Ecological Indicators 167 (2024) 112671

Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind

Mapping forest tree species and its biodiversity using EnMAP hyperspectral data along with Sentinel-2 temporal data: An approach of tree species classification and diversity indices

Rajesh Vanguri [a,*], Giovanni Laneve [b], Agata Hościło [c]

[a] Department of Astronautical, Electric and Energy Engineering (DIAEE), Sapienza University of Rome, Rome, Italy
[b] School of Aerospace Engineering, EOSIA Lab, Sapienza University of Rome, Italy
[c] Institute of Environmental Protection – National Research Institute, National Centre for Emissions Management, Warsaw, Poland

ARTICLE INFO

Keywords:
Tree species classification
Biodiversity indices
Hyperspectral data
Remote sensing
Sentinel-2
EnMap

ABSTRACT

Forests play a crucial role in maintaining ecological balance and biodiversity, making the accurate mapping of tree species and assessment of biodiversity indices essential for informed management decisions. This study introduces an innovative methodology that integrates EnMAP (Environmental Mapping and Analysis Program) hyperspectral data with Sentinel-2 multitemporal data to classify tree species in the biodiverse landscapes of Kampinos National Park and its surrounding regions in Poland.

We extract essential vegetation indices such as NDVI, NDMI, SAVI, and EVI from Sentinel-2 data to assess forest health and dynamics. The Sentinel-2 data is upscaled from 10 m to 30 m to align with EnMAP's spatial resolution, followed by precise co-registration of the images using QGIS. Utilizing a rich dataset from the National Forest Inventory, we extract spectral signatures of nine distinct tree species from both data sources. We employ five machine learning algorithms—Support Vector Machines (SVM), Random Forest (RF), CatBoost (CAT), Gradient Boosting Classifier (GBC), and XGBoost (XGB)—to enhance classification accuracy.

Through iterative experimentation with data reduction techniques and algorithm tuning, we achieve optimal performance across needle-leaved and broad-leaved species. The resulting tree species maps are validated through quantitative accuracy assessments against mixed-species polygons from the National Forest Inventory and ground truthing in the Kampinos National Park. Achieving an overall accuracy of 85% to 93%, our study demonstrates the efficacy of this integrated approach in tree species mapping. Furthermore, the tree species maps serve as a foundation for deriving key biodiversity indices—species richness, Shannon-Wiener Diversity Index, Simpson's Diversity Index, and a composite Biodiversity Index—providing insights into spatial biodiversity patterns and informing targeted conservation strategies. This study exemplifies the potential of combining advanced remote sensing techniques with field validation to enhance our understanding of forest ecosystems and guide sustainable management practices.

1. Introduction

Forests are crucial in upholding biodiversity and delivering ecosystem services that are fundamental for human welfare and the preservation of the environment. Accurate mapping of forest tree species composition and assessing biodiversity are crucial endeavours that hold significant implications for effective forest management, conservation efforts, and understanding the intricate relationships between forest ecosystems and their inhabitants (Langmaier et al., 2023; Nagendra, 2001; Pausas & Austin, 2001). It is also important in maintaining biodiversity, providing habitats for numerous plant and animal species, and contributing to ecological processes such as carbon sequestration, nutrient cycling, and climate regulation (Lindenmayer et al., 2000; Noss, 1999; Fahey et al., 2018; Liang et al., 2016). However, biodiversity is highly sensitive to natural and anthropogenic factors, such as climate change, deforestation, habitat fragmentation, and invasive species, making it imperative to monitor and quantify changes in forest composition and diversity over time (Foley et al., 2005; Sala et al., 2000; Vitousek et al., 1997; Brockerhoff et al., 2017).

Recent research has made significant strides in tree species

* Corresponding author.
E-mail address: rajesh.vanguri@uniroma1.it (R. Vanguri).

https://doi.org/10.1016/j.ecolind.2024.112671
Received 13 June 2024; Received in revised form 10 September 2024; Accepted 25 September 2024
Available online 1 October 2024

classification and biodiversity assessment through the application of advanced remote sensing technologies and machine learning algorithms. For instance, studies have shown that integrating hyperspectral and multispectral data can enhance the accuracy of species classification, allowing for the identification of subtle spectral differences among tree species (Fang et al., 2020; Marconi et al., 2022). Moreover, the use of deep learning techniques has emerged as a promising approach for automating tree species classification, significantly improving efficiency and precision in complex forest environments (Sun et al., 2019; Zhong et al., 2024). These advancements underscore the importance of continuous research and innovation in the field to address the challenges posed by changing environmental conditions. Despite these advancements, gaps remain in the application of these methods across different ecological contexts. The current research aims to bridge this gap by focusing on the biodiverse landscapes of Kampinos National Park in Poland.

Recent advancements in tree species classification have primarily leveraged data from airborne remote sensing technologies, such as the APEX hyperspectral data. APEX is a dispersive push-broom system with a 28-degree field of view developed by a Swiss–Belgian consortium under the framework of the ESA-PRODEX program (Itten et al., 2008). Studies utilizing APEX data have demonstrated its effectiveness in capturing detailed spectral information, enabling researchers to classify tree species with considerable accuracy. For instance, (Raczko and Zagajewski, 2017) employed APEX to classify common tree species in the Szklarska Poręba town, achieving an accuracy of 77 % using neural networks. Similarly, UAV-based hyperspectral sensors have gained traction in forest monitoring as well, providing high-resolution imagery that enhances classification precision. (Wang et al. 2023) Utilized UAV LiDAR and hyperspectral data in the Maoershan forest area, achieving a classification accuracy exceeding 79.91 % through various machine learning algorithms. These studies illustrate the significant progress made in the field, yet they also highlight the limitations associated with the scale and coverage of airborne data.

The introduction of the DLR Earth Sensing Imaging Spectrometer (DESIS), mounted on the International Space Station, marks a significant advancement in hyperspectral remote sensing. DESIS covers the spectral range of 402 – 1000 nm, which is narrower compared to airborne sensors. While previous studies have focused on tree species classification, recent research using DESIS has explored plant species richness prediction rather than direct species mapping. For example, the study (Guo et al., 2023) demonstrates the sensor's capability to assess biodiversity indirectly. However, while DESIS provides a broader spatial coverage, it often comes with lower spatial resolution compared to airborne sensors and also it covers a narrower spectral region compared to the EnMAP satellite which will cover 430—2450 nm.

The advent of hyperspectral remote sensing technology, combined with multitemporal multispectral data, has revolutionized our ability to map forest tree species and assess biodiversity at unprecedented scales (Dalponte et al., 2012; Fassnacht et al., 2016; Somers & Asner, 2014). Hyperspectral sensors capture detailed spectral information across hundreds of narrow wavebands, enabling the identification of unique spectral signatures for different tree species and their biochemical properties (Dalponte et al., 2012; Somers & Asner, 2014). When combined with multitemporal multispectral data, which captures the seasonal variations in vegetation phenology, these datasets provide a powerful tool for accurate tree species classification and biodiversity assessment (Grabska et al., 2019; Xi et al., 2021; Immitzer et al., 2012; Somers et al., 2011; Hościło and Lewandowska, 2019).

Mapping forest tree species and quantifying biodiversity have numerous applications in forest management, conservation planning, ecosystem monitoring, and ecological research (Hernández-Stefanoni et al., 2014; Rocchini et al., 2015; Turner et al., 2003). For instance, in the event of natural disasters or disturbances, forest managers can use these maps to identify areas where specific tree species have been impacted and plan targeted reforestation efforts to maintain ecosystem balance (Hemmerling et al., 2021; Pausas et al., 2004). Additionally, biodiversity indices derived from tree species maps can inform conservation strategies, highlighting areas of high biodiversity that require protection or restoration efforts (Gillespie et al., 2008; Nagendra & Rocchini, 2008; Féret & Asner, 2014).

Furthermore, accurate tree species mapping and biodiversity assessment contribute to our understanding of ecosystem processes, such as carbon sequestration, nutrient cycling, and habitat suitability for various species (Fahey et al., 2018; Liang et al., 2016; Paquette & Messier, 2011). This knowledge is crucial for developing sustainable forest management practices and mitigating the impacts of environmental changes (Brockerhoff et al., 2017; Lindner et al., 2010; Millar et al., 2007). Moreover, these techniques can aid in monitoring the dynamics of soil moisture and its relationship with climate patterns in semi-arid regions , further enhancing our understanding of ecosystem resilience and vulnerability.

In this study, we aim to leverage newly developed spaceborne hyperspectral satellite data to address the challenges faced by previous research. Unlike airborne sensors, spaceborne data offers the advantage of large coverage, enabling us to apply our models at a national level (Future work). However, this approach presents certain disadvantages, such as reduced resolution and the lack of structural visibility of trees compared to UAV-collected data. We have employed five different machine learning models, combining their strengths to enhance classification performance unlike the previous studies where they examined each model separately.

2. Study area, data used and preprocessing

2.1. Study area

The study area for this research encompasses the Kampinos National Park and its surrounding areas in Mazowian region, located near the city of Warsaw in central Poland (Fig. 1). The study area covers the area of Kampinos National Park is the second-largest national park in Poland, covering an area of approximately 385 square kilometers, it is a biosphere reserve and Natura 2000 site (Grabska-Szwagrzyk et al., 2024, Kampinos National Park website:https://kampn.gov.pl/kampinoski-park-narodowy#first). The park is characterized by a diverse range of habitats, including forests, wetlands, dunes, and meadows, supporting a rich biodiversity of flora and fauna, making it an important conservation area.

Approximately 70 % of Kampinos National Park is covered with forests, and of that forested area, around 67 % is comprised of pine trees, and the common tree species are pine (Pinus sylvestris), birch (Betula spp.), oak (Quercus spp.), and alder (Alnus spp.) (Andrzejewska et al. 2018). The age of the forests in the study area varies, with the presence of both old-growth and younger stands, providing a range of ecological niches and habitats for various species.

The Kampinos National Park is located in the area of the largest Polish hydrological node – point of convergence of valleys of Vistula, Bug, Narew, Wkra and Bzura rivers, and therefore is recognised as ecological corridor of high European importance. The climate in the Kampinos Forest is classified as continental, characterized by warm summers and cold winters, with large daily temperature amplitudes exceeding 30 °C. Additionally, ground frost occurs during the growing season, leading to lower average air temperatures and higher relative humidity (Owadowska et al., 2013).

The mean annual temperature ranges from 7 °C to 9 °C, with significant seasonal variations. Precipitation is moderate, with an average annual rainfall of approximately 550 mm, distributed throughout the year. The diverse habitats within the study area, including the forested areas, wetlands, and dunes, support a rich biodiversity of plant and animal species, making it an important conservation area.

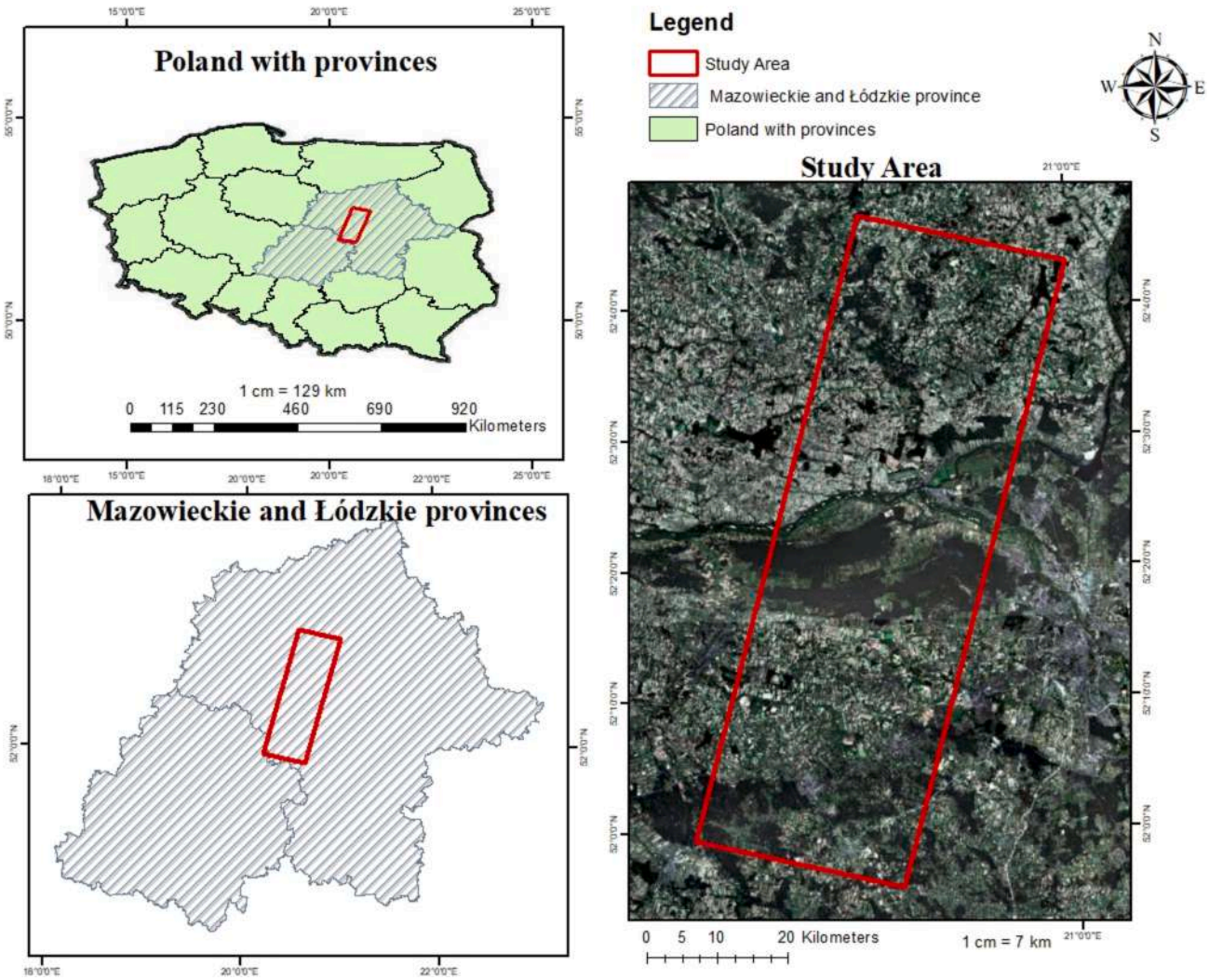

Fig. 1. Study Area with orthophotomap on the background.

2.2. Data used

For this study, we utilized both multispectral and hyperspectral satellite data to map forest tree species and assess biodiversity. Specifically, we employed 10 images from the Sentinel-2 multispectral satellite (Gatti and Bertolini, 2018) and 3 images from the EnMAP hyperspectral mission.

The Sentinel-2 mission, part of the Copernicus program, consists of two identical satellites (Sentinel-2A and Sentinel-2B) equipped with the MultiSpectral Instrument (MSI). The MSI captures data in 13 spectral bands ranging from visible to shortwave infrared, with spatial resolutions of 10 m, 20 m, and 60 m depending on the band (Drusch et al., 2012). The area of interest in our study covered two tiles (T34UDC and T34UDD), necessitating the use of 10 Sentinel-2 images to ensure complete coverage.

On the other hand, the Environmental Mapping and Analysis Program (EnMAP) is a hyperspectral satellite mission developed by the German Aerospace Center (DLR). EnMAP carries a hyperspectral imager that acquires data in 224 spectral bands spanning the visible to shortwave infrared range (420–2450 nm) with a spectral sampling of 6.5 nm in the visible and near-infrared and 10 nm in the shortwave infrared (Guanter et al., 2015, Segl et al., 2015). The sensor has a spatial resolution of 30 m and a swath width of 30 km (Stuffler et al., 2007). For our study area, we utilized 3 EnMAP hyperspectral images.

The combination of multispectral Sentinel-2 data, which provides frequent revisit times and broad spatial coverage, and hyperspectral EnMAP data, which offers detailed spectral information for accurate material identification, presents a powerful approach for mapping forest tree species and assessing biodiversity (Fassnacht et al., 2016).

2.2.1. Sentinel-2 data preprocessing

The Sentinel-2 data used in this study consisted of 10 images covering the two tiles, T34UDC and T34UDD, which encompass the area of interest. Table 1 provides details about the Sentinel-2 images, including the acquisition dates, tile numbers, and processing levels and Table 2 provides list of indices and bands used.

Since the area of interest spans across two tiles, a mosaic operation

Table 1
Sentniel-2 images used for study.

No.	Date	Tile Number	Processing Level
1.	2023-04-22	T34UDC	L2A
2.	2023-04-22	T34UDD	L2A
3.	2023-06-03	T34UDC	L2A
4.	2023-06-03	T34UDD	L2A
5.	2023-08-15	T34UDC	L2A
6.	2023-08-15	T34UDD	L2A
7.	2023-09-06	T34UDC	L2A
8.	2023-09-06	T34UDD	L2A
9.	2024-01-29	T34UDC	L2A
10.	2024-01-29	T34UDD	L2A

Table 2
List of Indices and Bands used from Sentinel-2.

S. No	Indices	S. No	Band
1.	Aerosol Free Vegetation Index (AFRI_1600)	1.	B2 (Blue-490 nm)
2.	Canopy Chlorophyll Content Index (CCCI)	2.	B3 (Green-560 nm)
3.	Green Chlorophyll Index (CIgreen)	3.	B4 (Red-665 nm)
4.	Red Edge Chlorophyll Index (CIrededge)	4.	B5 (Red Edge-705 nm)
5.	Enhanced Vegetation Index (EVI)	5.	B8 (NIR- 842 nm)
6.	Normalized Difference Vegetation Index with MIR (NDVI_MIR)	6.	B11 (SWIR-1 1610 nm)
7.	Normalized Difference Vegetation Index (NDVI)	7.	B12 (SWIR-2 2190 nm)
8.	Soil-Adjusted Vegetation Index (SAVI)		
9.	Normalized Difference Moisture Index (NDMI)		

was performed to combine the overlapping tiles into a single seamless image. The extent of the EnMAP images (List of them mentioned in Table 3) was used to extract the specific area of interest from the mosaicked Sentinel-2 data. This process resulted in a total of 5 Sentinel-2 images covering the study area, which were further utilized for subsequent analysis.

The preprocessing of Sentinel-2 data followed a systematic approach, as illustrated in the flowchart (Fig. 2).

I. The first step involved selecting the bands of interest, namely B2 (Blue), B3 (Green), B4 (Red), B5 (Red Edge), B8 (NIR), B11 (SWIR-1), and B12 (SWIR-2), which cover the visible, near-infrared, and shortwave infrared regions of the electromagnetic spectrum (Sentinel-2 2015).
II. In the second step, the spatial resolution of all selected bands was downscaled to 30 m to match the resolution of the EnMAP hyperspectral data. This step was necessary because the original spatial resolutions of the Sentinel-2 bands vary between 10 and 20 m, depending on the band.
III. The third step involved converting the digital numbers (DNs) of the selected bands to bottom-of-atmosphere (BOA) reflectance values.
IV. The fourth step involved calculating various vegetation indices from the selected bands. A total of nine vegetation indices (Huete et al., 2002) were computed, including:
- Aerosol Free Vegetation Index (AFRI_1600) (Karnieli et al., 2001): B8 − 0.66 * (B11/(B8 + 0.66 * B11))
- Canopy Chlorophyll Content Index (CCCI) (Gitelson et al., 2005): ((B8 − B5)/(B8 + B5))/((B8 − B4)/(B8 + B4))
- Green Chlorophyll Index (CIgreen) (Gitelson et al., 1996): (B8/B3) − 1
- Red Edge Chlorophyll Index (CIrededge) (Gitelson et al., 2003): (B8/B5) − 1
- Enhanced Vegetation Index (EVI): 2.5 * (B8 − B4)/((B8 + 6 * B4 − 7.5 * B2) + 1)
- NDVI with MIR (NDVI_MIR): (B12 − B8)/(B12 + B8)
- Normalized Difference Vegetation Index (NDVI): (B8 − B4)/(B8 + B4)
- Soil-Adjusted Vegetation Index (SAVI) (Huete, 1988): ((B8 − B4)/(B8 + B4 + 0.5)) * 1.5
- Normalized Difference Moisture Index (NDMI) (Gao, 1996): (B3 − B12)/(B3 + B12)

V. The final step involved stacking the selected bands (7 layers) and the calculated vegetation indices (9 layers) into a single 16-layer image for each Sentinel-2 scene.

Table 3
List of EnMAP images used.

No	Date	Processing level	Version
1.	20230928T102021	L2A	001
2.	20230928T102025	L2A	002
3.	20230928T102029	L2A	003

2.2.2. EnMAP data preprocessing

The EnMAP hyperspectral data underwent a streamlined pre-processing approach, as depicted in the flowchart (Fig. 3). The first step involved acquiring the EnMAP hyperspectral images covering the study area.

In the second step, co-registration of the EnMAP images with the Sentinel-2 data was performed using the Basic Pixel Alignment tool in the Co-registration plugin of QGIS. This step ensured that the EnMAP and Sentinel-2 data were spatially aligned, facilitating their integration and subsequent analysis.

The third and final step involved removing the zero bands present in the EnMAP data. Specifically, five bands in the EnMAP hyperspectral images contained zero values, which were identified and removed to optimize the data for further processing.

The version numbers indicate the order in which the images were processed, with Image 1 being the first version, followed by Image 2 and Image 3. So, in summary, these three images are part of the same tile (DT0000044865), captured in close succession, and processed in a specific order, as indicated by the acquisition times and version numbers.

After completing these pre-processing steps, the EnMAP hyperspectral data were ready for integration with the pre-processed Sentinel-2 data, enabling the subsequent tree species classification and biodiversity assessment.

2.2.3. Reference data preprocessing

The reference data for this study was obtained from the Polish Forest Data Bank (FDB), a national forest inventory database. The FDB contains information on forest stands, represented as polygons, with detailed attributes on the tree species composition and their respective coverage percentages within each stand.

Each FDB polygon represents a forest stand, and the attribute table provides the percentage share of different tree species within that stand. The species share is expressed using a scale ranging from 1 to 10, where 10 indicates a homogenous coverage of 100 % for a particular species. However, the precise spatial distribution of these species within the stand remains uncertain.

To ensure the reliability of the reference data, several preprocessing steps were undertaken. First, only the FDB polygons representing pure stands with a single species dominance of 90 % or more were selected. Due to the high frequency of Scots pine (Pinus sylvestris) in Polish forests, 10 % of the pure stands with a 100 % share of this species were randomly chosen for inclusion.

The next step involved aligning the reference samples with the actual forest mask any reference samples or their parts that fell outside the forest mask were removed to ensure the accuracy of the training and validation data.

The final reference dataset covers four different regions in Poland and includes nine tree species, as well as two subspecies of oak. This comprehensive reference data, combined with the preprocessing steps, provides a robust foundation for the subsequent tree species mapping and biodiversity assessment using the EnMAP hyperspectral and Sentinel-2 multispectral data (Table 4).

3. Methodology

The methodology followed a systematic approach, as illustrated in the flowchart (Fig. 4). The first step involved co-registering the EnMAP hyperspectral images with the Sentinel-2 data using the Basic Pixel Alignment tool in QGIS. This ensured spatial alignment between the two

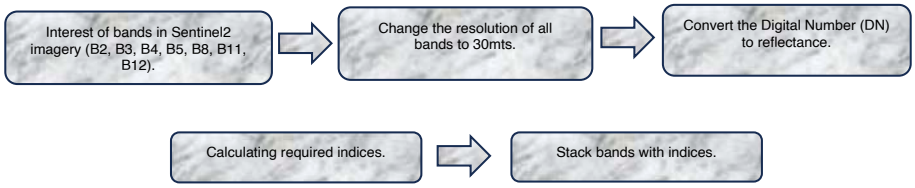

Fig. 2. Sentinel-2 images preprocessing flowchart.

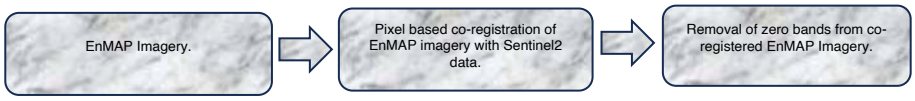

Fig. 3. EnMAP images preprocessing flowchart.

Table 4
Classes considered for training and testing.

No.	Tree Species	No. of Polygons	No. of Pixels
1.	Black locus (Robinia pseudoacacia L.)	6	25
2.	European Beech (European beech—Fagus sylvatica L.)	5	50
3.	Birch (silver birch—Betula pendula Roth and downy birch—Betula pubescens Ehrh.)	96	1610
4.	Oak	160	5883
4a.	Northern Red Oak (Quercus rubra)	2	29
4b.	Pedunculate Oak (English oak—Quercus robur L. and sessile oak—Quercus petraea (Matt.) Liebl.)	7	179
5.	Larch (European larch—Larix decidua Mill.)	12	533
6.	Alder (gray alder—Alnus incana (L.) Moench and black alder—Alnus glutinosa Gaertn.)	183	3884
7.	Aspen (Populus sp.)	5	69
8.	Pine (Scots pine – Pinus sylvestris L.)	30	7204
9.	Spruce (Norway spruce—Picea abies (L.) H. Karst)	8	97

datasets, facilitating their integration for subsequent analysis.

Next, the pre-processed EnMAP and Sentinel-2 data were stacked, resulting in a combined dataset with 299 features (219 from EnMAP and 80 from Sentinel-2) for each pixel. Reference data from the National Forest Inventory, which are in the form of polygons representing different tree species, were used to extract spectral signatures from the stacked dataset. Preprocessing steps were applied to this reference data, including replacing outliers with averages and handling missing values (NaNs) using the same approach.

The reference data were then separated into two data frames: one for the three needle-leaf species and another for the six broad-leaf species. Five different machine learning algorithms were chosen to evaluate their performance in tree species classification: Random Forest, Support Vector Machine (SVM), Gradient Boosting Classifier (GBC), CatBoost Classifier, and XGBoost.

Machine Learning Algorithms

- Random Forest (RF) (Breiman, 2001): An ensemble learning method that constructs multiple decision trees and combines their predictions, providing robust and accurate results for classification tasks.
- Support Vector Machine (SVM) (Cortes and Vapnik, 1995): A supervised learning algorithm that finds the optimal hyperplane separating different classes, making it effective for both classification and regression problems.
- Gradient Boosting Classifier (GBC) (Friedman, 2001): An ensemble technique that combines weak learners (decision trees) in an iterative manner, with each subsequent model attempting to correct the errors of the previous one.
- CatBoost Classifier (CBC) (Prokhorenkova et al., 2018): A gradient boosting algorithm that uses ordered target encoding and oblivious trees, making it efficient for handling categorical features and reducing overfitting.
- XGBoost (XGB) (Chen and Guestrin, 2016): An optimized implementation of gradient boosting that employs parallel processing and advanced regularization techniques, resulting in improved computational performance and model accuracy.

To address the high dimensionality of the feature space (299 features), three data reduction techniques were employed: Gaussian Random Projection, Independent Component Analysis (ICA), and Principal Component Analysis (PCA). These techniques aimed to reduce the number of features while preserving the most relevant information for classification. As shown in Fig. 5, ICA outperformed the other two methods and was selected for further processing.

The machine learning algorithms were trained and tested separately for needle-leaf and broad-leaf species, as the tuning parameters and class distributions differed between these two groups. Hyperparameter tuning was performed to optimize the performance of each algorithm. A forest mask Fig. 6 was used to differentiate between forest and non forest.

Once the algorithms were trained and tuned, they were applied to the full stacked image (approximately 90 km × 30 km) to predict tree species. First, a forest/non-forest filter was applied to the image, followed by separate broad-leaf and needle-leaf filters. The trained algorithms were then used to predict tree species on the filtered images, and the results were combined to produce a final classified image.

Following the successful production of the species map, this study aims to further analyse the forest ecosystem by calculating various biodiversity indices. These indices will provide valuable insights into the diversity of the forest, enabling a comprehensive understanding of the ecological dynamics within the study area. Specifically, this analysis will involve the calculation of the Shannon-Wiener Index, Simpson's Diversity Index, Species Richness, and biodiversity index. These indices will be derived from the species map, allowing for a detailed examination of the forest's biodiversity and its spatial distribution.

Shannon-Wiener Index: The Shannon-Wiener index, also known as Shannon entropy, quantifies the species diversity within a given area by considering both the number of species present and their relative abundance. It is calculated as $-\Sigma(p_i * \ln(p_i))$, where p_i represents the proportion of individuals of the i-th species in the total population. This index provides insights into the richness and evenness of species in the ecosystem (Pipinis & Radoglou, 2024; Ette et al., 2023).

Simpson's Diversity Index: Simpson's diversity index measures the dominance or evenness of species in a community. It is calculated as $1 - \Sigma(p_i^2)$, where p_i represents the proportion of individuals of the i-th species. A higher Simpson's index indicates lower diversity due to dominance by a few species, while a lower value signifies higher diversity (Ette et al., 2023; Purvis & Hector, 2000).

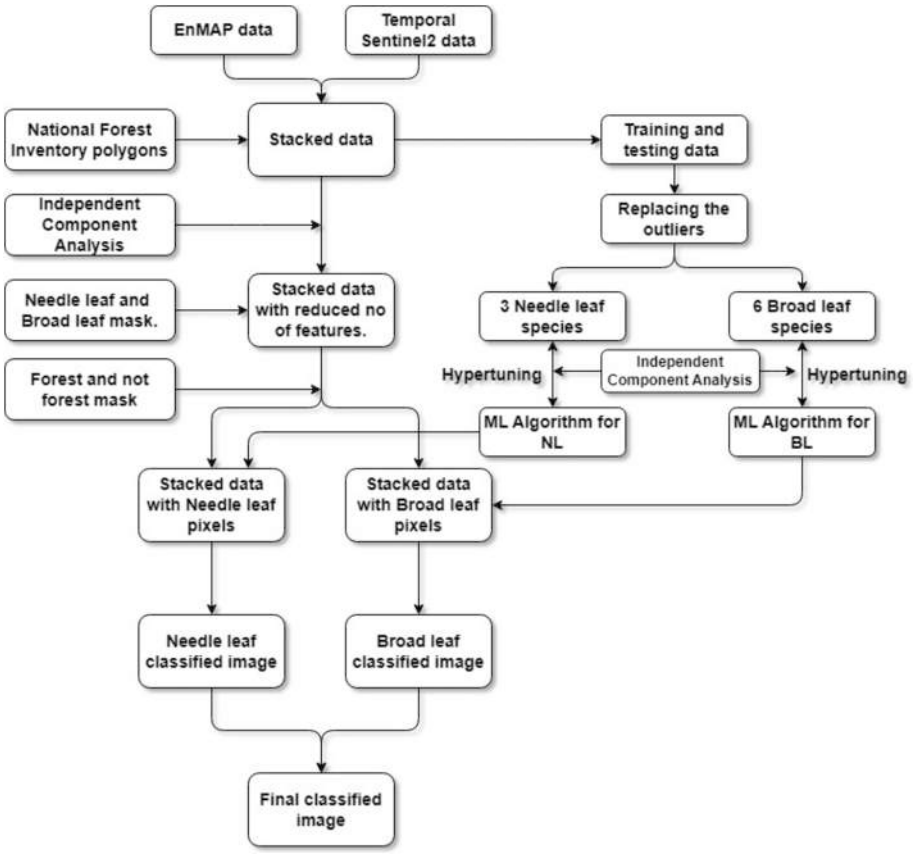

Fig. 4. Workflow for tree species classification.

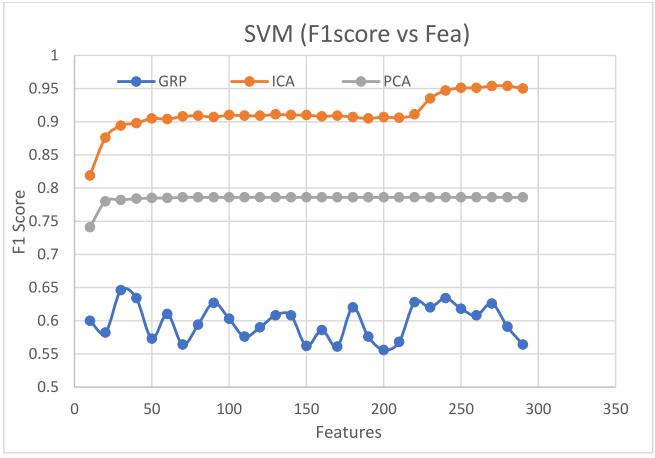

Fig. 5. Model for feature selection.

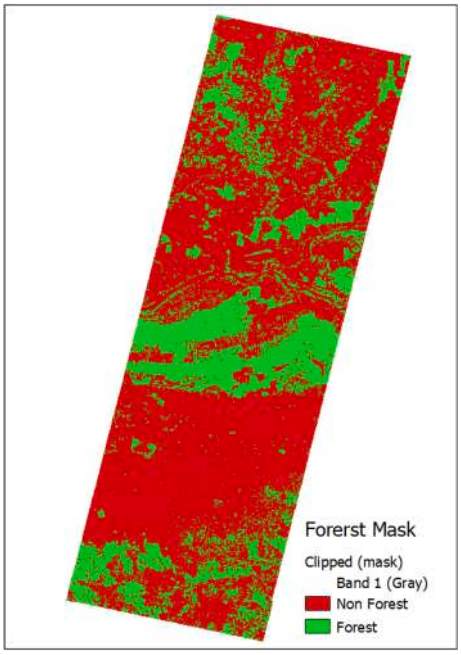

Fig. 6. Forest Mask.

Species Richness: Species richness simply refers to the number of different species present in a given area. It provides a basic measure of biodiversity by counting the total number of unique species observed within the study area. Higher species richness indicates greater biodiversity and ecosystem complexity (Heym et al., 2020; Laurila-Pant et al., 2015).

Biodiversity Index: The custom biodiversity index developed involves calculating the ratio of the number of different species observed within the 9-pixel area to the total number of species in the study area. For example, if out of the 9 pixels contain 2 different species, the index would be 2/9 (0.2222). This index offers a simplified measure of biodiversity based on the presence of unique species within the defined area.

To assess the performance of the tree species classification models, we employed several evaluation metrics, including overall accuracy, precision, recall, and F1-score (mainly F1-score presented as a result). The dataset was split into training and validation samples using a 70:30

ratio, with 70 % of the data allocated for training the models and 30 % reserved for testing their performance.

After training the models, we evaluated their performance on the validation dataset. Overall accuracy was calculated as the ratio of correctly classified instances to the total number of instances. Precision and recall were computed for each species to assess the model's ability to correctly identify true positives while minimizing false positives and false negatives. The F1-score, which is the harmonic mean of precision and recall, was also calculated to provide a balanced measure of model performance. Due to space constraints, we have included only the F1-score results in the Results section of this paper.

4. Results & discussion

4.1. Results

In this section, we present the classification results obtained from five different machine learning algorithms. Each algorithm was trained separately for needle leaf and broad leaf classes to account for their distinct characteristics. Notably, due to the limited number of classes in needle leaf species (three classes), no tuning was required for this part of the algorithm, resulting in consistently high accuracies exceeding 90 %. Conversely, tuning efforts were focused on the broad leaf part of the algorithms to optimize performance. We begin by providing an overview of the overall accuracy and class-specific accuracies achieved by each algorithm, followed by a detailed analysis of the classification outcomes.

The results of the classification models were analysed to evaluate the F1 scores across the five different algorithms with varying numbers of features (50–150). Fig. 7 illustrates the F1 scores of the algorithms as the number of features increases. It is observed that the F1 scores for all algorithms remain consistently high, ranging between 0.98 and 0.99 across the different feature sets.

The minimal variation in F1 scores indicates a high level of stability and consistency in the performance of the classification models as the number of features changes. This behaviour of F1 scores also suggests that the algorithms maintain similar levels of accuracy and robustness across the feature sets tested.

The results for broad leaf species indicate that the F1 score remains relatively consistent, with minimal fluctuations across the feature sets. This suggests that the optimal number of features for this classification task is 50, as the F1 score does not significantly improve or worsen with increased features.

Fig. 8 illustrates the F1 score variation across the different feature sets. The scores range from 0.86 to 0.93, with the majority of the algorithms exhibiting a stable performance. The Random Forest (RF), CatBoost (CBC), and Extreme Gradient Boosting (XGB) algorithms demonstrate a slight decrease in F1 score as the number of features increases, indicating that the optimal feature set is indeed 50.

In contrast, the performance of the algorithms trained with needle leaf species does not exhibit a clear trend with respect to the number of features. The F1 score remains relatively stable across the different feature sets, making it challenging to draw conclusions about the optimal number of features for this classification task.

The variation in F1 score with the number of features is a crucial aspect of evaluating the performance of the algorithms. The results presented in Fig. 8 provide valuable insights into the behaviour of the algorithms under different conditions, allowing for a more informed evaluation of their performance.

Although there is a slight increase in the F1 score for the needle-leaf model, this increase is not substantial. In contrast, Fig. 8 demonstrates that the broad-leaf model performs optimally with 50 features. Considering both of these observations, we have decided to select 50 features for our analysis.

The results for the broad leaf species classification presented above were obtained after the algorithms were tuned to optimize their performance.

By selecting the top 50 independent components, we minimized redundancy and focused on the most significant features for tree species classification. The feature contribution analysis revealed that, for the evergreen model, 51 % of the features were derived from EnMap data while 49 % originated from Sentinel-2 data. Similarly, for the deciduous model, the split was 52 % EnMap and 48 % Sentinel-2. This approach highlights the balanced and complementary contribution of both datasets in capturing essential spectral information.

The detailed analysis of the selected features indicated that certain Sentinel-2 bands and indices played a crucial role. By analysing the top 50 components, we observed that along with bands B11 and B12, and indices like NDVI, AFVI, CCCI, EVI, NDMI(M), and SAVI, were consistently among the highest contributors across multiple seasons. For

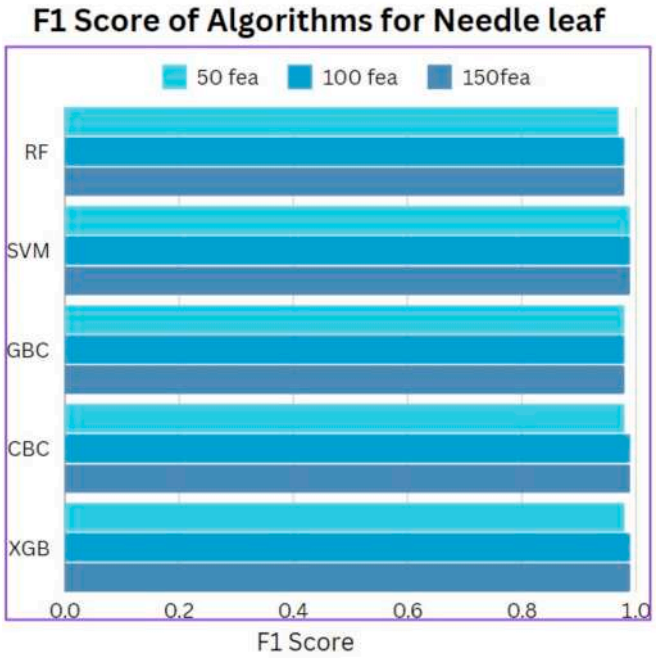

Fig. 7. Algorithms performance for Needle leaf.

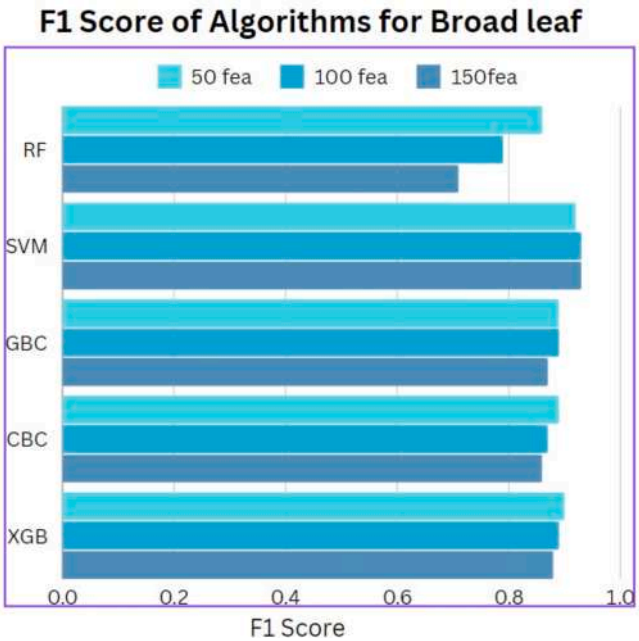

Fig. 8. Algorithms performance for broad leaf.

instance, the NDVI calculated from each of the five Sentinel-2 images, representing different seasons, was identified as a top contributor to the components as were the mentioned bands and indices..For the EnMap data, critical spectral bands for both evergreen and deciduous models were found within the ranges of 418 nm-454 nm (8 bands), 686 nm-756 nm (8 bands), and 879 nm-1128 nm (20 bands). Additionally, specific bands for the deciduous model included 1282 nm-1378 nm (9 bands), 1609 nm-1958 nm (11 bands), and 2384 nm-2400 nm (3 bands). For the evergreen model, the significant bands were 1294 nm-1342 nm (5 bands), 1738 nm-1977 nm (8 bands), and 2337 nm-2400 nm (3 bands). These findings highlight the importance of various spectral regions and indices in the classification of different tree species.

Significance of tuning Machine learning Models:

Tuning is essential to optimize the performance of machine learning algorithms and enhance their ability to accurately classify specific classes within a dataset. Fig. 9 illustrates the variation in F1 scores for all broad leaf species following the tuning process. Notably, the F1 scores for classes 1 (from 0 to 0.67), 2 (from 0.22 to 0.55), and 7(from 0.32 to 0.54) exhibited significant increases, highlighting the effectiveness of the tuning in enhancing the classification performance for these specific classes. In contrast, classes 3, 4, 4a, and 4b showed a slight increase in F1 scores, indicating a moderate improvement. Finally, the F1 score for class 6 remained consistent throughout the tuning process.

The analysis presented in Fig. 9 focuses on the F1 scores of broad leaf species after tuning, specifically for the Support Vector Machine (SVM) algorithm using 50 features. This detailed examination provides valuable insights into the impact of tuning on the classification performance of individual classes within the broad leaf species dataset.

To address the query on combining Sentinel-2 and EnMap data, we conducted an analysis comparing the classification performance using four different datasets: Sentinel-2 alone, EnMap alone, a combination of both and EnMap indices alone. The evaluation focused on the F1 scores of various broadleaf species classes. Our results, illustrated in the accompanying Fig. 10, demonstrate that the combined dataset consistently outperforms the individual datasets for the majority of classes. Specifically, for classes 1, 2, 3, 6, and 7, the combined data achieves better F1 scores compared to using Sentinel-2 or EnMap data alone. For class 4, all three datasets performed equally well. Interestingly, for classes 4a and 4b, Sentinel-2 data showed better performance. However, considering the overall performance across all classes, it is evident that the combined dataset yields the most robust results. Thus, we opted to use the combined Sentinel-2 and EnMap data to leverage the strengths of both datasets, ensuring a more accurate and reliable classification.

Regarding the use of indices derived from hyperspectral data, we conducted an additional analysis where we calculated 186 indices from

Fig. 9. SVM F1 score for broad leaf tree species Before and After tuning.

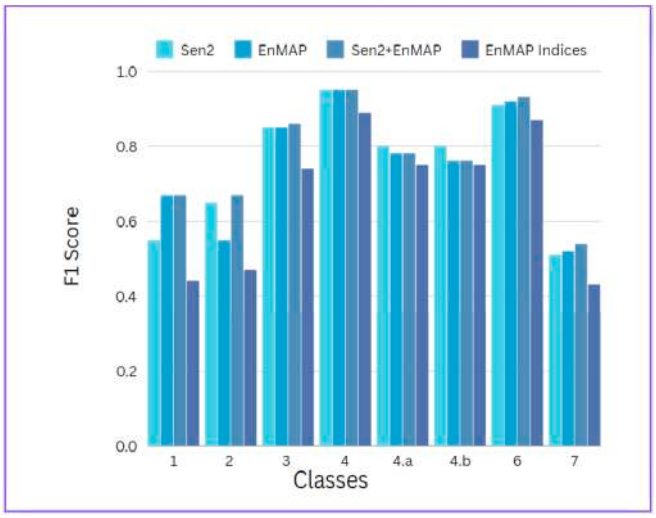

Fig. 10. F1score comparison between multiple datasets.

the EnMap hyperspectral data. These indices relate to the structural, biophysical, biochemical properties, and moisture or water content of the trees. The indices were sourced from the Index Database website (https://www.indexdatabase.de/db/ias.php), selecting EnMap as the sensor and vegetation as the application. Out of 256 available indices, only 186 were feasible for calculation. We applied the same data reduction process using Independent Component Analysis (ICA) and employed SVM for the classification of tree species classes based solely on these indices, using the top 50 components derived from ICA. The results of this analysis are depicted in the image above (Fig. 10). Due to the extensive number of indices used, we provide a direct reference to the source for detailed formulas and descriptions.

The classification results depicted in Fig. 11 showcase how various algorithms classified different tree species using a set of 50 features. As previously indicated, the decision to utilize 50 features was based on its superior performance in terms of overall accuracy. Therefore, the classified map presented in the document is specifically derived from the 50-feature set to ensure optimal accuracy and consistency in the classification results. The inclusion of this figure provides a visual representation of the classification outcomes achieved by the algorithms, emphasizing the significance of the selected feature set in enhancing the accuracy of the classification process.

After combining the outputs from Support Vector Machine (SVM), Random Forest (RF), and CatBoost (CBC) algorithms, a comprehensive analysis was conducted to determine the final classification results. Notably, SVM exhibited the best overall accuracy for the needle species, particularly excelling in classes 5, 8, and 9, corresponding to larch, pine, and spruce. These classes were directly sourced from the SVM output due to its superior performance in accurately classifying these species.

For classes 1, 2, 4.a, and 4.b,7 representing Black locus, European beech, Northern red oak, and pendunculate and Sessile oaks, and Aspen SVM demonstrated good class accuracy. Hence, these classes were also selected from the SVM output. To ensure accuracy, class 3 (birch) underwent a cross-check process, where it had to be classified as birch by both CatBoost and SVM algorithms to be categorized as birch.

Lastly, classes 4 and 6, corresponding to oak and alder, were sourced from the RF algorithm. The combination of these selections from SVM and RF, based on their individual strengths in classifying specific species, culminated in the final classification results. Fig. 12 visually represents the outcomes of this amalgamation, showcasing the effectiveness of leveraging multiple algorithms to enhance the accuracy and precision of the classification process.

Using Fig. 12 as a base map (Heym et al., 2020), multiple biodiversity indices (Pipinis and Radoglou, 2024) were generated to assess the

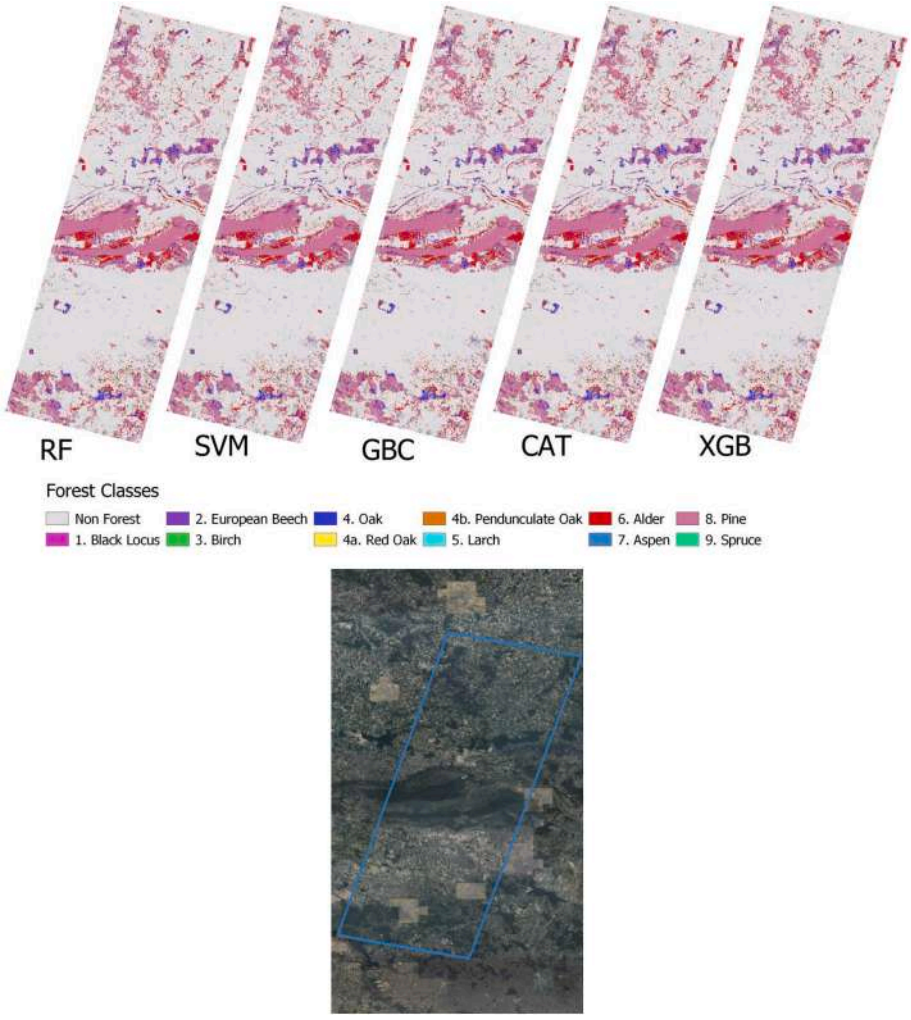

Fig. 11. The classification results for all the algorithms followed by the orthophotomap.

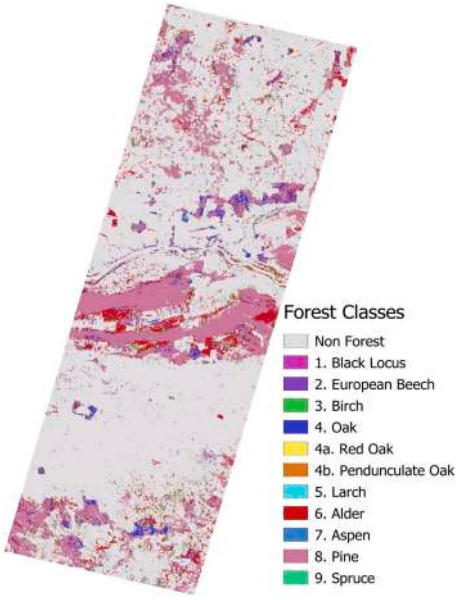

Fig. 12. Combination of 3 classified images.

forest's biodiversity (Ette et al.,2023, Purvis and Hector, 2000, Laurila-Pant et al., 2015). As explained previously these indices include the Shannon-Wiener index, Simpson's diversity index, species richness, and a custom biodiversity index, all calculated at a resolution of 90 m, where each pixel represents a 9-pixel area of the EnMAP image. Each index offers a distinct perspective on the distribution and abundance of species within the habitat.

Fig. 13 displays the results of these indices, providing valuable insights into the biodiversity patterns and species composition within the forest area.

4.2. Validation

To evaluate the performance of the classification model, a comprehensive validation process was undertaken. Initially, the Polish Forest Data Bank (FDB) was utilized to select stands with multiple tree species classes and varying percentages. The classification results were then compared to the actual species composition and percentages within each stand. This process was repeated for five stands, selected based on non-homogeneity, covering areas of 61,446; 16,601; 59,563; 162,321; and 110,646 square meters, respectively. The results are presented in Table 5 below.

Example Interpretation: For Stand ID 17–12-1–07-379-f, the actual data indicates that there are two classes: Forest Class 6 with an occupancy of 90 % and an Other class with an occupancy of 10 %.

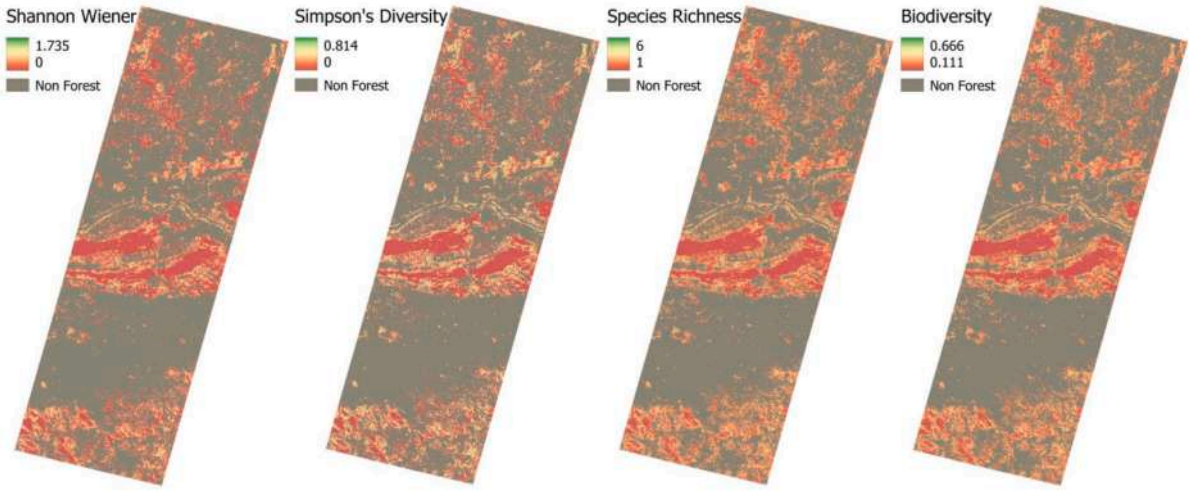

Fig. 13. Biodiversity Indices.

Table 5
Comparison of national forest inventory data (FDB) with classified image.

No		1	2	3	4	5
Stand ID		17-12-1-07-379-f	16-03-3-08-1-f	16-03-3-11-117-c	17-13-2-10-82-b	06-14-1-07-115-d
Class from FDB	**Class**	[6, other]	[3,4,6]	[4,8]	[3,4]	[3,4, other]
	%	[90,10]	[10,40,50]	[40,60]	[40,60]	[30,60,10]
SVM	**Class**	[4,6]	[4]	[3,4,6,8]	[4]	[3,4,6,8]
	%	[22.5,77.5]	[100]	[4.4,40,10,44.6]	[100]	[7.4,79,12,0.8]
RM	**Class**	[4,6]	[4,6]	[3,4,5,6,8]	[4,6]	[3,4,6,8]
	%	[4.2,95.7]	[6.3,93.7]	[1.4,19,1.4,34,43]	[5.5,95.5]	[1.6,74,23,0.8]
CAT	**Class**	[4,6]	[3,4,6]	[3,4,6,8,9]	[3,4,4b,6]	[3,4,6,8]
	%	[4.2,95.7]	[31,6.2,62.5]	[1.4,29.8,23.8,40,4.4]	[51,3,1,43]	[6.6,80,12,0.8]
GBC	**Class**	[2,4,4a,6,7]	[4a,7]	[2,4,5,6,8,9]	[4,4a,4b,7]	[3,4,6,8]
	%	[12,2,2,76,8]	[37.5,62.5]	[8.9,28,2.9,17.9,31.3,10.4]	[3,64,1,30]	[4.1,79,15,0.8]
XGB	**Class**	[2,4,6]	[2,3,4,6]	[3,4,5,6,8,9]	[2,3,4,4b,6]	[3,4,6,8]
	%	[1.4,12.6,85]	[6.25,6.25,50,37]	[1.4,37.3,2.9,16.4,35.8,5.9]	[3,0.5,87,1,7]	[5.7,78,14,0.8]

- The SVM model classified this stand as having 22.5 % coverage of Class 4 and 77.5 % coverage of Class 6, indicating a significant misclassification of Class 4.
- The RM & CAT models showed 4.2 % coverage of Class 4 and 95.7 % coverage of Class 6, which is a closer approximation to the actual occupancy.

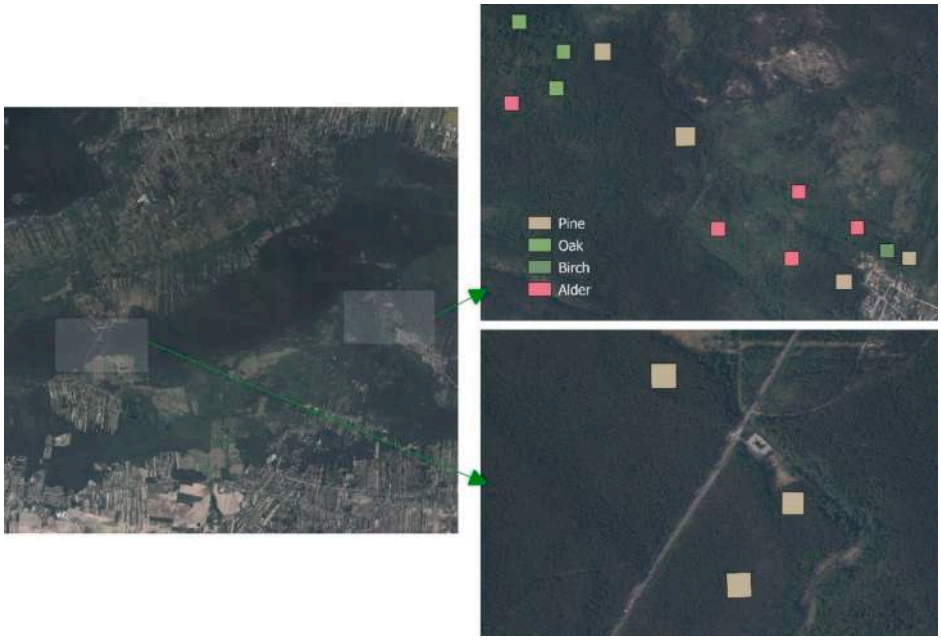

Fig. 14. Field visit sites with google images in the background.

- The GBC model reported 12 % for Class 2, 2 % for Class 4, 2 % for Class 4a, 76 % for Class 6, and 8 % for Class 7, suggesting a more diverse classification but still deviating from the actual data.
- The XGB model indicated 1.4 % for Class 2, 12.6 % for Class 4, and 85 % for Class 6, which aligns more closely with the actual occupancy of Class 6.

The validation process was further expanded to include the Kampinos National Park, which was not represented in the FDB reference data. To achieve this, a grid of 60x60 meter squares was created, covering the park's area. Four species, pine, oak, alder, and birch were selected for this validation exercise, as they are the dominant species within the national park.

Field visit was conducted on April 13, 2024, to two sites as illustrated in the Fig. 14b within the Kampinos National Park. The results confirmed that the classification model accurately identified the species within each square, with all species correctly classified.

However, due to the vast area of the national park and the restrictions on accessing certain areas, not all squares could be validated through field visits. To address this, the validation process was extended to include an online search of the Polish National Data website (National Forest Inventory: https://www.bdl.lasy.gov.pl/portal/mapy-en). This allowed for the identification of the stand corresponding to each square and the verification of the presence or absence of the species within that stand. The percentage of each species within the stand was also checked to ensure accuracy.

The validation process employed a combination of field visits and online research to ensure the reliability of the classification results. This comprehensive approach allowed for a thorough evaluation of the model's performance and provided valuable insights into its strengths and limitations.

4.3. Discussion

The classification results obtained using the five algorithms reveal an interesting phenomenon. Upon examining the classified images, it becomes apparent that a strip-like area in the middle of the image is classified as a different class, whereas in reality, it belongs to the same species. This effect is observed in the classified images of all algorithms except for the SVM classifier. Fig. 15 illustrates this phenomenon, with the left image showing the classified image from SVM and the right image showing the classified image from Random Forest. Both images correspond to the Kampinos National Park.

The Support Vector Classifier (SVC) with tuned parameters (C = 10, gamma='scale', kernel='rbf') demonstrated a unique ability to avoid the strip-like area misclassification observed in the classified images, unlike the other algorithms used in the study. The utilization of these specific parameters, such as a regularization parameter (C = 10), a gamma value ('scale'), and a radial basis function kernel ('rbf'), played a crucial role in enhancing the algorithm's performance. The regularization parameter helped prevent overfitting, the gamma value influenced the kernel coefficient, and the radial basis function kernel allowed for effective handling of non-linear relationships in the data (Mercier and Lennon, 2003, Camps-Valls and Bruzzone, 2005). These tuned parameters, combined with the inherent characteristics of the SVM algorithm, likely contributed to its success in accurately classifying the strip-like area and achieving superior performance compared to the other algorithms.

The Support Vector Machine (SVM) algorithm is designed to handle imbalanced datasets, including those with varying sample sizes. SVMs are known for their ability to focus on the most informative samples, which can help mitigate the effects of class imbalance. This is achieved through the use of a kernel function and the optimization of a margin around the decision boundary (Dalponte et al., 2015).

To study this effect more closely, we selected a square area within the strip and a square area just beside it, both covering 9 pixels each and representing the same tree species, pine. Despite having the same species, the spectral signatures of these areas differ, as evident from Fig. 16. This discrepancy suggests that the strip area has a higher reflectance, whereas the area beside it has a lower reflectance. This finding highlights the importance of considering the spectral signatures of different areas within the same species (Ballanti et al., 2016, Mercier and Lennon, 2003).

Notably, the SVM classifier appears to have avoided this effect, which is one of the reasons why we chose its output for the needle leaf species, in addition to its higher F1 score.

In our study, we thoroughly evaluated the performance of the Support Vector Machine (SVM) algorithm, both with and without parameter tuning. Our results indicate that the SVM consistently outperforms other machine learning algorithms, mainly in the anomaly area. Regardless of

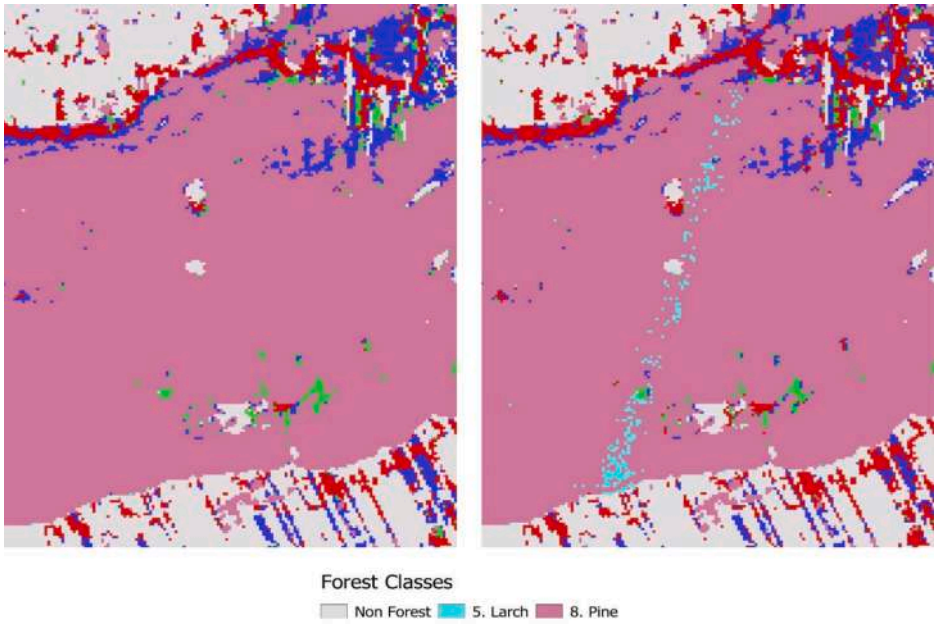

Fig. 15. Comparison between SVM and RM classified images.

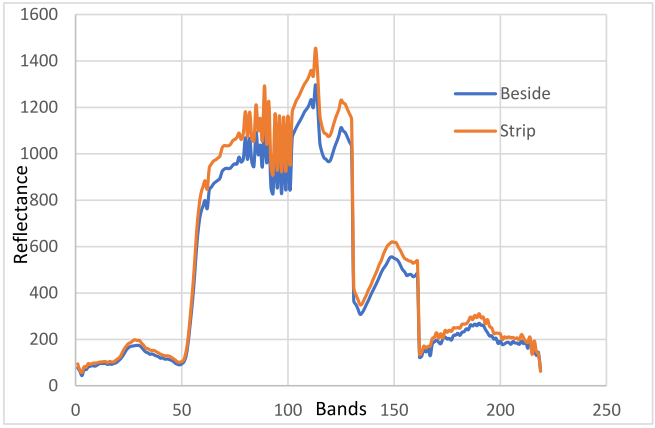

Fig. 16. Spectral signature comparison for pine species.

the tuning, the SVM demonstrated robust classification capabilities, effectively distinguishing between forest classes despite the presence of the anomaly shown strip-like area. This resilience suggests that the SVM is capable of overcoming the challenges posed by such anomalies, enabling it to perform classification more accurately than other machine learning methods employed in this study. Therefore, we believe that the superior performance of the SVM is not merely apparent but is indicative of its effectiveness in handling complex classification tasks in the presence of anomalies.

The combination of outputs from multiple algorithms, such as SVM, RF, and CBC, underscores the importance of leveraging diverse approaches to enhance classification accuracy (Camps-Valls and Bruzzone, 2005). The selection of algorithms based on their individual strengths in classifying specific species contributes to the overall accuracy and precision of the classification process. Furthermore, the visual representation of the classification outcomes provides valuable insights into the spatial distribution of tree species within the study area (Fig. 10, 11).

The final output, derived from the combination of multiple algorithm outputs as detailed above, is a culmination of diverse classification approaches. The accuracy of the overall image classification is intricately tied to the F1 score of each class and their weighted average. Through a meticulous evaluation process, the overall F1 score is determined to be 0.91, reflecting the robustness and effectiveness of the combined classification methodology in accurately delineating and categorizing the various classes (Ballanti et al., 2016) (Mercier and Lennon, 2003).

The biodiversity indices derived from the species map give us important insights into the forest's richness and diversity. The Shannon-Wiener index helps us understand the complexity of the ecosystem by looking at both the number of species and how evenly they are distributed. In our study, this index ranged from 0 to 1.735, indicating varying levels of diversity across the park. Higher values mean more species and a more balanced distribution, suggesting healthier ecosystems (Pipinis & Radoglou, 2024; Ette et al., 2023).

Simpson's diversity index, which ranges from 0 to 0.814 in our findings, shows us the balance between common and rare species. Lower values suggest a diverse ecosystem with many species, while higher values point to the dominance of a few species, which can signal ecological issues (Ette et al., 2023; Purvis & Hector, 2000).

Species richness, the simple count of different species, ranged from 1 to 6 in our study area. This measure highlights the variety of habitats and the range of species they support, underscoring the park's importance for biodiversity (Heym et al., 2020; Laurila-Pant et al., 2015).

Our custom biodiversity index, which ranges from 0.111 to 0.666, looks at the proportion of different species within small areas. This index helps identify biodiversity hotspots and assess the effectiveness of conservation efforts.

In summary, these biodiversity indices reveal the complex ecological dynamics in Kampinos National Park and emphasize the need for ongoing conservation to maintain the park's rich biodiversity. They provide a clear picture of the park's health, guiding future conservation and management strategies.

5. Conclusion

This study presents a comprehensive approach to tree species classification and biodiversity assessment using a combination of machine learning algorithms and remote sensing data. By leveraging the strengths of Support Vector Machine (SVM), Random Forest (RF), and CatBoost (CBC) algorithms, we achieved an overall F1 score of 0.91 in accurately classifying needle leaf and broad leaf tree species.

The results demonstrate the effectiveness of using 50 features for optimal classification performance, as increasing the number of features did not significantly improve the F1 scores. The tuning process proved valuable in enhancing the classification accuracy for specific broad leaf species classes, highlighting the importance of algorithm optimization.

An interesting finding emerged from the classified images, where a strip-like area was consistently misclassified by all algorithms except SVM. Further investigation revealed differences in spectral signatures within the same species, emphasizing the need to consider spatial variations in reflectance properties.

The combination of multiple algorithm outputs, with each contributing its strengths in classifying specific species, underscores the importance of leveraging diverse approaches to improve overall accuracy. The visual representation of the classification results provides valuable insights into the spatial distribution of tree species and aids in understanding the forest ecosystem's structure.

The biodiversity indices calculated from the classified maps, including Shannon-Wiener index, Simpson's diversity index, species richness, and a custom biodiversity index, offer a comprehensive assessment of the forest's biodiversity patterns. These indices can inform forest management decisions and support conservation efforts.

The validation process, which involved field visits and online data verification, ensures the reliability of the classification results. This multi-faceted approach, combining remote sensing data, machine learning algorithms, and field validation, demonstrates the potential of this methodology in accurately mapping tree species and assessing biodiversity at a landscape scale.

Despite the promising results, this study has some limitations that warrant consideration. Future research should focus on increasing coverage of EnMAP data which restricts the scope of species mapping, indicating the need for additional data acquisition, possibly exploiting the availability of other hyperspectral sensors (PRISMA), to expand the study area and increase the number of species considered. Future research efforts could explore the integration of neural networks and other advanced techniques to further enhance classification accuracy and address the challenges associated with mapping complex forest ecosystems.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used perplexity in order to check for grammatical mistakes and clarity of sentences. After using this tool, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

CRediT authorship contribution statement

Rajesh Vanguri: Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Giovanni Laneve:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Agata Hościło:** Writing – review & editing, Validation, Supervision, Methodology.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The work of Agata Hościło was supported by the National Science Centre in Poland [grant number 2021/41/B/ST10/04113].

References

Andrzejewska, A., Anna, K., Danuta, P.M., Dawid, M., Grzegorz, O., Tomasz, H., Łukasz, Ł., Dawid, M., Grzegorz, O., Adam, O., Magdalena, S., Mirosław, S., Maciej, S., Roman, K., Andrzej, M., Ewa, K., Beata, Ś.G., Ewa, M., 2018. Kampinoski Park Narodowy. Oficyna Wydawnicza FOREST 2018.
Ballanti, L., Blesius, L., Hines, E., Kruse, B., 2016. Tree species classification using hyperspectral imagery: a comparison of two classifiers. Remote Sens. 8, 445.
Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.
Brockerhoff, E.G., Barbaro, L., Castagneyrol, B., Forrester, D.I., Gardiner, B., González-Olabarria, J.R., Jactel, H., 2017. Forest biodiversity, ecosystem functioning and the provision of ecosystem services. Biodivers. Conserv. 26 (13), 3005–3035.
Camps-Valls, G., Bruzzone, L., 2005. Kernel-based methods for hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. 43 (6), 1351–1362.
Chen, T., Guestrin, C., 2016. XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794.
Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20 (3), 273–297.
Dalponte, M., Bruzzone, L., Gianelle, D., 2012. Tree species classification in the Southern Alps based on the fusion of very high geometrical resolution multispectral/hyperspectral images and LiDAR data. Remote Sens. Environ. 123, 258–270.
Dalponte, M., Ene, L.T., Marconcini, M., Gobakken, T., Næsset, E., 2015. Semi-supervised SVM for individual tree crown species classification. ISPRS J. Photogramm. Remote Sens. 110, 77–87.
Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Bargellini, P., 2012. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. Remote Sens. Environ. 120, 25–36.
Ette, J.-S., Sallmannshofer, M., Geburek, T., 2023. Assessing forest biodiversity: a novel index to consider ecosystem, species, and genetic diversity. Forests 14 (4), 709.
Fahey, T.J., Woodbury, P.B., Battles, J.J., Goodale, C.L., Hamburg, S.P., Ollinger, S.V., Woodall, C.W., 2018. Forest carbon storage: ecology, management, and policy. Front. Ecol. Environ. 8 (5), 245–252.
Fang, F., McNeil, B.E., Warner, T.A., Maxwell, A.E., Dahle, G.A., Eutsler, E., Li, J., 2020. Discriminating tree species at different taxonomic levels using multi-temporal WorldView-3 imagery in Washington DC, USA. Remote Sens. Environ. 246, 111811.
Fassnacht, F.E., Latifi, H., Stereńczak, K., Modzelewska, A., Lefsky, M., Waser, L.T., Ghosh, A., 2016. Review of studies on tree species classification from remotely sensed data. Remote Sens. Environ. 186, 64–87.
Féret, J.B., Asner, G.P., 2014. Mapping tropical forest canopy diversity using high-fidelity imaging spectroscopy. Ecol. Appl. 24 (6), 1289–1296.
Foley, J.A., DeFries, R., Asner, G.P., Barford, C., Bonan, G., Carpenter, S.R., Snyder, P.K., 2005. Global consequences of land use. Science 309 (5734), 570–574.
Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 29 (5), 1189–1232.
Gao, B.C., 1996. NDWI - A normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sens. Environ. 58 (3), 257–266.
Gatti, A., Bertolini, A., 2018. Sentinel-2 products specification document. Thales Alenia Sp 1–487.
Gillespie, T.W., Foody, G.M., Rocchini, D., Giorgi, A.P., Saatchi, S., 2008. Measuring and modelling biodiversity from space. Prog. Phys. Geogr. 32 (2), 203–221.
Gitelson, A.A., Kaufman, Y.J., Merzlyak, M.N., 1996. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. Remote Sens. Environ. 58 (3), 289–298.
Gitelson, A.A., Gritz, Y., Merzlyak, M.N., 2003. Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. J. Plant Physiol. 160 (3), 271–282.
Gitelson, A.A., Vina, A., Ciganda, V., Rundquist, D.C., Arkebauer, T.J., 2005. Remote estimation of canopy chlorophyll content in crops. Geophys. Res. Lett. 32 (8).
Grabska, E., Hostert, P., Leitão, P.J., Radeloff, V.C., 2019. Commonalities and differences of landsat-8 over landsat-7 derived phenological metrics in the dry tropics of West Africa. Remote Sens. (Basel) 11 (10), 1204.
Grabska-Szwagrzyk, E.M., Tiede, D., Sudmanns, M., Kozak, J., 2024. Map of forest tree species for Poland based on Sentinel-2 data. Earth Syst. Sci. Data Discuss. 1–21.
Guanter, L., Kaufmann, H., Segl, K., Foerster, S., Rogass, C., Chabrillat, S., Chlebek, C., 2015. The EnMAP spaceborne imaging spectroscopy mission for earth observation. Int. J. Remote Sens. 36 (18), 4661–4685.
Guo, Y., Mokany, K., Ong, C., Moghadam, P., Ferrier, S., Levick, S.R., 2023. Plant species richness prediction from DESIS hyperspectral data: A comparison study on feature extraction procedures and regression models. ISPRS J. Photogramm. Remote Sens. 196, 120–133.
Hemmerling, S.A., Querengesser, M., Jarvis, I., Morley, P.J., Buras, A., Grace, J., Koricheva, J., 2021. Responses of insect herbivores to climate change depend on plant community composition and species traits. Oecologia 196 (3), 671–686.
Hernández-Stefanoni, J.L., Gallardo-Cruz, J.A., Meave, J.A., Rocchini, D., Bello-Pineda, J., López-Martínez, J.O., 2014. Modeling α-and β-diversity in a tropical forest from remotely sensed and spatial data. Int. J. Appl. Earth Obs. Geoinf. 29, 93–102.
Heym, M., Uhl, E., Moshammer, R., Dieler, J., Stimm, K., Pretzsch, H., 2020. Utilising Forest Inventory Data for Biodiversity Assessment. Ecol. Ind. 121, 107196.
Hościło, A., Lewandowska, A., 2019. Mapping Forest Type and Tree Species on a Regional Scale Using Multi-Temporal Sentinel-2 Data. Remote Sens 11, 929.
Huete, A.R., 1988. A soil-adjusted vegetation index (SAVI). Remote Sens. Environ. 25 (3), 295–309.
Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sens. Environ. 83 (1–2), 195–213.
Immitzer, M., Atzberger, C., Koukal, T., 2012. Tree species classification with random forest using very high spatial resolution 8-band WorldView-2 satellite data. Remote Sens. (Basel) 4 (9), 2661–2693.
Itten, K.I., Dell'Endice, F., Hueni, A., Kneubuhler, M., Schlapfer, D., Odermatt, D., Seidel, F., 2008. APEX – The hyperspectral ESA Airborne Prism Experiment. Sensors 8, 6235–6259. https://doi.org/10.3390/s8106235.
Kampinos National Park website: https://kampn.gov.pl/kampinoski-park-narodowy#first.
Karnieli, A., Kaufman, Y.J., Remer, L., Wald, A., 2001. AFRI - Aerosol free vegetation index. Remote Sens. Environ. 77 (1), 10–21.
Langmaier, M., Hochbichler, E., Payrhuber, A., 2023. Importance of tree species composition and forest structure on recreational use – a case study for mountain forests in Upper Styria. Austrian J. For. Sci. 2023 (4), 249–278.
Laurila-Pant, M., Lehikoinen, A., Uusitalo, L., Venesjärvi, R., 2015. How to Value Biodiversity in Environmental Management? Ecol. Ind. 121, 107196.
Liang, J., Crowther, T.W., Picard, N., Wiser, S., Zhou, M., Alberti, G., Reich, P.B., 2016. Positive biodiversity-productivity relationship predominant in global forests. Science 354 (6309), aaf8957.
Lindenmayer, D.B., Margules, C.R., Botkin, D.B., 2000. Indicators of biodiversity for ecologically sustainable forest management. Conserv. Biol. 14 (4), 941–950.
Lindner, M., Maroschek, M., Netherer, S., Kremer, A., Barbati, A., Garcia-Gonzalo, J., Marchetti, M., 2010. Climate change impacts, adaptive capacity, and vulnerability of European forest ecosystems. For. Ecol. Manage. 259 (4), 698–709.
Marconi, S., Weinstein, B.G., Zou, S., Bohlman, S.A., Zare, A., Singh, A., White, E.P., 2022. Continental-scale hyperspectral tree species classification in the United States National Ecological Observatory Network. Remote Sens. Environ. 282, 113264.
Mercier, G., Lennon, M., 2003. Support vector machines for hyperspectral image classification with spectral-based kernels. In: IGARSS 2003. 2003 IEEE International Geoscience and Remote Sensing Symposium. Proceedings (IEEE Cat. No. 03CH37477). IEEE, pp. 288–290.
Millar, C.I., Stephenson, N.L., Stephens, S.L., 2007. Climate change and forests of the future: managing in the face of uncertainty. Ecol. Appl. 17 (8), 2145–2151.
Nagendra, H., 2001. Using remote sensing to assess biodiversity. Int. J. Remote Sens. 22 (12), 2377–2400.
Nagendra, H., Rocchini, D., 2008. High resolution satellite imagery for tropical biodiversity studies: the devil is in the detail. Biodivers. Conserv. 17 (14), 3431–3442.
Noss, R.F., 1999. Assessing and monitoring forest biodiversity: a suggested framework and indicators. For. Ecol. Manage. 115 (2–3), 135–146.
Owadowska, E., Szymura, T., Kunz, M., 2013. Kampinos forest biosphere reserve. In: Kunz, M., Nienartowicz, A. (Eds.), Biosphere Reserves in Poland. Nicolaus Copernicus University in Toruń, pp. 144–164.
Paquette, A., Messier, C., 2011. The effect of biodiversity on tree productivity: from temperate to boreal forests. Glob. Ecol. Biogeogr. 20 (1), 170–180.
Pausas, J.G., Austin, M.P., 2001. Patterns of plant species richness in relation to different environments: an appraisal. J. Veg. Sci. 12 (2), 153–166.
Pausas, J.G., Ouadah, N., Ferran, A., Gimeno, T., Vallejo, R., 2004. Fire severity and seedling establishment in Pinus halepensis woodlands, eastern Iberian Peninsula. Plant Ecol. 169 (2), 205–213.
Pipinis, E., Radoglou, K., 2024. Using biodiversity indices effectively: considerations for forest management. Ecologies 5 (1), 42–51.
Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V., Gulin, A., 2018. CatBoost: unbiased boosting with categorical features. Adv. Neural Inf. Proces. Syst. 31.
Purvis, A., Hector, A., 2000. Getting the measure of biodiversity. Nature 405, 212–219.
Raczko, E., Zagajewski, B., 2017. Comparison of support vector machine, random forest and neural network classifiers for tree species classification on airborne hyperspectral APEX images. Eur. J. Remote Sens. 50 (1), 144–154.
Rocchini, D., Hernández-Stefanoni, J.L., He, K.S., 2015. Advancing species diversity estimate by remotely sensed proxies: a conceptual review. Eco. Inform. 25, 22–28.
Sala, O.E., Chapin, F.S., Armesto, J.J., Berlow, E., Bloomfield, J., Dirzo, R., Wall, D.H., 2000. Global biodiversity scenarios for the year 2100. Science 287 (5459), 1770–1774.
Segl, K., Guanter, L., Gascon, F., Kuester, T., Rogass, C., Kaufmann, H., 2015. Sentinel-2 calibration and validation for the upcoming GEO/CEOS earth observation missions. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 8 (5), 2195–2209.

Somers, B., Asner, G.P., 2014. Tree species mapping in tropical forests using hyperspectral and LiDAR data fusion. In: Hyperspectral Remote Sensing of Tropical and Sub-Tropical Forests. CRC Press, pp. 93–114.
Somers, B., Asner, G.P., Tits, L., Coppin, P., 2011. Endmember variability in spectral mixture analysis: A review. Remote Sens. Environ. 115 (7), 1603–1616.
Stuffler, T., Kaufmann, C., Hofer, S., Förster, K.P., Schreier, G., Mueller, A., Bach, H., 2007. EnMAP-a hyperspectral imaging spectrometer for terrestrial applications from space. Acta Astronaut. 61 (1–6), 115–120.
Sun, Y., Huang, J., Ao, Z., Lao, D., Xin, Q., 2019. Deep learning approaches for the mapping of tree species diversity in a tropical wetland using airborne LiDAR and high-spatial-resolution remote sensing images. Forests 10 (11), 1047.
Turner, W., Spector, S., Gardiner, N., Fladeland, M., Sterling, E., Steininger, M., 2003. Remote sensing for biodiversity science and conservation. Trends Ecol. Evol. 18 (6), 306–314.
Vitousek, P.M., Mooney, H.A., Lubchenco, J., Melillo, J.M., 1997. Human domination of Earth's ecosystems. Science 277 (5325), 494–499.
Wang, B., Liu, J., Li, J., Li, M., 2023. UAV LiDAR and hyperspectral data synergy for tree species classification in the Maoershan Forest Farm region. Remote Sens. (Basel) 15 (4), 1000.
Xi, X., Pang, Y., Guo, Z., Feng, R., 2021. Mapping tree species in subtropical forests using an integrated airborne hyperspectral and LiDAR system. Remote Sens. Environ. 253, 112174.
Zhong, L., Dai, Z., Fang, P., Cao, Y., Wang, L., 2024. A Review: Tree Species Classification Based on Remote Sensing Data and Classic Deep Learning-Based Methods. Forests 15 (5), 852.