

# Environmental Monitoring of Land Use/ Land Cover by Integrating Remote Sensing and Machine Learning Algorithms

Firas K. Aljanabi<sup>1\*</sup>, Mert Dedeoğlu<sup>2</sup>, Cevdet Şeker<sup>3</sup>

<sup>1,2,3</sup>Soil Science and Plant Nutrition Department, College of Agriculture, Selcuk University, Konya, Turkey

<sup>1</sup><https://orcid.org/0000-0002-5122-3699>,

<sup>2</sup><https://orcid.org/0000-0001-8611-3724>,

<sup>3</sup><https://orcid.org/0000-0002-8760-6990>

\*Email: [Firas.aljanabi@lisansustu.selcuk.edu.tr](mailto:Firas.aljanabi@lisansustu.selcuk.edu.tr)

Article Info	Abstract
Received 01/12/2023 Revised 28/03/2024 Accepted 30/03/2024	Evaluation of the land use/ land cover (LULC) case over large regions is very important in a variety of domains, including natural resources such as soil, water, etc., and climate change risks and LULC change has emerged as a high anxiety for the environment. Therefore, we tested and compared the performance of three classification algorithms: Support Vector Machines (SVM), Random Trees (RT), and Maximum Likelihood (MaxL) to derive and extract LULC information for the district of Sarayönü/ Konya across five distinct classes: water, plantation, grassland, built-up, and bare land. Two remote sensing indices, the normalized difference vegetation index (NDVI) and the normalized difference water index (NDWI), were used as supplementary inputs for the classification of LULC. To evaluate the performance of the algorithms, a confusion matrix was employed. The average overall accuracy of support vector machines, random trees, and maximum likelihood algorithms was found 85.60%, 79.20%, and 74.80%, respectively, and 82.00%, 74.00%, and 68.50% for the Kappa coefficient. These results indicate that the support vector machines algorithm outperforms other algorithms in terms of accuracy. As a result of the research, it was determined that classification algorithms integrated with remote sensing in LULC change monitoring/determination could produce accurate classification maps that can be used as base data. This is due to the ability of machine learning algorithms to learn complex patterns, adapt to diverse data, and continuously improve, making them achieve higher accuracy compared to traditional classifiers. Therefore, their use was recommended for decision-makers.

**Keywords:** Geographic information system; Remote sensing; Land use/ land cover; Machine learning; Random trees; Support vector machine

## 1. Introduction

The comprehensive understanding of LULC across various scales holds significant potential for advancing investigations into global phenomena, including but not limited to droughts, floods, soil erosion, natural hazards, and climate change [1]-[3]. Precise analysis of LULC forms an integral aspect of sustainable development applications within any given region. The assessment and prediction of dynamics in LULC stand as practical tools for managing and comprehending landscape conversions [4]. The evaluation of changes in LULC proves

indispensable for multiple objectives, including the management of rapid and unchecked increase in population, and economic and industrial progress, particularly in developing countries undergoing substantial alterations in LULC [5]-[8]. Elaborated LULC maps emerge as crucial inputs across various scientific domains, encompassing climate change, natural resource management, and land monitoring. These maps assist in identifying suitable lands for agriculture and watershed management [9]-[10]. In recent years, remote sensing researchers have increasingly adopted

machine-learning classification techniques to generate LULC maps [11]- [12]. Notably, relevant previous studies and findings related to the topic described in the next section underscore the enhanced accuracy and reliability of machine learning classifiers compared to traditional statistical counterparts like the Minimum Distance and Parallelepiped classifiers, particularly when equipped with substantial training data [13].

The SVM algorithm's strong capacity to generalize complicated characteristics has allowed it to perform admirably [14]-[15]. Sentinel-2 data were used for LULC classification tasks in research by Abdi (2020), Four distinct machine learning techniques were used in the study to categorize eight different LULC classes. The total accuracy of the SVM algorithm was  $0.758 \pm 0.017$  [11]. Similarly, RF is a widely known machine-learning algorithm [16]. Due to their adaptability to solve both regression and classification issues while accounting for both categorical and persistent variables [17]. Sentinel-2A and Planet Scope data were used by Aguilera in a 2020 research to classify land cover using the pixel classification approach, the research entailed evaluating six different machine learning algorithms to see how effective they were. The decision trees algorithm is noteworthy for achieving a kappa value of 65.70% and an overall accuracy of 67.80%. Although moderate accuracy is attained, the study indicates that higher outcomes may be attained by refining the suggested models [18].

Numerous applications have taken use of the RF algorithm's versatility, such as object-oriented categorization, land use analysis, and tree cover modeling [19]-[22]. Sentinel-2 data and five distinct machine learning algorithms were used in research by Prasad et al. (2022) to evaluate the performance of these algorithms in categorizing six LULC classes. The RF algorithm achieved an overall accuracy of 77.20% with a kappa coefficient of 0.727. Additionally, an overall accuracy of 80.60% with a kappa coefficient of 0.767 was obtained from the combination of Sentinel-1 and -2 data. These results highlight how the combination models that have been suggested might improve the precision of LULC maps [23].

Several research works have underscored the significance of the MaxL algorithm in the process of categorization [24]-[25]. There are two concepts for the MaxL algorithm. First, it assumes that the multidimensional space's sample class pixels have a normal distribution. Second, it makes decisions by applying Bayes' theorem. When allocating each pixel to a class represented in the classification, this algorithm considers both the covariances and variances of the class signatures [24]. Salih and Salih (2015) conducted a study employing three different machine learning algorithms to classify seven classes of LULC using a method of image fusion of optical multi-

source remote sensing data. Notably, the MaxL algorithm achieved an overall accuracy of 71.09% [26].

Unsupervised learning models tend to have less accurate results, as unsupervised classifiers rely on similarity measures such as color information and pixel distances, potentially introducing imprecision [27]. Conversely, the model approach proposed in this study utilizes supervised learning models, resulting in higher accuracy for classification maps. Performance in classification is influenced by several factors, including the type of satellite imagery used, band combinations, and proper parameter selection or training data usage [28]. Sentinel-2 optical data have been pivotal in classification, segmentation, plant species identification, agricultural product assessment, and water quality studies [29]-[33]. Based on this, the preference for Sentinel-2 data in this study stems from its high spatial resolution, multiple spectral bands, and open data policy. Increasing training sample sizes for machine learning algorithms improves the model's understanding of data diversity and complexity, enhancing robustness and reliability for more accurate results, this indicates that many training samples are needed for accurate categorization [21]. Consequently, this study used 1250 samples to classify five LULC classes to achieve maximum accuracy compared to state-of-the-art models. The major objective of this study is to use three distinct machine learning algorithms (SVM, RT, and MaxL) integrated with NDVI and NDWI indices for LULC classification using multispectral satellite imagery from Sentinel-2. In this regard, this manuscript seeks to identify the machine learning algorithm that achieves the highest overall accuracy and kappa coefficient in a five-category LULC classification scenario (water, plantation, grassland, built-up, and bare land) for the Sarayönü district. As a result, this novel approach offers a systematic means to evaluate the actual situation of LULC through the integration of remote sensing indicators and machine learning algorithms, rendering this study both original and a valuable resource for decision-makers. The results gleaned from this study will aid decision-makers and interested parties in advancing sustainable development and the optimal management of natural resources in the region.

## 2. Material and Methods

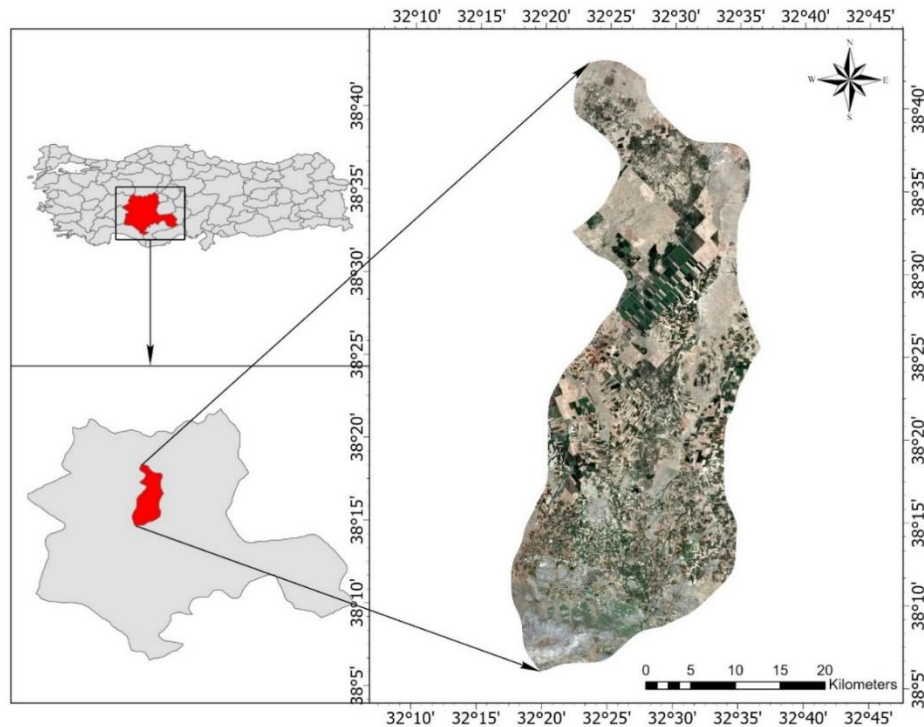
### 2.1. Materials

#### 2.1.1. Study Area

The study area occupies a geographical extent delineated by latitude  $38^{\circ}5'58''$  N to  $38^{\circ}42'51''$ N and longitude  $32^{\circ}17'37''$ E to  $32^{\circ}36'39''$ E, which represents Sarayönü district within Konya province/ Turkey Fig.1 This locale spans an area of approximately 1555.28 square kilometers. Situated at a distance of 48 kilometers from the city center, the district's

average elevation stands at 1068 meters above sea level. The district's terrain predominantly comprises limestone formations, nestled on the plains that extend from the southern reaches of the Cihanbeyli plateau. These plains, unmarred by extensive river incisions, exude a distinct topographical simplicity in essence, the region constitutes a closed basin in terms of its hydrological characteristics, devoid of major surface watercourses within the district. Significant inputs to

subterranean aquifers consist of rain-driven infiltration from Beşgöz and Buharcalı, which slowly seep from specific slopes, thus the emergence of Beşpınar spring water. However, it's worth noting that these water sources are vulnerable to evaporation and absorption. Despite the aforementioned hydrological characteristics, the district hosts a considerable expanse of agricultural land, evidence of the district's fertile soil quality.



**Figure 1.** Location of the study area.

### 2.1.2. Satellite Data

A quantities of earth monitoring data for the prior years including satellite images such as Sentinel-2 are stored in the ESA (European Space Agency) platform and this data can be accessed free of charge. In this study, multispectral Sentinel-

2A images with low cloud cover between (0 - 9.4) percent dated May/ 2023. Fig. 2 presents the properties of the bands for the Sentinel-2 images.

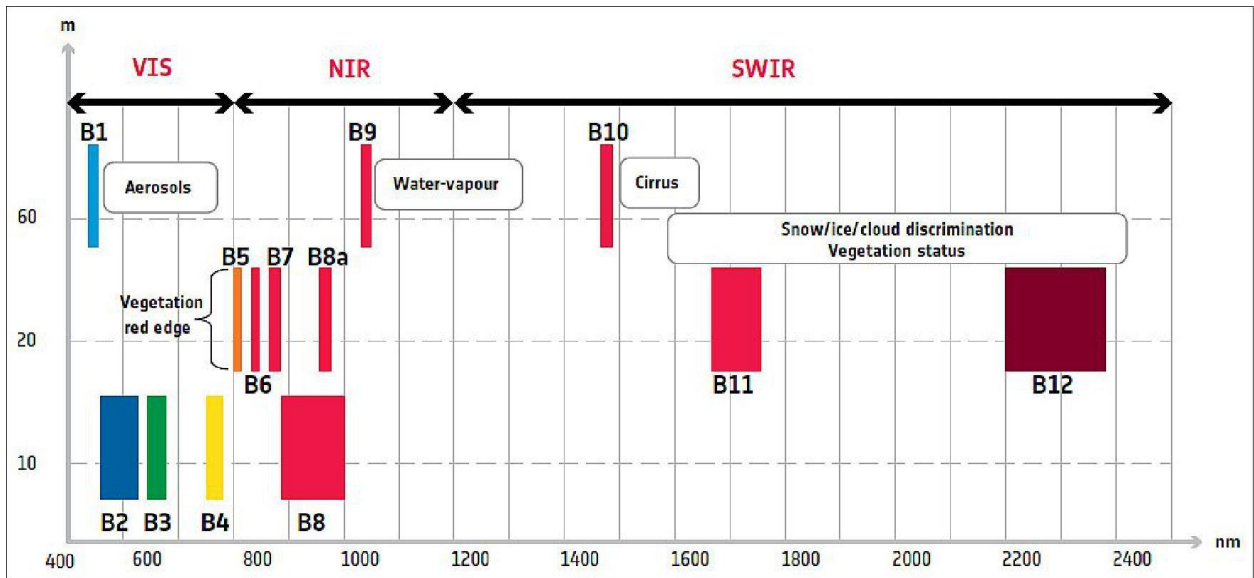


Figure 2. Spatial resolution for multispectral instrument (MSI) of Sentinel-2 (image credit: ESA).

2.2. Methods

This study employs pixel-based classification using machine learning algorithms integrated with remote sensing indicators.

This amalgamation enhances precision in (LULC) characterization. Fig. 3 illustrates the methodology of the study in the form of a flowchart.

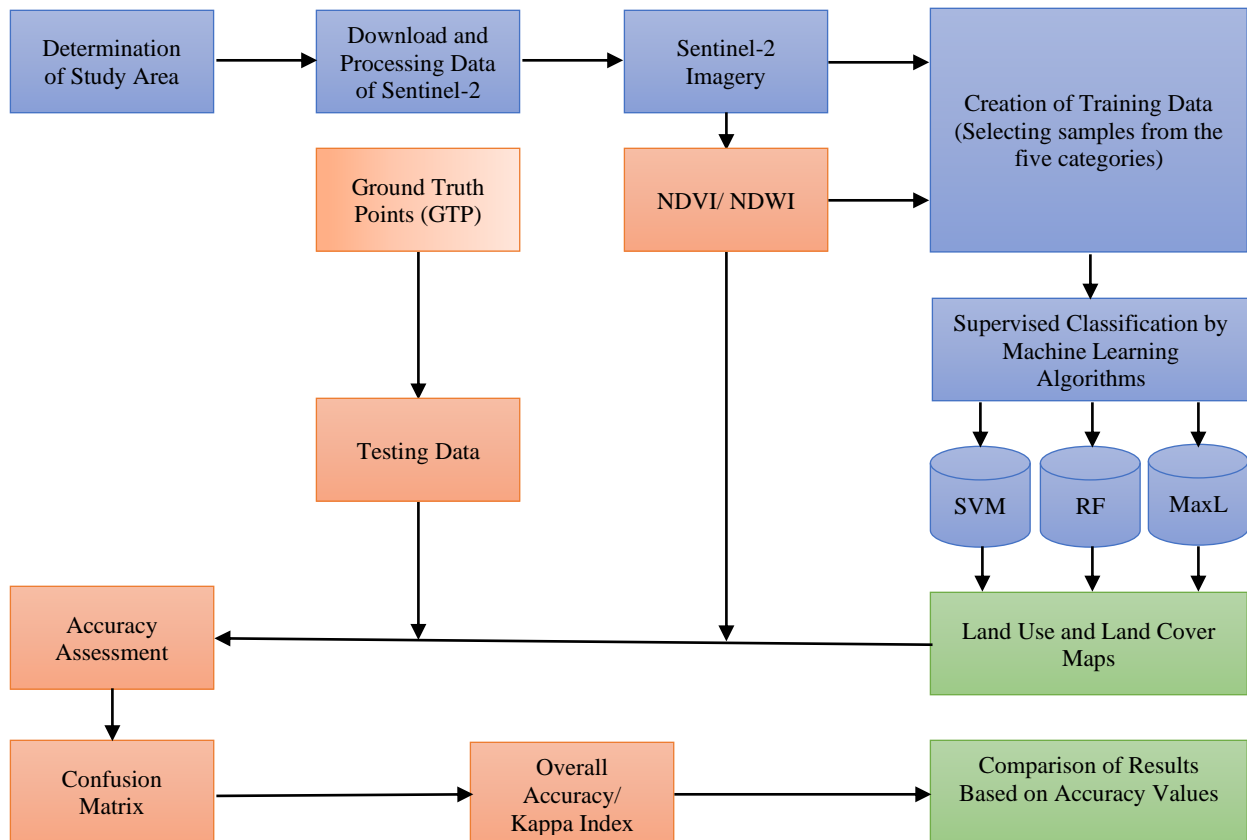


Figure 3. Flowchart of the methodology.

### 2.2.1. Processing of Satellite Data

The classification process began with obtaining corrected satellite images with low cloud cover between (0 - 9.4) percent of the datasets. These images served as primary inputs for the classification process, the image processing chain was performed using ArcGIS Pro software (Version 2.5) and the open-source SNAP software (Version 9.0.0). Training polygons were generated for five land use categories and these polygons were evenly distributed across the study area. The initial dataset, comprising 1250 samples (samples are segmented polygons from the image according to categories, as shown in Table 1), was divided into two subsets: 80% of the dataset, equivalent to 1000 polygons, was allocated for training purposes, while the remaining 20%, or 250 polygons, was reserved for testing [34]. Validation work was executed using ground truth points (GTP) for surveys of ground truth. In addition, supplementary data were acquired from NDVI and NDWI maps of the study area to recognize the LCLU classes and to validate the proposed methodology, the NDVI, and NDWI were calculated by following Equations (1) and (2):


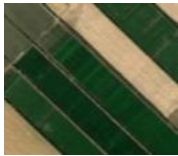


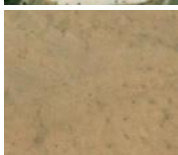
$$NDVI = (NIR - RED)/(NIR + RED) \quad (1)$$

$$NDWI = (G - NIR)/(G + NIR) \quad (2)$$

Equation (1) symbolizes the NDVI index, a measure of the density and health of vegetation. It is computed by dividing the difference in reflectance levels between the red (RED) and near-infrared (NIR) sections of the electromagnetic spectrum by their total. Typically, NDVI values fall between -1 and 1. High positive values (close to 1) indicate healthy and dense vegetation. Conversely, low or negative values (close to -1) represent non-vegetated surfaces, such as water bodies or barren land. Values (around 0) suggest the presence of soil, rocks, or built-up areas.

Equation (2) refers to the NDWI index, an additional measure for remote sensing that is used to identify water bodies in satellite images. The near-infrared (NIR) and green (G) spectral bands are where NDWI is most sensitive to the absorption and reflection properties of water. The values of NDWI range from -1 to 1. High positive values (close to 1) indicate the presence of water bodies. Low or negative values (close to -1) are associated with non-water features, such as land surfaces. Values (around 0) in this range may indicate mixed pixels or transitional areas between land and water.

**Table 1.** Description of the LULC classes identified

Class Name	Class Description	Example of class
Water	Area covered by water	
Plantation	Area covered by crops, forest	
Grassland	Area covered by sparse grass	
Built-up	Area covered by buildings, roads	
Bare land	Area without vegetation, bare soil	

### 2.2.2. Machine Learning Algorithms

**Support Vector Machine (SVM):** SVM, a supervised learning method widely utilized in a variety of remote sensing applications, was first introduced by Cortes and Vapnik (1995) [35], according to Vapnik's findings (1982) [36]. It is an alternate method for classifying images that enables precise classifications to be made using a collection of condensed training examples [37]-[38], to distinguish types in the training data, the algorithm iteratively searches for the best hyperplane boundary in an n-dimensional classification space. It then applies the same arrangement to a different evaluation dataset. The number of spectral bands in this context represents the dimensions, and the individual pixels in a multiband composite serve as the vectors [39]. According to Buddhiraju and Rizvi (2010) [40] and Mather and Tso (2016) [41], the fundamental benefit of SVM is its capacity to reduce classification errors by building a hyperplane between every pair of classes that maximizes the distance between the support vectors of every class. Cortes and Vapnik (1995) [35] provide a thorough mathematical explanation of this algorithm.

**Random Trees (RT):** RT is a machine-learning algorithm based on decision trees (DT). RT belongs to a type of machine learning algorithm that does group classification. The term "group" denotes a technique that averages the forecasts of

various base models to provide predictions. The Random Forest (RF) algorithm, later referred to as RT for trademark reasons, was originally conceived by Breiman (2001) [16] as a technique for merging several Classification and Regression Trees (CART). Since its introduction by Breiman (2001) [16], the RF framework has been considerably successful as a general-purpose classification and regression method [42]. Various studies have shown satisfactory performance by using the RT algorithm for LULC classification and other fields of remote sensing applications [43]-[45]. The number of parameters, which can be elucidated by 'm-try.' and trees, which can be elucidated by 'n-tree', are the two most crucial input variables for Random Forest (RF) algorithm. According to the literature, the best number of trees to count is between 100 and 500, and the best number of variables to count is the square of the set of variables [46]. This technique uses a large number of trees to improve accuracy in the land use modeling and image classification fields [47]. Breiman (2001) [16] stated that utilizing more trees than necessary is a waste of time and effort, and the overall performance of the model is not affected. In addition, Feng et al. (2015) [48] highlighted that the performance of Random Forest was precise by using 200 decision trees in their study.

**Maximum likelihood (MaxL):** The MaxL algorithm has been frequently used in supervised classification due to its accessibility and the lack of a lengthy training process [49]-[51]. MaxL algorithm is dependent on the likelihood that a pixel belongs to a particular class, for calculating the likelihood  $D$  of unknown measurement vector  $X$ , or weighted distance, belongs to one of the established categories,  $M_c$ , that is dependent on the Bayesian Equation [52].

$$D = \ln(a_c) - [0.5 \ln(|Cov_c|)] - [0.5(X - M_c)^T(Cov_c^{-1})(X - M_c)] \quad (3)$$

Where ( $D$ ) is the weighted distance (likelihood), ( $X$ ) is the measurement vector of the nominee pixel, ( $c$ ) is a particular class, ( $M_c$ ) is the sample's mean vector for class  $c$ , ( $a_c$ ) is the percentage chance that each potential pixel belongs to class  $c$ , ( $|Cov_c|$ ) is the covariance matrix's determinant for the class  $c$  sample's pixels, ( $Cov_c^{-1}$ ) is the inverse of  $Cov_c$  (matrix algebra)  $\ln$  = natural logarithm function, ( $T$ ) is the transposition function (matrix algebra)."

### 2.2.3. Accuracy Assessment

Following the completion of the classification process, an evaluation of accuracy was conducted to gauge the precision of the LULC maps. It is essential to determine whether the resulting map aligns with, surpasses, or falls short of predetermined classification accuracy criteria, which serves as a basis for assessing the correctness of the classification.

Researchers commonly employ an error matrix, often referred to as a confusion matrix or contingency table [34], [53], as a widely accepted technique to evaluate classification accuracy. In this matrix, known reference data is compared to the results of the related classification [54]. The confusion matrix, presented in a square format organized with rows and columns, quantifies the extent to which sample units, such as pixels, clusters of pixels, or polygons, have been accurately assigned to specific classes compared to their actual ground-based classifications, as elucidated by [55]. In this study, validation was performed using an error matrix, and the three algorithms used were validated, as depicted in Tables 2, 3, and 4 (see Results and Discussion section). It's worth noting that the ArcGIS Pro 2.5 program includes an integrated algorithm for generating the confusion matrix, computing the process of validating and assessing the accuracy of image classification. The overall accuracy (OA) and kappa coefficient (KC) are computed using the equations provided below:

$$O A = \left( \frac{P_c}{P_n} \right) \times 100 \quad (4)$$

Where ( $P_c$ ) is the number of pixels that are correctly categorized, and ( $P_n$ ) is the overall number of pixels.

$$K C = \frac{N \sum_{i=1}^r x_{ii} - \sum_{i=1}^r (x_{i+} \times x_{+i})}{N^2 - \sum_{i=1}^r (x_{i+} \times x_{+i})} \quad (5)$$

Here ( $N$ ) the overall number of observations, ( $r$ ) the count of columns and rows in the confusion matrix, ( $x_{ii}$ ) the count of observations in column  $i$  and row  $i$ , ( $x_{+i}$ ) the marginal sum of column  $i$ , ( $x_{i+}$ ) the marginal sum of row  $i$ .

User's accuracy (UA) for each class is calculated as the proportion of correctly classified pixels within that class relative to the total number of classified pixels. Conversely, the producer's accuracy (PA) is computed as the ratio of correctly classified pixels to the overall number of pixels in the reference data for each class [34],[56].

User's Accuracy =

$$\frac{\text{Number of correctly classified pixels in every category (diagonal)}}{\text{Total number of reference pixels in every category (row total)}} \times 100\% \quad (6)$$

Producer's Accuracy =

$$\frac{\text{Number of correctly classified pixels in every category (diagonal)}}{\text{Total number of reference pixels in every category (column total)}} \times 100\% \quad (7)$$

The equalized stratified random method was employed to create and generate the ground truth points (GTP) in ArcGIS



Pro 2.5 software on the Sentinel-2 image from analysis tools > image analyst tools > create accuracy assessment points Fig. 4.

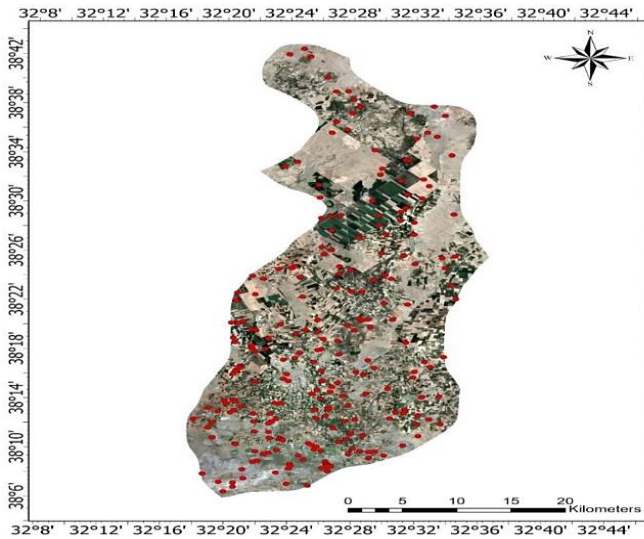


Figure 4. Overview of the ground truth points.

Fig. 5. demonstrates the widely employed indices that were used as additional inputs for the LULC classification, which are NDVI and NDWI for the study area which reflect the condition of vegetation and water respectively.

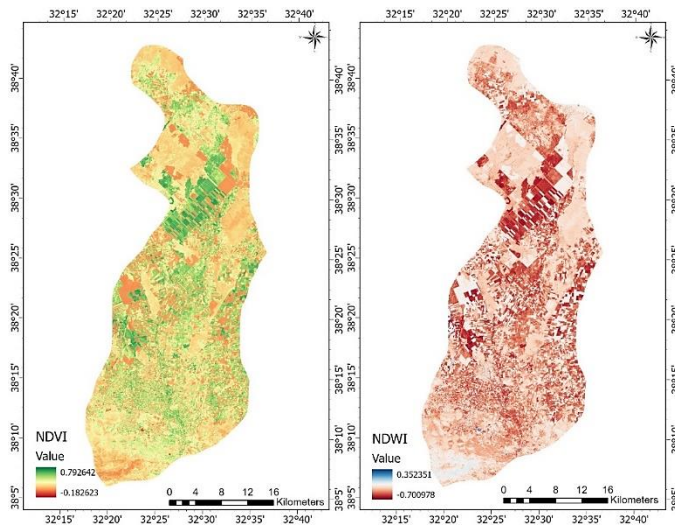


Figure 5. NDVI and NDWI of study area.

### 3. Results and Discussion

This study examines the effectiveness of three machine learning algorithms in the classification of LULC using Sentinel-2 Level-2A data at a 10-meter resolution. These algorithms utilize corrected images, incorporating atmospheric corrections that address Rayleigh scattering, as well as the absorbing and scattering effects of atmospheric gases, along with corrections for absorption and scattering caused by

aerosol particles. Fig. 6, 7, and 8 illustrate the application of these three machine learning algorithms (SVM, RT, and MaxL) in mapping LULC. To ensure fairness, an equal number of training and assessment samples were employed for all five LULC classes, preventing an overall accuracy bias toward classes with more extensive training samples [57].

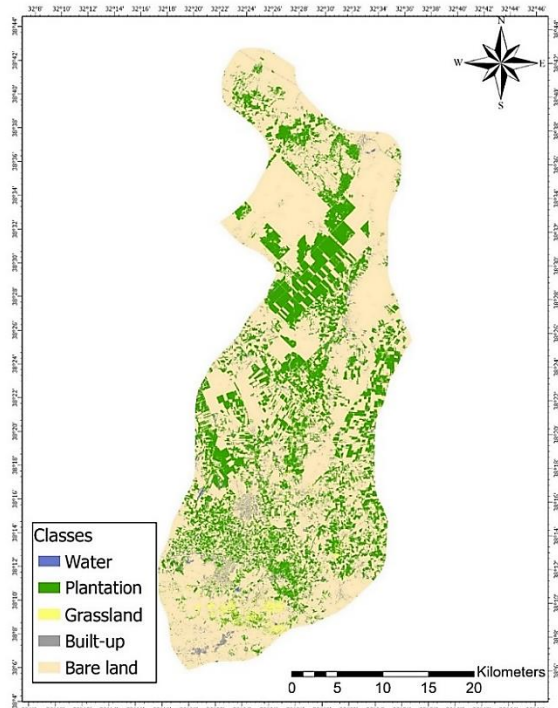


Figure 6. LULC map by SVM.

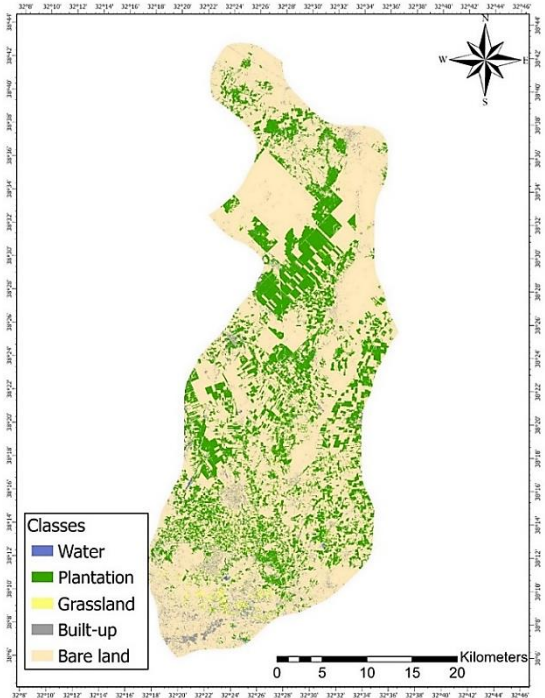


Figure 7. LULC map by RT.

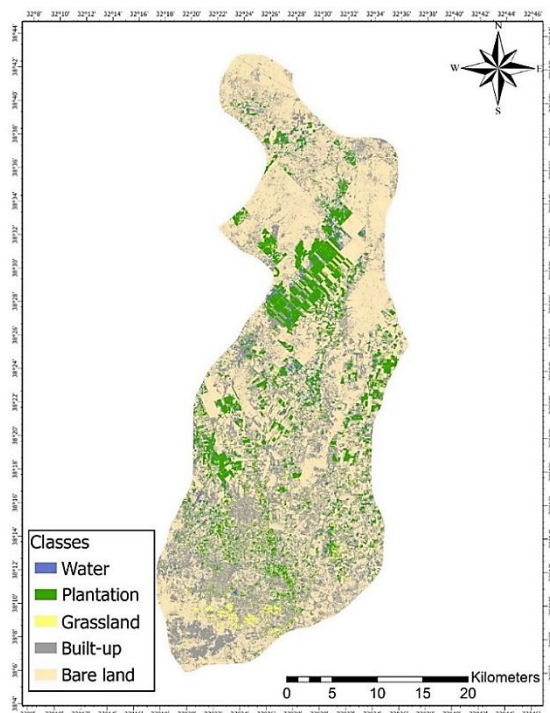


Figure 8. LULC map by MaxL.

The overall accuracy of the three algorithms was relatively different from each other. These precisions were deduced from the error matrix shown in Tables 2, 3, and 4. The most significant overall accuracy was produced by SVM (85.60%), closely followed by RT (79.20%), and finally MaxL (74.80%). While kappa coefficient was 82.00%, 74.00%, and 68.50%, respectively, as showed by other studies [14, 18, 26]. All three algorithms produced somewhat comparable maps that had an attractive appearance and represented the study area fairly well Fig. 5, Fig. 6, and Fig.7. However, relative to the other two algorithms, the MaxL overestimated class of built-up, and undervalued the classes that were left such as plantation except bare land Fig. 8. Despite of using 10 m resolution Sentinel-2 images, it is risque to recognize between bare land and built-up classes due to mixed pixels, especially in cases of closely-related categories, prior probability estimates can have a significant impact on classification results. Whereas when the land cover classes are well detached, using prior probability estimates produces unpretentious effects [58].

Table 2. Error matrix for SVM algorithm.

	Water	Plantation	Grassland	Built-up	Bare land	Total	UA	Kappa
Water	48	2	0	0	0	50	0.96	0
Plantation	0	47	1	0	2	50	0.94	0
Grassland	0	1	49	0	0	50	0.98	0
Built-up	0	0	11	21	18	50	0.42	0
Bare land	0	0	1	0	49	50	0.98	0
Total	48	50	62	21	69	250	0	0
PA	1	0.94	0.790	1	0.710	0	0.856	0
Kappa	0	0	0	0	0	0	0	0.82



**Table 3.** Error matrix for RT algorithm.

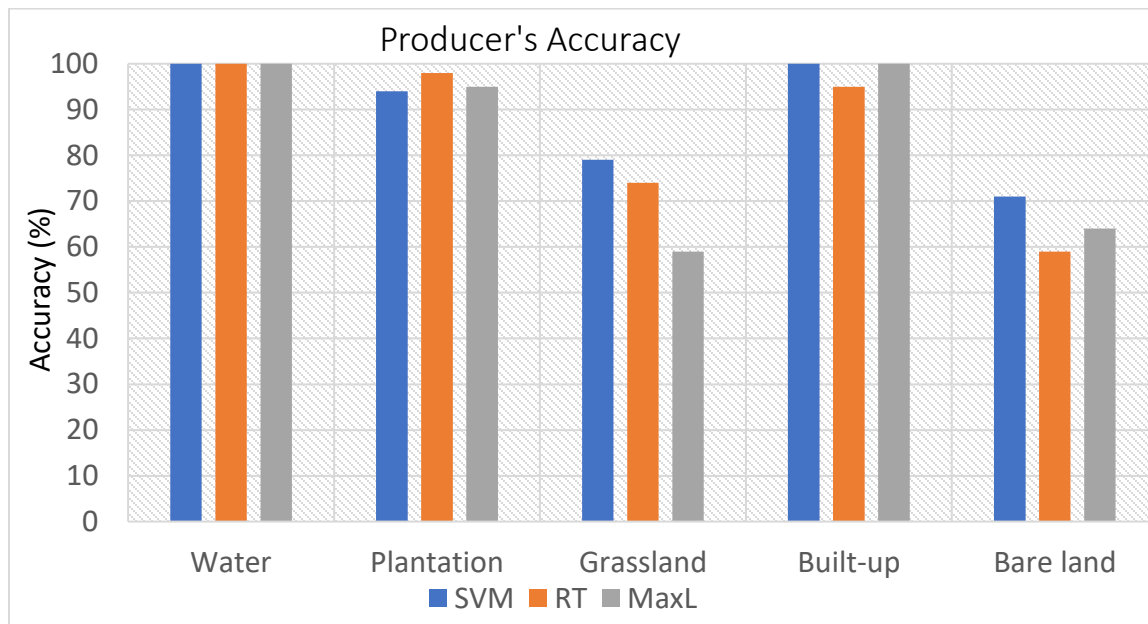
	Water	Plantation	Grassland	Built-up	Bare land	Total	UA	Kappa
Water	29	1	14	0	6	50	0.58	0
Plantation	0	49	1	0	0	50	0.98	0
Grassland	0	1	50	0	0	50	1	0
Built-up	0	0	2	21	27	50	0.42	0
Bare land	0	0	0	1	49	50	0.98	0
Total	29	50	67	22	82	250	0	0
PA	1	0.98	0.746	0.954	0.597	0	0.792	0
Kappa	0	0	0	0	0	0	0	0.74

**Table 4.** Error matrix for MaxL algorithm.

	Water	Plantation	Grassland	Built-up	Bare land	Total	UA	Kappa
Water	25	2	22	0	1	50	0.5	0
Plantation	0	41	7	0	2	50	0.82	0
Grassland	0	0	50	0	0	50	1	0
Built-up	0	0	4	22	24	50	0.424	0
Bare land	0	0	1	0	49	50	0.98	0
Total	25	43	84	22	76	250	0	0
PA	1	0.953	0.595	1	0.644	0	0.748	0
Kappa	0	0	0	0	0	0	0	0.685

The lowest producer's accuracy across all three algorithms was for grassland and bare land Fig. 9, which demonstrates how difficult it is to identify these classes from the other classes. Grassland was composed primarily of areas covered by sparse grass and was visually similar to plantations, especially those with regrown shrubs and crop cover. Thus, it is complicated to classify due to the similar spectral characteristics of the vegetation found in the study area [59]. On the other hand, the study area's bare land was covered with suburban elements like small towns and villages that were dotted with trees and sparse grass. In addition to that may be some similarities in the spectral properties of rooftops of buildings (Such as buildings that have tile roofs with spectral characteristics common to bare land) and some types of soils in the bare land in the study

area. However, it is challenging to separate these two types of surfaces based on their spectral signatures alone. Multispectral imagery and other techniques such as artificial surface index (ASI) may be used to classify built-up classes based on their unique spectral properties [60]. Therefore, there was an overlap in the classification of the built-up class and the bare land class, and this was evident in the results of the MaxL logarithm classification. The addition of land surface temperature (LST) in this study [61] could have enhanced the detection of bare land and built-up classes but the MSI instrument in Sentinel-2 does not contain a thermal band therefore it is not possible to derive the land surface temperature.



**Figure 9.** The producer's accuracy for SVM, RT, and MaxL algorithms.

#### 4. Conclusions

Within this study, the analysis encompassed the utilization of three distinct machine learning algorithms for the delineation of LULC conditions during May 2023. Anchored in Sentinel-2 data and GIS program tools, this endeavor specifically focused on the Sarayönü district of Konya province, situated in Central Anatolia, Turkey. The distinctive attributes of the Sentinel-2 multispectral satellite, which include three red-edged bands enabling the capture of plant chlorophyll content, a spatial resolution of 10 meters, and unfettered access to its data, set it apart from other Earth observation satellites.

The machine learning algorithms that underscored this investigation were subject to meticulous evaluation. Of these, two, namely SVM and RT, hold prominence within the realm of remote sensing. The third algorithm, MaxL while chiefly employed in statistical science, also finds application in the domain of remote sensing. The empirical results demonstrated that the SVM algorithm outperformed the other two algorithms RT and MaxL when it came to categorizing different categories within the study area, with an overall accuracy rate of 85.60% and a Kappa coefficient of 82.00%.

The implications derived from this manuscript are encapsulated as follows: Firstly, it substantiates the efficacy of leveraging open-source data, channeling it through diverse classification algorithms implemented via various programs, to yield a robust framework for characterizing the intricate nuances of LULC dynamics. Secondly, it proffers a strategic vision with the potential to guide planners in their decision-

making processes, facilitate ecosystem preservation, and promote sustainable land development strategies across analogous conditions. Thirdly, the study achieved commendable classification accuracies across all algorithms, except the MaxL algorithm which exhibited a relatively lower performance.

In subsequent investigations, we envisage a comparative analysis between the outcomes derived from Sentinel-2 data and analogous data from other temporal times, aimed at comprehensively monitoring LULC transformations within the study area. Such insights are envisioned to serve as a practical guide for effective land management decisions. Furthermore, the proposed methodology is poised for broader application across expansive regions, augmenting the inclusive monitoring and cartography of LULC dynamics within Central Anatolia, Turkey, and similar geographical conditions.

#### Conflict of interest

The authors report no conflicts of interest regarding the publication of this manuscript.

#### Author Contribution Statement

The first author proposed the research problem, which included using machine learning algorithms to classify the LULC with supervision from Dr. Mert Dedeoğlu and Dr. Cevdet Şeker. The authors confirmed the study findings by the confusion matrix and confirming their validity and efficacy. The introduction and manuscript style were developed by all

authors and the results were discussed and incorporated into the final version by all authors.

## References

- [1] 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*, vol. 13, no. 24, p. 13758, 2021, doi: <https://doi.org/10.3390/su132413758>.
- [2] S. Talukdar and S. Pal, "Effects of damming on the hydrological regime of Punarbhaba river basin wetlands," *Ecological Engineering*, vol. 135, pp. 61-74, 2019, doi: <https://doi.org/10.1016/j.ecoleng.2019.05.014>.
- [3] Y. Zhang, T. Ge, W. Tian, and Y.-A. Liou, "Debris flow susceptibility mapping using machine-learning techniques in Shigatse area, China," *Remote Sensing*, vol. 11, no. 23, p. 2801, 2019, doi: <https://doi.org/10.3390/rs11232801>.
- [4] J.-F. Mas, R. Lemoine-Rodríguez, R. González-López, J. López-Sánchez, A. Piña-Garduño, and E. Herrera-Flores, "Land use/land cover change detection combining automatic processing and visual interpretation," *European Journal of Remote Sensing*, vol. 50, no. 1, pp. 626-635, 2017, doi: <https://doi.org/10.1080/22797254.2017.1387505>.
- [5] D. Dutta, A. Rahman, S. Paul, and A. Kundu, "Changing pattern of the urban landscape and its effect on land surface temperature in and around Delhi," *Environmental monitoring and assessment*, vol. 191, pp. 1-15, 2019, doi: <https://doi.org/10.1007/s10661-019-7645-3>.
- [6] S. A. Al-Tamimi, "The Role of Urban Sprawl on Agricultural Uses of Land Surrounding the City of Baghdad," *Journal of Engineering and Sustainable Development*, vol. 18, no. 6, pp. 19-44, 2014. [Online]. Available: <https://jeasd.uomustansiriyah.edu.iq/index.php/jeasd/article/view/874>.
- [7] A. D. Abed, "Use of Geographic Information Systems in Controlling Land Use (Study Area -Samarra City)," *Journal of Engineering and Sustainable Development*, vol. 23, no. 5, 2019, doi: <https://doi.org/10.31272/jeasd.23.5.16>.
- [8] Shahfahad, B. Kumari, M. Tayyab, H. T. Hang, M. F. Khan, and A. Rahman, "Assessment of public open spaces (POS) and landscape quality based on per capita POS index in Delhi, India," *SN Applied Sciences*, vol. 1, pp. 1-13, 2019, doi: <https://doi.org/10.1007/s42452-019-0372-0>.
- [9] A. Rahman *et al.*, "Performance of different machine learning algorithms on satellite image classification in rural and urban setup," *Remote Sensing Applications: Society and Environment*, vol. 20, p. 100410, 2020, doi: <https://doi.org/10.1016/j.rsase.2020.100410>.
- [10] V. Sridhar and K. A. Anderson, "Human-induced modifications to land surface fluxes and their implications on water management under past and future climate change conditions," *Agricultural and Forest Meteorology*, vol. 234-235, pp. 66-79, 2017, doi: <https://doi.org/10.1016/j.agrformet.2016.12.009>.
- [11] A. M. Abdi, "Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data," *GIScience & Remote Sensing*, vol. 57, no. 1, pp. 1-20, 2020, doi: <https://doi.org/10.1080/15481603.2019.1650447>.
- [12] S. Talukdar *et al.*, "Land-Use Land-Cover Classification by Machine Learning Classifiers for Satellite Observations—A Review," *Remote Sensing*, vol. 12, no. 7, p. 1135, 2020, doi: <https://doi.org/10.3390/rs12071135>.
- [13] A. Schneider, "Monitoring land cover change in urban and peri-urban areas using dense time stacks of Landsat satellite data and a data mining approach," *Remote Sensing of Environment*, vol. 124, pp. 689-704, 2012, doi: <https://doi.org/10.1016/j.rse.2012.06.006>.
- [14] D. ARIKAN and F. YILDIZ, "Sentinel-2 Uydu Görüntülerinde Destek Vektör Makinesi ve Rastgele Orman Algoritmaları Kullanılarak Pikselle Tabanlı Arazi Sınıflandırması," *Osmaniye Korkut Ata Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, vol. 6, no. 2, pp. 1243-1260, 2023. [Online]. Available: <https://dergipark.org.tr/en/pub/okufbed/issue/78780/1123426>.
- [15] Y. Zhang, G. Wang, F.-I. Chung, and S. Wang, "Support vector machines with the known feature-evolution priors," *Knowledge-Based Systems*, vol. 223, p. 107048, 2021, doi: <https://doi.org/10.1016/j.knsys.2021.107048>.
- [16] L. Breiman, "Random forests," *Machine learning*, vol. 45, pp. 5-32, 2001, doi: <https://doi.org/10.1023/a:1010933404324>.
- [17] S. A. Woznicki, J. Baynes, S. Panlasigui, M. Mehaffey, and A. Neale, "Development of a spatially complete floodplain map of the conterminous United States using random forest," *Science of the total environment*, vol. 647, pp. 942-953, 2019, doi: <https://doi.org/10.1016/j.scitotenv.2018.07.353>.
- [18] M. A. Z. Aguilera, "Classification Of Land-Cover Through Machine Learning Algorithms For Fusion Of Sentinel-2a And Planetscope Imagery," in *2020 IEEE Latin American GRSS & ISPRS Remote Sensing Conference (LAGIRS)*, 2020: IEEE, pp. 246-253, doi: <https://doi.org/10.1109/lagirs48042.2020.9165632>.
- [19] E. A. Freeman, G. G. Moisen, J. W. Coulston, and B. T. Wilson, "Random forests and stochastic gradient boosting for predicting tree canopy cover: comparing tuning processes and model performance," *Canadian Journal of Forest Research*, vol. 46, no. 3, pp. 323-339, 2016, doi: <https://doi.org/10.1139/cjfr-2014-0562>.
- [20] Q. Guo *et al.*, "Urban tree classification based on object-oriented approach and random forest algorithm using unmanned aerial vehicle (uav) multispectral imagery," *Remote Sensing*, vol. 14, no. 16, p. 3885, 2022, doi: <https://doi.org/10.3390/rs14163885>.
- [21] S. a. Ibrahim, "Improving land use/cover classification accuracy from random forest feature importance selection based on synergistic use of sentinel data and digital elevation model in agriculturally dominated landscape," *Agriculture*, vol. 13, no. 1, p. 98, 2022, doi: <https://doi.org/10.3390/agriculture13010098>.
- [22] J. Svoboda, P. Štych, J. Laštovička, D. Paluba, and N. Kobliuk, "Random Forest Classification of Land Use, Land-Use Change and Forestry (LULUCF) Using Sentinel-2 Data—A Case Study of Czechia," *Remote Sensing*, vol. 14, no. 5, p. 1189, 2022, doi: <https://doi.org/10.3390/rs14051189>.
- [23] P. Prasad, V. J. Loveson, P. Chandra, and M. Kotha, "Evaluation and comparison of the earth observing sensors in land cover/land use studies using machine learning algorithms," *Ecological Informatics*, vol. 68, p. 101522, 2022, doi: <https://doi.org/10.1016/j.ecoinf.2021.101522>.
- [24] I. Alimuddin, "The application of Sentinel 2B satellite imagery using Supervised Image Classification of Maximum Likelihood Algorithm in Landcover Updating of The Mamminasata Metropolitan Area, South Sulawesi," in *IOP Conference Series: Earth and Environmental Science*, 2019, vol. 280, no. 1: IOP Publishing, p. 012033, doi: <https://doi.org/10.1088/1755-1315/280/1/012033>.
- [25] M. F. Baig, M. R. U. Mustafa, H. binti Takaijudin, and M. T. Zeshan, "Comparative analysis of support vector machine and maximum likelihood classifications using satellite images of Selangor, Malaysia," in *2021 Third International Sustainability and Resilience Conference: Climate Change*, 2021: IEEE, pp. 405-409, doi: <https://doi.org/10.1109/IEEECONF53624.2021.9668109>.
- [26] H. M. Salih and N. M. Salih, "The implementation of image classification and analysis of Mrsd using three different classifiers: a Case Study of Newcastle-UK," in *Published in the Journal of the Second Engineering Scientific Conference of the College of Engineering/University of Diyala. For the period from*, 2015, vol. 16, pp. 521-537. [Online]. Available: <https://djes.info/index.php/djes/article/view/346>.
- [27] R. Sivagami, R. Krishankumar, and K. Ravichandran, "A Comparative Analysis of Supervised Learning Techniques for Pixel Classification in Remote Sensing Images," in *2018 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, 2018: IEEE, pp. 1-4, doi: <https://doi.org/10.1109/wispnet.2018.8538518>.
- [28] M. Üstüner, "DESTEK VEKTOR MAKİNELERİ YONTEMİ İLE ARAZI KULLANIMI SINIFLANDIRILMASINDA KERNEL FONKSİYONLARINA AIT KARSILASTIRMALI PARAMETRE DUYARLIK ANALİZİ: RAPİDEYE VE SPOT ORNEGI," 2013. [Online]. Available: <http://dspace.yildiz.edu.tr/xmlui/handle/1/2848>
- [29] M. ALTUN and M. TÜRKER, "Çoklu zamanlı Sentinel-2 görüntülerinden tarımsal ürün tespiti: Mardin-Kızıltepe örneği," *Afyon Kocatepe Üniversitesi Fen Ve Mühendislik Bilimleri Dergisi*, vol. 21, no. 4, pp. 881-899, 2021, doi: <https://doi.org/10.35414/akufemubid.890436>.
- [30] M. K. Sbahi, A. R. T. Ziboon, and K. I. Hassoon, "Evaluation of the Efficiency of Agricultural Production in the Pivotal Farms Utilizing Remote Sensing Techniques," *Journal of Engineering and Sustainable Development*, vol. 23, no. 4, pp. 86-99, 2019, doi: <https://doi.org/10.31272/jeasd.23.4.6>.
- [31] M. Dedeoğlu, L. Başayığıt, M. Yüksel, and F. Kaya, "Assessment of the vegetation indices on Sentinel-2A images for predicting the soil productivity potential in Bursa, Turkey," *Environmental Monitoring and Assessment*, vol. 192, pp. 1-16, 2020, doi: <https://doi.org/10.1007/s10661-019-7989-8>.
- [32] M. Gomroki, M. Hasanlou, and P. Reinartz, "STCD-EffV2T Unet: Semi Transfer Learning EfficientNetV2 T-Unet Network for Urban/Land Cover Change Detection Using Sentinel-2 Satellite Images," *Remote Sensing*, vol. 15, no. 5, p. 1232, 2023, doi: <https://doi.org/10.3390/rs15051232>.

- [33] E. Tunca and E. KÖKSAL, "Sentinel 2 Uydu Görüntülerinden Bitki Türlerinin Makine Öğrenmesi ile Belirlenmesi," *ÇOMÜ Ziraat Fakültesi Dergisi*, vol. 9, no. 1, pp. 189-200, 2021, doi: <https://doi.org/10.33202/comuagri.842202>.
- [34] R. G. Congalton and K. Green, *Assessing the accuracy of remotely sensed data: principles and practices*. CRC Press, 2019.
- [35] C. Cortes and V. Vapnik, "Support-vector networks," in *machine learning*, vol. 20, 1995, pp. 273-297.
- [36] V. N. Vapnik, "Estimation of dependences based on empirical data. 1982," *NY: Springer-Verlag*, 1995, doi: <https://doi.org/10.1007/0-387-34239-7>.
- [37] P. Addesso, R. Conte, M. Longo, R. Restaino, and G. Vivone, "SVM-based cloud detection aided by contextual information," in *2012 Tyrrhenian Workshop on Advances in Radar and Remote Sensing (TyWRRS)*, 2012: IEEE, pp. 214-221, doi: <https://doi.org/10.1109/tywrrs.2012.6381132>.
- [38] I. Lizarazo, "SVM-based segmentation and classification of remotely sensed data," *International Journal of Remote Sensing*, vol. 29, no. 24, pp. 7277-7283, 2008, doi: <https://doi.org/10.1080/01431160802326081>.
- [39] G. Mountrakis, J. Im, and C. Ogole, "Support vector machines in remote sensing: A review," *ISPRS journal of photogrammetry and remote sensing*, vol. 66, no. 3, pp. 247-259, 2011, doi: <https://doi.org/10.1016/j.isprsjprs.2010.11.001>.
- [40] K. M. Buddhiraju and I. A. Rizvi, "Comparison of CBF, ANN and SVM classifiers for object based classification of high-resolution satellite images," in *2010 IEEE international geoscience and remote sensing symposium*, 2010: IEEE, pp. 40-43, doi: <https://doi.org/10.1109/igarss.2010.5652033>.
- [41] P. Mather and B. Tso, *Classification methods for remotely sensed data*. CRC Press, 2016.
- [42] M. Denil, D. Matheson, and N. De Freitas, "Narrowing the gap: Random forests in theory and in practice," in *International conference on machine learning*, 2014: PMLR, pp. 665-673. [Online]. Available: <https://proceedings.mlr.press/v32/denil14.html>.
- [43] E. Adam, O. Mutanga, J. Odindi, and E. M. Abdel-Rahman, "Land-use/cover classification in a heterogeneous coastal landscape using RapidEye imagery: evaluating the performance of random forest and support vector machines classifiers," *International Journal of Remote Sensing*, vol. 35, no. 10, pp. 3440-3458, 2014, doi: <https://doi.org/10.1080/01431161.2014.903435>.
- [44] F. F. Camargo, E. E. Sano, C. M. Almeida, J. C. Mura, and T. Almeida, "A comparative assessment of machine-learning techniques for land use and land cover classification of the Brazilian tropical savanna using ALOS-2/PALSAR-2 polarimetric images," *Remote Sensing*, vol. 11, no. 13, p. 1600, 2019, doi: <https://doi.org/10.3390/rs11131600>.
- [45] L. Ma, M. Li, X. Ma, L. Cheng, P. Du, and Y. Liu, "A review of supervised object-based land-cover image classification," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 130, pp. 277-293, 2017, doi: <https://doi.org/10.1016/j.isprsjprs.2017.06.001>.
- [46] M. Belgiu and L. Drăguț, "Random forest in remote sensing: A review of applications and future directions," *ISPRS journal of photogrammetry and remote sensing*, vol. 114, pp. 24-31, 2016, doi: <https://doi.org/10.1016/j.isprsjprs.2016.01.011>.
- [47] A. Liaw and M. Wiener, "Classification and regression by randomForest," *R News*, vol. 2, no. 3, pp. 18-22, 2002. [Online]. Available: <https://journal.r-project.org/articles/RN-2002-022/RN-2002-022.pdf>.
- [48] Q. Feng, J. Gong, J. Liu, and Y. Li, "Flood mapping based on multiple endmember spectral mixture analysis and random forest classifier—The case of Yuyao, China," *Remote Sensing*, vol. 7, no. 9, pp. 12539-12562, 2015, doi: <https://doi.org/10.3390/rs70912539>.
- [49] N. Currit, "Development of a remotely sensed, historical land-cover change database for rural Chihuahua, Mexico," *International Journal of Applied Earth Observation and Geoinformation*, vol. 7, no. 3, pp. 232-247, 2005, doi: <https://doi.org/10.1016/j.jag.2005.05.001>.
- [50] B. A. Osunmadewa, W. Z. Gebrehiwot, E. Csaplovics, and O. C. Adeofun, "Spatio-temporal monitoring of vegetation phenology in the dry sub-humid region of Nigeria using time series of AVHRR NDVI and TAMSAT datasets," *Open Geosciences*, vol. 10, no. 1, pp. 1-11, 2018, doi: <https://doi.org/10.1515/geo-2018-0001>.
- [51] M. Pal and P. M. Mather, "An assessment of the effectiveness of decision tree methods for land cover classification," *Remote sensing of environment*, vol. 86, no. 4, pp. 554-565, 2003, doi: [https://doi.org/10.1016/s0034-4257\(03\)00132-9](https://doi.org/10.1016/s0034-4257(03)00132-9).
- [52] C. Smith and N. Brown, "Erdas field guide. revised and expanded," ed: ERDAS®, Inc Atlanta, 1999.
- [53] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An introduction to statistical learning*. Springer, 2013.
- [54] S. Myeong, D. J. Nowak, P. F. Hopkins, and R. H. Brock, "Urban cover mapping using digital, high-spatial resolution aerial imagery," *Urban ecosystems*, vol. 5, pp. 243-256, 2001. [Online]. Available: <https://link.springer.com/article/10.1023/A:1025687711588>.
- [55] R. G. Congalton, "A review of assessing the accuracy of classifications of remotely sensed data," *Remote sensing of environment*, vol. 37, no. 1, pp. 35-46, 1991, doi: [https://doi.org/10.1016/0034-4257\(91\)90048-b](https://doi.org/10.1016/0034-4257(91)90048-b).
- [56] D. Lu and Q. Weng, "A survey of image classification methods and techniques for improving classification performance," *International Journal of Remote Sensing*, vol. 28, no. 5, pp. 823-870, 2007, doi: <https://doi.org/10.2139/ssrn.3349696>.
- [57] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263-1284, 2009, doi: <https://doi.org/10.1109/tkde.2008.239>.
- [58] B. Shivakumar and S. Rajashekararadhya, "Investigation on land cover mapping capability of maximum likelihood classifier: a case study on North Canara, India," *Procedia computer science*, vol. 143, pp. 579-586, 2018, doi: <https://doi.org/10.1016/j.procs.2018.10.434>.
- [59] O. Buck, V. E. G. Millán, A. Klink, and K. Pakzad, "Using information layers for mapping grassland habitat distribution at local to regional scales," *International Journal of Applied Earth Observation and Geoinformation*, vol. 37, pp. 83-89, 2015, doi: <https://doi.org/10.1016/j.jag.2014.10.012>.
- [60] Y. Zhao and Z. Zhu, "ASI: An artificial surface Index for Landsat 8 imagery," *International Journal of Applied Earth Observation and Geoinformation*, vol. 107, p. 102703, 2022, doi: <https://doi.org/10.1016/j.jag.2022.102703>.
- [61] A. Abdi, "Decadal land-use/land-cover and land surface temperature change in Dubai and implications on the urban heat island effect: A preliminary assessment," 2019, doi: <https://doi.org/10.31223/osf.io/w79ea>.