Carpathian Journal of Earth and Environmental Sciences February 2024, Vol. 19, No. 1, p. 103 – 113; DOI:10.26471/cjees/2024/019/282

USING MACHINE LEARNING ALGORITHMS FOR NATURAL HABITATS ASSESSMENT

Monica Liliana MARIAN[1], Daniel NASUI[1*], Ciprian Radu GHISE[2], Flavia POPOVICI [2], Cosmin SABO[1], Oana MARE ROSCA[1], Lucia MIHALESCU[1], Bogdan VASILESCU[1], & Zorica VOSGAN[1]

[1]*Technical University of Cluj Napoca, 430122, Baia Mare, Romania*
[2]*Indeco Soft, 430094, Baia Mare, Romania*

Abstract: The potential of AI to process and interpret large volumes of data can provide researchers with a powerful tool to understand and monitor biodiversity on a global scale. In this paper we aimed to identify dominant individual plant species in natural protected habitats. Mapping the dominant species from the targeted natural habitats was followed by testing machine learning algorithm for differentiating similar species using satellite images. In the end we validated the data generated by machine learning algorithms through extensive field observations. Using the Sentinel-2 mission 10m resolution data and comprehensive field mapping we managed to see different phenology variations between diverse types of plant communities. Using the NDVI and NDII vegetation indexes and Random Forest algorithm during the dominant species phenology stages for each consecutive 10-day periods between May 1st and September 10th, revealed distinct responses to climate fluctuations and environmental factors. The natural habitats different signatures are strongly influenced by their ecological and conservation status and are not yet suitable for identification, but could help improve AI's automatic detection for multiannual analysis if a favorable conservation trend is reached. The main achievement of the proposed methodology is the ability to differentiate between different species of deciduous trees, with machine learning training accuracy generally exceeding 95% and classification accuracy surpassing 90%.

Keywords: Biodiversity, Plant communities, Satellite images, Sentinel-2, NDVI, NDII, machine learning, Random Forest

1. INTRODUCTION

The potential of AI to help protect biodiversity is enormous. AI's ability to process and interpret large volumes of data can provide researchers with a powerful tool to understand and monitor biodiversity on a global scale. However, exploiting this potential is not without difficulties. From technical issues such as data quality and resolution to implementation and ethical issues, the use of AI in the study of biodiversity involves a number of complex challenges. We aim to make a significant contribution to understanding these issues by exploring how AI can be used to assess species biodiversity and identifying the main obstacles and difficulties that arise in this process.

Apart from governmental and regional agencies and private economic agents, a number of universities and research institutes have addressed the topic of satellite image interpretation algorithms for the purpose of finding the ecological indicators of vegetation.

Virtanen et al., (2004), succeeded in identifying vegetation types in northwestern Russia by associating data collected in the field with 30m cells of satellite images. Images from different dates were then spectrally standardized by multitemporal relative calibration using first- or second-order linear regression equations for each channel. Regression equations were calculated for the pixel data values taken from the overlapping areas of the images. Calibration rectangles, blocks of pixels whose values were used to calculate the calibration equations, were chosen from deep water, coniferous forests (low near-infrared values), and rocky sites to avoid seasonal bias effects.

Rapinel, et al., (2014), on the Atlantic coast of France confirmed 526 points identified in the field with the images produced by Worldview 2. The points were selected in squares of homogeneous vegetation (15 x 15 m) and recorded with a GPS (horizontal accuracy <2 m). The classification approach was based on automatic processing combining pixel and object-based classification methods. As a first step, a pixel-based supervised classification of the 8 spectral bands using only spectral information was applied. and in a second step, this classification was improved by making an object based on classified objects that integrates contextual, shape and texture criteria.

In conclusion, there have been few studies reporting both the identification of plant species in the field and the mapping of natural vegetation from remote sensing data.

In recent years, there has been an increase in research using algorithms on satellite images with a high spatial resolution, to completely replace traveling in the field, most of them being focused on neural networks (neural networks) and deep learning (deep learning).

Watanabe et al., (2020), suggest that the CNN (convolutional neural networks) technique and the cropped image method would be powerful tools for high-precision image-based vegetation mapping and show immense potential for reducing the efforts and costs needed for vegetation mapping.

Liu, et al., (2021), in northeastern China, used the DeepLabV3 Plus neural network to prove the impact of spatial resolution and spectral bands of remote sensing images on the classification accuracy of marsh vegetation.

The project "Biodiversity estimation using satellite images" coordinated by ETH Zürich University aims to automate the estimation of biodiversity in Switzerland, using machine learning (deep) and satellite images. By combining prompt and high-quality in situ data from Swiss institutions such as BLW, Agroscope, BAFU and WSL, which have been conducting field surveys with dense, large-scale satellite imagery for decades. The first experiments showed that the spectral resolution of the ESA Sentinel-2 satellites, designed for vegetation monitoring, partially allows the recognition of plant species on the ground. This project aims to directly map the distributions of distinct species and biodiversity in Switzerland to help protect the environment and measure the impact of agriculture on biodiversity (ETH Zurich, 2023).

The current trend is to improve the NDVI index, sometimes creating methods such as conditional generative adversarial network (cGAN) to simulate the NIR band from RGB bands of Sentinel-2 multispectral data (Yuan et al., 2023)

In recent months, high-precision models have been increasingly proposed, such as a workflow for using high-resolution satellite time series (Rapid Eye, Planet Scope) to assess spatially consistent vegetation patterns and their relationship with soil characteristics at the field scale (Mohr et al., 2023).

The assessment of ecological status in a territory combines several indices such as Normalized Differential Vegetation Index (NDVI), humidity (UMED), Normalized Differential Accumulation Index and Bare Soil Index (NDBSI) and Land Surface Temperature (LST) resulting in the ecological index of remote sensing (RSEI) to quantitatively assess changes in the ecological environment (Wang, et al., 2023).

Also, an ecological status assessment but transformed into a public database was carried out as well as a multitemporal data set at the protected area level characterizing the spatial patterns and temporal dynamics of ecosystem functioning in the Sierra Nevada Biosphere Reserve (Spain), captured by the Enhanced Spectral Vegetation Index (EVI, using the MOD13Q1.006 product from the MODIS sensor) from 2001 to 2018. The database holds, on an annual scale, a synthetic map of three-attribute Ecosystem Functional Type (EFT) classes ecosystem functions (EFAs): descriptors of annual primary production, seasonality and phenology of carbon gains. It also includes two indices of ecosystem functional diversity derived from the above datasets: EFT richness and EFT rarity (Cazorla, et al., 2023).

In this paper we aimed to identify dominant individual plant species in natural protected habitats, by using Sentinel-2 free images and indices trough AI trainings and extensive field mapping, in order to provide a tool for biodiversity conservation.

The objectives of the paper are:

(a) mapping of the dominant species from the targeted natural habitats;

b) testing the machine learning algorithm for differentiating similar species using satellite images);

(c) using the indices to validate the data generated by machine learning algorithms.

The main challenge of this study was to try to differentiate similar tree species (such as *Quercus robur* and *Quercus petraea*) as recent studies show that it is not possible to differentiate between forest types with different dominant tree species, because some assembled species can generate similar NDVI values or similar NDVI temporal trends (D'Andrea et al, 2022). A recent study found differences between the genera Pinus and Quercus, possibly related to the physiological features of the species and their phenology (Gallardo-Salazar et al, 2023), indicating that similar species cannot be separated using NDVI.

2. STUDY AREA

This study was conducted in an area

surrounding Baia Mare City (Figure 1) with various habitat types and different plant communities. It includes a mountainous area in the north, which diminishes into a plain one in the south (Figure 2-A). The slopes vary from 0 to 450 (Figure 2-B) while the Natura 2000 protected areas cover large areas (Figure 2-C). The land use shows the presence of constructed areas mainly in the south (in and around Baia Mare City) while forests cover most of study area (Figure 2-D).

The study area falls into the type of temperate - continental climate with oceanic influences. According to the Köppen classification, the studied area falls under the climate type Dfb (Beck, et al., 2018), where D corresponds to the continental climate (in which the temperature of the coldest month is less than 0^0C and the temperature of the hottest month exceeds 10^0C), f corresponds to precipitation uniformly distributed throughout the year (without a dry season) and b – the average temperature of the hottest month is below 22^0C and at least 4 months of the year have an average temperature higher than 10^0C (Peel, et al., 2007).

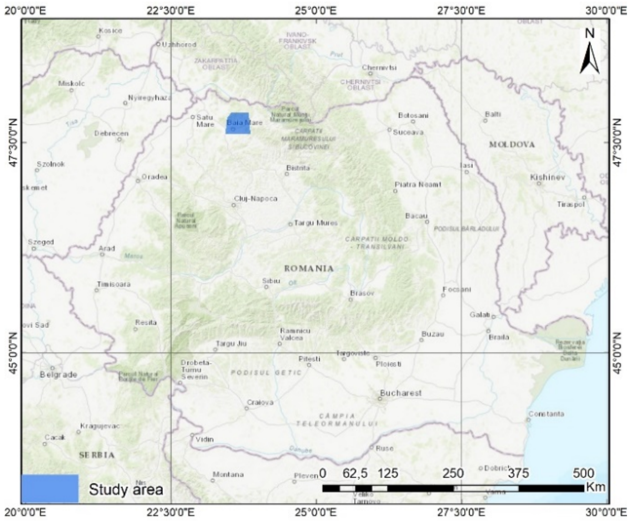

Figure 1. Study area location map

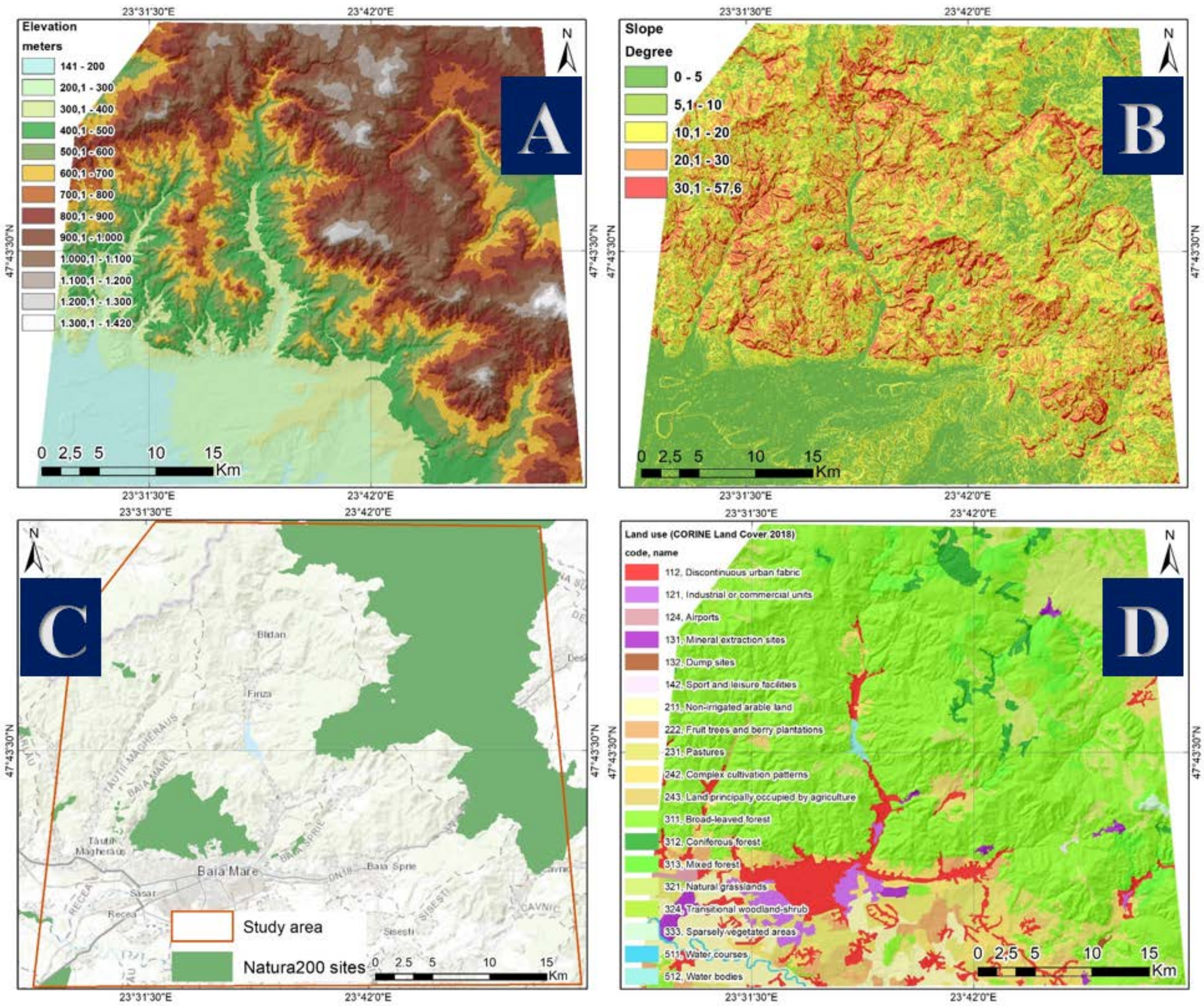

Figure 2. (A- Elevation map, B – Slope map, C Aspect map, D – CORINE Land Cover map)

3. MATERIALS AND METHODS

3.1. Vegetation mapping

The field trips took place during the vegetative periods of 2022 and 2023, for mapping habitats or for validating algorithm data. Even though at first, we started mapping a larger area, we eventually considered only the study area as it had the best satellite coverage (Figure 3).

The main methods for mapping the plant communities involve satellite images identification of sizable and uniform vegetation patches, field trips to the selected areas where we used the transect and quadrant methods. Basically, the transect method was used to cross large areas in search of homogenous vegetation patterns. Once reached we used the quadrant method to map the species distribution inside a 20/20 meters area. The size of the area varied between 10/10m (to match the size of the satellite image pixel) and 50/50m for combined habitats.

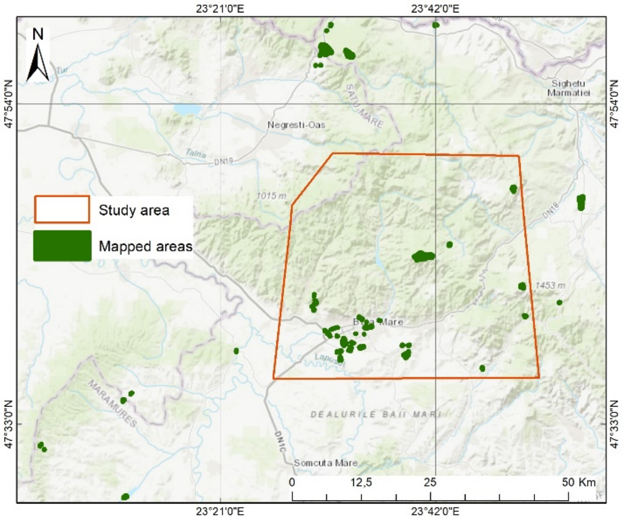

Figure 3. Mapped study area

In this article, the main species we analyzed are dominant species in their habitat (Council Directive 92/43/EEC): *Quercus petraea* (sessile oak) – in 9170 *Galio-Carpinetum* oak-hornbeam forests, *Quercus robur* (oak) – in 9160 Sub-Atlantic and medio-European oak, *Rhynchospora sp.* – in 7150 Depressions on peat substrates of the *Rhynchosporion* and *Vaccinium myrtillus* (European blueberry) – in 4060 Alpine and Boreal heaths.

The best phenology intervals for satellite readings were determined by the best vegetation period, species richness and biomass extent (Figure 4). These intervals do not denote the entire vegetation period, but the intervals where satellite readings would differentiate better similar species.

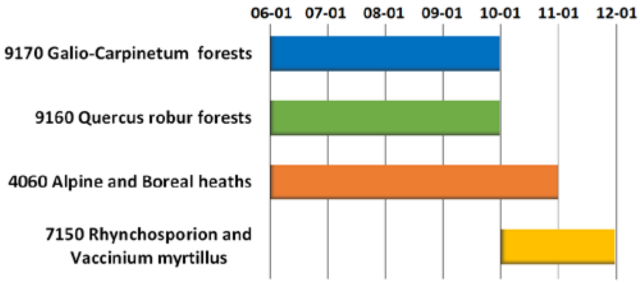

Figure 4. The best phenological intervals for satellite readings in selected habitats

3.2 Satellite indices

Sentinel-2 free images from the 2018-2023 period with a resolution of 10m/pixel were used for land use assessment. To achieve frequent revisits and high mission availability, two identical Sentinel-2 satellites (Sentinel-2A and Sentinel-2B) work together. The satellites are phased 180 degrees from each other on the same orbit. This allows for what would be a 10-day revisit cycle to be completed in 5 days (ESA, 2023). The available free data has far less readable images, as cloud cover in the study area is higher than 40% in the vegetated season.

The following indices have been found useful for Sentinel-2 data processing and land use identification: PSRI (Plant Senescence Reflectance Index), NDVI (Normalized Difference Vegetation Index), NDII (Normalized Difference Infrared Index), EVI (Enhanced Vegetation Index), ARI-1 & ARI-2 (Anthocyanin Reflectance Index), MCARI (Modified Chlorophyll Absorption in Reflectance Index) and NDBI (Normalized Difference Built-Up Index). All of the above were tried and tested on all the mapped areas, constantly trying to differentiate the similar species.

3.3 Data analysis

It is important to decide which set of metrics are effective by method when using machine learning techniques. In a previous experimental study, the performance of seven popular techniques including Logistic Regression, K-nearest Neighbors, Decision Tree, Random Forest, I Bayes, Support Vector Machine and Multilayer Perceptron using software metrics from Promise repository dataset usage were evaluated, experimental results showing that Support Vector Machine achieves a higher performance in class-level datasets and Multilayer Perception produces a better accuracy in method-level datasets among seven techniques above (Phuong Ha, 2019).

In this study, five machine learning algorithms were implemented: K-nearest Neighbors, Random Forest, Naive Bayes, Support Vector Machine and

Multilayer Perceptron. Comparative tests of classification quality were performed on the same data sets with the algorithm change.

4. RESULTS

A series of algorithm trainings were conducted using the biodiversity information to extract the values for the main indices identified as useful in the evaluation of the vegetation species.

Initially, training was conducted on a single 10-day period, and the classification results were modest, as follows:

- For the assessment of land use in the same 10-day period – incredibly superior results, with an accuracy above 90%.
- For land use assessment in another closed period (+- 10 days) acceptable results (60-70% correctly found areas).
- For the evaluation of land use in another more distant 10-day period, or in another year (even in the same period), accuracy was below 30%.

Identification of plant species in the study area, with training and classifications performed on the same data set, at various times, raises problems about the consistency of the classification. Although the learning process is performed in a unitary way, on distinct calendar days, the classification performs significantly better on certain calendar dates than on others. The use of the 10-day period and afterwards, consecutive 10-day periods during the vegetative season (May 1st to September 10th), strongly improved the end results. Here training accuracy generally exceeds 95% and classification accuracy for the same calendar year exceeds 90%.

4.1. Vegetation indices

NDWI and NDBI were found to be rather confusing and lower the quality of the results. Also, multiple species were identified with similar evolutions of the main vegetation indices (NDVI, PSRI, EVI), leading to a high degree of confusion between them.

Experimentally, the MCARI and ARI2 indices were found to be good discriminators between most of these species, which is why they were used in combination with the main indices, ensuring best results. The main problem of MCARI and ARI2 is that they are not normalized indices, there are situations where certain abnormal or exceptional values reduce the quality of the classifications.

Although the accuracy in ideal situations is better by using the two indices, for the elimination of anomalies and better results on the general case, they have been replaced by NDII.

It was found that a best effort / data volume / processing speed ratio is obtained by using NDVI+NDII, which is sufficient for most average classifications (Figure 5). While de NDVI normalizes green leaf scattering in Near Infra-red wavelengths with chlorophyll absorption in red wavelengths (NDVI, 2023) and it is more commonly used for measuring vegetation density and photosynthetic activity, the NDII is sensitive to changes in water content of plant canopies, with values that raise with increasing water content (NDII, 2023) being more sensitive to water content in vegetation and is used to monitor water stress in plants, soil moisture, and drought periods (Sriwongsitanon et al., 2016).

4.2. Machine learning algorithms

The Random Forrest algorithm results are clearly superior to other algorithms. The random forest model has been proved to be the best model in predicting some key plant parameters (Tian & Fu, 2022).

Multilayer Perceptron and Naive Bayes cannot manage missing values for training. K-nearest Neighbors and Support Vector Machine give comparable but significantly inferior results to Random Forrest.

	A	B	C	D	E	F	G	H	I
1	IdZone	IdSpecies	Species_name	PixelX	PixelY	year	month/10-day period	average_NDVI	Average_NDII
2	95590	127	Quercus petraea	14524	415	2021	51	0,40706956	0,10921349
3	95590	127	Quercus petraea	14524	415	2021	52	0,7096637	0,056008413
4	95590	127	Quercus petraea	14524	415	2021	53	0,87618875	0,29962279
5	95590	127	Quercus petraea	14524	415	2021	61	0,88127872	0,3427463
6	95590	127	Quercus petraea	14524	415	2021	62	0,8836792	0,40332124
7	95590	127	Quercus petraea	14524	415	2021	63	0,86911692	0,3440671
8	95590	127	Quercus petraea	14524	415	2021	71	0,69143853	0,261252295
9	95590	127	Quercus petraea	14524	415	2021	72	0,8865564	0,337882975
10	95590	127	Quercus petraea	14524	415	2021	73	0,854501775	0,44997835
11	95590	127	Quercus petraea	14524	415	2021	81	0,88854417	0,3006929
12	95590	127	Quercus petraea	14524	415	2021	82	0,815394228	0,299467758
13	95590	127	Quercus petraea	14524	415	2021	83	0,88656265	0,2750193
14	95590	127	Quercus petraea	14524	415	2021	91	0,848233358	0,312840578

Figure 5. 10-day periods of the vegetative stages, along the index's values

The chosen implementations can manage missing values themselves. An attempt was made to complete / generate the missing values, but the results were inferior to those with "no data".

4.3. Machine learning results

Each training is producing a land use classification for each day for the selected vegetation indexes, producing a 10/10m resolution raster file (Figure 6).

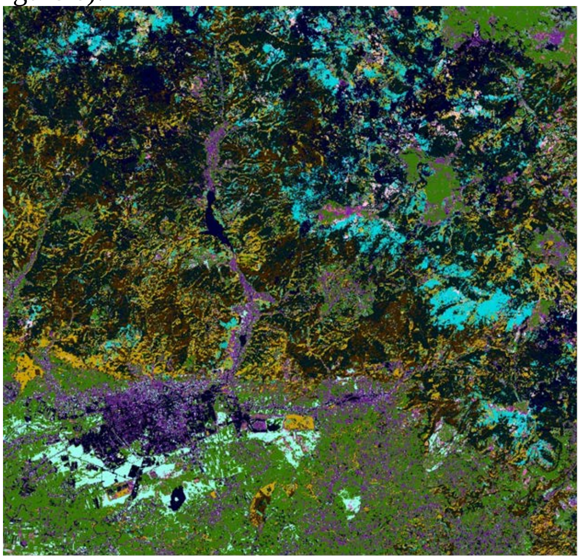

Figure 6. 10/10m resolution land use raster file

The data contained in each pixel can be exported as a csv file for each 10-day period of the vegetative period, along the index's values (Figure 5).

Furthermore, averaging all the pixels of a species for all the 10-day periods can give us a better view of the index's variation throughout the plant's phenology trough NDVI (Figure 7) and plants water content trough NDII (Figure 8).

The combination between NDVI and NDII indices gives a more precise correlation between plant phenology and water content variations (Figure 9)

For better data reading, a 6-degree polynomial trendline was used as the polynomial curvilinear trendline works well for large data sets with oscillating values that have more hills and valleys. The R-squared trendline shows better correlation in the combined NDVI-NDII dataset.

The challenging separation between *Quercus robur* and *Quercus petraea* is slightly visible at the beginning of the phenology period where the precipitation amount is higher, with clearer differentiation in the middle of summer, where drought periods are more often.

4.4. Selected habitats conservation status

The latest assessment of the EEA (2023) shows that habitats and species protected under the EU Habitats Directive have a predominantly unfavorable conservation status at 81% for habitats. The infield evaluation of the selected habitats is as follows:

- 9170 *Galio-Carpinetum* oak-hornbeam forests. The on-site analysis of oak and hornbeam forests, conducted for algorithm training purposes, has revealed an unfavorable and inadequate conservation status of these stands. The inadequate conservation status, marked by deficient compositional and structural parameters, is attributed to pronounced anthropogenic interferences. The anthropogenic impact of the past century, manifested through extensive clear-cutting of plots followed by trunk

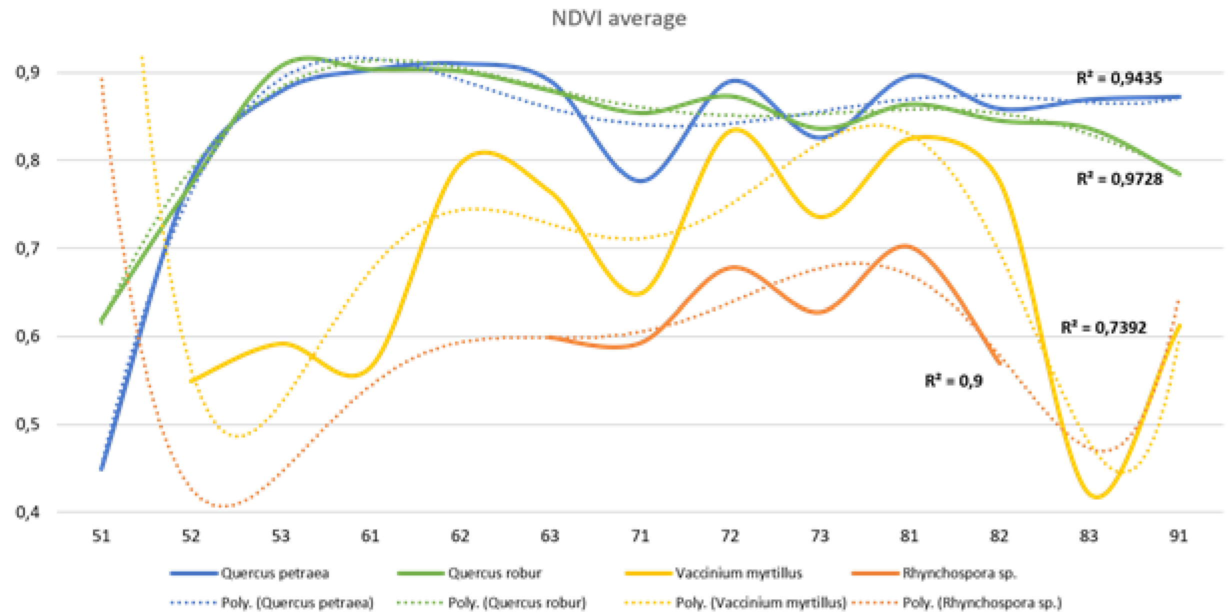

Figure 7. Index variation throughout the plant's phenology trough NDVI

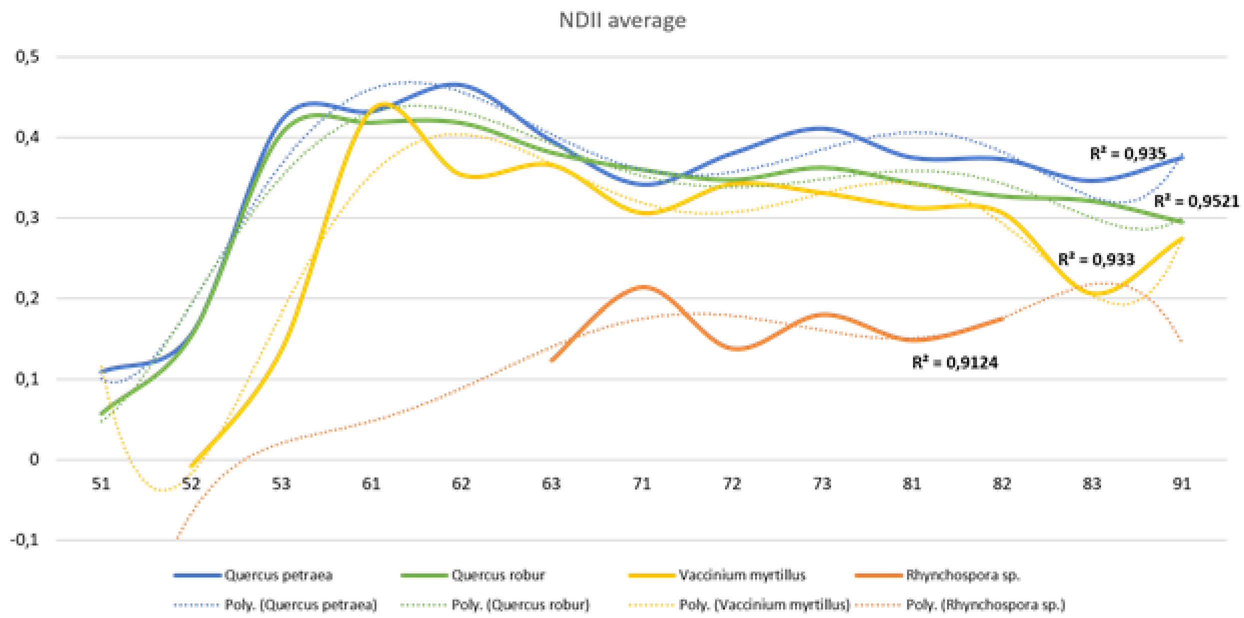

Figure 8. Plant water content variation trough NDII index

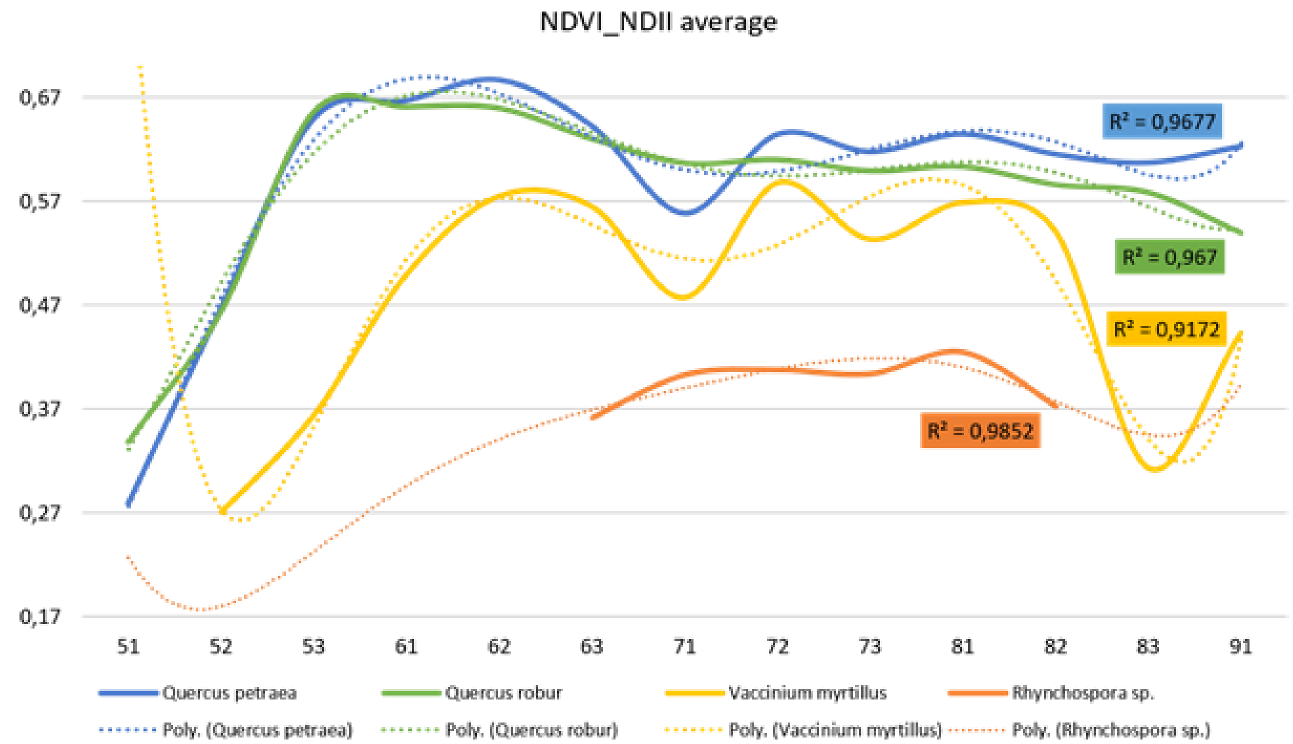

Figure 9. The combination between NDVI and NDII

regenerations, is now reflected in a shift in the ratio of dominant species, favoring hornbeam and disadvantaging oak. In addition to the change in the proportion of dominants, the forests lack a natural structure, with a reduced age distribution of trees; the entire population belongs to the same age group, lacking proper stratification and, consequently, optimal space occupancy. This results in the underutilization of resources within the ecosystem. In conclusion, the stands are characterized by a diminished number of characteristic species, a deficit in spatial architecture with few tree strata, and, consequently, a modification of crown density and distribution, directly impacting the leaf indices analyzed by algorithms.

- 9160 Sub-Atlantic and medio-European oak. Oak forests with a mixture of turkey oak, located on generally flat and low-lying terrain near human communities, exhibit indices of specific and structural diversity characteristic of anthropized ecosystems. Management through intensive exploitation of the forests, including occasional complete removal of the entire tree canopy, has resulted in the regeneration process not restoring structural and functional parameters to their initial levels. The presence of a single tree stratum, instead of 3-4, generates a different pattern of cohesion and crown distribution. It has also been observed that the two analyzed forest habitats, 9170 and 9160, although having different dominant species - sessile oak and pedunculate oak, respectively, with hornbeam as a mixed species for both - tend to homogenize under the consecutive regeneration pressures of strong anthropogenic influence. This homogenization is marked by the simplification of floristic composition and structural parameters.

- 7150 Depressions on peat substrates of the *Rhynchosporion* and *Vaccinium myrtillus*. Oligotrophic marshes have a distinct configuration compared to other types of regional habitats, characterized by substrate features on one hand and a complex of strongly acidophilic species on the other. Structurally, the peat layers accumulate large amounts of water, exhibit low floristic diversity, but photosynthetic indices depend significantly on the amount of bound water. This accumulated water in the abiotic part of the peatland is contingent upon the variation in the hydric regime throughout the year and the seasons. Monitoring oligotrophic marshes indicates a reduction in the amount of bound water, affecting the color and viability of the moss layer.
- 4060 Alpine and Boreal heaths. Blueberries and cranberries shrublands are plant formations that depend on substrate acidity, as well as on a thermal and hydric regime characteristic of high-altitude and subalpine mountain zones. Their conservation status in the analyzed areas is generally favorable, with the component species remaining unaffected. However, their surfaces are shrinking, both due to anthropogenic exploitation and global warming and aridification.

4.5. Data validation

For validation of algorithm predicted results we followed the location of large predicted dominant species with mapping of adjacent species. For example, *Quercus robur* was predicted by algorithm classification (yellow color) in a forest east of Baia Mare City (Figure 10). Alongside the correctly recognized species we also identified *Carpinus betulus* (brown color), which is a characteristic species for the 9160 Sub-Atlantic and medio-European oak habitats. Even though the spatial presence of hornbeam is spread out with small clumps of trees, it is correctly recognized in algorithm classification and its presence is a valuable indicator for habitat's identification. The other identified species are not characteristic of the above-mentioned habitat, showing that its conservation status is inadequate.

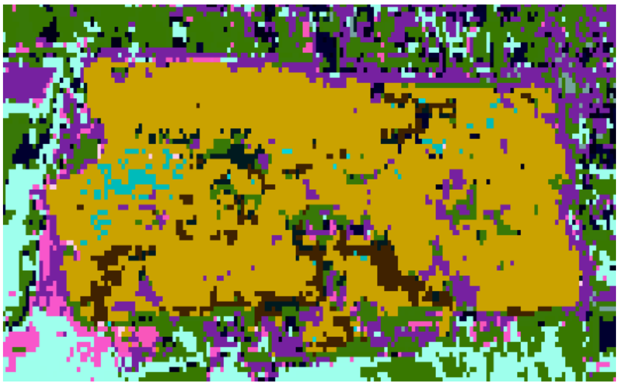

Figure 10. Confirmed presence of *Quercus robur* (yellow color) and *Carpinus betulus* (brown color)

5. DISCUSSION

Identifying plant communities at one stage of their phenological cycle using machine learning on Sentinel-2 10m pixel images is a straightforward task, with quite impressive results (Languille et al., 2017). Nevertheless, the challenge of this study was to show the same plant communities on each reading on satellite images all throughout their phenological year. Weather variations, varied species, phenological mismatch, conservation status and the ubiquitous cloud cover on satellite images, make the algorithm classification results vary vastly over the phenological cycles.

The fact that most of the analyzed plant communities belong to a not so large study area supports the idea that all of them had to adapt to the same environmental variations, hence their response becomes the only variable. The different plants' response to all the above-mentioned fluctuations gives a very distinct signature only when we analyze individual species that are dominant for large areas. Different forest types and tree species differ in their vegetation phenology, offering an opportunity to map and characterize forests based on the seasonal dynamic of vegetation indices and auxiliary data (Silveira et al., 2022, Han et al., 2023).

The results of this case study cannot be used as such in other areas, as different phenology responses will occur in different conditions of climate, geomorphology or habitats conservation status. But the proposed methodology, of using NDVI and NDII on the free Sentinel-2 satellite images could be used to clearly separate similar species of trees in protected habitats.

The NDVI/NDII average lines graphs show

different individual signatures (profiles) of the four plant species, which could prove valuable in identifying patterns in future years, as the type of response to climate variations and water content tends to be the same per species. Although species richness can be an informative measure of diversity, plant communities are dynamic systems under constant change (Aggemyr et al., 2018).

The mixture of species present in these four habitats could supply a habitat identification technique by averaging all the pixels from that specific area, only if a constant or favorable conservation status is reached.

We highlight first the inherent complexity of the process of species identification in the sphere of biodiversity. The difficulties met are much more substantial than were initially foreseen, given several factors of a variable nature. We see that the multiannual evolution of the species is divergent, even in the case of perennial species, depending on the weather and climate peculiarities.

Unfortunately, areas with very high biodiversity values are extremely rare, almost non-existent. The mixture of species in natural habitats further complicates the classification process, because even on a relatively small area, such as a 10 x 10-meter (1 pixel) Sentinel-2 plot, we can find a multitude of species.

One of the main challenges with the data is the small amount of data available, rarely exists imagery available for every 10-day period, often there are few points of intersection between the data, especially if trying to classify on a year other than the training year (Figure 11).

Year	M5d1	M5d2	M5d3	M6d1	M6d2	M6d3	M7d1	M7d2	M7d3	M8D1
2020	x	x		x			x	x	x	
2021		x	x		x	x		x		x
2022	x		x		x				x	x

Where M=month, d=10-day period

Figure 11. Data availability for every 10-day period

Despite these difficulties, a superior quality of the classifications of natural forest and marshy areas, meadows, with a good mapping of vegetation species, both in the lowlands and in the highlands, is found.

Machine learning (ML) algorithms have been increasingly used in biodiversity studies due to their ability to analyze large datasets and identify patterns that may not be easily discernible through traditional methods (Shivaprakash et al., 2022). Within our activities, a series of machine learning algorithm trainings were conducted using biodiversity information to extract the values for the main indices useful in the evaluation of dominant species.

Similarities were found in the value of the indices, in certain periods for certain types of forests.

Ultimately, a relevant barrier to using free ESA data is image resolution. Practitioners have requested the use of higher resolution data. One of the main problems is that Copernicus images only supply a general view of areas of interest but are insufficient for detailed analysis. In this context, activity-relevant Sentinel-2 satellite indices were also evaluated.

6. CONCLUSIONS

The main achievement of the proposed methodology is the ability to differentiate between different species of deciduous tree species. For biodiversity purposes, it could prove useful to be able to differentiate the yearly change in species percent cover.

This study supplies a comprehensive understanding of the challenges and potential solutions in using artificial intelligence algorithms for biodiversity assessment. Despite the difficulties, the way is being outlined for future improvements by continuing to investigate the factors that influence species classification and by adapting AI algorithm training methodologies.

Using the NDVI and NDII vegetation index and Random Forest algorithm during the plant's vegetation season for each consecutive 10-day periods between May 1st and September 10th, revealed dominant species different responses to weather variations, with the machine learning training accuracy generally exceeding 95% and classification accuracy surpassing 90%.

The plant communities' different signatures are not yet suitable for algorithm classification as the natural habitat's conservation status is gradually shifting to the point of unreliable multiannual identification. Maintaining or improving the favorable conservation status of natural habitats could help improve algorithm classification of natural habitats for multiannual analysis.

Acknowledgments

This research was funded by the research project GREENTOP – Agro-Biodiversity, Financing contract no.381/390057/01.10.2021, Project co-financed from the European Regional Development Fund through the Romania's Competitiveness Operational Program 2014-2020.

REFERENCES

Aggemyr, E., Auffret, A.G., Jädergård, L. et al. 2018 Species richness and composition differ in response to landscape and biogeography. Landscape Ecol 33, 2273–2284 (2018). https://doi.org/10.1007/s10980-018-0742-9

Beck, H., Zimmermann, N., McVicar, T. et al., 2018. *Present and future Köppen-Geiger climate classification maps at 1-km resolution*. Sci Data 5, 180214 https://doi.org/10.1038/sdata.2018.214

Cazorla, B. P., Cabello, J., Reyes, A., Guirado, E., Peñas, J., Pérez-Luque, A. J., & Alcaraz-Segura, D., 2023. *A remote-sensing-based dataset to characterize the ecosystem functioning and functional diversity in the Biosphere Reserve of the Sierra Nevada (southeastern Spain),* Earth Syst. Sci. Data, 15, 1871–1887, https://doi.org/10.5194/essd-15-1871-2023.

D'Andrea, Giuseppe, Václav Šimůnek, Maria Castellaneta, Zdeněk V., Stanislav V., Osvaldo Pericolo, Zito, R.,G., & Ripullone, F., 2022. *Mismatch between Annual Tree-Ring Width Growth and NDVI Index in Norway Spruce Stands of Central Europe,* Forests 13, no. 9: 1417. https://doi.org/10.3390/f13091417.

Gallardo-Salazar, J.L., Rosas-Chavoya, M., Pompa-García, M. et al., 2023. *Multi-temporal NDVI analysis using UAV images of tree crowns in a northern Mexican pine-oak forest.* J. For. Res. 34, 1855–1867. https://doi.org/10.1007/s11676-023-01639-w.

Han, J., Tao, V., Xie, Y., Li, H., Yi, H., & Guan, **X.,** 2023. *Validation of the TOA Products of the Baotou Sandy Site with Landsat8/OLI Considering BRDF Correction.* In IEEE Transactions on Geoscience and Remote Sensing, vol. 61, pp. 1-11, Art no. 5401611, doi: 10.1109/TGRS.2023.3260764.

Languille, F., Gaudel, A., Vidal, B., Binet, R., Poulain, V., & Trémas, T., 2017. *Sentinel-2B image quality commissioning phase results and Sentinel2 constellation performances.* Proc. SPIE 10423, Sensors, Systems, and Next-Generation Satellites XXI, 1042311 (29 September 2017); https://doi.org/10.1117/12.2277584.

Liu, M., Fu, B., Xie, S., He, H., Lan, F., Li, H., Lou, P., & Fan, D., 2021. *Comparison of multi-source satellite images for classifying marsh vegetation using DeepLabV3 Plus deep learning algorithm.* In Ecological Indicators 125: 1-21. 107562 https://doi.org/10.1016/j.ecolind.2021.107562.

Mohr, J., Tewes, A., Ahrends, H., & Gaiser, T., 2023. *Assessing the Within-Field Heterogeneity Using Rapid-Eye NDVI Time Series Data.* Agriculture, 13, 1029, 10.3390/agriculture13051029.

Peel, M. C., Finlayson, B. L. & McMahon, T. A., 2007. *Updated world map of the Köppen-Geiger climate classification.* Hydrology and Earth System Sciences 11, 1633–1644.

Phuong Ha, T. M., Hung Tran, D., My Hanh, L. T., & Thanh Binh, N., 2019. *Experimental Study on Software Fault Prediction Using Machine Learning Model.* 11th International Conference on Knowledge and Systems Engineering (KSE), Da Nang, Vietnam, pp. 1-5, doi: 10.1109/KSE.2019.8919429.

Rapinel, S., Clément, B., Magnanon, S., Sellin, V., & Hubert-Moy, L., 2014. *Identification and mapping of natural vegetation on a coastal site using a Worldview-2 satellite image.* Journal of Environmental Management, Elsevier, 144, pp.236-246. 10.1016/j.jenvman.2014.05.027.hal-01017152.

Silveira, E., M. O., Volker C. R., Martínez, G., J., Pastur, Martinuzzi, S., Politi, N., Lizarraga, L., & Rivera, L., O., 2022. *Forest Phenoclusters for Argentina Based on Vegetation Phenology and Climate.* Ecological Applications 32(3): e2526. https://doi.org/10.1002/eap.2526.

Sriwongsitanon, N., Gao, H., Savenije, H. H. G., Maekan, E., Saengsawang, S., & Thianpopirug, S., 2016. *Comparing the Normalized Difference Infrared Index (NDII) with root zone storage in a lumped conceptual model.* Hydrol. Earth Syst. Sci., 20, 3361–3377, https://doi.org/10.5194/hess-20-3361-2016

Shivaprakash, K.N.; Swami, N.; Mysorekar, S.; Arora, R.; Gangadharan, A.; Vohra, K.; Jadeyegowda, M.; Kiesecker, J.M., 2016. *Potential for Artificial Intelligence (AI) and Machine Learning (ML) Applications.* Biodiversity Conservation, Managing Forests, and Related Services in India. Sustainability 2022, 14, 7154. https://doi.org/10.3390/su14127154

Tian, Y.; & Fu, G., 2022. *Quantifying Plant Species α-Diversity Using Normalized Difference Vegetation Index and Climate Data in Alpine Grasslands.* Remote Sens. 14,5007. https://doi.org/10.3390/rs14195007

Virtanen, T., Kari Mikkola & Ari Nikula, 2004. *Satellite image-based vegetation classification of a large area using limited ground reference data: a case study in the Usa Basin, north-east European Russia*. Polar Research, 23 (1), 51-66.

Wang, J; Wang, J; & Xu, JQ., 2023. *Spatio-temporal variation and prediction of ecological quality based on remote sensing ecological index – a case study of Zhanjiang City, China*. FRONTIERS IN ECOLOGY AND EVOLUTION, vol. 11, 10.3389/fevo.2023.1153342

Watanabe, S., Kazuaki Sumi, & Ise, T., 2020. *Identifying the vegetation type in Google Earth images using a convolutional neural network: a case study for Japanese bamboo forests*. BMC Ecology, 20:65, https://doi.org/10.1186/s12898-020-00331-5

Yuan, X.; Tian, J.; & Reinartz, P., 2023. *Learning-Based Near-Infrared Band Simulation with Applications on Large-Scale Landcover Classification*. Sensors 23, 4179, 10.3390/s23094179.

Zhang, X.; & Song, P., 2021. *Estimating Urban Evapotranspiration at 10m Resolution Using Vegetation Information from Sentinel-2: A Case Study for the Beijing Sponge City*. Remote Sens. 13, 2048. https://doi.org/10.3390/rs13112048

Council Directive 92/43/EEC of 21 May 1992 on the

conservation of natural habitats and of wild fauna and flora, O.J. L206, 22.07.92
EEA. European Environment Agency's Habitats and species: latest status and trends. https://www.eea.europa.eu/en/topics/at-a-glance/nature/state-of-nature-in-europe-a-health-check/habitats-and-species-latest-status-and-trends (accessed Oct 25, 2023)
ETH Zurich. Biodiversity Estimation from Satellite Images. https://prs.igp.ethz.ch/research/current_projects/biodiversity_estimation_from_satellite_images.html (accessed Jul 27, 2023)
ESA. European Space Agency's Sentinel-2 mission guide. https://sentinel.esa.int/web/sentinel/missions/sentinel-2 (accessed Aug 17, 2023)
NDII (Normalized difference 819/1600). https://custom-scripts.sentinel-hub.com/custom-scripts/sentinel-2/ndii/ (accessed May 12, 2023)
NDVI.https://custom-scripts.sentinel-hub.com/custom-scripts/hls/ndvi/ (accessed May 12, 2023)

Received at: 19. 11. 2023
Revised at: 24. 12. 2023
Accepted for publication at: 28. 12. 2023
Published online at: 29. 12. 2023