

Tree Species Classification at the Pixel Level Using Super-resolved Satellite Imagery with Band-wise Normalization

Choe Taebin¹, Jeon Seungpyo²

Abstract: 정확한 수종 분포 정보는 산림 관리에 중요하나, 위성 영상은 해상도 및 변동성 한계를 지닙니다. 본 연구는 다중 위성 데이터(Sentinel-2, PlanetScope) 융합과 딥러닝을 이용해 한국 온대 혼효림(소백산, 지리산)의 고해상도(2.5m) 수종 분류 정확도 향상을 목표로 했습니다. 이를 위해 12 개 시점의 Sentinel-2/PlanetScope 데이터를 활용하여 PlanetScope 기반 초해상화로 2.5m 시계열 데이터셋을 구축하고, 영상 밝기 변화와 그림자 효과 완화를 위해 독자적인 밴드별 정규화(Band-wise Normalization) 기법을 제안 및 적용했습니다. 결과적으로, 제안된 밴드별 정규화는 모델의 강건성을 유의미하게 향상시키는 것을 확인했습니다. 주요 5 개 수종(소나무, 잣나무, 낙엽송, 신갈나무, 굴참나무)과 비산림 지역 분류를 위해 Random Forest, MLP, Transformer 등 다양한 머신러닝 모델 성능을 교차 검증으로 비교 분석했습니다. 수종 분류 정확도에는 공간 정보(2.5m 해상도 기준)보다 분광 및 다중 시계열 정보가 훨씬 중요했으며, 특히 여름철(7 월) 데이터가 판별력 향상에 기여했습니다. 다양한 모델 중 Spectral-Temporal 패턴 학습에 최적화된 Transformer 모델이 가장 우수한 분류 성능(전체 정확도 85.52%)과 파라미터 효율성을 보였습니다. 본 연구는 다중 위성 데이터 융합, 초해상화, 제안된 정규화 기법, 그리고 최적화된 Transformer 모델을 효과적으로 결합하여 복잡한 산림 환경에서도 고해상도의 정밀한 수종 지도를 제작할 수 있는 실용적이고 강건한 방법론을 성공적으로 제시했습니다.

Keywords: 수종 분류; 딥러닝; 원격탐사; Sentinel-2; PlanetScope; 데이터 융합; 다중 시계열 분석; 초해상화; Transformer 모델; 산림 모니터링

1. Introduction

정확한 수종 분포 정보는 생물 다양성 보전 및 평가 [1], 야생동물 서식지 분석, 지속 가능한 산림 경영 및 탄소 저장량 추정, 그리고 산불 위험 관리 등 광범위한 생태 및 환경 관리 분야의 기초 자료로 활용됩니다. 예를 들어, 특정 수종의 분포 변화는 기후 변화의 영향을 파악하는 지표가 될 수 있으며, 산림 병해충 확산 예측 및 방제 전략 수립에도 핵심적인 정보를 제공합니다.

전통적으로 수종 정보는 현장 조사를 통해 수집되어 왔습니다. 이 방식은 높은 정확도를 보장하지만, 넓은 지역, 특히 접근이 어려운 산악 지형을 대상으로 하기에는 많은 시간과 인력, 비용이 소요되는 한계가 있습니다. 대한민국 국토의 약

63%가 산림이고 상당 부분이 험준한 산악 지형임을 고려할 때 [7], 전국 단위의 상세하고 시의성 있는 수종 지도를 현장 조사만으로 구축하고 갱신하는 것은 현실적으로 어렵습니다. 이러한 한계를 극복하기 위해 지난 수십 년간 원격탐사(Remote Sensing) 기술이 효과적인 대안으로 주목받아 왔습니다.

초기 원격탐사 기반 수종 연구는 항공 라이다(LiDAR)나 항공 초분광/다중분광 영상을 주로 활용했습니다 [8, 9]. 항공 기반 센서는 매우 높은 해상도나 상세한 분광 정보를 제공하지만, 데이터 취득 비용, 넓은 지역 커버리지, 기상 조건 등의 제약으로 인해 국가 규모의 주기적인 모니터링에는 한계가 있습니다.

이러한 배경에서, 무료로 제공되며 주기적인 관측이 가능한 Sentinel-2 위성 데이터의 활용이 크게 증가했습니다 [12, 13]. Sentinel-2 는 수종 분류에 유용한 다중 분광 정보를 제공하지만, 공간 해상도(10-20m)의 한계는 여전히 존재합니다. 이를 보완하기 위해 PlanetScope 와 같은 고해상도 위성과의 데이터 융합 및 초해상화(Super-Resolution) 기법 적용이 시도되고 있습니다.

수종 분류 방법론 또한 머신러닝을 거쳐 최근에는 복잡한 시계열 데이터 패턴 학습에 강점을 보이는 딥러닝(Deep Learning) 모델로 발전하고 있습니다. 하지만 여전히 어떤 정보(분광, 공간, 시간) 조합이 가장 효과적인지, 그리고 산악 지형의 그림자나 영상 강도 변화와 같은 데이터 변동성에 어떻게 강인하게 대처할 것인지는 중요한 연구 과제로 남아있습니다.

이에 본 연구는 한국의 대표적인 온대 혼효림(소백산, 지리산 국립공원)을 대상으로, 다중 시계열 Sentinel-2 및 PlanetScope 데이터를 융합하고 딥러닝 모델을 적용하여 고해상도(2.5m) 수종 분류를 수행하고자 합니다. 구체적으로, 초해상화 기법을 적용하고 분광/공간/시간 정보의 중요도를 평가하며, 데이터 강건성 향상을 위한 밴드별 정규화 기법을 제안 및 검증합니다. 또한, Transformer 등 다양한 딥러닝 모델의 성능을 비교하여 최적 모델을 선정하고, 최종적으로 주요 5 개 수종 및 비산림 지역에 대한 정밀 수종 지도를 제작하는 것을 목표로 합니다. 본 연구 결과는 시간 정보의 중요성, 제안된 정규화 기법의 효과, 그리고 Transformer 기반 모델의 우수성을 입증하며, 복잡한 산림 환경에서 고해상도 수종 지도를 효과적으로 제작할 수 있는 가능성을 제시합니다.

2. Materials and Methods

2.1. Study Sites and Data Timeframes

2.1.1. Study Sites

이 연구는 한국의 대표적인 산림 지역인 소백산 국립공원과 지리산 국립공원에서 수행되었다. 두 지역은 해발고도의 범위가 넓고 다양한 수종이 분포하고 있어 수종 분류 연구에 적합한 환경을 제공한다. 또한, 국립공원으로 지정되어 보호·관리되고 있어 인위적 간섭이 비교적 적고, 장기적인 생태 연구에 적합한 장소이다.

소백산 국립공원에서는 가로 9km, 세로 9km 로 총 81km²의 지역을 연구 대상으로 선정하였으며, 위도와 경도 범위는 (36°93'N – 37°02'N, 128°44'E – 128°54'E)로 남한의 중위도에 위치한다. 이 지역의 해발고도는 250m 이하에서 1400m 이상까지 다양하며, 주요 수종으로는 신갈나무 (*Quercus mongolica*, QM) 와 굴참나무

(*Quercus variabilis*, QV) 등의 참나무류가 우세하다. 그 외에도 소나무 (*Pinus densiflora*, PD), 낙엽송 (*Larix kaempferi*, LK), 잣나무 (*Pinus koraiensis*, PK), 포플러(*Populus* spp.)등이 분포하고 있다.

지리산 국립공원에서는 가로 18km, 세로 9km 로 총 162km²의 지역 데이터를 사용하였으며, 위도와 경도 범위는 (35°29'N – 35°38'N, 127°59'E – 127°79'E)이다. 이 지역은 남한지역에서 상대적으로 저위도에 위치하며, 해발고도는 350m 이하에서 1900m 이상까지 분포한다. 주요 수종으로는 신갈나무와 굴참나무 등의 참나무류가 주로 분포하며, 고산지대에는 구상나무(*Abies koreana*) 군락이 존재한다. 또한, 소나무, 잣나무, 낙엽송 등이 함께 서식하고 있다.

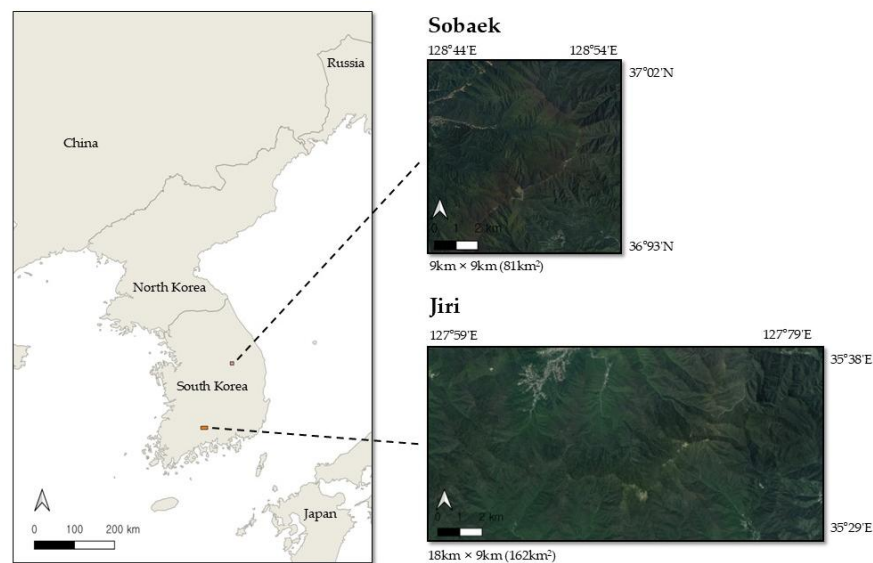


Figure 1. Study Sites

2.1.2. Data Timeframes

본 연구에서는 사용할 라벨링 데이터의 시점이 2022 년 기준이었기 때문에, 2022 년의 위성 데이터를 활용하였으며 총 12 개 시점의 영상을 활용하였다. 기본적으로 1 개월 간격으로 데이터를 수집하였으며, 수종 간 개엽 시기(leaf-out timing) 및 낙엽 시기(leaf senescence timing)에서 나타나는 수종별 특성 차이를 보다 정밀하게 포착하기 위해 특정 기간 동안 15 일 간격으로 추가 데이터를 확보하였다. 4 월 1 일부터 6 월 1 일까지는 개엽 시기에 나타나는 수종 간 차이를 더 잘 포착하기 위해 15 일 간격으로 데이터를 수집하였으며, 10 월 1 일부터 11 월 1 일까지는 낙엽 시기에 관찰할 수 있는 수종 간 차이를 정확히 포착하기 위해 동일한 방식으로 데이터를 추가 확보하였다. 결과적으로 본 연구에서 사용된 데이터 시점은 다음과 같다.

0201, 0301, 0401, 0415, 0501, 0515, 0601, 0701, 0901, 1001, 1015, 1101

단, 기상 조건 등으로 인해 해당 시점의 데이터를 가져오지 못할 경우 가용한 가장 가까운 시점의 데이터를 사용하였다.

2.2. Pre-Processing of Satellite Imageries

본 연구에서는 Sentinel-2 및 PlanetScope 위성 데이터를 사용하였다. Sentinel-2는 Copernicus 프로그램의 일환으로 운영되는 다중 분광(Multispectral) 위성이며, 총 13개의 분광 밴드를 제공한다. 본 연구에서는 Sentinel-2 Level-2A 제품을 활용하였으며, Sentinel-2A (S2A) 및 Sentinel-2B (S2B) 위성의 데이터를 혼용하여 사용하였다.

Sentinel-2 위성은 다양한 분광 밴드를 제공하는 장점이 있으며, 특히 수중 분류에 중요한 밴드들(RedEdge, NNIR, SWIR)을 포함하고 있어 본 연구에 적합하다. 이러한 밴드들이 위성영상을 이용한 수중 분류 Task에서 중요한 역할을 한다는 점은 다양한 선행 연구를 통해 입증된 바 있다.

한편, PlanetScope 위성 데이터는 150 개 이상의 마이크로 위성을 이용하여 고빈도로 높은 해상도의 데이터를 제공한다는 장점이 있다. 본 연구에서는 고해상도의 수중 지도 생성을 목표로 하였기 때문에, Sentinel-2 보다 높은 공간 해상도를 제공하는 위성 데이터를 추가로 활용할 필요가 있었다. 이중, PlanetScope 데이터는 교육 및 연구 목적으로 월 3,000 km² 한도 내에서 무료로 제공되므로 연구에 적절히 활용할 수 있었다.

본 연구에서 사용한 위성별 활용 밴드 및 공간 해상도를 Table X에 정리하였다. 이 중, Sentinel-2의 B02, B03, B04, B08 밴드는 Super-Resolution 과정에만 사용되었으며, 최종적인 수중 분류 모델의 입력 데이터로는 사용하지 않았다.

Table 1. Satellite Information

Satellite	Spatial Resolution (m)	Spectral Bands	Center Wavelength (nm)	Band Width (nm)
Sentinel-2	10	Blue (B4)	492.1	66
		Green (B3)	559.0	36
		Red (B2)	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
Planet	3	Blue	20	485
		Green	90	545
		Red	80	630
		Near-infrared	80	820

2.2.1 band wise Normalization

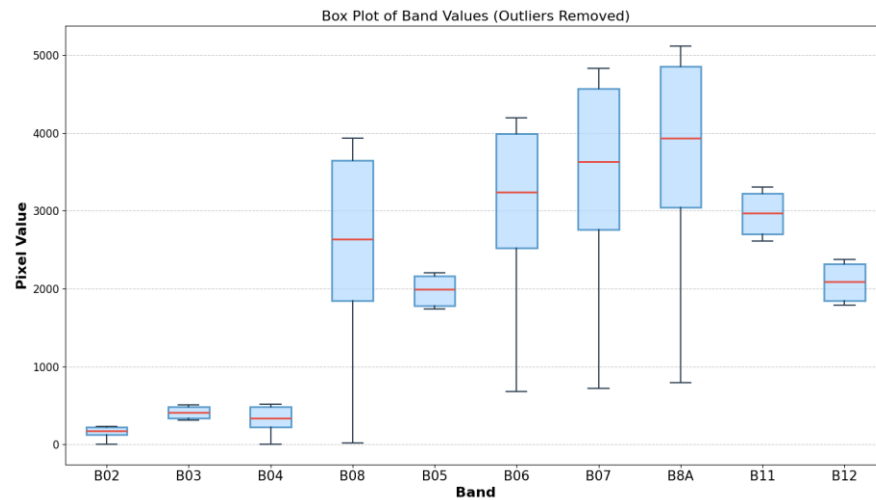


Figure 2. Distribution of values per band

우리는 수종 분류(Task)에서 각 밴드의 절대적인 강도(intensity)보다는 픽셀 내 밴드 간 상대적인 비율(즉, 색 비율)이 더욱 중요한 특징이라고 판단하였고 이러한 사전 지식을 모델에 반영하기 위해, 밴드별 정규화를 수행하였다. 이때, 가시광선 밴드(Visible bands)와 비가시광선 밴드(Invisible bands) 간의 스케일 차이가 크기때문에(Appendix 1 참고), 두 그룹을 함께 정규화할 경우 가시광선 밴드의 값이 상대적으로 작아져 영향력이 감소할 우려가 있다. 따라서, 각 그룹을 별도로 나누어 정규화를 적용하였다.

$$Band_{visible} = \{B2, B3, B4\}, Band_{invisible} = \{B5, \dots, B12\}$$

$$i := 2000 \times |Band_{visible}| \times \frac{\{i\}}{\sum (Band_{visible})}, i \in Band_{visible}$$

$$j := 2000 \times |Band_{invisible}| \times \frac{\{j\}}{\sum (Band_{invisible})}, j \in Band_{invisible}$$

이와 같은 정규화 방식은 강력한 inductive bias 를 모델에 부여하는 역할을 한다. 즉, 모델이 특정한 특징(픽셀 내 각 밴드 값의 상대적 비율)에 집중하도록 유도하여, 훈련 데이터에 존재하는 불필요한 특성에 과적합되는 것을 방지하고 일반화 성능(generalization performance)을 향상시킬 수 있다. 이 방식을 이용하면 모델의 절대적인 intensity 값에 대한 의존도를 줄이고, 시간적(다른 시기 적용) 및 공간적(다른 지역 적용) 일반화 성능이 향상될 것으로 기대된다. 이는 위성 데이터의 전반적인 intensity 값이 지역과 시점에 따라 크게 달라진다는 점을 고려했을 때, 매우 합리적인 설계라고 볼 수 있다. 추가적으로, 산악 지형으로 인해 발생하는 그림자 문제 역시 이러한 정규화 방법을 통해 완화할 수 있을 것으로 판단하였다. 이는 band wise normalization 을 이용하면 태양 복사 에너지의 감소로 인한 spectral 차원의 왜곡을 어느정도 보정할 수 있기 때문이다.

우리는 Band-wise Normalization 의 효과를 검증하기 위해, 전반적인 intensity 값에대한 sturdiness 와 그림자 지역에서의 분류 성능을 비교하는 실험을 수행하였다. 예상대로, Band-wise Normalization 을 적용한 경우 두 경우 모두 성능이 향상됨을 확인할 수 있었으며, 이에 대한 상세한 결과는 3.1.1 에서 논의한다.

2.2.2 Super Resolution

앞서 언급했듯이, Sentinel-2 데이터는 다양한 밴드를 제공한다는 장점이 있으나, 공간 해상도가 상대적으로 낮다는 단점이 있다. 본 연구에서는 고해상도 수종 지도(high-resolution forest classification map) 생성을 목표로 하므로, 더 높은 해상도의 데이터가 필요했다.

또한, 우리는 Forest Texture(숲의 질감) 정보의 필요성을 검증하고자 하였으나, Sentinel-2 데이터의 해상도가 20m/pixel 이기 때문에, PlanetScope 수준의 해상도에서는 유의미한 공간 정보(spatial information)가 존재하더라도, Sentinel-2 의 낮은 해상도가 이를 학습하는 데 방해요소로 작용할 가능성이 있다고 판단하였다. 따라서, Sentinel-2 데이터에 대해 Super-Resolution 기법을 적용하여 해상도를 향상시키는 접근 방식을 채택하였다.

Super-Resolution 은 Wald's Protocol 을 기반으로 수행되었으며, 방법론은 [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 (2020)]연구를 참고하였다. 본 연구에서 수행한 Super-Resolution 과정은 일반적인 Super-Resolution 작업과 차이가 있다. 기존 Super-Resolution 문제와 달리, 우리는 이미 고해상도의 특정 밴드 데이터를 보유하고 있으며, 이를 활용할 수 있다는 점에서 보다 향상된 결과를 기대할 수 있다. 개략적인 Super-Resolution 과정은 다음과 같다.

1. PlanetScope 데이터(3m/pixel)를 2.5m/pixel 로 보간(interpolation) (Bi-cubic interpolation)
2. PlanetScope 데이터와 Sentinel-2 데이터 모두, 1/8 해상도로 다운샘플링
3. Residual Convolutional Neural Network(RCNN)를 활용하여 $\times 8$ 해상도 복원 모델 학습
4. 학습된 모델을 다운 샘플링 되지 않은 Sentinel-2 데이터(20/px)에 적용하여 2.5m/pixel 해상도의 이미지 생성

Super-Resolution 결과는 Figure3 에서 제시된다. 자세한 방법론과 구현에 대한 사항은 해당 논문을 참고하길 바란다.

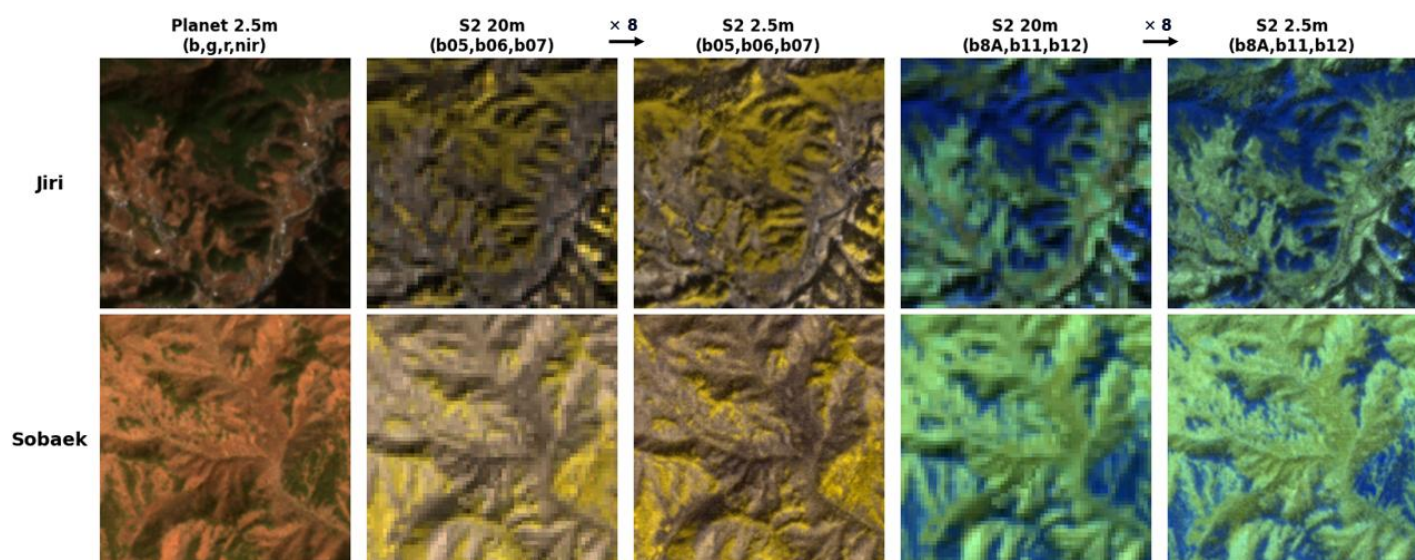


Figure 3. Super-Resolution Results**2.3. Label Data**

본 연구에서는 산림청에서 제공하는 임상도(Forest Type Map) 데이터를 활용하여 학습을 위한 레이블을 생성하였다. 우리나라에 가장 흔하게 분포하는 침엽수종인 소나무 (*Pinus densiflora*, PD), 잣나무 (*Pinus koraiensis*, PK), 낙엽송(*Larix kaempferi*, LK) 과 대표적인 활엽수종인 신갈나무 (*Quercus mongolica*, QM), 굴참나무(*Quercus variabilis*, QV) 총 5 가지 수종을 선정하였으며, 여기에 비산림(Non-Forest) 지역을 추가하여 총 6 개 클래스로 구분하였다. 임상도 데이터에서 해당 클래스로 분류된 픽셀을 선별하여 학습 데이터로 활용하였다.

임상도에는 기타 침엽수림, 기타 활엽수림, 기타 참나무림, 침·활 혼효림 등 단일 수종으로 매핑할 수 없는 지역이 상당수 포함되어 있다. 본 연구에서는 이러한 혼합 수종 지역을 제외하고, 단일 수종으로 분류된 픽셀만을 학습 데이터로 활용하였다. 이후, 학습된 모델을 혼합 수종 지역에 적용하여, 기존 ground truth 데이터인 임상도를 뛰어넘는 보다 세밀한 수종 지도를 생성하는 것이 본 연구의 목표 중 하나이다.

레이블 데이터를 선정하는 과정에서 데이터 간의 연관성을 최소화하기 위해 최소 2 픽셀 이상의 간격을 가진 샘플만을 사용하였다. 또한, 데이터 불균형 문제를 해결하기 위해 신갈나무, 소나무 등 일부 수종에서 과도하게 많은 샘플이 포함되지 않도록 균형 조정 sampling 작업을 수행하였다. 이때 각각의 sample 들이 최대한 독립적인 정보를 가질 수 있도록 샘플링 하였다. 최종적으로 선정된 학습 데이터의 개수는 Table 2 에 제시된다.

Table 2. Number of training samples per class

Species Code	Species	Definition	Number Of Data	Data Ratio(%)
NF		Non-Forest	42,691	18.3
PD	<i>Pinus Densiflora</i>	Tree species of the genre <i>Pinus</i>	35,855	15.4
PK	<i>Pinus Koraiensis</i>	Tree species of the genre <i>Pinus</i>	27,993	12.0
LK	<i>Larix Kaempferi</i>	Tree species of the genre <i>Larix</i>	44,162	18.9
QM	<i>Quercus Mongolica</i>	Tree species of the genre <i>Quercus</i>	45,700	19.6
QV	<i>Quercus Variabilis</i>	Tree species of the genre <i>Quercus</i>	37,051	15.9

2.4. Data Split

Train-Validation-Test 를 효과적으로 분리하기 위해, 각 연구 지역을 100 개의 격자(박스)로 분할하였다. 이 과정에서 지리산 지역은 두 개의 큰 정사각형 영역으로, 소백산 지역은 하나의 큰 정사각형 영역으로 설정한 후, 각각을 100 개의 작은 박스로 세분화하였다. 이를 통해 공간적으로 분리된 학습, 검증, 테스트 데이터를 구성함으로써 모델의 일반화 성능을 보다 정확하게 평가할 수 있도록 하였다.

• **Test Data**

각 지역에서 클래스 불균형이 심하지 않으며 연구 대상 수종(타겟 클래스)이 충분히 포함된 박스 2 개씩을 선정하여 Test 데이터로 활용하였다. 본 연구에서는 하이퍼파라미터 튜닝을 수행하지 않기 때문에 Validation 데이터만으로도 모델의

일반화 성능을 검증할 수 있었다. 따라서, Test 데이터는 전체 데이터의 약 2%만을
할당하였으며, 이는 추후 정성 평가 등의 용도로 활용될 예정이다.

- Train-Validation Data

Test 데이터를 제외한 나머지 박스들은 5-Fold Cross-Validation 을 위해 5 개의
Fold 로 분할하였다. 각 Fold 에서 Validation 데이터는 서로 겹치지 않는(mutually
exclusive) 방식으로 설정하였다. 즉, 총 5 가지 Train-Validation 조합을 생성하여
실험을 수행하였다.

구체적인 데이터 분할 구조는 Figure 4 에 제시된다.

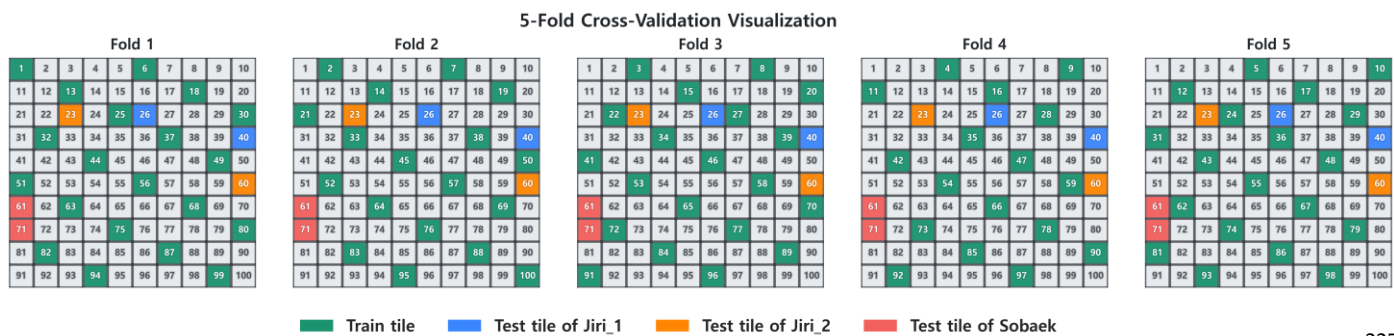


Figure 4. Dataset split layout

2.5. Evaluation Metrics

모델 개발 과정에서는 Ground Truth 데이터와의 일치도를 평가하기 위해 정량적인
Evaluation Metrics 를 활용하였다. 그러나 최종 모델 테스트에서는 현재 제공된
임상도(Ground Truth) 이상의 성능을 달성하는 것이 목표였기 때문에, 단순히
임상도와 일치도를 평가하는 것에 그치지 않았고, 전문가의 도움을 받아 모델이
생성한 수종 지도를 정성적으로 평가함으로써, 실제 수종 분포를 더 정확하게
반영하는지 검증하였다.

2.5.1. Quantitative Evaluation

본 연구에서는 모델의 성능을 보다 세밀하게 분석하고 직관적으로 이해하기 위해 네
가지 세분화된 분류 과제(classification tasks)를 정의하여 평가를 수행하였다. 각 분류
과제는 다음과 같다.

- Non-Forest vs. Forest: 비산림지(Non-Forest)와 산림지(Forest)를 구분하는 이진
분류(binary classification).
- Conifer vs. Broadleaf: 산림지로 올바르게 분류된 영역을 기준으로,
침엽수(conifer)와 활엽수(broadleaf)를 구분하는 이진 분류.
- Intra-Conifer Classification: 침엽수로 올바르게 분류된 영역을 기준으로
소나무(PD), 잣나무(PK), 낙엽송(LK)을 분류하는 다중 클래스 분류(multi-class
classification).
- Intra-Broadleaf Classification: 활엽수로 올바르게 분류된 영역을 기준으로
신갈나무(QM)와 굴참나무(QV)를 분류하는 이진 분류.

각 4 가지 분류 과제에 대해 Averaged Precision 과 Averaged Recall 을 평가 지표로
활용하였으며, 이를 통해 총 8 개의 정량적 평가 지표를 기반으로 모델 성능을
분석하였다.

본 연구에서는 Weighted Averaging 이 아닌 Macro-Averaging 을 사용하였다. 그 이유는 각 클래스 간 성능 편차를 보다 균형적으로 평가하기 위함이며, 특히 클래스 불균형이 모델 성능 평가에 미치는 영향을 최소화하고자 하였다.

2.5.2. Qualitative Evaluation

앞서 언급한 바와 같이, 본 연구의 주요 목표중 하나는 기존 임상도의 한계를 극복하고 보다 정밀한 수종 지도를 생성하는 것이다. 기존 임상도는 제작 과정에서 다양한 오차가 포함될 가능성이 있으며, 이로 인해 모델의 성능을 단순히 임상도와 정합성(correctness)만으로 평가하는 것은 적절하지 않다. 특히, 모델이 임상도의 오차를 그대로 학습하거나, 오히려 더 정교한 정보를 제공할 수 있음에도 불구하고, 단순히 기존 임상도와 일치 여부에 따라 성능이 과소평가되거나 과대평가될 위험이 존재한다. 따라서 모델의 성능을 보다 신뢰성 있게 평가하기 위해, 임상도와 정합성 분석뿐만 아니라 추가적인 보완 평가가 필요하다.

이에 본 연구에서는 보다 신뢰성 높은 평가를 수행하기 위해 전문가의 검토를 거친 대체 라벨링 데이터 구축 및 정성적 평가(qualitative evaluation) 를 진행하였다. 이를 위해 국토지리정보원에서 제공하는 고해상도 항공사진 데이터를 활용하여 산림 전문가와 협력을 통해 테스트 데이터에 대한 새로운 라벨링 작업을 수행하였으며, 이후 해당 데이터를 모델의 예측 결과와 비교하여 정합성을 평가하였다. 이 과정에서 테스트 지역의 수종 분포를 상세히 분석하고, 모델이 실제 산림 구조를 얼마나 정확하게 반영하는지 검토하였다.

2.6. Preliminary Experiments

본 연구에서는 모델 개발에 대한 인사이트를 얻기 위해 두 가지 사전 실험을 진행하였다. 첫 번째 실험은 Band-wise Normalization 의 효과를 분석하는 것이며, 두 번째 실험은 수종 분류 task 에서 Spectral, Temporal, Spatial dimension Information 이 기여하는 정도를 평가하는 것이다. 두 번째 실험에서는 특히 Spatial dimension Information 의 유용성에 초점을 맞추어 분석을 수행하였다. 이는 기존 연구들에서 공간적 정보의 활용 효과에 대한 상반된 결과가 보고되었기 때문이다. 일부 연구에서는 공간 정보를 활용하여 숲의 질감(forest texture)을 효과적으로 포착하는 것이 수종 분류에 어느정도 도움이 된다고 보고한 반면(Machine Learning for Tree Species Classification Using Sentinel-2 Spectral Information, Crown Texture, and Environmental Variables), 다른 연구에서는 공간적 특성이 큰 기여를 하지 않는다고 주장하였다 (Tree Species Classification at the Pixel Level Using Deep Learning and Multispectral Time Series in an Imbalanced Context). 이러한 상반된 결과를 고려할 때, 본 연구에서 활용하는 데이터셋에서도 공간적 정보가 유용한지 여부를 검증할 필요가 있다.

2.6.1. Effect of Band wise Normalization

본 연구에서는 Band-wise Normalization 의 효과를 검증하기 위해 두 가지 실험을 수행하였다. 첫 번째 실험은 shadow 지역에서의 분류 성능 향상 여부를 평가하는 실험이며, 두 번째 실험은 전체적인 intensity 변화에 대한 모델의 strudiness 를 측정하는 실험이다.

5. Improving Classification Performance in Shadowed Areas

대한민국의 산림은 대부분 산지로 이루어져 있으며, 지형의 굴곡으로 인해 위성 영상에서 상당한 shadow 영역이 발생한다. 이러한 그림자 영역은 태양 복사 에너지가 적게 도달하는 지역으로, 분광 정보의 왜곡을 초래하여 수종 분류 모델의 성능에 부정적인 영향을 미칠 수 있다. 따라서 대한민국의 산림을 대상으로 한 수종 분류 모델에서는 그림자 지역에 대한 효과적인 처리가 필수적이다.

본 실험에서는 Band-wise Normalization 이 그림자 지역에서의 분류 성능을 향상시킬 수 있는지 검증하고자 하였다. 실험을 위해 학습 데이터 중 모든 시기의 visible bands(B, G, R) 값의 평균이 하위 15% 이하인 박스를 추출하여 그림자 데이터로 선정하였다. 이후, 이 그림자 데이터를 이용하여 모델을 테스트하였다.

모델은 2.6 절에서 제시한 Light Model 을 사용하였으며, 두 경우 실험 조건(정규화 적용 여부)을 제외하고 동일한 설정으로 학습을 진행하였다. 모든 경우에서 검증(validation) 성능이 수렴할 때까지 학습을 진행한 후, 가장 높은 검증 성능을 기록한 시점의 모델을 비교하였다. 실험 결과는 3.1.1 에서 상세히 논의한다.

6. Evaluating Model Sturdiness Against Intensity Variations

모델의 일반화 성능을 높이기 위해서는 시간적(temporal) 및 공간적(spatial) 일반화 능력이 필수적이다. 특히, 위성 영상의 전반적인 밝기(intensity)는 촬영 시점의 태양 복사 에너지 양과 대기 상태 등의 요인에 따라 크게 달라질 수 있다. 즉, 데이터의 시점 혹은 지역에 따라 전체적인 밴드 값의 크기가 다를 수 있으며, 이러한 차이는 모델의 성능 저하를 초래할 가능성이 있다.

본 연구에서는 밴드별 정규화(Band-wise Normalization)가 이러한 intensity 변화에 대한 일반화 성능을 향상시킬 수 있는지 평가하고자 하였다. 이를 위해 시간적·공간적 변화가 포함된 별도의 검증 데이터가 존재하면 가장 좋겠지만, 본 연구에서는 수집한 데이터 내에서 이러한 환경을 인위적으로 조성하여 실험을 진행하였다.

구체적으로, 각 시기의 모든 밴드 값에 대해 0.8, 1.0, 1.2 중 하나를 랜덤하게 선택하여 곱하는 방식으로 데이터를 변형하였다. 이를 통해 일부 시기의 밴드 값은 전체적으로 작아지고, 일부 시기의 값은 커지는 효과를 부여하였다. 이렇게 변형된 데이터는 촬영 시점과 장소에 따라 달라지는 전반적인 Intensity 차이를 모사하는 역할을 한다.

이후, 조작된 데이터에 대해 밴드별 정규화를 수행한 경우와 수행하지 않은 경우로 나누어 비교하였다. 정규화를 수행한 경우에는 변형된 intensity 가 모델 입력에 반영되지 않으므로, 2.7 절에서 논의한 원래의 실험 결과를 그대로 사용할 수 있다. 따라서, 본 실험에서는 정규화를 수행하지 않은 경우에 대해 intensity 변동에 대한 모델의 강건성을 평가하는 것만으로 충분하였다. 이를 위해 원본 데이터로 학습을 진행한 후, intensity 를 변형한 데이터와 변형하지 않은 데이터에서 추론 성능을 비교하였다.

이러한 실험을 통해 정규화를 수행하지 않을 경우 모델이 intensity 변화에 얼마나 취약한지를 측정하였으며, 이를 바탕으로 Band-wise Normalization 을 활용하여 intensity 변화에 대한 모델의 강건성을 확보하는 것이 얼마나 중요한지 평가하였다. 구체적인 실험 결과 및 분석은 3.1.1 절에서 상세히 제시한다.

2.6.2 Dimensional Importance Comparison

본 실험은 수종 분류(task)에서 spectral, temporal, spatial 차원의 정보가 모델 성능에 미치는 영향을 정량적으로 평가하기 위해 수행되었다. 기존 연구들에서 spectral 정보가 수종 분류에 필수적인 요소로 활용되어 왔기 때문에, 본 실험에서는 spectral 정보를 제외하는 경우는 고려하지 않았다. 따라서, 실험은 다음과 같은 세 가지 케이스로 구성되었다.

- Case 1: Spectral
- Case 2: Spectral + Spatial
- Case 3: Spectral + Temporal

Case 1은 spectral 차원만을 이용했을 때 모델의 분류 성능을 검증하기 위한 실험이다. 이 실험을 통해 spectral 정보만으로 수종 분류가 어느 정도까지 가능한지를 평가하고자 하였다. Case 2는 spectral 정보에 spatial 정보를 추가했을 때 성능 향상이 발생하는지를 분석하는 것을 목표로 하였다. Forest Texture가 수종 분류 Task에서 모델이 학습할 수 있는 중요한 정보인지 확인하기 위해 spatial 정보를 보존한 상태에서 실험을 진행하였다. Case 3은 spectral 정보에 temporal 정보를 추가했을 때 모델의 성능 변화를 평가하기 위해 수행되었다. 시간에 따른 Spectral, Spatial dimension의 변화양상을 파악하는 것이 수종 분류에 유용한지 확인하고자 하였다.

모든 실험에서 입력 데이터는 위성영상의 5×5 픽셀($12.5m \times 12.5m$) 영역을 사용하였다. 그러나 Case 1과 Case 3에서는 spatial 정보를 제거하기 위해 해당 영역의 평균(mean) 값을 입력으로 사용하였다. 이를 통해 주변 픽셀 간의 texture 정보를 활용할 수 없도록 하여, 모델이 순수하게 Spectral 정보 또는 Spectral, Temporal 정보만을 이용하도록 제한하였다.

Temporal 차원의 정보는 특정 시점에서 spectral 및 spatial 정보가 어떻게 변화하는지를 나타낸다. 따라서, temporal 정보를 제거하는 방식에 대해서도 신중한 접근이 필요했다. 본 연구에서는 temporal 정보를 효과적으로 제거하기 위해 각 시점의 데이터를 개별적으로 모델에 입력하고, 모델의 예측 결과를 soft voting 방식으로 앙상블하는 기법을 사용하였다. 이를 통해 모델이 모든 시점의 spectral 및 spatial 정보를 학습하되, 예측 과정에서 다른 시점의 데이터를 활용할 수 없도록 제한하여 temporal 정보의 영향을 차단하였다.

각 실험 케이스에서 사용된 모델 구조는 다음과 같다. 모든 경우에서 band-wise normalization을 적용하여 절대적인 픽셀값보다는 다른 밴드와의 관계에 대한 정보를 학습하도록 했다. 또 모든 case에서 Spectral dimension의 정보를 학습하기 위한 방법으로 pectral Encoder(2.6에서 설명)를 활용하였다.

- Case 1

Spectral Encoder를 활용하여 spectral 차원의 정보를 학습한 후, encoding 결과를 바로 linear head에 연결하여 최종 분류를 수행하였다. 앞서 언급하였듯이 각 시기마다 하나의 모델을 두어 총 12개 모델을 앙상블하였다. 이 앙상블 모델은 총 325K개의 파라미터를 사용한다.

- Case 2

먼저, 각 밴드의 공간적 정보를 학습하기 위해 Fully Connected (FC) Layer를 활용하였다. 이 과정에서는 각 밴드를 개별적인 샘플(batch)처럼 처리하므로, 공간적 특징을 학습하는 동안 Spectral Dimension의 정보는 직접적으로 반영되지 않는다.

이후, 각 밴드별 이미지에서 대표값을 추출하여 최종적으로 Bands 차원의 벡터를 생성하고, 이를 Spectral Encoder 에 입력하여 스펙트럼 정보를 학습하도록 설계하였다. 마찬가지로, 각 시기별로 독립적인 모델을 두어 총 12 개의 모델을 앙상블하였으며, 이 앙상블 모델의 총 356K 의 파라미터를 사용한다.

- Case 3

각각 시기별로 별개의 Spectral encoder 를 가지며 이를 통해 encoding 된 (Time, embed_dim) 텐서를 flatten 한 후, 이들사이의 관계성을 학습하는 linear layer 를 통과하도록 하여 temporal Information 이 반영된 결과를 만들어 낼 수 있도록 설계하였다. 이 모델은 총 324K 개의 파라미터를 사용한다.

각 케이스별 성능 비교 결과는 3.1.2 에서 자세히 분석하였다. 추가로 case1 에서서 각 시기의 모델들이 각각 어떤 예측을 만들었는지를 분석하여 각 클래스를 구분하는데 어떤 시기의 데이터가 유용한 정보를 제공하였는지 분석 할 수 있을것이다. 이에 대한 논의는 4.1 에서 다룬다.

현재 Case 3 에서는 각 차원의 중요성을 효과적으로 비교하기 위해 Case 1 및 Case 2 와 동일하게 각 시점마다 개별적인 Spectral Encoder 를 사용하였다. 그러나 실험 결과, 개별 Spectral Encoder 를 사용하는 것이 성능상 특별한 이점을 제공하지 않았으며, 오히려 공유하는 방식이 파라미터 효율성 측면에서 유리하고 시간적 강건성 또한 향상될 가능성이 높다고 판단하였다. 따라서 이후 모델에서는 공유된 Spectral Encoder 를 적용하였으며, 이에 대한 자세한 논의는 4.2 절에서 다룬다.

2.7. Models

본 연구에서는 앞선 분석을 바탕으로 단일 픽셀 기반 분류(single-pixel classification) 방식을 채택하였으며, 이 접근법에서 가장 효과적인 모델 구조를 탐색하기 위해 여러 후보 모델을 설계하였다. 모델 비교 실험에서 Random Forest (RF) 및 Linear Model 을 제외한 모든 모델에서 공통적으로 Spectral Encoder 를 활용하였다.

- Random Forest

Random Forest 는 기존 연구에서 수종 분류의 기본 모델로 널리 사용되어 왔기 때문에 본 연구에서도 비교 모델로 선정하였다. 해당 모델은 모든 하이퍼파라미터를 기본 설정으로 두고 실험을 수행하였다.

- Linear Model

Linear Model 은 본 연구의 수종 분류 task 가 얼마나 단순한 구조에서도 수행될 수 있는지를 검증하기 위해 추가되었다. 해당 모델은 입력 차원인 time × bands 를 단일 linear layer 를 통해 num_classes 로 매핑하는 방식으로 설계되었으며, 총 파라미터 수는 726 개로 매우 간단한 구조를 갖는다.

- Multi-Layer Perceptron (MLP)

Multi-Layer Perceptron (MLP)은 추가적인 inductive bias 없이 충분한 파라미터를 제공했을 때 모델의 학습 성능을 평가하기 위해 도입되었다. 이 모델은 linear layer, batch normalization, ReLU, dropout 등의 구성 요소를 포함하는 심층 신경망 구조를 갖추고 있으며, 총 229K 개의 파라미터를 가진다. 입력 데이터는 flatten 된 형태로 모델에 전달되며, 여러 개의 hidden layer 를 통해 복잡한 비선형 관계를 학습하도록 설계되었다.

- Light Model

Light Model 은 Spectral Encoder 를 기반으로 가장 단순한 형태의 시간적 정보 학습 구조를 갖춘 모델이다. Spectral Encoder 를 통해 각 시점의 스펙트럼 정보를 추출한 후, (Time, embed_dim) 형태의 출력을 flatten 한 뒤 단일 linear layer 를 적용하여 최종 분류를 수행한다. 이 모델의 총 파라미터 수는 31K 개이며, spectral encoding 이후 temporal 차원의 정보를 학습하는 가장 단순한 형태의 모델로, 보다 복잡한 모델들과 비교하는 기준점 역할을 한다.

- Transformer Model

Transformer Model 은 Spectral Encoder 이후 시간적 정보(temporal information)를 보다 효과적으로 학습할 수 있도록 Transformer Encoder 를 추가한 모델이다. Spectral Encoder 의 출력인 (Time, embed_dim) 텐서를 Transformer Encoder 에 입력하여 시간적 패턴을 학습한 후, 출력을 flatten 한 뒤 linear head 를 적용하여 최종 분류를 수행한다.

이 모델의 총 파라미터 수는 98k 개로, Light Model 대비 temporal 차원의 정보를 더욱 정교하게 학습할 수 있다. Light Model 과 Transformer Model 을 비교함으로써, 시간적 관계를 학습하는 데 있어 단순 Linear layer 이상의 high-capacity 구조가 필요한지 여부를 평가할 수 있으며, 그중에서도 Transformer 기반 모델이 성능 향상에 얼마나 기여하는지 분석할 수 있다.

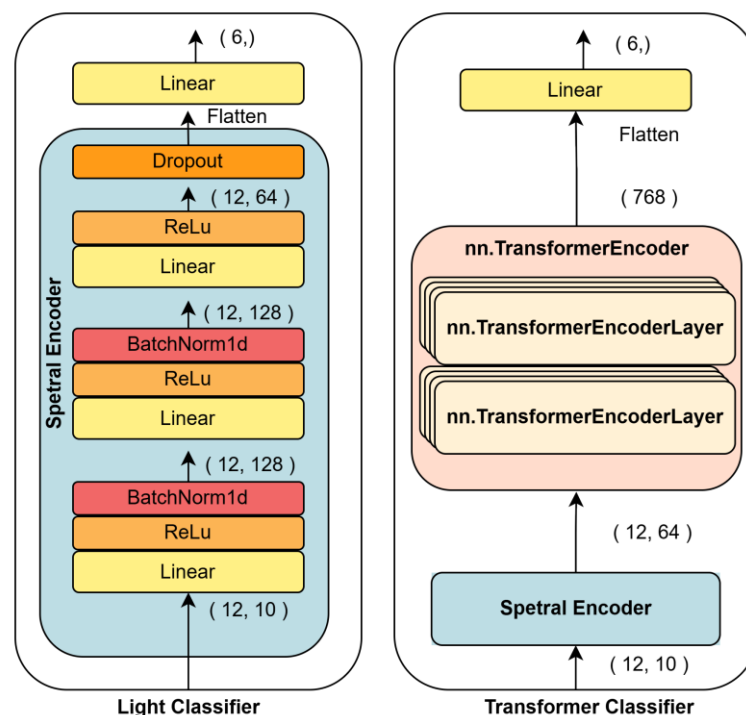


Figure 5. Architectures of main models

2.8. Model Comparison

모델 성능 비교를 위해 각 모델에 대해 5-Fold 교차 검증을 수행하였으며, RF(Random Forest)를 제외한 모든 모델은 동일한 학습 환경에서 훈련되었다. 학습 과정은 100 epochs 동안 진행되었으며, 대부분의 모델이 조기 수렴하는 경향을 보였지만, 충분한 학습을 보장하기 위해 100 epochs 까지 훈련을 지속하였다. 최종

성능 평가는 각 폴드에서 validation 성능이 가장 높았던 epoch 의 결과를 기준으로 수행하였다. 모든 모델은 CrossEntropyLoss 를 손실 함수로 사용하였으며, 최적화 기법으로는 AdamW 를 적용하였다. 또한, 학습률(learning rate)은 $3e-4$, weight decay 는 $5e-4$ 로 설정하여 실험을 진행하였다.

각 폴드마다 4 개의 task 에 대해 averaged recall 과 averaged precision 을 각각 측정하여 폴드당 총 8 개의 성능 지표를 수집하였다. 따라서, 5-Fold 교차 검증을 통해 총 40 개의 성능 지표를 확보하였다. 수집된 성능 지표를 바탕으로 통계적 유의성을 평가하기 위해 Friedman test 를 수행하였으며, 모델 간의 성능 차이를 보다 정밀하게 분석하기 위해 post-hoc Nemenyi test 를 추가적으로 실시하였다.

Friedman test 는 여러 모델이 동일한 데이터셋에서 수행되었을 때, 모델 간의 성능 차이가 통계적으로 유의미한지를 검증하는 비모수적 검정(non-parametric test) 방법이다. 이 테스트에서 모델들의 평균 랭크 차이가 유의미한 수준으로 나타날 경우, 추가적인 다중 비교 분석이 필요하다. 이에 따라, 사후 검정(post-hoc test)으로 Nemenyi test 를 적용하여 모델 간 pairwise 비교를 수행하였다.

Nemenyi test 는 여러 모델 간의 평균 랭크 차이를 비교하여 특정 모델이 다른 모델보다 통계적으로 유의미하게 더 나은 성능을 보이는지를 확인하는 검정 방법이다. 본 연구에서는 Nemenyi test 를 통해 모델 간 통계적으로 유의미한 성능 차이가 존재하는지를 분석하였으며, 이를 기반으로 최적의 모델을 선정하였다. 모델 비교 실험의 결과 및 상세 분석 내용은 3.2 절에서 제시한다.

3. Results

3.1. Results of Preliminary Experiments

3.1.1 Effect of Band wise Normalization

1. Improving Classification Performance in Shadowed Areas

본 연구에서는 band-wise normalization 이 그림자 영역에서의 분류 성능에 미치는 영향을 정량적으로 분석하였다. 이를 위해, 그림자 영역에서 Normalization 을 적용한 모델과 적용하지 않은 모델의 성능을 비교하였다.

Table 3. Performance under shadow conditions with and without normalization

	Non-Forest Vs. Forest	Intra-Conifer	Intra-Broadleaf	Conifer Vs. Broadleaf	Overall F1-score	Overall Accuracy
Unnormalized	88.62	87.52	65.94	92.73	73.91	71.66
Normalized	91.13	88.91	69.49	93.25	76.64	75.01

Table 3 의 A 결과를 보면, Band-wise Normalization 을 적용한 모델이 모든 평가 항목에서 더 높은 성능을 보였다. 특히, Overall F1-score 는 0.739 에서 0.766 으로 증가하였으며, Overall Accuracy 또한 0.716 에서 0.750 으로 향상되었다. 이는 Normalization 을 통해 그림자 영역에서의 분광 정보 왜곡이 보정되었으며, 이에 따라 모델이 보다 일관된 특성을 학습할 수 있었음을 시사한다.

세부적으로 살펴보면, Intra-Broadleaf(0.659 → 0.694)와 Intra-Conifer(0.875 → 0.889)와 같은 세부 분류에서 더욱 뚜렷한 성능 향상이 나타났다. 이는 그림자 영역에서의 분광 정보 왜곡이 세부적인 수종 구분에 더 큰 영향을 미치기 때문으로 해석할 수 있다. 즉, 동일한 범주 내에서 미세한 스펙트럼 차이를 구별하려면 보다 정밀한 스펙트럼 정보가 필요하지만, 그림자 지역에서 발생하는 왜곡이 이러한 구분을 더욱 어렵게 만들기 때문이다. 이러한 상황에서 Band-wise Normalization 을 적용함으로써 왜곡이 일정 부분 보정되었고, 그 결과 모델의 세부적인 수종 분류 성능이 향상된 것으로 판단된다.

결론적으로, 본 연구에서는 Band-wise Normalization 기법이 그림자 영역에서의 수종 분류 성능 향상에 효과적임을 실험적으로 검증하였다. 이러한 결과는 산악 지형이 많은 우리나라의 환경에서 특히 중요한 의미를 가지며, 본 기법을 활용함으로써 지형적 요인으로 인해 발생하는 그림자 영역에서도 보다 일관된 분광 정보를 확보할 수 있음을 시사한다.

2. Evaluating Model Sturdiness Against Intensity Variations

Table 4. Performance by normalization and perturbation

Normalization	Perturbation	Non-Forest Vs. Forest	Intra- Conifer	Intra- Broadleaf	Conifer Vs. Broadleaf	Overall F1-score	Overall Accuracy
x	x	98.15	87.39	79.61	95.18	82.49	84.01
	o	95.27	77.91	70.82	90.11	70.56	71.44
o	x	98.01	87.57	79.81	95.32	82.65	83.74
	o						

실험 결과, intensity 변화를 적용했을 때 밴드별 정규화를 수행하지 않은 모델의 성능이 크게 저하되는 것으로 나타났다. 원본 데이터에서 전체 정확도(Overall Accuracy)는 84.01%였으나, intensity 변형이 적용된 데이터에서는 71.44%로 감소하였다. 마찬가지로, 전체 F1-score 도 82.49 에서 70.56 으로 하락하며, 모델이 intensity 변화에 민감하게 반응함을 확인할 수 있었다. 이는 모델이 절대적인 반사율 값에 의존하는 경향이 있으며, 이러한 의존성이 환경 변화에 취약하게 만드는 요인으로 작용할 수 있음을 의미한다.

한편, 밴드별 정규화를 적용한 모델은 intensity 변화가 단순한 상수배 형태로 주어질 경우 입력값의 상대적 분포가 유지되므로 성능 저하가 발생하지 않았다. 이 두 실험 결과를 비교해 보면, 모델의 intensity sturdiness 를 확보하고 일반화 성능을 높이기 위해서는 Band-wise Normalization 과 같은 전반적인 intensity 변동에 강인한 모델 설계 요소가 반드시 필요함을 확인할 수 있다.

그러나 실제 환경에서의 intensity 변화는 단순한 상수배 형태로만 나타나지 않는다. 위성 영상의 촬영 시점과 공간적 차이에 따라 Spectral 특성이 비선형적으로 변화할 가능성이 있으며, 이러한 변화는 단순한 Band-wise Normalization 기법만으로 완전히 보정되지 않을 수 있다. 그럼에도 불구하고, 본 실험을 통해 intensity 변화가 모델 성능에 미치는 부정적인 영향을 정량적으로 평가할 수 있었으며, 이러한 변동성에 강건한 모델 설계가 필요함을 명확히 시사한다.

3.1.2 Dimensional Importance Comparison

Table 5. Performance by case

Case	Non-Forest Vs. Forest	Intra-Co- nifer	Intra- Broadleaf	Conifer Vs. Broadleaf	Overall F1-score	Overall Accuracy
Case1	97.63	89.36	78.11	94.12	82.06	82.68
Case2	97.38	88.97	77.24	94.04	81.45	82.20
Case3	98.00	91.45	81.95	94.87	84.95	85.40

실험 결과, 첫 번째로 주목할 만한 점은 Spectral 차원의 정보만으로도 80% 이상의 Overall Accuracy 를 달성할 수 있었다는 것이다. 이는 수종 분류에서 Spectral 정보가 매우 중요한 역할을 한다는 점을 강하게 시사한다. 실제로 산림을 항공사진으로 관찰해 보면, 수종별로 색상이 다를 수 있음을 확인할 수 있다. 예를 들어, 침엽수는 일반적으로 어두운 녹색을 띠는 반면, 활엽수는 상대적으로 밝은 색을 나타낸다. 이러한 차이는 개엽기와 낙엽기에는 더욱 두드러지며, 여기에 가시광선을 이외의 Invisible Bands 까지 활용하면 수종 분류에 유용한 추가 정보를 제공할 수 있을 것으로 보인다.

둘째, 공간적(spatial) 차원의 정보는 모델 성능 향상에 유의미한 기여를 하지 않는 것으로 나타났다. 공간적 정보를 학습하도록 설계된 경우, 모델의 성능이 오히려 저하되는 경향을 보였다. 이는 본 연구에서 사용된 2.5m/px 해상도에서는 개별 나무 내부의 잎과 가지 패턴과 같은 미세한 텍스처 정보를 포착하기 어렵기 때문인 것으로 판단된다. 대신, 인접한 나무들 간의 배치에서 발생하는 forest texture 를 학습하는 것이 가능할 수 있다. 그러나 본 연구에서 사용된 데이터셋과 클래스에서는 이러한 forest texture 의 차이가 수종 분류에 유의미한 도움을 제공하지 못함을 확인하였다.

셋째로, Temporal 차원의 정보는 모델의 분류 성능을 유의미하게 향상시키는 것으로 나타났다. 이는 개별 시점에서 얻은 Spectral 정보뿐만 아니라, 시간에 따른 변화를 함께 고려하는 것이 수종 분류에 중요한 역할을 한다는 점을 시사한다. 특히, Intra-Broadleaf, Intra-Conifer 등의 상대적으로 어려운 task 에서 성능 향상 폭이 두드러지게 나타났으며, 이를 통해 동일한 범주 내의 세부 분류에서는 단일 시점의 Spectral 정보만으로는 충분하지 않고 Temporal 정보를 활용하는 것이 필수적이라는 점을 확인할 수 있었다.

이러한 결과를 종합하면, 수종 분류 모델 설계 시에 무엇보다 Spectral 차원의 정보를 확실히 학습할 수 있도록 하는 것이 중요하고 Spatial 정보보다는 Temporal 정보를 적극적으로 활용하는 것이 더욱 효과적인 접근법이라는 결론을 도출할 수 있다. 본 연구에서는 이러한 실험 결과를 바탕으로, 1 픽셀 단위의 입력을 사용하면서 Spectral, Temporal 정보를 효과적으로 반영할 수 있는 모델 구조를 채택하였다.

3.2. Results of Model Comparison

Table 6 는 각 모델별로 개별 분류 과제에서의 모든 폴드(Fold)에 대한 평균 F1-score, 평균 Overall F1-score, 및 평균 Overall Accuracy 를 정리한 결과이다.

Table 5. Performance by Model

Model	Non-Forest Vs. Forest	Intra-Co- nifer	Intra- Broadleaf	Conifer Vs. Broadleaf	Overall F1-score	Overall Accuracy
-------	--------------------------	--------------------	---------------------	-----------------------------	---------------------	---------------------

Linear	97.26	84.15	71.49	93.93	77.29	78.97
Random Forest	97.70	86.26	76.45	94.36	80.30	81.65
MLP	97.96	89.16	79.47	95.65	83.92	84.79
Light	97.91	88.41	79.75	95.08	82.89	83.36
Transformer	98.15	89.05	80.79	95.28	83.82	84.88

먼저, Linear 모델만을 활용하더라도 Non-Forest vs. Forest 및 Conifer vs. Broad-leaf 분류에서 높은 성능을 기록하였다. 이는 Spectral 및 Temporal 정보의 결합만으로도 단순한 모델에서 일정 수준 이상의 분류 성능을 확보할 수 있음을 시사한다. 반면, Intra-Conifer 및 Intra-Broadleaf 분류에서는 보다 높은 모델 용량(model capacity)이 요구되는 것으로 나타났다. 이는 동일한 범주 내에서 세부 수종을 구별하는 것이 상대적으로 높은 난이도를 가지며, 선형 모델만으로는 충분한 분류가 어렵다는 점을 의미한다.

실제로, MLP, Light Model, Transformer Model 과 같은 비선형 모델들은 이러한 고난도 분류 과제에서 성능이 개선되는 경향을 보였다. 특히, MLP 와 Transformer Model 은 Intra-Conifer 및 Intra-Broadleaf 분류에서 가장 높은 수준의 성능을 기록하였다. 특히, Intra-Broadleaf 분류 과제에서는 Transformer 모델과 Light Model 이 MLP 보다 더 우수한 성능을 보였으며, 이는 QV 와 QM 분류에서 Spectral Encoding 을 먼저 수행한 후, Temporal 차원에서의 관계를 학습하는 방식이 효과적으로 작용하였기 때문으로 분석된다. 이러한 결과는 향후 다양한 활엽수종을 분류하는 문제에서도 적용 가능한 방법론을 제시한다.

모델 간 성능 차이가 통계적으로 유의미한지 검증하기 위해 Friedman test 를 수행한 결과, 유의수준 0.05 에서 모든 classifier 간 성능 차이가 없다는 귀무가설이 기각되었다. 즉, 모델 간 성능 차이는 통계적으로 유의미한 것으로 나타났다. 이에 따라 post-hoc Nemenyi test 를 수행하였으며, 각 모델 간의 상호 성능 차이를 분석하였다.

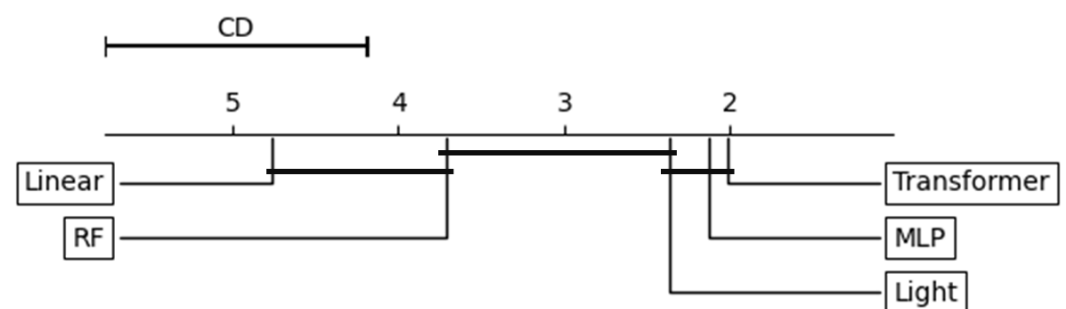


Figure 6. CD-plot

분석 결과, Transformer 모델이 전반적으로 가장 우수한 성능을 보였으나, MLP 및 Light Model 과의 성능 차이는 통계적으로 유의미하지 않은 수준(Critical Distance 이하)으로 나타났다. 주목할 점은 Transformer 모델의 파라미터 수가 MLP 의 약 40% 수준, Light Model 은 MLP 의 약 13% 수준에 불과하다는 점이다. 이는 Spectral Encoder 를 활용하여 분광 정보를 먼저 학습하는 방식이 모델의 효율성을 높이는 데 기여했음을 시사한다.

이러한 결과를 바탕으로, 본 연구에서는 최종적으로 Transformer 모델을 선택하였다. 또한, 보다 경량화된 모델이 필요한 경우, Spectral Encoder 에 Linear Head 를 결합한 Light Model 이 효과적인 대안이 될 수 있음을 확인하였다. 추가적으로, 본 연구에서는 Spectral Encoder 에 1D-Convolution 및 GRU 기반의 헤드를 결합하는 다양한 변형 모델을 실험하였으며, 대부분 유사한 수준의 성능을 보였다.

3.2. Final Model Performance

3.2.1 Confusion Matrix

우리는 최종적으로 transformer 모델을 사용하여 예측을 수행하였다. 분류 결과를 나타내는 confusion matrix 는 다음과 같았다.

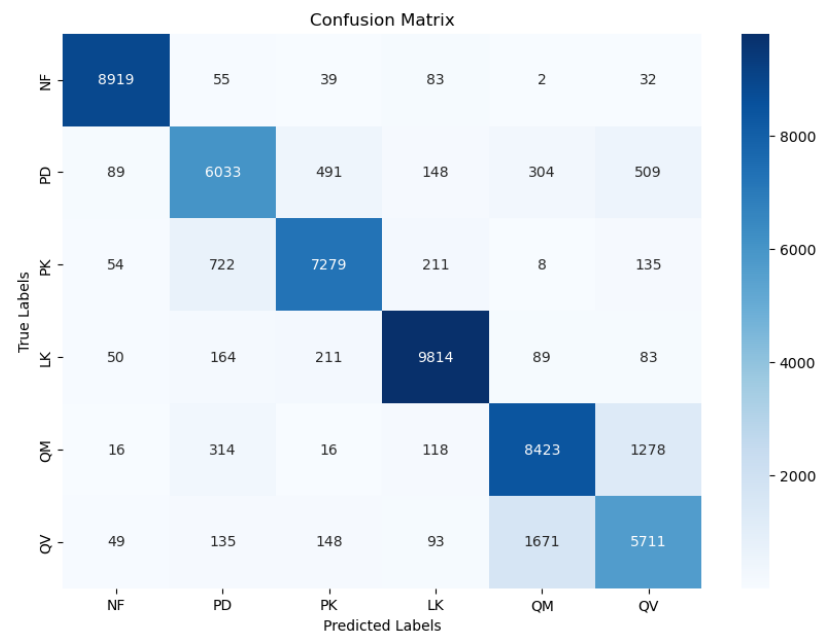


Figure 7. Confusion Matrix

분류 결과를 분석한 결과, 대부분의 샘플이 대각선상에 위치하며 전반적으로 우수한 분류 성능을 보였다. 특히 NF 와 LK 클래스에서는 오분류(confusion)가 거의 발생하지 않았으며, 모델이 해당 수종을 효과적으로 구별하는 것으로 나타났다.

반면, QM 과 QV 클래스에서는 상대적으로 높은 오분류율이 관찰되었다. 이는 두 수종이 모두 참나무속(Quercus)에 속하며, 스펙트럼(Spectral) 차원에서 많은 유사성을 가지기 때문으로 판단된다. 실제로 해당 수종들은 생태적 특성과 생육 환경이 유사하여, 일부 교잡(Hybridization)이 발생할 가능성이 있으며, 이에 따라 명확한 구분이 어려울 수 있다.

그럼에도 불구하고, 제안된 Transformer 모델은 이 두 수종을 80% 이상의 정확도로 분류하는 성능을 보였다. 이는 본 연구의 방법론이 높은 난이도의 수종 분류에서도 효과적으로 작동함을 시사하며, 향후 더 다양한 수종에 대한 데이터 구축과 모델 학습을 진행할 경우, 스펙트럼 특성이 보다 뚜렷이 구분되는 수종들 간에는 더욱 높은 분류 성능을 기대할 수 있을 것이다.

3.2.2 Qualitative Evaluation

Figure 8 는 본 연구에서 개발한 모델의 예측 결과를 테스트 지역에 적용한 사례를 보여준다. 이를 학습에 사용된 Ground Truth 데이터 및 전문가 라벨링 데이터와 비교하여 모델의 성능을 정성적으로 평가하였다.

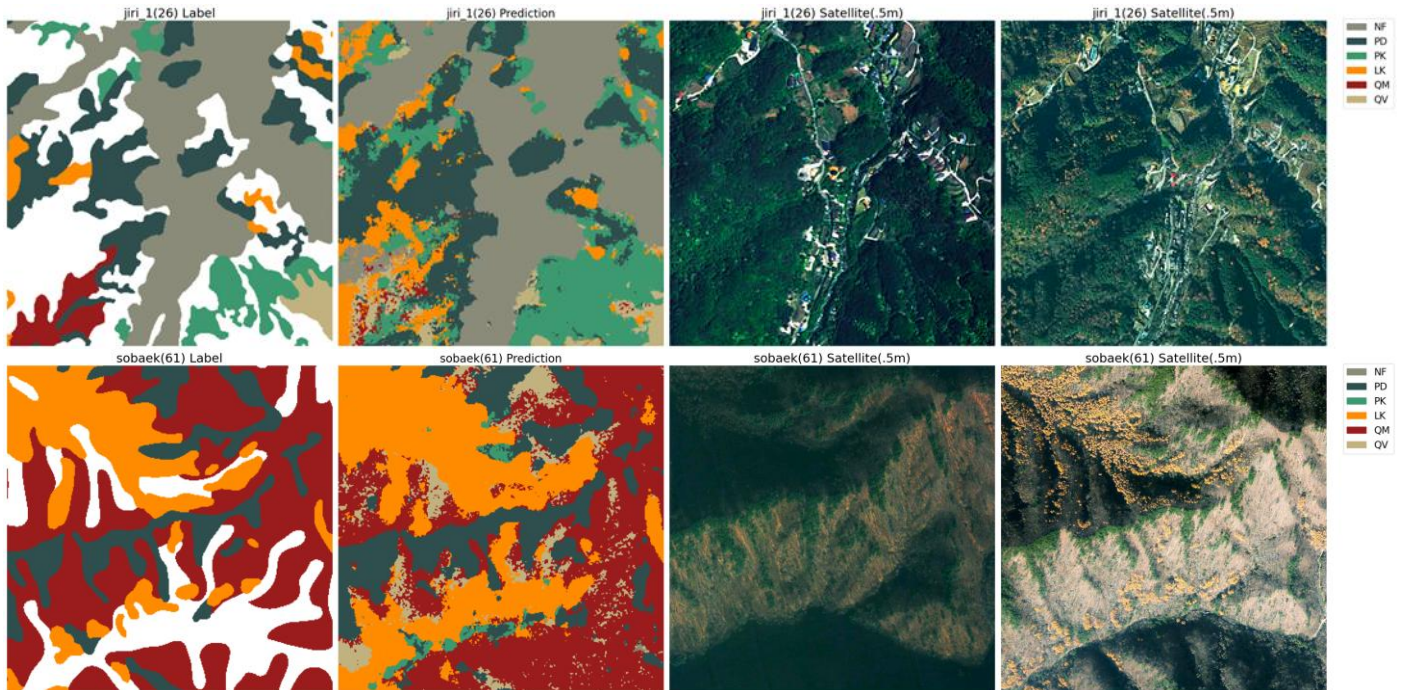


Figure 8. Prediction results on test regions

본 연구에서는 12 개의 시기별 Sentinel-2 위성 데이터를 기반으로 각각의 spectral 모델을 학습하고 이를 앙상블하여 최종 결과를 도출하였다. 이를 통해 각 클래스의 분류에서 특정 시기의 중요성을 평가할 수 있었다. 클래스별로 각 시기의 모델이 생성한 확률 분포를 합산하여 평균 확률 분포를 분석하였으며, 특히 분류가 어려운 두 쌍의 클래스, 즉 *Pinus densiflora* (PD)와 *Pinus koraiensis* (PK), 그리고 *Quercus mongolica* (QM)와 *Q. variabilis* (QV)에 대해 시기별 예측 확률 차이를 분석하였다.

Figure 8 은 이를 시각화 한것이다. 본 연구에서 생성한 수종 분류 지도는 Ground Truth 데이터의 해상도를 초과하는 수준으로 예측되었으며, 이에 따라 기존 라벨링 데이터에서 혼효림으로 매핑된 지역에서도 수종이 혼재된 양상을 보다 세밀하게 관찰할 수 있었다. 또한, 전문가 라벨링 데이터와 비교했을때, 제안된 모델의 예측 결과는 단순히 새로운 패턴을 생성하는 것이 아니라 실제로 수종 분류 지도의 해상도와 표현력을 향상시키는 역할을 수행함을 확인할 수 있었다.

이를 통해, 본 연구는 별도의 라벨링 데이터를 구축하지 않고도 기존의 임상도 데이터를 활용하여 보다 정밀한 수종 분류 지도를 생성할 수 있음을 보여주었다. 즉, 기존 임상도를 활용하여 그 이상의 품질을 갖는 수종 지도를 생성할 수 있도록 모델을 학습함으로써, 임상도의 정밀도를 효과적으로 향상시키는 방법론을 제시하였다.

4. Discussion

4.1. Evaluating Temporal Importance Per Classes

본 연구에서는 12 개의 시기별 Sentinel-2 위성 데이터를 기반으로 각각의 spectral 모델을 학습하고 이를 앙상블하여 최종 결과를 도출하였다. 이를 통해 각 클래스의 분류에서 특정 시기의 중요성을 평가할 수 있었다. 클래스별로 각 시기의 모델이 생성한 확률 분포를 합산하여 평균 확률 분포를 분석하였으며, 특히 분류가 어려운 두 쌍의 클래스, 즉 *Pinus densiflora* (PD)와 *Pinus koraiensis* (PK), 그리고 *Quercus mongolica* (QM)와 *Q. variabilis* (QV)에 대해 시기별 예측 확률 차이를 분석하였다. Figure 9 는 이를 시각화 한것이다.

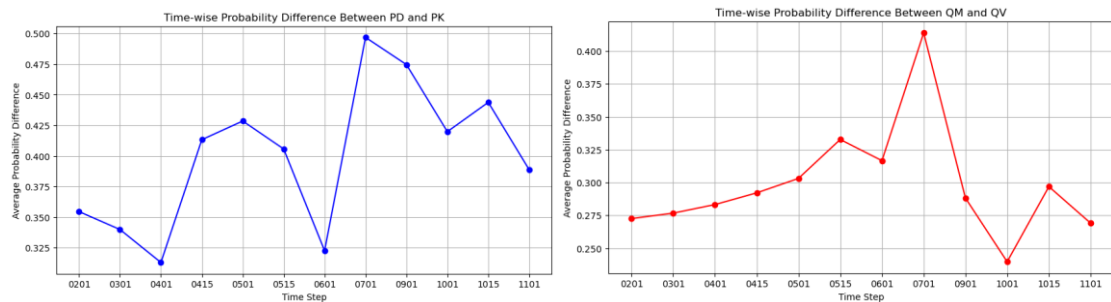


Figure 9. Variation in prediction probabilities over time

본 연구에서 7 월 1 일 데이터를 사용한 모델이 *Pinus densiflora* (PD)와 *Pinus koraiensis* (PK), *Quercus mongolica* (QM)와 *Q. variabilis* (QV) 간 가장 큰 확률 차이를 보였다. 이는 여름철 데이터가 수종 간 스펙트럼 차이를 극대화하기 때문으로 해석된다. 이는 여름철에 수종의 생리적 변화와 광합성 활성도가 최고조에 이르러 스펙트럼 반사율에서 뚜렷한 차이를 만들어내기 때문이다. 실제로, [1]Freudenberg et al. (2025)의 연구에서는 다계절 Sentinel-2 데이터를 사용하여 수종 분류를 수행했으며, 여름철 데이터가 높은 분류 정확도를 제공한다고 보고되었다. 또한, [2]Hemmerling et al. (2021)은 다계절 Sentinel-2 데이터로 17 개 수종을 분류하면서 여름철 데이터가 특정 수종 간의 차이를 명확히 구분할 수 있는 중요한 시점임을 밝혔다

실제로 실험 결과, 7 월 데이터는 단일 시점의 분광 정보만으로도 validation accuracy 0.76 이상의 성능을 보여주었으며, 이는 여름철 스펙트럼 데이터가 수종 분류에서 매우 효과적임을 입증한다. 따라서, 여러 시점의 데이터를 사용할 수 없을 경우 대한민국에서는 장마가 시작되기 전인 7 월 초의 데이터를 이용하면 제한된 자원으로도 효율적인 데이터 구축이 가능할 것이다.

본 연구에서는 위성 데이터의 스펙트럼 차원 정보를 효과적으로 학습하기 위해 Spectral Encoder 를 활용한 모델 구조를 채택하였다. 최종적으로 선정된 Transformer 모델과 Light Model 모두 모든 시점에서 공유 Spectral Encoder 를 사용하였으며, 이는 다음과 같은 실험 결과에 기반한 설계 결정이다.

4.2. Shared vs. Individual Spectral Encoder

이 실험에서는 모든 시점에서 Spectral Encoder 를 공유하는 모델과 각 시점별로 개별적인 Spectral Encoder 를 갖는 모델을 설계하여 비교 실험을 수행하였다. 실험에 사용된 기본 모델 구조는 Light Model 이며, Spectral Encoder 의 공유 여부를 제외한 모든 모델 구조는 동일하게 유지하였다. 두 모델 모두 동일한 데이터셋을 사용하여

학습과 테스트를 진행하였으며, validation score 가 수렴할 때까지 충분한 학습을
수행하였다. 실험 결과는 다음과 같다.

Table 5. Performance based on shared vs. individual Spectral Encoder

Model	Non-Forest Vs. Forest	Intra-Co- nifer	Intra- Broadleaf	Conifer Vs. Broadleaf	Overall F1-score	Overall Accuracy
individual	97.62	91.64	79.88	94.29	83.82	84.34
shared	98.29	91.28	81.25	95.14	84.96	85.52

실험 결과, 각 시점별로 독립적인 Spectral Encoder 를 사용하는 경우 훨씬 많은
파라미터(individual: 325k, shared: 31k)를 사용했음에도 불구하고 대부분의 과업에서
성능이 저하되는 현상이 관찰되었다. 이는 spectral encoding 방법이 시간에 대해
불변(invariant)함을 시사한다. 또한, 시간적으로 불변한 encoding 방법을 학습하는 데
있어 공유된 Encoder 가 더 유리한 것으로 판단된다. 이는 개별 Spectral Encoder 가
특정 시점의 정보만을 학습함에 따라, 여러 시기에 걸쳐 공통적으로 나타나는
스펙트럼 차원의 일반적인 특징을 파악하지 못하기 때문으로 해석된다. 반면, 여러
시기의 데이터를 이용하여 하나의 공유된 Encoder 를 학습하면 일반적인 Spectral
Encoding 방식을 효과적으로 학습할 수 있으며, 이로 인해 성능이 향상된 것으로
보인다. 따라서 모든 시점에서 동일한 Spectral Encoder 를 사용하는 것이 더
효과적인 전략임을 확인할 수 있다. 이러한 구조는 다음과 같은 장점을 제공한다:

1. 여러 시기의 데이터를 이용한 일반적인 특징 학습: 하나의 공유된 Spectral En-
coder 를 통해 다양한 시기의 데이터를 학습함으로써 보다 일반화된 스펙트럼
정보를 추출할 수 있다. 이는 입력 시점이 달라져도 동일한 Spectral Encoding
방식을 적용할 수 있어 시간적 변화에 강건한 모델을 구축할 수 있다.
2. 파라미터 효율성: 모든 시점에서 동일한 Spectral Encoder 를 공유함으로써 모델의
파라미터 수를 크게 줄일 수 있다. 이는 CNN 기반 인코더와 유사하게 입력
데이터 크기와 관계없이 일정한 파라미터 수로 스펙트럼 정보를 추출할 수
있도록 한다.

4.3. Constraint and Futureworks

본 연구에서는 PlanetScope 데이터의 사용량 제한 및 가용한 컴퓨팅 자원의
한계로 인해 대규모 데이터셋을 구축하지 못하였다. 그러나 시간적·공간적 일반화
성능을 향상시키기 위해서는 다양한 시기 및 지역에서 획득한 데이터를 추가적으로
학습하는 것이 필수적이다. 이를 통해 보다 강건한 일반화 성능을 갖춘 모델을
구축할 수 있을 것으로 예상된다.

또한, 본 연구에서는 소수의 수종에 대해 충분한 학습 데이터를 확보하지
못하여 5 가지 수종만을 대상으로 분석하였다. 만약 리기다소나무(*Pinus rigida*),
상수리나무(*Quercus acutissima*), 편백나무(*Chamaecyparis obtusa*) 등 다른 주요
수종에 대한 충분한 데이터를 확보할 수 있다면, 더 많은 클래스를 구분할 수 있는
모델을 구축할 수 있을 것이다. 특히, 본 연구에서는 신갈나무(*Quercus mongolica*)와
굴참나무(*Quercus variabilis*)가 0.8 이상의 Accuracy 로 분류 가능함을 확인하였으며,
이에 기반하여 더 큰 차이를 보이는 수종 간 분류에서도 효과적인 성능을 기대할 수
있다.

위성영상을 활용한 수종 지도 제작의 주요 장점은 동적인 변화에 따른 갱신이 용이하다는 점이다. 그러나 본 연구에서 제안한 모델은 12 시점의 위성 영상을 입력으로 사용하며, 특정 시점에서 변화가 발생할 경우 모델의 예측이 모호해질 가능성이 존재한다. 예를 들어, 어떤 지역이 초기 시점에는 산림이었으나, 이후 벌목 등의 요인으로 인해 비산림으로 변화한 경우, 모델이 이를 어떻게 해석할지 불분명하다. 이는 본 연구에서 고려하지 않은 부분으로, 모델이 중간 변화에 취약할 가능성을 시사한다.

이를 해결하기 위한 방법으로는 다음과 같은 두 가지 접근법이 가능하다.

1. 데이터 차원의 접근법: 변화가 포함된 다양한 시계열 데이터를 확보하여 모델이 변화 이후의 클래스를 정확히 학습하도록 유도할 수 있다.
2. 모델 차원의 접근법: 각 시점별 이미지간의 거리(distance)를 정의하여, 특정시점에서 큰 변화가 감지될 경우 이를 별도로 처리하도록 한다. 이 방식으로 통해서 변화가 발생한 지역에 대하여 변화 결과를 반영하는 예측을 하도록 할 수 있다.

물론 효과적인 처리를 위해서는 이 두가지 차원의 접근법을 모두 사용해야 할 것이다.

추가적으로, 본 연구에서는 현재 사용한 위성 영상 해상도에서 forest texture의 유용성이 크지 않다고 판단하였으나, 더 고해상도의 영상 데이터를 활용할 경우 texture 정보가 유의미한 역할을 할 가능성이 있다. 향후 연구에서는 보다 높은 해상도를 갖는 영상 데이터를 활용하여 forest texture 정보를 충분히 이용하는 모델을 도입할 필요가 있다.

향후 연구에서는 본 연구에서 사용한 위성 데이터를 활용하여 흉고직경(DBH), 임분고(Tree height), 산림 밀도(Forest density) 등의 회귀 분석(regression)을 수행하고, 이를 바탕으로 산림이 저장하고 있는 탄소량을 추정하는 연구로 확장할 계획이다.

5. Conclusions

본 연구는 다중 시계열 Sentinel-2 및 PlanetScope 위성 데이터를 융합하고 딥러닝 모델을 적용하여, 한국의 복잡한 온대 혼효림(소백산, 지리산 국립공원)에 대한 2.5m 고해상도 수종 분류 프레임워크를 성공적으로 개발했다. 핵심적으로, 독자 제안한 밴드별 정규화 기법은 위성 영상의 그림자 및 강도 변화에 대한 모델 강건성을 효과적으로 개선했다. 또한, 수종 분류에는 공간 정보보다 분광 정보와 다중 시계열 정보가 결정적인 역할을 하며, 특히 유사 수종 구분과 여름철(7월) 데이터의 중요성을 확인했다. 모델 구조 측면에서는 Spectral-Temporal 정보를 통합 학습하는 Transformer 기반 모델이 파라미터 효율성과 분류 성능(Overall Accuracy 85.52%) 모두에서 가장 우수함을 입증했다.

본 연구는 접근이 어려운 산악 지형에 대해 위성 데이터와 딥러닝 기술을 효과적으로 결합하여, 정밀하고 시의성 있는 대규모 수종 분포 지도를 제작할 수 있는 실용적인 방법론을 제시했다는 데 중요한 의의가 있다. 이는 향후 원격탐사 기반의 정밀 산림 모니터링 및 관리, 생태계 연구 발전에 기여할 수 있는 기반을 마련한다.

다만, 본 연구는 제한된 연구 지역과 일부 주요 수종을 대상으로 수행되었으며, 730
벌목과 같은 급격한 산림 변화가 모델 예측에 미치는 영향에 대한 심층적인 검증은 731
부족했다. 향후 연구에서는 더 넓은 지역과 다양한 수종을 포괄하는 대규모 732
데이터셋을 구축하고, 시간적 동적 변화를 효과적으로 감지하고 처리할 수 있는 모델 733
아키텍처를 개발할 필요가 있다. 또한, 초고해상도 위성 영상에서의 공간 텍스처 734
정보 활용 가능성을 재탐색하고, 본 프레임워크를 홍고직경, 임분고, 탄소 저장량 등 735
주요 산림 자원 파라미터 추정 연구로 확장하여 산림 관리 및 기후 변화 대응 전략 736
수립에 실질적인 기여를 할 수 있도록 발전시켜 나갈 계획이다. 737