Semi-Supervised European Forest Types Mapping using High-Fidelity Satellite Data

Bohdan Yailymov[1], Hanna Yailymova[1,2], Nataliia Kussul[1,2,3] and Andrii Shelestov[1,2]

[1] *Space Research Institute NAS of Ukraine and SSA of Ukraine, Glushkov ave 40, 4/1, 03187, Kyiv, Ukraine*

[2] *National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Beresteiskyi ave 37, 03056, Kyiv, Ukraine*

[3] *University of Maryland, College Park, MD 20742, US*

Abstract

Accurate and up-to-date forest type maps are crucial for effective monitoring and management of forest ecosystems across Europe. However, the availability of up to date high-resolution forest type maps has been limited. This study introduces an innovative semi-supervised approach for mapping European forest types by harnessing the power of high-resolution Sentinel-1 and Sentinel-2 satellite data from the Copernicus program. The novelty of the approach lies in the integration of various data sources for training dataset creation and the utilization of the Random Forest classifier on the Google Earth Engine cloud computing platform. This innovative combination enables efficient processing and classification of vast amounts of satellite imagery for large-scale forest type mapping. In particular, the LUCAS Copernicus 2018 and 2022 datasets were employed for training and validation, ensuring the robustness of the classification model. The resulting forest type map for 2022 has a fine spatial resolution of 10 meters and distinguishes between three key classes: broadleaved, coniferous, and mixed forests. Accuracy assessment using independent validation data demonstrated the reliability of the proposed approach, yielding an impressive overall accuracy of 93%. Comparative analysis with existing forest products revealed both consistencies and differences, underscoring the dynamic nature of forest ecosystems. The generated map fills a gap in up to date geospatial information on European forest types, empowering informed decision-making in forest management, conservation efforts, and environmental impact assessment. This study demonstrates the potential of synergizing cutting-edge remote sensing, cloud computing, and machine learning technologies to tackle complex environmental challenges at a continental scale, paving the way for future advancements in forest monitoring and management.

Keywords

Forest type classification, Sentinel-1, Sentinel-2, Random Forest, Google Earth Engine, Europe

1. Introduction

This study was carried out within the scope of the Horizon Europe project Satellites for Wilderness Inspection and Forest Threat Tracking (SWIFTT - https://swiftt.eu/), the main tasks of which are developing and improving models for forest health monitoring and damage detection (windthrow damage, tree health, forest volumes and data for fire risk mapping and early fire detection). Having an up-to-date geospatial map of forest types is critical to effectively monitor forest damage such as fires [1], logging, windstorms, disease [2], and other natural and anthropogenic forest events [3]. Geospatial information makes it possible to accurately determine the location of forest ecosystems [4] and determine their types. This provides the possibility of prompt response to potential threats to forests, which allows preserving biodiversity and ecological balance [5]. Geospatial mapping also aids in resource management and conservation planning by providing the ability to track changes in forest cover over time [6]. The use of modern technologies of geospatial analysis allows collecting,

ProfIT AI 2024: 4th International Workshop of IT-professionals on Artificial Intelligence (ProfIT AI 2024), September 25–27, 2024, Cambridge, MA, USA

yailymov@gmail.com (B. Yailymov); anna.yailymova@gmail.com (H. Yailymova); nataliia.kussul@gmail.com (N. Kussul); andrii.shelestov@gmail.com (A. Shelestov)

0000-0002-2635-9842 (B. Yailymov); 0000-0001-6116-8294 (H. Yailymova); 0000-0002-9704-9702 (N. Kussul); 0000-0001-9256-4097 (A. Shelestov)

processing and analyzing data in real time, which is key to timely response to events that may affect the health and sustainability of forests. Therefore, within the SWIFTT project there is a pressing need for the timely and cost-effective creation of contemporary forest maps, particularly those that classify European forests by type.

In the next section, existing products containing forest maps or forest type maps for European countries were analyzed. Their main gap is that there is no single product for 2022 that would include different types of forests (coniferous, deciduous, mixed) and would have a high spatial resolution (10 meters). The task of the SWIFTT project was to obtain a modern map of forest types for the entire territory of Europe using time series of open Sentinel-1,2 satellite data with a maximum spatial resolution of 10 meters.

1.1. Overview of Available Forest Cover Products

At present, there exists a range of geospatial solutions featuring forest mapping layers. The Copernicus program marks a significant stride forward by integrating satellite imagery and machine learning algorithms to generate highly detailed and accurate forest classification maps. A noteworthy example is the ESA WorldCover initiative, which has released comprehensive global land cover products for 2020 [7] and 2021 [8], boasting a spatial resolution of 10 meters. Leveraging Sentinel-1 and Sentinel-2 data, these products are crafted and verified in near-real time. The classification scheme encompasses 11 distinct classes, aligning with the Land Cover Classification System (LCCS) devised by the Food and Agriculture Organization (FAO) of the United Nations (UN). Independent validation conducted by Wageningen University demonstrates an overall accuracy of 74.4% for the WorldCover product [9] across the globe, with continent-specific accuracy rates ranging from 68% to 81%. While this map includes a forest class, it lacks a division into specific forest types, which is important information for monitoring the dynamics of the state of forests and their damage. For our task within the project, the division of the forest into different types of forests is important, because there are different pests for coniferous and deciduous forests and these forests behave differently throughout the year.

The University of Maryland's Global Land Analysis and Discovery (GLAD) laboratory, in collaboration with Global Forest Watch (GFW), offers annually updated global forest loss data, utilizing Landsat time-series imagery with a resolution of 30 meters [10]. This dataset spans from 2000 to 2022 and is segmented into 10×10-degree tiles, each containing seven files. The data, represented in unsigned 8-bit values, has a spatial resolution of 1 arc-second per pixel, or roughly 30 meters per pixel at the equator. It includes a tree cover map expressed as a percentage per grid cell (ranging from 0 to 100), a forest gain map for 2000-2012, and a forest loss map for 2000-2022. The disadvantage lies in its 30-meter spatial resolution, inability to differentiate between coniferous and deciduous forests, and variations in forest mask determination across countries, leading to some inaccuracies compared to other products.

Additionally, Landsat satellite data has been employed to derive annual forest disturbances among 35 European countries, covering the period 1986-2020. This includes maps of disturbance severity (up to 2016) and a forest mask [11], utilizing the LandTrend time-series segmentation approach in the Google Earth Engine environment [5].

The CGLS-LC100 product, part of the Copernicus Global Land Service, offers a global land cover map based on PROBA-V 100 m satellite data for 2015-2019, with various forest types represented [12]. However, its 100-meter spatial resolution remains a limitation, though the product maintains an accuracy of 80%.

Copernicus Land Monitoring Service provides the Forest Type (2018) product [13] for the EEA39 countries, offering 10-meter rasters with a forest classification into three thematic classes (non-forest, broadleaved forest, coniferous forest) with a minimum 90% accuracy for both forest classes, albeit only for 2018. In [14] have created a global forest cover product for 2017 utilizing Sentinel-1 satellite data, including distribution between broadleaved and coniferous forests.

There's been a notable increase in geospatial products formed by integrating existing datasets. For instance, two earth-observation products [15], [16] have been combined with statistical data to produce a new pan-European forest map at a 1 km spatial resolution, aligning with official forest inventory statistics at national and/or regional levels.

Given the SWIFTT project aimed at creating a classification map based on Sentinel satellite data with a 10-meter spatial resolution, products like WorldCover (which has the appropriate spatial resolution and updated to the nearest 2022) and Forest Type 2018 (which has the appropriate spatial resolution and has a distribution between coniferous and deciduous forests) serve as pertinent references for both training dataset creation and validation of the resulting products, considering their suitable spatial resolution and updated information up to 2022.

Within the scope of the project, monitoring of all forests in Europe is planned. Therefore, this paper analyzes various algorithms and cloud platforms for building a forest classification map. Taking into account the amount of satellite data and the speed of the algorithm for large-scale construction of the forest map, the Random Forest algorithm was chosen in free Google Earth Engine cloud platform, since all satellite data are already ready for use in it. The obtained forest maps and area was compared with existing products.

2. Study Area

The forest type classification maps were created for European countries. The information about countries boundary were the Geographic Information System of the Commission [17]. A critical aspect of this study area lies in its varied topography, including mountainous terrains that introduce shadows and elevation variations. These geographical features can complicate the interpretation of satellite imagery, influencing the accuracy of forest classification. Additionally, the presence of snow cover during certain seasons further adds complexity to the satellite data analysis, impacting the differentiation between land cover types. Also, taking into account the area of all countries included in the study area - 6.5 MLN km^2, the question of solving the problem of big data arises. To optimize the work, the countries were divided into groups based on their geo-location, size of the countries, and availability of the training data. The distribution of countries by groups is shown in the Table 1 and in the Figure 1.

Table 1
Country division by groups

Group number	List of countries
1	Spain, Portugal
2	France, Belgium, Luxemburg
3	Switzerland, Netherland, Germany
4	Slovakia, Liechtenstein, Hungary, Czechia, Austria
5	Slovenia, Italy, Croatia, Malta
6	Bosnia and Herzegovina, Albania, Bulgaria, Serbia, Greece, North Macedonia, Romania, Montenegro, Cyprus, Turkey
7	Poland, Ukraine
8	Lithuania, Latvia, Estonia
9	Finland
10	Sweden, Denmark, Norway
11	United Kingdom, Ireland, Iceland

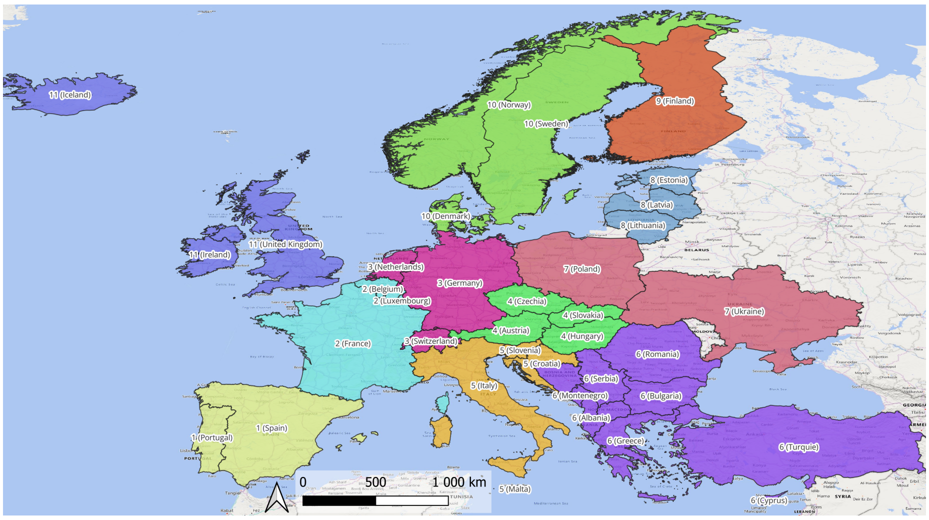

Figure 1: Country division by groups.

3. Satellite Data Used

For each group from the Table 1, 12-day mean composites of SAR Sentinel-1 satellite data with VV, VH bands with 10-meters spatial resolution were created [18]. The SAR Sentinel-1 preprocessing steps were: apply orbit, border noise removal, thermal noise removal, radiometric calibration, orthorectification, filter box 3x3 were performed as preliminary processing. As a result, the 12-day time series of Sentinel-1 was formed for every group of countries.

Moreover, in the classification process, Sentinel-2 data with preprocessing Level-2A and a spatial resolution of 10 meters are employed, utilizing four bands: Red (B4), Green (B3), Blue (B2), and Infrared (B8). The revisit time of Sentinel-2 is every 5 days; however, due to significant cloud cover, three composites are generated for each group of countries. These composites represent the median value of all available data within every 5-day period for the respective bands. To mitigate the impact of clouds on optical data, a Scene Classification Map (SCL) band with a spatial resolution of 20 meters is utilized for cloud masking. The specific dates utilized for both Sentinel-1 and Sentinel-2 are detailed in Table 2.

Table 2
Satellite data used

Satellite	Dates (2022)		
Sentinel-1	01.03 - 13.03; 13.03 - 25.03; 25.03 - 06.04; 06.04 - 18.04; 18.04 - 30.04; 30.04 - 12.05; 12.05 - 24.05;	24.05 - 05.06; 05.06 - 17.06; 17.06 - 29.06; 29.06 - 11.07; 11.07 - 23.07; 23.07 - 04.08; 04.08 - 16.08;	16.08 - 28.08; 28.08 - 09.09; 09.09 - 21.09; 21.09 - 03.10; 03.10 - 15.10; 15.10 - 27.10
Sentinel-2	01.03 - 01.06;	01.06 - 01.08;	01.08 - 30.09

Consequently, our input consists of a stack of raster images, comprising a time series of radar and optical satellite data, encompassing a total of 76 spectral bands. This includes 32 VV bands, 32 VH bands from Sentinel-1, and three composites featuring four bands each (red, green, blue, and infrared) from Sentinel-2.

4. Train and Validation Data

For training and testing the creation of forest type maps for Europe, the LUCAS Copernicus 2018 open dataset [2] serves as the primary resource. Despite being based on 2018 data, this dataset remains suitable for forest type classification due to the relatively slow change in forest types over time. A five-year span is not considered extensive for a land cover type like forests. To update this dataset for 2022, the global land cover data from WorldCover 2021 [8] was utilized. Samples that exhibited a change in class between 2018 and 2021 were subsequently excluded from consideration. The great advantage of LUCAS Copernicus 2018 data set is that for each sample there were 5 photos that confirm the correctness of the class that is entered for this sample.

As we can see from Figure 2, some countries are not covered by the given data set (Norway, Switzerland, Serbia, Montenegro, Albania, Bosnia and Herzegovina, North Macedonia, Turquie, Iceland, and Ukraine). Therefore, additionally, based on the image interpretation of the Sentinel-2, forest and another land cover samples were added for those areas where there was a lower coverage with the initial Lucas Copernicus data set. Also, two classes — rocks and snow — were added to the data set, because due to the mountainous terrain, there was a confusion of forest classes due to shadows and slopes.

The Figure 3 shows an example of the difference between broadleaved and coniferous forest in summer and winter season, as well as an example of a classification map that separates them. This confirms the need to use a time series of satellite data and necessarily composites of optical data.

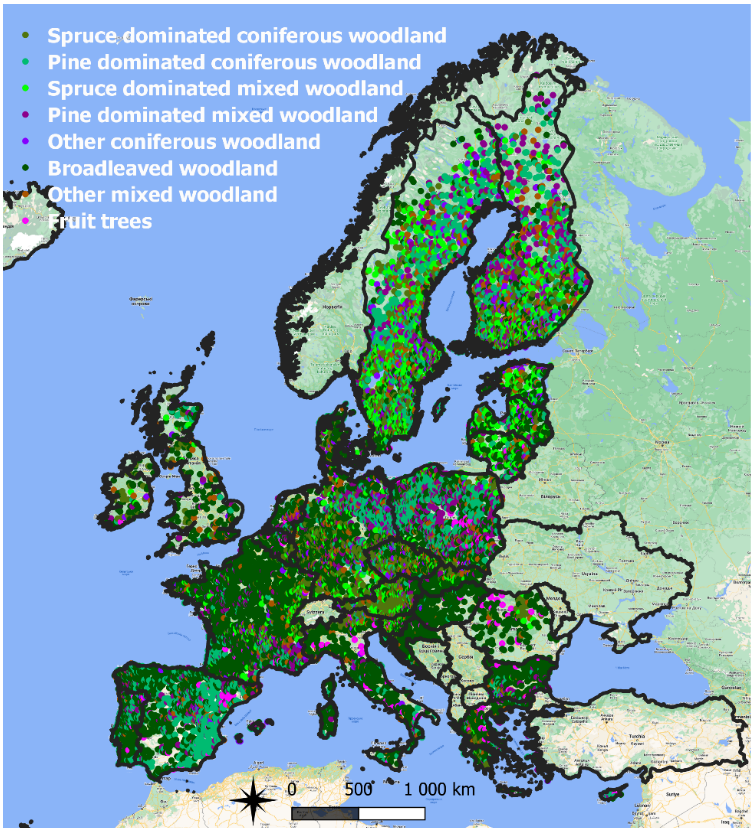

Figure 2: The preprocessed LUCAS Copernicus 2018 dataset serves as the foundation for the forest type classification in 2022.

For Ukraine, we used a data set that we have been collecting along the roads to create a land cover map [19]. In 2022, while collecting data along the roads, we also separated different types of forests (coniferous, broadleaved and mixed), which gave us the opportunity to build a map of forest types for the territory of Ukraine. Note that no other data, such as LUCAS 2018, forest type map 2018, which divide forests into different types and have a high spatial resolution, were not available for

Ukraine. Besides that, this dataset was extended with new samples using satellite image interpretation approach. The total number of samples in the dataset for Ukraine is 934 forests and 7862 other land and is shown in the Table 3. The generated data set was divided into training and test dataset in a ratio of 80:20 for each oblast of Ukraine.

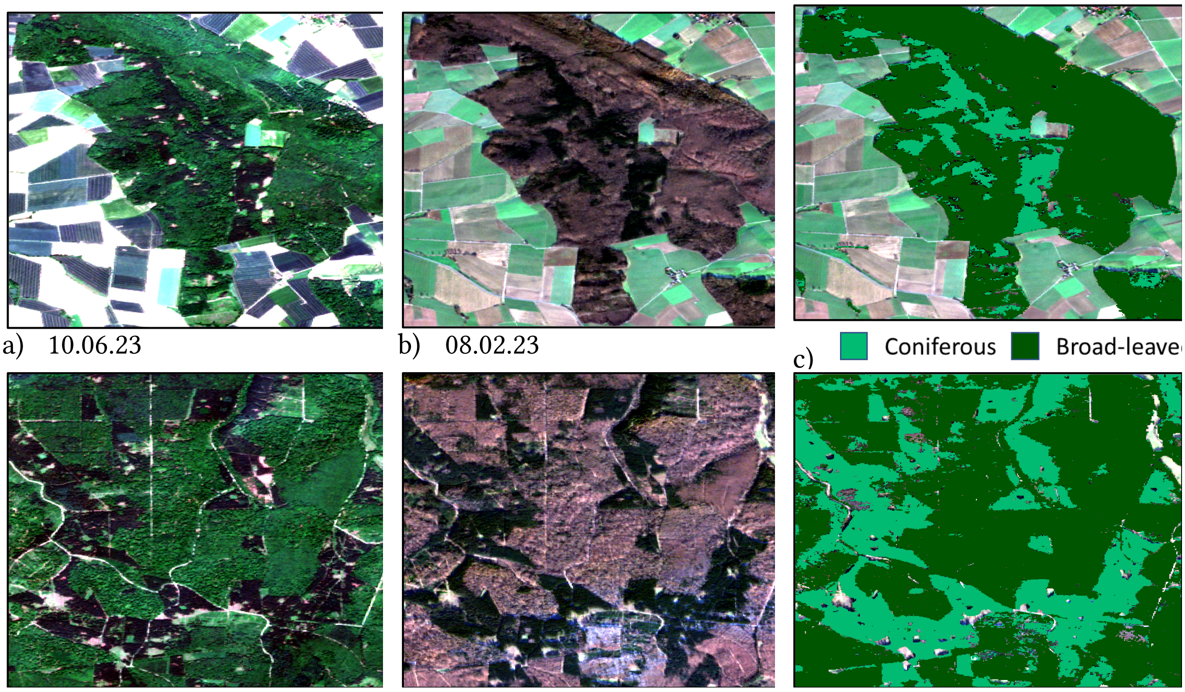

Figure 3: An example of broadleaved and coniferous forest in summer (a) and winter (b) season, and forest type map (c) to separate them.

Table 3
Dataset for Ukraine 2022

Land cover classes	Train	Test	Total
Artificial	567	131	698
Cropland	3274	759	4033
Gardens	245	56	301
Broadleaved forest	420	94	514
Coniferous forest	216	43	259
Mixed forest	133	28	161
Grassland	1427	343	1770
Bare land	242	50	292
Water	623	145	768
Total Forest	769	165	934
Total	7147	1649	8796

The Table 4 displays the distribution of data across European countries by class, while Figure 4 illustrates the geospatial distribution of the resulting dataset for three types of forests (broadleaved, coniferous, and mixed) within selected groups, as Table 1. To train the model and validate the resulting product, the dataset was divided into an 80:20 ratio within each distinct group.

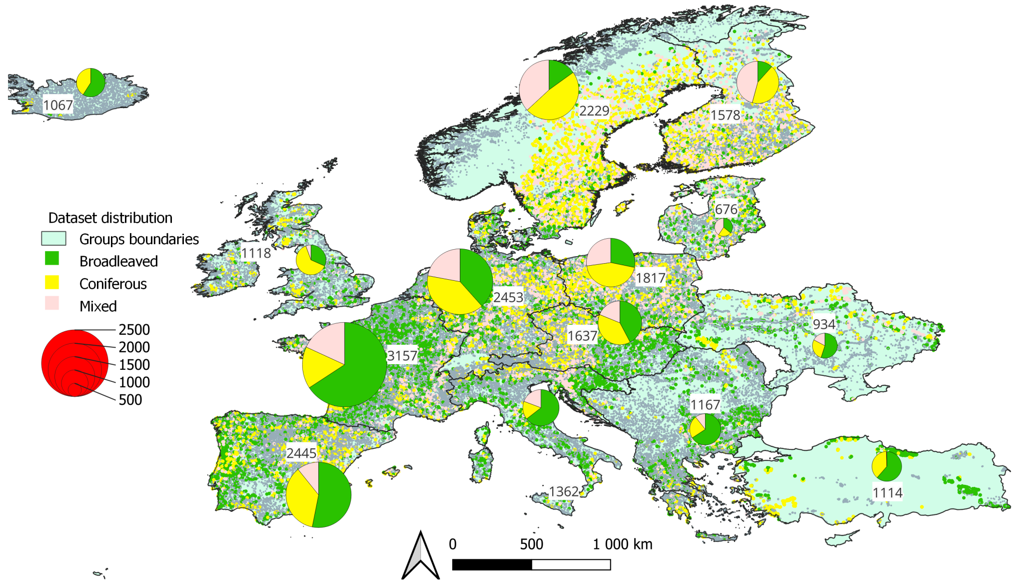

Figure 4: The geospatial distribution of the prepared data set for 3 types of forests (broadleaved, coniferous and mixed) by selected groups.

Around the beginning of October 2023, after the SWIFTT project was started and forest type classification map were created, the LUCAS preliminary micro data (2022) [20] was published as open data source. This dataset is similar to LUCAS Copernicus 2018 but contains more samples. In this regard, we have used this dataset for validation, and for the correctness of the experiment, we removed from this data set those elements that were used for training and validation of the obtained forest map and used the obtained data set as a completely independent data set for testing the obtained product. The last column of Table 4 shows the distribution of elements by land cover classes and forest types of LUCAS 2022.

Table 4
Train and validation data used (Lucas 2018 and Lucas 2022)

Class	Name	Lucas 2018			Lucas 2022
		Train	Test	Total	Test (Total)
1	Artificial	9901	2880	12781	3519
2	Cropland	16774	5207	21981	85216
3	Gardens	476	122	598	4085
4	Broadleaved woodland	7690	2426	10116	58855
5	Spruce dominated coniferous woodland	2977	899	3876	8804
6	Pine dominated coniferous woodland	2553	1098	3651	11889
7	Other coniferous woodland	347	150	497	2821
8	Spruce dominated mixed woodland	1112	500	1612	5532
9	Pine dominated mixed woodland	1228	483	1711	4598
10	Other mixed woodland	951	340	1291	4478
11	Grassland	26295	8120	34415	91969
12	Water	4355	1837	6192	10268
13	Bare, Rocks	12500	3798	16298	3138
14	Permanent Snow	1172	347	1519	0
	Total Forest	16858	5896	22754	96977
	Total	88331	28207	116538	295172

5. Methodology

Within the project, we have developed an automatic technology for forest type classification map for Europe in the Google Earth Engine cloud platform, which is repeated for each individual group of countries listed in the Table 1.

The prepared stack of satellite Sentinel-1 and Sentinel-2 composites (section 3) together with prepared pre-filtered LUCAS Copernicus 2018 train data (section 4) were used as input data for forest type classification [21], [22]. All data (satellite and train data set) is contained in cloud and we don’t need extra resource to train the classifier. For each separate group of countries, we trained different Random Forest models due to GEE capacity with 100 number of trees.

The model generates a raster georeferenced image as output, with each pixel representing the respective land cover or forest type class. Notably, this approach offers an advantage as it enables the creation of maps even in regions with limited training data, such as Turkey and Iceland. This is achieved through the utilization of a common Random Forest model for each group of countries. Leveraging the cloud platform Google Earth Engine facilitates seamless scalability and utilization of the model across extensive areas, particularly throughout Europe, as demonstrated by previous studies [23], [24].

The next steps no require large computing resources and can be done on a personal computer (PC).

The list of classes on the received map corresponds to the list of classes in the educational data (Table 4). Aggregation of the received forest types into three generalizing classes (Broadleaved, Coniferous, and Mixed) took place according the [20] (Table 5). All other classes not related to forest cover were set to 0.

Table 5
Forest type aggregation according LUCAS classification

Description	Class	LUCAS	Description	Class	LUCAS
Broadleaved woodland	4	C10	Broadleaved woodland	1	C10
Spruce dominated coniferous woodland	5	C21	Coniferous woodland	2	C20
Pine dominated coniferous woodland	6	C22			
Other coniferous woodland	7	C23			
Spruce dominated mixed woodland	8	C31	Mixed woodland	3	C30
Pine dominated mixed woodland	9	C32			
Other mixed woodland	10	C33			
Non forest	1 - 3, 11 -14	-	Null	0	-

A portion of the LUCAS 2018 validation dataset (outlined in section 4), along with the LUCAS 2022 data, served as independent data for validating the aggregated forest type map. To evaluate the accuracy of the land cover classification maps, a confusion matrix [25] derived from an independent test sample was utilized. This matrix is presented as a rectangular table, where each cell represents the number of pixels n_{ij} belonging to the actual class "i" but classified as class "j" on the classification map. Additionally, the assessment included metrics such as Overall Accuracy (OA), Producer Accuracy (PA), and User Accuracy (UA). UA and PA values are ways of representing the accuracies of the individual classes. The UA value is the probability that the pixel class on the classification map corresponds to the sample class in the test data, whereas PA indicates the probability that a pixel from the test data is recognized correctly on the map. By analyzing the correlation between these quantities, the map user can obtain information on reliability of class recognition on the map, as well as assess the quality of the geospatial product itself.

Overall Accuracy (OA) is an indicator of the overall quality of land cover map. In fact, this is the ratio of the sum of the elements of the main diagonal (i.e., correctly classified pixels) and the sum of all elements in the error matrix.

The User, Producer, and Overall accuracies are calculated according to the formulas 1-3:

$$PA_j = \frac{n_{jj}}{\sum_{i=1}^{q} n_{ij}}; \tag{1}$$

$$UA_i = \frac{n_{ii}}{\sum_{j=1}^{q} n_{ij}}; \tag{2}$$

$$OA = \frac{\sum_{i=1}^{q} n_{ii}}{\sum_{i=1}^{q}\sum_{j=1}^{q} n_{ij}}, \tag{3}$$

where q is the number of classes on the land cover map and test data.

One more metric for the assessment of the classifier quality is F1-score. It reduces the two other metrics, UA and PA, down to one number, and it is defined as an average-weighted harmonic mean value between them (formula 4).

$$F_1 = 2\frac{\mathrm{PA} * \mathrm{UA}}{\mathrm{PA} + \mathrm{UA}} \tag{4}$$

The above-described methodology for obtaining a classification map of forest types for Europe is schematically presented in the Figure 5.

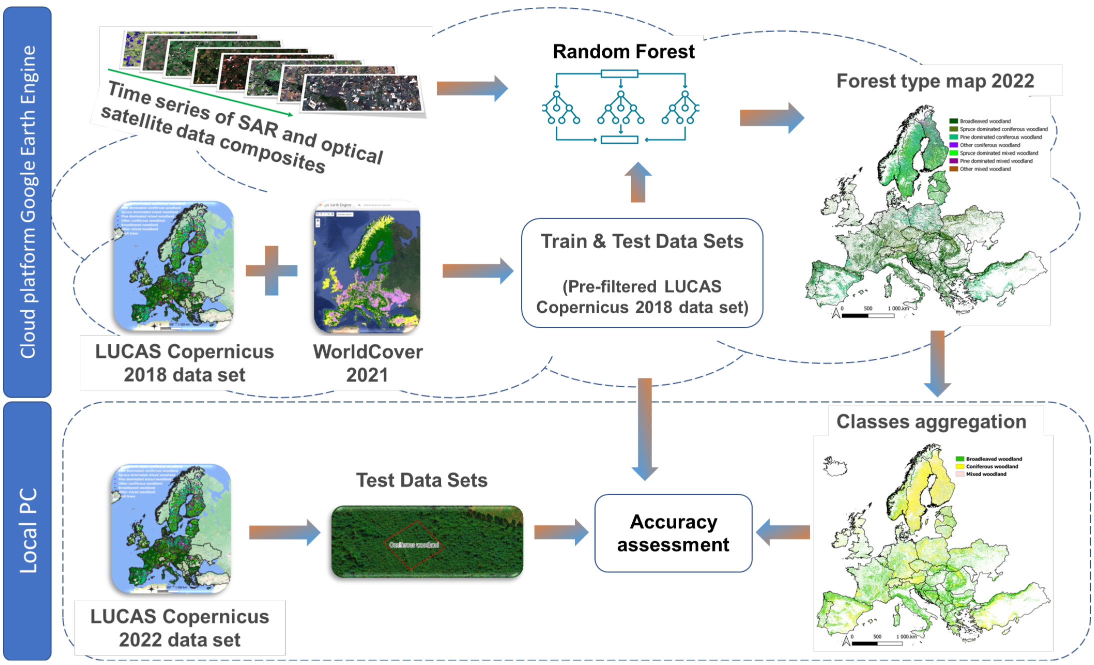

Figure 5: Workflow of forest type map creation.

6. Machine Learning Model Selection

The Random Forest classifier was chosen as a basis in this work, because it showed relatively good results compared to deep neural networks, it is built into the GEE cloud platform and can be used in contrast to deep networks, and it is also significantly superior in algorithm execution time. In the study [26], researchers introduced and investigated two modifications of Random Forest (RF) models and two modifications of Convolutional Neural Network (CNN) U-Net for forest type classification. The raster Forest Type (2018) map [13] served as training labels. During the development of these models, it was observed that RF might not be the optimal tool for addressing this problem, primarily due to its architectural constraints, which hinder training the model on large datasets. However, for point-based datasets, this method remains reasonable due to its relatively low learning and execution time. The Random Forest model exhibited an overall accuracy ranging from 86% to 86.5% for the pilot territories in Germany. Conversely, within this study, the U-Net model demonstrated significantly improved overall accuracy, ranging from 91% to 91.7%, along with higher prediction speed. Nonetheless, as with deep learning methods, the learning process of the U-Net model extends to tens of hours, unlike the minutes required for RF. In this work, all experiments were conducted on the CREODIAS cloud platform (within ORCE project [27]), which contains ready-to-use satellite data, but still requires time and resources for their partial pre-processing (cloud extraction, filtering, etc.), unlike the GEE cloud platform.

In the work [28] authors compare the productivity of cloud platforms AWS and CREODIAS for the land cover classification for Ukraine using Multilayer perceptron. On average, it takes 400 minutes and 50-80 euros to build a classification map based on a time series of satellite data for 1 year, covering an area of about 200,000 km^2. Taking into account the territory area of 6.5 MLN km2, for which we need to obtain a forest type classification map, the necessary time to obtain the map will be about 9 days of continuous work of the instance and 1700 - 2600 euros for one iteration of obtaining the classification map. Given the complexity of the mountainous terrain in some countries, sometimes there is a need to adjust the sampling in difficult areas and rerun the training model several times.

Therefore, taking into account the area for which we need to obtain a classification map, in this work the authors decided to use the Random Forest algorithm in the GEE cloud platform.

7. Results

The main outcome of this study is the forest type classification map for the year 2022 covering the European territory, featuring a spatial resolution of 10 meters. This map was generated utilizing a time series of Synthetic Aperture Radar (SAR) Sentinel-1 and optical Sentinel-2 satellite data. The resulting map includes 3 forest type classes (broadleaved, coniferous and mixed) and presented in the Figure 6. The Table 6 shows the accuracy assessment of the obtained map on the test data, which was formed on the basis of the Lucas 2018 data set.

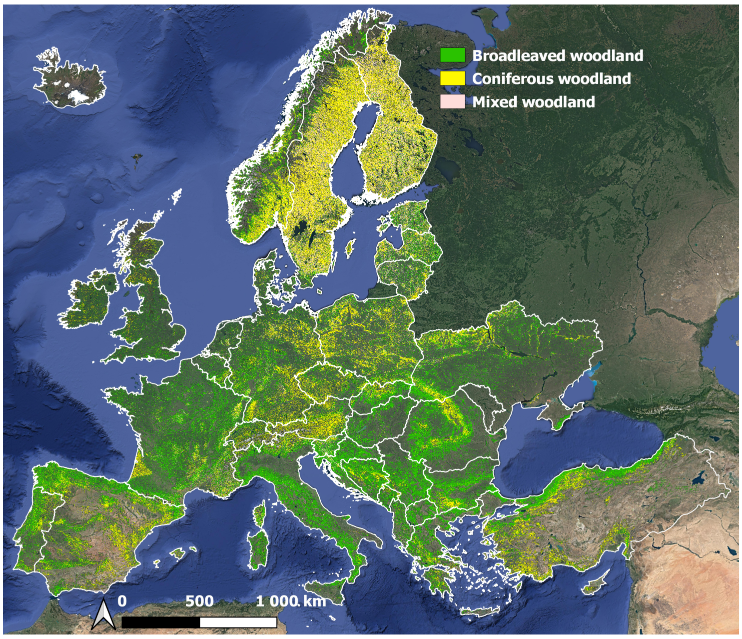

Figure 6: Forest type map 2022 for European countries.

Table 6
Overall accuracy of forest map for Europe based on LUCAS 2018 and LUCAS 2022 test dataset

	Test LUCAS 2018			LUCAS 2022		
Class	UA	PA	F1	UA	PA	F1
Forest	87,9	92,8	90,3	86,2	77,5	81,6
Broadleaved woodland	76,1	84,3	80	69,6	68,6	69,1
Coniferous woodland	75,9	78,2	77,1	65,7	57,1	61,1
Mixed woodland	64,1	39,1	48,5	41,4	24,8	31
Non forest	98,1	99	98,6	89,5	93,9	91,7
Overall accuracy	93,2			88,5		

The class of mixed forests turned out to be the most problematic, since it contains various types of trees that can be contained in other classes. Accordingly, the confusion for this class is the highest, and the recognition accuracy is the lowest. Accordingly, since it may contain trees of other classes, it also leads to an underestimation of the accuracy of other classes of forest types.

In contrast to the Lucas 2018 dataset, the Lucas 2022 dataset contains sampling elements on a grid. Very often there are cases when the point, which is responsible for the forest class, falls into a forest strip or in a city in a cluster of a few of trees. Or a point from the Lucas 2022 data set is very close to the cluster of trees, but does not physically fall into the forest mask, as well as the sample could be incorrect. In such cases, given the spatial resolution of the satellite data we work with, the created classification map may not correspond to the class given in the Lucas 2022 set. These examples certainly affect the classification accuracy estimation, but in this case we have an accuracy estimate from the bottom of our product, which is 88.5% on independent data, which is a good indicator at the level of Europe as a whole.

Also, we compared the resulting areas of forest types and total forest in 2022 with the corresponding existing global products. With Forest type 2018, the areas of broadleaved and coniferous forest are compared (Figure 7). There is a decrease in the area of coniferous forests compared to 2018. This may be due to the fact that coniferous forests suffer more from bark beetles and various diseases. Trees dry out, and as a result, they are cut down.

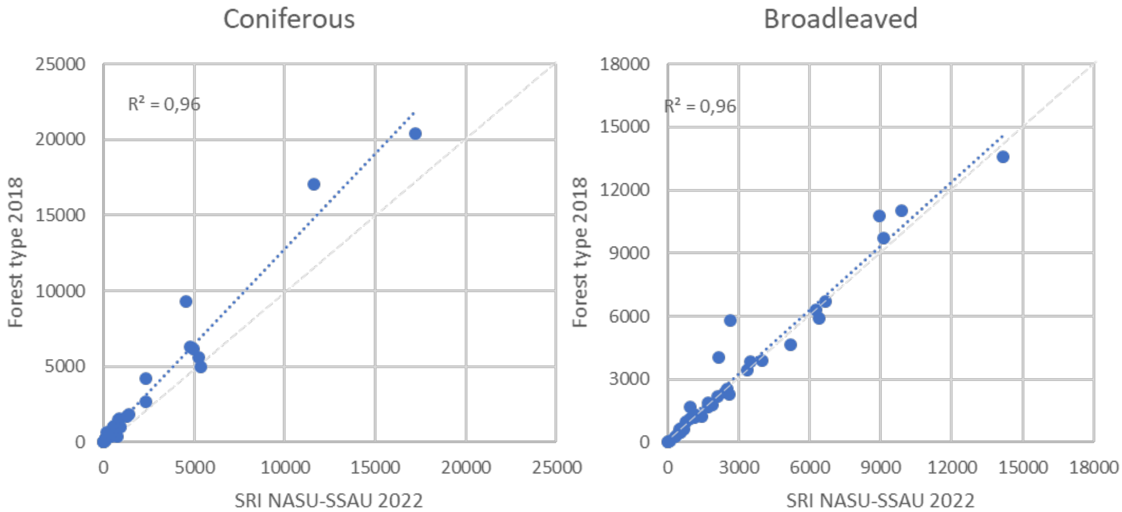

Figure 7: Area comparison of SRI coniferous and broadleaved forest 2022 with Forest type 2018 product.

If we compare the total area of the forest (Figure 8, Figure 9), there is a difference between the Hansen 2022 data set (Hansen underestimates the forest compared to our product). For Hansen's data set, a threshold of 30% of pixels belonging to the forest was chosen, below which we consider it inappropriate to take. The reason for the lower forest area in Hansen's data set is that every year deforestation is subtracted from the forest area, and the last layer of forest growth (gain layer) was only in 2012.

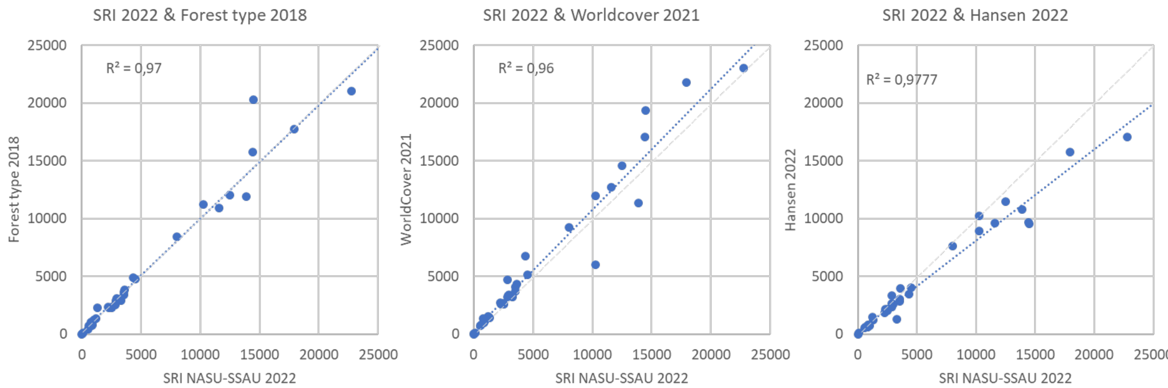

Figure 8: Area comparison of SRI forest 2022 with Forest type 2018 product, WorldCover 2021 and Hansen 2022.

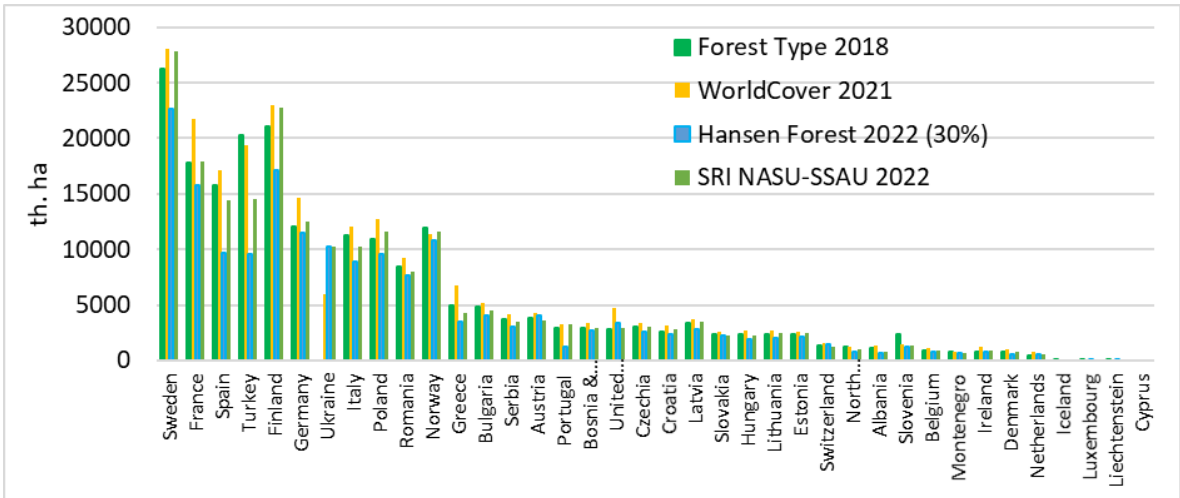

Figure 9: Area comparison of SRI forest 2022 with Forest type 2018 product, WorldCover 2021 and Hansen 2022.

8. Product Availability

The source codes are available for downloading at the link: https://github.com/IPT-MMDA/Forest_type_classification. The code is available in Google Earth Engine cloud platform: for forest type classification - https://code.earthengine.google.com/0eb34e8a84988ad3dcea334d4ef59805, for resulting forest type map using - https://code.earthengine.google.com/baa95cfdd09e0630e21f6aa4ffe87eaa, forest type visualization within Google App: https://ee-swiftt.projects.earthengine.app/view/foresttype.

9. Discussion

The results obtained in this study demonstrate the effectiveness of employing satellite data and machine learning techniques for large-scale forest type mapping across Europe. The application of the Random Forest classifier in the Google Earth Engine cloud platform has proven to be a reliable and efficient approach, addressing the challenges associated with processing and analyzing of big satellite data.

One of the key strengths of the proposed methodology lies in its ability to overcome the limitations posed by the lack of ground truth data in certain regions. By leveraging the LUCAS

Copernicus dataset and satellite image interpretation, we were able to create a comprehensive training dataset, ensuring adequate representation of different forest types and land cover classes across the entire study area.

The resulting forest type map exhibits a high overall accuracy of 93% when evaluated against an independent test dataset. However, it is essential to acknowledge the challenges encountered in accurately distinguishing mixed forests from other classes. The inherent complexity of mixed forests, comprising various tree species, likely contributed to the lower accuracy observed for this particular class.

It is noteworthy that the accuracy assessment based on the LUCAS 2022 dataset revealed certain inconsistencies, potentially attributable to the grid-based sampling approach employed in this dataset. The mismatch between the spatial resolution of the satellite data and the sampling points may have resulted in instances where the assigned class did not align with the actual land cover observed in the satellite imagery. Nevertheless, the overall accuracy of 88.5% obtained using the LUCAS 2022 dataset provides a conservative estimate of the product's performance, further validating its reliability.

The comparison of the obtained forest areas with existing global products, such as the Forest Type 2018 and WorldCover 2021, highlights the dynamic nature of forest ecosystems and the importance of regularly updating forest maps. The observed decrease in coniferous forest areas compared to 2018 may be attributed to factors such as pest infestations, diseases, and subsequent deforestation efforts. Conversely, the differences in total forest area compared to the Hansen dataset can be explained by the consideration of deforestation and the limited availability of forest gain data in the Hansen product after 2012.

10. Conclusions

This study successfully developed and implemented a semi-supervised approach for mapping European forest types using high-resolution satellite data from the Copernicus program. The integration of Sentinel-1 and Sentinel-2 data, combined with the Random Forest classifier deployed on the Google Earth Engine cloud platform, enabled efficient and accurate classification of forest types across the European continent.

The resulting forest type map, with a spatial resolution of 10 meters, provides a comprehensive and up-to-date representation of the distribution of broadleaved, coniferous, and mixed forests for the year 2022. This product addresses a critical need for contemporary geospatial information, facilitating informed decision-making in various domains, including forest management, conservation efforts, and the assessment of environmental impacts.

The accuracy assessment, conducted using independent validation data from the LUCAS Copernicus dataset, demonstrated an overall accuracy of 93%, highlighting the reliability of the proposed methodology. While the mixed forest class exhibited lower accuracy due to its inherent complexity, the overall performance of the classification model was satisfactory.

By leveraging cloud computing resources and machine learning techniques, this study overcame the challenges associated with processing and analyzing large volumes of satellite data, enabling the efficient generation of a high-resolution forest type map for the entire European region.

Overall, this study demonstrates the potential of integrating cutting-edge technologies and data sources for addressing complex environmental challenges at a continental scale. The developed methodology can be further refined and applied to other regions, contributing to a better understanding of global forest dynamics and supporting sustainable forest management practices.

Future research could explore the integration of additional data sources, such as LiDAR or hyperspectral imagery, to enhance the accuracy of forest type classification, particularly for mixed forests. Additionally, the development of automated monitoring systems, leveraging the capabilities of cloud computing and machine learning, could facilitate the timely detection of forest disturbances and support proactive conservation efforts.

Acknowledgements

The study was supported by the HORIZON Europe project SWIFTT No. 101082732, "Satellites for Wilderness Inspection and Forest Threat Tracking."

References

[1] B. Yailymov, et al., Google Earth Engine framework for satellite data-driven wildfire monitoring in Ukraine. Fire, 2023, 6(11), 411.

[2] G. Chen, R. Meentemeyer, Remote sensing of forest damage by diseases and insects. Remote sensing for sustainability, 2016, 145-162.

[3] M. Matsala, et al, War drives forest fire risks and highlights the need for more ecologically-sound forest management in post-war Ukraine. Scientific reports, 2024, 14(1), 4131.

[4] P. Verkerk, et al, Spatial distribution of the potential forest biomass availability in Europe. Forest Ecosystems, 2019, 6(1), 1-11.

[5] C. Senf, R. Seidl R., Mapping the forest disturbance regimes of Europe. Nature Sustainability, 2021, 4(1), 63-70.

[6] M. Palmero-Iniesta, et al, Changes in forest landscape patterns resulting from recent afforestation in Europe (1990–2012): Defragmentation of pre-existing forest versus new patch proliferation. Annals of Forest Science, 2020, 77(2), 1-15.

[7] D. Zanaga,et al, ESA WorldCover 10 m 2020 v100. URL: https://doi.org/10.5281/zenodo.5571936.

[8] D. Zanaga, et al, ESA WorldCover 10 m 2021 v200. URL: https://doi.org/10.5281/zenodo.7254221.

[9] R. Kerchove, Release of the 10 m worldcover map. URL: https://blog.vito.be/remotesensing/release-of-the-10-m-worldcover-map.

[10] M. Hansen, et al, High-Resolution Global Maps of 21st-Century Forest Cover Change. Science 342, 2013, 850–53. URL: https://glad.earthengine.app/view/global-forest-change.

[11] C. Senf, European forest disturbance maps (1.0.0), 2020. [Data set]. Zenodo. https://doi.org/10.5281/zenodo.3924381.

[12] M. Buchhorn, et al., Copernicus Global Land Cover Layers-Collection 2. Remote Sensing, 2020, 12, 108, 1044. doi:10.3390/rs12061044.

[13] Forest Type 2018 (raster 10 m), 2020. URL: https://land.copernicus.eu/pan-european/high-resolution-layers/forests/forest-type-1/status-maps/forest-type-2018.

[14] A. Dostálová, et al, European wide forest classification based on sentinel-1 data, Remote Sensing, 2021, 13(3), 337.

[15] P. Kempeneers, et al, Data fusion of different spatial resolution remote sensing images applied to forest-type mapping. IEEE Transactions on Geoscience and Remote Sensing, 2011, 49(12), 4977-4986.

[16] R. Päivinen, et al, Combining earth observation data and forest statistics, Joensuu: European Forest Institute. 2001.

[17] Countries, 2020. URL: https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units/countries.

[18] N. Kussul, et al, Data assimilation technique for flood monitoring and prediction. International Journal Information Theories & Applications. 2008, 15, 76-83.

[19] F. Waldner, et al. Roadside collection of training data for cropland mapping is viable when environmental and management gradients are surveyed. *International Journal of Applied Earth Observation and Geoinformation* 80 (2019): 82-93.

[20] LUCAS preliminary micro data 2022. URL: https://ec.europa.eu/eurostat/web/lucas/database/2022.

[21] A. Shelestov, et al, Cloud approach to automated crop classification using Sentinel-1 imagery. IEEE Transactions on Big Data, 2019, 6(3), 572-582. DOI: 10.1109/TBDATA.2019.2940237.

[22] A. Shelestov, et al, Cloud-Based Technologies for Data Processing in Ukraine: International Context. In International Scientific and Technical Conference-Modern Challenges in

Telecommunications, 2021, pp. 101-118. Cham: Springer International Publishing. DOI: 10.1007/978-3-031-16368-5_5.
[23] N. Kussul, et al, Land cover changes analysis based on deep machine learning technique. Journal of Automation and Information Sciences, 2016, 48(5).
[24] A. Shelestov, et al, Large scale crop classification using Google earth engine platform. In 2017 IEEE international geoscience and remote sensing symposium (IGARSS), 2017, pp. 3696-3699. DOI: 10.1109/IGARSS.2017.8127801.
[25] R. Congalton, A review of assessing the accuracy of classifications of remotely sensed data. Remote sensing of environment, 1991, 37(1), 35-46.
[26] Y. Salii, et al, Machine learning models and technology for classification of forest on satellite data. In IEEE EUROCON 2023-20th International Conference on Smart Technologies, 2023, pp. 93-98.
[27] A. Shelestov, E. Siemens, OCRE project: consortium, main goals, approaches and opportunities. In 2022 12th International Conference on Dependable Systems, Services and Technologies (DESSERT), 2022.
[28] N. Kussul, et al, Cloud Platforms and Technologies for Big Satellite Data Processing, in International scientific and practical conference Information and communication technologies and sustainable development, 2023, pp. 303-321. Cham: Springer Nature Switzerland.
[29] R. Andrimont, et al, LUCAS Copernicus 2018. Earth Observation relevant in-situ data on land cover throughout the European Union. Earth System Science Data Discussions, 2020, 1-19, https://doi.org/10.5194/essd-13-1119-2021.