TURKGEO
TURKISH JOURNAL OF GEOSCIENCES

ISSN: *2717-7696*
Research Article
Vol; 5, Issue; 1, pp. 12-28, 2025
DOI: *10.48053/turkgeo.1613103*
Journal homepage:
https://dergipark.org.tr/en/pub/turkgeo

Assessment of the Performance of Sentinel-2A MSI and Landsat 9 OLI Images in Land Cover/Use Classification by Comparing Machine Learning Algorithms: A Case Study of Soma District, Turkiye

Sajaweddin Sadid[*1], **Kaan Kalkan** [2]

[1] *Eskisehir Technical University, Institute of Earth and Space Sciences, Remote Sensing and Geographic Information Systems, Eskisehir, Turkey*
[2]*TÜBİTAK Space Technologies Research Institute, Ankara, Turkey*

Keywords
Land cover
Remote sensing
Machine learning algorithms
Classification

ABSTRACT

The combination of rapid, uncontrolled population growth and economic and industrial development has significantly accelerated land use/land cover (LULC) changes. Assessing these changes is one of the most effective ways to understand and manage land transformation. The advancement of remote sensing technology and increased accessibility to satellite data have made it more feasible to produce accurate and up-to-date LULC maps through the development of classifier algorithms. This has enabled better assessment and management of ecosystem and land use changes. The main objective of this study is to evaluate the performance of four machine learning algorithms—RF, SVM, CART, and GTB—using Sentinel-2 and Landsat 9 satellite images for the Soma district of Turkiye, on the Google Earth Engine (GEE) platform. In the study, a kernel function was applied to the SVM algorithm. Downloaded satellite images were visually inspected, and Google Earth Pro images were utilized to create training and test samples. Sentinel-2 and Landsat 9 images were classified using these training data and machine learning algorithms on the GEE platform. In the evaluation of the results, an error matrix was generated for the classified images, using the test samples for validation. The evaluation showed that the overall accuracy of the SVM algorithm, using the kernel function, was 92.6% for Sentinel-2 and 87% for Landsat 9, placing it third in terms of accuracy. The GTB algorithm provided the highest overall accuracy, with 94.4% for Sentinel-2 and 88.3% for Landsat 9. The RF algorithm achieved 93.2% accuracy for Sentinel-2 and 87% for Landsat 9, matching the accuracy of SVM for Landsat 9. CART demonstrated the lowest performance, with 86.4% accuracy for Sentinel-2 and 91.4% for Landsat 9. Additionally, Sentinel-2 imagery performed better than Landsat 9 across all algorithms due to its higher spatial resolution and spectral characteristics. This study provides valuable insights for local and provincial planners, authorities, and decision-makers regarding proper land management and the production of reliable LULC maps, especially in mining regions.

1. INTRODUCTION

The dynamic process of land use/land cover change worldwide is an indispensable source of concern, signaling global environmental change. It represents one of the most remarkable regional anthropogenic degradations of the environment. At the same time, even though land is an extremely precious natural resource and has all the material wealth for mankind, the extraction of natural resources through mining activities inevitably leads to changes in land use and land cover (Garai & Narayana, 2018). In this context, knowledge of land use and land cover at various scales contributes to the study of a wide range of current and future global phenomena, including drought, flooding, erosion, and climate change.

* **Corresponding Author**

*(qarluq.sadid.2018@gmail.com) 0000-0002-0031-4531
(kalkaan@gmail.com) ORCID ID 0000-0002-2732-5425

Received: 04/01/2025; Accepted: 25/03/2025

Cite this article

Sadid, S., & Kalkan, K. (2025). Assessment of the Performance of Sentinel-2A MSI and Landsat 9 OLI Images in Land Cover/Use Classification by Comparing Machine Learning Algorithms: A Case Study of Soma District, Turkiye. Turkish Journal of Geosciences 5(1), 12-28.

Continuous and accurate analysis of land use and land cover (LULC) is an integral part of sustainable development activities in any region. Earth science and remote sensing organizations have a traditional interest in accurate and up-to-date monitoring, providing valuable data for understanding the relationships between people and their environment (Praticò et al., 2021). Detailed and accurate land cover maps are recognized as an important input in various scientific studies, such as climate change impacts on streamflow and water budgets (Sridhar et al., 2019|; Kumar et al., 2022), geomorphology (Sujatha and Sridhar, 2018; Jayappa et al., 2006), groundwater management (Xiao et al., 2022b), social information management of natural resources (Sridhar et al., 2021) and agricultural land monitoring (Sierra-Soler et al., 2015). LULC maps contribute to the control of land degradation in agriculture and watershed management in general or because of mining activities (Cihlar, 2000; Chen et al., 2020).

Remote sensing satellite imagery is one of the primary data sources that facilitates access to vast amounts of data and information for Earth observation and monitoring. Remote sensing techniques are widely used to study urban expansion, land degradation caused by mining activities, and changes in land cover. Satellites provide a variety of sensors, image resolutions, collection methods, and spatiotemporal characteristics to observe the Earth from the surface to space (Mangkhaseum and Hanazawa, 2021). Today, with tools such as Google Earth Engine (GEE), remote sensing, GIS technology, Google Earth, and machine learning algorithms, the spatial mapping of land use, land cover, and other Earth surface features can now be done more quickly and accurately (Pande, 2022).

Producing low-resolution land cover maps over large areas requires processing large volumes of data, which demands significant storage capacity, substantial processing power, and flexibility in using various approaches. These needs have been met with the release of Google Earth Engine (GEE), which makes the technology freely available to everyone (Gorelick et al., 2017). GEE provides free access to satellite imagery from sources such as Landsat, Sentinel, and MODIS.

Recently, with population growth and industrial development, mining activities have been increasing to meet energy demands, and the use of modern mining techniques and heavy equipment worldwide, especially in Turkiye, is causing significant changes in land cover, ecology, and hydrology (Garai and Narayana, 2018). In general, mining activities lead to changes in topography (Manna and Maiti, 2014) and drainage patterns (Manna and Maiti, 2016; Kříbek et al., 2023), as well as environmental impacts such as landscape alteration and degradation (Tadesse et al., 2017; Y. Shi et al., 2024), soil erosion and degradation (Xiao et al., 2022a), and broader environmental changes (Tripathi et al., 2024; Mehta et al., 2024).

As a developing country, Turkiye's energy needs are increasing daily. Consequently, to meet the demand for energy production through thermal power plants, the coal mining industry is steadily ramping up its production, driven by the reliance on coal for electricity generation. The Soma coal basin, situated in the Soma district of Manisa province, contains one of the most significant coal reserves in the country. This basin is vital for lignite coal production and contributes substantial value to the Turkish economy. It plays a critical role in fulfilling Turkiye's energy requirements and serves as the primary fuel source for thermal power plants in the region. However, sustainable mining practices and effective environmental management strategies are essential for ensuring the long-term availability of coal and safeguarding the environmental health of the basin. In this context, it is crucial to study the impact of mining on land use and land cover changes in the region to mitigate the environmental impacts of mining activities and facilitate efficient land management and decision-making processes (Wang et al., 2019).

When we examine the literature, the researcher has used various remote-sensing images for LULC classification. These studies focused on the classification and extraction of features such as cultivated land, shrubland, agriculture, bare land, and water bodies (Wang et al., 2019; Pande, 2022; Praticò et al., 2021). Lu and Weng (2007) showed that the results of land use/cover mapping depend not only on the suitability of the imagery but also on the appropriate choice of classification methods. Various classification approaches have been developed and evaluated for land use/cover assessment using remotely sensed data. Various classifiers have been used in land use/cover assessment, including unsupervised algorithms such as the ISODATA clustering algorithm or K-means, parametric supervised algorithms such as maximum likelihood (ML), and machine learning algorithms such as artificial neural networks (ANN), k-nearest Neighbors (kNN), decision trees (DT), support vector machine (SVM), random forest (RF), and Classification and Regression Trees (CART) (Friedl and Brodley, 1997; Waske and Braun, 2009; Li et al., 2014; Shao and Lunetta, 2012; Chen et al., 2020) (Oo et al., 2022; Zhao et al., 2024). Nonparametric approaches, especially machine learning-based algorithms, have received significant attention in remote sensing studies over the last decade (Thanh Noi and Kappas, 2018; Oo et al., 2022).

Many studies have been conducted to determine the best algorithm for land use/land cover classification by evaluating their performance. However, the results obtained can vary significantly. In a land cover classification study conducted by (Dixon and Candade, 2008) using Landsat TM data, SVM performed well, while Maximum Likelihood performed much worse. Moreover, previous studies

have shown that SVM outperforms traditional classification algorithms such as maximum likelihood (ML), k-nearest neighbor (kNN), and neural networks (NN) in hyperspectral remote sensing classification (Huang et al., 2002; Melgani and Bruzzone, 2004; Pal and Mather, 2005). Oo et al. (2022) compared four machine learning algorithms, namely Maximum Likelihood, Random Forest, Support Vector Machine, and CART for Land Cover Classification in Gold Mining. As a result of the study, the RF algorithm was found to have the highest accuracy. Thanh Noi and Kappas. (2018) used RF, K-Nearest Neighbor, and SVM algorithms for land cover classification. At the end of the study, SVM was found to have the highest accuracy among the three machine learning algorithms.

There is no denying that in a mining area where many human activities take place, there will be many disturbances that can change land use/land cover. In this context, it has been proven that there is a critical relationship between long-term open-cast mining activities and land use/land cover changes in the mining basin (Paraskevis et al., 2021). Vorovencii. (2024) applied Maximum Likelihood, Minimum Distance, and SVM algorithms using Landsat image series in the assessment of long-term land cover changes in the Jiului Valley mining basin in Romania. As a result of the classification, forest, residential areas, mining areas, and water bodies in the region have increased, while pasture, agricultural land, and garbage dump areas have decreased.

A review of the literature reveals a scarcity of studies that compare and evaluate the performance of machine learning algorithms utilizing Landsat and Sentinel-2 imagery, particularly in mining regions. This gap underscores the importance of comparing and assessing the effectiveness of four specific classifiers: Random Forest (RF), Support Vector Machine (SVM), Classification and Regression Trees (CART), and Gradient Boosting Trees (GBT) in the context of land use/land cover (LULC) mapping. With the growing demand for reliable LULC data derived from extensive satellite imagery, understanding the performance of various machine learning methods on widely utilized cloud-based platforms, such as Google Earth Engine (GEE), has become increasingly crucial. In this study, we employ four machine learning classifiers—RF, SVM, GBT, and CART—to systematically evaluate the performance of Landsat-9 OLI and Sentinel-2A satellite imagery for LULC classification in the Soma district of Manisa, Turkiye, from June 1 to August 30, 2023. The objective of this research is twofold: to determine which satellite imagery provides superior performance in LULC classification and to identify which of the machine learning classifiers yields the best classification results.

2. STUDY AREA

This study focuses on the Soma district, which covers a total area of 838.9 km^2. As a coal basin, the Soma district faces various environmental challenges, including mining waste, pollution of water resources, and soil contamination. Additionally, it was selected as the study area for several reasons, such as the expansion of urban areas due to the increasing population. Soma is in Manisa province in the Aegean Region of Turkiye, positioned at approximately 39°11′18″ north latitudes and 27°36′32″ east longitudes (Figure 1). As one of the most important coal basins in Turkiye, Soma attracts attention for its natural beauty and historical richness, holding both local and regional significance. Situated 100 km northeast of Izmir and 50 km from the Aegean coast, Soma's strategic location plays a vital role in its economic and cultural activities. Its proximity to Izmir and distance from the Aegean coast enhance Soma's prominence, making it a distinct and noteworthy region.

The physical environment of Soma has been drastically altered by ongoing lignite exploitation in the open coal basins to the northeast and southwest, as well as by the thermal power plant operating near the city center (Karadağ, 2006). While this change has resulted in a series of environmental problems due to mining activities and misuse, it has also significantly impacted the socio-economic structure of the city, introducing a new urban population with diverse urban life tendencies and expectations. The elevation of Soma varies between 59 m and 1209 m above sea level. The main geological formations in the area include clay, marn, and limestone, along with andesite and basalt rocks (https://eli.tki.gov.tr/soma). Due to its location under the influence of the Mediterranean climate, the average annual temperature in Soma is estimated to be between 15.3 °C and 16.1 °C, with annual precipitation totaling 554.4 mm. Soma's landscape structure and land cover dynamics exhibit complex patterns, comprising forests, agricultural lands, shrubland, open-pit coal mining activities, settlements, bare areas, and water bodies.

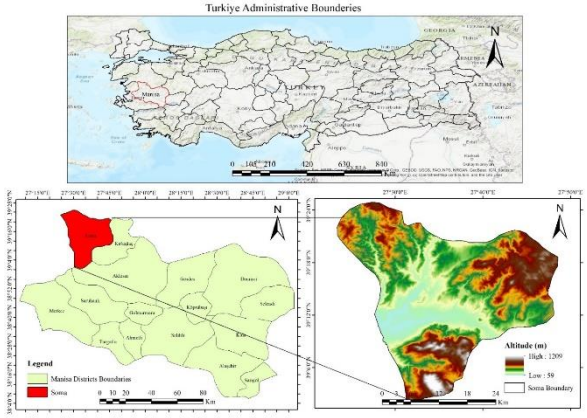

Figure 1. Geographical location of the study

3. MATERIAL and METHODS

3.1. Dataset

A vast amount of Earth observation data (EOD) from the past 40 years, including popular satellite imagery such as Sentinel, Landsat, and MODIS, as well as other geospatial data like climate and demographic information, is stored on the cloud-based Google Earth Engine (GEE) platform. Landsat and Sentinel data are accessible through GEE from the U.S. Geological Survey (USGS) and the European Space Agency (ESA). In this study, Sentinel-2 MSI and Landsat 9 OLI satellite images from 2023 were used to evaluate four machine learning classifiers for land use/land cover classification. To capture the period of vegetation growth and obtain cloud-free satellite images the months of June to August were selected and averaged. The Landsat 5 and Landsat 8 sensors have a spatial resolution of 30 meters and a swath width of 185 kilometers (http://landsat.gsfc.nasa.gov). Sentinel-2 sensors feature spatial resolutions of 10, 20, and 60 meters, depending on the spectral band, with a swath width of 290 kilometers (https://sentinel.esa.int). Table 1 below shows the characteristics of the satellite imagery.

Table 1. The primary features of Landsat-9 and Sentinel-2

Sensor	Collection	Bands	Spectral Range (μm)	Pixel Size (m)	Spectral Resolution	Radiometric Resolution (bit)	Revisit time (days)	Cloud Cover (%)
Landsat-9	Surface Reflectance Level 2, Collection 2, Tier 1	B1:Coastal/aerosol	0.433–0.453	30	11 Bands	16	16	> 5
		B2: Blue	0.450–0.515	30				
		B3: Green	0.525–0.600	30				
		B4: Red	0.630–0.680	30				
		B5: NIR	0.845–0.885	30				
		B6: SWIR1	1.560–1.660	30				
		B7: SWIR2	2.100–2.300	30				
		B8: Panchro.	0.500–0.680	15				
		B9: cirrus	1.360–1.390	30				
		B10: TIRS 1	10.30–11.30	100				
		B11: TIRS2	11.50–12.50	100				
Sentinel-2	Surface Reflectance Multi Spectral Instrument, Level 2A	B1: Coastal	0.433–0.453	60	13 Bands	12	5	> 5
		B2: Blue	0.457–0.522	10				
		B3: Green	0.542–0.577	10				
		B4: Red	0.650–0.680	10				
		B5: Red Edge	0.679–0.718	20				
		B6: Red Edge	0.732–0.747	20				
		B7: Red Edge	0.773–0.793	20				
		B8: NIR	0.784–0.899	10				
		B8A:Red Edge	0.855–0.885	20				
		B9: Water vapor	0.935–0.955	60				
		B10: Cirrus	1.36–1.39	60				
		B11: SWIR1	1.565–1.655	20				
		B12: SWIR2	2.10–2.28	20				

3.2. Image Preprocessing

In the Google Earth Engine (GEE) platform, machine learning algorithms were used comparatively to evaluate the performance of Landsat 9 OLI and Sentinel-2A MSI imagery to create a land cover/use map of the Soma district in Manisa province, Turkiye, which is the study area, for the months of June to August 2023 (Figure 2). Image preprocessing in the GEE platform is an important step that allows for image cropping, cloud masking, and the creation of topographically corrected image composites. Following image enhancement, a crucial step is to extract the normalized difference vegetation index (NDVI), normalized difference water index (NDWI), and normalized difference built-up index (NDBI), which are considered inputs for the classification algorithms. These indices are then combined with the spectral bands. Therefore, NDVI, NDWI, and NDBI are calculated based on the band values of Landsat 5, Landsat 8, and Sentinel-2 images according to the equations provided below, and the resulting maps are shown in figure 2.

The corrected Landsat and Sentinel-2 images with the lowest amount of cloud cover were used as the initial inputs for classification. The first step after importing the satellite data into Google Earth Engine (GEE) is to remove cloud shadows and cloud cover. Using a cloud mask, pixels contaminated by clouds or those with cloud-free conditions were eliminated from all available images (Mateo-García et al., 2018). In the second stage, Landsat and Sentinel data for each selected year were combined into a single image using a mean filter to create a composite image. Each pixel is assigned an average value from the entire image stack, resulting in one composite image for the entire collection. In this study, the normalized difference vegetation index (NDVI), normalized difference water index (NDWI), and normalized difference built-up index (NDBI) were calculated for the selected years and combined with the spectral bands as additional spectral features to enhance classification accuracy. NDVI is calculated using the near-infrared (NIR) and red bands (Townshend & Justice, 2007), NDWI is calculated using the green and shortwave infrared (SWIR)

bands (McFeeters, 1996), and NDBI is calculated using the SWIR and NIR bands (Zha et al., 2003), as shown in the following equations.

$$NDVI = \frac{NIR - RED}{NIR + RED} \quad (1)$$

$$NDWI = \frac{GREEN - SWIR}{GRESSN + SWIR} \quad (2)$$

$$NDBI = \frac{SWIR - NIR}{SWIR + NIR} \quad (3)$$

As a final step, the Random Forest (RF), Classification and Regression Trees (CART), Support Vector Machine (SVM), and Gradient Tree Boosting machine learning algorithms in GEE were used to train the classifiers for Landsat -9 OLI and Sentinel-2 images.

3.3. Training and Validation Sample Datasets

In this study, a set of geometric points (features) for each category (Water, Forest Area, Settlement Area, Agricultural Area, Vegetation, Bare Area, and Mining Area) was collected based on visual interpretation of the original Landsat and Sentinel-2 composite images, as well as high-resolution images from Google Earth. Training data were then generated using this collection of features for each category. This training data was divided into 80% for training the machine learning classifier models and 20% for testing to evaluate classification accuracy.

Table 2. Land cover classes are defined in this study

Class	Number of Point	Description
Water	100	Includes rivers, lakes, and other water bodies.
Forest area	107	Land dominated by trees.
Residential area	107	Areas used for residential, commercial, industrial, or mixed-use purposes.
Agriculture area	108	Includes all cultivated land and areas used for agricultural purposes.
Vegetation	100	Grassland areas including shrubs, meadows, and pastures.
Bare land	103	Typically includes soil, sand, or rock with minimal vegetation.
Mining Area	104	Designated areas for mining activities, including coal, sand, gravel pits, and other extractive industries.
Total	**729**	

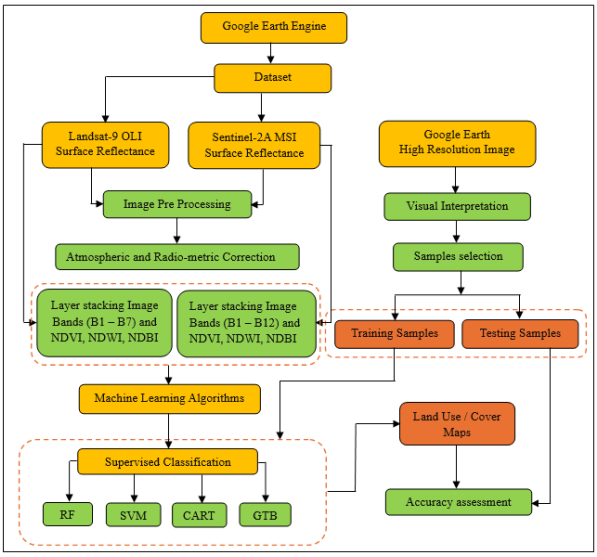

Figure 2. Flowchart for process of LULC classification within the GEE Platform

3.4. Supervised Classification Models

After collecting the training data and incorporating features that store the known class labels and predictor values, the classifiers need to be trained. The Google Earth Engine (GEE) platform offers several classifier packages for supervised classification using machine learning algorithms. However, in this study, four classifiers—Random Forest (RF), Support Vector Machine (SVM), Classification and Regression Trees (CART), and Gradient Tree Boost (GTB)—were used to evaluate the performance of Sentinel-2 and Landsat 9 data in land cover and land use classification. At this stage, the machine learning algorithms were trained using the collected training data and applied to the Landsat-9 OLI and Sentinel-2A composite image bands to produce a classified map of the study area. Below is an overview of each classifier algorithm used in this study.

Random Forest (RF): The Random Forest (RF) algorithm is rooted in the theory of bagging ensemble learning, introduced by Prof. Leo Breiman at the University of California in 1996 (Breiman, 2001). The RF algorithm has been extensively applied to solve various environmental problems and has demonstrated versatility in handling diverse data types, including satellite imagery and digital datasets (Abdullah et al., 2017). Due to its robustness in feature selection, RF is recognized as an effective and reliable supervised machine-learning technique (Xu & Cheng, 2024). to build the RF classifier model, two critical parameters are required: (1) n – the number of decision trees, commonly referred to as "trees," and (2) m – the number of attributes considered at each split, known as "trials" (Oo et al., 2022). The classification decision is made through a majority vote, where each decision tree contributes to the result based on its accuracy. The final

classification is determined by aggregating the votes from all trees in the forest. The following formula is used to determine the final classification result (Xu & Cheng, 2024).

$$H(x)=\text{voting} \sum_{i=1}^{k} I(h_i(x) = \gamma \quad (4)$$

Here, H(x) is the classification combination model; hi is the classification model of the decision tree; I() is the indicator function (this value is considered to be 1 when present in the parameter set and 0 when not present in the set) and γ represents the target variable (or output variable).

Support Vector Machine (SVM): Support Vector Machine (SVM) is a non-parametric algorithm first proposed by Vapnik and Chervonenkis (Vapnik & Chervonenkis, 1971). The SVM method constructs a hyperplane based on the largest margin between the given training sample sets and then classifies the segmented objects into recognized LULC (Land Use/Land Cover) classes. According to Knorn et al. (2009) and Shi & Yang (2015), the radial basis function (RBF) kernel of the SVM classifier is widely used in land cover classification studies and demonstrates strong performance. When implementing an SVM classifier with an RBF kernel, two key parameters need to be tuned: the optimal cost parameter (C) and the kernel width parameter (γ) (Qian et al., 2015; Ballanti et al., 2016). The C parameter controls the trade-off between allowing misclassification in the training data and creating a simpler decision boundary, thus adjusting the rigidity of the classifier. The γ parameter influences the smoothness of the class-separating hyperplane (Melgani & Bruzzone, 2004). While larger values of γ can lead to overfitting, they also affect the shape of the class-separating hyperplane, potentially impacting classification accuracy (Huang et al., 2010).

Classification and Regression Tree (CART): Breiman et al. developed the decision tree (DT) model, a widely used non-parametric method (Breiman et al., 1984). The classification and regression tree (CART) algorithm is a particularly common decision tree used for various purposes. A tree is constructed in CART through a binary iterative partitioning process, where the training sample set is split into subsets based on an attribute value test, and this process is repeated for each resulting subset (Shao & Lunetta, 2012; Oo et al., 2022). The tree-growing process concludes when further partitioning of subsets is no longer feasible. In the CART algorithm, the maximum depth of the tree is a crucial tuning parameter that determines the model's complexity. Generally, a greater depth allows the creation of a more complex tree, which may improve overall classification accuracy. However, an excessive number of nodes can lead to overfitting, where the model becomes too tailored to the training data, potentially reducing its generalization capability.

Gradient Tree Boost (GTB): Gradient Tree Boosting (GTB) is a powerful ensemble machine-learning technique used for regression and classification problems. It produces a prediction model typically in the form of an ensemble of weak prediction models using decision trees and has demonstrated significant success across a wide range of practical applications (Natekin & Knoll, 2013). However, the GTB classifier limits the complexity of individual decision trees by restricting them to weaker prediction models. The algorithm minimizes the loss function step-by-step through gradient descent optimization, improving classification accuracy by iteratively combining weak learners into a stronger ensemble of trees (Friedman, 2002). Unlike other ensemble classifiers, GTB adapts to the residuals of the regression tree at each iteration by using negative gradient loss values (Ouma et al., 2023). This method stochastically reduces the correlation between trees by generating new trees based on a selected training subset of the data.

4. RESULTS

4.1. Accuracy Assessment and Comparisons

An essential phase in all classification projects is the validation and accuracy assessment stage. This phase is a critical component of modeling and mapping, used to evaluate the effectiveness and scientific significance of the classifier models. The purpose of accuracy assessment is to compare the classified image with another data source, typically ground truth data. To assess the performance of the classifier models, confusion or error matrices are constructed by cross-referencing the classified land cover/use results with the test samples. Four commonly used metrics in remote sensing image classification—producer accuracy, user accuracy, overall accuracy, kappa coefficient—and the F1-score (equations 5–9) are used to evaluate the accuracy of the classified land cover/use (Hong et al., 2023). Producer Accuracy measures the degree of correctness and refers to the proportion of instances that truly belong to a specific class among all instances classified as that class. On the other hand, User Accuracy is the proportion of instances classified as a particular class relative to all instances that genuinely belong to that class (Ouma et al., 2023). Overall Accuracy represents the ratio of correctly classified instances to the total number of instances. The Kappa index assesses the level of agreement between the predicted and actual classes while accounting for the likelihood of correct classification occurring by random chance. Finally, the F1-measure, which is the harmonic mean of Producer and User Accuracy, is calculated to evaluate performance at both the classifier and class levels.

$$UA = \frac{K_{ii}}{K_{i+}} \quad (5)$$

$$PA = \frac{K_{jj}}{K_{j+}} \quad (6)$$

$$OA = \frac{\sum_{i=1}^{n} K_{ii}}{K_{i+}} \quad (7)$$

$$K = \frac{T \sum_{i=1}^{n} K_{ii} - \sum_{i,j=1}^{n} (K_{i+} K_{j+})}{T^2 - \sum_{i,j=1}^{n} (K_{i+} K_{j+})} \quad (8)$$

$$F1 - Score = 2 \times \frac{UA \times PA}{UA + PA} \quad (9)$$

Where PA = producer accuracy; UA = user accuracy; OA = overall accuracy; K = Kappa index; n = number of classes; Kii = number of correctly classified classes; Ki+ = i number of pixels in row and Kj+ =i number of pixels in column i; and T = number of pixels used for accuracy assessment.

4.2. Land Cover Classification

In this study, Landsat-9 OLI and Sentinel-2A satellite imagery were comparatively applied on the Google Earth Engine (GEE) platform for land cover mapping of the Soma district in Turkiye using supervised classification methods, including Random Forest (RF), Support Vector Machine (SVM), Classification and Regression Tree (CART), and Gradient Tree Boosting (GTB) algorithms. The spectral bands of Landsat-9 OLI and Sentinel-2A surface reflectance images, acquired between June 1, 2023, and August 30, 2023, were utilized for land cover classification. Due to cloud cover, the mean composite of the images within this period was used. The classification was conducted using all Near-Infrared (NIR), Visible-Near Infrared (VNIR), and Shortwave Infrared (SWIR) bands, while Thermal Infrared (TIR) bands were excluded.

Additionally, the Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index (NDWI), and Normalized Difference Built-up Index (NDBI) were calculated and included as additional bands, as they significantly enhance the accuracy of land cover classification results (Figure 3). NDVI is commonly used to assess vegetation density, with values typically ranging from -1 to +1. In this study, the NDVI values varied between -0.625 and 0.936, as shown in Figure 3.

NDWI is employed to detect the presence of water, with values generally ranging from -1 to +1; in this study, NDWI values ranged from -0.857 to 0.800. NDBI is used to identify built-up areas and typically varies between -1 and +1; in this study, NDBI values ranged from -0.583 to 0.710 (Figure 3).

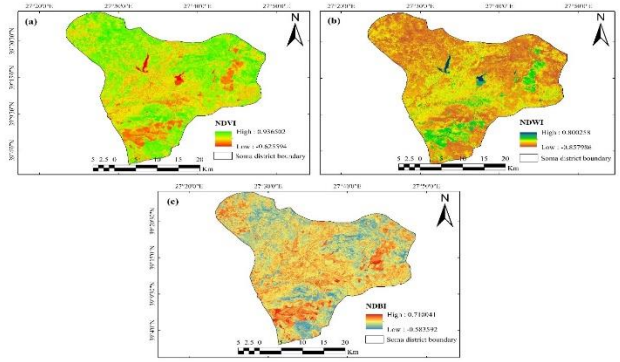

Figure 3. (a) Normalized Difference Vegetation Index, (b) Normalized Difference Water Index and (c) Normalized Difference Build up Index.

4.3. Comparison of Accuracy Analysis Results of Machine Learning Algorithms

To analyze the accuracy of land cover/use classification results from Sentinel-2 and Landsat-9 images using RF, SVM, CART, and GTB methods, five traditional metrics commonly used in remote sensing image classification were employed: producer accuracy, user accuracy, F1 score, overall accuracy, and kappa coefficient. These accuracy metrics are derived from the confusion matrix, which shows the agreement between the classified results and the reference data. Of the collected data, 80% was allocated for training and 20% for testing, while validation points were used to calculate the confusion matrix. The relevant results for Sentinel-2 and Landsat-9 images are presented in Tables 2 and 3 a,b below.

Table 3. Producer accuracy, user accuracy and F1 score values for machine learning classifier in land cover/use classification (PA = Producer Accuracy, UA = User Accuracy and F1 = F1-Score)

Class	(a) Landsat-9 OLI											
	RF			SVM			CART			GTB		
	UA	PA	F1	UA	PA	F1	UA	PA	F1	UA	PA	F1
Water	100	96	97.95	100	96	97.95	100	96	97.95	100	96	97.95
Forest area	93.33	86.66	93.33	86.66	86.66	86.66	100	86.66	92.85	87.5	93.33	90.32
Built area	80.95	76	73.91	82.60	76	79.16	89.47	68	77.27	95	76	84.44
Agriculture land	80.76	95.65	85.71	75.86	95.65	84.61	73.07	82.60	77.55	80.76	91.30	85.71
Shrubland	87.47	90.90	91.30	86.95	90.90	88.88	84.61	100	91.66	83.33	90.90	86.95
Bare land	89.47	90	87.17	100	90	94.73	88.88	80	84.21	90	90	90
Mining area	80.76	76	82.35	82.60	76	79.16	79.31	92	85.18	84	84	83.99
Class	**(b) Sentinel-2A**											
	RF			**SVM**			**CART**			**GTB**		
	UA	PA	F1	UA	PA	F1	UA	PA	F1	UA	PA	F1
Water	100	100	100	100	100	100	100	100	100	100	100	100
Forest area	100	96.29	98.11	96.29	96.29	98.11	100	96.29	100	100	96.29	100
Built area	95.45	100	95.87	87.5	100	93.33	90.9	95.23	95.2	100	90.47	100

Agriculture land	87.5	77.77	91.68	95.23	74.07	97.56	90.47	70.37	95.0	92	85.18	95.83
Shrubland	90.9	100	93.52	90.47	95	95.00	90.9	100	95.2	90.9	100	95.23
Bare land	82.6	86.36	88.92	80.76	95.45	89.36	74.07	90.9	85.1	86.95	90.9	93.02
Mining area	95.65	95.65	95.97	100	91.3	95.45	95.45	91.3	97.7	92	100	95.83

Table 3a compares the producer accuracy (PA), user accuracy (UA), and F1 score values for machine learning algorithms—Random Forest (RF), Support Vector Machine (SVM), Classification and Regression Trees (CART), and Gradient Tree Boosting (GTB)—in land cover/use classification using Landsat-9 OLI satellite images. As a result of the accuracy analysis, the water area class is classified correctly, as it has the highest values of user accuracy, producer accuracy, and F1 score across all classification algorithms. In the forest area class, the CART classifier shows the highest UA value (100), although its PA value is like those of the other methods. The RF classifier demonstrates a high F1 score (93.33) and generally provides balanced results. The SVM and GTB classifiers exhibit slightly lower performance compared to the others, with SVM showing lower UA and F1 score values.

In the residential area class, the GTB classifier achieves the highest UA (95), and its F1 score (84.44) is also the highest compared to the other methods. The CART classifier shows a lower PA (68) but maintains a high UA. Both the RF and SVM classifiers perform lower than the others, with RF exhibiting a particularly low F1 score.

In the agricultural area class, both the RF and GTB classifiers achieved the highest F1 score (85.71). SVM and CART classifiers show lower F1 scores, with CART having notably lower UA and PA values. In the shrubland class, CART achieves the highest F1 score (91.66) and PA value (100), while RF and SVM show similar results, with RF being slightly higher in terms of F1 score. The GTB classifier performed slightly lower than the others. In the bare land class, the SVM classifier outperforms the other classifiers, achieving higher UA (100), PA (90), and F1 score (94.73). In contrast, the CART classifier performs poorly, showing lower values for UA, PA, and F1. In the mining area class, the CART classifier performed better than the other classifiers, showing the highest PA (92) and F1 score (85.18) (Table 3 a,b).

When analyzing the producer accuracy (PA), user accuracy (UA), and F1 score values for land cover/use classification using Sentinel-2 satellite imagery, as presented in Table 3b, it is evident that the water area class is classified correctly, having the highest values of user accuracy, producer accuracy, and F1 score (100) across all classifiers. In the forest area class, all classifiers demonstrate high performance; however, CART and GTB excel, achieving the highest values (100) for both user accuracy and F1 score. In the residential area class, the GTB classifier attains the highest F1 score (100) and user accuracy (100), while RF shows the highest producer accuracy (100). In contrast, the SVM and CART classifiers perform worse than both RF and GTB in terms of user accuracy and F1 score. For the farmland class, SVM and CART classifiers achieve the highest F1 values of 97.56 and 95.0, respectively, but their user accuracy values of 74.07 and 70.37 are lower than those of the other classifiers. In the shrubland class, the GTB classifier outperforms the others, with user and producer accuracy and F1 values of 90.9, 100, and 95.23, respectively.

In the bare area class, the GTB classifier also surpasses the others, achieving user and producer accuracy and F1 values of 86.95, 90.9, and 93.02, respectively. In the mining area class, RF and CART classifiers yield more balanced results compared to the SVM and GTB classifiers, with F1 score values of 95.97 and 97.7, respectively (Table 3b). Overall, Landsat 9 and Sentinel 2 satellite images were classified more accurately, particularly in the water area class, where user accuracy, producer accuracy, and F1 score values were highest among the four classifiers used for land cover/use classification. Sentinel-2A satellite imagery outperforms Landsat-9 in land cover/use classification, as evidenced by its higher user accuracy, producer accuracy, and F1 score values across the four classifier methods. This superior performance may be attributed to Sentinel-2A's higher resolution and better spectral coverage.

Table 4. Confusion matrix, overall accuracy and kappa coefficient results for each classifier algorithm

(a) Confusion matrix, overall accuracy and kappa results for each classification algorithm on Landsat 9 OLI image.

ML Algorithms	Class	Water	Forest Area	Built Area	Agriculture Land	Shrubland	Bare land	Mining Area	Total
RF Classifier	Water	**24**	0	0	1	0	0	0	25
	Forest area	0	**14**	0	0	1	0	0	15
	Built area	0	0	**17**	2	0	1	5	25
	Agriculture land	0	0	0	**21**	1	1	0	23
	Shrubland	0	1	0	0	**21**	0	0	22
	Bare land	0	0	2	0	1	**17**	0	20
	Mining area	0	0	2	2	0	0	**21**	25
	Overall Accuracy	**0.870**							
	Kappa	**0.848**							
SVM Classifier	Water	**24**	0	0	1	0	0	0	25
	Forest area	0	**13**	0	0	2	0	0	15

	Built area	0	0	**19**	2	0	0	4	25
	Agriculture land	0	0	0	**22**	1	0	0	23
	Shrubland	0	2	0	0	**20**	0	0	22
	Bare land	0	0	1	1	0	**18**	0	20
	Mining area	0	0	3	3	0	0	**19**	25
	Overall Accuracy	**0.870**							
	Kappa	**0.848**							
CART Classifier	Water	**24**	0	0	0	0	0	1	25
	Forest area	0	**13**	0	0	2	0	0	15
	Built area	0	0	**17**	3	0	0	5	25
	Agriculture land	0	0	1	**19**	1	2	0	23
	Shrubland	0	0	0	0	**22**	0	0	22
	Bare land	0	0	1	2	1	**16**	0	20
	Mining area	0	0	0	2	0	0	**23**	25
	Overall Accuracy	**0.864**							
	Kappa	**0.841**							
GTB Classifier	Water	**24**	0	0	1	0	0	0	25
	Forest area	0	**14**	0	0	1	0	0	15
	Built area	0	0	**19**	2	0	0	4	25
	Agriculture land	0	0	0	**21**	1	2	0	23
	Shrubland	0	2	0	0	**20**	0	0	22
	Bare land	0	0	0	0	2	**18**	0	20
	Mining area	0	0	1	2	0	1	**21**	25
	Overall Accuracy	**0.883**							
	Kappa	**0.864**							

(b) Confusion matrix, overall accuracy and kappa results for each classification algorithm on Sentinel-2A image.

ML Algorithms	Class	Water	Forest Area	Built Area	Agriculture Land	Shrubland	Bare land	Mining Area	Total
RF Classifier	Water	**23**	0	0	0	0	0	0	23
	Forest area	0	**26**	0	0	1	0	0	27
	Built area	0	0	**21**	0	0	0	0	21
	Agriculture land	0	0	0	**21**	1	4	1	27
	Shrubland	0	0	0	0	**20**	0	0	20
	Bare land	0	0	0	3	0	**19**	0	22
	Mining area	0	0	1	0	0	0	**22**	23
	Overall Accuracy	**0.932**							
	Kappa	**0.921**							
SVM Classifier	Water	**23**	0	0	0	0	0	0	23
	Forest area	0	**26**	0	0	1	0	0	27
	Built area	0	0	**21**	0	0	0	0	21
	Agriculture land	0	0	1	**20**	1	5	0	27
	Shrubland	0	1	0	0	**19**	0	0	20
	Bare land	0	0	0	1	0	**21**	0	22
	Mining area	0	0	2	0	0	0	**21**	23
	Overall Accuracy	**0.926**							
	Kappa	**0.914**							
CART Classifier	Water	**23**	0	0	0	0	0	0	23
	Forest area	0	**26**	0	0	1	0	0	27
	Built area	0	0	**20**	0	0	0	1	21
	Agriculture land	0	0	0	**19**	1	7	0	27
	Shrubland	0	0	0	0	**20**	0	0	20
	Bare land	0	0	0	2	0	**20**	0	22
	Mining area	0	0	2	0	0	0	**21**	**23**
	Over Accuracy	**0.914**							
	Kappa	**0.899**							
GTB Classifier	Water	**23**	0	0	0	0	0	0	23
	Forest area	0	**26**	0	0	1	0	0	27
	Built area	0	0	**19**	0	0	0	2	21
	Agriculture land	0	0	0	**23**	1	3	0	27
	Shrubland	0	0	0	0	**20**	0	0	20
	Bare land	0	0	0	2	0	**20**	0	22
	Mining area	0	0	0	0	0	0	**23**	23
	Overall Accuracy	**0.944**							
	Kappa	**0.935**							

Table 4a above presents the error matrices, overall accuracy, and kappa coefficient for land cover/use classification of Landsat-9 OLI image using four different classifiers (RF, Support SVM, CART, and GTB). The GTB classifier was found to have the highest overall accuracy of 88.3% compared to the other classifiers. The CART classifier, on the other hand, has the lowest overall accuracy of 86.4%. Moreover, the GTB classifier performed better as it had the highest Kappa value of

0.864. The CART classifier, on the other hand, had the lowest Kappa value of 0.841 and performed worse than the other classifiers (Table 4a).

As shown in Table 4b, the error matrices, overall accuracy, and kappa coefficient are presented for the land cover/use classification of the Sentinel-2 image using four different classifiers: RF, SVM, CART, and GTB. The GTB classifier achieved the highest overall accuracy at 94.4%, outperforming the other classifiers. In contrast, the CART classifier recorded the lowest overall accuracy at 91.4%. Furthermore, the GTB classifier also excelled in terms of the kappa coefficient, attaining the highest value of 0.935. Conversely, the CART classifier had the lowest kappa value at 0.899, indicating its poorer performance relative to the other classifiers (Table 4b). In the classifications performed on Landsat-9 and Sentinel-2 satellite images, the GTB classifier achieved the highest overall accuracy and kappa coefficient for both images, outperforming the other methods. The CART classifier exhibited the lowest performance across both satellite images. Furthermore, Sentinel-2 provided higher overall accuracy results compared to Landsat-9. This difference can be attributed to the higher resolution and broader spectral coverage of Sentinel-2.

In this study the reference points were selected based on a stratified random sampling approach to ensure representation from all classes. However, due to the limited extent of certain land cover types, the number of points in some categories, such as water, was inherently lower. To mitigate any bias, we ensured that the selection process was proportional to the availability of each class within the study area. The number of accurate assessment points (163) was determined based on Abbas Al-Aarajy et al. (2024) while we recognize that increasing the number of points could enhance the robustness of the analysis, logistical constraints and the quality of available reference data influenced our decision to use 163 points. Nevertheless, we ensured that the selected points were well distributed and representative of the land cover classes to provide a reliable accuracy assessment.

We appreciate the reviewer's calculation regarding the number of required accuracy assessment points. In our study, the number of points was determined following standard practices in remote sensing literature. However, we agree that increasing the number of points could provide a more precise estimate of classification accuracy. In future study, it must expand the sample size to further improve the reliability of the accuracy assessment.

As a result of the study, the importance of a well-structured accuracy assessment is realized and to strengthen it, a more comprehensive review of the relevant literature will be conducted in future studies. Regarding the selection of sample points, we will examine its applicability by comparing it with other commonly used sampling strategies to provide the most reliable and unbiased accuracy assessment by considering the binomial sampling method. We will also critically evaluate and improve our methodology to address the deficiency in drawing conclusions only from existing accuracy analysis methods and to minimize possible biases and ensure the robustness of our results.

4.4. Assessment of Land Cover Change

The land cover/use classification using Landsat-9 OLI and Sentinel-2A images was conducted on the GEE platform by applying four different classifiers: RF, SVM, CART, and GTB. The classification resulted in seven classes: water, forest, settlement, agriculture, shrubland, open space, and mining area. The areas covered by each class are presented in square kilometers (km^2) and as a percentage (%) in Table 5 below.

Table 5. Comparison of the area covered by the classes

Class	(a) Landsat-9 OLI							
	RF		SVM		CART		GTB	
	Area (Km^2)	Area (%)	Area (Km^2)	Area (%)	Area (Km^2)	Area (%)	Area (Km^2)	Area (%)
Water	8.167	0.987	5.902	0.987	9.999	1.208	8.272	1.000
Forest area	117.330	14.180	133.604	14.180	128.520	15.533	111.363	13.459
Built area	52.870	6.390	48.544	6.390	65.891	7.963	42.161	5.095
Agriculture land	210.482	25.438	229.960	25.438	174.935	21.142	222.563	26.898
Shrubland	290.193	35.072	275.570	35.072	276.016	33.358	294.881	35.638
Bare land	112.070	13.544	102.353	13.544	121.607	14.697	111.995	13.535
Mining area	36.313	4.389	31.492	4.389	50.456	6.098	36.189	4.374

Class	(b) Sentinel-2A							
	RF		SVM		CART		GTB	
	Area (Km^2)	Area (%)	Area (Km^2)	Area (%)	Area (Km^2)	Area (%)	Area (Km^2)	Area (%)
Water	8.056	0.974	5.928	0.716	10.411	1.258	7.205	0.871
Forest area	163.227	19.727	161.350	19.500	164.523	19.884	158.581	19.166
Built area	53.750	6.496	47.137	5.697	41.608	5.029	47.311	5.718
Agriculture land	256.481	30.998	237.407	28.692	229.837	27.777	241.456	29.182
Shrubland	224.124	27.087	241.049	29.132	214.605	25.937	244.793	29.585
Bare land	97.004	11.724	111.298	13.451	134.689	16.278	102.751	12.418
Mining area	24.782	2.995	23.255	2.811	31.751	3.837	25.328	3.061

When analyzing the areas covered by the classes resulting from the classification of the Landsat-9 OLI image in Table 5a, it is determined that shrubland occupies the most land in the Soma region, accounting for 35.072%, 35.072%, 33.358%, and 35.638% in the RF, SVM, CART, and GTB classifiers, respectively. The agricultural area ranks second, covering 25.438%, 25.438%, 21.142%, and 26.898% in the RF, SVM, CART, and GTB classifiers, respectively. The forest area ranks third, with land cover percentages of 14.180%, 14.180%, 15.533%, and 13.459% in the RF, SVM, CART, and GTB classifiers, respectively. The bare area ranks fourth, covering 13.544%, 13.544%, 14.697%, and 13.535% of the land area, respectively. The settlement area ranks fifth, with 6.39%, 6.390%, 7.963%, and 5.095% coverage, respectively. The mining area ranks sixth, with land coverage of 4.389%, 4.389%, 6.098%, and 4.374%, respectively. The water area was found to cover the least amount of land, with 0.987%, 0.987%, 1.208%, and 1.000% in the RF, SVM, CART, and GTB classifiers, respectively (Table 5a). The GTB and CART classifiers predicted larger areas for most classes, while the RF and SVM classifiers were generally more conservative. This difference may be attributed to the varying abilities of each classifier to process spectral information and distinguish between land types.

As a result of the classification of the Sentinel-2A image presented in Table 5b, the agricultural area covers the largest percentage of land, with 30.998%, 28.692%, 27.777%, and 29.182% in the RF, SVM, CART, and GTB classifiers, respectively. Scrubland ranks second, occupying 27.087% in RF, 29.132% in SVM, 25.937% in CART, and 29.585% in GTB. The SVM and CART classifiers showed higher percentages in shrubland compared to the other classifiers. The forest area ranks third, covering 19.727% in RF, 19.500% in SVM, 19.884% in CART, and 19.166% in GTB. The bare area ranks fourth, with coverage of 11.724%, 13.451%, 16.278%, and 12.418%, respectively. Settlements ranked fifth, occupying 6.496% in RF, 5.697% in SVM, 5.029% in CART, and 5.718% in GTB. Mining areas are in sixth place, occupying 2.995% in RF, 2.811% in SVM, 3.837% in CART, and 3.061% in GTB. The water area occupies the least land, with 0.974% in RF, 0.716% in SVM, 1.258% in CART, and 0.871% in GTB. Notably, the CART classifier appears to give the highest percentage value (1.258%) for the water area (Table 5b).

When comparing the areas covered by the classes resulting from the classification of Landsat-9 OLI and Sentinel-2A images, it is observed that Landsat-9 OLI, with a spatial resolution of 30 meters, is successful in classifying larger areas. However, it may be more limited than Sentinel-2A in classifying detailed and smaller areas. Sentinel-2A, with a spatial resolution of 10-20 meters, is better able to classify smaller and finer structural differences in detail. Among the classifiers, GTB and CART predicted larger areas in both satellite images, while RF and SVM provided more limited results.

Figures 4a and 4b below show the land cover/use thematic maps for the Soma district, which was determined as the study area using four machine learning perceptron, namely RF, SVM, CART, and GTB, for the classification of Sentinel-2 and Landsat-9 images. Radial Basis Function (RBF) was used as the kernel function in the SVM classifier. RBF is a function widely used in complex and nonlinear classification problems. By transforming the data, the RBF creates more distinct boundaries between classes, thereby improving the accuracy of machine learning models.

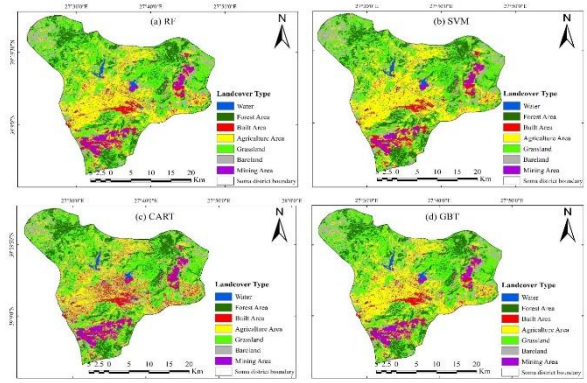

Figure 4a. Land cover/use map obtained from Landsat-9 OLI image

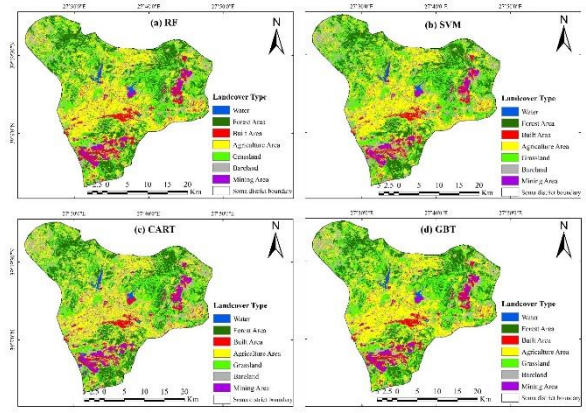

Figure 4b. Land cover/use map obtained from Sentinel-2A image

5. DISCUSSION

This study evaluated the performance of Sentinel-2 and Landsat-9 imagery for land use and land cover (LULC) mapping of coal deposits in the Soma region of Turkiye by comparing four machine learning algorithms: Random Forest (RF), Support Vector Machine (SVM), Classification and Regression Trees (CART), and Gradient Boosting Trees (GBT). Sentinel-2 and Landsat-9 images were downloaded from the Google Earth Engine (GEE) platform, where geometric and atmospheric corrections were applied. Subsequently, training samples were

collected, and machine learning algorithms were employed for LULC classification through visual interpretation and Google Earth Pro. Among these algorithms, the radial basis function (RBF), commonly used in non-parametric SVM algorithms, was utilized.

The classification results indicate that the GBT algorithm outperformed the other machine learning models in terms of overall accuracy and kappa coefficient for both Sentinel-2 and Landsat-9 imagery. Additionally, Sentinel-2 imagery exhibited slightly higher classification accuracy across all four machine learning algorithms compared to Landsat-9. Accuracy assessment metrics, including user accuracy, producer accuracy, and F1 score, were calculated from the error matrix to evaluate classification performance. The study identified seven distinct LULC classes in the Soma district: water, forest area, residential area, agricultural area, scrub area, bare area, and mining area. The water class was distinguished with the highest accuracy for both Sentinel-2 and Landsat-9 imagery. The forest area class ranked second in classification accuracy across both sensors. However, the mining area class exhibited the lowest classification accuracy, primarily due to its high spectral similarity with bare terrain and areas undergoing land degradation caused by mining activities.

The comparative analysis revealed that while advanced machine learning algorithms such as RF and SVM generally provide high classification accuracy for various terrain classes, GBT performed better in distinguishing the mining area class. This advantage is attributed to the GBT algorithm's ability to iteratively combine decision trees, allowing for improved classification performance without assuming a specific data distribution. RF and SVM, as widely recognized supervised classification algorithms, typically yield high accuracy for LULC mapping but may struggle with classes that have overlapping spectral signatures.

To contextualize the findings, the results of this study were compared with previous research on LULC classification. Shao and Lunetta (2012) applied the SVM algorithm to MODIS time series data for the Albemarle-Pamlico Estuarine System (APES) region in the United States and found that SVM outperformed neural networks (NN) and CART, achieving an overall accuracy of 77-80%. Similarly, Zhao et al. (2024) compared RF, SVM, and CART for LULC assessment in Mardan, Pakistan, using Sentinel-2 imagery, reporting overall accuracy values of 96.25%, 97%, and 98.68%, respectively, with CART performing best. Thanh Noi and Kappas (2018) evaluated RF, k-nearest neighbor (kNN), and SVM in Vietnam's Red River Delta, finding that SVM consistently achieved the highest accuracy, ranging from 90% to 95%. Other studies, such as those by Ghayour et al. (2021) and Talukdar et al. (2020), further corroborate the effectiveness of machine learning algorithms for LULC classification, with RF and SVM consistently yielding high accuracy in various geographic settings.

The performance of machine learning algorithms in classification studies is influenced by both satellite data and the characteristics of the study area. Decision tree-based models, such as CART, are prone to overfitting and are highly sensitive to variations in training datasets (Prasad et al., 2006). SVM algorithms, while computationally efficient in high-dimensional spaces, often require extensive fine-tuning and can be challenging to interpret (Huang et al., 2010). The findings of this study align with previous research, confirming that supervised classification algorithms such as RF, SVM, CART, and GBT provide reliable and high-accuracy results for Sentinel-2 and Landsat-9 imagery. Notably, the GBT algorithm demonstrated superior classification performance, particularly for the mining area class, which is often difficult to distinguish from residential and bare land classes.

This study also highlights environmental changes in the Soma region, particularly the decline in green areas due to expanding mining activities. These findings emphasize the need for effective land management strategies to mitigate the environmental impacts of mining. Sustainable land use policies should consider factors such as forest quality, vegetation cover, climate change, and human activities to ensure ecological balance and minimize land degradation.

Despite the study's contributions, certain limitations should be acknowledged. The classification accuracy of some LULC classes, particularly mining areas, was lower due to their spectral similarity with other land cover types. This challenge underscores the difficulty of distinguishing land cover categories with overlapping spectral signatures, even with advanced machine learning algorithms. Additionally, the study relied on a limited number of accuracy assessment points (163), which may have introduced bias, especially for underrepresented classes such as water. While stratified random sampling was used to ensure proportional representation, increasing the sample size could enhance result generalizability. Furthermore, since this study focused on a single region (Soma), its findings may not be directly applicable to other geographic locations with distinct LULC characteristics. Future studies should address these limitations by incorporating larger sample sizes, exploring additional spectral indices, and testing the methodology in diverse environments to enhance robustness and applicability.

Overall, this study provides valuable insights for local and regional land management authorities by demonstrating the efficacy of machine learning algorithms in LULC mapping. The findings contribute to the development of more accurate land classification techniques, ultimately aiding policymakers and researchers in making informed decisions regarding sustainable land management.

Future research should focus on refining classification methodologies, integrating additional data sources, and leveraging advanced remote sensing techniques to improve LULC mapping accuracy.

6. CONCLUSIONS

This study evaluated the performance of four supervised machine learning algorithms—Random Forest (RF), Support Vector Machine (SVM), Classification and Regression Trees (CART), and Gradient Boosting Trees (GBT)for Land Use Land Cover (LULC) classification using Sentinel-2 and Landsat 9 imagery on the Google Earth Engine (GEE) platform. Additionally, NDVI, NDWI, and NDBI indices were incorporated into the classification process to enhance class separation. The radial basis function (RBF) kernel was applied with SVM to assess its impact on classification accuracy.

The accuracy assessment results indicate that the GBT algorithm achieved the highest overall classification performance for both Sentinel-2 and Landsat 9 images, followed closely by RF. The SVM algorithm with the RBF kernel showed overall accuracy rates of 92.6% for Sentinel-2 and 87% for Landsat 9, performing better than CART but slightly lower than RF and GBT. Sentinel-2 imagery consistently outperformed Landsat 9 in class separability, likely due to its higher spatial resolution and spectral characteristics. The classification performance was further evaluated based on user accuracy and producer accuracy, revealing that the water class had 100% user accuracy across all algorithms. However, spectral similarity between certain classes, such as mining areas and bare land, led to reduced user accuracy in some cases, ranging from 80% to 95%.

To address this spectral similarity issue, future research should explore additional spectral indices or feature selection methods to improve class distinction. While NDVI, NDWI, and NDBI were used in this study, additional indices—such as soil-adjusted vegetation index (SAVI), built-up indices, or texture-based features—could be investigated to enhance classification accuracy. Moreover, the effectiveness of these indices should be systematically compared to determining their impact on resolving class confusion.

This study contributes to the literature by integrating multiple machine learning algorithms and satellite data sources while analyzing classification challenges in a complex LULC setting. Future studies should further assess classification performance under varying geomorphic and environmental conditions, incorporating additional datasets to refine classification methodologies. Developing more accurate and adaptable classification approaches will support reliable LULC mapping, which is essential for effective land management and policymaking. To overcome the limitations identified in this study, future research should focus on increasing the sample size for accuracy assessment to enhance the robustness of the results. Additionally, a comprehensive review of sampling strategies, including the binomial sampling method, should be conducted to ensure a more reliable and unbiased accuracy assessment.

Author Contributions

Sajaweddin Sadid: Conceptualization, Methodology, Software, Visualization, Writing- Original draft preparation. **Kaan Kalkan:** Data curation, Writing- Original draft preparation. Visualization, Supervision, Writing- Reviewing and Editing.

Conflicts of Interest

It has been declared that no conflicts of interest exist. The authors declare that no funding was received for this study.

REFERENCES

Abdullah, A. Y., Masrur, A., Sarfaraz, M., & Adnan, G. (2017). Spatio-Temporal Patterns of Land Use / Land Cover Change in the Heterogeneous Coastal Region of Bangladesh between 1990 and 2017. *Remote Sensing*, *11*(790), 1–26. https://doi.org/10.3390/rs11070790

Ahmed A. Zaeen, Khaleel I. Abood, Khalid H. Abbas Al-Aarajy *(2024)* Supervised Classification Accuracy Assessment Using Remote Sensing and Geographic Information System. *TEM Journal 13(1) 396-403*

Ballanti, L., Blesius, L., Hines, E., & Kruse, B. (2016). Tree Species Classification Using Hyperspectral Imagery : A Comparison of Two Classifiers. *Remote Sensing*, *8*(445), 1–18. https://doi.org/10.3390/rs8060445

Breiman, L., Frıedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and Regression Tree*. Wadsworth & Brooks.

Chen, W., Li, X., & Wang, L. (2020). Fine land cover classification in an open pit mining area using optimized support vector machine and world view-3 imagery. *Remote Sensing*, *12*(1), 12–14. https://doi.org/10.3390/RS12010082

Cihlar, J. (2000). Land cover mapping of large areas from satellites: Status and research priorities. *International Journal of Remote Sensing*, *21*(6–7), 1093–1114. https://doi.org/10.1080/014311600210092

Dixon, B., & Candade, N. (2008). Multispectral landuse classification using neural networks and support vector machines: One or the other, or both? *International Journal of Remote*

Sensing, *29*(4), 1185–1206. https://doi.org/10.1080/01431160701294661

Friedl, M. A., & Brodley, C. E. (1997). Decision Tree Classification of land cover from remote sensig data. *Remote Sens. Envıron.*, *61*, 399–409.

Garai, D., & Narayana, A. C. (2018). Land use/land cover changes in the mining area of Godavari coal fields of southern India. *Egyptian Journal of Remote Sensing and Space Science*, *21*(3), 375–381.https://doi.org/10.1016/j.ejrs.2018.01.002

Ghayour, L., Neshat, A., Paryani, S., Shahabi, H., Shirzadi, A., Chen, W., Al-Ansari, N., Geertsema, M., Amiri, M. P., Gholamnia, M., Dou, J., & Ahmad, A. (2021). Performance evaluation of sentinel-2 and landsat 8 OLI data for land cover/use classification using a comparison between machine learning algorithms. *Remote Sensing*, *13*(7). https://doi.org/10.3390/rs13071349

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, *202*, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Hong, F., He, G., Wang, G., Zhang, Z., & Peng, Y. (2023). Monitoring of Land Cover and Vegetation Changes in Juhugeng Coal Mining Area Based on Multi-Source Remote Sensing Data. *Remote Sensing*, *15*(13). https://doi.org/10.3390/rs15133439

Huang, C., Davis, L. S., & Townshend, J. R. G. (2010). An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing ISSN:*, *23*(4), 725–749. https://doi.org/10.1080/01431160110040323

Jayappa, K. S., Mitra, D., & Mishra, A. K. (2006). Coastal geomorphological and land-use and land-cover study of Sagar Island, Bay of Bengal (India) using remotely sensed data. *International Journal of Remote Sensing*, *27*(17), 3671–3682. https://doi.org/10.1080/01431160500500375

Karadağ, A. (2006). Linyit işletmeleri ve termik santralin ardından Somada değişen çevre, kent ve kimlik. *Ege Coğrafya Dergisi*, *15*(2006), 31–50.

Knorn, J., Rabe, A., Radeloff, V. C., Kuemmerle, T., Kozak, J., & Hostert, P. (2009). Remote Sensing of Environment Land cover mapping of large areas using chain classi fi cation of neighboring Landsat satellite images. *Remote Sensing of Environment*, *113*(5), 957–964. https://doi.org/10.1016/j.rse.2009.01.010

Kříbek, B., Nyambe, I., Sracek, O., Mihaljevič, M., & Knésl, I. (2023). Impact of Mining and Ore Processing on Soil, Drainage and Vegetation in the Zambian Copperbelt Mining Districts: A Review. *Minerals*, *13*(3). https://doi.org/10.3390/min13030384

Kumar, M., Denis, D. M., Kundu, A., Joshi, N., & Suryavanshi, S. (2022). Understanding land use/land cover and climate change impacts on hydrological components of Usri watershed, India. *Applied Water Science*, *12*(3), 1–14. https://doi.org/10.1007/s13201-021-01547-6

Leo Breiman. (2001). Random Forests. *Machine Learning*, *45*, 5–32. https://doi.org/10.1007/978-3-030-62008-0_35

Li, C., Wang, J., Wang, L., Hu, L., & Gong, P. (2014). Comparison of classification algorithms and training sample sizes in urban land classification with landsat thematic mapper imagery. *Remote Sensing*, *6*(2), 964–983. https://doi.org/10.3390/rs6020964

Lu, D., & Weng, Q. (2007). A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing*, *28*(5), 823–870. https://doi.org/10.1080/01431160600746456

Mangkhaseum, S., & Hanazawa, A. (2021). Comparison of Machine Learning Classifiers for Land Cover Changes using Google Earth Engine. *Proceedings of the 2021 IEEE International Conference on Aerospace Electronics and Remote Sensing Technology, ICARES 2021*, 1–7. https://doi.org/10.1109/ICARES53960.2021.9665186

Manna, A., & Maiti, R. (2014). Opencast Coal Mining Induced Defaced Topography of Raniganj Coalfield in India - Remote Sensing and GIS Based Analysis. *Journal of the Indian Society of Remote Sensing*, *42*(4), 755–764. https://doi.org/10.1007/s12524-014-0363-y

Manna, A., & Maiti, R. (2016). Alteration of Surface Water Hydrology by Opencast Mining in the Raniganj Coalfield Area, India. *Mine Water and the Environment*, *35*(2), 156–167. https://doi.org/10.1007/s10230-015-0342-8

Mateo-García, G., Gómez-Chova, L., Amorós-López, J., Muñoz-Marí, J., & Camps-Valls, G. (2018).

Multitemporal Cloud Masking in the Google Earth Engine. *Remote Sensing*, *10*(1079), 2–18. https://doi.org/10.3390/rs10071079

McFEETERS, S. K. (1996). International Journal of Remote Sensing The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing*, *17*(7), 1425-1432.

Mehta, S. A., Ashish, Solanki, M., & Seth, A. (2024). A Characterization of Land-use Changes in the Proximity of Mining Sites in India. *ACM Journal on Computing and Sustainable Societies*, *2*(1), 1–23. https://doi.org/10.1145/3624774

Melgani, F., & Bruzzone, L. (2004a). *Classification of Hyperspectral Remote Sensing Image With Support Vector Machines* (Vol. 42, Issue August, pp. 1778–1790).

Melgani, F., & Bruzzone, L. (2004b). Classification of Hyperspectral Remote Sensing Images With Support Vector Machines. *IEEE Transactıons On Geoscıence And Remote Sensıng*, *42*(8), 1778–1790.

Oo, T. K., Arunrat, N., Sereenonchai, S., Ussawarujikulchai, A., Chareonwong, U., & Nutmagul, W. (2022). Comparing Four Machine Learning Algorithms for Land Cover Classification in Gold Mining: A Case Study of Kyaukpahto Gold Mine, Northern Myanmar. *Sustainability (Switzerland)*, *14*(17). https://doi.org/10.3390/su141710754

Ouma, Y. O., Keitsile, A., Nkwae, B., Odirile, P., Moalafhi, D., & Qi, J. (2023). Urban land-use classification using machine learning classifiers: comparative evaluation and post-classification multi-feature fusion approach. *European Journal of Remote Sensing*, *56*(1). https://doi.org/10.1080/22797254.2023.2173659

Pal, M., & Mather, P. M. (2005). Support vector machines for classification in remote sensing. *International Journal of Remote Sensing*, *26*(5), 1007–1011. https://doi.org/10.1080/01431160512331314083

Pande, C. B. (2022). Land use/land cover and change detection mapping in Rahuri watershed area (MS), India using the google earth engine and machine learning approach. *Geocarto International*, *37*(26), 13860–13880. https://doi.org/10.1080/10106049.2022.2086622

Paraskevis, N., Servou, A., Roumpos, C., & Pavloudakis, F. (2021). Spatiotemporal interactions between surface coal mining and land cover and use changes. *Journal of Sustainable Mining*, *20*(2), 72–89. https://doi.org/10.46873/2300-3960.1053

Praticò, S., Solano, F., Di Fazio, S., & Modica, G. (2021). Machine learning classification of mediterranean forest habitats in google earth engine based on seasonal sentinel-2 time-series and input image composition optimisation. *Remote Sensing*, *13*(4), 1–28. https://doi.org/10.3390/rs13040586

Qian, Y., Zhou, W., Yan, J., Li, W., & Han, L. (2015). Comparing Machine Learning Classifiers for Object-Based Land Cover Classification Using Very High Resolution Imagery. *Remote Sensing*, *7*, 153–168. https://doi.org/10.3390/rs70100153

Shao, Y., & Lunetta, R. S. (2012). Comparison of support vector machine, neural network, and CART algorithms for the land-cover classification using limited training data points. *ISPRS Journal of Photogrammetry and Remote Sensing*, *70*, 78–87. https://doi.org/10.1016/j.isprsjprs.2012.04.001

Shi, D., & Yang, X. (2015). Support Vector Machines for Land Cover Mapping from Remote Sensor Imagery. In *Monitoring and Modeling of Global Changes: A Geomatics Perspective* (pp. 265–279). Springer.

Shi, Y., Fan, X., Ding, X., & Sun, M. (2024). An Assessment of Ecological Sensitivity and Landscape Pattern in Abandoned Mining Land. *Sustainability (Switzerland)*, *16*(3). https://doi.org/10.3390/su16031105

Sierra-Soler, A., Adamowski, J., Qi, Z., Saadat, H., & Pingale, S. (2015). High accuracy Land Use Land Cover (LULC) maps for detecting agricultural drought effects in rainfed agro-ecosystems in central Mexico. *Journal of Water and Land Development*, *26*(1), 19–35. https://doi.org/10.1515/jwld-2015-0014

Sridhar, V., Ali, S. A., & Sample, D. J. (2021). Systems analysis of coupled natural and human processes in the mekong river basin. *Hydrology*, *8*(3), 1–20. https://doi.org/10.3390/hydrology8030140

Sridhar, V., Kang, H., & Ali, S. A. (2019). Human-induced alterations to land use and climate and their responses for hydrology and water management in the Mekong River Basin. *Water*

(Switzerland), *11*(6). https://doi.org/10.3390/w11061307

Sujatha, E. R., & Sridhar, V. (2018). Spatial prediction of erosion risk of a small mountainous watershed using RUSLE: A case-study of the palar sub-watershed in Kodaikanal, South India. *Water (Switzerland)*, *10*(11), 1–17. https://doi.org/10.3390/w10111608

Tadesse, L., Suryabhagavan, K. V., Sridhar, G., & Legesse, G. (2017). Land use and land cover changes and Soil erosion in Yezat Watershed, North Western Ethiopia. *International Soil and Water Conservation Research*, *5*(2), 85–94. https://doi.org/10.1016/j.iswcr.2017.05.004

Talukdar, S., Singha, P., Mahato, S., & Pal, S. (2020). *Land-Use Land-Cover Classification by Machine Learning Classifiers for Satellite Observations—A Review*.

Thanh Noi, P., & Kappas, M. (2018). Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. *Sensors (Basel, Switzerland)*, *18*(1). https://doi.org/10.3390/s18010018

Townshend, J. R. G., & Justice, C. O. (2007). International Journal of Remote Sensing Analysis of the dynamics of African vegetation using the normalized difference vegetation index. *International Journal of Remote Sensing*, *7*(11), 1435–1445. https://doi.org/10.1080/01431168608948946

Tripathi, A. K., Aruna, M., Parida, S., Nandan, D., Elumalai, P. V., Prakash, E., Isaac JoshuaRamesh Lalvani, J. S. C., & Rao, K. S. (2024). Integrated smart dust monitoring and prediction system for surface mine sites using IoT and machine learning techniques. *Scientific Reports*, *14*(1), 1–12. https://doi.org/10.1038/s41598-024-58021-x

Vapnik, V. N., & Chervonenkis, A. Y. (1971). On the Uniform Convergence of Relative Frequencies of Events to Their Probabilities. In *Theory of Probability and Its Applications* (pp. 264–280). Springer International Publishing Switzerland. https://doi.org/10.1007/978-3-319-21852-6

Vorovencii, I. (2024). Long-term land cover changes assessment in the Jiului Valley mining basin in Romania. *Frontiers in Environmental Science*, *12*(January), 1–23. https://doi.org/10.3389/fenvs.2024.1320009

Wang, L., Jia, Y., Yao, Y., & Xu, D. (2019). Accuracy assessment of land use classification using support vector machine and neural network for coal mining area of Hegang city, China. *Nature Environment and Pollution Technology*, *18*(1), 335–341.

Waske, B., & Braun, M. (2009). Classifier ensembles for land cover mapping using multitemporal SAR imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, *64*(5), 450–457. https://doi.org/10.1016/j.isprsjprs.2009.01.003

Xiao, W., Ren, H., Sui, T., Zhang, H., Zhao, Y., & Hu, Z. (2022a). A drone- and field-based investigation of the land degradation and soil erosion at an opencast coal mine dump after 5 years' evolution of natural processes. *International Journal of Coal Science and Technology*, *9*(1). https://doi.org/10.1007/s40789-022-00513-0

Xiao, Y., Hao, Q., Zhang, Y., Zhu, Y., Yin, S., Qin, L., & Li, X. (2022b). Investigating sources, driving forces and potential health risks of nitrate and fluoride in groundwater of a typical alluvial fan plain. *Science of the Total Environment*, *802*, 149909. https://doi.org/10.1016/j.scitotenv.2021.149909

Xu, H., & Cheng, W. (2024). Landscape Analysis and Ecological Risk Assessment during 1995–2020 Based on Land Utilization/Land Coverage (LULC) and Random Forest: A Case Study of the Fushun Open-Pit Coal Area in Liaoning, China. *Sustainability*, *16*(6), 2442. https://doi.org/10.3390/su16062442

Zha, Y., Gao, J., & Ni, S. (2003). International Journal of Remote Sensing Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *International Journal of Remote Sensing*, *24*(3), 583–594.

Zhao, Z., Islam, F., Waseem, L. A., Tariq, A., Nawaz, M., Islam, I. U., Bibi, T., Rehman, N. U., Ahmad, W., Aslam, R. W., Raza, D., & Hatamleh, W. A. (2024). Comparison of Three Machine Learning Algorithms Using Google Earth Engine for Land Use Land Cover Classification. *Rangeland Ecology and Management*, *92*, 129–137. https://doi.org/10.1016/j.rama.2023.10.007