf Area Index_A	0.062
ES_A	Standard deviation of elevation_A	0.049
CRI1_S	Carotenoid Reflectance Index 1_S	0.048
VREI1_S	Vogelmann Red Edge Index 1_S	0.045
ARI1_N	Anthocyanin Reflectance Index 1_N	0.044
ES_N	Standard deviation of elevation_N	0.043
PRI_N	Photochemical Reflectance Index_N	0.042
RGRI_N	Red Green Ratio Index_N	0.042
VREI1_N	Vogelmann Red Edge Index 1_N	0.041
SIPI_N	Structure Insensitive Pigment Index_N	0.039

LAI_N	Leaf Area Index__N	0.038
MRESR_S	Modified Red Edge Simple Ratio_S	0.036
PC1_S	First Principal Component_S	0.034
PC1_N	First Principal Component_N	0.032
MCARI_N	Modified Chlorophyll Absorption Ratio Index_N	0.031
CC_A	Canopy Cover_A	0.028
ED9_A	Ninth decile of elevation density_A	0.027
PSRI_S	Plant Senescence Reflectance Index_S	0.027
RGRI_S	Red Green Ratio Index_N_S	0.026
ED9_N	Ninth decile of elevation density_N	0.025
SIPI_S	Structure Insensitive Pigment Index_S	0.025
PRI_S	Photochemical Reflectance Index_S	0.024
IP75_A	75th percentile of return intensity_A	0.023
IP75_N	75th percentile of return intensity_N	0.023
IS_N	Standard deviation of return intensity_N	0.022
CC_N	Canopy cover_N	0.022
IS_A	Standard deviation of return intensity_A	0.021
AD	Absolute distance	0.017

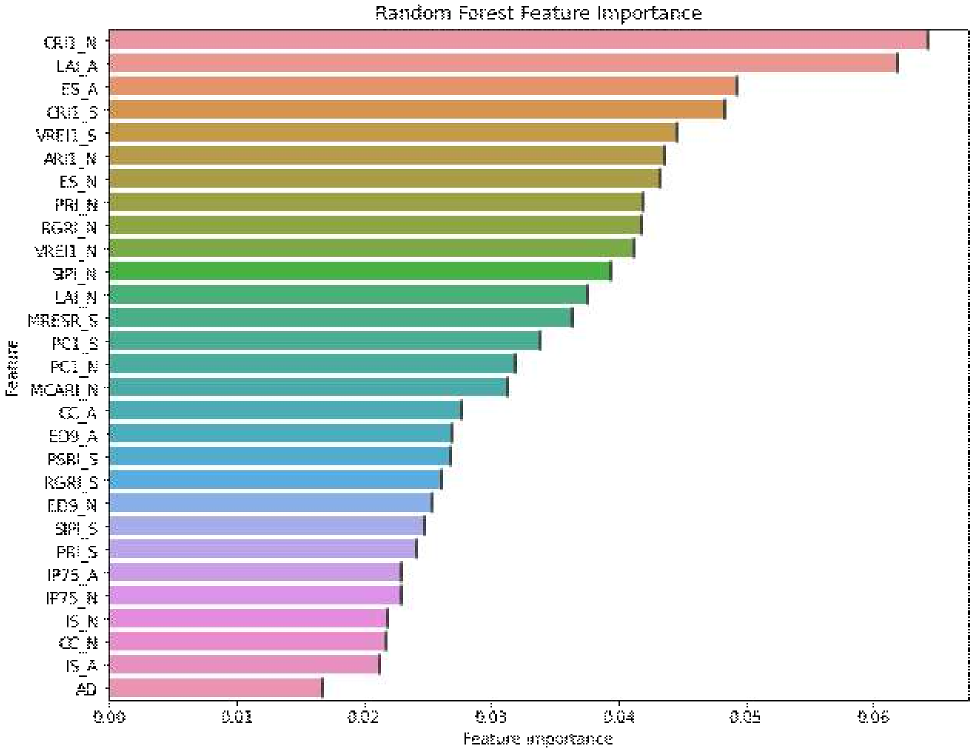


Fig. 4.2 Ranking of important independent variables

4.3. Estimating Aboveground Biomass

4.3.1. Estimating Individual Tree Height and DBH

The estimated individual tree through local maxima segmentation of the CHM model created by fusing LiDAR images of leaf-on and leaf-off was 928,015 trees. Individual TH values calculated by dividing the research area's central point into left and right halves revealed a significant height variation. Comparing 250 K geological maps (Korea Institute of Geoscience and Mineral Resources), most of the Gwanak Mountain area west of the research site comprised Granite, and trees with an average height of 5.14 m were distributed due to shallow soil depth and many cliffs and bedrock. In addition, trees with an average TH of 7.33 m were distributed in the eastern Cheonggyesan Mountain area, which comprises gneiss and is relatively deep in the weathered earth core. The difference in TH of the research area vegetation due to differences in topography and bedrock was revealed, and the average height of all research area trees was approximately 6.4 m.

Table 4.5 The statistic of individual tree height

	Mean	Std. dev.	Freq.
Left	5.14	2.24	393,211
Right	7.33	2.37	534,804
Total	6.40	2.56	928,015

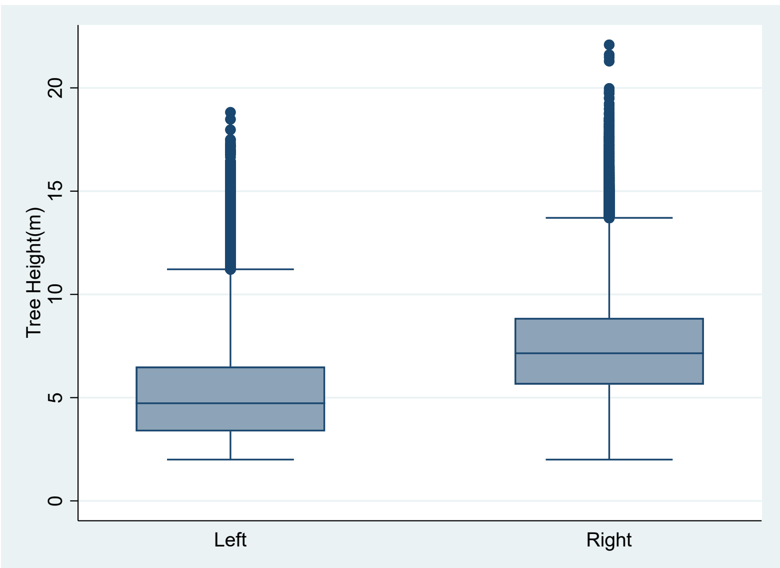


Fig. 4.3 Comparison of tree height in the research area

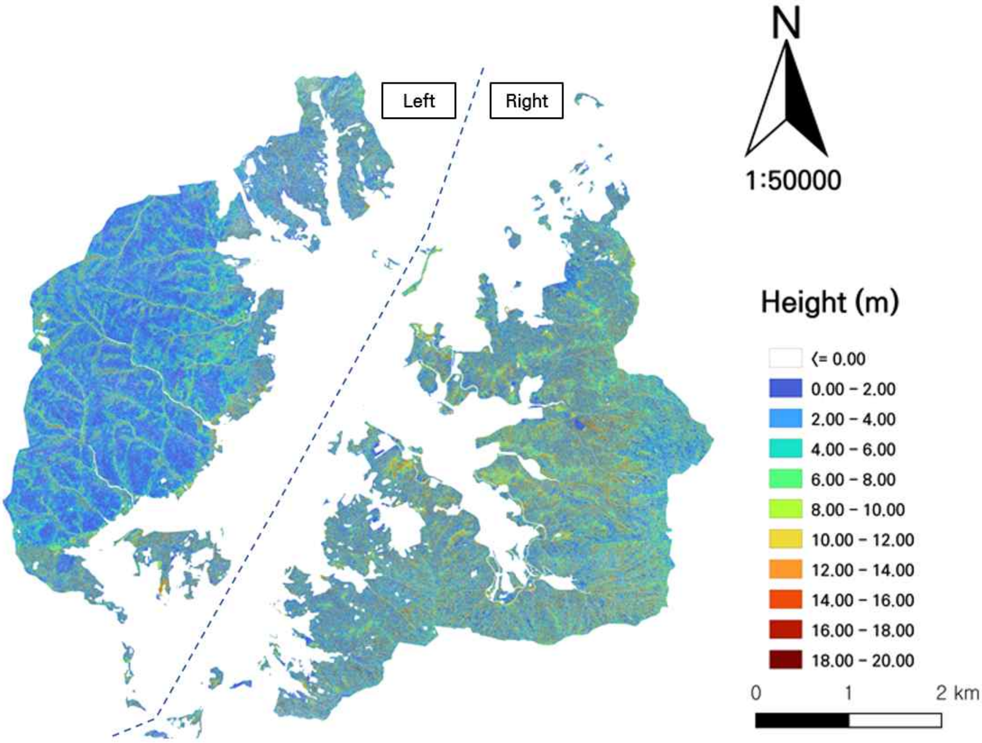


Fig. 4.4 Distribution of individual tree heights in the research area

4.3.2. Estimating Aboveground Biomass

Tree species values were assigned to individual tree crown area polygons based on the results of classifying multitemporal multidatasets using the RF algorithm. The DBH of the independent tree was derived using Logistic TH-DBH relational expression. Different a–c parameters were applied according to TH, DBH, and tree species to calculate the biomass of stems, branches, and leaves and AGB by adding them to the allometric equations of tree volume, biomass, and stand yield table. Stem biomass had an average of 29.6 kg, a standard deviation of 44.9 kg, a maximum of 587.9 kg, and a minimum of 2.7 kg, with a total of 27,504,963.4 kg (27,505 t). Branch biomass had an average of 16.7 kg, a standard deviation of 43.4 kg, a maximum of 431.6 kg, a minimum of 2.0 kg, and a total of 15,454,894.9 kg (15,455 t). Leaf biomass was calculated with an average of 2.6 kg, a standard deviation of 4.2 kg, a maximum of 39.5 kg, and a minimum of 0.6 kg, with a total of 2,390,777.8 kg (2391 t). The average AGB of stem, branch, and leaf biomass was 48.9 kg, with a standard deviation of 91.5 kg, a maximum of 954.5 kg, and a minimum of 7.3 kg. A total of 45,350,636 kg (45,351 t) of AGB was calculated for the research area. Fig. 4.5 shows the outcomes of mapping the research area biomass calculation result (Table 4.6) by combining the biomass result and attribute value of the indivisible tree crown area polygon. According to the AGB distribution at the research area, DBH calculated based on TH of trees was used as input values, so AGB was calculated to be higher in forests near Seoul Grand Park and Cheonggye Mountain in the east than forests in Gwanak Mountain in the west.

Table 4.6 Biomass calculation result for tree part

Biomass (kg)	Mean	Stdev	Max	Min	Sum
Stems	29.6	44.9	587.9	2.7	27,504,963.4
branch	16.7	43.4	431.6	2.0	15,454,894.9
leaves	2.6	4.2	39.5	0.6	2,390,777.8
Above ground	48.9	91.5	954.5	7.3	45,350,636.0

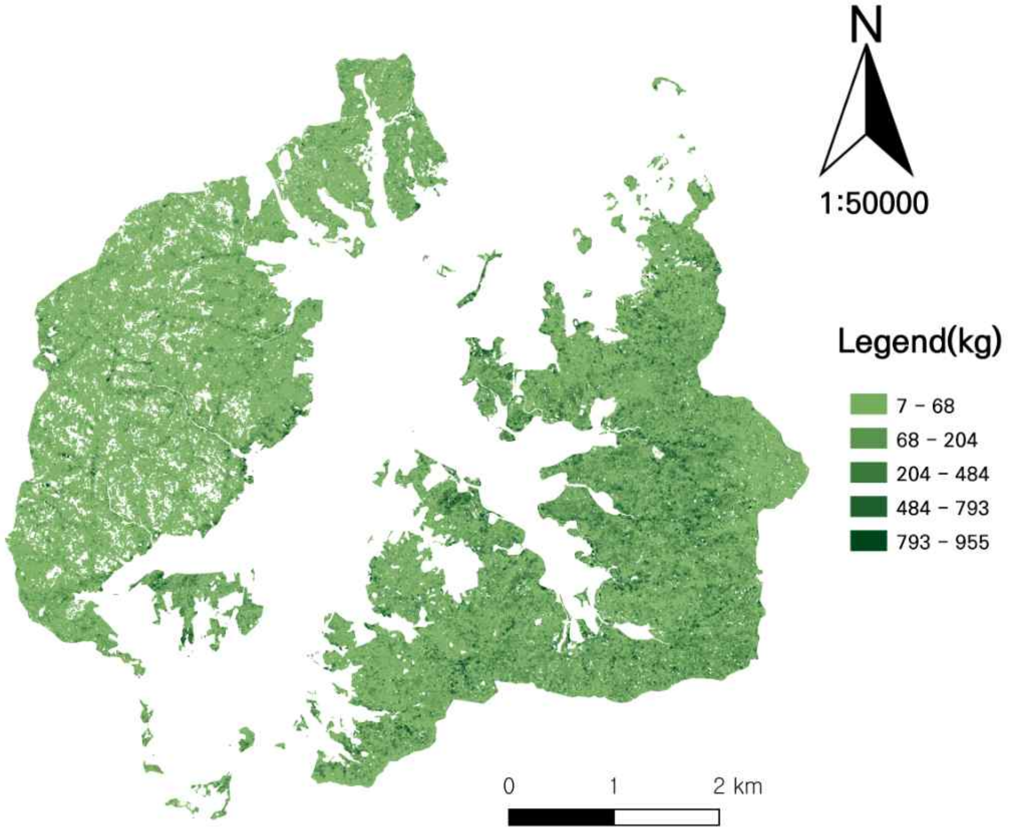


Fig. 4.5 The map of aboveground biomass estimation in the city of Gwacheon

Chapter 5. Discussion

5.1. Discussion

5.1.1. Tree Species Classification

This research suggests that classifying urban forest species by fusing AHI and airborne LiDAR data acquired over two different periods could be more accurate than traditional aerial image-based research methods. The average accuracy of the five machine learning classifiers in multitemporal multidatasets was 71%, exceeding those of single temporal multidatasets (leaf-on:57%; leaf-off:61%) and multitemporal single datasets (AHI:64%; airborne LiDAR 55%). Notably, employing the seasonal characteristics of forests can improve classification accuracy because the spectral reflectance of leaves and degree of leaf development in forests vary in seasons. In addition, combining data can improve classification results because two distinct sensors can collect a variety of attribute information for the target species. The canopy leaf surface reflectance is strongly prominent in AHI, whereas airborne LiDAR can confirm the tree's vertical structure and height-based volume. Datasets that fuse different periods and sensors may have high classification accuracy because they use the overall features of the tree as variables for tree classification. Therefore, compared to existing aerial or satellite images that classify coniferous and deciduous trees, the combination of different sensors in the studied two periods resulted in more than 80% classification accuracy for tree species.

The RF classification algorithm had the highest average accuracy (76%) because it uses ensemble methods that synthesize the classification results of multiple DTs to prevent overfitting, leading to high classification accuracy. In addition, RF makes non-relational predictions between trees through randomness and bagging during the selection of variables, hence reducing their sensitivity to the noise within the dataset. When segmenting nodes in the tree, 29 features derived from AHI and airborne LiDAR are randomly selected and optimal features are found, thus reducing bias while creating a suitable classification model. For this reason, high-dimensional datasets with 29 variables are thought to be classified with distinctly higher classification accuracy compared to other classifiers. The LGBM algorithm, which showed the second-highest classification accuracy, had an average classification accuracy of 70% because it finds the optimal classifier via error reduction of the previous model residuals by weighting the error-containing data while learning several weak learners sequentially. However, DT, SVM, and LR exhibited relatively low accuracy in classifying urban forest species using many independent variables owing to their limitations in reducing errors and calculating optimal models when learning and classifying high-dimensional datasets using various features of AHI and airborne LiDAR.

5.1.2. Important Independent Variables

The classification of the multitemporal multidataset using the RF classifier indicated that variables with important effects on tree classification did not appear to have a bias toward one sensor at a time, suggesting that some important independent variables may contribute to improving the accuracy of tree classification by period and sensor. Based on the analysis that the CRI of the hyperspectral image acquired in the fall foliage season was analyzed as the most important independent variable, the temporal difference and spectral reflectance difference of fall foliage increased by tree species in the corresponding period and can be detected through AHI. Fig. 5.1, which was captured from the ground in November when the AHI data were acquired, shows that the fall foliage season differs for each tree species and that it is relatively easier to distinguish the boundaries of deciduous trees during the leaf-off period. In the fall foliage season, the ARI also recorded a high importance of 0.044, and the classification result of the PRI showed a high contribution of 0.042, suggesting that species classification accuracy can be improved by utilizing the timing of chlorophyll destruction and anthocyanin synthesis. The CRI in the leaf-on period recorded a high importance of 0.045, suggesting that carotenoid concentration differs between species in both the fall foliage season and leaf growth period.

Fig. 5.1 Ground field image at the time of acquisition of Airborne Hyperspectral Imaging (2021.11.07.) (Left: Munwon Children's Park, Right: Airdrie Park)

Notably, VREI using the wavelength band of the red edge area could improve classification accuracy in all periods of November and September, suggesting that this can be used as an independent variable for classification by utilizing the difference in spectral reflectance for each species in the red edge area, which is a section where the spectral reflectance of vegetation increases rapidly in the infrared area.

The leaf area index obtained from an airborne LiDAR image showed a difference between April's leaf-off period (0.062) and November's leaf-on period (0.038). During the leaf-on period, which is the full-leaf period, the vegetation density increased, resulting in many overlaps at the boundary and making it difficult to distinguish between tree species boundaries. In addition, the high importance of the standard deviation variable in the height metrics of LiDAR points (leaf-off: 0.049; leaf-on: 0.043) may have affected the classification because the height of the point data for each species differed. Meanwhile, intensity is expected to act as an important variable in the classification process because it contains information on the difference in light scattering according to leaf surface conditions. However, in this research, the difference in intensity of laser pulses by season and aircraft course was reflected in the data acquisition process, indicating different attribute information for each data sampling point and low importance in the classification process.

Chapter 6. Conclusion

6.1. Overall Summary

6.1.1. Overall Summary

This research targeted the temperate urban forest and generated five datasets by dividing the AHI and airborne LiDAR images obtained during the two periods by period and sensor. Through interband and metrics calculations, an independent variable containing 29 species for tree classification was derived, and machine learning classification was performed for 10 representative species using independent and dependent variable values for 165,216 points by sampling 16,522 random points for each species. As a result, the average accuracy of the five classifiers of the multitemporal multidataset was the highest (71%), and the RF classifier had the highest average accuracy of the datasets for the five classifiers (76%). When classifying multitemporal multidatasets with RF, accuracy was highest (accuracy: 83.3%; Kappa: 0.82). Using the model trained with the RF classifier, the multitemporal multidataset with the highest classification accuracy was classified, and the TH-DBH relational equation and biomass calculation formula were applied to calculate 45,351 t of AGB in 928,015 tree crown areas at the research

6.2. Implications and Limitations

6.2.1. Implications of Research

This research is significant in that it quantified the classification of species and biomass for calculating carbon storage to cope with climate change policies and classified species based on machine learning analysis using the advantages of high-resolution hyperspectral and LiDAR images as an advanced method of traditional investigation. Although the spatial resolution of previous spectral satellite images was limited to 10–30 m and the number of available bands was limited to 10, this research used high-resolution image data with 1 m spatial resolution and 127 hyperspectral images to reflect the three-dimensional forest characteristics of the research area by combining airborne LiDAR and two-dimensional plane data. This research is expected to help preserve urban forests and manage climate change-vulnerable species, as it classifies urban forest species with over 80% accuracy. In addition, sampling was conducted based on forest field survey data to enable a more detailed tree boundary classification than the existing 1:5000 scale forest type map, and classification was performed in urban forests as opposed to test bed research area. The results of the area tree species classification are unique because the crown region was derived and analyzed as a vector polygon region, as opposed to the raster pixel format used in previous studies. In the individual tree crown area, the independent variable value was input as a median value by zone statistics to reduce noise and misclassification caused by fine pixels, thereby improving classification accuracy, and the highest point height of the TH was input as an attribute value to derive DBH. Carbon absorption in urban forests can be quantified in response to climate change policies using AGB, making it possible to establish and apply urban ecological status for sustainable urban development and urban management. The importance of the independent variables contributing to the classification can be calculated by extracting vegetation indices and

PCA derived from high-resolution hyperspectral images along with independent variables that reflect the situation at the area calculated by height and intensity metrics.

In addition, establishing DBH using hyperspectral and LiDAR datasets in the future and obtaining the corresponding time series data from various periods would contribute to quantifying the urban forest carbon cycle by measuring the growth and growth rate of major urban forest species and calculating their biomass.

6.2.2. Limitations of Research

Certain species exhibited lower categorization accuracy, which is a limitation of this research. Owing to vegetation production at various layers in actual forests, object classification accuracy at the boundary was deemed to be quite low due to overlap with adjacent trees. Although the vegetation area with at least 2 m TH was selected for tree species classification, this has a limitation because the case of vegetation that does not grow high due to bedrock and soil characteristics is not reflected. Furthermore, because data classification accuracy verification was conducted as a confusion matrix, comparing the tree heights derived from airborne LiDAR, the location of the tree crown area, and the classification results to the actual site is crucial. CHM was extracted from the airborne LiDAR dataset, and local maximum filtering was performed based on the height point to set the crown area because various vegetations are mixed in actual temperate forests. For horizontally growing trees, accurately calculating the tree crown area was difficult; hence, the research may result in an underestimation or overestimation of the actual one. DBH was derived from TH-DBH relational expression, and the results were obtained using the DBH-biomass calculation formula used in previous studies; hence, it may differ from the actual field when calculating DBH through actual tree height. Although a verified model was used, a re-verification should be undertaken because the verification comparing the actual biomass amount to the calculated biomass amount at this research area was omitted. To accurately calculate forest biomass, additional research is required to establish a regression equation between two variables that can estimate DBH for each species in the forest using tree heights derived from airborne LiDAR for various species. In addition, when defining the remote sensing image acquired in November, the timing was not unified in that the November image was regarded as leaf-on in AHI and leaf-off in airborne LiDAR. However, the November dataset was advantageous since it could be defined based on specific situations because it was acquired when both leaves and fall foliage existed. Regarding domestic forests, obtaining AHI and airborne LiDAR images in August–September was difficult due to the influence of the continuous rainy season and typhoons, suggesting that the November datasets can be used more for classification. In the future, improved accuracy for tree species classification can be realized by developing parameters that classify domestic forests with a large number of deciduous and coniferous trees based on their structural forms and by applying classification algorithms. Notably, if a vegetation index is developed using airborne LiDAR–hyperspectral images, forest monitoring research employing remote sensing data will be actively conducted in urban forest research.

References

▌ Domestic Literature

(1) Report

Gwacheon city, (2019), 2020~2024 Climate Change Adaptation Second Action Plan

Korea Forest Service, (2021), 2020 Forest Basic Statistics

Ministry of Land, Infrastructure and Transport (South Korea). (2021). Statistics of Urban Planning

Ministry of Security and Public Administration. (2022.10) Resident Registry Population. https://jumin.mois.go.kr/

Park Eun-jin, & Jwa Seung-hee. (2009). Quantification of CO2 Uptake by Urban Trees and Greenspace Management for C Sequestration. Gyeonggi Research Institute, 1-1.

Son, Y. M., Kim, R. H., Lee, K. H., Pyo, J. K., Kim, S. W., Hwang, J. S., ... & Park, H. (2014). Carbon emission factors and biomass allometric equations by species in Korea. Korea Forest Research Institute.

(2) Research Paper

An-Jin chang, & Hyung-Tae Kim. (2008). Study of Biomass Estimation in Forest by Aerial Photograph and LiDAR Data. Journal of the Korean Association of Geographic Information Studies, 11(3), 166-173

Cho, H. K., & Shin, M. Y. (2009). Estimation of Forest Biomass based upon Satellite Data and National Forest Inventory Data. Korean Journal of Remore Sensing, 25(4), 311-320.

Eun Jin Park, & Kyu Yi Kang. (2010). Estimation of C Storage and Annual CO2 Uptake by Street Trees in Gyeonggi-do. Korean Journal of Environment and Ecology, 24(5), 591-600.

Hyunggab Cho, & Kyu-Sung Lee. (2014). Comparison between Hyperspectral and Multispectral Images for the Classification of Coniferous Species. Korean Journal of Remote Sensing, 30(1), 25-36.

Jo, Hyun-Kil, & Ahn, Tae-Won. (2012). Carbon Storage and Uptake by Deciduous Tree Species for Urban Landscape. Journal of the Korean Institute of Landscape Architecture, 40(5), 160-168.

Jo, Hyun-Kil. (1999). Carbon Uptake and Emissions in Urban Landscape, and the Role of Urban Greenspace for several Cities in Kangwon Province. Journal of Korean institute of landscape architecture, 27(1), 39-53.

Joon Hyung Park, Su Young Jung, Kwang Soo Lee, Chang Hwan Kim, Yong Bae Park, & Byung Oh Yoo. (2016). Classification of Regional Types for Pinus densiflora stands Using Height-DBH Growth in Korea. Journal of Korean Society of Forest Science, 105(3), 336-341.

Lee, D. G., Yun, E. J., Kim, E. Y., & Park, E. J. (2006). Gwacheon City Street Tree Status Analysis and Maintenance Plan. In Proceedings of the Korean Institute of Landscape Architecture Conference (pp. 66-69). The Korean Institute of Landscape Architecture.

Lee, Hyun Jik, & Ru, Ji Ho. (2012). Application of LiDAR Data &

High-Resolution Satellite Image for Calculate Forest Biomass. Journal of Korean Society for Geospatial Information Science, 20(1), 53-63.

Lim, Ye Seul, Eo, Yang Dam, Jeon, Min Cheol, Lee, Mi Hee, & Pyeon, Mu Wook. (2016). Experiments of Individual Tree and Crown Width Extraction by Band Combination Using Monthly Drone Images. Journal of Korean Society for Geospatial Information Science, 24(4), 67-74.

Park, Jung-Seo., Seo, Jin-Jae & Go, Je-Woong. (2016). The Study on Improving Accuracy of Land Cover Classification using Spectral Library of Hyperspectral Image. Journal of Cadastre & Land InformatiX, 46(2), 239-251.

Seo Yeon Ok, Young Jin Lee, Dae-Kyun Nor, Kim Sung Ho, Choi Jung Kee, & LEE， Woo-Kyun. (2011). Height-DBH Growth Models of Major Tree Species in Chungcheong Province. Journal of Korean Society of Forest Science, 100(1), 62-69.

Soo-Kyung Kwon, ·Kyoung-Min Kim, & Joongbin Lim. (2021). A Study on Pre-evaluation of Tree Species Classification Possibility of CAS500-4 Using RapidEye Satellite Imageries. Korean Journal of Remote Sensing, 37(2), 291-304.

(3) Thesis

Choi, Heejoon. (2018). Estimation of the Urban Forest Growth Using Multi-temporal LiDAR Dataset (Doctoral dissertation, Department of Landscape Architecture Graduate School of Environmental Studies Seoul National University)

(4) Other Materials

KFS (Korea Forest Service), NIFoS (National Institute of Forest Science). 2021. Tree volume, biomass, and stand yield table 2021. Seoul, Republic of Korea: KFS and NIFoS (in Korean)

Korea Forest Service: forest type map(1:5000)

Korea Institute of Geoscience and Mineral Resources, Geo Big Data Open Platform, 250K geological maps

National Assembly of the Republic of Korea, 2020, Creation And Management Of Urban Forest Act(Act No. 17420, jun. 9, 2020)

▌International Literature

(1) Research Paper

A. Ramezan, C., A. Warner, T., & E. Maxwell, A. (2019). Evaluation of sampling and cross-validation tuning strategies for regional-scale machine learning classification. Remote Sensing, 11(2), 185.

Alonzo, M., Bookhagen, B., & Roberts, D. A. (2014). Urban tree species mapping using hyperspectral and lidar data fusion. Remote Sensing of Environment, 148, 70-83.

Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.

Cao, Q. V. (2004). Predicting parameters of a Weibull function for modeling diameter distribution. Forest science, 50(5), 682-685.

Daughtry, C. S., Walthall, C. L., Kim, M. S., De Colstoun, E. B., & McMurtrey

Iii, J. E. (2000). Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. Remote sensing of Environment, 74(2), 229-239.

De Almeida, C. T., Galvao, L. S., Ometto, J. P. H. B., Jacon, A. D., de Souza Pereira, F. R., Sato, L. Y., ... & Longo, M. (2019). Combining LiDAR and hyperspectral data for aboveground biomass modeling in the Brazilian Amazon using different regression algorithms. Remote Sensing of Environment, 232, 111323.

De Almeida, D. R. A., Broadbent, E. N., Ferreira, M. P., Meli, P., Zambrano, A. M. A., Gorgens, E. B., ... & Brancalion, P. H. (2021). Monitoring restored tropical forest diversity and structure through UAV-borne hyperspectral and lidar fusion. Remote Sensing of Environment, 264, 112582.

Degerickx, J., Roberts, D. A., McFadden, J. P., Hermy, M., & Somers, B. (2018). Urban tree health assessment using airborne hyperspectral and LiDAR imagery. International journal of applied earth observation and geoinformation, 73, 26-38.

Din, M., Zheng, W., Rashid, M., Wang, S., & Shi, Z. (2017). Evaluating hyperspectral vegetation indices for leaf area index estimation of Oryza sativa L. at diverse phenological stages. Frontiers in plant science, 8, 820.

Englhart, S., Jubanski, J., & Siegert, F. (2013). Quantifying dynamics in tropical peat swamp forest biomass with multi-temporal LiDAR datasets. Remote sensing, 5(5), 2368-2388.

Esposito, G., Salvini, R., Matano, F., Sacchi, M., Danzi, M., Somma, R., & Troise, C. (2017). Multitemporal monitoring of a coastal landslide through SfM-derived point cloud comparison. The Photogrammetric Record, 32(160), 459-479.

Gamon, J. A., & Surfus, J. S. (1999). Assessing leaf pigment content and activity with a reflectometer. The New Phytologist, 143(1), 105-117.

Gamon, J., Serrano, L., & Surfus, J. S. (1997). The photochemical reflectance index: an optical indicator of photosynthetic radiation use efficiency across species, functional types, and nutrient levels. Oecologia, 112(4), 492-501.

Gitelson, A. A., Merzlyak, M. N., & Chivkunova, O. B. (2001). Optical properties and nondestructive estimation of anthocyanin content in plant leaves¶. Photochemistry and photobiology, 74(1), 38-45.

Gitelson, A. A., Zur, Y., Chivkunova, O. B., & Merzlyak, M. N. (2002). Assessing carotenoid content in plant leaves with reflectance spectroscopy¶. Photochemistry and photobiology, 75(3), 272-281.

Huang, S., Price, D., & Titus, S. J. (2000). Development of ecoregion-based height–diameter models for white spruce in boreal forests. Forest ecology and management, 129(1-3), 125-141.

Jennings, S. B., Brown, N. D., & Sheil, D. (1999). Assessing forest canopies and understorey illumination: canopy closure, canopy cover and other measures. Forestry: An International Journal of Forest Research, 72(1), 59-74.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. Advances in neural information processing systems, 30.

Kim, J. H. (2019). Multicollinearity and misleading statistical results. Korean journal of anesthesiology, 72(6), 558-569.

Knapp, N., Fischer, R., Cazcarra-Bes, V., & Huth, A. (2020). Structure metrics to generalize biomass estimation from lidar across forest types from different continents. Remote Sensing of Environment, 237, 111597.

Koch, B. (2010). Status and future of laser scanning, synthetic aperture radar and hyperspectral remote sensing data for forest biomass assessment. ISPRS Journal of Photogrammetry and Remote sensing, 65(6), 581-590.

Korpela, I., Ørka, H. O., Maltamo, M., Tokola, T., & Hyyppä, J. (2010). Tree species classification using airborne LiDAR–effects of stand and tree parameters, downsizing of training set, intensity normalization, and sensor type. Silva Fennica, 44(2), 319-339.

Kotsiantis, S. B. (2013). Decision trees: a recent overview. Artificial Intelligence Review, 39(4), 261-283.

Kwak, C., & Clayton-Matthews, A. (2002). Multinomial logistic regression. Nursing research, 51(6), 404-410.

L., Coops, N. C., Aven, N. W., & Pang, Y. (2017). Mapping urban tree species using integrated airborne hyperspectral and LiDAR remote sensing data. Remote Sensing of Environment, 200, 170-182.

Li, Q., Wong, F. K. K., & Fung, T. (2021). Mapping multi-layered mangroves from multispectral, hyperspectral, and LiDAR data. Remote Sensing of Environment, 258, 112403.

Mayra, J., Keski-Saari, S., Kivinen, S., Tanhuanpaa, T., Hurskainen, P., Kullberg, P., ... & Vihervaara, P. (2021). Tree species classification from airborne hyperspectral and LiDAR data using 3D convolutional neural networks. Remote Sensing of Environment, 256, 112322.

Merzlyak, J., et al. “Non-destructive Optical Detection of Pigment Changes During Leaf Senescence and Fruit Ripening.“ Physiologia Plantarum 106 (1999): 135-141.

Merzlyak, M. N., Gitelson, A. A., Chivkunova, O. B., & Rakitin, V. Y. (1999). Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening. Physiologia plantarum, 106(1), 135-141.

N. M. Singer and V. K. Asari, “DALES Objects: A Large Scale Benchmark Dataset for Instance Segmentation in Aerial Lidar,“ in IEEE Access, vol. 9, pp. 97495-97504, 2021, doi: 10.1109/ACCESS.2021.3094127.

Nilson, T. (1971). A theoretical analysis of the frequency of gaps in plant stands. Agricultural meteorology, 8, 25-38.

Penuelas, J., Filella, I., & Gamon, J. A. (1995). Assessment of photosynthetic radiation?use efficiency with spectral reflectance. New Phytologist, 131(3), 291-296.

Prins, A. J., & Van Niekerk, A. (2021). Crop type mapping using LiDAR, Sentinel-2 and aerial imagery with machine learning algorithms. Geo-spatial Information Science, 24(2), 215-227.

Ratkowsky, D. A., & Reedy, T. J. (1986). Choosing near-linear parameters in the four-parameter logistic model for radioligand and related assays. Biometrics, 575-582.

Sims, D. A., & Gamon, J. A. (2002). Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. Remote sensing of environment, 81(2-3), 337-354.

Song, Y., & Ryu, Y. (2015). Seasonal changes in vertical canopy structure in a temperate broadleaved forest in Korea. Ecological research, 30(5), 821-831.

Tong, S., & Chang, E. (2001, October). Support vector machine active learning for image retrieval. In Proceedings of the ninth ACM international conference on Multimedia (pp. 107-118).

Vogelmann, J. E., Rock, B. N., & Moss, D. M. (1993). Red edge spectral measurements from sugar maple leaves. TitleREMOTE SENSING, 14(8), 1563-1575.

Weinstein, B. G., Marconi, S., Aubry-Kientz, M., Vincent, G., Senyondo, H., & White, E. P. (2020). DeepForest: A Python package for RGB deep learning tree crown delineation. Methods in Ecology and Evolution, 11(12), 1743-1751.

Yan, P., Han, Q., Feng, Y., & Kang, S. (2022). Estimating LAI for Cotton Using Multisource UAV Data and a Modified Universal Model. Remote Sensing, 14(17), 4272.

Zhu, X., Skidmore, A. K., Wang, T., Liu, J., Darvishzadeh, R., Shi, Y., ... & Heurich, M. (2018). Improving leaf area index (LAI) estimation by correcting for clumping and woody effects using terrestrial laser scanning. Agricultural and forest meteorology, 263, 276-286.

(2) Book

Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). The elements of statistical learning: data mining, inference, and prediction (Vol. 2, pp. 1-758). New York: springer.

초록

다중시기 항공 LiDAR와 초분광 영상을 활용한 도시림 수종 분류 및 바이오매스 추정

김 대 열
서울대학교 환경대학원 환경조경학과

기후변화가 전 세계적 관심사로 부각되고 도시에 거주하는 인구의 비율이 증가함에 따라 대기오염과 열섬 현상을 완화하고 바이오매스 생성, 생물다양성 보존, 탄소 저장 등 다양한 편익을 제공하는 공간으로서 도시림의 중요성이 증가하고 있다. 산림을 구성하는 수종에 따라서 바이오매스 산정량에 따른 탄소 흡수량과 축적량이 다르기 때문에 도시림이 제공하는 편익을 정량적으로 계산하고 기후변화 취약종을 관리하기 위해서는 정확한 수종 분류가 필요하다. 전통적인 산림 모니터링의 경우 산림청에서 항공 영상을 이용한 판독과 현장 조사를 통해 임상도를 제작하여 관리하고 있지만 많은 노동력과 시간이 필요하고 항공 사진으로는 도시림 식생의 수직구조를 파악할 수 없기 때문에 대상지에서 생장하는 수종을 분류하고 경계를 정확하게 구분하는 방법이 요구되고 있다. 선행연구에 따르면 항공 LiDAR에서 파생된 산림구조 특성과 초분광영상의 분광 반사율을 이용하는 효과적인 산림 모니터링 연구가 많이 진행되고 있다. 최근에는 측량 기술의 발달로 인해 고밀도의(10 point/m^2) LiDAR 점군 데이터를 획득이 가능하게 되었고 오픈소스 소프트웨어서도 점군 데이터의 활용이 용이하게 되었으며 초분광 영상의 경우 다중 분광 영상에 비해 확대된 식생지수 목록과 전처리 및 보정 알고리즘 등이 개발되었다.

본 연구에서는 전통적인 산림조사 방법을 개선하기 위해 두시기의 초분광 영상과 항공 LiDAR 영상을 결합하여 각 자료가 가지는 특징과 식생의 계절적 특성 변화를 활용하여 수종 분류의 정확도와 효율성을 높이고 환경 계획에 활용 가능한 지도를 제작하여 도시림의 수종 분포를 파악하고자 하였으며 최종 분류 결과를 기반으로 대상지의 지상부 바이오매스를 산정하는 것을 목표로 하였다. 대상지는 북위 37° 23' ~ 37° 27', 동경 126° 57' ~ 127° 02'의 경기도 과천시 도시림으로 면적은 2,034 ha이고 10종의 주요 수종이 성립하고 있다. 분류를 위한 현장 조사 자료는 8월부터 10월에 취득된 산림조사결과 데이터를 사용하였고, 항공 LiDAR는 Leaf-on (11월), Leaf-off (4월) 시기에 취득되었고 항공 초분광의 경우 Leaf-on (9월), Leaf-off (11월)에 취득된 데이터셋을 사용하였다.

항공 LiDAR와 항공 초분광 데이터셋은 전처리 과정을 통해 보정되었으며 도시림의 식생 영역을 대상으로 항공 초분광 영상의 PC1 밴드, 잎의 색소 및 광합성 특성과 관련된 식생지수와 항공 LiDAR 영상의 높이, 반사강도 메트릭스 계산을 통해 수종 분류를 위한 독립변수 29개를 추출하였다. 대상지의 대표 수종 10종을 대상으로 16,522개의 랜덤 포인트를 생성하여 결측치를 제외한 총 165,216개의 데이터셋을 생성하였고 현장 조사에서 획득된 수종 정보를 기반으로 로지스틱 회귀 (LR), 서포트 벡터 머신 (SVM), 의사결정나무 (DT), 랜덤포레스트 (RF), Light Gradient Boosting Machine (LGBM)의 5개의 머신러닝 분류 모델을 학습하여 분류와 검증을 수행하였다.

머신러닝 학습을 통한 분류 결과 다중시기 다중 데이터셋의 5개 분류기 평균 정확도는 71%으로 단일시기 다중 데이터셋 (leaf-on: 57%; leaf-off: 61%)와 다중시기 단일 데이터셋 (항공 초분광: 64%; 항공 LiDAR: 55%)에 비해 높게 나타났다. 5개의 데이터셋의 머신러닝 분류기별 정확도 비교 결과 RF의 평균 정확도는 76%으로 LGBM (70%), DT (61%), SVM (60%), LR (39%)에 비해 높게 나타났다. 결과적으로 다중시기 다중 데이터셋을 RF 기법을 이용한 분류의 정

확도가 83.3%로 (Kappa: 0.80) 가장 높은 것으로 나타났다. 수종 분류에 기여하는 주요 독립변수는 11월에 취득된 항공 초분광 데이터셋에서 추출된 Carotenoid 반사 지수 (Importance: 0.064)와 4월에 취득된 항공 LiDAR 영상의 엽면적 지수 (Importance: 0.062)로 추정되었다.

개별 수목 추출 알고리즘을 통해 추출한 928,015개의 수관 영역을 Modified Logistic 수고-흉고직경 관계식을 사용하여 개체목의 흉고직경을 도출하고 수고 및 흉고직경을 입목재적·바이오매스 및 임분수확표의 부위별 상대생장식에 대입하여 2 m 이상의 교목을 대상으로 지상부 바이오매스를 도출한 결과 총 45,351 t의 바이오매스를 산정하였다.

항공 LiDAR와 항공 초분광을 활용한 수종 분류는 시 단위의 도시림에서 80% 이상의 정확도로 수종 분류를 수행할 수 있음을 시사하였으며 단일 시기에 촬영된 영상에 비해 잎의 생장 시기, 갈변 시기, 낙엽 시기에 따라 촬영된 영상을 결합하여 실제 산림의 계절적 특징을 반영했을 때 분류 정확도가 증가하는 것이 뚜렷하게 나타났다. 분류 결과를 시각화한 수종 지도를 토대로 기후변화 취약종을 관리하고 지상부 바이오매스를 산정하여 탄소 흡수량과 저장량을 추정하는 연구에 기여할 수 있을 것으로 사료된다.

주요어 : 항공 LiDAR, 항공 초분광, 도시림, 수종 분류, 지상부 바이오매스, 다중시기

학　번 : 2021-22194