# Machine learning based combinatorial analysis for land use and land cover assessment in Kyiv City (Ukraine)

**Vadym Belenok**[a,*] **Liliia Hebryn-Baidy**[a] **Nataliia Bielousova**[a] **Valeriy Gladilin**[b] **Sergíy Kryachok**[c] **Andrii Tereshchenko**[a] **Sofiia Alpert**[d] **and Sergii Bodnar**[e]

[a]National Aviation University, Department of Aerospace Geodesy and Land Management, Kyiv, Ukraine

[b]Bila Tserkva National Agrarian University, Department of Geodesy, Cartography and Land Management, Bila Tserkva, Ukraine

[c]Chernihiv Polytechnic National University, Department of Geodesy, Cartography and Land Management, Chernihiv, Ukraine

[d]Scientific Centre for Aerospace Research of the Earth of the Institute of Geological Science of the National Academy of Sciences of Ukraine, Department of Geoinformation Technologies in Remote Sensing of the Earth, Kyiv, Ukraine

[e]Taras Shevchenko National University of Kyiv, Department of Geodesy and Cartography, Geography Faculty, Kyiv, Ukraine

**Abstract.** The main goal of this study is to evaluate different models for further improvement of the accuracy of land use and land cover (LULC) classification on Google Earth Engine using random forest (RF) and support vector machine (SVM) learning algorithms. Ten indices, namely normalized difference vegetation index, normalized difference soil index, index-based built-up index, biophysical composition index, built-up area extraction index (BAEI), urban index, new built-up index, band ratio for built-up area, bare soil index, and normalized built up area index, were used as input parameters for the machine learning algorithms to improve classification accuracy. The combinatorial analysis of the Sentinel-2 bands and the aforementioned indices allowed us to create four combinations based on surface reflectance characteristics. The study includes data from April 2020 to September 2021 and April 2022 to June 2022. The multitemporal Sentinel-2 data with spatial resolutions of 10 m were used to determine the LULC classification. The major land use classes such as water, forest, grassland, urban areas, and other lands were obtained. Generally, the RF algorithm showed higher accuracy than the SVM. The overall accuracy for RF and SVM was 86.56% and 84.48%, respectively, and the mean Kappa was 0.82 and 0.79, respectively. Using the combination 2 with the RF algorithm and combination 4 with the SVM algorithm for LULC classification was more accurate. The additional use of vegetation indices allowed to increase in the accuracy of LULC classification and separate classes with similar reflection spectra.  [DOI: 10.1117/1.JRS.17.014506]

**Keywords:** land use and land cover; Sentinel-2; vegetation index; random forest; support vector machine; Google Earth engine.

Paper 220428G received Jul. 17, 2022; accepted for publication Jan. 2, 2023; published online Jan. 24, 2023.

## 1 Introduction

Anthropogenic activities have a great influence on the Earth's land surface. The lands are intensively used for the needs of energy extraction, management of natural resources, food production, and the development of urban areas.[1] The transformation of natural ecosystems, because of human activity, changes the biophysical properties of the Earth's surface.[2] Rapid urban growth is a global problem, with direct or indirect expressed in environmental consequences such as local air pollution, biodiversity loss, soil degradation, water crises, reduction of green areas in

*Address all correspondence to Vadym Belenok, belenok.vadimnau.edu.ua

1931-3195/2023/$28.00 

the cities and all of which are leading to gradual accumulation of climate change.[3,4] Land use land cover (LULC) is a fundamental concept of sustainable development. That is why LULC mapping has been recognized as the main task for deriving information about ecological monitoring at global, regional, and local scales.[5] The generation of global LULC products that are stable and consistent over time while also reflecting the land surface seasonality, allows rational management of natural resources.[6] LULC mapping is one of the most important and earliest applications of remote sensing technology that is constantly developing and improving. Remote sensing is the technology for effective and accurate land cover mapping at large scales. There are numerous advantages such as repeatability of observations, cost-effectiveness, and tracking changes over time.[7] Updates and improvements in data types allow for significant progress in land cover mapping across a range of spatial resolutions.[8] Review of articles[9–11] shows that the contemporary high adoption and application of Sentinel-2 can be attributed to the high spatial and temporal resolution. The free access policy drives the increasing use of Sentinel-2 data, especially in the developing countries. Many studies have focused on comparing the performance of LULC data derived from Sentinel-2 and Landsat imagery.[12–17] In the current and past research,[18,19] we have been using Sentinel-2 imagery, as this data is fully relevant to the research objectives. Generating various-resolution LULC maps require massive amounts of data as huge storage capacities, high processing power, and the flexibility to apply diverse approaches are all required. To solve all of these problems, we used the Google Earth Engine (GEE), which was used for easier implementation. This combines huge amounts of remote sensing data from multiple sources with a high-performance computer service that allows you to quickly and easily calculate satellite images.[20–23]

Mapping of the LULC classification is based on remote sensing imagery using nonparametric machine learning algorithms such as random forest (RF) and support vector machine (SVM). These are mostly used to provide accurate land cover maps.[24,25] The RF classifier is suitable for classifying multispectral data, less sensitive to the quality of the training samples, to overfitting, this is due to the large number of decision trees than other machine learning classifiers.[9,26–30] The performance of different machine learning algorithms on the satellite image in a rural or urban setup shows the SVM with an overall accuracy (OA) of 0.969 comparatively better results than the other two algorithms.[15] In turn, the classification of a complex tropical forest with a combination of dry and rainy seasons showed high accuracy with the SVM classifier (80.3%), whereas RF obtained a lower accuracy (63.9%).[10] The classification results are clearly influenced by the type of surface and the complexity of the conditions. The authors[31–33] use the RF classifier to map landscapes in a rapidly urbanizing region and to assess the impact of urbanization on the environment. According to Ref. 32, transformation of various LULC types into urban/built-up areas based on RF (91% to 98%) is statistically significant. The urban/built-up areas are continually increasing at the expense of agricultural and forest lands. This proves the accuracy of the RF classifier for LULC classification of urban areas.

To increase the accuracy of spectral classification LULC, approaches based on various spectral vegetation indices (VI) are used. The research results confirm the effectiveness of the VI as Normalized difference vegetation index (NDVI), NDBI, and NDWI with the corresponding Sentinel-2 red-edge bands using RF classifiers (94.85%) to increase the accuracy of classification LULC.[12,14] The creation of input composites comprising of spectral bands and EVI, NDBI, NDMI, NDVI, and normalized difference soil index (NDSI) indices allows it to improve the LULC classification of the city using the RF (88% to 89%) classifier.[31,34–36] The addition of these indices improves the classification in terms of sensitivity for determining the state of water on different surfaces.[11,37,38] For accurate mapping of a built-up area and bare soil, those that have similar spectral characteristics should be applied to a different VI. The applying of such spectral indices as BAEI, bare soil index (BSI), new built-up index (NBI), NDBI, urban index (UI), and index-based built-up index (IBI) in combination with spectral bands shows significantly higher classification accuracy LULC.[39–42] The use of normalized built-up area index (NBAI), band ratio for built-up area (BRBA), and BSI indices in combination with Sentinel-2 bands made it possible to better distinguish the classes of a built-up area and bare soil.[16,43] The NDSI index (newly developed spectral index) is compared with the BRBA, NBAI, NBI, and NDBI spectral indices to map the urban built-up area more accurately.[44] However, the authors used Landsat images with 30 m resolution. It is worth noting that the use of the biophysical composition index (BCI), PISI,

and CBI indices to characterize the urban environment allows for the separation of impervious surfaces, bare soil, and vegetation.[45–48] To accurately extract built-up area and bare land with RF classification is conducted using multisensor features, including temperature, night-time light, backscattering, and topography.[49] Applying very high-resolution satellite imagery and incorporating geometry, texture, and spectral information with object-based image classification is an alternative approach that requires more additional information to accurately identify urban classes.[13,24,50,51] An interesting approach for mapping the built-up area (building detection) is a classification model that combines spectral, height, and textural characteristics. However, it is desirable to use high-resolution images such as WorldView-2 and Spot-5 in order to achieve the required accuracy, which is not always possible.[52–55]

The main aim of the present study was to analyze ways to improve the accuracy of the LULC classification with machine learning algorithms RF and SVM based on multispectral Sentinel-2 satellite images. The decisive technical novelty of the study is that four spectral combinations were proposed. The combinations use 10 VI that are justified from a physical point of view. In doing so, the indices were chosen, in particular those that solve the subtask of separating LULC types with similar spectral reflectance properties, i.e., built-up and other lands (bare land). This study investigated these four combinations for RF and SVM, evaluating their performance in terms of Kappa coefficient, user, and producer accuracy values. Thus, we found the best of the proposed combinations, which most accurately classify the LULC of the urban area.

## 2 Materials and Methods

### 2.1 *Materials*

#### 2.1.1 *Characteristics of the study area*

Kyiv is the capital of an independent Ukraine that has been developing rapidly recently. At the same time, urban planning activities are often chaotic, not taking into account the additional systemic overloads in the performance of metropolitan functions. Additional development negatively affects not only the environment but also the health of the population and its natural rhythm of life.[56] Kyiv is located on the border of two natural and climatic zones of Ukraine – Polissia (northern part) and Forest steppe (southern territories).[57] The location affects its characteristics – vegetation, distribution, species, and on the composition of the soil. A minor part of Kyiv's greenery is the remnants of natural forests, which previously occupied the territory of the modern city. However, a larger part consists of artificial plantations created during the construction of the urban area.[19] The study area Kyiv municipality is shown in Fig. 1. It is constructed using geographic information technology, as described in Refs. 58 and 59.

As of February 2022, Kyiv was one of the ten most populated cities in Europe. According to the Kyiv General Department of State Statistics Committee of Ukraine, the population as of February 1, 2022, was 2,950,702 people part of the existing population and 2,909,395 people are the permanent population. This period constantly grew due to migration.[60] The factor of such growth was the progressive urbanization of the surrounding territories.[19] The economically active population was 16666.8 thousand people, including 14957.3 thousand employed and 1709.5 thousand unemployed. The economically inactive population was 295 thousand. The highest employment of the population by types of economic activity was professional, scientific, and technical activities – 127 thousand and education – 119.9 thousand people.[60] The average monthly temperature in the city in January is −0.9°C and in July, 25.6°C. The annual rainfall in 2021 was 456.3 mm.[61]

The analysis of the structure of the city's land fund shows that the main place belongs to built-up lands, which occupy an area of 37.00 thousand hectares, which is 44.2% of the total area of the city, and lands belonging to forest plantations with an area of 35.10 or 41.7%.[57] By Order No. 1425 of the Kyiv City Council dated June 22, 2021, the project of the City Target Program for the Use and Protection of Lands of the City of Kyiv for 2022 to 2025 was approved. One of the goals of this program is to update cartographic materials and digital orthophoto plans.[62]

The rapid expansion of Kyiv city is due to the increase of the built-up area at the expense of green areas of the city and the development of suburban green landscapes. The reduction of

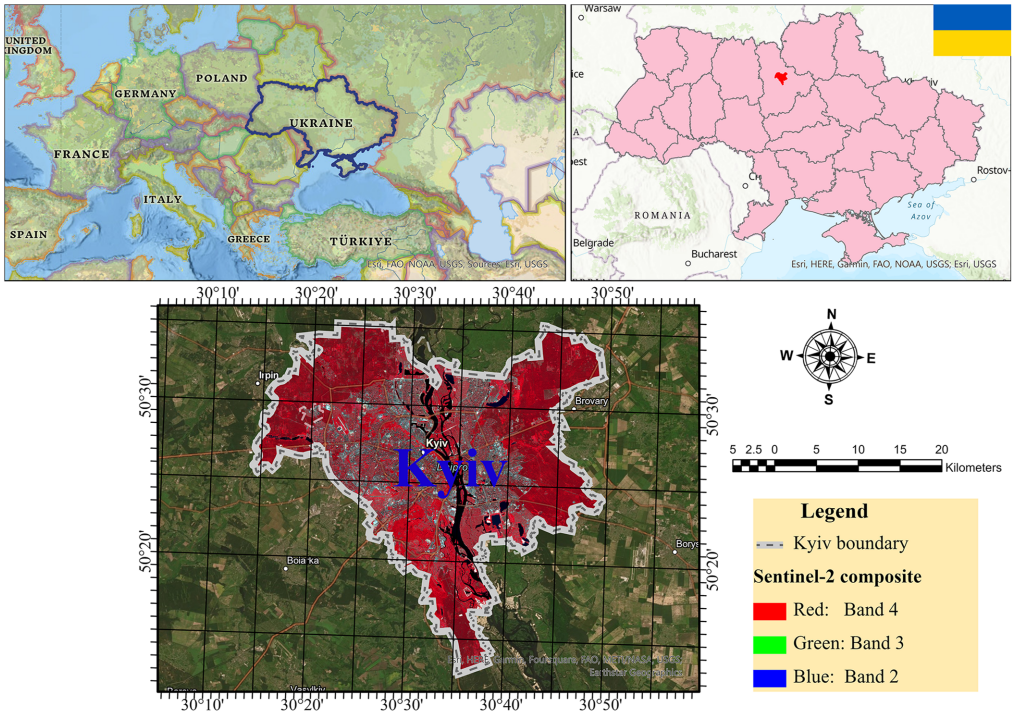


**Fig. 1** Study area. Source: an original study in ArcGIS Pro.

green zones and public spaces, chaotic and illegal development, and traffic that increases air pollution, all indicate the imperfection of the existing urban space planning system and require effective management decisions at the state level. According to the Kyiv General Plan 2040, planning decisions are outdated and based on Soviet approaches. Accurate and timely mapping of the city's LULC will strengthen planning and design decisions of urban planning policy, in particular, improve zoning and reduce the rate of environmental pollution. In this study, five main LULC classes were selected such as water, forest, grassland, built-up, and other lands. This classification allowed the best interpretation of the results of the study in accordance with the above problems.

Therefore, the study of land resources of the Kyiv based on satellite data is an important component for effective management of the city territory within the framework of increased urbanization.

### 2.1.2 *Google Earth engine*

In this study, to determine land use and land cover, we decided to use publicly available data from Sentinel-2 multispectral instrument (MSI). Sentinel-2 MSI data are available for free downloading from Copernicus Open Access Hub.[63] The disadvantage of this service is the ability to download only a single scene for a specific date. Since this study needed to use images for a long period, it was decided not to download individual scenes but to create mosaics of scenes using GEE cloud computing.[64] It not only allows downloading satellite data for a selected time interval but also contains powerful tools for processing remote sensing data and conducting geoinformation analysis.[65] Many limitations associated with loading, storing, and processing satellite data at different times, which usually arise when analyzing a large amount of spatial data, can be easily removed using GEE. The GEE platform consists of two main components, the GEE Explorer (for viewing datasets) and the GEE Playground, which work together. The Google EEP app, a JavaScript API, is used to download and analyze large satellite images and to perform complex geostatistical and geospatial operations.[54] The GEE is accessed over the Internet and requires a Google account to work with the code.

The data catalog contains an extensive repository of publicly available geospatial datasets, including observations from various satellite and aerial imaging systems in optical and non-optical

**Table 1** Characteristics of Sentinel-2 MSI data. Source: Ref. 64.

| Bands names | Pixel size | Wavelength | Description |
|---|---|---|---|
| B1 | 60 m | 443.9 nm (S2A)/442.3 nm (S2B) | Aerosols |
| B2 | 10 m | 496.6 nm (S2A)/492.1 nm (S2B) | Blue |
| B3 | 10 m | 560 nm (S2A)/559 nm (S2B) | Green |
| B4 | 10 m | 664.5 nm (S2A)/665 nm (S2B) | Red |
| B5 | 20 m | 703.9 nm (S2A)/703.8 nm (S2B) | Red edge 1 |
| B6 | 20 m | 740.2 nm (S2A)/739.1 nm (S2B) | Red edge 2 |
| B7 | 20 m | 782.5 nm (S2A)/779.7 nm (S2B) | Red edge 3 |
| B8 | 10 m | 835.1 nm (S2A)/833 nm (S2B) | Near infrared (NIR) |
| B8A | 20 m | 864.8 nm (S2A)/864 nm (S2B) | Red edge 4 |
| B9 | 60 m | 945 nm (S2A)/943.2 nm (S2B) | Water vapor |
| B11 | 20 m | 1613.7 nm (S2A)/1610.4 nm (S2B) | Short wavelength infrared 1 (SWIR 1) |
| B12 | 20 m | 2202.4 nm (S2A)/2185.7 nm (S2B) | Short wavelength infrared 2 (SWIR 2) |

wavelength ranges, environmental variables, weather and climate data, land cover, and topographic and socio-economic data. All this data is preprocessed into a ready-to-use one.[20] GEE can be used to process more than 40 years of global-scale satellite imagery.

### 2.1.3 *Data*

Sentinel-2 MSI is high-resolution optical data. Sentinel-2 covers 13 bands ranging from 443 to 2190 nm and offers a spatial resolution of 10, 20, and 60 m. Characteristics of Sentinel-2 MSI Level-2A data that were used are shown in Table 1.

The Sentinel-2 instrument acquires measurements at 12 bits. These measurements are converted to reflectances and stored as 16-bit integers in the S2 product. The revisit frequency of every single Sentinel-2 satellite is 10 days and the combined constellation revisit is 5 days.[66]

Sentinel-2 MSI data in GEE is available for two levels of processing. Its Level-1C orthorectified Top-Of-Atmosphere reflectance (available since March 23, 2015) and Level-2A orthorectified atmospherically corrected surface reflectance (available since March 28, 2017). The authors used Level-2A data, which are already atmospherically corrected. This means that their pixels have surface reflectance values. This data is in the collection GEE ee. ImageCollection('COPERNICUS/S2_SR').

The Sentinel-2 Level-2A data were computed by running sen2cor.[64] An atmospheric correction was applied to Top-Of-Atmosphere Level-1C orthoimage products. Level-2A's main output is an orthoimage bottom-of-atmosphere corrected reflectance product, i.e., the surface reflectance values.[66] The Sen2Cor processor algorithm is a combination of state-of-the-art techniques for performing atmospheric corrections, which have been tailored to the Sentinel-2 environment together.[67] Thus, the Level-2 data contain 12 unsigned integer 16-bits spectral bands representing surface reflectance values scaled by 10,000 (unlike in Level-1C data, there is no band 10).

## 2.2 *Methods*

### 2.2.1 *Preprocessing of Sentinel-2 MSI data in GEE*

GEE was used to download Sentinel-2 MSI data. Data in GEE are organized into collections. We used ee.ImageCollection ("COPERNICUS/S2_SR"). According to recommendations from,[68] for areas where the vegetation changes during the seasons, it is necessary to use images obtained

**Table 2** Dependence of the number of scenes on the percentage of cloud cover. Source: an original study in GEE.

| | Number of scenes Sentinel-2 MSI Level-2A | | |
|---|---|---|---|
| Cloudy pixel percentage | 01 Apr to 30 Sept. 2020 | 01 Apr to 30 Sept. 2021 | 01 Apr to 30 June 2022 |
| ≤1% | 38 | 22 | 1 |
| ≤5% | 68 | 36 | 16 |
| ≤10% | 89 | 50 | 23 |
| ≤20% | 117 | 82 | 31 |
| ≤30% | 136 | 100 | 40 |

within the growing season. Therefore, in our article, we used Sentinel-2 MSI data for the study period from the beginning of April to the end of September 2020 to 2022 (until June 30, 2022). The next step in GEE was to perform cloud masking since the pixels under the cloud cover do not carry useful information about the land cover, and their use in data processing will lead to deliberately erroneous results.

Level-2A data and several other specific bands are not photographic images. Among these, three quality assessment (QA) bands are present, one of which (QA60) is a bit band with cloud mask information, which we used to perform cloud masking. To obtain more accurate results when performing cloud masking using the QA60 band, the image collection was first filtered by the cloudy pixel percentage value, to use only a relatively cloud-free granule. We investigated the dependence of the number of images in the collection within the growing seasons of 3 years on a series of percentage values of cloudy pixels. The results are given in Table 2.

Based on the data in Table 2, it was decided to use image pellets in our study, where the percentage of cloudiness does not exceed 10%. Thus, we have a collection of 162 images, to which the cloud masking algorithm was applied based on the QA60 band. Also, the original pixel values were multiplied by $10^{-4}$ to scale surface reflectance 0...1.

In the next step, a mosaic image was created in GEE from the data collection using the median function for pixel values. The resulting mosaic of the mean spectral reflectance values was cropped in GEE to the area of the polygonal feature that matches the Kyiv city, downloaded from the humanitarian data exchange, and imported into GEE.[69] The result is shown in Fig. 2.

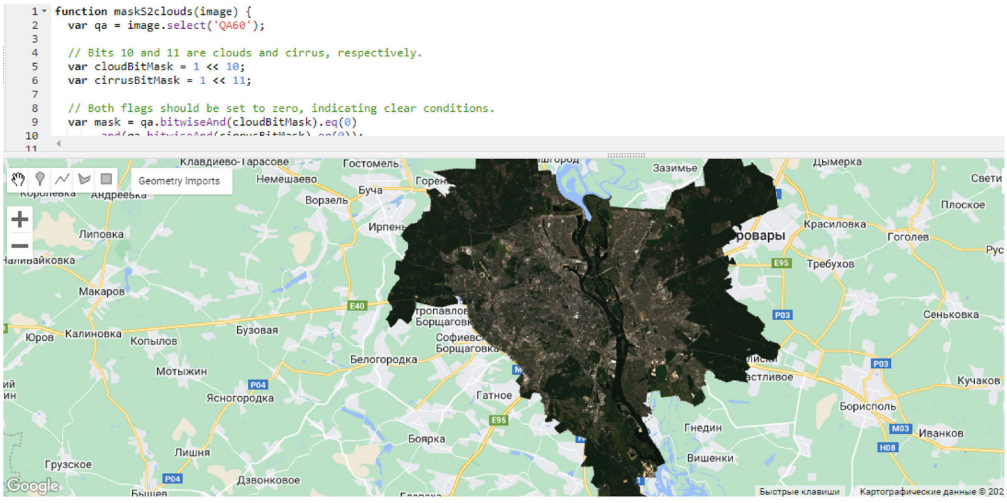


**Fig. 2** Sentinel-2 MSI satellite data mosaic of the mean spectral reflectance values plotted in GEE for the months from April to September 2020 to 2022 for the territory of Kyiv city (True Color Composite R-G-B = 4-3-2). Source: an original study in GEE.

### 2.2.2 *Machine learning algorithms*

Different classification methods are widely used for processing remote sensing data. Among them are commonly used algorithms for unsupervised classification (cluster analysis), probabilistic supervised classification algorithms, object-based methods, and machine learning algorithms. According to Ref. 29, the RF algorithm is generally immune to data noise and overfitting and is extremely useful for the classification of remotely sensed data. Also, the RF algorithm can successfully process high-dimensional data and typically provide higher accuracy than other approaches, such as maximum likelihood methods, single decision trees, and single-layer neural networks.[26,30,70] At the same time, the RF classifier is more reliable than a single decision tree.[25] RF algorithm is a machine learning technique where RF classifiers build multiple decorrelated random decision trees that are loaded and aggregated to classify a dataset using a predictive model from all decision trees.[27]

SVM is a type of supervised learning algorithm, which is used to solve regression and classification problems.[14] SVM classifiers create an ideal hyperplane during the training phase that separates the few classes with the fewest misclassified pixels. SVM is used to select endpoints/vectors that will help create a hyperplane. These endpoints are called support vectors.[71] Based on the above advantages of machine learning algorithms, in this work, we used the RF and SVM algorithms implemented in GEE to extract LULC.

### 2.2.3 *Application of vegetation indices*

Carrying out classification using machine learning algorithms, as well as using any other supervised classification methods, involves the creation of training spectral signatures. It was decided to use VI as the source of additional information about LULC when creating training spectral signatures. The practice of using VI to identify LULC types is widely known. At the same time, with the help of VI, both purposefully individual cover types and all those available in the study area can be determined. It is being implemented using the VI by highlighting certain surface properties that are not visible in the original image bands.[72] In Ref. 36, based on NDVI for satellite data $1 \times 1$ 11 cover types were identified for the continents using the maximum likelihood supervised classification method. The authors[73] studied the influence of the presence/absence of various spectral bands on the accuracy of water extraction, the combination of which is the VI. The second role of VIs, which was envisaged in our work, is their use as additional channels when performing classification using machine learning algorithms to improve the classification accuracy of Sentinel-2 data.

The similarity of the reflectance spectra of surfaces that refer to artificially created objects and surfaces of other hard-surfaced objects, such as bare earth and sand, is a known problem that occurs when extracting LULC using classification methods.[45] It is connected with the wide variety of materials from which built-up areas are created, such as concrete, asphalt, metal, plastic, and glass.[46,55] Dry soil and light impenetrable surfaces are similar in spectral characteristics, and the shadows from tall buildings and tree crowns are similar to the spectra of dark impenetrable surfaces.[47]

Thus, to solve the above two questions, we used VIs that had positive recommendations from other authors in terms of distinguishing different types of LULCs and, especially, the possibility of separating the reflectance spectra of built-up and bare soils. Considering the results obtained in Refs. 45, 49, and 74, we decided to use such VIs as NDVI, NDSI, IBI, BCI, BAEI, UI, NBI, BRBA, BSI, and NBAI. Normalized difference vegetation index (NDVI) is calculated by the equation

$$\mathrm{NDVI} = \frac{\rho_{\mathrm{NIR}} - \rho_{\mathrm{Red}}}{\rho_{\mathrm{NIR}} + \rho_{\mathrm{Red}}} = \frac{\mathrm{Band\,8} - \mathrm{Band\,4}}{\mathrm{Band\,8} + \mathrm{Band\,4}}, \tag{1}$$

where $\rho_{\mathrm{NIR}}$ is the near-infrared wavelength reflectance value; $\rho_{\mathrm{Red}}$ is the red wavelength reflectance value (according to Table 1).

Here and further band in the equations indicate the band numbers in Sentinel-2 MSI data.

NDSI is calculated by the equation [37]

$$\mathrm{NDSI} = \frac{\rho_{\mathrm{SWIR\,2}} - \rho_{\mathrm{Blue}}}{\rho_{\mathrm{SWIR\,2}} + \rho_{\mathrm{Blue}}} = \frac{\mathrm{Band\,12} - \mathrm{Band\,2}}{\mathrm{Band\,12} + \mathrm{Band\,2}}, \tag{2}$$

where $\rho_{\mathrm{Blue}}$ is the blue wavelength reflectance value; $\rho_{\mathrm{SWIR\,2}}$ is the SWIR 2 wavelength reflectance value.

IBI is calculated by the equation

$$\begin{aligned}\mathrm{IBI} &= \frac{\frac{2\cdot\rho_{\mathrm{SWIR1}}}{\rho_{\mathrm{SWIR1}}+\rho_{\mathrm{NIR}}} - \left(\frac{\rho_{\mathrm{NIR}}}{\rho_{\mathrm{NIR}}+\rho_{\mathrm{Red}}} - \frac{\rho_{\mathrm{Green}}}{\rho_{\mathrm{Green}}+\rho_{\mathrm{SWIR1}}}\right)}{\frac{2\cdot\rho_{\mathrm{SWIR1}}}{\rho_{\mathrm{SWIR1}}+\mathrm{NIR}} + \left(\frac{\rho_{\mathrm{NIR}}}{\rho_{\mathrm{NIR}}+\rho_{\mathrm{Red}}} - \frac{\rho_{\mathrm{Green}}}{\rho_{\mathrm{Green}}+\rho_{\mathrm{SWIR1}}}\right)} \\ &= \frac{\frac{2\cdot\mathrm{Band\,11}}{\mathrm{Band\,11}+\mathrm{Band\,8}} - \left(\frac{\mathrm{Band\,8}}{\mathrm{Band\,8}+\mathrm{Band\,4}} - \frac{\mathrm{Band\,3}}{\mathrm{Band\,3}+\mathrm{Band\,11}}\right)}{\frac{2\cdot\mathrm{Band\,11}}{\mathrm{Band\,11}+\mathrm{Band\,8}} + \left(\frac{\mathrm{Band\,8}}{\mathrm{Band\,8}+\mathrm{Band\,4}} - \frac{\mathrm{Band\,3}}{\mathrm{Band\,3}+\mathrm{Band\,11}}\right)},\end{aligned} \tag{3}$$

where $\rho_{\mathrm{Green}}$ is the green wavelength reflectance value; $\rho_{\mathrm{SWIR1}}$ is SWIR 1 wavelength reflectance value.

BCI is calculated by the equation[48]

$$\mathrm{BCI} = \frac{\frac{\mathrm{TC\,1}+\mathrm{TC\,3}}{2} - \mathrm{TC\,2}}{\frac{\mathrm{TC\,1}+\mathrm{TC\,3}}{2} + \mathrm{TC\,2}}, \tag{4}$$

where TC1, TC2, and TC3 are the Tasseled Cap components for Sentinel-2 MSI images.

The Tasseled Cap transformation coefficients for Sentinel-2 MSI data that we used in our work are given in Table 3.

BAEI is calculated by the equation[44]

$$\mathrm{BAEI} = \frac{\rho_{\mathrm{Red}} + 0.3}{\rho_{\mathrm{Green}} + \rho_{\mathrm{SWIR\,11}}} = \frac{\mathrm{Band\,4} + 0.3}{\mathrm{Band\,3} + \mathrm{Band\,11}}. \tag{5}$$

UI is calculated by the equation[38]

$$\mathrm{UI} = \frac{\rho_{\mathrm{SWIR2}} - \rho_{\mathrm{NIR}}}{\rho_{\mathrm{SWIR2}} + \rho_{\mathrm{NIR}}} = \frac{\mathrm{Band\,12} - \mathrm{Band\,8}}{\mathrm{Band\,12} + \mathrm{Band\,8}}. \tag{6}$$

**Table 3** Tasseled Cap transformation coefficients for Sentinel-2 MSI data. Source: Ref. 75.

| | Sentinel-2 MSI Bands | | | | | |
|---|---|---|---|---|---|---|
| Components | B1 | B2 | B3 | B4 | B5 | B6 |
| Brightness | 0.0356 | 0.0822 | 0.136 | 0.2611 | 0.2964 | 0.3338 |
| Greenness | −0.0635 | −0.1128 | −0.168 | −0.348 | −0.3303 | 0.0852 |
| Wetness | 0.0649 | 0.1363 | 0.2802 | 0.3072 | 0.5288 | 0.1379 |
| Brightness | B7 | B8 | B8A | B9 | B11 | B12 |
| Greenness | 0.3877 | 0.3895 | 0.475 | 0.0949 | 0.3882 | 0.1366 |
| Wetness | 0.3302 | 0.3165 | 0.3625 | 0.0467 | −0.4578 | −0.4064 |

NBI is calculated by the equation[43]

$$\mathrm{NBI} = \frac{\rho_{\mathrm{Green}} \times \rho_{\mathrm{NIR}}}{\rho_{\mathrm{Red}}} = \frac{\mathrm{Band\,3} \times \mathrm{Band\,8}}{\mathrm{Band\,4}}. \tag{7}$$

BRBA is calculated by the equation[16]

$$\mathrm{BRBA} = \frac{\rho_{\mathrm{Red}}}{\rho_{\mathrm{NIR}}} = \frac{\mathrm{Band\,3}}{\mathrm{Band\,8}}. \tag{8}$$

BSI is calculated by the equation[76]

$$\begin{aligned}\mathrm{BSI} &= \frac{(\rho_{\mathrm{Red}} + \rho_{\mathrm{Green}}) - (\rho_{\mathrm{Red}} + \rho_{\mathrm{Blue}})}{(\rho_{\mathrm{NIR}} + \rho_{\mathrm{Green}}) + (\rho_{\mathrm{Red}} + \rho_{\mathrm{Blue}})} 100 + 100 \\ &= \frac{(\mathrm{Band\,4} + \mathrm{Band\,3}) - (\mathrm{Band\,4} + \mathrm{Band\,2})}{(\mathrm{Band\,8} + \mathrm{Band\,3}) + (\mathrm{Band\,4} + \mathrm{Band\,2})} 100 + 100.\end{aligned} \tag{9}$$

NBAI is calculated by the equation[17]

$$\mathrm{NBAI} = \frac{\rho_{\mathrm{SWIR\,2}} - \frac{\rho_{\mathrm{SWIR\,1}}}{\rho_{\mathrm{Green}}}}{\rho_{\mathrm{SWIR\,2}} + \frac{\rho_{\mathrm{SWIR\,1}}}{\rho_{\mathrm{Green}}}} = \frac{\mathrm{Band\,12} - \frac{\mathrm{Band\,11}}{\mathrm{Band\,3}}}{\mathrm{Band\,12} + \frac{\mathrm{Band\,11}}{\mathrm{Band\,3}}}. \tag{10}$$

## 2.3 *Performing Sentinel-2 MSI Data Classification Based on Machine Learning Algorithms on GEE*

We went directly to performing classification using machine learning algorithms on GEE after creating a cloudless Sentinel-2 MSI data Level-2A mosaic image in the mean spectral reflectance values for April to September 2020 to 2021 and April to June 2022 for the territory of Kyiv. Studying the effectiveness of various machine learning algorithms for LULC extraction is one of the objectives of our classification. Another, more significant task is to improve approaches to the definition of LULC.

Based on the results of the classification for solving the first problem, we performed an accuracy assessment and compared the results obtained. And to solve the second problem, we used four combinations of channels as source images for the classification of machine learning algorithms on GEE, as shown in Table 4.

As additional bands, the VIs shown in Table 4 have been attached to the GEE to create combinations two to four to the original 12 bands Sentinel-2 MSI given in Table 1.

It was necessary to prepare training samples or signatures (usually called region of interest, ROI, or area of interest, AOI) at the beginning of the classification process. Samples were created for our study area following landscape elements:

Class 1 – Water (may include rivers, lakes, reservoirs, and other hydrographic objects);

Class 2 – Forest;

**Table 4** Band combinations to perform classification. Source: original study.

| Combination number | Bands used |
|---|---|
| Combination 1 | Only 12 bands of Sentinel-2 MSI given in Table 1 |
| Combination 2 | 12 bands of Sentinel-2 and VI: UI, NDSI, IBI, BCI, BAEI |
| Combination 3 | 12 bands of Sentinel-2 and VI: NDVI, NBI, BRBA, NBAI, BSI |
| Combination 4 | 12 bands of Sentinel-2 and all VI: UI, NDSI, IBI, BCI, BAEI, NDVI, NBI, BRBA, NBAI, BSI |

Class 3 – Grassland;

Class 4 – Built-up (may include buildings, road, and other human-created objects);

Class 5 – Other lands (bare soil, sand).

The process of creating ROIs in GEE is to digitize image pixels that correspond to the class for which the ROI is being created. Digitization in GEE can be done with dots, lines, and polygons. The GEE algorithm converts each image pixel that is included in the ROI into a separate ROI element if lines or polygons are used. In this case, the ROI pixels should be as uniform as possible so as not to make further mistakes in the classification process. Each class must have at least 50 training samples for classification, according to the recommendations.[14,77] Therefore, we created a total of 739 training samples, including polygonal samples (Table 5). The number of training samples for each LULC class roughly corresponded to the representation of this class among all classes for the study area. We also distributed ROIs evenly across the territory of Kyiv.

The number of decision trees to create is a required input parameter for RF work on GEE. We used 100 as recommended.[14] The important input parameters of the SVM algorithm are[64]

- kernelType – one of LINEAR, POLY, radial basis function (RBF), or SIGMOID;
- gamma – the gamma value in the kernel function. Defaults to the reciprocal of the number of features. Valid for POLY, RBF, and SIGMOID kernels;
- cost – the cost (C) parameter. Defaults to 1. Only valid for C-SVC, epsilon-SVR, and nu-SVR.

As recommended,[71] we used the RBF kernelType, gamma = 0.5, cost = 10.

We proceeded to perform the classification using machine learning algorithms RF and SVM after all the necessary input data and parameters were prepared.

## 2.4 *Accuracy Assessment of Classification Results*

After the classification is completed, it is necessary to evaluate its accuracy. To do this, we created additional samples: 508 points, evenly distributed over the image. Also, the forest and grassland classes were merged into a common vegetation class before testing. Thus, was created classes such as

Class 1 – Water;

Class 2 –Vegetation;

Class 3 – Built-up;

Class 4 – Other lands.

**Table 5** Number of training samples. Source: original study.

| | Training samples | | |
|---|---|---|---|
| | | Number | |
| Class name | Total | Points | Polygons |
| Water | 161 | 49 | 112 |
| Forest | 135 | 88 | 47 |
| Grassland | 197 | 143 | 54 |
| Built-up | 164 | 121 | 43 |
| Other lands | 82 | 82 | — |
| Total | 739 | 483 | 256 |

To create the above-mentioned additional samples, we performed field interpretation of those parts of the image that caused doubts. We also used historical Google Earth Pro images and topographic maps.

To perform the accuracy assessment of the obtained LULC maps in the GEE, we compiled a confusion matrix, according to which the most commonly used descriptive statistics for assessing the classification were calculated: OA, producer's accuracy (PA) user's accuracy (UA) and the Kappa coefficient (Kappa hat). Formulas for calculating them can be found, for example, in Ref. 78. The results are presented in Tables 6–13.

The classification results using the RF algorithm are shown in Fig. 3.

The classification results using the SVM algorithm are shown in Fig. 4.

Recent images in various band combinations for comparison with the original Sentinel-2 images were shown in Fig. 5. The final LULC maps were designed in ArcGIS. It is important to analyze each class on LULC maps.

Analyzing Fig. 3, combinations 3 and 4 misclassify built-up and other lands as water (combinations 1 and 2 did not have such errors). Combinations 1 and 2 had more errors in the classification of water as vegetation than combinations 3 and 4. All combinations had no errors

**Table 6** Confusion matrix for RF LULC classification for combination 1. Source: an original study in GEE.

| | | Reference data | | | | | |
|---|---|---|---|---|---|---|---|
| RF combination 1 | | Water | Vegetation | Built-up | Other lands | Total | UA (%) |
| Map data | Water | 84 | 4 | 0 | 0 | 88 | 95.45 |
| | Vegetation | 0 | 112 | 5 | 13 | 130 | 86.15 |
| | Built-up | 2 | 6 | 145 | 33 | 186 | 77.96 |
| | Other lands | 0 | 6 | 3 | 104 | 113 | 92.04 |
| | Total | 86 | 128 | 153 | 150 | 517 | |
| | PA (%) | 97.67 | 87.5 | 94.77 | 69.33 | | |
| | OA (%) | 86.07% | | | | | |
| | Kappa | 0.81 | | | | | |

**Table 7** Confusion matrix for RF LULC classification for combination 2. Source: an original study in GEE.

| | | Reference data | | | | | |
|---|---|---|---|---|---|---|---|
| RF combination 2 | | Water | Vegetation | Built-up | Other lands | Total | UA (%) |
| Map data | Water | 84 | 4 | 0 | 0 | 88 | 95.45 |
| | Vegetation | 0 | 116 | 4 | 10 | 130 | 89.23 |
| | Built-up | 2 | 2 | 154 | 28 | 186 | 82.8 |
| | Other lands | 0 | 6 | 6 | 101 | 113 | 89.38 |
| | Total | 86 | 128 | 164 | 139 | 517 | |
| | PA (%) | 97.67 | 90.63 | 93.9 | 72.66 | | |
| | OA (%) | 88.01% | | | | | |
| | Kappa | 0.84 | | | | | |

**Table 8** Confusion matrix for RF LULC classification for combination 3. Source: an original study in GEE.

| | | Reference data | | | | | |
|---|---|---|---|---|---|---|---|
| RF combination 3 | | Water | Vegetation | Built-up | Other lands | Total | UA (%) |
| Map data | Water | 80 | 0 | 6 | 2 | 88 | 90.91 |
| | Vegetation | 0 | 114 | 1 | 15 | 130 | 87.69 |
| | Built-up | 2 | 3 | 144 | 37 | 186 | 77.42 |
| | Other lands | 0 | 3 | 6 | 104 | 113 | 92.04 |
| | Total | 82 | 120 | 157 | 158 | 517 | |
| | PA (%) | 97.56 | 95.00 | 91.72 | 65.82 | | |
| | OA (%) | | | 85.49% | | | |
| | Kappa | | | 0.80 | | | |

**Table 9** Confusion matrix for RF LULC classification for combination 4. Source: an original study in GEE.

| | | Reference data | | | | | |
|---|---|---|---|---|---|---|---|
| RF combination 4 | | Water | Vegetation | Built-up | Other lands | Total | UA (%) |
| Map data | Water | 81 | 1 | 5 | 1 | 88 | 92.05 |
| | Vegetation | 0 | 116 | 3 | 11 | 130 | 89.23 |
| | Built-up | 3 | 2 | 146 | 35 | 186 | 78.49 |
| | Other lands | 0 | 4 | 4 | 105 | 113 | 92.92 |
| | Total | 84 | 123 | 158 | 152 | 517 | |
| | PA (%) | 96.43 | 94.31 | 92.41 | 69.08 | | |
| | OA (%) | | | 86.65% | | | |
| | Kappa | | | 0.82 | | | |

**Table 10** Confusion matrix for SVM LULC classification for combination 1. Source: an original study in GEE.

| | | Reference data | | | | | |
|---|---|---|---|---|---|---|---|
| SVM combination 1 | | Water | Vegetation | Built-up | Other lands | Total | UA (%) |
| Map data | Water | 81 | 5 | 2 | 0 | 88 | 92.05 |
| | Vegetation | 0 | 116 | 1 | 13 | 130 | 89.23 |
| | Built-up | 2 | 11 | 129 | 44 | 186 | 69.35 |
| | Other lands | 0 | 5 | 2 | 106 | 113 | 93.81 |
| | Total | 83 | 137 | 134 | 163 | 517 | |
| | PA (%) | 97.59 | 84.67 | 96.27 | 65.03 | | |
| | OA (%) | | | 83.56% | | | |
| | Kappa | | | 0.78 | | | |

**Table 11** Confusion matrix for SVM LULC classification for combination 2. Source: an original study in GEE.

| | | Reference data | | | | | |
|---|---|---|---|---|---|---|---|
| SVM combination 2 | | Water | Vegetation | Built-up | Other lands | Total | UA (%) |
| Map data | Water | 81 | 4 | 3 | 0 | 88 | 92.05 |
| | Vegetation | 4 | 114 | 1 | 11 | 130 | 87.69 |
| | Built-up | 3 | 9 | 140 | 34 | 186 | 75.27 |
| | Other lands | 0 | 9 | 1 | 103 | 113 | 91.15 |
| | Total | 88 | 136 | 145 | 148 | 517 | |
| | PA (%) | 92.05 | 83.82 | 96.55 | 69.59 | | |
| | OA (%) | | | 84.72% | | | |
| | Kappa | | | 0.79 | | | |

**Table 12** Confusion matrix for SVM LULC classification for combination 3. Source: an original study in GEE.

| | | Reference data | | | | | |
|---|---|---|---|---|---|---|---|
| SVM combination 3 | | Water | Vegetation | Built-up | Other lands | Total | UA (%) |
| Map data | Water | 84 | 1 | 1 | 2 | 88 | 95.45 |
| | Vegetation | 2 | 113 | 0 | 15 | 130 | 86.92 |
| | Built-up | 3 | 3 | 131 | 49 | 186 | 70.43 |
| | Other lands | 0 | 4 | 2 | 107 | 113 | 94.69 |
| | Total | 89 | 121 | 134 | 173 | 517 | |
| | PA (%) | 94.38 | 93.39 | 97.76 | 61.85 | | |
| | OA (%) | | | 84.14% | | | |
| | Kappa | | | 0.79 | | | |

**Table 13** Confusion matrix for SVM LULC classification for combination 4. Source: an original study in GEE.

| | | Reference data | | | | | |
|---|---|---|---|---|---|---|---|
| SVM combination 4 | | Water | Vegetation | Built-up | Other lands | Total | UA (%) |
| Map data | Water | 84 | 1 | 3 | 0 | 88 | 95.45 |
| | Vegetation | 9 | 109 | 1 | 11 | 130 | 83.85 |
| | Built-up | 7 | 7 | 148 | 24 | 186 | 79.57 |
| | Other lands | 2 | 7 | 3 | 101 | 113 | 89.38 |
| | Total | 102 | 124 | 155 | 136 | 517 | |
| | PA (%) | 82.35 | 87.9 | 95.48 | 74.26 | | |
| | OA (%) | | | 85.49% | | | |
| | Kappa | | | 0.80 | | | |

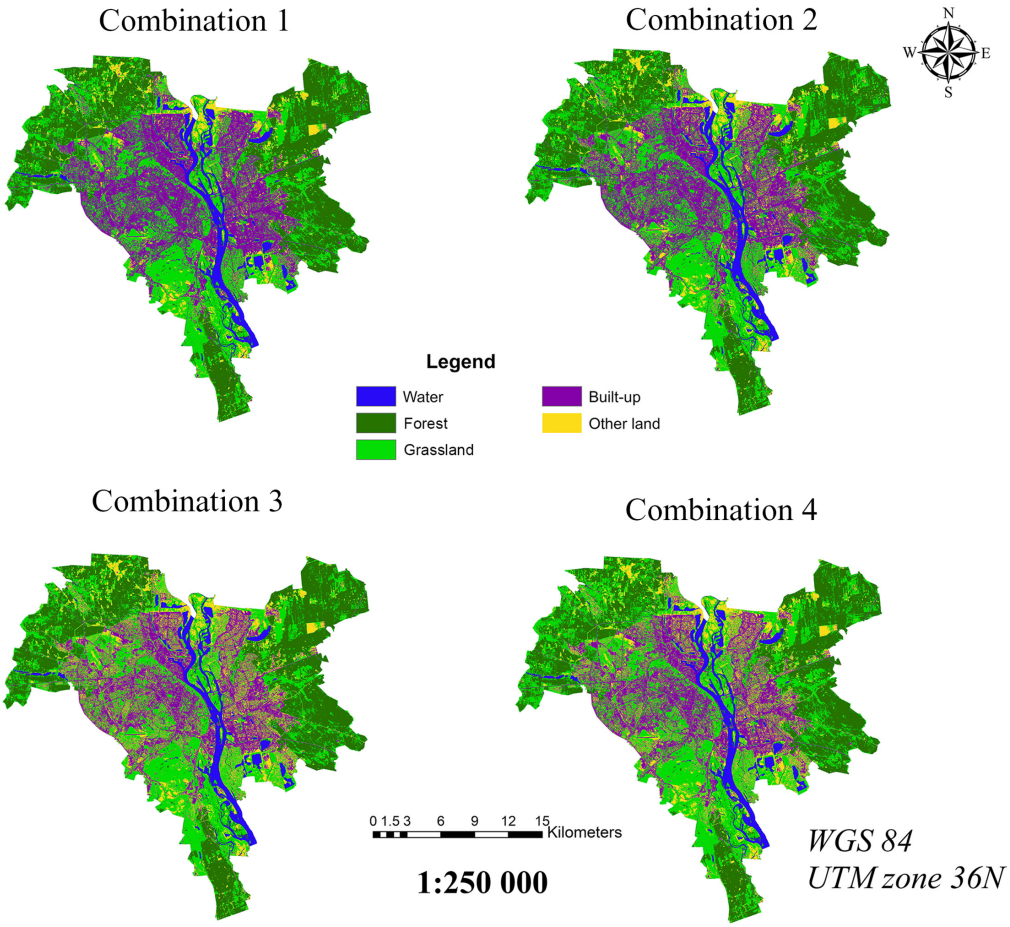


**Fig. 3** LULC maps of Sentinel-2 MSI images using RF and four combinations of 12 bands and VI for Kyiv city. Source: an original study in GEE.

in the classification of vegetation as water. All combinations had small errors in the classification of built-up and other lands as vegetation (the lowest number of built-up samples misclassified as the vegetation was in combination 3). All combinations had errors in the classification of water, vegetation, and other lands as built-up. At the same time, the largest number of samples, which are actually vegetation, which was erroneously classified as built-up, belongs to combination 1 (for other classes, the number is approximately equal). All combinations had no errors in the classification of other lands as water but had errors in the classification of vegetation and built-up as other lands. The smallest number of samples of other lands erroneously classified as built-up belong to combination 2, and the largest number belong to combination 3.

Analyzing Fig. 4, combinations 1, 2, and 4 had no errors in the classification of water as other lands. Combination 3 includes a minimum number of vegetation and built-up samples that were erroneously classified as water. Only combination 1 had no errors in the classification of water as vegetation. All combinations had few erroneous samples in the built-up as vegetation classification and all combinations had many other lands samples erroneously classified as vegetation. Combination 4 showed the most errors in the classification of water as built-up. Only combination 4 had errors in the classification of water as other lands. At the same time, it had a significantly smaller number of samples built-up erroneously classified as other lands. In addition, combination 3 showed the greatest number of such errors. The number of samples of other lands erroneously classified as built-up was approximately equal for all combinations.

The mean OA for RF and SVM was 86.56% and 84.48%, respectively. The average Kappa value was 0.82 and 0.79, which indicates that the RF algorithm showed a higher result than SVM (Table 14).

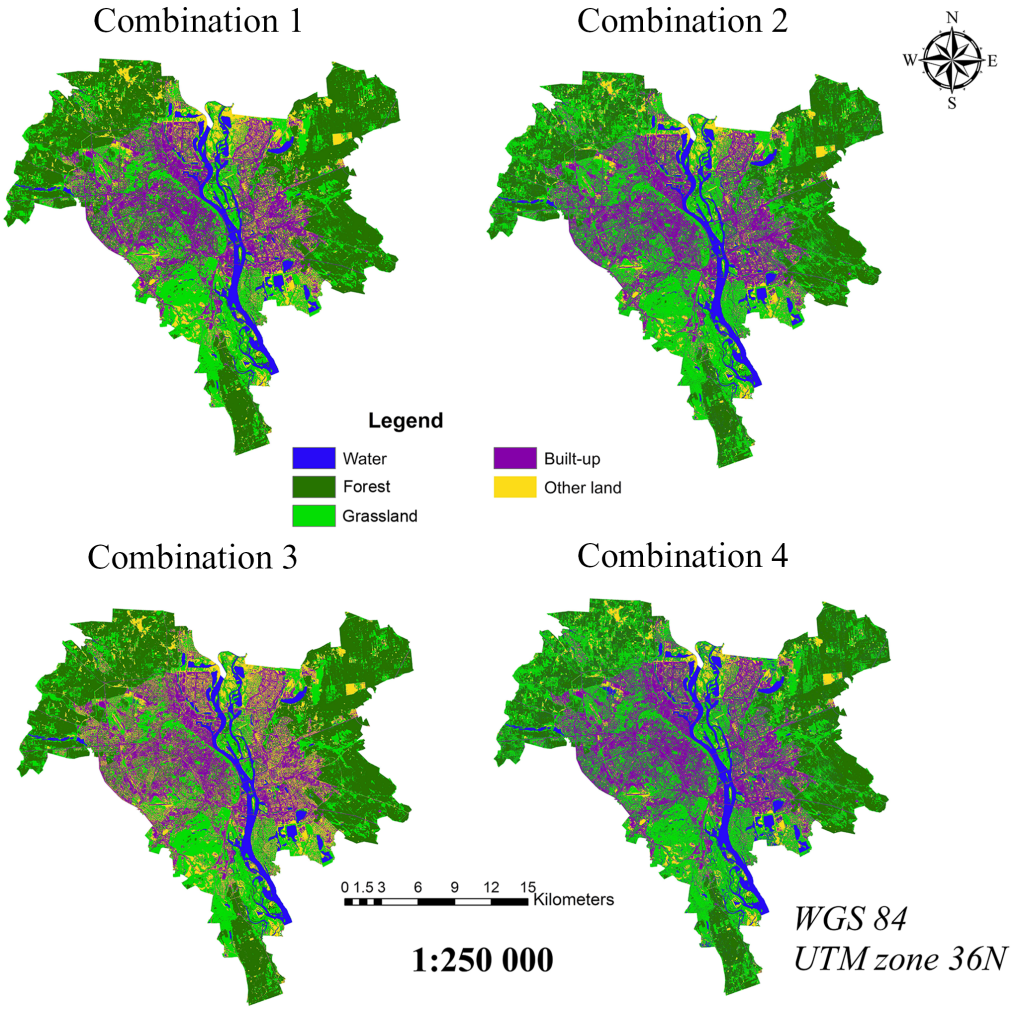


**Fig. 4** LULC maps of Sentinel-2 MSI images using SVM and four combinations of 12 bands and VI for Kyiv city. Source: an original study in GEE.

In Table 14, combination 2 has the best performance, and combination 4 differs quite a bit from it. User's and producer's accuracy for each type of the LULC, obtained through RF and SVM algorithms for each combination shown in Figs. 6 and 7.

In Fig. 6, both algorithms (RF and SVM) class water has the best UA, and built-up has the lowest UA. At the same time, if we consider all combinations, then within each class UA for each combination turned out to be different. It is possible to use one or the other combination, depending on what types of LULC need to highlight. For example, if the research is aimed at isolating a certain type of LULC. In Fig. 7, RF algorithm class water and vegetation have the best PA, and the SVM algorithm is the water and built-up. The lowest results for both algorithms (RF and SVM) were obtained for other lands.

## 2.5 *Discussion*

The band combinations proposed in this paper solve the additional task of separating LULC types with similar spectral reflectance properties, i.e., built-up and other lands (bare soils/sand). According to the data in Tables 6–14, combinations 2 and 4 (Table 4) most accurately classify built-up and other lands and have the highest OA for RF is 88.01%, 86.65%, and for SVM −84.72%, 85.49%, respectively. In turn, combinations 1 and 3 (Table 4) show worse results of classification of built-up and other lands (Tables 6–14) and have lower OA for RF is 86.07%, 85.49% and for SVM 83.56%, 84.14%, respectively. In our opinion, these results

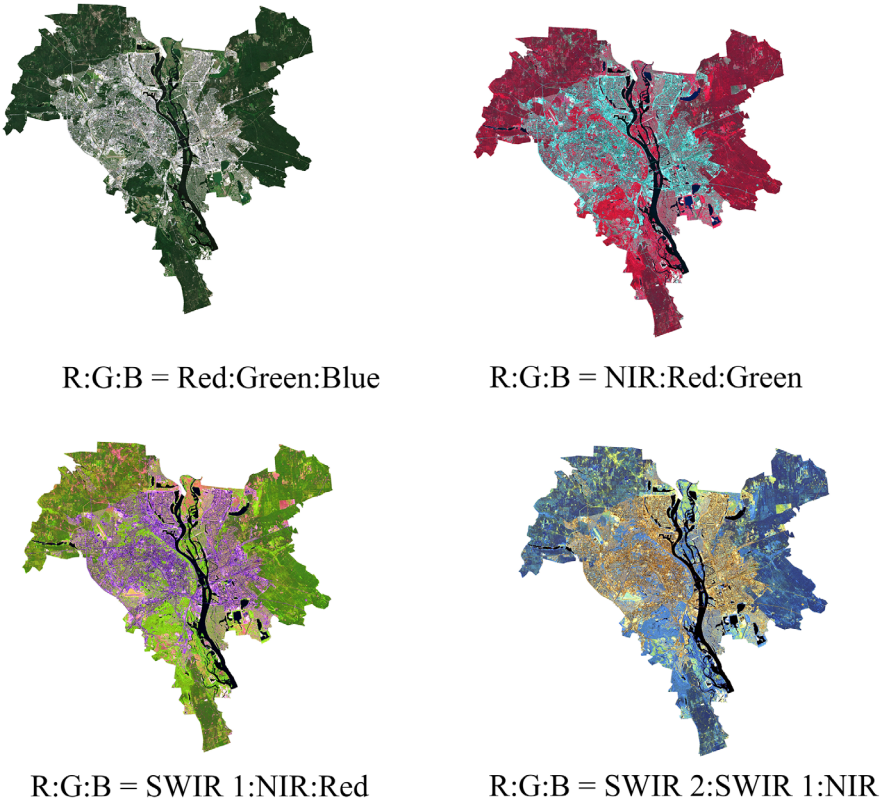


**Fig. 5** Composite Sentinel-2 MSI images. Source: an original study in GEE.

**Table 14** Average values of OA and Kappa for all combinations. Source: original study

| Average statistic values | Combinations | | | |
|---|---|---|---|---|
| | Comb. 1 | Comb. 2 | Comb. 3 | Comb. 4 |
| OA (%) | 84.82% | 86.36% | 84.82% | 86.07% |
| Kappa | 0.80 | 0.82 | 0.80 | 0.81 |

indicate the low efficiency of some VI included in classification 3 and, respectively, included in combination 4, which may reduce its accuracy (Table 4).

Although, according to Ref. 39, the combination of bands Sentinel 2 data with BRBA,[16] NBAI, and BSI[76] indices allows accurate classification of built-up area and bare soil. This does not exactly correspond to our research since these indices are included in combinations 3 and 4, which showed lower accuracy. Note, that the results of the authors[17] showed that the use of the NBAI index reduces the classification accuracy for artificial surfaces. The authors recommend using the perpendicular impervious surface index (PISI) for more accurate separating of artificial surfaces (built-up) from soil (bare land) and vegetation. Similar studies[40,74] where BAEI, BSI, NBAI, NBI, NDBI, and UI indices were used showed good results to improve the classification by SVM for built-up and bare soil. This is consistent with current research, namely inclusion in the combination 2 VI UI and BAEI (Tables 7 and 11). The results of the authors[37] showed that the NDSI allows to classify the built-up area with greater accuracy, which is consistent with our research, where this index was included in combination 2 (Tables 4, 7, 9, and 11).

It is worth noting that in[42,44] the authors successfully apply the BAEI index to classify urban intensity, whereas the NBI index better identifies discrete types of urban land use.

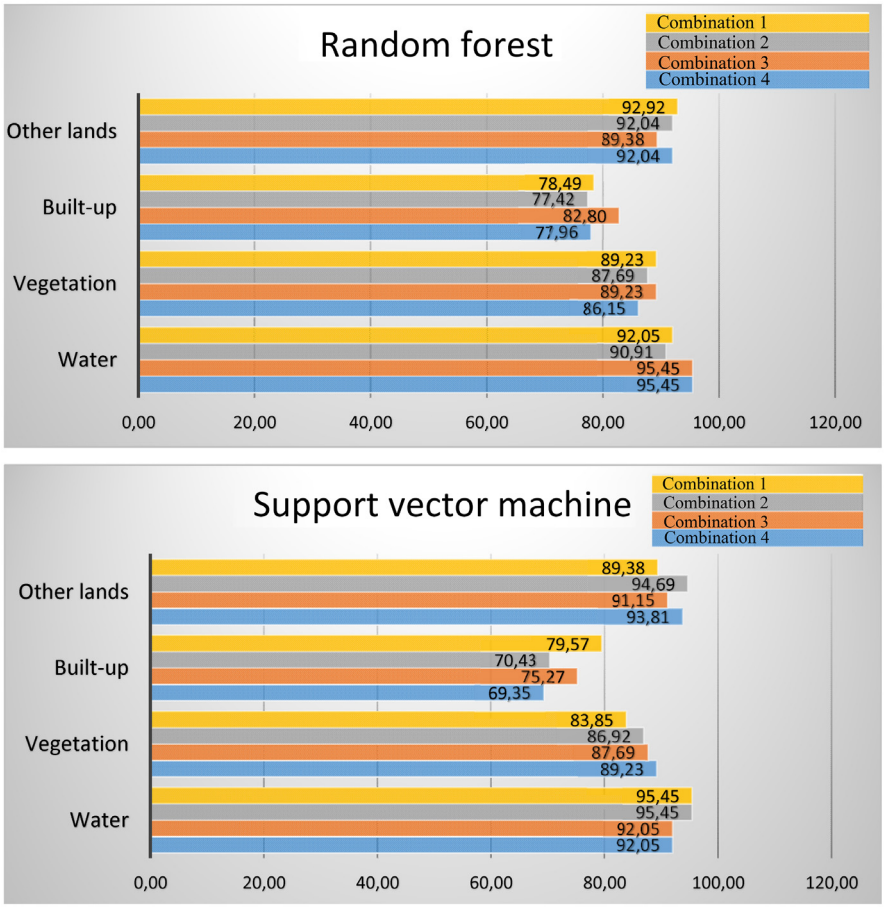


**Fig. 6** User's accuracy for each LULC.

The use of NBI and NDVI indices is successfully used in the analysis of temperature indicators LULC classifications, in particular, urban lands.[43] Application of the IBI, created on the basis of SAVI, MNDWI, and NDBI indices,[41] in our opinion, significantly improves the accuracy of the combination 2 and 4 (Table 4), it immediately identifies three major urban components of vegetation, water, built-up land while effectively suppressing background noise. The authors of Ref. 38 use UI and IBI indices in combination with other indices to separate built-up lands from dry vegetation. According to Ref. 48, it is proved that the vegetation index BCI is effective both for separating impervious surfaces (built-up) and bare soil and for quantifying vegetation abundance when compared with NDVI, which coincides with the results of our work. However, more attention should be paid to the NDVI, as its existence is controversial. According to Ref. 12, LULC classification, including NDVI index and others, based on RF shows more accurate results. While in this study, combination 2 shows better results without NDVI index. The analysis of the combination in Tables 7 and 9 showed that the presence of NDVI reduces the accuracy with RF. However, with SVM there is a slight increase in the accuracy (Tables 11 and 13). Some authors recommend using NDVI mainly for the classification of forest, as an example of forests.[11] For an effective distinction between built-up area and bare land, the authors[49] suggest RF. This classification is conducted using multi-sensor features, including temperature, night-time light, backscattering, topography, optical spectra, and NDVI, EVI, NDWI indices time-series metrics[34] and synthetically incorporating geometry, and texture data.[13,25,50,53]

One of the main objectives of our study was to investigate the performance of different machine learning algorithms RF and SVM for the classification of LULC on the GEE platform. Thus, the RF-based LULC classification is more accurate than SVM according to OA and Kappa (Table 14). Although the OA of RF exceeds the OA of SVM by only 2.08%, which indicates the effectiveness of both machine learning algorithms.

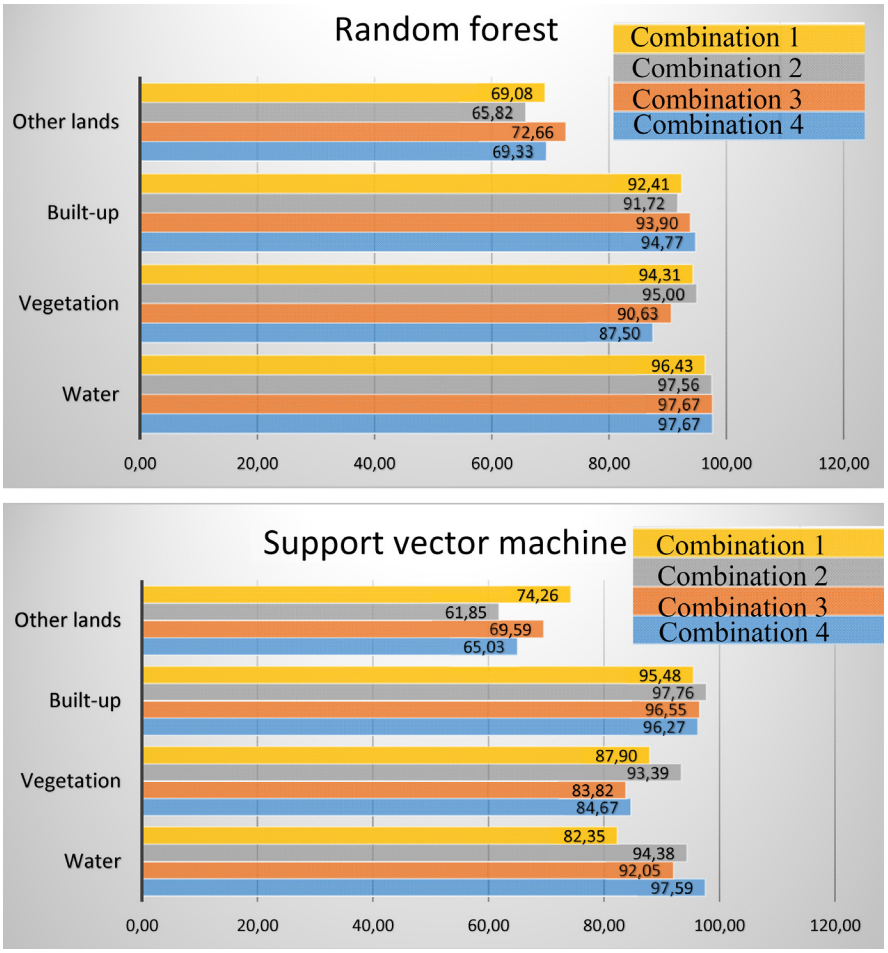


**Fig. 7** Producer's accuracy for each LULC.

Similar results were obtained[14] where for Sentinel-2 and Landsat-8 images, RF outperformed CART and SVM, however, the authors used only spectral bands. Comparing the performance of two ML algorithms RF and SVM applied on Landsat-8, Sentinel-2, and Planet images,[15] the authors note that Sentinel 2 with SVM (OA – 0.96%) was better for classification in rural and urban areas. Similar results were obtained by authors[10] among the classifiers, SVM produced the greatest accuracies of OA – 80.3%, although RF, which had similar accuracies of OA – 80.0%, was simpler to train and apply. While in the present study, accuracy makes a significant difference. Tracked results when among the compared classifiers as ANN, and SVM, RF achieved the lowest OA with 62%, but these researches were obtained for the forest land.[24] For effective cropland classification, the use of the RF algorithm demonstrates the best accuracy.[29] From Fig. 3, which shows LULC maps of Sentinel-2 images using RF and four combinations of 12 bands and VI, it follows from this that the reflectance of built-up and other lands coincides with a reflectance of water and in part with vegetation. Similar results showed[14] that water was slightly misclassified as built-up and barren land. RFs, combine numerous soft linear boundaries at the surface of the decision. SVM performs well if the input training data is sparse, making it a better choice when less data is available.[24] Each algorithm has its own set of benefits and drawbacks. RF is more resilient and less impacted by parameters, whereas SVM is sensitive to hyperparameters.[28] The obtained results indicate the effectiveness of the application for machine learning methods. It is worth noting that the exact classification methodology and the use of the selected combination for classification will depend on the types of LULC that will be detected and the topographic features of the studied area.

## 3 Conclusions

The article studies an approach to improve the accuracy of LULC mapping by RF and SVM machine learning methods based on Sentinel-2 MSI Level-2A data on the GEE platform. This classification used mosaic image, averaged for data for vegetation periods from April, 2020 to September, 2021 and April, 2022 to June, 2022 for the territory of Kyiv (Ukraine). Four combinations were considered to perform classification using RF and SVM learning algorithms. Ten vegetation indices (NDVI, NDSI, IBI, BCI, BAEI, UI, NBI, BRBA, BSI, and NBAI) were included in the combinations, which made it possible to solve the urgent problem of separating LULC classes with similar spectral reflective properties (built-up and other lands). Thus, eight LULC maps were constructed for water, forest, grassland, built-up and other lands. The confusion matrices were compiled and descriptive statistics were calculated to evaluate the classification accuracy. Comparison of the results showed that the type of classifier used and the set of initial bands for them influenced the classification accuracy of LULC. RF algorithm showed higher accuracy than SVM. The mean OA for RF and SVM was 86.56% and 84.48%, respectively, and the mean Kappa was 0.82 and 0.79, respectively. For the RF algorithm, the best LULC classification accuracy was achieved using a combination of 12 bands of Sentinel-2 and UI, NDSI, IBI, BCI, and BAEI (the highest accuracy of all band combinations for both machine learning algorithms), and for the SVM algorithm, using combination of 12 bands of Sentinel-2 and all VI. Combination of 12 bands of Sentinel-2 and NDVI, NBI, BRBA, NBAI, and BSI showed the lowest accuracy for the RF algorithm and combination only 12 bands of Sentinel-2 showed the lowest accuracy for the SVM algorithm. As for the four considered combinations, for the RF algorithm, the best LULC classification accuracy was achieved using combination 2, and for the SVM algorithm, using combination 4. The combination 3 showed the lowest accuracy for the RF algorithm and the combination 1 showed the lowest accuracy for the SVM algorithm.

Important to note the increase in UA and PA to determine built-up and other lands. Thus, the additional use of VI makes it possible to increase the accuracy of creating LULC maps and the accuracy of distinguishing classes with similar reflection spectra.

## Acknowledgments

The author, Liliia Hebryn-Baidy, express gratitude to the Technische Informations bibliothek (TIB), Leibniz Information Centre for Science and Technology, and University Library, for support to this research through the Scholarship to promote academic qualification. **Author Contribution**: Vadym Belenok: conceptualization, methodology, software, data curation, writing – original draft preparation, visualization, project administration. Liliia Hebryn-Baidy: conceptualization, methodology, writing – original draft preparation, writing – review and editing, project administration. Nataliia Bielousova: formal analysis, validation, supervision. Valeriy Gladilin: formal analysis, validation, supervision. Sergíy Kryachok: formal analysis, validation, supervision. Andrii Tereshchenko: formal analysis, validation, supervision. Sofiia Alpert: formal analysis, validation, supervision, resources. Sergii Bodnar: formal analysis, validation.

## References

1. P. Meyfroidt et al., "Ten facts about land systems for sustainability," *PNAS* **119**(7), 1–12 (2022).
2. E. Barbiroglio, "Land use puts huge pressure on Earth's resources. Here's what needs to change," https://ec.europa.eu/research-and-innovation/en/horizon-magazine/land-use-puts-huge-pressure-earths-resources-heres-what-needs-change (access 14 May 2022).
3. J. Liu et al., "Systems integration for global sustainability," *Science* **347**(6225), 1258832 (2015).
4. S. Seifollahi-Aghmiuni et al., "Urbanisation-driven land degradation and socioeconomic challenges in peri-urban areas: insights from Southern Europe," *J. Environ. Soc.* **51**, 1446–1458 (2022).

5. P. C. Pandey et al., "Land use/land cover in view of earth observation: data sources, input dimensions, and classifiers - a review of the state of the art," *Geocarto Int.* **36**(9), 957–988 (2019).
6. S. Bontemps et al., "Multi-year global land cover mapping at 300 m and characterization for climate modelling: achievements of the Land Cover component of the ESA Climate Change Initiative," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **XL-7/W3**, 323–328 (2015).
7. J. Cihlar, "Land cover mapping of large areas from satellites: status and research priorities," *Int. J. Remote Sens.* **21**(6-7), 1093–1114 (2000).
8. A. Belward and J. Skøien, "Who launched what, when and why; trends in global land-cover observation capacity from civilian earth observation satellites," *ISPRS J. Photogramm. Remote Sens.* **103**, 115–128 (2015).
9. D. Phiri et al., "Sentinel-2 data for land cover/use mapping: a review," *Remote Sens.* **12**(14), 2291 (2020).
10. H. T. T. Nguyen et al., "Land use/land cover mapping using multitemporal Sentinel-2 imagery and four classification methods—a case study from Dak Nong, Vietnam," *Remote Sens.* **12**(9), 1367 (2020).
11. G. L. Spadoni et al., "Analysis of Normalized Difference Vegetation Index (NDVI) multi-temporal series for the production of forest cartography," *Remote Sens. Appl.: Soc. Environ.* **20**, 100419 (2020).
12. V. Nasiri et al., "Land use and land cover mapping using Sentinel-2, Landsat-8 satellite images, and Google Earth engine: a comparison of two composition methods," *Remote Sens.* **14**(9), 1977 (2022).
13. A. Tassi and M. Vizzari, "Object-oriented LULC classification in Google Earth learning algorithms," *Remote Sens.* **12**(22), 3776 (2020).
14. K. N. Loukika, V. R. Keesara, and V. Sridhar, "Analysis of land use and land cover using machine learning algorithms on Google Earth engine for Munneru River Basin, India," *Sustainability* **13**(24), 13758 (2021).
15. A. Rahman et al., "Performance of different machine learning algorithms on satellite image classification in a rural and urban setup," *Remote Sens. Appl.: Soc. Environ.* **20**, 100410 (2020).
16. K. Antil and M. Pal, "Comparison of landsat 8 and sentinel 2 data for accurate mapping of built-up area and bare soil," in *38th Asian Conf. Remote Sens, – Space Appl,: Touching Human Lives, ACRS 2017* (2017).
17. C. Li et al., "A comparative analysis of index-based methods for impervious surface mapping using multiseasonal Sentinel-2 satellite data," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **14**, 3682–3694 (2021).
18. V. Belenok et al., "Geoinformation mapping of anthropogenically transformed landscapes of Bila Tserkva (Ukraine)," *Acta Sci. Pol. Formatio Circumiectus* **21**(1), 69–84 (2022).
19. V. Belenok et al., "Investigating anthropogenically transformed landscapes with remote sensing," *Remote Sens. Appl.: Soc. Environ.* **24**, 100635 (2021).
20. N. Gorelick et al., "Google Earth engine: planetary-scale geospatial analysis for everyone," *Remote Sensing Environmental* **202**, 18–27 (2017).
21. N. Sidhu, E. Pebesma, and G. Câmara, "Using Google Earth engine to detect land cover change: Singapore as a use case," *Eur. J. Remote Sens.* **51**(1), 486–500 (2018).
22. H. Tamiminia et al., "Google Earth engine for geo-big data applications: a meta-analysis and systematic review," *ISPRS J. Photogramm. Remote Sens.* **164**, 152–170 (2020).
23. M. Amani et al., "Google Earth engine cloud computing platform for remote sensing big data applications: a comprehensive review," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **13**, 5326–5350 (2020).
24. E. Raczko and B. Zagajewski, "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 (2017).
25. J. C.-W. Chan and D. Paelinckx, "Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery," *Remote Sens. Environ.* **112**(6), 2999–3011 (2008).

26. M. Belgiu and L. Drăguţ, "Random forest in remote sensing: a review of applications and future directions," *ISPRS J. Photogramm. Remote Sens.* **114**, 24–31 (2016).
27. L. Breiman, "Random forests," *Mach. Learn.* **45**(1), 5–32 (2001).
28. C. Lei et al., "A comparison of random forest and support vector machine approaches to predict coal spontaneous combustion in gob," *FUEL* **239**, 297–311 (2019).
29. P. Teluguntla et al., "A 30-m landsat-derived cropland extent product of Australia and China using random forest machine learning algorithm on Google Earth Engine cloud computing platform," *ISPRS J. Photogramm. Remote Sens.* **144**, 325–340 (2018).
30. R. G. Rejith et al., "GIS-based machine learning algorithms for mapping beach placer deposits in the southwest coast of India using Landsat-8 OLI images," *J. Appl. Remote Sens.* **16**(1), 012011 (2022).
31. A. Shivani and H. Harini, "Classification of Indian cities using Google Earth Engine," *J. Land Use Sci.* **14**(4–6), 425–439 (2019).
32. B. F. Frimpong and F. Molkenthin, "Tracking urban expansion using random forests for the classification of Landsat imagery (1986–2015) and predicting urban/built-up areas for 2025: a study of the Kumasi Metropolis, Ghana," *Land* **10**(1), 44 (2021).
33. A. Ghosh, R. Sharma, and P. K. Joshi, "Random forest classification of urban landscape using Landsat archive and ancillary data: combining seasonal maps with decision level fusion," *Appl. Geogr.* **48**, 31–41 (2014).
34. J. Ye et al., "Analysis on land-use change and its driving mechanism in Xilingol, China, during 2000–2020 using the Google Earth engine," *Remote Sens.* **13**(24), 5134 (2021).
35. M. F. Baqa et al., "Monitoring and modeling the patterns and trends of urban growth using urban sprawl matrix and CA-Markov Model: a case study of Karachi, Pakistan," *Land* **10**(7), 700 (2021).
36. R. S. Defries and J. R. Townshend, "NDVI-derived land cover classifications at a global scale," *Int. J. Remote Sens.* **15**(17), 3567–3586 (1994).
37. W. Takeuchi and Y. Yasuoka, "Development of normalized vegetation, soil and water indices derived from satellite remote sensing data. C-9.4," in *Remote Sens. Appl., 25th ACRS*, Chiang Mai (2004).
38. R. C. Estoque and Y. Murayama, "Classification and change detection of built-up lands from Landsat-7 ETM+ and Landsat-8 OLI/TIRS imageries: a comparative assessment of various spectral indices," *Ecol. Indic.* **56**, 205–217 (2015).
39. M. Pal and K. Antil, "Comparison of Landsat 8 and Sentinel 2 data for accurate mapping of built-up area and bare soil," in *38th Asian Conf. Remote Sens.*, New Delhi, India (2017).
40. P. S. Rahar and M. Pal, "Comparison of various indices to differentiate built-up and bare soil with Sentinel 2 data," *Appl. Geomat. Civil Eng.* **33**, 501–509 (2019).
41. H. Xu, "A new index for delineating built-up land features in satellite imagery," *Int. J. Remote Sens.* **29**(14), 4269–4276 (2008).
42. P. Lynch, L. Blesius, and E. Hines, "Classification of urban area using multispectral indices for urban planning," *Remote Sens.* **12**, 2503 (2020).
43. X.-L. Chen et al., "Remote sensing image-based analysis of the relationship between urban heat island and land use/cover changes," *Remote Sens. Environ.* **104**(2), 133–146 (2006).
44. S. Bouzekri, A. A. Lasbet, and A. Lachehab, "A new spectral index for extraction of built-up area using Landsat-8 data," *J. Indian Soc. Remote Sens.* **43**(4), 867–873 (2015).
45. G. Sun et al., "Combinational Build-Up Index (CBI) for effective impervious surface mapping in urban areas," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **9**(5), 2081 (2016).
46. D. Lu, E. Moran, and S. Hetrick, "Detection of impervious surface change with multitemporal Landsat images in an urban-rural frontier," *ISPRS J. Photogramm. Remote Sens.* **66**(3), 298–306 (2011).
47. Q. Weng, X. Hu, and D. Lu, "Extracting impervious surfaces from medium spatial resolution multispectral and hyperspectral imagery: a comparison," *Int. J. Remote Sens.* **29**(11), 3209–3232 (2008).
48. C. Deng and C. Wu, "BCI: a biophysical composition index for remote sensing of urban environments," *Remote Sens. Environ.* **127**, 247–259 (2012).

49. J. Xu et al., "Extraction of built-up area using multi-sensor data—a case study based on Google earth engine in Zhejiang Province, China," *Int. J. Remote Sens.* **42**(2), 389–404 (2021).
50. G. Cai et al., "Detailed urban land use land cover classification at the metropolitan scale using a three-layer classification scheme," *Sensors* **19**(14), 3120 (2019).
51. S. W. Myint et al., "Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery," *Remote Sens. Environ.* **115**(5), 1145–1161 (2011).
52. L. G. E.-D. Taha and R. E. Ibrahim, "A machine learning model for improving building detection in informal areas: a case study of Greater Cairo," *Geomat. Environ. Eng.* **16**(2), 39–58 (2022).
53. H. S. Jaber, M. A. Shareef Shareef, and Z. F. Merzah, "Object-based approaches for land use-land cover classification using high-resolution quick bird satellite imagery (a case study: Kerbela, Iraq)," *Geod. Cartogr.* **48**(2), 85–91 (2022).
54. A. Shetty, P. Umesh, and A. Shetty, "An exploratory analysis of urbanization effects on climatic variables: a study using Google Earth Engine," *Model. Earth Systems and Environ.* **8**(1), 1363–1378 (2022).
55. U. Heiden et al., "Determination of robust spectral features for identification of urban surface materials in hyperspectral remote sensing data," *Remote Sens. Environ.* **111**(4), 537–552 (2007).
56. "Passport of the city of Kyiv," 2021, https://dei.kyivcity.gov.ua/content/kmu.html (access 19.03.2022).
57. "Ecological passport of the city of Kyiv," 2019, https://ecodep.kyivcity.gov.ua/content/ekologichnyy-pasport.html (access 07.03.2022).
58. V. I. Zatserkovny et al., "Black sea level change monitoring using altimetry data and geo-information technologies," in *XVIIIth Int. Conf. Geoinf. – Theor. and Appli.Aspects: May 13–16, 2019*, European Association of Geoscientists & Engineers, Kyiv, pp. 13–16 (2019).
59. D. Liashenko et al., "Landslide GIS modelling with QGIS software," in *XIV Int. Sci. Conf. Monit. of Geol. Processes and Ecol. Cond. of the Environ.: Nov. 10–13, 2020*, European Association of Geoscientists & Engineers, Kyiv (2020).
60. Senior Department of Statistics in Kyiv, "State Statistics Service of Ukraine," Population, 2022, http://www.kyiv.ukrstat.gov.ua/ (access 9.03.2022).
61. State Service of Ukraine for Emergencies, "Central geophysical observatory named after Boris Sreznevsky," http://cgo-sreznevskyi.kyiv.ua/index.php?fn=k_klimat&f=kyiv (access 07.03.2022).
62. Official Portal of Kyiv City, "The project of the City Target Program for the Use and Protection of Lands of the City of Kyiv for 2022-2025, No. 1425 of the Kyiv City Council dated June 22, 2021, The official portal of the city of Kyiv," https://kyivcity.gov.ua/news/kivrada_zatverdila_misku_tsilovu_programu_vikoristannya_ta_okhoroni_zemel_na_2022-2025_roki/ (access 09.03.2022).
63. The European Space Agency, "Copernicus open access hub," https://scihub.copernicus.eu/ (accessed 30 June 2022).
64. "Google Earth Engine," https://earthengine.google.com/ (accessed 30 June 2022).
65. G. Mateo-García et al., "Multitemporal cloud masking in the Google Earth Engine," *Remote Sensing* **10**(7), 1079 (2018).
66. The European Space Agency (ESA), "Radiometric resolutions," https://sentinels.copernicus.eu/ (accessed 30 June 2022).
67. M. Main-Knorn et al., "Sen2Cor for sentinel-2," *Proc. SPIE* **10427**, 2278218 (2017).
68. M. Popov et al., "Assessing long-term land cover changes in the watershed by spatio-temporal fusion of classifications based on probability propagation: the case of Dniester river basin," *Remote Sens. Appl.: Soc. Environ.* **22**, 100477 (2021).
69. "The Humanitarian data exchange," https://data.humdata.org (accessed 30 June 2022).
70. H. Zhou et al., "A hybrid approach of combining random forest with texture analysis and vdvi for desert vegetation mapping based on UAV RGB data," *Remote Sensi.* **13**(10), 1891 (2021).

71. C. W. Hsu, C. C. Chang, and C. J. Lin, "A practical guide to support vector classification," Technical Report; Department of Computer Science and Information Engineering, University of National Taiwan: Taipei, Taiwan, pp. 1–12 (2003).
72. L. Leroux et al., "Land cover mapping using Sentinel-2 images and the semi-automatic classification plugin: a Northern Burkina Faso Case Study," in N. Baghdadi, C. Mallet, and M. Zribi, Eds., *QGIS and Applications in Agriculture and Forest*, Vol. 2, ISTE Ltd (2018).
73. E. Głowienka and K. Michałowska, "Analyzing the impact of simulated multispectral images on water classification accuracy by means of spectral characteristics," *Geomat. Environ. Eng.* **14**(1), 47–58 (2020).
74. P. S. Rahar and M. Pal, "Comparison of various indices to differentiate built-up and bare soil with sentinel 2 data," in J. K. Ghosh and I. da Silva, Eds., *Applications of Geomatics in Civil Engineering*, pp. 501–509, Springer, Singapore (2020).
75. R. Nedkov, "Orthogonal transformation of segmented images from the satellite Sentinel-2," *C. R. Acad. Bulg. Sci.* **70**(5), 687–692 (2017).
76. G. S. Bhunia, P. Kumar Shit, and H. R. Pourghasemi, "Soil organic carbon mapping using remote sensing techniques and multivariate regression model," *Geocarto Int.* **34**(2), 215–226 (2019).
77. T. M. Lillesand, R. W. Kiefer, and J. Chipman, *Remote Sensing and Image Interpretation*, 7th ed., Wiley (2015).
78. R. Congalton, "Assessing positional and thematic accuracies of maps generated from remotely sensed data," in Remote Sensing Handbook, Data Characterization, Classification, and Accuracies, P. Thenkabail, Ed., Vol. I, pp. 583–601, CRC/Taylor & Francis, Boca Raton, Florida (2015).

**Vadym Belenok** received his degree candidate of physical and mathematical sciences (comparable to PhD) on remote-controlled airspace research from National Aviation University, Kyiv, Ukraine, in 2015. Now, he is an associate professor of the Aerospace Geodesy and Land Management Department at National Aviation University, Kyiv, Ukraine. His main research interests include geodesy, GIS, and remote sensing.

**Liliia Hebryn-Baidy** received her degree candidate of technical sciences (comparable to PhD) on cadastre and the lands monitoring from Lviv Polytechnic National University, Ukraine, in 2018. Now, she is an associate professor of the Aerospace Geodesy and Land Management Department at National Aviation University, Kyiv, Ukraine. Her main research interests include satellite imagery processing, remote aerospace research methods, land and real estate valuation, and land monitoring.

**Nataliia Bielousova** received her degree doctor of economics on specialty development of productive forces and regional economy from V.I. Vernadsky Taurida National University, Ukraine, in 2021. Now, she is a professor of the Aerospace Geodesy and Land Management Department at National Aviation University, Kyiv, Ukraine. Her main research interests include rational use of natural resources, tourism, and rehabilitation geography.

**Valeriy Gladilin** received his degree candidate of technical sciences (comparable to PhD) on geodesy, photogrammetry, and cartography from Kyiv National University of Construction and Architecture, Ukraine, in 1995. Now, he is an associate professor of the Geodesy and Land Management Department at Bila Tserkva National Agrarian University, Ukraine. His main research interests include geodesy and land management, land management design, land market, territorial planning and balanced land use, and determination of deformations of industrial equipment by geodetic measurements.

**Sergíy Kryachok** received his degree candidate of technical sciences (comparable to PhD) on geodesy, photogrammetry and cartography from Kyiv National University of Construction and Architecture, Ukraine, in 2001. Now, he is an associate professor of the Department of Geodesy, Cartography and Land Management at Chernihiv Polytechnic National University, Ukraine. His main research interests include geodesy, photogrammetry, surveying instruments, and vertical geodetic measurements.

**Andrii Tereshchenko** received his degree candidate of physical and mathematical sciences (comparable to PhD) in astrometry, celestial mechanics from National Aviation University, Kyiv, Ukraine in 2011. Now, he is an associate professor of the Aerospace Geodesy and Land Management Department at National Aviation University, Kyiv, Ukraine. His main research interests include dynamics of self-gravitating systems, celestial mechanics, digital image processing, economic, and mathematical modeling.

**Sofiia Alpert** received her degree candidate of technical sciences (comparable to PhD) on remote-controlled airspace research from the Scientific Centre for Aerospace Research of the Earth of the Institute of Geological Science of the National Academy of Sciences of Ukraine in 2016. Now, she is a researcher at the Department of Geoinformation Technologies in Remote Sensing of the Earth of Scientific Centre for Aerospace Research of the Earth of the Institute of Geological Science of the National Academy of Sciences of Ukraine. Her main research interests include processing and quality assessment of multispectral aerospace images interpretation, and satellite images classification.

**Sergii Bodnar** is educated as a geographer and cartographer. Currently, he works as an assistant in the Department of Geodesy and Cartography, Geography Faculty, Taras Shevchenko National University of Kyiv. His research interests include photogrammetric measurements of architectural monuments and historical and cultural heritage, digital photogrammetry and UAV surveys, large-scale topographic surveys, and engineering geodesy. He has published more than 50 scientific and educational works.