remote sensing

MDPI

Article

From Coarse to Crisp: Enhancing Tree Species Maps with Deep Learning and Satellite Imagery

Taebin Choe[1], Seungpyo Jeon[1], Byeongcheol Kim[1] and Seonyoung Park[1]*

[1] Department of Applied Artificial Intelligence, Seoul National University of Science and Technology, 232 Gongneung-ro, Nowon-gu, Seoul 01811, Republic of Korea

* Correspondence: sypark@seoultech.ac.kr

Abstract: Accurate, detailed, and up-to-date tree species distribution information is essential for effective forest management and environmental research. However, existing tree species maps have limitations in resolution and update frequency, making it difficult to meet modern demands. To overcome these limitations, this study proposes a novel framework that utilizes existing low-resolution national tree species maps as 'weak labels' and fuses multi-temporal Sentinel-2 and PlanetScope satellite imagery data. Specifically, a Super-Resolution (SR) technique, referencing PlanetScope imagery, was first applied to Sentinel-2 data to enhance its resolution to 2.5 m/px. Then, these enhanced Sentinel-2 bands were combined with PlanetScope bands to construct the final multi-spectral, multi-temporal input data. Deep Learning (DL) model training data was built by strategically sampling information-rich pixels from the national tree species map. Applying the proposed methodology to Sobaeksan and Jirisan National Parks in South Korea, the performance of multiple models (Linear, RF, MLP, Light, Transformer) was compared. The MLP model demonstrated optimal performance, achieving over 85% Over Accuracy (OA) and over 81% accuracy even in classifying spectrally similar and difficult-to-distinguish Quercus mongolica (QM) and Quercus variabilis (QV). Furthermore, while spectral and temporal information were confirmed to contribute significantly to tree species classification, the contribution of spatial (texture) information was experimentally found to be limited at the 2.5 m resolution level. This study presents a practical method for creating high-resolution tree species maps scalable to the national level by fusing existing tree species maps with Sentinel-2 and PlanetScope imagery, without requiring separate costly field surveys. Its significance lies in establishing a foundation that can contribute to various fields such as forest resource management, biodiversity conservation, and climate change research.

Keywords: Tree species classification; Deep Learning; Remote Sensing; Sentinel-2; PlanetScope; Super-Resolution; Tree species map; Quercus classification

Academic Editor: Firstname Lastname

Received: date
Revised: date
Accepted: date
Published: date

Citation: To be added by editorial staff during production.

1. Introduction

Accurate and up-to-date tree species distribution information is essential foundational data for effective modern forest management and environmental research. This information plays a crucial role in establishing biodiversity conservation strategies and evaluating their effectiveness [1], analyzing and predicting habitats for specific wild-life species, and formulating sustainable forest management plans. Furthermore, it contributes to climate change mitigation efforts by improving the accuracy of forest carbon storage estimations [2,3] and provides critical information for predicting wild-fire spread

based on species-specific characteristics and developing effective suppression strategies. Changes in the distribution patterns of specific tree species communities can also serve as important indicators of the impact of climate change on forest ecosystems, linking to forest health analysis based on accurate species information [4] or supporting reforestation decisions [5]. Moreover, precise location information for key tree species is indispensable for predicting the spread of forest pests and diseases, such as pine wilt disease, and establishing proactive control measures. Therefore, securing timely and detailed tree species distribution data is paramount for enhancing the effectiveness of a wide range of ecological and environmental management fields.

Historically, tree species distribution information has been primarily obtained through traditional field surveys and aerial photograph interpretation [6,7], often compiled in the form of tree species maps[8]. While these maps have contributed to past forest resource assessment and management, they exhibit several limitations in meeting modern demands for precision management and analysis. One of the major issues is the limitation of spatial resolution. Existing tree species maps often simplify areas with mixed tree species into single dominant species stands due to their low resolution. Furthermore, areas without a clear dominant species are frequently categorized broadly as 'broadleaf forest', 'coniferous forest', or 'mixed coniferous-broadleaf forest', failing to provide specific information about the individual species present. As such, tree species maps often struggle to reflect the complex and fine-scale patterns of species distribution within actual forests. Additionally, their long production and up-to-date cycles result in low temporal resolution, posing challenges in promptly identifying dynamic changes in forest ecosystems. The considerable time and effort required for field surveys and manual interpretation act as constraints in rapidly acquiring up-to-date national-scale information.

To overcome the limitations of these traditional methods, remote sensing technology has emerged as a powerful tool for forest monitoring over the past few decades [9]. Early approaches focused on mapping broad forest types using medium-resolution satellite imagery like Landsat, applying pixel-based classification algorithms or simple statistical techniques. Subsequent advancements, driven by the availability of freely accessible high-resolution multispectral imagery such as Sentinel-2 [10] and increased computing power, have spurred significant progress. Researchers have increasingly adopted Machine Learning (ML) algorithms like Random Forest (RF)[11] and Support Vector Machines (SVM) [12-14]. These algorithms often demonstrate improved performance over traditional classifiers by combining rich spectral information extracted from satellite data with, occasionally, texture feature s[15] or environmental variables[16]. More recently, DL techniques, particularly Convolutional Neural Networks (CNNs) and their variants like U-Net, have shown remarkable potential in tree species classification [17]. DL models can automatically learn hierarchical features from spectral, temporal, and spatial patterns within images, often outperforming conventional ML methods, especially when dealing with complex forest structures or utilizing multi-temporal datasets to capture seasonal variations among species [18-20].

Despite these technological advancements, two major practical challenges hinder the application of state-of-the-art satellite-based tree species mapping techniques to actual large-scale areas, particularly at the national level. The first is the difficulty in obtaining accurate and extensive labeled data required for training supervised ML/DL models. Building reliable reference data necessitates large-scale field sampling and meticulous manual labeling, which incurs substantial costs and time, thereby creating a significant barrier to extending research areas to a national scale. The second challenge concerns the spatial resolution of commonly used satellite data. While Sentinel-2 provides valuable multispectral bands (10-20m resolution) for tree species classification, this resolution is often insufficient to distinguish individual tree crowns or capture fine-scale species

distribution within mixed forests. Although some studies have shown promising results, many utilize original resolution data or employ traditional up-sampling methods, consequently limiting the detail level of the resulting tree species maps.

This study introduced a novel and practical approach designed to simultaneously address the critical limitations of training data acquisition and spatial resolution in large-scale tree species mapping. To overcome the challenge of costly labeled data generation, we utilized readily available national-level tree species maps as initial 'weak' labels, developing an efficient sampling strategy specifically designed to select information-rich pixels and mitigate potential noise and data imbalance issues [21] inherent in such sources. Concurrently, to address the spatial resolution constraints of commonly used satellite data, we applied an advanced DL-based Super-Resolution (SR) technique to Sentinel-2 satellite imagery, significantly enhancing its spatial detail to a 2.5 m/pixel level prior to classification. A DL-based tree species classification model was then trained using this combination of strategically sampled weak labels derived from existing maps and the super-resolved Sentinel-2 imagery. Finally, the trained model was applied across the entire study area using the super-resolved imagery to generate a precise tree species map at the enhanced 2.5 m/pixel resolution.

The primary contribution of this research is the development and successful validation of a practical 'recipe' for significantly improving the quality and resolution of existing national-level tree species maps without requiring substantial additional investment in field data collection. We demonstrated the effectiveness of synergistically combining weak supervision (using existing maps) with SR; the classification achieved over 85% Overall Accuracy (OA) for the five major tree species in South Korea when compared against the initial weak labels. Notably, even the challenging classification between Quercus mongolica (QM) and Quercus variabilis (QV) reached an accuracy exceeding 81%. Additionally, visual analysis of the predictions for test areas confirmed that the method effectively provides high-resolution information on mixed forests, de-tails unavailable in the original tree species map. This confirms the potential of the approach to producing high-quality tree species information cost-effectively. By extending this methodology nationwide and potentially utilizing larger training datasets, we anticipate enabling the classification of a wider variety of species, ultimately enabling the creation of high-resolution (e.g., 2.5 m/px) tree species maps over large areas, which will substantially benefit th diverse forest management and environmental re-search fields outlined earlier.

2. Study Site and Datasets

This study was conducted in Sobaeksan National Park and Jirisan National Park, representative forest areas in South Korea. Both sites offer suitable environments for tree species classification research due to their wide elevation ranges and diverse species distributions. Furthermore, being designated and managed as national parks, they experience relatively low anthropogenic disturbance, making them suitable locations for long-term ecological research. Figure 1 shows the specific study areas. In Sobaeksan National Park, an area of 9 km by 9 km, totaling 81 km², was selected as the study site, with latitude and longitude ranging from (36°56'N – 37°00'N, 128°26'E – 128°32'E), located in the mid-latitudes of South Korea. The elevation in this area varies from below 250m to over 1400m. Dominant tree species include oaks such as QM and QV. Other species present include Pinus densiflora (PD), Larix kaempferi (LK), Pinus koraiensis (PK), and Poplar (Populus spp.).

In Jirisan National Park, data covering an area of 18 km by 9 km, totaling 162 km², were used, with latitude and longitude ranging from (35°17'N – 35°22'N, 127°35'E – 127°47'E). This area is located at a relatively lower latitude in South Korea, with elevations

ranging from below 350 m to over 1,900 m. Major tree species are predominantly oaks like QM and QV, with Abies koreana communities existing in high-altitude areas. PD, PK, and LK also inhabit the area.

For efficient data processing and training, the Jirisan area was divided into square grids, creating two sub-regions, Jirisan 1 and Jirisan 2. Including the Sobaeksan area, the data was organized into a total of three regional units. This division aims not only for efficient memory management but also enables the application of a consistent data split strategy, thereby enhancing the consistency and efficiency of the entire training and evaluation process.

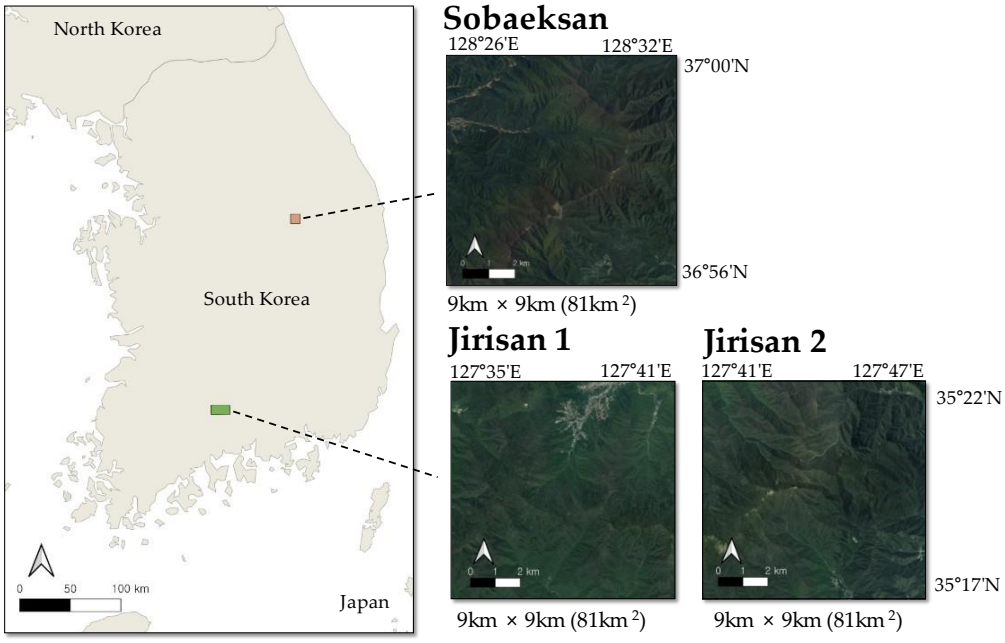

Figure 1. Location of the study areas. The map on the left shows the locations of Sobaeksan (orange) and Jirisan (green) National Parks within South Korea. The panels on the right are satellite images of the 9 km × 9 km regions: Sobaeksan, Jirisan 1, and Jirisan 2.

2.1. Data Timeframes

Since the reference label data used in this study was based on the year 2022, satellite data from 2022 were utilized, incorporating imagery from a total of 12 time points. Data was collected basically at monthly intervals. To capture species-specific characteristic differences more precisely during leaf-out timing and leaf senescence timing, additional data were acquired at 15-day intervals during specific periods. From April 1st to June 1st, data were collected every 15 days to better capture inter-species differences during leaf-out. Similarly, from October 1st to November 1st, data were additionally acquired in the same manner to accurately capture observable inter-species differences during leaf senescence. Consequently, the data time points used in this study are as follows. However, if data for a specific date could not be obtained due to weather conditions or other factors, the data available closest date's data was used. **Table 1** below details the selected target dates and the actual data acquisition dates for each satellite and site.

Table 1. Data Acquisition Date

Target date (2022)		02/01	03/01	04/01	04/15	05/01	05/15	06/01	07/01	09/01	10/01	10/15	11/01
Jirisan	Planet Scope	01/31	03/03	03/29	04/17	05/01	05/18	05/31	07/02	09/07	10/01	10/15	11/01

Sobaeksan	Senti-nel-2	01/27	03/03	04/04	04/17	05/04	05/14	06/03	07/01	09/09	10/01	10/14	11/05
	Planet Scope	01/31	03/03	03/29	04/16	05/03	05/16	06/01	07/03	09/08	09/30	10/18	11/04
	Senti-nel-2	01/29	03/03	04/04	04/17	05/04	05/17	06/01	07/03	08/27	09/24	10/19	11/05

2.2. Satellite Imagery

This study utilized Sentinel-2 and PlanetScope satellite data. Sentinel-2, operated as part of the Copernicus program, is a multispectral satellite providing a total of 13 spectral bands. We used Sentinel-2 Level-2A products, incorporating data from both Sentinel-2A and Sentinel-2B satellites. Sentinel-2 offers the advantage of providing various spectral bands, particularly including bands crucial for tree species classification as red edge, near infrared (NIR), and short-wave infrared (SWIR), making it suitable for this research. The importance of these bands in satellite image-based tree species classification tasks has been demonstrated in various previous studies.

Meanwhile, PlanetScope [22] satellite data utilizes over 150 micro-satellites (Dove series) to provide high-resolution imagery with an average revisit period of less than one day. This study employed PlanetScope imagery tagged as PSScene captured by the SuperDove constellation. This data consists of orthorectified images with a ground sampling distance of 4.1m and a spatial resolution of 3 m/px. SuperDove satellites, based on the latest Dove-R platform launched from 2020, offer an enhanced 8-multispectral band configuration and improved radiometric calibration performance compared to the previous Dove-Classic and Dove-R.

The data used in this study were provided through Planet's Education & Research Program, which restricted usage to only 4 of the 8 bands: Red, Green, Blue, and NIR. Given the study's objective of analyzing the precise spatial distribution of tree species, satellite data with higher spatial resolution than Sentinel-2 was required, and PlanetScope imagery was a suitable alternative meeting these conditions. Notably, the program allowed free data access within a monthly limit of 3,000 km², enabling its use in a manner consistent with research objectives.

The specific satellite bands and spatial resolutions used in this study are summarized in **Table 2**. Among these, Sentinel-2 bands B02, B03, B04, and B08 were used only for the SR process. The final input data for the tree species classification model consisted of PlanetScope's blue, green, red, and NIR bands and Sentinel-2's B05, B06, B07, B8A, B11, and B12 bands.

Table 2. Satellite Information.

Satellite	Spatial Resolution (m)	Spectral Bands	Center Wavelength (nm)	Band Width (nm)
Sentinel-2	10	blue (B2)	492.1	66
		green (B3)	559.0	36
		red (B4)	664.9	31
		NIR (B8)	832.9	106
	20	Vegetation red edge (B5)	703.8	16
		Vegetation red edge (B6)	739.1	15

		Vegetation red edge (B7)	779.7	20
		Narrow NIR (B8A)	864.0	22
		SWIR (B11)	1610.4	94
		SWIR (B12)	2185.7	185
PlanetScope	3	blue	490	25
		green	566	19
		red	665	15
		NIR	865	20

3. Methodology

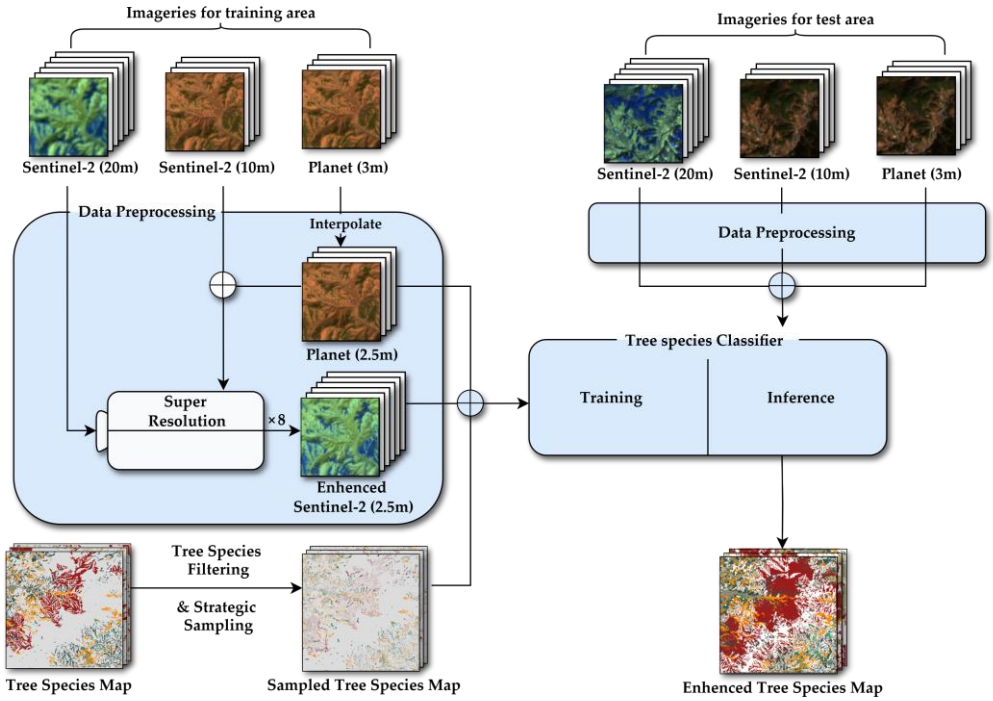

Figure 2. Overall methodological workflow. Schematic overview of the processing pipeline, from data acquisition (Sentinel-2, PlanetScope, tree species map) and preprocessing (SR, Strategic sampling) to model training/inference and generation of the enhanced tree species map.

3.1. Super Resolution

As previously mentioned, Sentinel-2 data has diverse bands, but its spatial resolution is coarse for tree species classification. Since this study aims to generate a high-resolution tree species map, data with higher resolution was necessary. Furthermore, we intended to verify the necessity of Forest Texture information. However, given Sentinel-2's 20 m/px resolution, we determined that even if significant spatial information exists at the PlanetScope resolution level, Sentinel-2's lower resolution could hinder the learning process. Therefore, we adopted an approach applying a SR technique to the Sentinel-2 data to enhance its resolution.

SR was performed based on Wald's Protocol, referencing the methodology from [23] presents a method to increase the resolution of Sentinel-2 satellite data using PlanetScope satellite data. The SR process performed in their study differs from typical SR tasks. Unlike

conventional SR problems, this approach benefits from already having high-resolution data for specific bands, allowing for potentially improved results. The general process is as follows:

1. Interpolate PlanetScope data (3 m/pixel) to 2.5 m/pixel (Bi-cubic interpolation).
2. Downsample both PlanetScope and Sentinel-2 data to 1/8 resolution. The text continues here.
3. Train a Residual Convolutional Neural Network model for ×8 resolution restoration.
4. Apply the trained model to the non-downsampled Sentinel-2 data (20 m/pixel) to generate 2.5 m/pixel resolution images.

The SR results are presented in **Figure 3**.

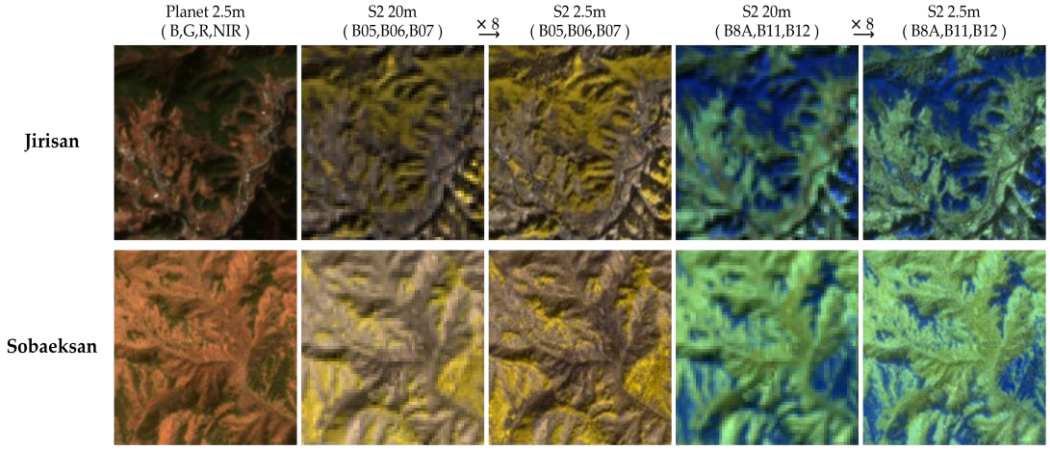

Figure 3. Super-Resolution (SR) results for Sentinel-2 imagery in the Jirisan and Sobaeksan area. Comparison between the original 20m bands (B05-B06-B07 and B8A-B11-B12) with the 2.5m resolution results generated by applying 8x SR. PlanetScope 2.5m imagery (blue, green, red, and NIR) is shown on the far left for reference.

3.2. Label Data

In this study, labels for training were generated using the tree species map data provided by the Korea Forest Service (https://map.forest.go.kr, accessed 18 December 2024). We selected a total of five tree species: the most common coniferous species in Korea, PD, PK, and LK, and the representative broadleaf species QM and QV. Adding a non-forest (NF) class resulted in a total of six classes. Pixels classified as these specific classes in the tree species map were selected and utilized as training data. The tree species map contains a significant number of areas that cannot be mapped to a single species, such as coniferous forests, broadleaf forests, and mixed coniferous-broadleaf forests. In this study, these mixed species areas were excluded, and only pixels classified as a single species were used for training.

Table 3 presents the distribution of data points for each tree species class per region, showing both the original counts and proportions before sampling on the left side and the counts and proportions after applying the sampling strategy on the right side. The left side of Table 3 reveals the initial data distribution. Using this data directly for training could lead to performance degradation for minority classes due to severe class imbalance. Particularly, if certain species are concentrated in specific regions, the model might be overfit to patterns unique to those regions, potentially failing to accurately classify species distributed in minority areas. Furthermore, using all data would significantly increase the computational resources required for training.

To address these issues, this study introduced a sampling strategy that limits the number of samples per class within each region. Specifically, if the number of pixels for a particular species class within a region exceeded 15,000, 15,000 samples were uniformly

extracted from that class. If the count was 15,000 or less, the entire data for that class was used.

Furthermore, to prevent sample concentration in specific spatial areas within a single region, samples were evenly selected by sequentially iterating through spatially distinct areas (box numbers, see Section 3.3). This method prevents training biased towards specific areas and ensures the inclusion of data from diverse spatial locations in a balanced manner, preserving diversity of information.

Figure 4 illustrates an example of sampling performed for the QM class in the Jirisan 1 region. As can be seen in the figure, samples were extracted evenly across the entire region. The resulting number and proportion of samples per class in the final dataset constructed this way are summarized on the right side of **Table 3**. After sampling, the smallest class, PK, constitutes 12.0% of the total samples, while the largest class, QM, accounts for 19.6%, indicating a significant reduction in class imbalance compared to the original distribution shown on the left side. It can also be observed that the regional distribution bias has been greatly improved.

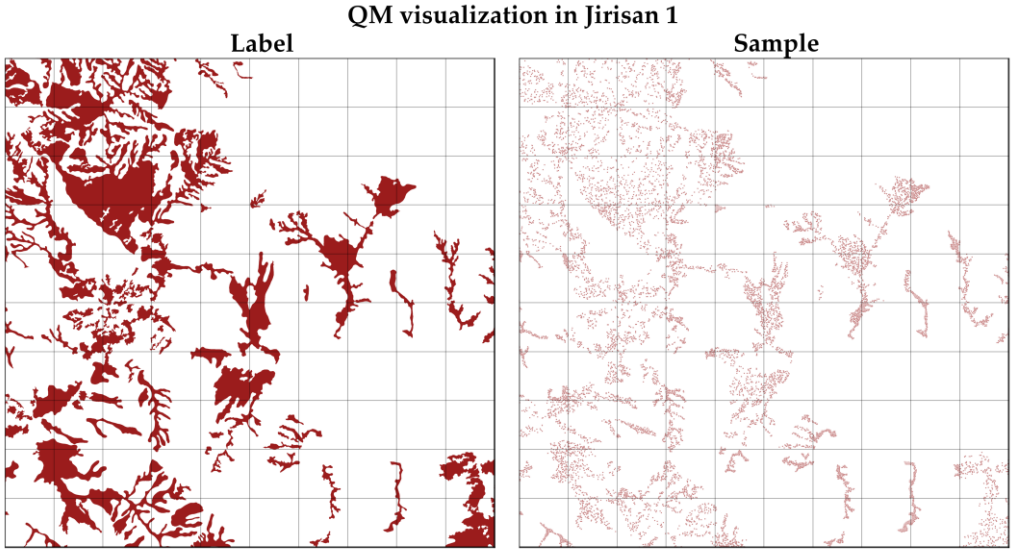

Figure 4. The visualization of strategic sampling applied on Jirisan. It visualizes QM distribution on the left side and the spatially uniformly extracted training samples on the right side.

Table 3. The results of strategic sampling.

Species code	Region	Number of original data	Ratio of original data (%)	Number of sampled data	Ratio of sampled data (%)
NF	Jirisan 1	42,258	3.86	14,614	6.26
	Jirisan 2	12,224	1.12	12,224	5.24
	Sobaeksan	22,398	2.04	15,853	6.79
	Total	76,880	7.02	42,691	18.29
PD	Jirisan 1	41,445	3.78	14,864	6.37
	Jirisan 2	5,881	0.54	5,402	2.31
	Sobaeksan	113,645	10.37	15,589	6.68
	Total	160,971	14.69	35,855	15.36
PK	Jirisan 1	14,822	1.35	13,971	5.98
	Jirisan 2	3,512	0.32	3,435	1.47

	Sobaeksan	10,587	0.97	10,587	4.53
	Total	28,921	2.64	27,993	11.98
LK	Jirisan 1	21,852	1.99	15,522	6.65
	Jirisan 2	13,994	1.28	13,333	5.71
	Sobaeksan	79,311	7.24	15,307	6.56
	Total	115,157	10.51	44,162	18.92
QM	Jirisan 1	243,180	22.2	15,406	6.6
	Jirisan 2	189,166	17.27	15,374	6.59
	Sobaeksan	138,665	12.66	14,920	6.39
	Total	571,011	52.13	45,700	19.58
QV	Jirisan 1	69,056	6.3	15,200	6.51
	Jirisan 2	66,881	6.1	15,077	6.46
	Sobaeksan	6,774	0.62	6,774	2.9
	Total	142,711	13.02	37,051	15.87

3.3. Selected Models Using Single-pixel Classification

This study adopted a single-pixel classification approach. To explore the most effective model architecture for this approach, various candidate models were designed, including a Linear model, RF, MLP [24], a Light model, and Transformer [25]. The **Linear** model was introduced as a baseline model to approach the tree species classification task in its simplest form. This model flattens the (times, bands) input data into a 1D vector and directly maps it to class classification results through a single linear layer. The input data undergoes prior band-wise normalization, designed to enable the model to focus on the relative changes in band values at each time step. This normalization helps the model learn useful features more effectively, even with limited representational power. The total number of parameters for this model is 726, representing a very small and constrained structure. Training was performed for 100 epochs using the AdamW optimizer [26] and CosineAnnealingLR scheduler [27]. AdamW (Adam with decoupled weight decay) is a variant of the Adam optimizer that applies the weight decay term separately from the gradient update, thereby reflecting the L2 regularization effect more accurately. In this study, a learning rate of 5e-4 and weight decay of 5e-4 were set, values found to perform well in preliminary experiments. The CosineAnnealingLR scheduler gradually decreases the learning rate following a cosine function shape, allowing for rapid convergence with a large learning rate early in training and promoting stable convergence by reducing the learning rate later. The period for the cosine annealing scheduler was set to adjust the learning rate over the entire 100 epochs. This training configuration was applied identically to all subsequent DL models.

RF has shown stable performance in various classification problems and has been widely used due to its relatively light structure, interpretability, and efficiency. Particularly in the field of tree species classification, it has been utilized as a fundamental comparison model in many previous studies. Therefore, it was adopted in this study as a representative traditional machine learning technique. In this experiment, all hyperparameters for the RandomForestClassifier were kept at their default settings. Specifically, the number of trees was set to 100, and the default 'gini' coefficient was used as the splitting criterion. The maximum depth of the trees was not limited to allow sufficient data splitting, and the minimum number of samples required to split a node was set to 2, configured to allow maximum splitting at each node. These settings aim to maximize the expressive power of the individual models. Compared to the linear classifier, the RF model is an ML technique capable of learning more powerful non-linear relationships, potentially achieving relatively high performance even with a simple structure. Thus, this study uses RF to

gauge the performance limits of traditional ML methods and, through comparison with the subsequently introduced DL-based models, examines whether DL can substantially contribute to performance improvement in the single-pixel based tree species classification problem.

MLP was introduced to evaluate performance when the model learns by itself with a sufficient number of parameters, without imposing a specific inductive bias. This model is based on a typical deep neural network structure, where each layer consists of a linear layer, batch normalization[28], SiLU activation function [29], and dropout [30]. This block structure is repeated three times, and the total number of parameters in the model is approximately 98k. Input data is passed to the model in a flattened form and is configured to learn complex non-linear relationships through multiple hidden layers. This model contrasts with the Light model and Transformer structure, which incorporate inductive bias into the model architecture to learn spectral and temporal dimensions separately. Comparison with these models provides clues as to which method is more effective.

The spectral encoder is used commonly in both the Light model and the Transformer model, serving to extract information from the spectral dimension of the input data. This module first linearly transforms the input dimension to a hidden dimension using a Linear layer, projecting the data into a higher-dimensional feature space. Subsequently, Batch Normalization is applied to stabilize the distribution of each feature and enhance efficiency in learning. Then, the SiLU activation function is introduced to perform non-linear transformation, helping to model more complex patterns effectively. By repeatedly applying a second Linear layer, Batch Normalization, and SiLU, the model progressively reduces the characteristics of the input signal, summarizing them into an embedding dimension dimension. Dropout, applied immediately after the final Linear layer, serves to reduce the risk of overfitting and improve generalization performance. Configured this way, the spectral encoder extracts meaningful features from the spectral dimension.

The **Light** model is built upon the spectral encoder and features the simplest structure for learning temporal information. This model uses the spectral encoder to extract spectral information for each time point, then transforms it into a sequence representation of shape (times, embedding dimension). Subsequently, this sequence is flattened into a single vector, and a single linear layer is applied to perform the final classification. The total number of parameters is approximately 31K. It is designed to process temporal information with a minimal structure, without complex temporal modeling. Therefore, this model serves as a baseline for comparison with more sophisticated time-based architecture, allowing analysis of performance differences based on the complexity of temporal modeling.

The **Transformer** model is designed to learn temporal information more effectively by adding a Transformer encoder after the spectral encoder. The sequence representation of shape (times, embedding dimension), extracted by the spectral encoder, is input into the Transformer encoder to learn time-series patterns (internally using the self-attention mechanism). Its output is then flattened, and a linear head performs the final classification. This model has a total of 98K parameters and can process information along the time axis more elaborately than the Light model. The purpose of this model is to evaluate how much high-capacity architecture like the Transformer-based model can contribute to performance improvement compared to a simple linear layer-based approach for temporal information processing.

Figure 5 shows the detailed structures of the MLP model, Light model, and Transformer.

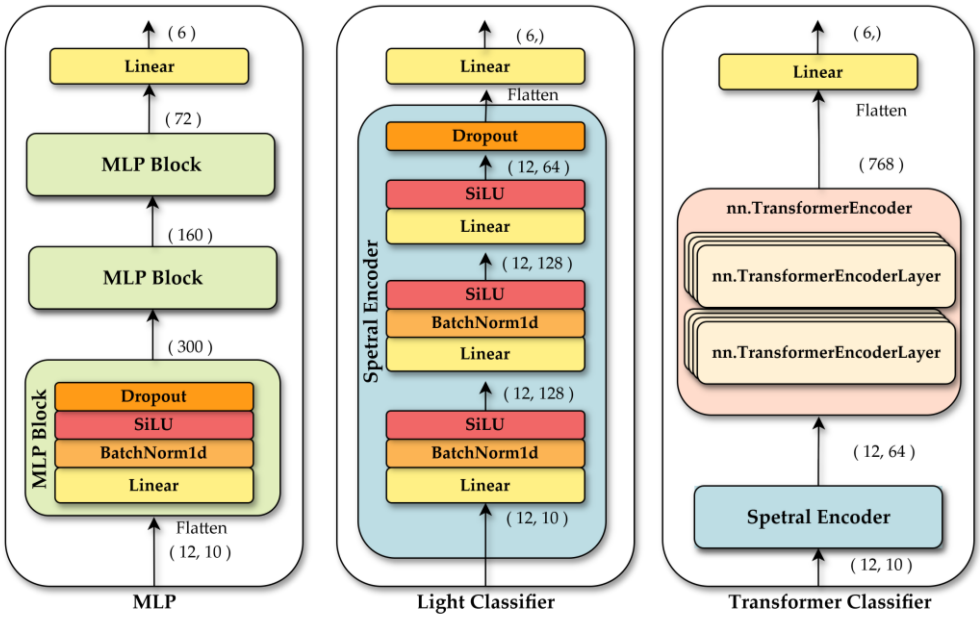

Figure 5. Architecture of the deep learning models used for single pixel-based tree species classification: MLP, Light classifier, and Transformer classifier.

3.4. Data Split

To effectively separate train-validation-test datasets, each region was divided into 100 grid cells (boxes). In this process, the three study were each subdivided into 100 small boxes. This allowed for the construction of spatially separated training, validation, and testing datasets, enabling a more accurate evaluation of the model's generalization performance. From each region, two boxes that exhibited minimal class imbalance and contained sufficient representation of the target tree species were selected for the test data. Since no hyperparameter tuning was performed in this study, the validation process did not involve fitting the model to the validation data; the validation data alone could verify the model's generalization ability. Therefore, the test data was allocated only about 2% of the total data, intended for future use such as qualitative evaluation. The remaining boxes, excluding the test data, were divided into five mutually exclusive train and validation splits for 5-fold cross-validation. In each fold, the validation data was set to be mutually exclusive, allowing experiments to be conducted with a total of five train-validation combinations. The specific data split structure is presented in Figure 6.

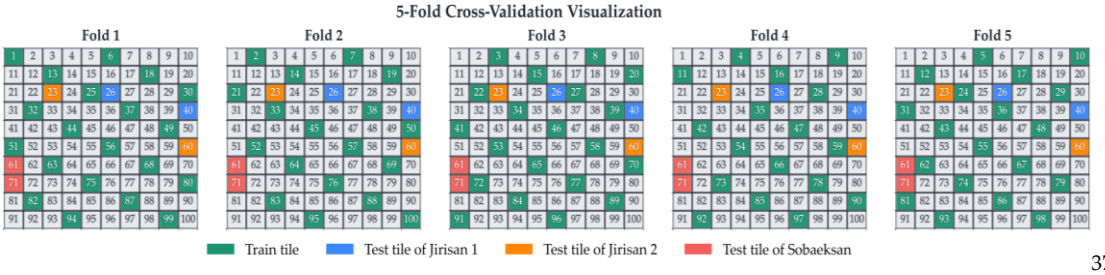

Figure 6. Visualization of the 5-fold cross-validation data split. The study area is divided into 100 tiles, showing the allocation of training tiles (green) and region-specific test tiles (Jirisan 1: blue, Jirisan 2: orange, Sobaeksan: red) for each fold.

3.5. Evaluation Metrics

During model development, quantitative Evaluation Metrics were used to assess the agreement with Ground Truth data. To analyze model performance in more detail and

facilitate intuitive understanding, this study defined four specific classification tasks for evaluation:

- **Non-Forest vs. Forest**: A binary classification task distinguishing NF areas from Forest areas across the entire study area;
- **Conifer vs. Broadleaf**: A binary classification task distinguishing Conifer from Broadleaf within areas correctly classified as Forest;
- **Intra-Conifer Classification**: Classification among three coniferous species—PD, PK, and LK—within areas correctly classified as Conifer.
- **Intra-Broadleaf Classification:** Classification between QM and QV within areas classified as Broadleaf.

For each of these four classification tasks, evaluation metrics such as accuracy, Macro-Averaged Precision, and Macro-Averaged Recall were calculated and utilized. For Precision and Recall, Macro-Averaging was used instead of Weighted Averaging to evaluate performance across classes more equitably. This was done because, although the sampling process somewhat mitigated imbalance, differences in the number of instances between classes still existed, and this approach minimizes the impact of such subtle imbalances on model performance evaluation.

3.6. Model Comparison

To compare model performance, 5- fold cross-validation was performed for each model. Final performance evaluation was based on the prediction results from the epoch showing the best validation performance for each fold. For all folds, Average Precision and Average Recall were measured for each of the four defined classification tasks (Non-Forest vs. Forest, Conifer vs. Broadleaf, Intra-Conifer Classification, Intra-Broadleaf Classification). This resulted in a total of 8 performance metrics per fold for each model. Consequently, through 5- fold cross-validation, a total of 40 quantitative performance metrics (5 folds × 4 tasks × 2 metrics) were obtained per model, forming the basis for inter-model performance comparison and analysis.

To assess the statistical significance of the collected performance metrics, the Friedman test was conducted. To analyze the performance differences between models more precisely, a post-hoc Nemenyi test was additionally performed. The Friedman test is a non-parametric test method used to verify whether performance differences between multiple models evaluated on the same dataset are statistically significant. If this test indicates a significant difference in the average ranks of the models, further multiple comparison analysis is needed. Accordingly, the Nemenyi test was applied as a post-hoc test to perform pairwise comparisons between models. The Nemenyi test compares the average rank differences between multiple models to determine if one model performs statistically significantly better than others. In this study, the Nemenyi test was used to analyze whether statistically significant performance differences exist between models, based on which the optimal model was selected. The results and detailed analysis of the model comparison experiments are presented in Section 4.1.

4. Results

4.1. Results of Model Comparison

Table 4. Comparison of models used.

Model	Non-Forest Vs. Forest	Conifer Vs. Broadleaf	Intra-Conifer	Intra-Broadleaf	Overall Accuracy	Overall F1-score
Linear	98.41	94.15	87.51	71.93	79.03	78.03

RF	98.71	94.47	87.93	77.10	81.36	80.02
MLP	**98.94**	**95.75**	**91.01**	**81.43**	**85.38**	**84.36**
Light	98.85	94.97	89.97	79.31	83.60	82.21
Transformer	98.89	95.17	90.47	81.60	84.78	83.55

Table 4 summarizes the average accuracy, overall F1-score, and OA from the 5-fold cross-validation results for each classification model. Even the simple Linear model achieved very high accuracy in higher-level classifications like Non-Forest vs. Forest and Conifer vs. Broadleaf, recording 98.41% and 94.15% accuracy, respectively. Its OA across all classes was also around 79% on average. This suggests that combining spectral and temporal information from satellite imagery is effective even with basic linear separation for classification tasks with relatively clear distinguishing features, such as forest/non-forest and conifer/broadleaf discrimination. Therefore, if high interpretability is crucial for these higher-level tasks, the Linear model could be a viable option.

However, the **Linear** model showed a marked decline in performance for tasks requiring finer distinctions within the same parent group, such as Intra-Conifer and Intra-Broadleaf classification. Its accuracy on these tasks was 87.51% and 71.93%, respectively, significantly lower than other non-linear models. This indicates that fine-grained classification tasks involve subtle and complex inter-class feature distinctions, demanding higher model expression power and capacity. In essence, the limited capability of the linear model proved insufficient for learning these complex patterns adequately.

The **RF** model demonstrated overall performance improvement compared to the Linear model. Notably, the accuracy for the most challenging Intra-Broadleaf classification improved by approximately 5.17 percentage points, from 71.93% to 77.10%. The overall F1-score and OA also increased to 80.02% and 81.36%, respectively. This suggests that RF, as an ensemble method, can capture some non-linear features and interactions in the data through the combination of decision trees. Nevertheless, limitations in fine-grained classification performance were still observed.

The **Light** model and **Transformer** model, utilizing the Spectral encoder, showed improved performance over RF. The improvement was particularly noticeable in detailed species classification (Intra-Conifer, Intra-Broadleaf), achieving accuracies of 89.97%, 79.31% (Light model) and 90.47%, 81.60% (Transformer), respectively. This validates the effectiveness of the DL-based approach for learning spectral and temporal information in the tree species classification task. Meanwhile, the Transformer model outperformed the Light model across all tasks, suggesting the importance of effectively learning temporal information using a module with sufficient capacity after spectral encoding.

The **MLP** model exhibited the best overall performance in this evaluation. It recorded the highest performance in Non-Forest vs. Forest (98.94% accuracy), Conifer vs. Broadleaf (95.75% accuracy), and Intra-Conifer (91.01% accuracy) classifications. Even in the most difficult Intra-Broadleaf classification, its accuracy of 81.43% was nearly tied with the Transformer model for the top performance. Consequently, MLP surpassed all other models in both overall F1-score (84.36%) and OA (85.38%), achieving the best comprehensive performance. Notably, it recorded slightly higher performance even compared to the Transformer model, which has a very similar number of parameters. This might suggest that an approach allowing the model to directly find patterns useful for tree species classification from the training data, rather than injecting specific inductive biases into the model, was potentially more suitable for the dataset and objective of this study.

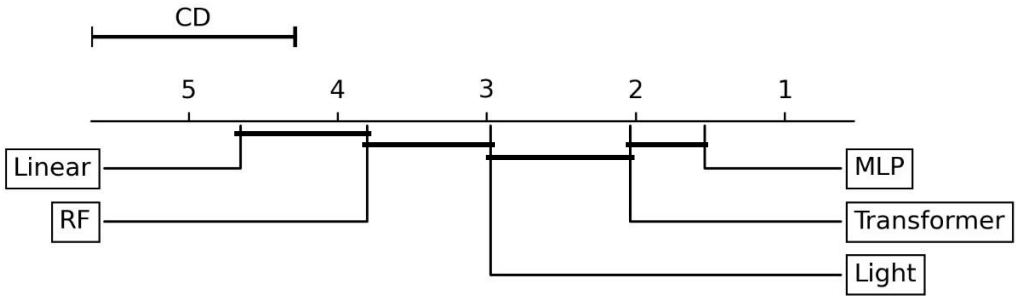

Figure 7. Nemenyi post-hoc test results (α=0.05) for comparing the performance of the five classification models. Models are ranked by average performance rank, and the Critical Distance (CD) indicates statistical significance. The connecting bar indicates a group of models with no statistically significant difference.

In conclusion, DL-based models generally showed superior classification performance compared to traditional machine learning models, with the MLP model consistently demonstrating the highest performance. To verify the statistical significance of these performance differences, a Friedman test was conducted. The result rejected the null hypothesis that there is no difference in performance among the models at a significant level of 0.05 ($p < 0.001$), confirming that the performance differences are statistically significant. Accordingly, a post-hoc Nemenyi test was performed for detailed analysis of performance rankings and significant differences between models.

The critical distance (CD) plot in Figure 7 visualizes the Nemenyi test results. In this plot, models are ranked according to their average performance rank. A statistically significant performance difference exists between two models if the difference in their average ranks exceeds the CD. The analysis revealed that MLP achieved the best average rank. The connecting bars in the plot indicate groups of models where the average rank difference is within the CD, meaning there is no statistically significant performance difference. Specifically, the performance difference between the MLP and Transformer models is not statistically significant.

However, the Nemenyi test results clearly show that these top-performing models as MLP and Transformer exhibits statistically significant performance differences compared to the other models, namely the Light model, RF, and Linear model. The average ranks of MLP and Transformer were significantly higher (by more than the CD value) than those of the Light model, RF, and Linear model. Furthermore, the Light model also demonstrated statistically significantly better performance than both RF and the Linear model. In summary, the DL-based models (MLP, Transformer, Light model) generally performed better than the traditional machine learning techniques (RF, Linear), with MLP and Transformer forming the top-performing group, as statistically confirmed.

Although the Nemenyi test did not find a statistically significant difference between MLP and Transformer, MLP 1) recorded the highest values for most individual classification tasks and overall evaluation metrics, and 2) achieved the best average rank in the Nemenyi test. Therefore, based on the aggregation of this evidence, MLP was ultimately selected as the most suitable model for achieving the tree species classification goals of this study and was chosen for subsequent analysis and result generation.

4.2. Final Model Performance

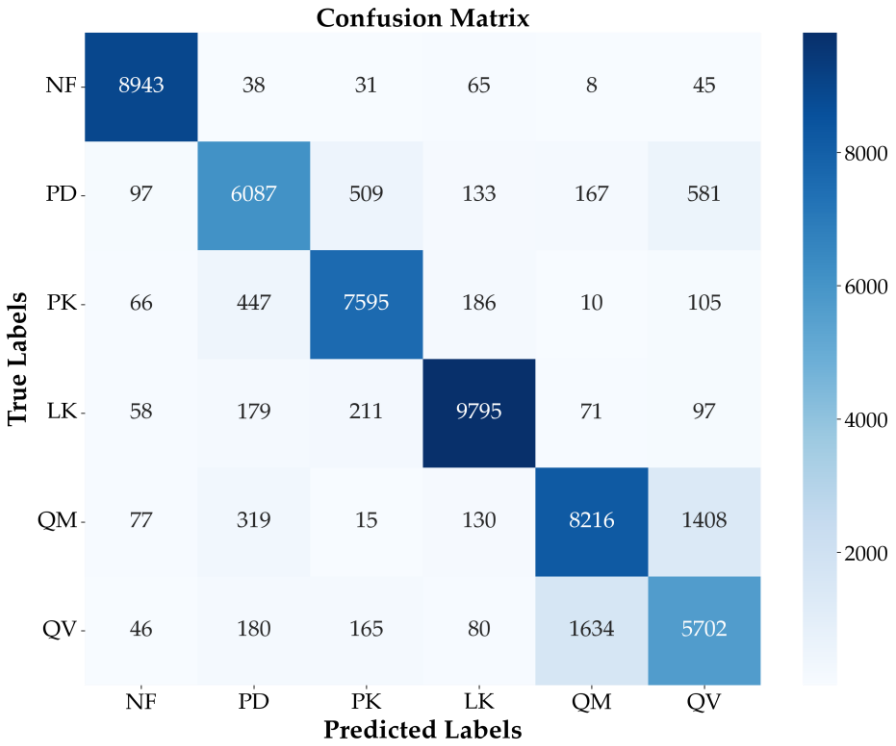

Figure 8. Confusion matrix for the best performing fold of the MLP model. Illustrates the classification performance across the six classes (NF, PD, PK, LK, QM, QV) and misclassification trends, particularly between similar species (PD/PK, QM/QV).

Figure 8 shows the confusion matrix for a single fold that exhibited the best performance among the 5- fold cross-validation results for the finally selected MLP model. This is presented to provide a detailed understanding of the model's overall classification performance, prediction accuracy for each class, and inter-class misclassification patterns. The diagonal elements of the confusion matrix represent the number of samples correctly classified for each class, while the off-diagonal elements indicate the number of samples of a specific class misclassified as another class.

The analysis revealed that the MLP model demonstrated generally excellent classification performance. Particularly, the NF class showed the highest classification accuracy, with 8943 samples correctly classified and very few misclassifications into other classes. Additionally, the LK class was also classified very effectively, recording a high True Positive count of 9795. This suggests the model has strengths in distinguishing non-forest areas and specific coniferous species (larch).

However, confusion still occurred between certain classes. Notably, frequent mutual misclassifications were observed between PD and PK (PD to PK: 509 cases, PK to PD: 447 cases). This suggests that the model might have struggled to distinguish between these two species belonging to the same Pinus genus due to similarities in their spectral or temporal characteristics, reflecting the difficulty of Intra-Conifer classification. In broadleaf classification, the confusion between QM and QV was the most prominent (QM to QV: 1,408 cases, QV to M: 1,634 cases). This accounted for the largest proportion of all misclassification instances, clearly indicating that distinguishing between these two Quercus genus species was the most challenging task in this study. Both species belong to the Quercus genus, possessing very similar spectral characteristics, as well as similar ecological traits and habitat preferences, posing inherent difficulties for satellite image-based classification. The possibility of hybridization in some areas could also contribute to making clear distinction even more difficult. Despite these inherent similarities and classification

challenges, the model achieved significant performance in distinguishing the two classes, correctly classifying a substantial number of samples for both QM and QV (8,216 and 5,702 True Positives, respectively).

Nevertheless, for QV, the proportion of samples misclassified as QM (1,634 cases) was relatively high compared to correctly classified samples (5,702 cases), suggesting that reducing misclassification for this class is a key task for future model performance improvement. Additionally, some confusion between other classes was also observed, such as the conifer PD being misclassified as the broadleaf QV (581 cases), and conversely, the broadleaf QM being misclassified as the conifer PD (319 cases).

Overall, the confusion matrix analysis in Figure 8 confirms that while the MLP model achieved high overall classification performance, challenges remain, particularly in distinguishing fine-grained species with similar characteristics within the same genus (e.g., PD/PK within Pinus, QM/QV within Quercus).

4.3. Enhanced Tree Species Map

Figure 9 and **Figure 10** show examples where the original tree species map was improved using the prediction results from the model developed in this study (Figures omitted). The prediction and improvement target areas commonly addressed in both figures are: (1) Areas classified as one of the five main target species in the original tree species map, (2) NF areas, and (3) Areas indicated by hatching in the figures. These hatched areas correspond to mixed forest areas in the original tree species map where the species composition was complex, preventing classification as a specific single species, and were thus broadly categorized as 'broadleaf forest', 'coniferous forest', or 'mixed coniferous-broadleaf forest'. One of the key contributions of this study's model is generating fine-grained tree species prediction information at the pixel level for precisely these hatched areas, where the original tree species map failed to provide detailed information, thereby significantly enhancing the map's information detail.

Figure 9 validates this improvement effect by showing the model's prediction results for an independent, pre-separated test area. All pixels in this area consist of unseen data not used for model training, thus serving to visually evaluate the model's generalization performance and species prediction accuracy in a real-world scenario. As can be observed in the figure, the model's predictions for areas classified as single species in the original map show high agreement with the actual labels. Furthermore, it clearly demonstrates that the model successfully generated detailed predictions within the hatched areas as described earlier, overcoming the limitations of the original map and improving information detail. Comparing the provided aerial imagery (https://map.kakao.com, https://map.ngii.go.kr/ms/map/NlipMap.do, accessed 30 March 2025) with the prediction results allows for visual verification that the model's predictions reasonably reflect the actual vegetation distribution.

Figure 10 is the final enhanced tree species map generated by applying the model predictions to the entire study area. Based on the performance demonstrated in the test area shown in **Figure 9**, this represents the final output where the quality and information detail of the tree species map have been improved across the entire study site. Through the enhanced tree species map in **Figure 10**, it is now possible to ascertain the specific tree species distribution status within the previously broadly categorized hatched areas, enabling a more precise and fine-grained understanding of the vegetation distribution throughout the entire study area.

In summary, this study effectively produced a high-precision 2.5 m/pixel resolution tree species map by training a model with meaningful information sampled from a relatively coarse-grained original tree species map and applying it to the entire study area.

These results suggest that the proposed methodology can effectively contribute to enhancing the resolution and information detail of existing tree species maps.

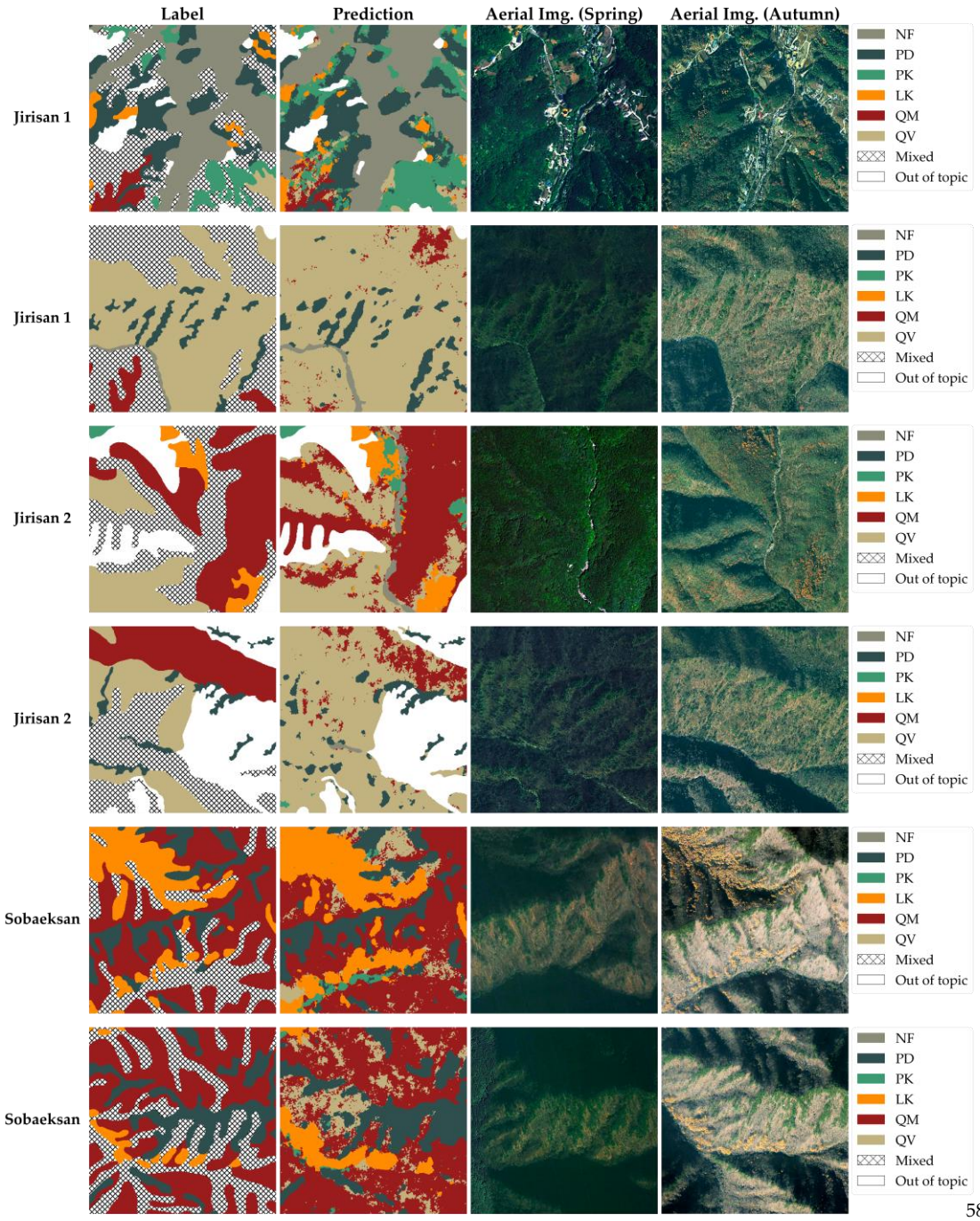

Figure 9. Comparison of the original tree species map (Label) and model prediction (Prediction) in test areas (unseen data), with aerial imagery for reference. Demonstrates the model providing detailed information within the original mixed forest (hatched) areas.

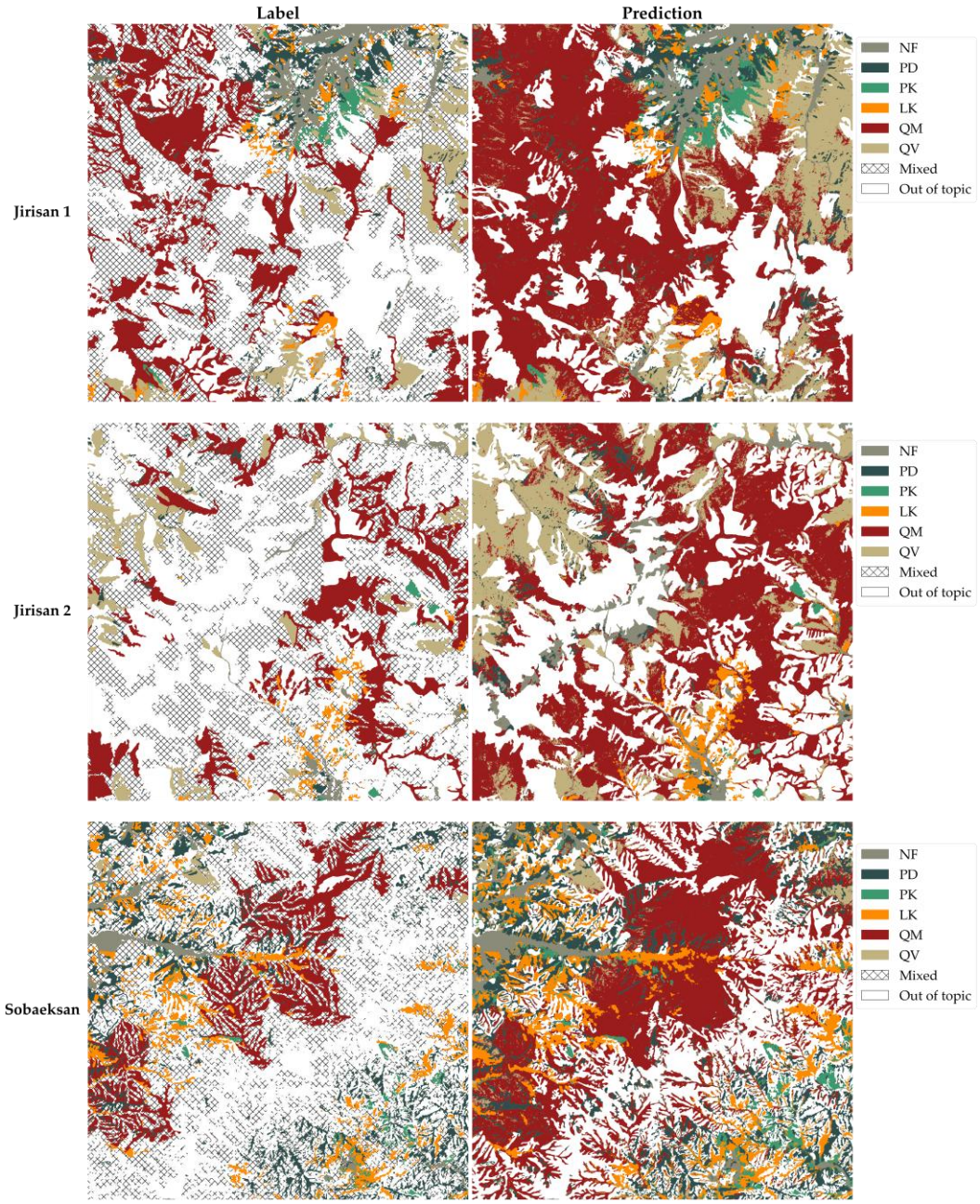

Figure 10. Enhanced tree species map. Comparison between the original tree species map (Label) and the final, enhanced 2.5m resolution model-predicted tree species map (Prediction) for the entire study areas (Jirisan 1, Jirisan 2, Sobaeksan).

5. Discussion

5.1. Detailed Analysis of Test Area Results

Figure 11 presents magnified views of selected notable areas from the test areas shown in **Figure 9**, facilitating a detailed analysis of the prediction results from the model proposed in this study. It encompasses both instances where the model's predictions were successful as good case and instances where the prediction accuracy warrants further examination as bad case.

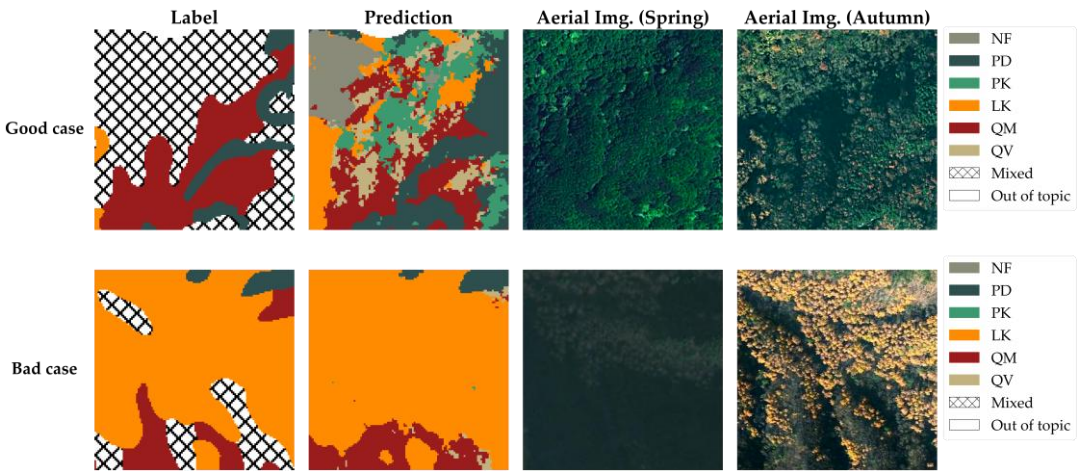

Figure 11. Test area prediction results which are comparisons of good cases (detailing mixed forests) and bad cases (LK boundaries).

First, the area presented as the good case serves as a representative example where the model provides significantly more detailed information compared to the original tree species map. As previously mentioned, the hatched areas on the original tree species map, which could not be classified under a single species, likely represent areas where identifying a specific dominant species was challenging or where multiple species coexist in a mixed forest setting. Indeed, the model predicts these areas as complex mosaics comprising various classes, including QM, QV, PD, PK, LF, and NF. This reveals the specific vegetation composition within the mixed forest, information unattainable from the original map. Naturally, the possibility that these areas represent communities of other species not included in the model's training data cannot be entirely ruled out. However, considering that the six main classes classified by the model (QM, QV, PD, PK, LF, and NF) occupy over 97% of the total area below 1000m elevation within the Jirisan 1 region, it is highly probable that these hatched areas are indeed mixed forests composed of these dominant species. Therefore, the model demonstrates its potential to contribute to a more precise understanding of forest resources by effectively generating high-resolution information on the detailed species distribution within mixed forests—details unavailable in the original map.

Conversely, the area presented as the bad case exemplifies instances where the model's predictions appear less accurate in reflecting the actual forest conditions compared to the original tree species map. When compared with the aerial photograph of the corresponding area, the boundary of the LK community delineated on the original map appears to align more closely with the actual vegetation boundary. The model's prediction exhibits a tendency to classify certain areas as LK, even where the aerial photograph suggests otherwise. This discrepancy is likely attributable primarily to the spatial resolution limitations of the satellite imagery used as input for the model. While the satellite imagery used in this study was significantly enhanced to 2.5 m/px resolution through SR, it still possesses considerably lower resolution compared to ultra-high-resolution aerial imagery (e.g., 0.25 m/px). Traditional methods for creating forest maps involve skilled personnel manually interpreting such ultra-high-resolution aerial photographs to delineate boundaries. Consequently, for delineating the boundaries of relatively distinct, single species stands like the LK community shown, the original map might occasionally offer greater accuracy than the model's output.

Nonetheless, as demonstrated by the good case, the model holds a distinct advantage: it can generate detailed distribution information for complex mixed forest areas—areas challenging for manual processing—thereby complementing the limitations of the

original map. Furthermore, it leverages the fine spectral information from satellite imagery (across visible and non-visible wavelengths) to potentially capture subtle vegetation differences—often indiscernible to the human eye—and classify them objectively. Thus, despite potential inaccuracies in delineating certain boundaries, the model remains a valuable and effective tool, offering the high-resolution, detailed information required for understanding and managing complex forest structures. Utilizing the original tree species map data and the high-resolution maps generated by this model in a complementary fashion could allow for effective acquisition of both accurate boundary delineation for single species stands and detailed internal composition information for mixed forests.

5.2. Dimensional Importance Comparison

This study aimed to quantitatively analyze the respective impacts of spectral, temporal, and spatial dimensions on model performance in the tree species classification task. To achieve this, we first defined the meaning of the information contained within each dimension and then designed experiments to analyze the influence of each dimension in isolation.

Spectral information refers to information obtainable through the relationships between different bands, encompassing characteristics like color, brightness, and saturation in both visible and non-visible light spectra. This spectral information reflects the physiological state or compositional differences of various vegetation types and has been considered a core element in previous remote sensing-based classification studies.

Spatial information relates to information derived from relationships with neighboring pixels. However, at the 2.5 m/pixel resolution used currently, capturing fine texture information originating from leaves or branches within individual trees is limited. Therefore, this study focused on analyzing whether the forest texture, formed by the arrangement of individual trees, is significant for tree species classification.

Temporal information encompasses information identified through relationships between different time points, including interactions between spectral and spatial dimensions changing along the time axis. This information, reflecting vegetation growth cycles or seasonal changes, can play a crucial role in tree species classification using multi-temporal data.

Based on these definitions and the established importance of spectral information, we designed our experimental setup. Since spectral information has been widely recognized as key for tree species classification in previous research, the condition of removing spectral information was excluded from our experiments. Thus, this study focused on the following three experimental conditions: (Case 1) Using only spectral information, (Case 2) Using integrated spectral and spatial information, (Case 3) Using integrated spectral and temporal information.

To implement these experimental conditions, specific techniques were employed to selectively exclude dimensional information. To remove spatial information, a mean blur technique was applied, replacing pixel values within a patch with their average value, thereby eliminating spatial texture information. In practice, all cases received input from the same spatial size (10m × 10m). However, Case 2, utilizing spatial information, used the detailed 2.5m resolution, while Case 1 and 3 simplified the same area to 10m resolution, intentionally blocking texture information.

Similarly, to remove temporal information, we adopted an approach where data from each time point was trained independently. That is, the model was configured to learn spectral and spatial information only on a per-time-point basis, preventing the learning of inter-timepoint relationships. The prediction results from the models trained for each time point were then ensembled using a soft voting method to derive the result. This

design strategy aims to leverage individual information from each time point as much as possible while excluding temporal relationships.

We now detail the specific configurations for each experimental case. Case 1 served as the baseline, using only spectral information to determine its standalone contribution to tree species classification. In this condition, independent spectral encoders were trained for each time point, followed by classification via a linear head. A total of 12 models were ensembled using soft voting for the final prediction. The entire model ensemble has approximately 325K parameters.

Next, Case 2 evaluated performance improvement by adding spatial information to spectral information. The objective was to assess whether spatial information, represented as forest texture, significantly contributes to tree species classification. For this, a structure was designed where spatial information for each band was learned using fully connected (FC) layers, after which representative values per band were extracted and used as input to the spectral encoder. FC layers were used instead of CNNs because the small input patch size allowed learning relationships between all pixels using FC layers without excessive parameter usage. A total of 12 models were ensembled using soft voting for the final prediction. The entire model ensemble has approximately 356K parameters.

Finally, Case 3 aimed to analyze the performance change pattern by adding temporal information to spectral information. For fair comparison with Cases 1 and 2, Case 3 also did not share the spectral encoder; each time point had its own separate spectral encoder. Internally, it operates very similarly to Case 1, but the final step involves flattening the encoded results from each time point and then using a linear layer to learn temporal Information at the last layer, instead of employing individual classification heads followed by soft voting. It has approximately 325K parameters. **Table 5** presents experimental results.

Table 5. The comparison of dimensional importances.

Case	Non-Forest Vs. Forest	Conifer Vs. Broadleaf	Intra-Conifer	Intra-Broadleaf	Overall Accuracy	Overall F1-score
Case 1	98.82	95.11	90.21	79.41	83.66	83.10
Case 2	98.82	94.43	89.02	79.11	82.65	82.06
Case 3	98.90	95.04	92.29	82.16	85.64	85.15

Analysis of the results presented in Table 5 revealed several key points. First, it was noteworthy that over 80% OA could be achieved using only spectral dimension information. This strongly suggests that spectral information plays a very crucial role in tree species classification. Indeed, observing forests from aerial photographs reveals color differences among tree species. For instance, conifers generally appear dark green, while broadleaf trees tend to have lighter colors. These differences become even more pronounced during the leaf-out and senescence periods. Utilizing Invisible Bands beyond visible light likely provides additional useful information for species classification.

Second, regarding spatial dimension information did not appear to significantly contribute to model performance improvement. In most tests, the performance of Case 2 was lower than Case 1. This indicates that spatial information acted as a hindrance to tree species classification. This suggests that, for the dataset and classes used in this study, differences in forest texture do not provide meaningful assistance for classification. This aligns with the conclusion from [20]

Third, temporal dimension information was found to significantly enhance the model's classification performance. This implies that considering changes over time, in

addition to the spectral information obtained at individual time points, plays an important role in tree species classification. Particularly, the performance improvement was pronounced in relatively difficult tasks such as Intra-Broadleaf and Intra-Conifer classification. This confirmed that for fine-grained classification within the same category, spectral information from a single time point alone is insufficient, and utilizing temporal information is essential.

Synthesizing these results leads to the conclusion that when designing a tree species classification model, it is paramount to ensure robust learning of spectral dimension information, and actively utilizing temporal information is a more effective approach than incorporating spatial information. These findings provide experimental justification for the multi-spectral, multi-temporal, single-pixel input approach adopted in this study's model design

5.3. Comparison with Previous Studies

This study proposes an integrated framework that targets existing national tree species maps for improvement while simultaneously utilizing them as weak labels and applies SR techniques to multi-temporal satellite imagery to generate high-resolution (2.5 m/px) tree species maps. This approach presents several key differentiators and contributions compared to existing tree species classification research.

First, it offers a practical solution to the challenges of label data acquisition and scalability. Acquiring accurate label data is a core task in supervised tree species classification, but large-scale data construction based on field surveys is often impractical for broad-area application due to immense costs and effort. Against this backdrop, many previous studies have utilized national tree species maps as a labeling basis. However, due to the inherent error potential within the maps themselves, some research attempted to improve accuracy by directly correcting or re-labeling them. Yet, such improvement efforts also demand significant amounts of time and effort, still limiting broad-area applicability. Furthermore, previous studies often acknowledged the errors in original tree species maps but lacked in-depth discussion on effective data sampling and utilization strategies to mitigate these errors. Such efficient data utilization strategies are essential for large-area applications. Therefore, instead of directly correcting the map, this study treats it as a 'weak label', proposes methods to effectively utilize the training data considering inherent errors, and validates its effectiveness. The proposed methodology establishes a foundation immediately applicable and scalable nationwide without additional labeling work, representing significant differentiation. This is a pragmatic approach from the perspective that sufficient quantity and diversity of data can sometimes be more crucial than perfect label accuracy in DL model training.

Second, it effectively overcomes the spatial resolution limitations of Sentinel-2 imagery. While Sentinel-2 data provides valuable multi-spectral and time-series information useful for tree species classification, its 10-20m spatial resolution constrains detailed species distribution mapping. The results from [16] and [20] were limited to the 10 m resolution level due to this constraint. Previous study [14] used both RapidEye (5 m) and Sentinel-2, upsampling the latter to 5 m, but this employed a traditional interpolation technique (nearest neighbor) and did not involve an advanced SR fusion process referencing high-resolution imagery (RapidEye). This study applies a state-of-the-art SR technique (Section 3.1), using PlanetScope (3 m) imagery as reference data, to Sentinel-2 imagery, ultimately constructing an ultra-high-resolution 2.5 m/px dataset. Utilizing this as input for the classification model actively addresses the spatial resolution issue. Furthermore, similar to [20], we perform pixel-level classification, independently analyzing and predicting the spectral-temporal information of each pixel, thereby eliminating the ambiguity that can arise in patch-based approaches (predicting dominant species in a patch vs. predicting the

center pixel). This enables the generation of detailed classification results that clearly represent complex mixed forest boundaries or small-scale stands.

Third, it successfully performs classification between QM and QV. Oak species are widely distributed in Korea (approx. 980 thousand ha [31]), with QM (414 thousand ha) and QV (235 thousand ha) being dominant species, accounting for about 66% of the total oak forest area [32]. However, these two species exhibit very similar spectral and ecological characteristics, making remote sensing-based classification challenging. Consequently, previous studies often grouped various oak species together as 'oak species' or treated QM and QV as a single class. This study overcomes this limitation by leveraging the expressive power of the 2.5m ultra-high-resolution multi-temporal dataset and DL models to successfully classify QM and QV with high accuracy (over 81%, see **Figure 8**). This suggests effective learning of more subtle spectral-temporal characteristic differences not attempted in prior work. Furthermore, this holds significant meaning as it demonstrates the potential applicability of the proposed methodology to classification problems involving other spectrally similar tree species.

Comprehensively, this study presents an integrated and practical framework capable of substantially improving the spatial resolution of existing tree species maps by organically combining ultra-high-resolution techniques and pixel-level DL classification. Particularly, the result of distinguishing the challenging major domestic oak species, QM and QV, with high accuracy validates the superiority and potential of the proposed methodology. This constitutes a significant academic and practical contribution compared to preceding studies.

5.4. Constraint and Future Works

Due to limitations in PlanetScope data usage and available computing resources, this study could not construct a large-scale dataset. However, acquiring additional data from diverse time periods and geographical locations for training is essential to enhance temporal and spatial generalization performance. This is expected to enable the construction of models with more robust generalization capabilities.

Furthermore, this study analyzed only five tree species due to insufficient training data for a few minor species. If sufficient data for other major species such as Pinus rigida, Quercus acutissima, and Chamaecyparis obtusa could be secured, it would be possible to build a model capable of distinguishing more classes. Given that this study confirmed the feasibility of classifying QM and QV, which have very similar spectral and ecological characteristics, with over 81% accuracy, even more effective performance can be anticipated for classifying species with greater differences.

A key advantage of creating tree species maps using satellite imagery is the ease of updating them in response to dynamic changes. However, the model proposed in this study uses 12 time points of satellite imagery as input, and there is a possibility that the model's prediction might become ambiguous if changes occur at specific time points. For instance, if an area was initially a forest but later changed to non-forest due to factors like logging, it is unclear how the model would interpret this. This aspect was not considered in the current study, suggesting a potential vulnerability of the model to intermediate changes.

Two approaches can address this. The first is a data-driven approach: providing the model with training data that includes various changes, thereby inducing it to correctly predict the class after the change. However, this method is limited by the need for sufficient data covering change areas. Events causing forest changes like logging, afforestation, and wildfires occur infrequently and vary widely in their patterns, making the construction of large-scale time-series data encompassing these changes potentially difficult in practice. Therefore, it needs to be pursued in parallel with the second approach, which is

model driven. This approach involves defining an inter-timepoint distance metric to quantify differences between satellite images at various times. If a significant change is detected at a specific time point, the model is designed to handle that time point separately. This allows for predictions reflecting the post-change information for areas where changes have occurred. By simultaneously applying both data- and model-centric approaches, a tree species classification model sensitive to the dynamic changes in forests can be implemented. This is proposed as a future task.

Additionally, while this study concluded that forest texture is not highly useful at the currently used satellite image resolution, there is a possibility that texture information could play a significant role if higher-resolution image data were utilized. Future research should explore incorporating models that sufficiently utilize forest texture information using higher-resolution imagery.

6. Conclusions

This study proposed a novel methodology for generating 2.5m/pixel high-resolution tree species classification maps by applying DL-based models to multi-temporal, multi-spectral Sentinel-2 and PlanetScope satellite imagery. Notably, it presented a method to create more precise, high-resolution tree species maps by utilizing existing noisy, low-resolution tree species map data as labeling data through appropriate sampling, even without high-resolution ground truth labels. This method holds significant practical value as it can be easily scaled to most locations in South Korea where tree species map data currently exists.

Applying this methodology to Sobaeksan and Jirisan National Parks, the study demonstrated its practical applicability, achieving over 85% OA and over 81% performance even for difficult-to-distinguish species like QM and QV. Furthermore, the study provided experimental evidence supporting the model design choices. It confirmed that the spectral dimension contributes most significantly to tree species classification performance and combining it with temporal information leads to meaningful performance enhancement, whereas spatial information might hinder model learning. This supports the effectiveness of a multi-spectral, multi-temporal, single pixel-based model design for tree species classification. Additionally, it was experimentally verified that a simple MLP structure exhibited superior classification performance compared to models with structurally incorporated inductive biases.

However, this study faced limitations due to constraints on PlanetScope data usage and computing resources, resulting in insufficient spatio-temporal diversity in the data and analysis being conducted on only a small number of species. Future work involving the construction and training on extensive time-series data from various locations could enable the classification of more species and is expected to yield superior generalization performance.

In conclusion, this study presents a practical and scalable method for generating high-resolution tree species classification maps based on multi-spectral, multi-temporal satellite imagery using low-resolution tree species map data. Its effectiveness has been demonstrated through real-world case studies, highlighting its academic and practical significance.

Author Contributions: Writing—original draft, T.C., S.J.; Writing—review & editing, B.C.; Supervision, S.P. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by …

Data Availability Statement: PlanetScope data can be downloaded from PlanetScope Education and Research (E&R) Program (https://api.planet.com, accessed on 18 December 2024). The Sentinel-2 satellite data were provided as part of the European Union's Copernicus Programme and are

publicly available from the Copernicus Data Space Ecosystem (https://dataspace.copernicus.eu, accessed on 18 December 2024).The tree species map of The Republic of Korea can be downloaded in https://map.forest.go.kr (accessed on 18 December 2024).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Cavers, S.; Cottrell, J. The basis of resilience in forest tree species and its use in adaptive forest management in Britain. *Forestry: An International Journal of Forest Research* **2015**, *88*, 13-26, doi:10.1093/forestry/cpu027.
2. Wang ChuanKuan, W.C. Biomass allometric equations for 10 co-occurring tree species in Chinese temperate forests. **2006**, doi:10.1016/j.foreco.2005.10.074.
3. Návar, J. Allometric equations for tree species and carbon stocks for forests of northwestern Mexico. *Forest ecology and Management* **2009**, *257*, 427-434, doi:10.1016/j.foreco.2008.09.028.
4. Mouret, F.; Morin, D.; Martin, H.; Planells, M.; Vincent-Barbaroux, C. Toward an operational monitoring of oak dieback with multispectral satellite time series: A case study in Centre-Val de Loire region of France. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2023**, *17*, 643-659, doi:10.1109/JSTARS.2023.3332420.
5. MacKenzie, W.H.; Mahony, C.R. An ecological approach to climate change-informed tree species selection for reforestation. *Forest Ecology and Management* **2021**, *481*, 118705, doi:10.1016/j.foreco.2020.118705.
6. Taylor, D.; Kent, M.; Coker, P. Vegetation description and analysis: a practical approach. *The Geographical Journal* **1993**, *159*, 237. ISBN 0471948101.
7. Crisco, W.A. *Interpretation of aerial photographs*; US Department of the Interior, Bureau of Land Management: 1983; Volume 287.
8. Kim, K.; Lee, S. Distribution of major species in Korea (based on 1: 5000 forest type map). *National Institute of Forest Science: Seoul, Republic of Korea* **2013**, 15.
9. Fassnacht, F.E.; Latifi, H.; Stereńczak, K.; Modzelewska, A.; Lefsky, M.; Waser, L.T.; Straub, C.; Ghosh, A. Review of studies on tree species classification from remotely sensed data. *Remote sensing of environment* **2016**, *186*, 64-87, doi:doi: 10.1016/j.rse.2016.08.013.
10. Drusch, M.; Del Bello, U.; Carlier, S.; Colin, O.; Fernandez, V.; Gascon, F.; Hoersch, B.; Isola, C.; Laberinti, P.; Martimort, P. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote sensing of Environment* **2012**, *120*, 25-36, doi:10.1016/j.rse.2011.11.026.
11. Breiman, L. Random forests. *Machine learning* **2001**, *45*, 5-32, doi:10.1023/A:1010933404324.
12. Belgiu, M.; Drăguţ, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS journal of photogrammetry and remote sensing* **2016**, *114*, 24-31, doi:10.1016/j.isprsjprs.2016.01.011.
13. Cha, S.; Lim, J.; Kim, K.; Yim, J.; Lee, W.-K. Uncovering the Potential of Multi-Temporally Integrated Satellite Imagery for Accurate Tree Species Classification. *Forests* **2023**, *14*, 746, doi:10.3390/f14040746.
14. Lim, J.; Kim, K.-M.; Jin, R. Tree species classification using hyperion and sentinel-2 data with machine learning in South Korea and China. *ISPRS International Journal of Geo-Information* **2019**, *8*, 150, doi:10.3390/ijgi8030150.
15. Maillard, P. Comparing texture analysis methods through classification. *Photogrammetric Engineering & Remote Sensing* **2003**, *69*, 357-367, doi:10.14358/PERS.69.4.357.
16. Lim, J.; Kim, K.-M.; Kim, E.-H.; Jin, R. Machine learning for tree species classification using sentinel-2 spectral information, crown texture, and environmental variables. *Remote Sensing* **2020**, *12*, 2049, doi:10.3390/rs12122049.
17. Zhong, L.; Dai, Z.; Fang, P.; Cao, Y.; Wang, L. A review: Tree species classification based on remote sensing data and classic deep learning-based methods. *Forests* **2024**, *15*, 852, doi:10.3390/f15050852.
18. Kowalski, K.; Senf, C.; Hostert, P.; Pflugmacher, D. Characterizing spring phenology of temperate broadleaf forests using Landsat and Sentinel-2 time series. *International Journal of Applied Earth Observation and Geoinformation* **2020**, *92*, 102172, doi:10.1016/j.jag.2020.102172.

19. Xi, Y.; Ren, C.; Tian, Q.; Ren, Y.; Dong, X.; Zhang, Z. Exploitation of time series sentinel-2 data and different machine learning algorithms for detailed tree species classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2021**, *14*, 7589-7603, doi:10.1109/JSTARS.2021.3098817.
20. Mouret, F.; Morin, D.; Planells, M.; Vincent-Barbaroux, C. Tree Species Classification at the Pixel Level Using Deep Learning and Multispectral Time Series in an Imbalanced Context. *Remote Sensing* **2025**, *17*, 1190, doi:10.3390/rs17071190.
21. Johnson, J.M.; Khoshgoftaar, T.M. Survey on deep learning with class imbalance. *Journal of big data* **2019**, *6*, 1-54, doi:10.1186/s40537-019-0192-5.
22. Team, P. Planet application program interface: In space for life on Earth. *Planet* **2017**. https://api.planet.com.
23. Latte, N.; Lejeune, P. PlanetScope radiometric normalization and sentinel-2 super-resolution (2.5 m): A straightforward spectral-spatial fusion of multi-satellite multi-sensor images using residual convolutional neural networks. *Remote Sensing* **2020**, *12*, 2366, doi:10.3390/rs12152366.
24. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning representations by back-propagating errors. *nature* **1986**, *323*, 533-536, doi:10.1038/323533a0.
25. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Advances in neural information processing systems* **2017**, *30*, doi:10.48550/arXiv.1706.03762.
26. Loshchilov, I.; Hutter, F. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101* **2017**, doi:10.48550/arXiv.1711.05101.
27. Loshchilov, I.; Hutter, F. Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983* **2016**, doi:10.48550/arXiv.1608.03983.
28. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the International conference on machine learning, 2015; pp. 448-456. doi:10.48550/arXiv.1502.03167
29. Ramachandran, P.; Zoph, B.; Le, Q.V. Searching for activation functions. *arXiv preprint arXiv:1710.05941* **2017**, doi:10.48550/arXiv.1710.05941.
30. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research* **2014**, *15*, 1929-1958, doi:10.5555/2627435.2670313.
31. Lee, S.J.; Yim, J.S.; Son, Y.M.; Son, Y.; Kim, R. Estimation of forest carbon stocks for national greenhouse gas inventory reporting in South Korea. *Forests* **2018**, *9*, 625, doi:10.3390/f9100625.
32. Park, J.; Chung, S.; Kim, S.; Lee, S. Estimation of site index curve for 6 oak species in Korea. *Journal of Agriculture & Life Science* **2020**, *54*, 27-33, doi:10.14397/jals.2020.54.3.27.