European Journal of Artificial Intelligence and Machine Learning
Vol 3 | Issue 3 | July 2024
ISSN 2796-0072

RESEARCH ARTICLE

Assessment of the Accuracy of Various Machine Learning Algorithms for Classifying Urban Areas through Google Earth Engine: A Case Study of Kabul City, Afghanistan

Karimullah Ahmadi [1,2,*]

ABSTRACT

Accurate identification of urban land use and land cover (LULC) is important for successful urban planning and management. Although previous studies have explored the capabilities of machine learning (ML) algorithms for mapping urban LULC, identifying the best algorithm for extracting specific LULC classes in different time periods and locations remains a challenge. In this research, three machine learning algorithms were employed on a cloud-based system to categorize urban land use of Kabul city through satellite images from Landsat-8 and Sentinel-2 taken in 2023. The most advanced method of generating accurate and informative LULC maps from various satellite data and presenting accurate outcomes is the machine learning algorithm in Google Earth Engine (GEE). The objective of the research was to assess the precision and efficiency of various machine learning techniques, such as random forest (RF), support vector machine (SVM), and classification and regression tree (CART), in producing dependable LULC maps for urban regions by analyzing optical satellite images of sentinel and Landsat taken in 2023. The urban area was divided into five classes: built-up area, vegetation, bare-land, soil, and water bodies. The accuracy and validation of all three algorithms were evaluated. The RF classifier showed the highest overall accuracy of 93.99% and 94.42% for Landsat-8 and Sentinel-2, respectively, while SVM and CART had lower overall accuracies of 87.02%, 81.12%, and 91.52%, 87.77%, with Landsat-8 and Sentinel-2, respectively. The results of the present study revealed that in this classification and comparison, RF performed better than SVM and CART for classifying urban territory for Landsat-8 and Sentinel-2 using GEE. Furthermore, the study highlights the importance of comparing the performance of different algorithms before selecting one and suggests that using multiple methods simultaneously can lead to the most precise map.

Keywords: Google Earth Engine, Machin Learning algorithms, Sentinel-2 and Landsat-8, Urban land use/cover.

Submitted: March 28, 2024
Published: July 15, 2024

10.24018/ejai.2024.3.3.40

[1] *Department of Aerospace Research of the Earth, Photogrammetry, Moscow State University of Geodesy and Cartography, Russia.*
[2] *Department of Civil Engineering, Nangarhar University, Afghanistan.*

****Corresponding Author:***
e-mail: ahmadi.niazai33@gmail.com

1. Introduction

In recent years, the classification of urban land cover using remote sensing imagery has become a significant area of research. The accurate and ongoing analysis of land cover and use is essential for sustainable development efforts in all regions. As urbanization continues to rapidly expand worldwide, obtaining information on urban land use quickly has become a crucial topic. Machine learning methods and the Google Earth Engine platform have been successfully applied and developed in various fields thanks to the powerful capabilities of remote sensing technology. Remote sensing data offers a wide range of information on a global scale, enabling the characterization and modeling of urban environments and providing assistance for environmental data processing, ecological planning and management, and other purposes [1]. Remote sensing has multiple applications within a wide range of areas, including mineralogy, natural resource management, agriculture, and urban planning, among others [2]. The primary goal of land cover/use classification is to examine the spatial information in an image and assign it to a land use category. While there are numerous techniques for creating LULC maps, satellite imagery, and remote sensing offer several advantages, including broad coverage, cheap cost, quick analysis, and the capacity to depict phenomenon aspects using various electromagnetic radiation regions [3]–[5]. The development of remote sensing technology has resulted in rising satellite imageries with medium to high resolutions. Nevertheless, generating maps of land use at low resolution over large areas requires substantial storage capacity, powerful processing capabilities, and the capacity to implement diverse methodologies [6]. The introduction of the Google Earth Engine (GEE) addressed these needs by integrating a large volume of remote sensing data from various sources into a cloud-based platform. This incredibly efficient computing tool allows for efficient and rapid processing of satellite imagery [7]–[10]. The data was integrated and analyzed in multiple stages to produce the end outcome. Utilizing remote sensing data and advanced machine learning techniques like CART, SVM, and RF, accurate land use and land cover classification were achieved [11]–[14].

GEE is widely used in research fields related to land use and land cover due to its extensive capabilities [7]. Zhao *et al*. [15] conducted research evaluating the effectiveness of several ML models (CART, SVM, and RF) in GEE for creating accurate land use maps through Sentinel-2 imagery. The results showed that GEE processed the satellite imagery quickly and provided excellent support for further analysis, with RF performing better than SVM and CART. Feizizadeh *et al*. [16] utilized ML algorithms on the GEE platform to perform LULC mapping and change detection analysis using Landsat multi-temporal satellite image. Their study demonstrated the effectiveness of ML algorithms for time series LULC mapping on the GEE platform. The GEE platform provides various ML algorithms for the supervised classification of image data, with three models (RF, SVM, and CART) utilized in this research to map urban areas.

Advanced machine learning methods have become favored in recent years for monitoring and assessing urban territory and natural hazards ML models are becoming more and more popular because of their increased accuracy and flexibility [5], [17], [18] being used in producing land use and land cover maps [19], [20]. Mao *et al*. [21] developed machine learning techniques for classifying urban land use by comparing the RF, SVM, and ANN algorithms. According to the findings, RF had the greatest impact on the classification of urban land use, whereas SVM and ANN had relatively little influence. Loukika *et al*. [9] utilized three machine learning algorithms, SVM, RF, and CART, to classify land use and land cover on the GEE platform and evaluated their performance using accuracy assessments. The results revealed that RF classifiers exhibited superior accuracy compared to both SVM and CART classifiers. Ouma *et al*. [22] conducted a study comparing four machine learning algorithms, including CART, RF, Gradient Tree Boosting (GTB), and SVM, for urban LULC classification. The findings indicated that RF and SVM were the most effective algorithms for mapping built-up areas based on overall accuracy. Talukdar *et al*. [19] assessed the performance of six ML methods-RF, SVM, ANN, Fuzzy ARTMAP, SAM, and MD. The findings indicated that all classifiers exhibited comparable accuracy with slight differences, with RF achieving the highest accuracy and MD performing the least effectively among the parametric classifiers.

Selecting the appropriate algorithm is a common challenge faced by users because it depends on various factors such as field situation, available data, and spectral similarity between the classes [23]. The primary aim of this study is to employ optical satellite imageries such as Landsat-8 and Sentinel-2 multispectral for the classification of urban LULC. This will be done using three different machine learning algorithms-random forest (RF), support vector machine (SVM), and classification and regression tree (CART) on the Google Earth Engine (GEE) platform. The study aims to compare the effectiveness and accuracy of these ML techniques for each algorithm and identify which algorithms are more accurate and suitable for similar conditions. Finally; the overall accuracy, Kappa coefficient, producer accuracy, and user accuracy are evaluated to compare and determine the results.

2. Materials and Methods

2.1. Study Area

Kabul, the largest city and capital of Afghanistan, is located in the eastern part of the country at 34°31′31″ North latitude and 69°10′42″ East longitude Fig. 1. The city covers an area of 1038 km^2 and sits at an elevation of approximately 5,900 feet (1,800 m) in a valley between the Asamaie and Sherdawaza mountain ranges [24]. Kabul is divided into 22 districts and has seen its population grow from 1.5 million in 2001 to around 5 million in 2017, making it one of the fastest-growing cities globally [25]. The rapid urbanization has strained the city's infrastructure, leading to around 70% of housing being developed illegally [25], [26]. This illegal housing, home to approximately 6 million people, is contributing to rising air pollution levels in the city [25]. Kabul experiences a climate ranging from dry to semi-arid, with warm summers and cold winters where temperatures can drop below −10 °C in winter and reach 40 °C in summer [27].

2.2. Data

The cloud-based GEE platform has stored an enormous quantity of Earth observation data (EOD) over the last 40 years. This includes satellite imagery from popular systems like Sentinel and Landsat, along with other geographical data such as climate and demographics. Sentinel-2 and Landsat-8 can be accessed through USGS in GEE. In the present research, less than 10% of the cloud cover of Sentinel-2 and Landsat-8 was used. In order to classify land use and land cover, only seven multispectral bands (1–7) with a spatial resolution of 30 m from Landsat-8 and six bands (2, 3, 4, 8, 11, 12) with spatial resolution of 10 and 20 m of sentinel-2 have been used. A total of 25 images of Sentinel-2 and 11 images of Landsat-8 for the 2023 year

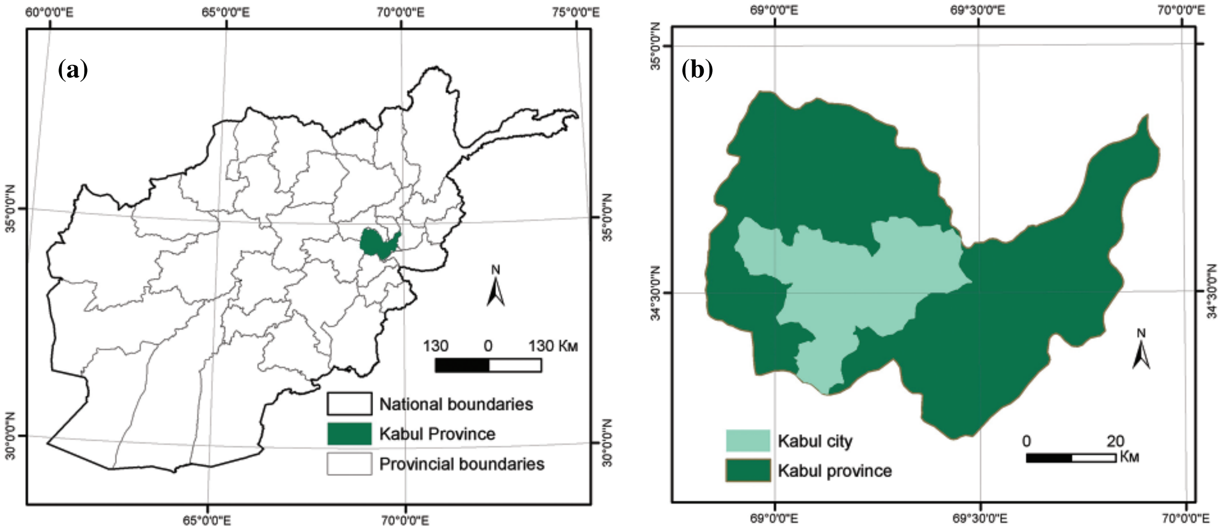

Fig. 1. Study area: (a) the geographic location of Afghanistan and (b) the geographic location of Kabul city.

TABLE I: The Dataset Characters that were Used in the Present Study

Sentinel-2B MSI			Landsat-8 OLI/TIR		
Band	Central wavelength (µm)	Spatial resolution (m)	Band	Central wavelength (µm)	Spatial resolution (m)
1	0.443	60	1	0.443	30
2	0.492	10	2	0.483	30
3	0.559	10	3	0.560	30
4	0.665	10	4	0.660	30
5	0.705	20	5	0.865	30
6	0.740	20	6	1.650	30
7	0.783	20	7	2.220	30
8	0.842	10	8	0.640	15
8a	0.865	20	–	–	–
9	0.940	60	9	1.375	30
10	1.375	60	10	10.900	100
11	1.610	20	11	12.000	100
12	2.190	20	–	–	–

were used in this study. Table I shows the data used in this research.

2.3. *Research Methods*

The method employed in this study was entirely carried out using the Google Earth Engine platform, with the method flowchart depicted in Fig. 2. Initially, satellite data were imported into GEE to eliminate cloud cover, and the yearly means function for Landsat and Sentinel satellite imagery was applied for the year 2023. Subsequently, training data for land use and land cover classification were generated and validated. Approximately 1200 samples were collected across the study areas for this purpose, with 70% utilized in the classification process and 30% reserved for accuracy assessment, which involved calculating overall accuracy and kappa statistics. The final step involved classifying the data using ML methods present in GEE, such as RF, SVM, and CART.

The urban area is classified into five classes: built-up area, vegetation, barren area, soil, and water bodies. The built-up area consists of asphalt, concrete, roofs, concrete high-rises, medium-rise houses, non-standard earthen houses, and metal. The vegetation class consists of agricultural farms, pastures, lawns, and urban green areas; the barren area class consists of rocky terrain and mountains of stones; the soil class consists of typically tin soil and lands without vegetation; and the water class consists of rivers, canals, ditches, and ponds.

2.4. *Machine Learning Algorithms*

This study assesses three machine learning algorithms to determine their accuracy and appropriateness for mapping urban LULC. The evaluated algorithms are SVM, RF, and CART, each of which is succinctly outlined:

- *Support Vector Machine (SVM):* A family of ML algorithms that is mostly utilized to solve problems associated with the classification of anomaly detection problems, or regression [28]. SVM is commonly employed in image and land classification mapping because of its superior accuracy, quick computation speed, and strong generalization capacity compared to traditional learning approaches [21], [29], [30]. It is an alternative algorithm that may be applied to remote sensing data to address a variety of classification weaknesses [31], [32]. The SVM algorithm creates a classification

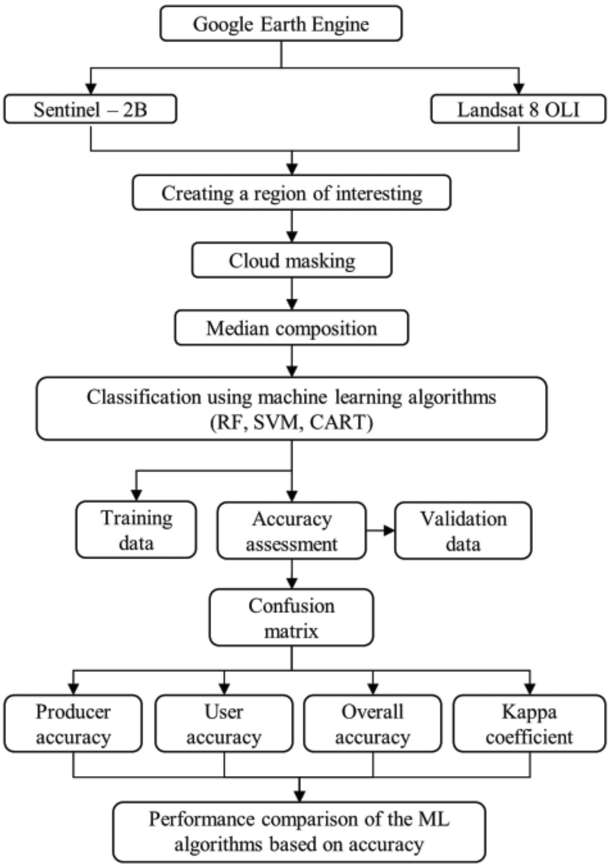

Fig. 2. Flowchart for the present study.

function by determining the best hyperplane to divide training data into classes through a specific process:

- *Random Forest (RF):* The Random Forest technique constructs numerous decision trees and is a widely used algorithm due to its precision and accuracy, ease of use, and adaptability [33]. Its ability to handle both classification and regression tasks, coupled with its nonlinear nature, makes it highly versatile and applicable to various data and scenarios. The term "forest" is used because it creates a collection of decision trees [34]. The results from these trees are then combined to produce the most precise predictions. While a single decision tree has limited outcomes and categories, the forest ensures greater accuracy by utilizing a larger number of categories and decisions. Additionally, the model introduces randomness by selecting the best feature from a random subset of features.
- *Classification and regression trees (CART):* Use a binary recursive partitioning method to analyze data, accommodating both categorical and continuous variables as predictors and targets without requiring data binning. CART produces a collection of pruned trees, with each potentially being the optimal tree. The best tree is selected by assessing the predictive accuracy of each tree in the pruning sequence using separate test data. The original data is used throughout the process [35]. The CART decision tree algorithm is highly dependent on the training data, meaning that any alterations to the training dataset can result in varied classification outcomes [36].

2.5. *Accuracy Assessment*

The accuracy assessment is essential to validating and assessing classification models by comparing them to ground truth data during modeling and mapping endeavors. This process helps determine the models' effectiveness and scientific significance. In summary, the validation and accuracy assessment stages are critical in every classification project. Its goal is to evaluate the models' efficacy and scientific importance by comparing the classified image to ground truth data from another data source.

In order to use supervised classification in ML approach for mapping urban LULC requires training samples for classification input. The training sample model utilized in this research consisted of five classes such as built-up area (614 points), vegetation (254 points), barren area (241 points), soil (151 points), and water bodies (46 points). The training sample model is utilized as input to direct the formation of a decision tree in the classification process using machine learning procedures.

3. Result and Discussion

3.1. *The LULC Classification Using Three Different ML Algorithms*

The research examines various machine learning techniques, such as SVM, RF, and CART, to map urban land use and land cover using optical data, including Sentinel-2 and Landsat-8 imagery on the Google Earth Engine. Using the cloud mask technique on the Google Earth Engine, cloud-contaminated pixels were eliminated from the corrected images, which had the least amount of cloud cover. Additionally, widely used indices like NDBI and NDVI are incorporated as supplementary inputs for LULC classification, representing built-up and vegetation characteristics, respectively. The median algorithm on the Google Earth Engine is used to composite satellite images for all selected years.

The training and validation sets were established, with around 1200 training samples gathered in the designated study regions. 70% of these samples were utilized for classification purposes, while the remaining 30% were reserved for accuracy evaluation, including the computation of overall accuracy and kappa statistics. The classification process involved applying SVM, RF, and CART machine learning algorithms on the same training and validation datasets available in GEE.

The urban area is classified into five classes as shown in Fig. 3, such as built-up area, vegetation, bare land, soil, and water bodies. The built-up area consists of asphalt, concrete, roofs, concrete high-rises, medium-rise houses, non-standard earthen houses, and metal. The vegetation class consists of agricultural farms, pastures, lawns, and urban green areas; the barren area class consists of rocky terrain and mountains of stones; the soil class consists of typically tin soil and lands without vegetation; and the water class consists of rivers, canals, ditches, and ponds.

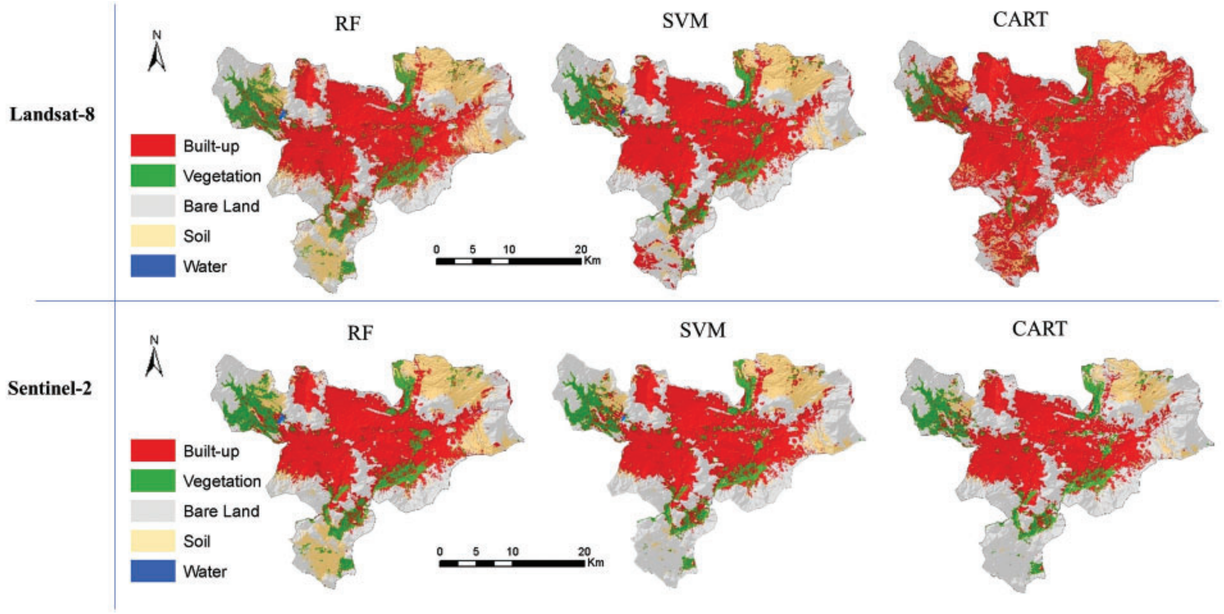

Fig. 3. The LULC map was created using the RF, SVM and CART algorithms in GEE platform for the Kabul city.

TABLE II: Accuracy Matrix Utilizing Landsat-8 Data for the RF, SVM, and CART Classification Algorithm

ML algorithm	Class name	B	V	S	BL	W	Total	UA
RF	Built-up (B)	254	5	7	4	0	270	94
	Vegetation (V)	5	237	1	2	2	247	96
	Soil (S)	8	2	127	5	0	142	89
	Barren land (BL)	5	3	5	196	0	209	94
	Water (W)	0	2	0	0	62	64	97
	Total	272	249	140	207	64		
	PA	93	95	91	95	97	OA = **93.99**	K = **0.92**
SVM	Built-up (B)	218	12	27	13	0	270	81
	Vegetation (V)	7	226	6	7	1	247	91
	Soil (S)	11	2	124	5	0	142	87
	Barren land (BL)	17	4	7	181	0	209	87
	Water (W)	1	0	1	0	62	64	97
	Total	254	244	165	206	63		
	PA	86	93	75	88	98	OA = **87.02**	K = **0.83**
CART	Built-up (B)	204	8	38	17	3	270	76
	Vegetation (V)	6	228	4	8	1	247	92
	Soil (S)	32	4	95	11	0	142	67
	Barren land (BL)	19	8	12	170	0	209	81
	Water (W)	0	2	3	0	59	64	92
	Total	261	250	152	206	63		
	PA	87	91	63	83	95	OA = **81.12**	K = **0.75**

3.2. Accuracy Assessment and Performance of Various ML Algorithms

Classification was first performed using supervised techniques such as SVM, RF, and CART on the Sentinel and Landsat imageries. The classification results are depicted in Fig. 3, with corresponding accuracy details in Tables II and III. The ML algorithm of RF outperformed SVM and CART, with Sentinel-2 images showing greater precision than Landsat images. The overall accuracy of the RF, SVM, and CART methods for Landsat were 93.99%, 87.02%, and 81.12%, respectively (Table II), while for sentinel, the overall accuracy of the RF, SVM, and CART methods were 94.42%, 91.52%, and 87.77%, respectively (Table III). Based on Landsat image for RF, SVM, and CART methods, the kappa coefficients were 0.92, 0.83, and 0.75, respectively, (Table II), while for Sentinel-2, they were 0.93, 0.89, and 0.84, respectively (Table III). Producer and user accuracy were highest for the RF classifier in both Landsat and Sentinel imageries compared to SVM and CART, as shown in Fig. 4.

This study compared several machine learning algorithms accuracy utilizing optical imageries of Sentinel-2 and Landsat-8, the same number of data used for validation and training data, and other information to assess how well they classifying urban areas. The assessment of algorithm accuracy was based on the analysis of LULC classifications. According to the results, Sentinel outperformed Landsat in terms of accuracy, while the RF method showed the greatest overall accuracy and Kappa coefficient for both satellite imageries. The SVM algorithm ranked

TABLE III: Accuracy Matrix Utilizing Senitinel-2 Data for the RF, SVM, and CART Classification Algorithm

ML algorithm	Class name	B	V	S	BL	W	Total	UA
RF	Built-up (B)	261	3	2	4	0	270	97
	Vegetation (V)	5	237	1	2	2	247	96
	Soil (S)	9	3	128	2	0	142	90
	Barren land (BL)	7	4	5	193	0	209	92
	Water (W)	0	1	2	0	61	64	95
	Total	282	248	138	201	63		
	PA	93	96	93	96	97	OA = **94.42**	K = **0.93**
SVM	Built-up (B)	243	7	12	8	0	270	90
	Vegetation (V)	5	226	6	9	1	247	91
	Soil (S)	6	5	123	8	0	142	87
	Barren land (BL)	3	3	5	198	0	209	95
	Water (W)	0	1	0	0	63	64	98
	Total	257	242	146	223	64		
	PA	95	93	84	89	98	OA = **91.52**	K = **0.89**
CART	Built-up (B)	224	16	22	8	0	270	83
	Vegetation (V)	11	223	7	5	1	247	90
	Soil (S)	7	12	113	9	1	142	80
	Barren land (BL)	6	2	5	196	0	209	94
	Water (W)	0	2	0	0	62	64	97
	Total	248	255	147	218	64		
	PA	90	87	77	90	97	OA = **87.77**	K = **0.84**

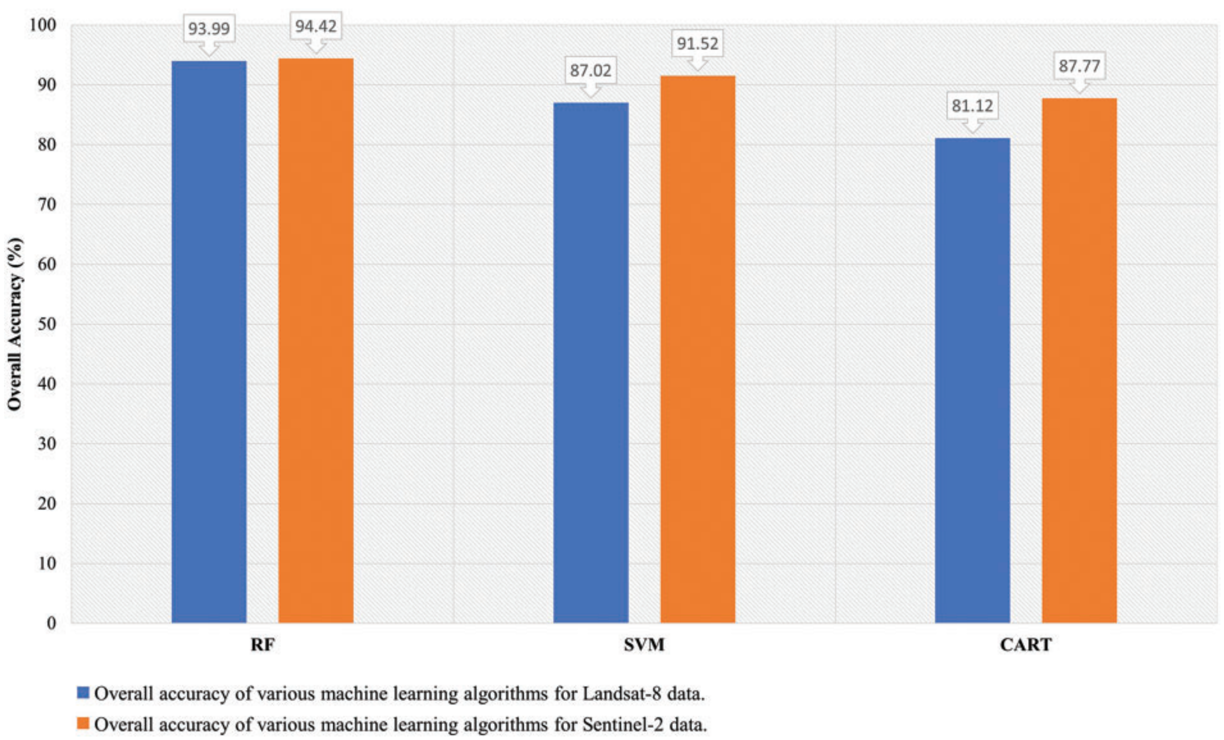

Fig. 4. The overall accuracy of Landsat-8 and Sentinel-2 for various machine learning algorithms.

second in accuracy for both satellites following RF, while CART exhibited the lowest performance across both satellite datasets (Figs. 5 and 6).

After comparing various machine learning algorithms, it was found that the RF algorithm had the most accurate classification for built-up areas in both satellite datasets. For Sentinel-2 data, the RF algorithm had a user accuracy of 97% and a producer accuracy of 93%, while for Landsat-8 data, it had a user accuracy of 94% and a producer accuracy of 93%. The SVM and CART algorithms had lower accuracy for built-up areas, with the SVM algorithm having a user accuracy of 90% and 83% and a producer accuracy of 95% and 90% for both Landsat and Sentinel imageries, respectively. The user accuracy of the CART method was 81% and 76%, and the producer accuracy was 86% and 87%, respectively, for Sentinel and Landsat images. These findings suggest that it's important to compare the performance of different algorithms before choosing one and that using multiple methods simultaneously and comparing their findings can create the most precise map.

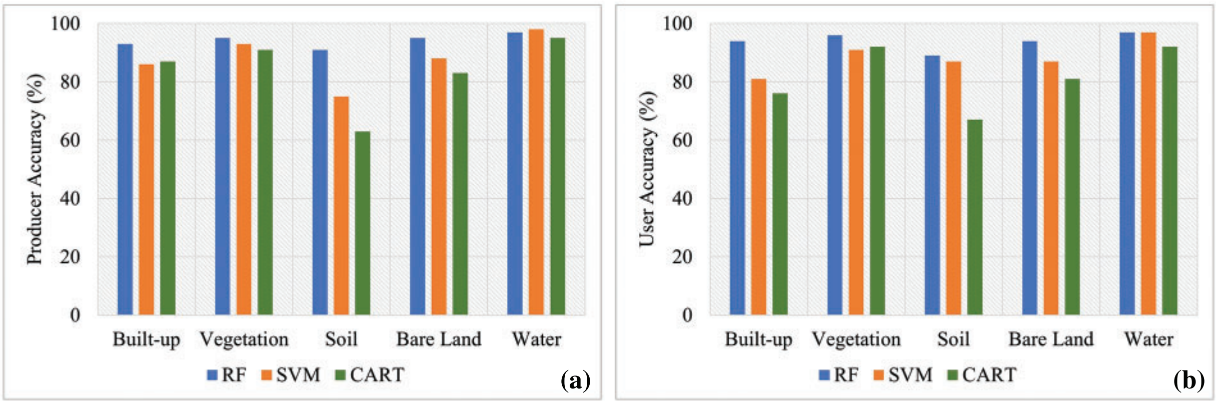

Fig. 5. Accuracy for each land class using RF, SVM, and CART classifiers of the Landsat-8 image: (a) producer accuracy and (b) user accuracy.

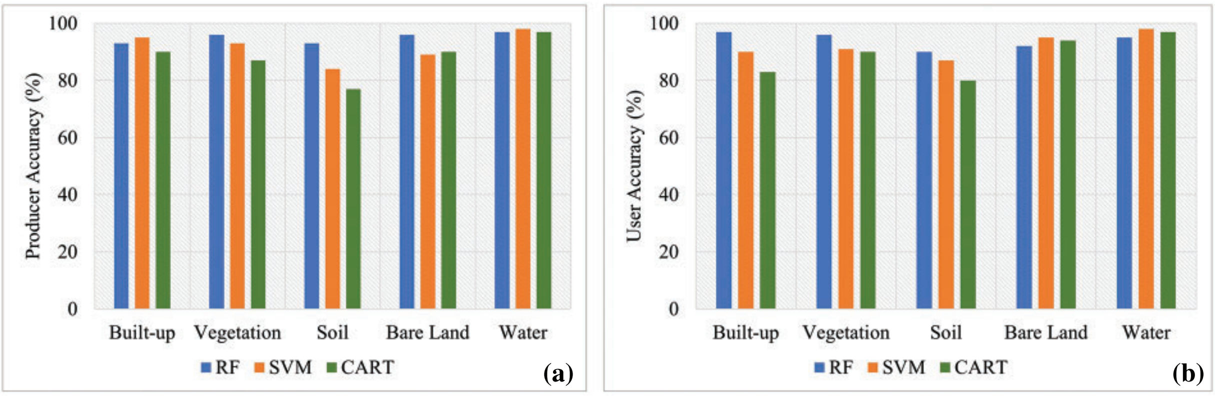

Fig. 6. Accuracy for each land class using RF, SVM, and CART classifiers of the Sentinel-2 image: (a) producer accuracy and (b) user accuracy.

4. Conclusion

This study's primary goal was to assess the comparisons of several machine learning algorithms in terms of their accuracy for classification, producer, user, and overall accuracy, as well as the Kappa coefficients. The study also aimed to assess the capability of optical satellite imageries of Landsat-8 and Sentinel-2, as well as the performance of a Google Earth engine-based classification in the Kabul city area. The Landsat and Sentinel satellite images on Google Earth Engine were used to evaluate the performance of the RF, SVM, and CART ML methods for classifying urban LULC. The accuracy of each algorithm was assessed for individual classes using an error matrix. The Sentinel-2 images showed superior performance compared to Landsat-8 due to its higher resolution. The RF algorithm demonstrated consistent and accurate classification of urban area data based on the confusion matrix and overall accuracy results. RF achieved overall accuracies of 93.99% for Landsat-8 and 94.42% for Sentinel-2, while SVM and CART had overall accuracies of 87.02% and 81.12% for Landsat-8, and 91.52% and 87.77% for Sentinel-2, respectively. The findings indicated that RF was the most suitable classifier for urban area classification with optical satellite imageries of Landsat and Sentinel using GEE. It is important to consider factors such as the condition of the study area, thematic accuracy, training data quality, and mapping requirements when selecting the most appropriate classifier. The dependability and flexibility of the Google Earth Engine make it a viable option for satellite image classification and analysis compared to other commercial software. These results emphasize the importance of comparing algorithm performance before choosing one, and utilizing multiple methods simultaneously and comparing their outcomes can lead to the most accurate mapping results.

Acknowledgment

The author would like to specifically thank the USGS Earth Explorer science team for providing Landsat data for this research. He is grateful to the Government of the Russian Federation for providing a scholarship for a PhD degree at the Moscow State University of Geodesy and Cartography. Moreover, he would like to express his deepest appreciation to his brother, Mansoor Ahmadi, a specialist in software engineering, for his financial support.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] Nelson PR, Maguire AJ, Pierrat Z, Orcutt EL, Yang D, Serbin S, *et al.* Remote sensing of tundra ecosystems using high spectral resolution reflectance: opportunities and challenges. *J Geophys Res: Biogeosci*. 2022;127(2):e2021JG006697.

[2] Ouchra H, Belangour A, Erraissi A. A comparative study on pixel-based classification and object-oriented classification of satellite image. *Int J Eng Trends Technol*. 2022;70(8):206–15.

[3] Topaloğlu RH, Sertel E, Musaoğlu N. Assessment of classification accuracies of Sentinel-2 and Landsat-8 data for land cover/use mapping. *Int Arch Photogramm, Remote Sens Spat Inform Sci*. 2016;41:1055–9.

[4] Shahabi H, Ahmad BB, Mokhtari MH, Zadeh MA. Detection of urban irregular development and green space destruction using normalized difference vegetation index (NDVI), principal component analysis (PCA) and post classification methods: a case study of Saqqez city. *Int J Phys Sci*. 2012;7(17):2587–95.

[5] Ghayour L, Neshat A, Paryani S, Shahabi H, Shirzadi A, Chen W, *et al.* Performance evaluation of sentinel-2 and landsat 8 OLI data for land cover/use classification using a comparison between machine learning algorithms. *Remote Sens*. 2021;13(7):1349.

[6] Xie S, Liu L, Zhang X, Yang J, Chen X, Gao Y. Automatic land-cover mapping using landsat time-series data based on google earth engine. *Remote Sens*. 2019;11(24):3023.

[7] Gorelick N, Hancher M, Dixon M, Ilyushchenko S, Thau D, Moore R. Google earth engine: planetary-scale geospatial analysis for everyone. *Remote Sens Environ*. 2017;202:18–27.

[8] Sidhu N, Pebesma E, Câmara G. Using google earth engine to detect land cover change: singapore as a use case. *Eur J Remote Sens*. 2018;51(1):486–500.

[9] Loukika KN, Keesara VR, Sridhar V. Analysis of land use and land cover using machine learning algorithms on google earth engine for Munneru River Basin. *India Sustain*. 2021;13(24):13758.

[10] Kolli MK, Opp C, Karthe D, Groll M. Mapping of major land-use changes in the Kolleru Lake freshwater ecosystem by using landsat satellite images in google earth engine. *Water*. 2020;12(9):2493.

[11] Nery T, Sadler R, Solis-Aulestia M, White B, Polyakov M, Chalak M. Comparing supervised algorithms in Land Use and land cover classification of a landsat time-series. *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE, 2016.

[12] Bar S, Parida BR, Pandey AC. Landsat-8 and sentinel-2 based forest fire burn area mapping using machine learning algorithms on GEE cloud platform over Uttarakhand, Western Himalaya. *Remote Sens Appl: Soc Environ*. 2020;18:100324.

[13] Liu D, Chen N, Zhang X, Wang C, Du W. Annual large-scale urban land mapping based on Landsat time series in google earth engine and OpenStreetMap data: a case study in the middle Yangtze River basin. *ISPRS J Photogramm Remote Sens*. 2020;159:337–51.

[14] Tassi A, Vizzari M. Object-oriented lulc classification in google earth engine combining snic, glcm, and machine learning algorithms. *Remote Sens*. 2020;12(22):3776.

[15] Zhao Z, Islam F, Waseem LA, Tariq A, Nawaz M, Islam IU, *et al.* Comparison of three machine learning algorithms using google earth engine for land use land cover classification. *Rangel Ecol Manag*. 2024;92:129–37.

[16] Feizizadeh B, Omarzadeh D, Garajeh MK, Lakes T, Blaschke T. Machine learning data-driven approaches for land use/cover mapping and trend analysis using google earth engine. *J Environ Plan Manage*. 2023;66(3):665–97.

[17] Duro DC, Franklin SE, Dubé MG. A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. *Remote Sens Environ*. 2012;118:259–72.

[18] Ma L, Li M, Ma X, Cheng L, Du P, Liu Y. A review of supervised object-based land-cover image classification. *ISPRS J Photogramm Remote Sens*. 2017;130:277–93.

[19] Talukdar S, Singha P, Mahato S, Pal S, Liou Y-A, Rahman A. Land-use land-cover classification by machine learning classifiers for satellite observations—A review. *Remote Sens*. 2020;12(7):1135.

[20] Mohammadi A, Baharin B, Shahabi H. Land cover mapping using a novel combination model of satellite imageries: case study of a part of the Cameron Highlands, Pahang, Malaysia. *Appl Ecol Environ Res*. 2019;17(2):1836–48.

[21] Mao W, Lu D, Hou L, Liu X, Yue W. Comparison of machine-learning methods for urban land-use mapping in Hangzhou city. *China Remote Sens*. 2020;12(17):2817.

[22] Ouma Y, Nkwae B, Moalafhi D, Odirile P, Parida B, Anderson G, *et al.* Comparison of machine learning classifiers for multitemporal and multisensor mapping of urban LULC features. *Int Arch Photogramm, Remote Sens Spat Inf Sci*. 2022;43:681–9.

[23] Thanh Noi P, Kappas M. Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery. *Sensors*. 2017;18(1):18.

[24] Wafa W, Hairan MH, Waizy H. The impacts of urbanization on Kabul City's groundwater quality. *Int J Adv Sci Technol*. 2020;29(4):10796–809.

[25] Chaturvedi V, Kuffer M, Kohli D. Analysing urban development patterns in a conflict zone: a case study of Kabul. *Remote Sens*. 2020;12(21):3662.

[26] Ahmadi K, Sahak AS, Sahak AT. Evaluation of Urban Sprawl and Land Use/Cover variation patterns through remote sensing data: a case study in Kabul City, Afghanistan. *Eur J Environ Earth Sci*. 2023;4(6):10–20.

[27] Qutbudin I, Shiru MS, Sharafati A, Ahmed K, Al-Ansari N, Yaseen ZM, *et al.* Seasonal drought pattern changes due to climate variability: case study in Afghanistan. *Water*. 2019;11(5):1096.

[28] Maxwell AE, Warner TA, Fang F. Implementation of machine-learning classification in remote sensing: an applied review. *Int J Remote Sens*. 2018;39(9):2784–817.

[29] Liu X, He J, Yao Y, Zhang J, Liang H, Wang H, *et al.* Classifying urban land use by integrating remote sensing and social media data. *Int J Geogr Inf Sci*. 2017;31(8):1675–96.

[30] Zhang Y, Liu J, Wan L, Qi S. Land cover/use classification based on feature selection. *J Coast Res*, 2015;(73):380–5.

[31] Huang C, Davis L, Townshend J. An assessment of support vector machines for land cover classification. *Int J Remote Sens*. 2002;23(4):725–49.

[32] Kamal M, Jamaluddin I, Parela A, Farda NM. Comparison of Google Earth Engine (GEE)-based machine learning classifiers for mangrove mapping. *Proceedings of the 40th Asian Conference Remote Sensing, ACRS*, 2019.

[33] Borra S, Thanki R, Dey N. *Satellite Image Analysis: Clustering and Classification*. Springer; 2019.

[34] Ouchra H, Belangour A, Erraissi A. Machine learning algorithms for satellite image classification using Google Earth Engine and Landsat satellite data: morocco case study. *IEEE Access*. 2023;(11):71127–42.

[35] Steinberg D, Colla P. CART: classification and regression trees. In *The Top Ten Algorithms in Data Mining*. pp. 179, 2009 Apr 9.

[36] Bishop CM, Nasrabadi NM. *Pattern Recognition and Machine Learning*, vol. 4. Springer; 2006.