

제한된 데이터 환경의 전이학습 기반 XGBoost 정주지 수목 탐지 모델 A Transfer Learning-Based XGBoost Settlement Tree Detection Model for Data-Scarce Environments

이준우¹⁾ · 김동범²⁾ · 김근한³⁾ · 전철민⁴⁾

Lee, Joonwoo · Kim, Dongbeom · Kim, Geunhan · Jun, Chulmin

Abstract

This study proposes a methodology for detecting carbon sinks (trees) within residential areas using a transfer learning-based XGBoost machine learning model, aiming to develop a tree detection model applicable even in municipalities lacking up-to-date tree information. Seasonal baseline models were constructed using public tree location data from Seoul and Sentinel-2 satellite imagery and then transferred to Daegu's data to explore the model's predictive performance in environments with limited labeled data. The summer model (August) demonstrated the highest performance, with the transfer model improving accuracy by 0.022 compared to the baseline. Interpretability of the prediction results was ensured through SHAP analysis. The tree distributions predicted by the transfer model showed high spatial agreement with actual tree locations, especially in densely vegetated urban street areas within residential zones. This study provides a foundation for future research directions, including high-resolution image-based detection, applications in data-scarce environments, and integration with greenhouse gas inventory frameworks.

Keywords : Tree Detection, Transfer Learning, XGBoost, Sentinel-2, SHAP

초 록

본 연구는 최신 수목 정보가 부족한 지자체에서도 활용 가능한 수목 탐지 모델을 개발하기 위해, 전이학습 기반의 XGBoost 기계학습 모델을 활용하여 정주지 내 탄소흡수원(수목)을 탐지하는 방법론을 제안한다. 서울시의 공공 수목 위치 정보와 Sentinel-2 위성영상을 기반으로 계절별 기초모델을 구축하고, 이를 대구시 데이터에 전이 학습함으로써 소규모 라벨링 데이터 환경에서도 높은 예측 성능을 확보할 수 있는 가능성을 탐색하였다. 특히 여름철(8월) 데이터를 활용한 모델이 가장 우수한 성능을 보였으며, 기초모델 대비 전이모델의 정확도(Accuracy)가 0.022 향상되었다. SHAP 분석을 통해 예측 결과의 해석 가능성도 확보하였다. 전이모델로 예측된 수목 분포는 정주지 내 도심 가로수 밀집 지역을 중심으로 나타났으며, 실제 수목 위치와 높은 정합도를 보였다. 본 연구는 향후 고해상도 영상 기반 탐지, 극소량 데이터 환경에 대한 적용성 검증, 온실가스 인벤토리와 연계 연구 등 다양한 후속 연구에 기여할 수 있는 기반을 제공한다.

핵심어 : 수목 탐지, 전이학습, XGBoost, Sentinel-2, SHAP

Received 2025.04.03, Revised 2025.05.09, Accepted 2025.06.30

1) Member, MS student, Dept. of Geoinformatics, University of Seoul (E-mail: leejoon924@uos.ac.kr)

2) Researcher, Division for Environmental Planning, Water and Land Research Group, Korea Environmental Institute (Email: dbkim@kei.re.kr)

3) Research Specialist, Division for Environmental Planning, Water and Land Research Group, Korea Environmental Institute (Email: ghkim@kei.re.kr)

4) Corresponding Author, Member, Professor, Dept. of Geoinformatics, University of Seoul (E-mail: cmjun@uos.ac.kr)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. 서론

전 세계적으로 기후 위기에 대한 대응이 시급해지면서, 탄소중립 정책의 필요성과 실행력 확보가 주요 과제로 대두되고 있다. IPCC (Intergovernmental Panel on Climate Change)의 제 6차 평가보고서(AR6)에 따르면, 인간 활동으로 인해 1850-1900년 대비 지표면 온도는 약 1.1℃ 상승했으며, 2019년 온실가스 배출량은 1990년 대비 약 54% 증가한 것으로 나타났다(IPCC, 2023). 이에 따라 2015년 파리협정을 기점으로 국가별 온실가스 감축 의무가 강화되었으며, 감축뿐만 아니라 흡수량 산정을 포함한 체계적 탄소 관리가 강조되고 있다. 이러한 배경으로 국토교통부에서는 ‘탄소공간지도 시스템’을 구축하여 국가 단위의 탄소 중립 수준을 진단하고 있으며, 2022년 시행된 「탄소중립·녹색성장 기본법」을 통해 지자체에 대한 탄소중립 의무를 강화하며(기후위기 대응을 위한 탄소중립·녹색성장 기본법 제 1장 제 4조 1항), 지자체 온실가스 인벤토리 산정에 대한 요구를 증대시켰다. 이에 따라, 지역 단위에서의 온실가스 배출과 흡수량 산정의 정확도 확보가 정책적 요구로 부상하고 있다.

IPCC 가이드라인에 따르면, LULUCF (Land Use, Land Use Change, and Forestry) 부문은 탄소흡수원(예: 도시 수목, 산림 등)을 포함하는 주요 인벤토리 항목이며, 사람이 밀집해 생활하는 도시 공간 등의 정주지 내 수목 분포에 대한 정량적 정보는 인벤토리 산정을 위한 핵심 요소로 간주된다. 그러나 남한에서는 정주지의 공간적 정의에 대한 불명확성과 활동 자료의 부족은 해당 부문 온실가스 인벤토리 산정의 신뢰도를 저해하는 요인으로 작용하고 있다(Choi *et al.*, 2024). 특히 도시 지역은 전 세계 탄소 배출량의 약 75%를 차지하는 것으로 보고되며(Seto *et al.*, 2012), 정주지 내 탄소흡수원의 정보 확보는 지역 단위 온실가스 인벤토리의 정확도 제고를 위한 필수적 기반이다.

정주지 내 수목의 탄소 저장량 또는 흡수량을 정량화한 국내의 선행연구들이 다수 수행되었다. Guo *et al.*,(2014)은 중국 전역의 산림 외 수목(Trees Outside Forests, TOF)을 대상으로 30년간의 바이오매스 탄소 축적 변화를 분석하였으며, TOF가 국가 차원의 탄소 순환에서 유의미한 비중을 차지함을 실증하였다. Tang *et al.*,(2016)은 베이징의 가로수를 대상으로 탄소 저장 및 흡수 능력을 평가하였으나, 도시 수목이 도시 전체의 온실가스 배출을 상쇄하는 데에는 제한적인 기여만을 할 수 있음을 지적하였다. Raciti *et al.*,(2014)은 미국 보스턴시를 대상으로 고해상도 원격탐사 데이터와 현장조사를 통합하여 수관면적과 탄소 저장량을 추정하였고, 공간 해

상도에 따라 도시 지역 탄소량이 과소추정될 수 있음을 강조하였다. Nowak *et al.*(2013)은 미국 도시지역 전역을 대상으로 수목 탄소 저장 및 흡수량을 정량화하고, 도시림의 기후 변화 대응 기여도를 수치화하였다. 이러한 해외 사례와 달리, 국내의 경우 도시 내 탄소흡수원을 대상으로 한 정량적 연구는 상대적으로 적은 편이다. Choi *et al.*,(2024)은 국가 MRV (Monitoring-Reporting-Verification) 시스템과 토지이용변화를 활용해 정주지 범위를 정의하고, 해당 부문의 온실가스 인벤토리 산정을 위한 활동 데이터 구축 방안을 제시하였다. 또한, Lee *et al.*,(2023)은 위성영상 기반 식생지수와 XGBoost (eXtreme Gradient Boosting) 기법을 결합하여 수목 탐지 모델을 개발하였다. 그러나 기존 연구들은 대부분 현장조사에 의존하거나, 대규모 학습 데이터 확보를 전제로 한 고비용 접근법에 한정되는 경향이 있다. 한편, 앞서 언급한 국외 연구들 또한 대부분 국가 단위의 인벤토리 기반 또는 고해상도 현장조사 자료에 기반하여 수행되었으며, 제한된 수목 정보만을 보유한 지자체에서의 실질적인 적용 가능성은 충분히 검토되지 않았다. 따라서, 소규모 지자체에서도 활용 가능한, 데이터 요구량이 낮고 적용성이 높은 정주지 내 수목 탐지 모델 개발의 필요성은 여전히 존재한다.

이러한 한계를 극복하기 위한 대안으로 최근 다양한 분야에서 전이학습(transfer learning)이 주목받고 있다. 전이학습은 충분한 데이터로 학습된 모델의 지식을, 유사한 구조를 갖는 데이터가 부족한 영역에 적용함으로써 예측 성능을 확보하고 학습 비용을 절감할 수 있는 기계학습 기법이다. Jiang *et al.*,(2021)은 탄소 지중 저장소 모의 실험에 전이학습을 적용하여 고비용의 수치모형 없이도 높은 예측 정확도를 유지하면서 학습 비용을 80% 절감하였다. Wan *et al.*,(2024)은 SCADA (Supervisory Control And Data Acquisition) 시스템 데이터를 활용하여 한 풍력 터빈에서 학습된 XGBoost 기반 모델을 다른 터빈에 적용함으로써, 해상 풍력 터빈의 조기 화재를 효율적으로 탐지하였다. Fong *et al.*,(2020)은 대기질 예측 문제에서 전이학습을 통해 데이터가 부족한 지역 관측소의 예측 정확도를 높이고 학습 시간을 단축하였다. 도시 수목 탐지 분야에서도 전이학습의 활용 가능성이 제시되고 있다. Velasquez *et al.*,(2022)는 Google Street View 이미지를 활용해 도시 내 가로수를 자동 탐지하고 정밀하게 위치를 추정하는 전이학습 기반 모델을 제안하였으며, Xie *et al.*,(2022)는 드론 영상에 전이학습 기반 알고리즘을 적용하여 도시 수림 분포를 81.6%의 정확도로 탐지하였다. 이처럼 전이학습은 제한된 학습 데이터를 보유한 도시 또는 지자체에서도 정주지 내 탄소흡수원을 효과적으로 탐지할 수 있는 실질적인 대안으로

가능할 수 있음이 선행연구를 통해 보고되고 있다.

본 연구는 제한된 수목 정보만으로 정주지 내 탄소흡수원을 효과적으로 탐지할 수 있는 전이학습 기반 탐지 모델을 제안한다. 구체적으로, 서울특별시 수목 정보를 기반으로 학습된 기초 모델을 전이 학습 기법을 통해 대구광역시로 적용하고, 정주지 내 수목 분포를 추정함으로써 온실가스 인벤토리 산정에 기여하고자 한다.

이를 위해서 2장에서는 연구의 절차, 데이터 구축 방안, 기계학습 모델 구축 방법과 결과 해석 방법론을 다룬다. 3장에서는 기초모델, 전이모델의 성능과 SHAP (Shapley Additive Explanations) 방법론을 활용한 모델 결과 해석, 예측 결과 비교를 다루며 4장은 토의와 결론으로 구성하였다.

2. 방법론

2.1. 연구 흐름

본 연구는 정주지 내 수목 탐지를 위해 전이학습 기반의 기계학습 모델을 활용하였다. 전체적인 연구 절차는 Fig. 1과 같다. 먼저 서울특별시와 대구광역시를 대상으로 Sentinel-2 위성영상과 공공 수목 위치 데이터를 수집하여 수목-식생지수 학습 데이터를 구성하였다. 이후, 서울시 데이터를 기반으로 4개의 계절별 기초 모델을 학습하였으며, 동일한 시기의 대구시 데이터를 이용하여 전이학습을 수행하였다. 전이된 4개 모델의 성능을 비교하여 최종 모델을 선정하고, 해당 모델에 대해 SHAP 기법을 통해 변수 영향도를 분석하였다. 마지막으로, 수목 존재 확률 지도를 제작하고, 수목-비수목 이진 항목에 대한 예측지도를 생성하였다.

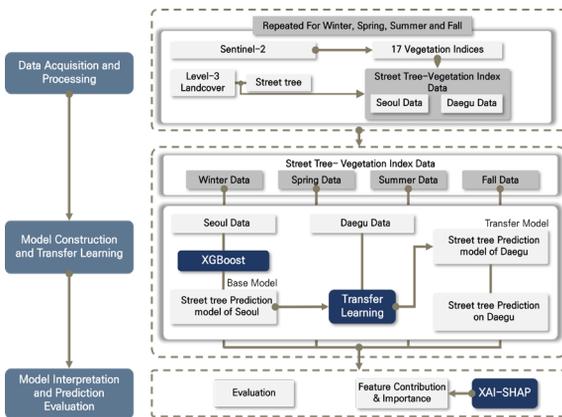


Fig. 1. Research procedure

2.2. 데이터 취득 및 처리

본 연구의 대상지는 서울특별시와 대구광역시로, 두 지역 모두 도시화가 높아 시가지 지역을 포함한 다양한 토지피복 특성을 보유하고 있다(Fig. 2). 현재 남한 내 정주지의 합의된 정의가 불분명함으로 다양한 도시 특성을 포함한 해당 대도시를 선정하였다. 위성영상은 Sentinel-2 Level 2A를 사용하였으며, 2024년 기준 4개 시점(1월, 4월, 8월, 11월)의 영상을 확보하였다. 영상은 모두 구름 피복률 10% 이하의 조건으로 선별되었다.

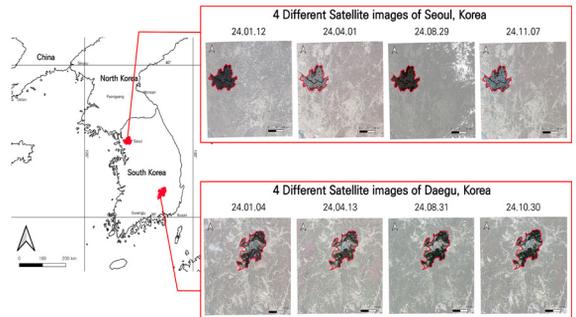


Fig. 2. Locations of the study areas and seasonal satellite images used for model training. Four Sentinel-2 satellite images selected across different seasons in 2024 for Seoul (top) and Daegu (bottom). Red boundaries indicate the administrative limits of each city

수목 위치정보는 가로수 식재 연도(2002~2003년)를 기준으로 서울시 약 56,000그루, 대구시 약 6,500그루의 공공 데이터를 활용하였다(Fig. 3). 현재 활용 가능한 가장 최신의 식재 연도 자료는 해당 시점으로 제한되며, 이는 본 연구가 최신 수목 데이터 확보가 어려운 지자체에서도 적용 가능한 수목 탐지 기법을 개발하고자 하는 연구 목적과 일치한다. 비수목 데이터는 환경부 토지피복지도에서 수목이 존재하지 않는 토지 유형⁵⁾을 추출하여 구성하였다. 기계학습 모델 학습 시 데이터 불균형을 방지하기 위하여, 수목 데이터와 동일한 샘플 수로 무작위 샘플링을 통한 수목:비수목 비율을 1:1로 설정하였다.

본 연구의 목적에 따라 활용한 모델은 수목-비수목 이진 분류 문제이며, 종속 변수는 수목의 존재 유무로 설정하였다. 독립변수는 NDVI (Normalized Difference Vegetation Index), RVI (Ratio Vegetation Index), NDMI (Normalized Difference Moisture Index), NDWI (Normalized Difference

5) 환경부 토지피복지도에서의 비수목 토지 유형으로는 단독주거시설, 공동주거시설, 공업시설, 상업업무시설, 혼합지역, 도로가 있다.

Table 1. Information of variables used as input dataset for XGBoost

Variable Type	Variables	Formula	Reference
Dependent	Public street tree location of Daegu, Korea		(https://www.data.go.kr/data/15109655/fileData.do)
	Public street tree location of Seoul, Korea		(https://data.seoul.go.kr/dataList/OA-1325/S/1/datasetView.do)
Independent	Plant Senescence Reflectance Index(PSRI)	$\frac{red - blue}{red - edge2}$	(Zhang <i>et al.</i> , 2018)
	Ratio Vegetation Index(RVI)	$\frac{NIR}{red}$	(Jordan, 1969; Pearson and Miller, 1972)
	Normalized Difference Vegetation Index red-edge 1(NDVIre1)	$\frac{NIR - red - edge1}{NIR + red - edge1}$	(Sharma <i>et al.</i> , 2015)
	Normalized Difference Vegetation Index red-edge 2(NDVIre2)	$\frac{NIR - red - edge2}{NIR + red - edge2}$	(Fernández-Manso <i>et al.</i> , 2016)
	Normalized Difference Vegetation Index red-edge 1 narrow(NDVIre1n)	$\frac{NIR - narrow - red - edge1}{NIR - narrow + red - edge1}$	(Fernández-Manso <i>et al.</i> , 2016)
	Normalized Difference Vegetation Index red-edge 2 narrow(NDVIre2n)	$\frac{NIR - narrow - red - edge2}{NIR - narrow + red - edge2}$	(Fernández-Manso <i>et al.</i> , 2016)
	Normalized Difference Vegetation Index red-edge 3 narrow(NDVIre3n)	$\frac{NIR - narrow - red - edge 3}{NIR - narrow + red - edge 3}$	(Navarro <i>et al.</i> , 2017)
	Normalized Difference Vegetation Index(NDVI)	$\frac{NIR - red}{NIR + red}$	(Rouse Jr <i>et al.</i> , 1973)
	Normalized Difference Red Edge(NDRE)	$\frac{NIR - red - edge3}{NIR + red - edge3}$	(Klemas and Smart, 1983; Nasiri <i>et al.</i> , 2022)
	Enhanced Vegetation Index(EVI)	$G^* \frac{red - edge3 - red}{L + red - edge3 + C_1 * red - C_2 * blue}$,where G : Gain factor, C_1, C_2 : coefficients of the aerosol resistance term, L : Soil adjustment factor	(Liu and Huete, 1995; Jiang, Z. <i>et al.</i> , 2008; Majasalmi and Rautiainen, 2018)
	Soil Adjusted Vegetation Index(SAVI)	$(1 + L) * \frac{NIR - red}{NIR + red + L}$,where L : Soil adjustment factor	(Huete, 1988; Erdenesumbee Sulderdene, <i>et al.</i> , 2017)
	Normalized Difference Moisture Index(NDMI)	$\frac{NIR - SWIR1}{NIR + SWIR1}$	(Xiao <i>et al.</i> , 2018; Lastovicka <i>et al.</i> , 2020)
	Moisture Stress Index(MSI)	$\frac{SWIR1}{NIR}$	(Hunt, 1989; Dotzler <i>et al.</i> , 2015)
	Normaliezed Difference Water Index(NDWI)	$\frac{NIR - green}{NIR + green}$	(Mcfeeters, 1996; Du <i>et al.</i> , 2016)
	Green Chlorophyll Index(GCI)	$\frac{NIR}{green - 1}$	(Gitelson <i>et al.</i> , 2003; Attaf <i>et al.</i> , 2024-04-15)
	Leaf Chlorophyll Index(LCI)	$\frac{NIR - red - edge1}{NIR + red}$	(Datt, 1999; Thenkabail <i>et al.</i> , 2000)
Structure Insensitive Pigment Index(SIPI)	$\frac{NIR - Aerosol}{NIR + red}$	(Pen Uelas <i>et al.</i> , 1995)	

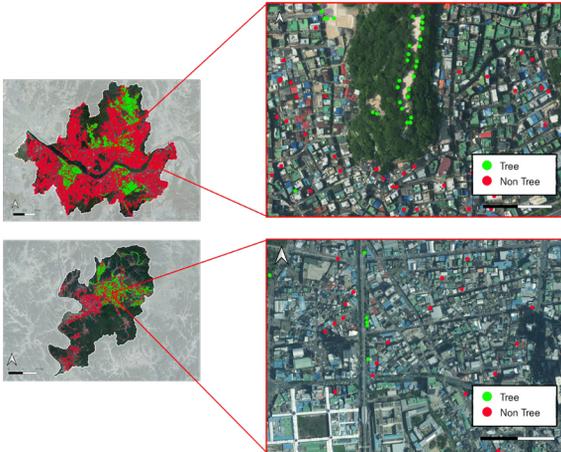


Fig. 3. Spatial distribution of tree and non-tree sample points used for model training. The maps show the locations of street trees (green) and sampled non-tree points (red) in Seoul (top) and Daegu (bottom)

Water Index) 등 위성영상의 각 밴드 정보를 활용한 17종의 식생지수를 포함한다(Table 1). 기계학습 모형의 입력자료로 구성하기 위해, 수목 위치 데이터와 해당 지점의 식생지수 값을 공간적으로 결합하여 정형화된 수목-식생지수 데이터를 구축하였다(Fig. 4). 정형 데이터는 지역(서울, 대구) 및 시기(계절 별 4개 시점)를 기준으로 총 8개 세트로 구축하였다.

2.3 모델 생성 및 전이학습

본 연구에서 기계학습 모델은 수목-비수목 이진 분류 문제

를 해결하기 위해 XGBoost⁶⁾ 기반의 기계학습 모델을 활용하였다. 모델 학습은 먼저 서울시 수목-식생지수 데이터를 기반으로 계절별 4개의 기초모델을 구축하는 방식으로 이루어졌다. 각 모델은 앞서 구축된 17종의 식생지수를 기반으로 한 독립변수로 학습되었으며, 분류 성능을 위해 무작위 탐색 기반의 하이퍼파라미터 튜닝을 수행하였다.

튜닝의 대상이 되는 XGBoost의 하이퍼파라미터는 Table 2에 나열된 총 8가지로, 모델의 부스팅되는 총 트리의 개수 ($n_estimators$), 각 트리의 최대 깊이(max_depth), 추후 학습될 트리에 대한 이전 트리들의 영향력을 조절하는 학습률 ($learning_rate$), 샘플링을 통해 트리를 더 일반화되게 만드는 $subsample$, $colsample_bytree$, 학습 시 각 트리 내 리프 노드의 수를 조절하는 정규화 항 $gamma$, 각 트리 내 리프 노드의 가중치에 대하여 L1, L2 정규화를 적용하는 정규화 항 reg_alpha , reg_lambda 이다(Chen and Guestrin, 2016).

Table 2. Hyperparameter search ranges for XGBoost model training

Hyperparameters	Search Options
$n_estimators$	200, 500, 800, 1000, 2000
max_depth	4, 6, 8, 10
$learning_rate$	0.01, 0.03, 0.05, 0.07, 0.1
$subsample$	0.7, 0.75, 0.8, 0.85, 0.9
$colsample_bytree$	0.7, 0.75, 0.8, 0.85, 0.9
$gamma$	0, 0.1, 0.3, 0.5
reg_alpha	0, 0.1, 0.5, 1
reg_lambda	0.1, 0.5, 1, 2

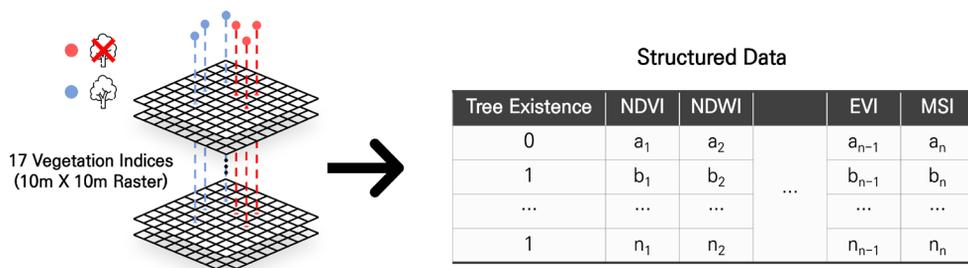


Fig. 4. Construction of tree-vegetation index structured dataset. 17 vegetation indices (10 m resolution) were extracted for each tree and non-tree location and merged into a tabular format. The final structured dataset includes binary labels for tree presence and corresponding vegetation index values used for model input

6) XGBoost는 Chen and Guestrin(2016)이 제안한 대표적인 Gradient Boosting 기반 앙상블 학습 알고리즘으로, 순차적으로 생성된 다수의 결정 트리를 통해 예측 성능을 향상시키는 방식이다. 각 트리는 이전 트리의 오차를 보완하며 학습되고, 전체 예측값은 트리별 출력의 가중합으로 계산된다. 또한 정규화 항을 포함한 목적함수를 사용해 과적합을 억제하며, 빠른 학습 속도와 대용량 데이터에 대한 효율적 처리로 다양한 머신러닝 실무 및 경진대회에서 널리 사용되고 있다.

하이퍼파라미터 탐색은 1000개의 하이퍼파라미터 조합을 탐색하도록 설정하였으며, 5회 교차검증을 기반으로 평가 기준은 정확도를 적용하였다. 모델 학습 시 과적합을 방지하고 일반화 성능을 확보하기 위해 정규화 파라미터와 subsample, colsample_bytree 비율 또한 함께 최적화하였다. 이 과정을 통해 계절별로 식생 특성을 반영한 기초 XGBoost 모델 4종을 도출하였다.

이후 동일한 구조와 하이퍼파라미터를 유지한 상태에서, 학습량(num_boost_round)만 조정하여 대구시 데이터를 활용한 전이학습을 수행하였다. 전이학습은 계절적 특성을 대응시키기 위하여 14일 내 시간적 차이를 가진 시기의 대구시 수목-식생지수 데이터를 서울시 기초모델에 연결하여 진행하였다(Table 3). 이는 식생지수 기반 feature space를 공유하면서도, 지역 및 환경적 특성이 상이한 지자체 간의 모델 일반화 가능성을 탐색하고자 한 것이다(Kou *et al.*, 2020).

Table 3. Seasonal training data configuration for base and transfer models. Each model was trained using season-matched Sentinel-2 imagery from Seoul (base) and Daegu (transfer)

Representing Season	Date of Base Model Training Data	Date of Transfer Model Training Data
Winter	2024.01.12	2024.01.04
Spring	2024.04.01	2024.04.13
Summer	2024.08.29	2024.08.31
Fall	2024.11.07	2024.10.30

본 연구는 특히 데이터 수급이 제한적인 정주지 환경에서도 적용 가능한 수목 탐지 기법 개발을 목표로 하기 때문에, 전이학습을 통해 적은 양의 라벨링된 수목 데이터를 가진 지역에서도 신뢰성 있는 예측 모델을 제시한다. 전이된 4개 모델의 예측 성능을 비교한 후, 최종 모델을 선정하고 후속 해석 및 시각화에 활용하였다.

2.4 모델 해석 및 평가

전이학습을 통해 도출된 최종 모델에 대해, 모델의 예측 메커니즘을 해석하고 예측 결과를 평가하기 위하여 XAI (eXplainable AI) 방법 중 SHAP 기법을 활용하였다. SHAP 기법은 각 독립변수가 예측 결과에 미치는 영향을 정량적으로 측정하며, 모델의 의사결정 과정에 대한 해석 가능성을 제공한다(Lundberg and Lee, 2017; Nohara *et al.*, 2022). 본 연구

에서는 수목 존재 여부를 결정하는 데 기여한 주요 식생지수를 파악하고, 변수 간 상호작용 구조를 분석하였다. 이를 통해 모델이 단일 지표 기준이 아닌, 식생 지수 간 복합적인 관계를 바탕으로 예측을 수행하고 있음을 탐색한다.

3. 연구 결과

3.1 기초모델 성능

서울시 수목-식생지수 데이터를 기반으로 4개의 XGBoost 기초모델을 학습하였다. 각 모델은 해당 계절의 식생 특성을 반영한 입력데이터로 구성되었으며, 성능 평가는 Accuracy, Precision, F1-Score를 기준으로 수행되었다. Table 4는 계절별 학습 데이터셋에 대한 튜닝된 하이퍼파라미터 조합을 나타낸다. 각 계절에 따라 상이한 최적의 파라미터 조합이 도출되었으며, 해당 설정은 이후 전이학습에도 동일하게 유지되었다.

Table 4. Tuned hyperparameters for seasonal XGBoost models. Each column shows the optimal set of hyperparameters obtained for the pretrained model trained on tree-vegetation index data from each season. These parameter settings were maintained for transfer learning

Hyperparameter	Winter (2024.01.12)	Spring (2024.04.01)	Summer (2024.08.29)	Fall (2024.11.07)
n_estimators	1000	1500	1000	1500
max_depth	8	8	8	8
learning_rate	0.03	0.05	0.05	0.05
subsample	0.8	0.8	0.8	0.8
colsample_bytree	0.75	0.7	0.75	0.7
gamma	0.1	0.1	0.5	0.1
reg_alpha	0.5	1.0	1	1
reg_lambda	0.5	0.5	0.5	0.5

모든 계절에서 F1-Score가 Precision과 Accuracy보다 높게 나타났으며, 특히 8월 모델이 가장 우수한 성능을 보였다(Table 5). 해당 모델은 3가지 평가 지표에서 모두 가장 높은 성능 수치를 기록하였고, 훈련셋과 검증셋 간 Accuracy 차이도 약 0.05 수준으로 과적합 현상이 보이지 않았다. 이는 여름철 식생 특성이 수목 탐지에 보다 유리한 정보를 제공함을 의미한다.

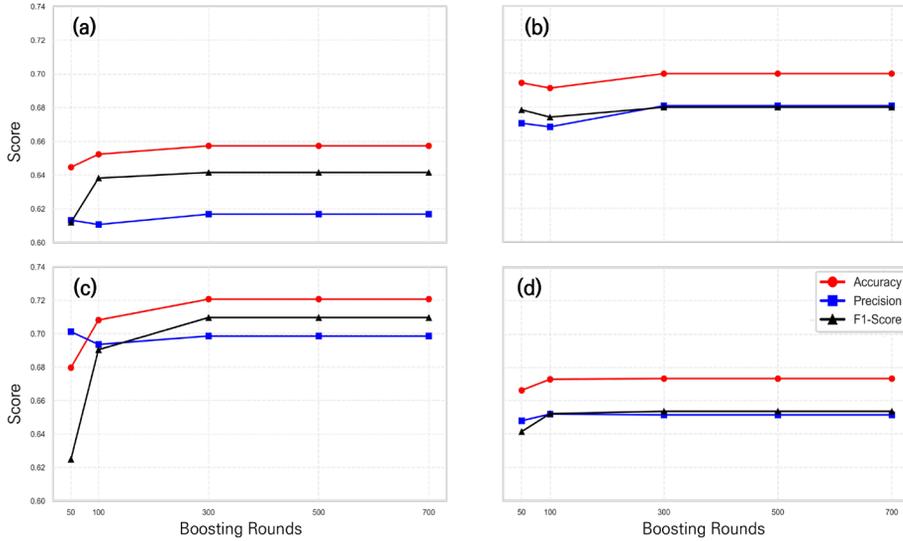


Fig. 5. Performance comparison of transfer models by training volume across four seasons: (a) January, (b) April, (c) August, and (d) November. Accuracy is shown in black, precision in red, and F1-score in blue line. The plots illustrate how performance metrics respond to increasing training iterations

Table 5. Performance metrics of seasonal base models trained with Seoul data. Four base models were trained using Sentinel-2 images from different seasons

Performance metrics	Winter (2024.01.12)	Spring (2024.04.01)	Summer (2024.08.29)	Fall (2024.11.07)
Train Accuracy	0.799	0.842	0.853	0.813
Test Accuracy	0.742	0.798	0.809	0.764
Test Precision	0.740	0.788	0.797	0.756
Test F1-score	0.748	0.803	0.817	0.771

3.2 전이모델 성능

기초 모델에서 학습된 파라미터 구조를 유지한 상태로, 각 계절별 서울시 모델에 대구시 데이터를 매칭시켜 전이학습을 수행하였다. 학습량(num_boost_round)은 성능 변화 관찰을 위해 조정되었으며, 최적 학습량은 계절에 따라 상이하게 나타났다.

Fig. 5는 계절별 학습량에 따른 전이 모델의 성능을 나타낸다. 모든 계절에서 학습량이 100을 초과하면 성능 향상이 정체되는 경향을 보이며, 학습량이 100일 때 성능과 과적합의 균형이 가장 우수한 것으로 나타났다. 학습량이 300 이상

시, 모든 계절에 있어서 더 이상의 성능 향상을 보이지 않았다. 이에 본 연구에서는 8월의 전이학습 모델을 최종 전이모델로 선정하였다.

대구시 수목-식생지수 데이터를 대상으로 모델 성능을 산출한 결과, 전이모델은 기초모델 대비 전반적인 예측 성능이 향상되었다(Table 6). 전이모델의 훈련셋 Accuracy는 0.815, 검증셋 Accuracy는 0.708로 산출되었다. 이 중 검증셋의 Accuracy가 기초모델의 검증셋 Accuracy인 0.686 대비 증가한 점을 확인하였고, F1-Score는 0.648에서 0.690으로 향상되었다. Precision은 0.703에서 0.694로 소폭 감소하였으나, 이는 모델이 보다 보수적인 판단을 수행함에 따라 발생한 자연스러운 결과이다.

Table 6. Classification performance of pretrained and transfer models for binary tree detection

	Base model	Transfer model
Accuracy	0.686	0.708
Precision	0.703	0.694
F1-score	0.648	0.690
Recall	0.601	0.710

전이학습을 통한 정주지 내 수목탐지의 성능향상을 확인

한 후, 대구시 수목-식생지수 데이터만을 학습한 XGBoost 기계학습 모델 대비 성능을 확인하였다. 이를 위하여 본 연구의 기초모델을 제작한 것과 동일한 방식의 학습 및 하이퍼파라미터 튜닝을 통해 대구광역시의 정주지 내 수목을 대상으로 하는 XGBoost 수목 탐지 모델을 제작하였다. Table 7는 Table 2와 동일한 탐색 옵션으로 튜닝된 하이퍼파라미터이다.

Table 7. Tuned hyperparameters for XGBoost model.
Each column shows the optimal set of hyperparameters obtained for the model trained on tree-vegetation index data from Daegu

Hyperparameter	XGBoost Classifier Trained on Daegu Tree-Vegetation Index data
n_estimators	100
max_depth	7
learning_rate	0.05
subsample	0.8
colsample_bytree	0.75
gamma	0.3
reg_alpha	1
reg_lambda	0.5

Table 8은 대구학습모델의 성능을 나타낸다. 학습과 평가가 동일 지역에서 수행되었기 때문에 높은 정확도를 보이는 것은 자연스러운 결과로 해석할 수 있다. 그러나 이러한 성능은 충분한 학습 데이터를 확보할 수 있다는 전제가 있을 때만 가능하며, 실제 지자체 현장에서는 고비용의 현장조사 및 데이터의 부재로 인해 이러한 조건을 만족시키기 어렵다. 이는 탄소흡수원 정보 확보를 위한 정량적 수단이 제한적임을 의미하며, 특히 「탄소중립·녹색성장 기본법」에 따라 지자체 차원의 온실가스 인벤토리 산정이 의무화된 상황에서는 그 실효성 확보가 중요한 과제로 부상하고 있다.

Table 8. Classification performance of XGBoost model trained on tree-vegetation index data from daegu

	XGBoost Classifier Trained on Daegu Tree-Vegetation Index data
Train Accuracy	0.982
Test Accuracy	0.786
Test Precision	0.785
Test F1-score	0.787

이에 본 연구는 제한된 학습 데이터를 갖는 지역에서도 식생 탐지가 가능하도록, 데이터가 풍부한 지역에서 학습된 모델을 다른 지역으로 적용하는 전이학습 기반의 모델을 제안하였다. 실험 결과, 전이학습 모델은 절대적인 평가 지표 수치에서는 동일 지역 학습 모델보다 낮았으나, 훈련셋 Accuracy(0.815)와 검증셋 Accuracy 간 차이가 크지 않아 보다 안정적인 일반화 성능을 보였다.

3.3 SHAP을 통한 모델 해석

Fig. 6은 전이모델의 예측구조를 해석하기 위해 SHAP 기법을 적용한 결과이다. RVI가 가장 큰 영향력을 가지는 변수로 나타났고, 그 뒤 NDMI와 SIPI 순으로 나타났다. RVI 값이 높을수록 수목 존재 확률이 증가하는 양의 방향성을 보였으며, NDMI는 수분 함유량을 반영하는 지표로써 수분이 많을수록 수목 존재 가능성이 높게 예측되었다. NDWI는 수체 탐지에 사용되는 지표로 수분 함량에 민감하게 반응하지만, 시가화 지역에서 근적외선을 흡수하는 건축물에 높은 값을 가질 수 있다는 특성과 교차되어 해석이 복잡한 변수였으나, SHAP 분석에서는 수목 존재 확률에 양의 방향성을 가지는 것으로 나타났다.

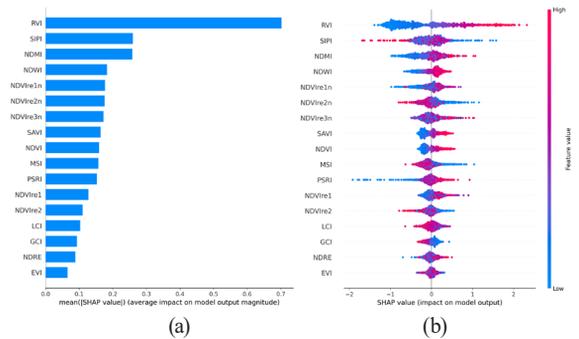


Fig. 6. SHAP summary plots for the transfer model: (a) average importance of independent variables based on mean absolute SHAP values, (b) SHAP value distributions for each variable. Pink indicates higher input values and blue indicates lower ones

변수 간 상호작용 분석에서는 전이모델이 변수를 단독으로 활용하지 않고, 주요 변수들과의 상호작용을 고려하여 예측하였다(Fig. 7). NDWI를 단독으로 활용하기 보다는 NDMI, RVI 등 주요 변수들과의 결합을 통해 수목 여부를 판단하는데 기여하는 영향을 보였다. NDWI가 낮고 NDMI가 높은 경우, NDWI가 보완적 역할을 수행하였으며, 반대로 RVI가 높

은 경우 NDWI나 NDMI의 영향이 낮더라도 수목으로 예측되는 경향을 보였다.

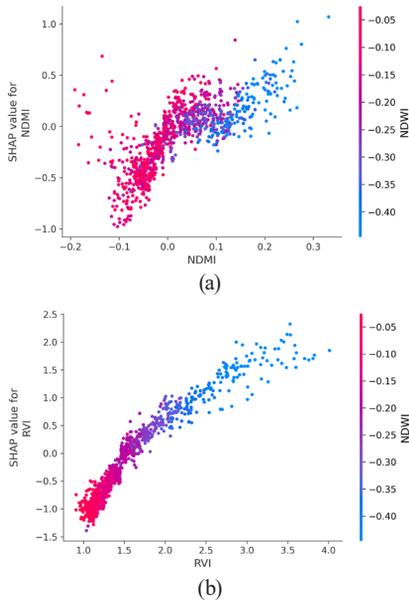


Fig. 7. SHAP-based interaction analysis between independent variables: (a) NDMI and NDWI, (b) RVI and NDWI. The X-axis represents the value of a primary variable, and the Y-axis shows its SHAP value, indicating the variable's contribution to the model output. Color denotes the value of a secondary interacting variable, allowing the visualization of interaction effects between the two

3.4 예측 결과 및 비교 평가

전이모델의 예측 성능과 공간적 분포 양상을 탐색하기 위해, 대구광역시 대상으로 수목 존재 확률 지도 및 이진 예측 지도를 생성하였다. Fig. 8은 기초 모델과 전이모델을 활용하

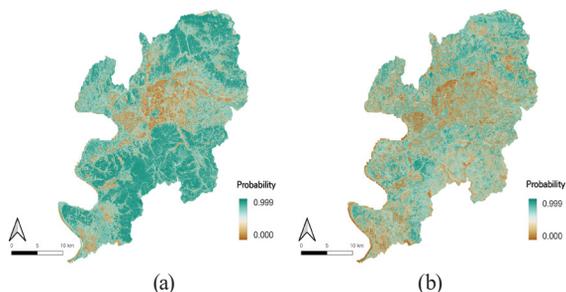


Fig. 8. Predicted probability of tree presence in Daegu: (a) using the base model, (b) using the transfer model

여 대구광역시 수목 존재 확률 지도를 생성한 결과이다. 전반적으로 기초 모델은 산림지역과 같은 비정주지 구간에 대해서 예측을 집중하였고, 전이모델은 기초모델 대비 예측 값이 도심 가로수 밀집 지역에 집중되는 경향을 보였다.

본 연구에서 대구광역시 내 수목으로 분류된 격자의 예측 확률값의 분포를 분석한 결과(Fig. 9), 전이모델은 기초모델보다 더 높은 평균값과 낮은 표준편차를 나타내며 확률분포의 분리도가 뚜렷하였다. 기초모델의 경우, 수목 전체 확률값이 전반적으로 균일하게 분포되며 정규분포에 가까운 형태를 보인 반면, 전이모델은 우측으로 비대칭적인 형태의 분포를 보였다. 분류 성능 50% 이상을 기준으로 수목-비수목을 분류하는 판단 기준을 산출하였을 때, 기초모델에서는 판단 기준이 0.585, 전이모델에서는 0.825로 산정되어, 수목 존재에 대한 판단 기준이 모델 간 명확하게 구분됨을 확인할 수 있다.

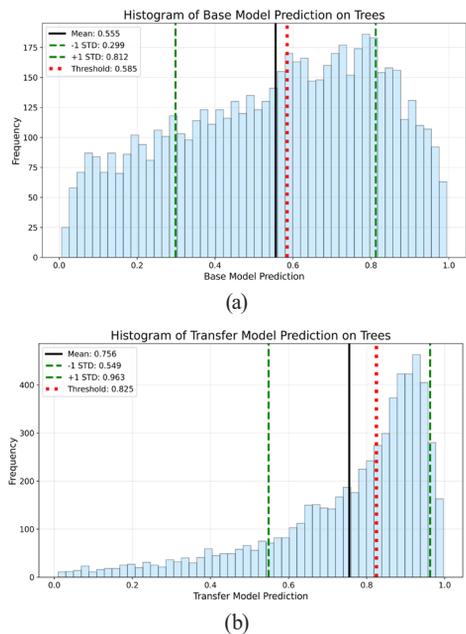


Fig. 9. Histogram of predicted tree existence probabilities: (a) base model, (b) transfer model. The solid black, green dashed, and red dashed lines indicate the mean probability, ± 1 standard deviation, and the 50% threshold of predicted tree probability, respectively

이진 분류 결과는 Fig. 10에 시각화하였다. 두 모델 간 예측 대상의 공간적 특성이 상이하게 나타났다. 기초 모델은 수목이 밀집된 산림 지역을 대부분 수목으로 분류한 반면, 전이모

텔은 도심 내 주요 가로수 밀집 구역에 대한 탐지에 집중되는 경향을 보였다.

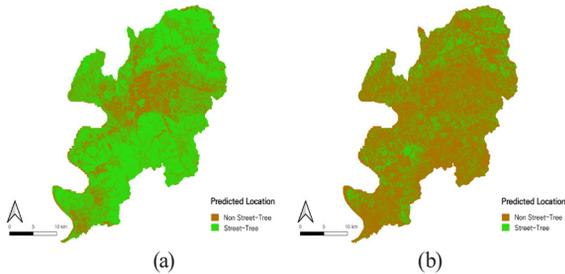


Fig. 10. Binary classification maps of street trees in Daegu: (a) base model, (b) transfer model. The maps show predicted locations of street trees (green) and non-tree areas (brown) using a 50% probability threshold

4. 토의

4.1 기초모델과 전이모델 간 예측 성향 비교

기초모델과 전이학습 간의 예측 성향 차이를 정성적으로 비교하기 위해, 대구광역시 내 두 지점(선돌공원 인근, 계명대역 인근)을 사례로 선정하였다. 먼저 중부내륙고속지선과 대명천이 평행하게 지나가는 선돌공원 인근 사례에서는 기초모델이 True color 영상상 수목으로 추정되는 영역보다 넓은 지역을 수목으로 예측하였으며, 건축물을 수목으로 오인한 사례

도 다수 확인되었다(Fig. 11). 계명대역 인근에서도 유사한 경향이 나타났다. 기초모델은 수목 예측 영역이 넓게 퍼져 있는 반면, 전이모델은 보다 집중된 예측 결과를 보였으며, 두 모델 모두 그림자 영역에 대해서는 오인 탐지가 발생하였다. 이러한 비교 결과를 종합하면, 전이학습 기반의 전이모델은 정주지 내 수목 탐지에 있어 기초모델보다 더 신중하고 정밀한 예측 성향을 보이는 것으로 확인되었다. 수목 밀집 지역에서는 기초 모델이 광범위한 탐지를 수행한 반면, 전이모델은 개별 수목 단위의 탐지 특성을 보였다. 또한 도로, 상업, 주거 지역 등 정주지 환경에서는 기초모델 대비 향상된 정도를 보였으며, 건축물이 밀집된 복잡한 공간 구조에서도 보다 높은 예측 성능을 나타냈다.

4.2 대구학습모델과 전이모델 간 예측 성향 비교

대구학습모델과 전이모델 간의 예측 성향 차이를 정성적으로 파악하기 위하여 기초 모델 및 전이모델에 적용된 동일한 50% 판단 기준을 산출하여 적용하였다. 이는 4.1과 동일한 지점을 사례로 선정하여 수행되었다. Fig. 12는 그 결과이다. 대구학습모델은 전이모델 대비 보수적인 예측을 수행하였으며, Fig. 12 (a)의 선형으로 구성된 공원에 대한 탐지 결과를 통해 수목이 선형적으로 군집되어 있는 지역에 대하여 오탐지가 되는 경향을 확인하였다. 허나, 수목이 선형적으로 군집된 영역을 제외한 탐지 영역에 대하여 비슷한 탐지 결과가 산출되었다. 본 연구에서 제안된 전이모델의 기초 모델 대비 향상된 성



Fig. 11. Qualitative comparison of model predictions in two sample areas: (a) Seondol Park, (b) Daegu Keimyung Station. Each row shows true color imagery, prediction of the base model, and prediction of the transfer model from left to right. The base model tends to overpredict tree locations and misclassify buildings, while the transfer model yields more conservative and spatially consistent predictions

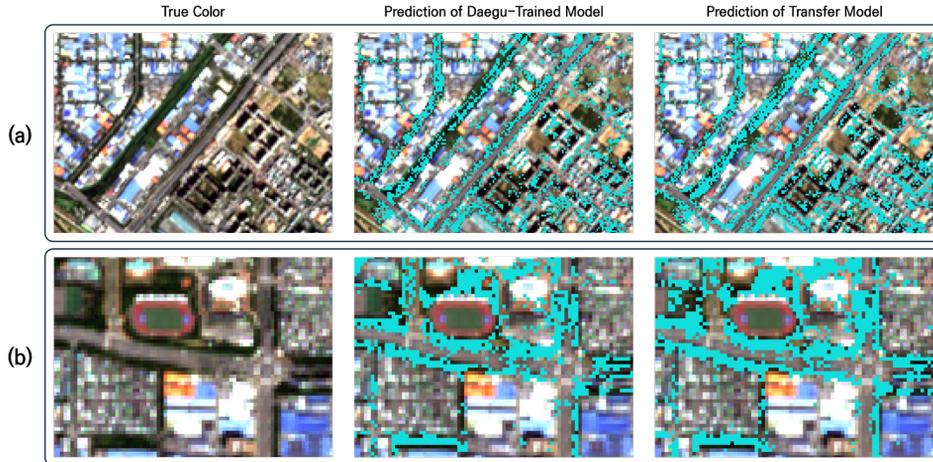


Fig. 12. Qualitative comparison of model predictions in two sample areas: (a) Seondol Park and (b) Daegu Keimyung Station. Each row shows true color imagery, prediction of XGBoost Classifier trained on Daegu tree-vegetaion index data, and prediction of the transfer model from left to right. The Daegu-trained model tends to underpredict tree locations on linear vegetation cluster

능과 대구학습모델 대비 안정적인 일반화 성능과 같은 특성은 정주지 내 수목 정보를 활용한 탄소흡수원 탐지의 정확성과 신뢰성을 높이는 데 기여할 수 있으며, 제한된 수목 위치 정보만을 보유한 지자체에서도 온실가스 인벤토리 산정의 정량적 기반으로 활용될 수 있음을 시사한다.

4. 결론

본 연구는 최신 수목 정보가 부족한 정주지가 포함된 지자체에서도 활용 가능한 수목 탐지 방법론을 제시하고자, 전이 학습 기반 XGBoost 모델을 구축하여 정주지 내 탄소흡수원(수목)의 공간적 분포를 탐지하는 모델을 제안하였다. 기초 모델의 학습 데이터로는 서울특별시의 약 56,000 그루의 가로수 위치 정보를 활용하였으며, 이를 대구광역시 내 6,500그루의 가로수 위치정보에 대하여 전이학습시킨 모델의 일반화 가능성과 적용성을 탐색하였다. 식생지수의 계절적 특성을 고려하기 위해 각 4개의 계절에 대한 기초, 전이모델들을 학습 및 비교하였으며, 본 연구에서는 8월(여름)이 전이학습 효과에 제일 적합한 계절로 분석되었다. 전이학습의 정량적 성능을 파악하기 위해 기초모델, 전이모델 그리고 대구학습모델의 대구광역시에 대한 예측 성능을 비교한 결과, 기초모델 대비 전이모델은 정확도(Accuracy)는 0.022 향상되고 보다 보수적인 판단이 수행된 것, 대구학습모델 대비 안정적인 일반화 성능을 보인 것을 확인하였다. 또한, 대구광역시 데이터만을 이용

하여 학습된 XGBoost 모델과의 비교를 통해 전이모델이 보다 안정적인 일반화 성능을 보인 것을 확인하였다. 이후 수행된 SHAP 분석 결과 전이모델의 예측은 합리적인 결과를 보였다. 전이모델로 예측된 수목의 공간적 특성은 정주지의 수목탐지에 집중하는 경향을 보였다.

본 연구에서 제안한 전이학습 기반 수목 탐지 모델은 정주지 내 탄소흡수원 탐지에 있어 가능성을 보였으나 다음과 같은 한계점이 존재한다. 첫째, 수목 위치정보는 서울시 및 대구시의 2002~2003년도 가로수 데이터로 기반으로 구축되었다. 이로 인해 위성영상(2024년 기준)과의 시간적 불일치가 존재한다. 산림청 고시 제2021-130호에 고지된 경상북도의 평균생장률이 2.6%인 것을 참고하여 대구광역시 내 2003년 대비 2024년에 추가적으로 식재된 수목의 비율을 계산할 시, 약 1.7%의 수목이 더 식재되었을 수 있다. 이렇듯 해당 시차는 예측 정확도에 영향을 줄 수 있는 요인이다. 그러나 이러한 제약은 동시에, 최신 현장조사 기반 데이터가 부족한 지자체 환경에서 적용 가능한 기법 개발이라는 본 연구의 목적을 강조한다고도 할 수 있다. 둘째, 예측 정확도는 도로변 및 건축물 외곽 수목에 대해서는 상대적으로 높았으나, 산림처럼 수목이 군집된 지역 내부에 대해서는 예측 성능이 낮게 나타났다. 이는 해당 지역들이 본 연구의 적용 범위인 정주지 외의 공간 특성을 가지기 때문으로 모델의 설계 목표와 일치하는 결과로 해석할 수 있다. 셋째, 본 연구에서 활용한 대구시 수목 데이터는 약 6,500 그루로, 서울시 대비 약 12% 수준이다. 이치

럼 제한된 학습량에서도 예측 성능을 유지되었으나, 1,000 그루 이하 수준의 극소량 라벨링 데이터 환경에서도 유사한 성능을 확보할 수 있는지는 향후 연구를 통해 추가적으로 검증이 필요하다.

본 연구의 결과 및 한계를 바탕으로 다음과 같은 후속 연구 방향을 제안한다. 첫째, 위치 정확도 및 해상도 개선이 요구된다. 현재 10m 해상도의 위성영상을 개별 수목을 정확히 식별하기에는 공간 해상도가 제한적이며, 수목 위치정보와의 불일치 가능성도 존재한다. 향후에는 고해상도 위성영상(예: WorldView, KOMPSAT 등)을 활용한 탄소흡수원 탐지 연구가 필요하다. 둘째, 극소량 데이터 환경에 대한 성능 검증이 요구된다. 본 연구보다 더 제한된 수목 정보를 보유한 지자체에 대한 전이학습 기반 탐지기법의 적용 가능성을 검토할 필요가 있다. 셋째로, 온실가스 인벤토리와의 연계 가능성 평가를 할 필요가 있다. 본 연구에서 제시된 수목 탐지 결과를 기반으로, 정주지 단위에서 온실가스 인벤토리 산정까지 연계하는 적용성을 평가할 수 있는 실증 연구가 필요하다.

감사의 글

이 성과는 정부(환경부)의 재원으로 한국환경산업기술원의 탄소중립 특성화대학원 사업의 지원을 받아 수행된 연구임

References

- Attaf, D., Mahi, H., and Mansour, D. (2024). Investigating the Relationship Between SAR and Optical Data Indices for Cereal Monitoring, *Proceedings of 2024 IEEE Mediterranean and Middle-East Geoscience and Remote Sensing Symposium (M2GARSS)*, IEEE, 15-17 April, Oran, Algeria, pp. 338-341.
- Cantarello, E., Newton, A. C., and Hill, R. A. (2011), Potential effects of future land-use change on regional carbon stocks in the UK, *Environmental Science & Policy*, Vol. 14, No. 1. pp. 40-52.
<https://doi.org/10.1016/j.envsci.2010.10.001>
- Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system, *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Association for Computing Machinery, 13-17 August, San Francisco, United States, pp. 785-794.
<https://doi.org/10.1145/2939672.2939785>
- Choi, S., Kim, M., Son, Y., Jeon, S., Lee, K., Kim, W., Lee, S., and Lee, W. (2024), Development of activity data for greenhouse gas inventory in settlements in south korea, *Land*, Vol. 13, No. 4. 497p.
<https://doi.org/10.3390/land13040497>
- Datt, B. (1999), A New Reflectance Index for Remote Sensing of Chlorophyll Content in Higher Plants: Tests using Eucalyptus Leaves, *Journal of Plant Physiology*, Vol. 154, No. 1. pp. 30-36.
[https://doi.org/10.1016/S0176-1617\(99\)80314-9](https://doi.org/10.1016/S0176-1617(99)80314-9)
- Dotzler, S., Hill, J., Buddenbaum, H., and Stoffels, J. (2015), The Potential of EnMAP and Sentinel-2 Data for Detecting Drought Stress Phenomena in Deciduous Forest Communities, *Remote Sensing*, Vol. 7, No. 10. pp. 14227-14258.
<https://doi.org/10.3390/rs71014227>
- Du, Y., Zhang, Y., Ling, F., Wang, Q., Li, W., and Li, X. (2016), Water Bodies' Mapping from Sentinel-2 Imagery with Modified Normalized Difference Water Index at 10-m Spatial Resolution Produced by Sharpening the SWIR Band, *Remote Sensing*, Vol. 8, No. 4. pp. 354-19.
<https://doi.org/10.3390/rs8040354>
- Fernández-Manso, A., Fernández-Manso, O., and Quintano, C. (2016), SENTINEL-2A red-edge spectral indices suitability for discriminating burn severity, *International Journal of Applied Earth Observation and Geoinformation*, Vol. 50, pp. 170-175.
<https://doi.org/10.1016/j.jag.2016.03.005>
- Fong, I. H., Li, T., Fong, S., Wong, R. K., and Tallon-Ballesteros, A. J. (2020), Predicting concentration levels of air pollutants by transfer learning and recurrent neural network, *Knowledge-Based Systems*, Vol. 192, pp. 105622-10.
<https://doi.org/10.1016/j.knosys.2020.105622>
- Friedl, M. A. and Davis, F. W. (1994), Sources of variation in radiometric surface temperature over a tallgrass prairie, *Remote sensing of Environment*, Vol. 48, No. 1. pp. 1-17.
[https://doi.org/10.1016/0034-4257\(94\)90109-0](https://doi.org/10.1016/0034-4257(94)90109-0)
- Friedman, J. H. (2001), Greedy function approximation: A gradient boosting machine, *Annals of Statistics*, Vol. 29, No. 5. pp. 1189-1232.

- Gitelson, A. A., Gritz, Y., and Merzlyak, M. N. (2003), Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves, *Journal of Plant Physiology*, Vol. 160, No. 3. pp. 271-282.
- Guo, Z. D., Hu, H. F., Pan, Y. D., Birdsey, R. A., and Fang, J. Y. (2014), Increasing biomass carbon stocks in trees outside forests in China over the last three decades, *Biogeosciences*, Vol. 11, No. 15. pp. 4115-4122.
<https://doi.org/10.1078/0176-1617-00887>
- Huang, S., Tang, L., Hupy, J. P., Wang, Y., and Shao, G. (2021), A commentary review on the use of normalized difference vegetation index (NDVI) in the era of popular remote sensing, *Journal of Forestry Research*, Vol. 32, No. 1. pp. 1-6.
<https://doi.org/10.1007/s11676-020-01155-1>
- Huete, A. R. (1988), A soil-adjusted vegetation index (SAVI), *Remote Sensing of Environment*, Vol. 25, No. 3. pp. 295-309.
[https://doi.org/10.1016/0034-4257\(88\)90106-X](https://doi.org/10.1016/0034-4257(88)90106-X)
- Hunt, E. R. and Rock, B. N. (1989), Detection of changes in leaf water content using near-and middle-infrared reflectances, *Remote Sensing of Environment*, Vol. 30, No. 1. pp. 43-54.
[https://doi.org/10.1016/0034-4257\(89\)90046-1](https://doi.org/10.1016/0034-4257(89)90046-1)
- IPCC (2023), Climate Change 2023: Synthesis Report, Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, IPCC Publication No. AR6-9789291691647, IPCC, Geneva, 184p.
<https://10.59327/IPCC/AR6-9789291691647>
- Jeong, S., Hwang, J., Kim, I., and Jeon, S. (2024), Analysis of management practices in Korea for enhanced greenhouse gas inventory in LULUCF sector, *Journal of the Korean Society of Environmental Restoration Technology*, Vol. 27, No. 6. pp. 1-14. (in Korean with English abstract)
<https://doi.org/10.13087/kosert.2024.27.6.1>
- Jiang, S., Tang, H., Fu, P., and Jo, H. (2021) A Transfer Learning-Based Surrogate Model for Geological Carbon Storage with Multi-Fidelity Training Data, *Climate Change AI*, Pittsburgh, <https://www.climatechange.ai/papers/neurips2021/24> (last date accessed: 3 April 2025)
- Jiang, Z., Huete, A. R., Didan, K., and Miura, T. (2008), Development of a two-band enhanced vegetation index without a blue band, *Remote Sensing of Environment*, Vol. 112, No. 10. pp. 3833-3845.
<https://doi.org/10.1016/j.rse.2008.06.006>
- Johnson, J. M., and Khoshgoftaar, T. M. (2019), Survey on deep learning with class imbalance, *Journal of Big Data*, Vol. 6, No. 1. pp. 1-54.
<https://doi.org/10.1186/s40537-019-0192-5>
- Jordan, C. F. (1969), Derivation of leaf-area index from quality of light on the forest floor, *Ecology*, Vol. 50, No. 4. pp. 663-666.
<https://doi.org/10.2307/1936256>
- Klemas, V. and Smart, R. (1983), The influence of soil salinity, growth form, and leaf moisture on the spectral radiance of Photogram, *Eng. Remote Sens*, Vol. 49, No. 1. pp. 77-83.
- Kou, R., Fang, B., Chen, G., and Wang, L. (2020), Progressive domain adaptation for change detection using season-varying remote sensing images, *Remote Sensing*, Vol. 12, No. 22.
<https://doi.org/10.3390/rs12223815>
- Lastovicka, J., Svec, P., Paluba, D., Kobiuk, N., Svoboda, J., Hladky, R., and Stych, P. (2020), Sentinel-2 data in an evaluation of the impact of the disturbances on forest vegetation, *Remote Sensing*, Vol. 12, No. 12.
<https://doi.org/10.3390/rs12121914>
- Lee, J., Han, Y., Lee, J., Park, J., and Kim, G. (2023), Development of tree detection methods for estimating LULUCF settlement greenhouse gas inventories using vegetation indices, *Korean Journal of Remote Sensing*, Vol. 39, No. 6_3. pp. 1721-1730. (in Korean with English abstract)
<https://doi.org/10.7780/kjrs.2023.39.6.3.5>
- Liu, H. Q. and Huete, A. (1995), A feedback based modification of the NDVI to minimize canopy background and atmospheric noise, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 33, No. 2. pp. 457-465.
<https://10.1109/TGRS.1995.8746027>
- Liu, W., Liu, W. D., and Gu, J. (2020), Predictive model for water absorption in sublayers using a joint distribution adaptation based XGBoost transfer learning method, *Journal of Petroleum Science and Engineering*, Vol. 188.

- <https://doi.org/10.1016/j.petrol.2020.106937>
Lundberg, S. M. and Lee, S. (2017), A unified approach to interpreting model predictions, *Advances in Neural Information Processing Systems* 30, NIPS, 4-9 December, Long beach, USA, pp. 4765-4774.
<https://doi.org/10.48550/arXiv.1705.07874>
- Majasalmi, T. and Rautiainen, M. (2018), The potential of sentinel-2 data for estimating biophysical variables in a boreal forest: A simulation study, *Remote Sensing Letters*, Vol. 7, No. 5. pp. 427-436.
<http://dx.doi.org/10.1080/2150704X.2016.1149251>
- Mcfeeters, S. K. (1996), The use of the normalized difference water index (NDWI) in the delineation of open water features, *International Journal of Remote Sensing*, Vol. 17, No. 7. pp. 1425-1432.
<https://doi.org/10.1080/01431169608948714>
- Nasiri, V., Darvishsefat, A. A., Arefi, H., Griess, V. C., Sadeghi, S. M. M., and Borz, S. A. (2022), Modeling forest canopy cover: A synergistic use of sentinel-2, aerial photogrammetry data, and machine learning, *Remote Sensing*, Vol. 14, No. 6.
<https://doi.org/10.3390/rs14061453>
- Navarro, G., Caballero, I., Silva, G., Parra, P. C., Vázquez, Á., and Caldeira, R. (2017), Evaluation of forest fire on madeira island using sentinel-2A MSI imagery, *International Journal of Applied Earth Observation and Geoinformation*, Vol. 58, pp. 97-106.
<https://doi.org/10.1016/j.jag.2017.02.003>
- Nohara, Y., Matsumoto, K., Soejima, H., and Nakashima, N. (2022), Explanation of machine learning models using shapley additive explanation and application for real data in hospital, *Computer Methods and Programs in Biomedicine*, Vol. 214, 106584p.
<https://doi.org/10.1016/j.cmpb.2021.106584>
- Nowak, D. J., Greenfield, E. J., Hoehn, R. E., and Lapoint, E. (2013), Carbon storage and sequestration by trees in urban and community areas of the united states, *Environmental Pollution*, Vol. 178, pp. 229-36.
<https://doi.org/10.1016/j.envpol.2013.03.019>
- Pearson, R. L. and Miller, L. D. (1972), Remote mapping of standing crop biomass for estimation of the productivity of the shortgrass prairie, *Proceedings of the eighth international symposium on remote sensing of environment*, 2-6 October, Ann Arbor, Michigan, USA