Hindawi
Journal of Electrical and Computer Engineering
Volume 2022, Article ID 9092299, 16 pages
https://doi.org/10.1155/2022/9092299

Research Article

Analysis of Machine Learning Techniques for Sentinel-2A Satellite Images

Eman A. Alshari [1,2] and Bharti W. Gawali [2]

[1] *Thamar University, Computer Science & Information Technology Department, Dhamar, Yemen*
[2] *Dr. Babasaheb Ambedkar Marathwada University, Aurangabad 431004, India*

Correspondence should be addressed to Eman A. Alshari; em.alshari3@gmail.com

Received 21 March 2022; Revised 13 April 2022; Accepted 15 April 2022; Published 16 May 2022

Academic Editor: Yang Li

This article presents the comparative analysis of classification techniques to assign land use and land cover classes from different strategies (pixel-based, object-based, rule-based, distance-based, and neural-based) with a Sentinel-2A satellite image for 2016. The study area is the Sana'a city of Yemen which covers about 18,796.88 km^2 land area. This research aims to present the fundamentals of supervised machine learning approaches, including their limitations and strengths and experimentation for twelve classifiers. The outcome of experimentation showed that the Random Forest could be a good choice as a classifier for object-based strategy. In contrast, DTC and SVM were efficient in rule-based and pixel-based strategies. Results also showed that the highest accuracy was with object-based strategy, followed by rule-based and then pixel-based and distance-based strategies.

1. Introduction

Artificial intelligence techniques play a significant role in LULC classification by spotting patterns in data. For many years, the popularity of AI has been growing, with a considerable percentage of necessary research employing AI in its operations; in the work of Alshari and Gawali (2021) [1], remote sensing aids in gathering data about the Earth through satellites.

Machine learning helps to solve a wide range of real-world computer problems. Two types of supervised methods are classification and regression (Singh and others, 2021) [2]. Classification is a technique for categorizing data into strategies (Paul and others, 2021) [3]. The primary purpose is to determine the type or class. Regression is a method for predicting a single output value using training data. When an algorithm divides data into two strategies, this is known as binary classification. Choosing between more than two classes is called multiclass classification (Alshari and Gawali, 2021) [4].

Disclosure of land change is a factor in conserving land and considering management and development (Khwarahm, 2021) [5]. LULC map is vital in arranging executives (Makwinja and others, 2021) [6] and monitoring local, territorial, and national programs (Nayak, 2021) [7]. Land use and land cover data are needed for strategy-making business (Sarif and Gupta, 2021) [8] and regulatory purposes. With their spatial subtleties, the information is, in like manner, vital for ecological security and spatial arranging (Xie and others, 2021) [9]. Land use arrangement is indispensable because it gives information that can be utilized to demonstrate (Sang and others, 2021) [10], particularly, the one managing climate; for example, models manage environmental change and strategies improvements (Bhattacharya and others, 2021) [11].

Sentinel-2A of multispectral satellites is used in this study. Sentinel-2A is of medium resolution and is the first optical Earth observation satellite from European Space Agency [12].

The software used in this study is SAGA GIS, which is free, open-source software used on Windows and Linux computers. "SAGA" is an acronym for System for Automated Geoscientific Analysis. According to the SAGA website, a GIS was created to make the application of spatial algorithms simple and effective [13–15]. It includes an easy-to-use user interface with various visualization possibilities and a rich, increasing collection of geoscientific methodologies [16].

The supervisory machine learning classifiers' principles were narrowed down and explored, and their strengths and limits were revealed in this study. The significance of this research and its limitations lies in applying strategies supervised machine learning, identifying constraints, stability, and weaknesses, as well as the opportunities and problems that each technology presents. The practice of this study was with twelve classifiers for five types of supervisory machine learning techniques to compare the differences in classification accuracy between the strategies for land-use changes described, which helps to search for flaws in supervisory learning to improve it. All these points were essential for future users and researchers. The results showed that the object-based (OB) method is better than other classification methods and superior to all approaches. The comparison results from this study showed that the RF classifier (object-based) was the first best result, having given overall accuracy of 99.92% with Sentinel-2A for map of 2016. Random Forest (RF) classifier gave the highest classification accuracy in the twelve classifiers. The results also indicated differences in performance between the twelve classifiers in the same year, same season, and same weather condition with the different satellites. The best four classifiers among the twelve are RF, KNN, DTC, and SVM. Results also showed that object-based strategy gave the highest accuracy, followed by rule-based and then pixel-based and distance-based strategies.

The critical contributions from this study are as follows: (i) implementation of the twelve classifiers related to the five strategies of machine learning using Sentinel-2A satellite and (ii) presenting the analytical comparison of classification techniques to assign land use and land cover classes from different strategies (pixel-based, object-based, rule-based, distance-based, and neural-based) with a Sentinel-2A satellite image for the year 2016. The structure of this article is as follows: the introduction of this study is given in Section 1; related work and comparison between supervised methods for various strategies of AI techniques and multiple classifiers are discussed in Section 2; methodology and materials related to research area in a case study and data collection, LULC preprocessing, and digital classification are in Section 3; accuracy estimation and kappa coefficient are in Section 4; the results and discussion are given in Section 5; and finally conclusion is given in Section 6.

2. Related Work

Several studies on machine learning have lately been published, and each study has a specific purpose that was discussed [17–47]. This study demonstrated that the purpose it addressed had never been addressed previously. It focused on limiting the fundamentals of supervised machine learning textbooks and studying and extracting their strengths and flaws. This research looked at a lot of materials about applying supervisory machine learning algorithms to classify land-use changes. This study analyzed several pieces of literature related to the classification of land-use changes using machine learning algorithms. Machine learning is significantly popular [17] owing to its widespread use, as evidenced by the previous literature, because of ease, flexibility, speed, and low cost compared to deep learning and all artificial intelligence techniques. Machine learning techniques' prediction performance can be considerably improved through parameter modification [18]. Algorithms and tasks for machine learning can be complex [19], selecting the best learning algorithm for the application at hand [20, 21]. Choosing the incorrect learning algorithm will create unanticipated outcomes [22], resulting in a loss of effort and the model's efficacy and accuracy [23, 24]. According to previous studies, despite drawing academic interest and their desire to learn more about current land changes, the field of LULCC development remains underutilized [25]. Investigations and studies are necessary for various ways to boost knowledge discovery utilizing artificial intelligence (AI) techniques [26], which provided a significant drive for our effort [25, 26]. Comparing various methods used in this study is very important, as described in Table 1 [38, 39].

Table 2 describes the comparison of the features of twelve classifiers implemented in this study which are Random Forest (RF), Decision Tree Classification (DTC), Maximum Likelihood Classifier (MLC), Spectral Angle Mapper (SAM), Support Vector Machine (SVM), K-Nearest Neighbour (KNN), Minimum Distance Classification (MDC), Artificial Neural Networks (ANN), Mahalanobis, Maximum Entropy, Parallelepiped, and Normal Bayes.

Computational Complexity. A modern approach is a concise, in-depth examination of the subject, including cryptography and quantum computation. It takes x minutes to train it on n points. What if you train it on kn issues instead? If the training time has now increased to kx, the training time is now linear. It can be more at times. The new training time may be k^2x. The training time would be labelled quadratic in the number of points in this situation. Do not anticipate being able to execute this procedure on millions of points if you have a long training period for a few thousand points. Assessing the complexity of a machine learning algorithm is more complex than it appears. It may be implementation-dependent, data properties may lead to other methods, and training time typically depends on some parameters provided to the algorithm. Another point to consider is that the learning algorithms are complicated and reliant on other algorithms. The following approximations are obtained by multiplying n by the number of training samples, p by the number of features, and n trees by the number of trees (for approaches based on various trees), as described in Table 3 [29, 45, 46].

3. Methodology

The method for accomplishing this study's aims is the following: Review theory about all strategies in machine learning is surveyed and features and characters of all varieties are determined. Twelve classifiers are implemented

Table 1: Strategies of AI techniques.

Type of strategy	Characteristics and related methods	Classifier
1. Pixel-based	(i) Divide pixels into groups. It is usually straightforward to put into practice. One of the critical disadvantages of pixel-based classification is that it ignores information from neighbouring pixels to help identify the target pixel's class more correctly. Consequently, pixels in a class with a lot of spectral heterogeneity can be divided into several classes [27, 28]. (ii) One of the main drawbacks of pixel-based strategy is that it ignores information from nearby pixels which may aid in accurately identifying the target pixel's class. As a result, pixels in a class with substantial spectral heterogeneity may be labelled as separate classes [29]. (iii) There is the issue of mismatched pixels. Pixel-based approaches for extracting low-level characteristics are commonly utilized. Pixels in the overlapping region are misclassified due to class confusion, and the picture is classified based on spectral information [25, 26, 30]	Maximum Likelihood Classifier (MLC), Spectral Angle Mapper (SAM), Support Vector Machine (SVM), Mahalanobis, Maximum Entropy, Parallelepiped, Normal Bayes
2. Object-based	(A) Image pixels are collected into spectrally homogenous image objects using an image segmentation approach, and then the individual objects are classified [31]. (B) This is a hyperparameter used to estimate the relevance of variables. Rather than single pixels, the approach finds clusters of pixels that reflect presently existing things in a GIS database [32]	Random Forest (RF), K-Nearest Neighbour (KNN)
3. Rule-based	(i) A set of classification rules make up a rule-based classifier. Rules use the absence and presence of the term. In this classifier, we establish specific criteria for producing rules, and these rules are formed during training [33]. Because a record may not trigger any restrictions, simplified rules may no longer be comprehensive. (ii) A solution for mutually exclusive practices is as follows: A set of rules are organized logically. Voting techniques are used to deal with an unorganized rule set [34]. (iii) Early AI research uses master data, agreed-upon rules, and reasoning processes to extract meaningful information from large amounts of data [35]	DTC
4. Distance-based	Distance-based algorithms are nonparametric techniques that may be used for classification. These algorithms sort items into strategies depending on how unlike they are, as measured by distance functions [30]. Some of the most recent uses of distance-based algorithms are also explored. The distance between pixels in the highlight space is regulated [31, 36]	MDC
5. Neural-based	(A) Size and complexity: It is smaller and more complicated. It is incapable of complex pattern recognition [32]. (B) Information storage is replaceable, implying that new data may be replaced by old data [33]. (C) ANN offers numerous benefits, but it also has certain drawbacks, such as extended training times, high computational costs, and weight adjustment. (D) Artificial neural networks can give input for parallel processing, which implies that they can perform several tasks at once. (E) Artificial neural networks have been met with opposition [34]. It is a mathematical model that's been applied neutrally [35]. (F) It has many linked processing components known as neurons to perform all activities. (G) Information held in the neurons is just the weighted connection of neurons [36, 37]	ANN

for 92 images of size database. It used multispectral Sentinel-2A 20 m resolution and comparison between the results evaluation and analysis to justify the results. The following is a quick rundown of the satellite data processing methodology: Data was collected from a downloaded image product. The processing procedure begins with identifying real-world data as follows: Extract images, identify the location of the study area, identify the composite band for

Table 2: Limitations, stability, and weaknesses of twelve classifiers implemented in this study [28].

	Classifier	Features of classifier
1	RF	The effectiveness of random forests is due to several points that are as follows: Logging below a given node in a subset of the necessary features [35]. During the development of trees, the random forest contributes to the unpredictability of the model. When dividing a node, it finds the best property from a random set of attributes rather than looking for a key component. As a result, more variance leads to a better model [25, 26, 36]. It has powerful predictive performance with little processing and easy to understand. It is implemented using algorithms with built-in feature selection techniques [34]. The success of the random forest model is mainly because many non-correlated (trees) models outperform any single component models. Random forests are frequently used for trait selection in the data science process due to the low correlation between models [35].
2	KNN	KNN is a supervised approach that predicts the output of data points using a labelled input dataset. It is one of the most basic machine learning algorithms, and it may be used for a wide range of issues. It is primarily based on visual resemblance. The training data is saved and only used to produce real-time predictions to learn from it. As a result, the KNN technique is substantially quicker than earlier training-based algorithms [32]
3	DTC	The Decision Tree Classifier's key benefit is its ability to use a variety of feature subsets and decision rules at different categorization stages. A generic Decision Tree comprises one root node, several internal and leaf nodes, and branches [33]
4	SVM	SVM generates a decision boundary, or a hyperplane, between two classes to split or categorize them. SVM is also used in object identification and picture classification. The SVM method aims to find the optimal line or decision boundary for organizing n-dimensional space so that the following data points may be readily classified. A hyperplane is a boundary that is the optimum option. Why is the SVM classifier, in particular, the most effective classification approach for binary classification tasks? The best is determined by the data utilized and the circumstance; hence, no classifier can always be the best for all data and problems [34]
5	Mahalanobis	The Mahalanobis distance measures the degree of correlation between variables. You may have observed, for example, that gas mileage and displacement are significantly connected. As a result, the Euclidean distance computation contains much duplicate information [35] Although the Mahalanobis-distance-based technique is motivated by classification prediction confidence, we discover that its improved performance is due to nonclassification information [36]
6	Normal Bayes	The Normal Bayes method is a classification approach for binary and multiclass data categorization. Naive Bayes outperforms numerical input variables when it comes to categorical input variables. It may be used to forecast data and make predictions based on historical data Advantages. The class of the test dataset may be predicted quickly and easily. It is also good at predicting multiclass. A Naive Bayes classifier outperforms competing models such as logistic regression and takes less training data when the independence criteria are fulfilled [37]
7	MLC	The Maximum Likelihood Classifier is a popular remote sensing classification method that categorizes pixels with the highest likelihood into the proper class. The likelihood Lk is the posterior probability of a pixel belonging to class k [40]
8	Maximum Entropy	If only one parameter about a probability distribution is known, the principle of Maximum Entropy is a model generation criterion that involves picking the most unexpected (Maximum Entropy) prior assumption [41]
9	ANN	ANN learning approach teaches you how to achieve a complicated goal or optimize a given dimension over a long period. Nonparametric ANN techniques offer a particular advantage over statistical classification methods in that they do not require a priori knowledge of the distribution model of input data [42]
10	MDC	It gives classification with the fewest total parameters and computing demand but at the cost of accuracy. MDC's purpose is to categorize as many patterns as possible properly. The MDC technique identifies centroids of classes and calculates distances between them and the test pattern [43]
11	SAM	It is a method for matching picture spectra to known spectra or an automated end member. The SAM classification yields a picture displaying the best match at each pixel [44]
12	Parallelepiped	It uses the class signature threshold to identify whether a pixel belongs to a class, is mixed up with other styles, or is unclassified [45]

satellite processing, identify construct layers form, and finally identify the classifiers. The statistical results and overall classification accuracy are calculated for each image for each satellite. The methodology steps of this study for the Sana'a region are presented in Figure 1.

3.1. Research Area. The city of Sana'a is one of the largest cities in Yemen which is located in the governorate of the same name, and this city is the case study for this article [28]. The city of Sana'a is located at 15°N 44°C or 15.369445 latitudes 44.1191006 with 15°22′ 10.0020′N and 44°11′ 27.6216″E in GPS coordinates [39]. The total area of the city of Sana'a is 18,796.88 km^2 (49 sq mi) in this study. The population was 2,545,000 in 2017 [35]. It is surrounded by two mountains (Jabal Naqum from the east and Jabal Eiban from the west), and it is also surrounded by the province from all sides [45]. The city is around 2,200 meters above ocean level. In Figure 2, the study area of the case study is clear.

Table 3: Computational complexity for some of machine learning methods [46].

Algorithm	Classification/regression	Training	Prediction
Decision Tree	C + R	$O(n^2p)$	$O(p)$
Random Forest	C + R	$O(n^2pn_{trees})$	$O(pn_{trees})$
Random Forest	Implementation	$O(n^2pn_{trees})$	$O(pn_{trees})$
Extremely Random Trees	C + R	$O(npn_{trees})$	$O(npn_{trees})$
Gradient Boosting	C + R	$O(npn_{trees})$	$O(pn_{trees})$
Linear Regression	R	$O(p^2n + p^3)$	$O(p)$
SVM (Kernel)	C + R	$O(n^2p + n^3)$	$O(n_{sv}p)$
K-Nearest Neighbours (naive)	C + R	−−	$O(np)$
Neural Network	C + R	?	$O(pn_{L1} + n_{L1}n_{L2} + \ldots)$
Naive Bayes	C	$O(np)$	$O(p)$

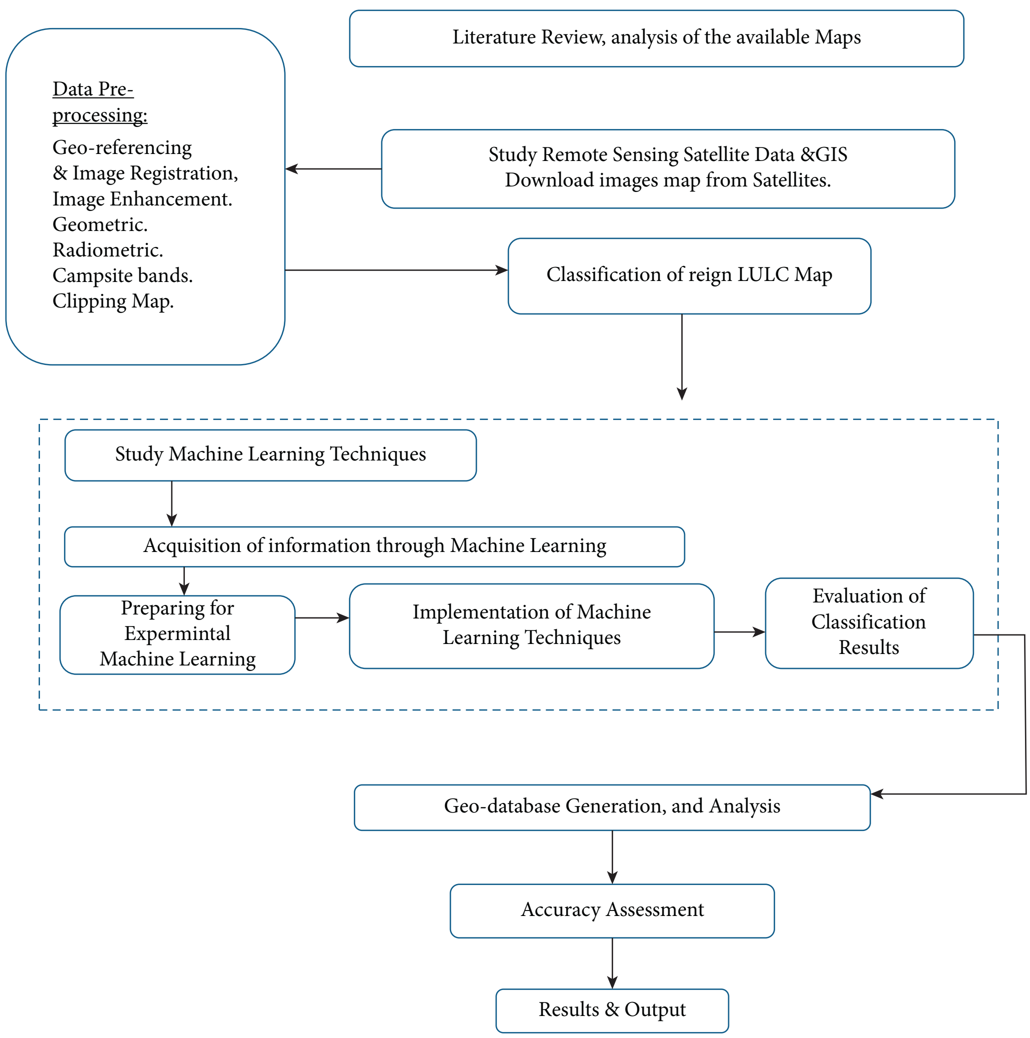

Figure 1: Workflow diagram for proposed methodology.

3.2. Data Collection. This study used images obtained from USGS of the Sana'a region. The survey images of the SOI Toposheet on the size of 1 : 50000 scales were used to prepare the base map [16]. In this study, the data was collected for Sentinel-2A satellite in 2016. The sensor is Sentinel-2A, allowing the calibration and comparison process in changing the land. The images generally consisted of maps of various types, dates, scales, and times. This study used pictures collected from Sentinel-2A (10 m) of multispectral resolution satellites. The image data were collected in December of 2016. Twelve images are contained for Sentinel-2A as dataset in this study as described in Table 4.

3.3. LULC Preprocessing. It is the primary stage and essential task in the LULC process, as well as the coordinate

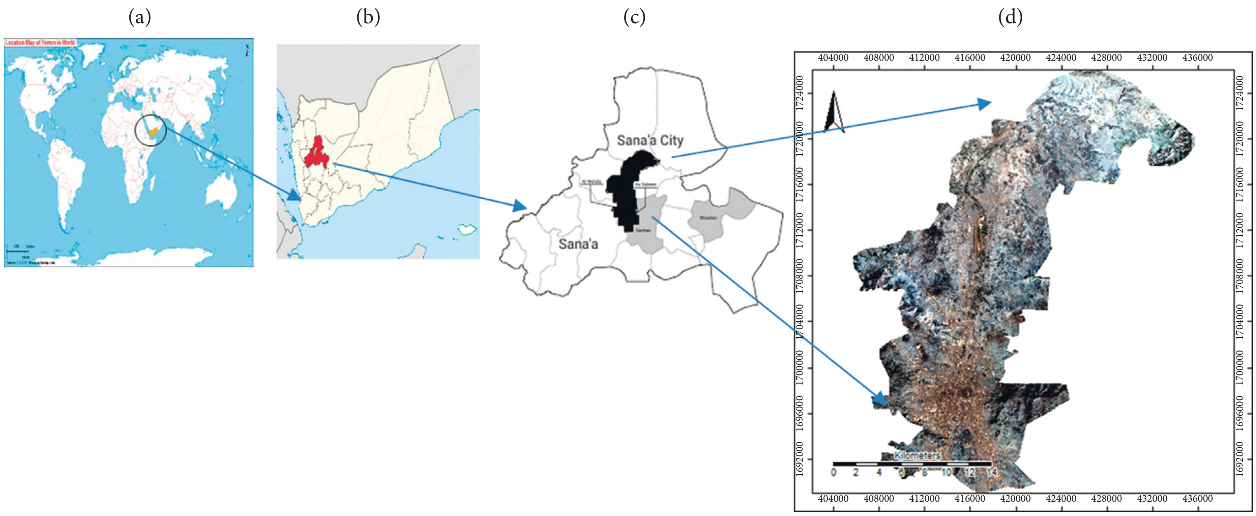

Figure 2: The location of Sana'a. (a) The world; (b) Yemen country; (c) Sana'a governorate; (d) Sana'a city.

Table 4: Downloaded images specification of this study.

Data acquired	Sentinel-2A satellite (2016)
Sensor	Sentinel-2A
Spatial resolution	10 m
The time of the season	December
Number of images	12 images

reference system for defining and cutting the map into specific areas. The preprocessing process includes studying the location of the case study exactly, as evident in this study (Figure 3).

Identifying the data after being downloaded from satellites under remote sensing technology is evident in Figure 4. The information subject to preprocessing is divided into the images shown in WGS84 or WGS84/UTM. It offers the preprocessing corrections for band 432. The preprocessing contains valid data with a geometric and radiometric correction, presented in this study [33] with QGIS and SAGA software. These operations improve satellite imagery for classification and rectify the degraded image to generate a more authentic portrayal of the actual scene [33].

3.4. Digital Classification. This section explains the approach used in the general level LULC planning action for Sana'a city and the specific outcomes obtained using multispectral medium goal satellite data. Our investigation reveals that the LULC in Sana'a saw considerable modifications in 2016. This data source can be used to containerize Sana'a's city sizes and contribute to territorial and global environmental models in the long run. There are two groups for classification models. Every group contains six models in LULC 2016 of the database for the proposed model to train, validate, and test the methods. The band classification used in this study is RGB 432. There are six samples for six parameters for creating models classes: High Land, Mountains, Land Area, Built-up, Vegetation, and Bare Land. The vegetation has been merged with the area of the agricultural land. The samples are created depending on RGB color composites of Sentinel-2A images, for example, the class Vegetation (red pixels in color composite RGB = 432) detailed changes in the region. There are twelve models for twelve classifiers described as groups in Figures 5 and 6, and the description of the twelve classifiers is detailed in Figures 5(a)–5(f) and Figures 6(a)–6(f).

4. Accuracy Assessment

Accuracy, confusion matrix, log-loss, and AUC-ROC are the four metrics used to assess classifier performance. This article employed the confusion matrix and the A kappa coefficient for accuracy assessment (Figures 7 and 8); the confusion matrix results for all methods used in this study are shown. A confusion matrix (sometimes called an error matrix) (Table 5) shows how well a classification model or classifier performs on a set of test data for which the proper values are known. The confusion matrix is simple, but the related terminology might be confusing. A confusion matrix is a tool for comparing two raster datasets' differences. An error matrix is the most frequent way of presenting the precision of the characterization result, the correctness of users and producers, and the insights gained from mistake lattices. The classes to which pixels in an array correspond for validation (ground truth) are used in the confusion matrix's columns. The confusion matrix is calculated by the following steps: the first step is to validate the dataset using the projected outcome values, the second step is to predict all of the rows in the test dataset, and the third step is to determine the anticipated outcomes and forecasts.

In this study, the SAGA GIS software used for LULC classification automatically calculates a confusion matrix and kappa coefficient with Excel for calculating statistical values. After removing the extent of performance anticipated by

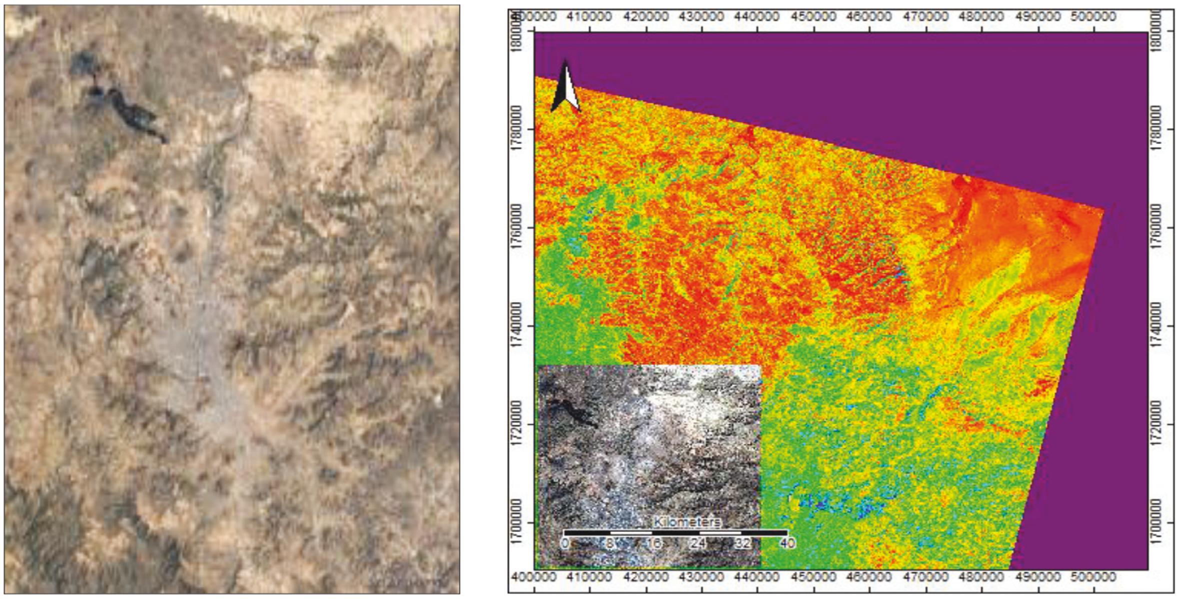

Figure 3: Sana'a on Google map and dataset of Sentinel-2A satellite.

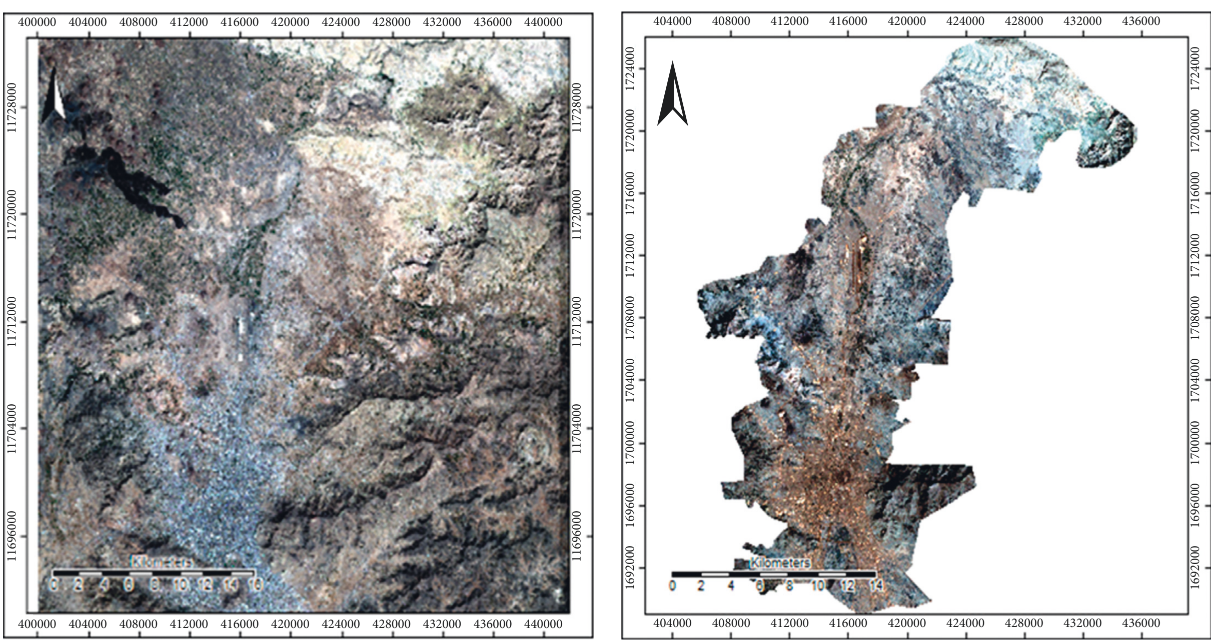

Figure 4: Selection of dataset from Sentinel-2A satellite and clipping of area study.

change, the kappa measurement merges the off-slanting components of the mistake frameworks and addresses arrangement [42]. The transaction is the perfect agreement when the Kappa coefficient equals 1. When it is close to zero, the deal is not much better than what you would anticipate by chance [43]. The kappa coefficient ranges from 0 to 1, with values above 0.7 deemed satisfactory. Simultaneously, individuals with a value of 0.4 or less identify an external link between the described image and the ground truth [44]. Table 5 shows details of the kappa values, and Table 6 shows the overall accuracy and kappa coefficient with the Sentinel-2A satellite calculated in this investigation.

5. Results and Discussion

The process has done LULC classification and the comparison of overall accuracy for LULC type to twelve classifiers described in Table 7. The object-based strategy was the first-best

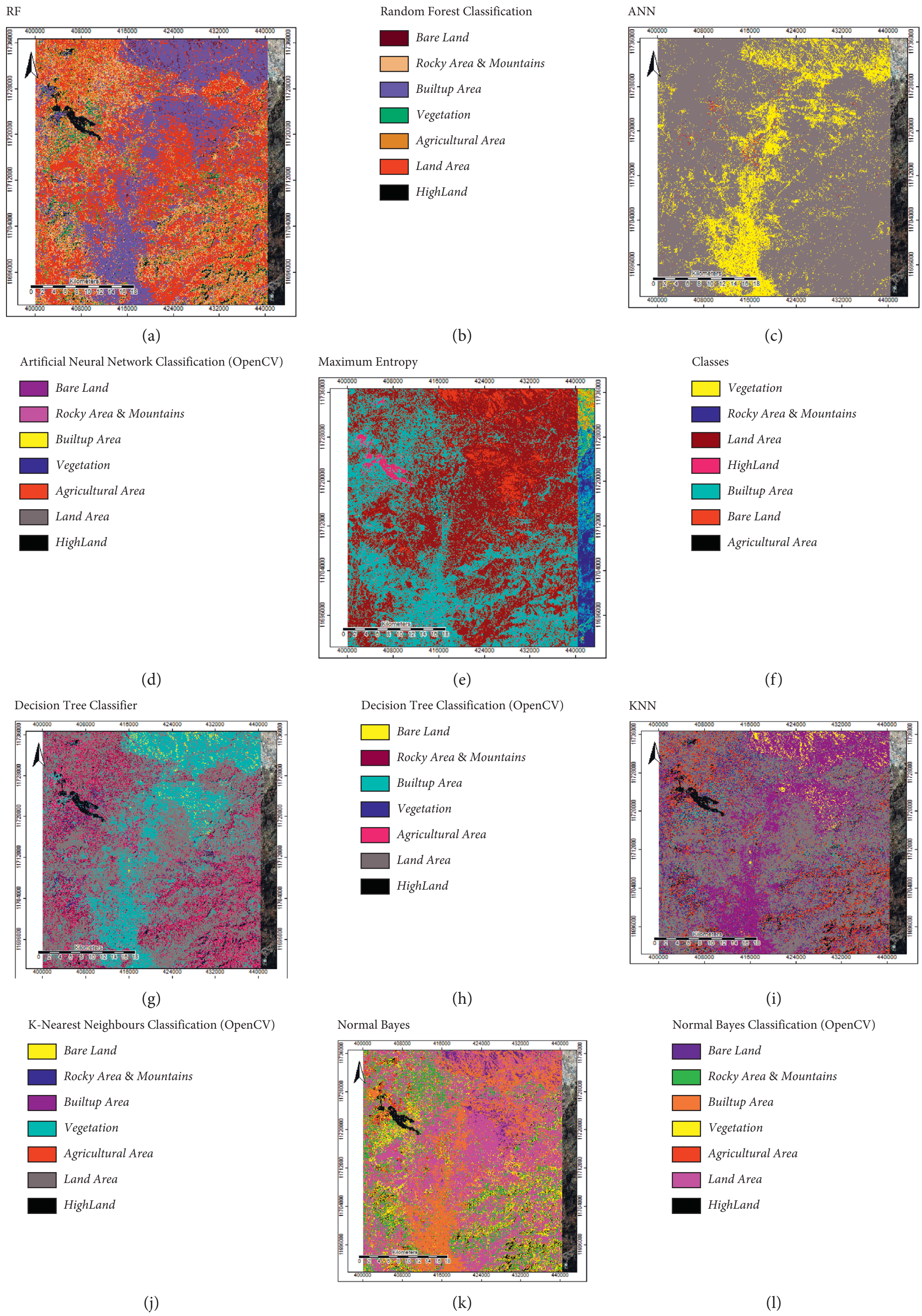

Figure 5: Group 1: land changes classification using machine learning methods with Sentinel-2A. (a, b) RF classifier. (c, d) ANN classifier. (e, f) ME classifier. (g, h) DTC classifier. (i, j) KNN classifier. (k, l) NB classifier.

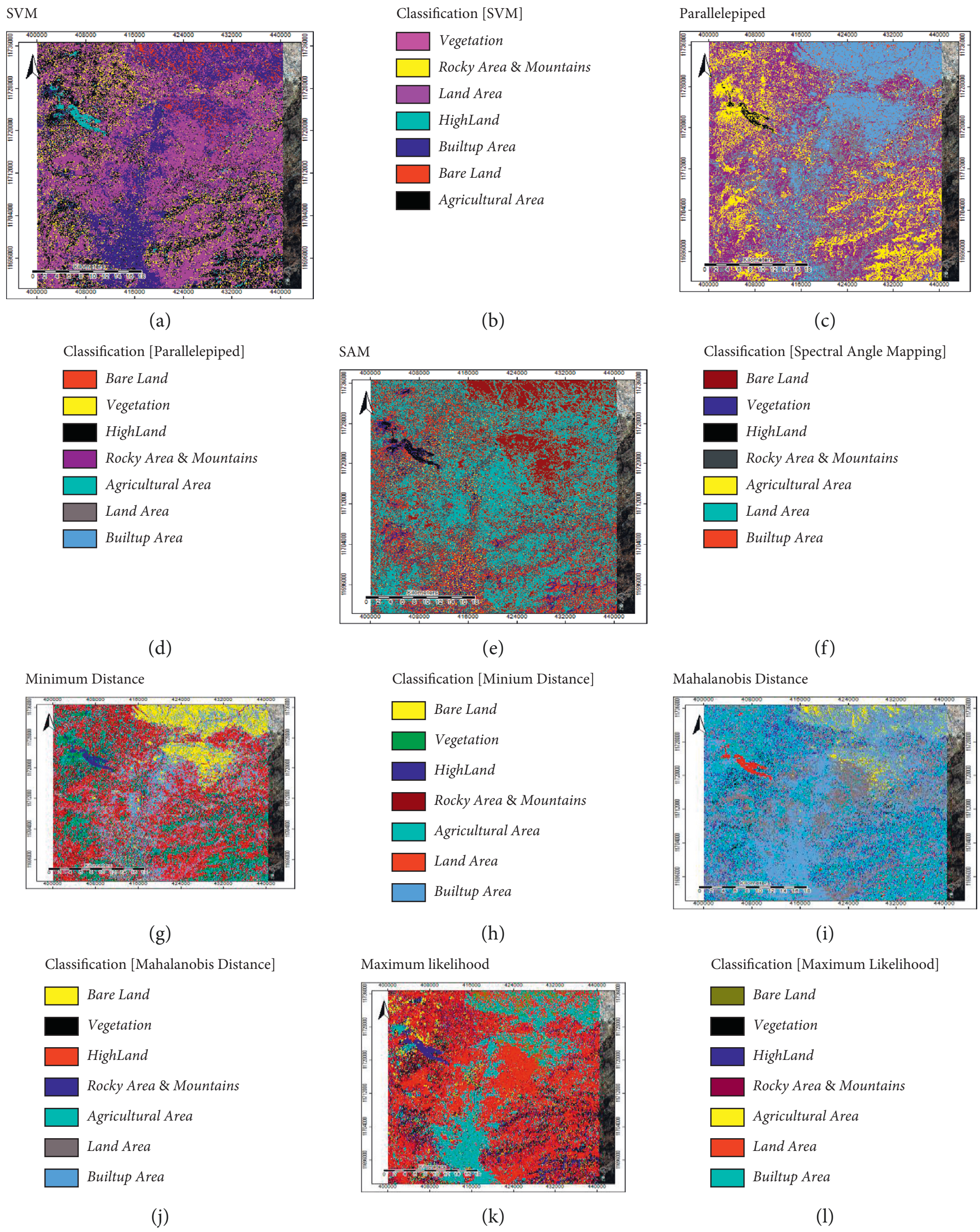

Figure 6: Group 2: land changes classification using machine learning methods with Sentinel-2A. (a, b) SVM classifier. (c, d) Parallelepiped. (e, f) SAM classifier. (g, h) Maximum Distance Classifier. (i, j) Mahalanobis distance classifier. (k, l) MLC.

result category of RF classifier with 99.92%. The rule-based strategy was the second-best result category of the DTC classifier with 91.49%. After that, the pixel-based strategy was the third-best result category of SVM with 84.56%. It mentions results of land changes classification of Sana'a City with RF classifier and Sentinel-2A satellite in 2016 in Tables 7–9; it offers statistical results ultimately. This section will discuss the analysis of the highest and the lowest accuracy recorded in the results of this study. When using multispectral satellite that has good accuracy like Sentinel-2A satellite (10 m) resolution, the objects on the ground are seen close and are easy to identify quickly by the object-oriented algorithm. Random Forests provide the highest accuracy for many reasons. RF offers a superior method for working with missing data. Among all the available classification methods, missing values are substituted by the variable appearing in the most

RF

CLASS	HighLand	Land Area	Agricultural	Vegetation	Builtup	Mountains	Bare Land	SumUser	AccUser
High Land	447	0	0	0	0	0	0	447	100
Land Area	0	903	0	0	0	1	0	904	99.88938
Agricultural	0	0	558	0	1	0	0	559	99.82111
Builtup	0	2	1	0	1693	0	0	1696	99.82311
Mountains	0	0	0	0	0	386	0	386	100
Bare Land	0	0	0	0	0	0	31	31	100
Unclassified	0	0	0	0	0	0	0	0	0
SumProd	447	905	559	206	1694	387	31		
AccProd	100	99.77901	99.82111	100	99.94097	99.7416	100		

KNN

CLASS	HighLand	Land Area	Agricultural	Vegetation	Builtup	Mountains	Bare Land	SumUser	AccUser
High Land	439	0	7	6	9	0	0	461	95.22777
Land Area	0	845	19	0	36	59	0	959	88.11262
Agricultural	3	6	487	31	22	28	0	577	84.40208
Builtup	1	25	18	4	1609	3	0	1660	96.92771
Mountains	0	29	18	0	15	296	0	358	82.68156
Bare Land	0	0	0	0	0	0	31	31	100
Unclassified	0	0	0	0	0	0	0	0	0
SumProd	447	905	559	206	1694	387	31		
AccProd	98.21029	93.37017	87.11986	80.09709	94.98229	76.48579	100		

DTC

CLASS	HighLand	Land Area	Agricultural	Vegetation	Builtup	Mountains	Bare Land	SumUser	AccUser
High Land	433	0	10	5	0	0	0	448	96.65179
Land Area	0	848	7	2	17	37	0	911	93.08452
Agricultural	2	6	496	74	30	30	0	638	77.74295
Builtup	6	7	7	3	1625	6	0	1654	98.24668
Mountains	0	44	29	0	21	314	0	408	76.96078
Bare Land	0	0	0	0	0	0	31	31	100
Unclassified	0	0	0	0	0	0	0	0	0
SumProd	447	905	559	206	1694	387	31		
AccProd	96.86801	93.70166	88.72988	59.2233	95.9268	81.13695	100		

SVM

CLASS	HighLand	Land Area	Agricultural	Vegetation	Builtup	Mountains	Bare Land	SumUser	AccUser
High Land	419	0	2	19	8	0	0	448	93.52679
Land Area	0	808	11	0	40	160	0	1019	79.29343
Agricultural	12	10	462	99	30	36	0	649	71.18644
Builtup	0	45	28	3	1591	11	0	1678	94.81526
Mountains	0	42	34	0	24	180	0	280	64.28571
Bare Land	0	0	0	0	0	0	31	31	100
Unclassified	0	0	0	0	0	0	0	0	0
SumProd	447	905	559	206	1694	387	31		
AccProd	93.73602	89.28177	82.64759	41.26214	93.91972	46.51163	100		

Mahalanobis

CLASS	HighLand	Land Area	Agricultural	Vegetation	Builtup	Mountains	Bare Land	SumUser	AccUser
High Land	405	0	3	2	2	0	0	412	98.30097
Land Area	0	708	0	0	11	59	0	778	91.00257
Agricultural	5	27	449	65	22	56	0	624	71.95513
Builtup	17	120	67	14	1652	97	0	1967	83.98577
Mountains	0	50	17	0	5	175	0	247	70.8502
Bare Land	0	0	0	0	0	0	31	31	100
Unclassified	0	0	0	0	0	0	0	0	0
SumProd	447	905	559	206	1694	387	31		
AccProd	90.60403	78.23204	80.322	60.67961	97.52066	45.21964	100		

Normal Bayes

CLASS	HighLand	Land Area	Agricultural	Vegetation	Builtup	Mountains	Bare Land	SumUser	AccUser
High Land	423	0	6	2	7	0	0	438	96.57534
Land Area	0	750	3	0	29	45	0	827	90.68924
Agricultural	1	2	236	11	31	12	0	293	80.54608
Builtup	0	32	23	4	1579	5	0	1643	96.10469
Mountains	0	117	115	2	41	315	0	590	53.38983
Bare Land	0	0	0	0	0	0	31	31	100
Unclassified	0	0	0	0	0	0	0	0	0
SumProd	447	905	559	206	1694	387	31		
AccProd	94.63087	82.87293	42.21825	90.7767	93.21133	81.39535	100		

Figure 7: Group 1 confusion matrix tables for machine learning methods of Sentinel-2A satellite.

MLC

CLASS	HighLand	Land Area	Agricultural	Vegetation	Builtup	Mountains	Bare Land	SumUser	AccUser
High Land	423	0	6	2	7	0	0	438	96.57534
Land Area	0	750	3	0	29	45	0	827	90.68924
Agricultural	1	2	236	11	31	12	0	293	80.54608
Builtup	0	32	23	4	1579	5	0	1643	96.10469
Mountains	0	117	115	2	41	315	0	590	53.38983
Bare Land	0	0	0	0	0	0	31	31	100
Unclassified	0	0	0	0	0	0	0	0	0
SumProd	447	905	559	206	1694	387	31		
AccProd	94.63087	82.87293	42.21825	90.7767	93.21133	81.39535	100		

Maximum Entropy

CLASS	HighLand	Land Area	Agricultural	Vegetation	Builtup	Mountains	Bare Land	SumUser	AccUser
High Land	405	0	13	12	34	0	0	464	87.28448
Land Area	0	783	6	0	107	141	0	1037	75.50627
Agricultural	0	0	116	0	3	0	0	119	97.47899
Builtup	41	122	424	194	1547	246	0	2574	60.10101
Mountains	0	0	0	0	3	0	0	3	0
Bare Land	0	0	0	0	0	0	31	31	100
Unclassified	0	0	0	0	0	0	0	0	0
SumProd	447	905	559	206	1694	387	31		
AccProd	90.60403	86.51934	20.75134	0	91.32231	0	100		

ANN

CLASS	HighLand	Land Area	Agricultural	Vegetation	Builtup	Mountains	Bare Land	SumUser	AccUser
High Land	0	0	0	0	0	0	0	0	
Land Area	446	859	350	188	114	380	10	2347	36.59992
Agricultural	1	0	179	11	7	0	0	198	90.40404
Builtup	0	46	24	1	1570	7	21	1669	94.0683
Mountains	0	0	5	5	2	0	0	12	0
Bare Land	0	0	0	0	0	0	0	0	
Unclassified	0	0	0	0	0	0	0	0	0
SumProd	447	905	559	206	1694	387	31		
AccProd	0	94.91713	32.02147	0.485437	92.68005	0	0		

MDC

CLASS	HighLand	Land Area	Agricultural	Vegetation	Builtup	Mountains	Bare Land	SumUser	AccUser
High Land	380	0	0	19	1	0	0	400	95
Land Area	0	484	11	0	126	88	0	709	68.26516
Agricultural	2	4	265	55	118	24	0	468	56.62393
Builtup	0	154	5	0	994	1	0	1154	86.13518
Mountains	0	263	94	4	373	272	0	1006	27.03777
Bare Land	0	0	0	0	64	0	31	95	32.63158
Unclassified	0	0	0	0	0	0	0	0	0
SumProd	447	905	559	206	1694	387	31		
AccProd	85.01119	53.48066	47.40608	62.13592	58.67769	70.28424	100		

SAM

CLASS	HighLand	Land Area	Agricultural	Vegetation	Builtup	Mountains	Bare Land	SumUser	AccUser
High Land	370	0	47	13	12	0	0	442	83.71041
Land Area	0	696	0	0	62	86	0	844	82.46446
Agricultural	0	0	92	16	178	1	0	287	32.05575
Builtup	1	12	219	37	747	64	0	1080	69.16667
Mountains	0	193	115	13	377	234	0	932	25.1073
Bare Land	0	4	0	0	0	0	31	35	88.57143
Unclassified	0	0	0	0	0	0	0	0	0
SumProd	447	905	559	206	1694	387	31		
AccProd	82.77405	76.90608	16.45796	61.65049	44.09681	60.46512	100		

Parallelepiped

CLASS	HighLand	Land Area	Agricultural	Vegetation	Builtup	Mountains	Bare Land	SumUser	AccUser
High Land	349	0	0	0	1	0	0	350	99.71429
Land Area	0	207	0	0	222	0	0	429	48.25175
Agricultural	0	9	19	0	332	0	0	360	5.277778
Builtup	0	0	0	0	622	0	0	622	100
Mountains	0	682	93	0	464	364	0	1603	22.70742
Bare Land	0	0	0	0	0	0	31	31	100
Unclassified	0	0	0	0	0	0	0	0	0
SumProd	447	905	559	206	1694	387	31		
AccProd	78.07606	22.87293	3.398927	100	36.71783	94.05685	100		

Figure 8: Group 2 confusion matrix tables for machine learning methods of Sentinel-2A satellite.

Table 5: Strength of agreement of a kappa coefficient [44].

No.	Kappa value	The degree of agreement
1	<0.00	Low
2	0.00–0.20	Medium
3	0.21–0.40	Good
4	0.41–0.60	Very good
5	0.61–0.80	Excellent
6	0.80–1.00	Very excellent

Table 6: Overall accuracy and kappa coefficient with Sentinel-2A satellite.

Sentinel-2A (2016)			
No.	Classifier	Overall accuracy (%)	Kappa coefficient
1	RF	99.92	0.998823
2	KNN	91.56	0.977268
3	DTC	91.49	0.952622
4	SVM	84.56	0.933182
5	Mahalanobis	83.83	0.736954
6	Normal Bayes	83.26	0.919499
7	MLC	83.26	0.906808
8	Maximum Entropy	68.15	0.788352
9	ANN	61.69	0.736954
10	MDC	60.39	0.668331
11	SAM	54.32	0.533520
12	Parallelepiped	42.52	0.116987

Table 7: Overall accuracy for classifiers with Sentinel-2A satellite.

	Type of strategies	Classifier	Accuracy (%)
1	Object-based	RF	99.92
		KNN	91.56
2	Rule-based	DTC	91.49
3	Pixel-based	SVM	84.56
		Mahalanobis	83.83
		MLC	83.26
		Normal Bayes	83.26
		Maximum Entropy	68.15
		SAM	54.32
		Parallelepiped	42.52
4	Distance-based	MDC	60.39
5	Neural-based	ANN	61.69

particular node. The Random Forest technique can also handle big data with numerous variables running into thousands. It can automatically balance datasets when a class is more infrequent than other classes in the data. The method also handles variables fast, making it suitable for complicated tasks. In comparison, the ANN classifier performed poorly with the Sentinel-2A satellite, probably due to the ANN's features and its classification procedure based solely on the training of the data. ANN has rarely employed the reason behind this. Artificial Neural Networks (ANN) attempt to identify land classes through training the data on strategies of land in most cases and most circumstances.

Table 8: Group 1 area and percentages of land changes for machine learning in Sentinel-2A satellite (2016).

No.	Name	RF		KNN		DTC		SVM		Mahalanobis		Normal Bayes	
		Area, km^2	%	Area, km^2	%	Area, km^2	%	Area, km^2	%	Area, km^2	%	Area, km^2	%
1	High land	49807900	2.65	59806000	3.18	47879400	2.55	35915200	3.80	29958400	2.27	41708400	2.22
2	Land area	725187700	38.58	766080400	40.76	701479200	37.32	818676800	86.70	527685000	39.91	750954500	39.95
3	Agricultural area	322988700	17.18	326778200	17.38	337729300	17.97	396037600	41.94	334353400	25.29	270312300	14.38
4	Built-up area	552998800	29.42	495807400	26.38	551666100	29.35	420439800	44.53	838959700	63.46	407204500	21.66
5	Mountains	214799500	11.43	198756400	10.57	226929200	12.07	171002900	18.11	109322700	8.27	363201000	19.32
6	Bare land	13905900	0.74	32460300	1.73	14005500	0.75	37616300	3.98	39409500	2.98	46308000	2.46
Total		1879688500	100.00	1879688700	100.00	1879688700	100.00	944277500	100.00	1322045300	100.00	1879688700	100.00

Table 9: Group 2 area and percentages of LULC for machine learning in Sentinel-2A satellite (2016).

No.	Name	MLC		Maximum Entropy		ANN		MDC		SAM		Parallelepiped	
		Area, km^2	%	Area, km^2	%	Area, km^2	%	Area, km^2	%	Area, km^2	%	Area, km^2	%
1	High land	41708400	2.27	33302400	4.53	193600	0.01	22922800	1.57	25655800	1.38	15333800	0.82
2	Land area	750954500	40.86	1010829000	137.62	1487337700	79.13	397663200	27.25	727984700	39.26	299130200	16.05
3	Agricultural area	270312300	14.71	2674700	0.36	14122000	0.75	366576900	25.12	138350600	7.46	406113800	21.80
4	Built-up area	407204500	22.15	727986600	99.12	376240400	20.02	488692800	33.49	283015300	15.26	418275700	22.45
5	Mountains	363201000	19.76	3825100	0.52	1795100	0.10	441238100	30.24	409992100	22.11	730496000	39.21
6	Bare land	46308000	2.52	101070800	13.76	0	0.00	162594900	11.14	294690100	15.89	9256600	0.50
Total		1837980300	100.00	734486400	100.00	1879688800	100.00	1459102700	100.00	1854032800	100.00	1863272300	100.00

6. Conclusion

This study analyzed the nature and qualities of the data and the performance of the learning algorithms to determine the effectiveness and efficiency of a machine-learning-based solution. The performances of twelve supervised machine learning classifiers on various categorization methodologies were explored in this work. It is observed that object-based classification using the Random Forest strategy produced the best results. Obtaining exact LULC maps in a variety of circumstances is difficult in general. When comparing the classification results, it was discovered that when the proper parameters are combined with the auxiliary data, the object-based classification technique provides high-accuracy LULC. This study found that this is still an open topic with much room for research and thought to improve land categorization tools. The basics should be the focus of future study of this field to devise proper methods for overcoming the challenges that this field presents, such as intraclass variation, which is the first factor to consider. Finally, any classification method's effectiveness is heavily dependent on a thorough understanding of the procedures and classifiers, the landscape's features, and the user's competence. This study provides a prediction model for future city planners for a better ecosystem. This study showcases the result of twelve machine learning classifiers which will be helpful for future researchers to select satellite images and the learner's algorithms according to their application.

Data Availability

The data of the executables for 12 classifiers and statistical results in Excel used to support the findings of this study have been deposited in the Google Drive repository (https://drive.google.com/drive/u/0/my-drive). The figures and tables data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] E. A. Alshari and B. W. Gawali, "Development of classification system for LULC using remote sensing and GIS," *Global Transitions Proceedings*, vol. 2, no. 1, 17 pages, 2021.

[2] R. K. Singh, V. S. P. Sinha, P. K. Joshi, and M. Kumar, "A multinomial logistic model-based land use and land cover classification for the South Asian Association for Regional Cooperation nations using Moderate Resolution Imaging Spectroradiometer product," *Environment, Development and Sustainability*, vol. 23, no. 4, pp. 6106–6127, 2021.

[3] S. P. Paul, P. A. Heaton, H. Nagendra, and N. Lele, "Physiological periosteal thickening in the long bones may mimic non-accidental injury," *British Journal of Hospital Medicine*, vol. 82, no. 11, p. 1, 2021.

[4] E. A. Alshari and B. W. Gawali, "Evaluation of the potentials and challenges of land observation satellites," *Global Transitions Proceedings*, vol. 2, no. 1, pp. 73–79, 2021.

[5] N. R. Khwarahm, "Spatial modeling of land use and land cover change in Sulaimani, Iraq, using multitemporal satellite data," *Environmental Monitoring and Assessment*, vol. 193, no. 3, 218 pages, 2021.

[6] R. Makwinja, E. Kaunda, S. Mengistou, and T. Alamirew, "Impact of land use/land cover dynamics on ecosystem service value-a case from Lake Malombe, Southern Malawi," *Environmental Monitoring and Assessment*, vol. 193, no. 8, 2021.

[7] S. Nayak, "Land use and land cover change and their impact on temperature over central India," *Letters in Spatial and Resource Sciences*, vol. 14, no. 2, 2021.

[8] M. O. Sarif and R. D. Gupta, "Spatiotemporal mapping of land use/land cover dynamics using remote sensing and GIS approach: a case study of Prayagraj City, India (1988–2018)," *Environment, Development, and Sustainability*, vol. 24, pp. 1–33, 2021.

[9] F.-D. Xie, X. Wu, L.-S. Liu, Y.-l. Zhang, and B. Paudel, "Land use and land cover change within the Koshi River Basin of the central Himalayas since 1990," *Journal of Mountain Science*, vol. 18, no. 1, pp. 159–177, 2021.

[10] X. Sang, Q. Guo, X. Wu et al., "The effect of DEM on the land use/cover classification accuracy of landsat OLI images," *Journal of the Indian Society of Remote Sensing*, vol. 49, no. 7, pp. 1507–1518, 2021.

[11] R. K. Bhattacharya, N. Das Chatterjee, and K. Das, "Land use and land cover change and its resultant erosion susceptible level: an appraisal using RUSLE and Logistic Regression in a tropical plateau basin of West Bengal, India," *Environment, Development and Sustainability*, vol. 23, no. 2, pp. 1411–1446, 2021.

[12] A. T. Angessa, B. Lemma, and K. Yeshitela, "Land-use and land-cover dynamics and their drivers in the central highlands of Ethiopia with special reference to the Lake Wanchi watershed," *Geojournal*, vol. 86, no. 3, pp. 1225–1243, 2021.

[13] M. S. Navin and L. Agilandeeswari, "Multispectral and hyperspectral images based land use/land cover change prediction analysis: an extensive review," *Multimedia Tools and Applications*, vol. 79, no. 39-40, Article ID 29751, 2020.

[14] H. Dibs, H. A. Hasab, J. K. Al-Rifaie, and N. Al-Ansari, "An optimal approach for land-use/land-cover mapping by integration and fusion of multispectral Landsat OLI images: a case study in Baghdad, Iraq. Water, Air," *& Soil Pollution*, vol. 231, no. 9, pp. 1–15, 2020.

[15] İ. Atay Kaya and E. Kut Görgün, "Land use and land cover change monitoring in Bandırma (Turkey) using remote sensing and geographic information systems," *Environmental Monitoring and Assessment*, vol. 192, no. 7, pp. 430–518, 2020.

[16] Carleton, "Land use change assessment in SAGA GIS," 2021, https://dges.carleton.ca/CUOSGwiki/index.php/Land_Use_Change_Assessment_in_SAGA_GIS.

[17] X. Xu, S. Shrestha, H. Gilani, M. K. Gumma, B. N. Siddiqui, and A. K. Jain, "Dynamics and drivers of land use and land cover changes in Bangladesh," *Regional Environmental Change*, vol. 20, no. 2, pp. 54–11, 2020.

[18] Z. B. Shi, Y. Li, and T. Yu, "Short-term load forecasting based on LS-SVM optimized by bacterial colony chemotaxis algorithm," in *Proceedings of the 2009 International Conference on Information and Multimedia Technology*, pp. 306–309, Jeju, Republic of Korea, December 2009.

[19] S. N. MohanRajan, A. Loganathan, and P. Manoharan, "Survey on land use/land cover (LU/LC) change analysis in remote sensing and GIS environment: techniques and challenges," *Environmental Science and Pollution Research*, vol. 27, no. 24, pp. 29900–29926, 2020.

[20] F. Rojas, C. Rubio, M. Rizzo, M. Bernabeu, N. Akil, and F. Martín, "Land use and land cover in irrigated drylands: a long-term analysis of changes in the mendoza and tunuyán river basins, Argentina (1986–2018)," *Applied Spatial Analysis and Policy*, vol. 13, no. 4, pp. 875–899, 2020.

[21] N. Saddique, T. Mahmood, and C. Bernhofer, "Quantifying the impacts of land use/land cover change on the water balance in the afforested River Basin, Pakistan," *Environmental Earth Sciences*, vol. 79, no. 19, pp. 448–513, 2020.

[22] B. Ekumah, F. A. Armah, E. K. A. Afrifa, D. W. Aheto, J. O. Odoi, and A. R. Afitiri, "Assessing land use and land cover change in coastal urban wetlands of international importance in Ghana using Intensity Analysis," *Wetlands Ecology and Management*, vol. 28, no. 2, pp. 271–284, 2020.

[23] W. R. Becker, T. B. Ló, J. A. Johann, and E. Mercante, "Statistical features for land use and land cover classification in Google Earth Engine," *Remote Sensing Applications: Society and Environment*, vol. 21, Article ID 100459, 2021.

[24] N. Das, P. Mondal, S. Sutradhar, and R. Ghosh, "Assessment of variation of land use/land cover and its impact on land surface temperature of Asansol subdivision," *The Egyptian Journal of Remote Sensing and Space Science*, vol. 24, no. 1, pp. 131–149, 2021.

[25] Z.-b. Shi, T. Yu, Q. Zhao, Y. Li, and Y.-B. Lan, "Comparison of algorithms for an electronic nose in identifying liquors," *Journal of Bionics Engineering*, vol. 5, no. 3, pp. 253–257, 2008.

[26] M. Zhang, J. Li, Y. Li, and R. Xu, "Deep learning for short-term voltage stability assessment of power systems," *IEEE Access*, vol. 9, Article ID 29711, 2021.

[27] Y. Hu, W. Li, D. Wright et al., "Artificial Intelligence Approaches," 2019, https://arxiv.org/abs/1908.10345.

[28] S. Kumar, S. Shwetank, and K. Jain, "Development of spectral signature of land cover and feature extraction using artificial neural network model," in *Proceedings of the 2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*, pp. 113–118, Greater Noida, India, February 2021.

[29] S. Samudrala, M. P. nand, S. Mohammad, and R. Vaddi, "Change detection in land use-land cover using convolutional neural network," in *International Conference on Communication, Computing and Electronics Systems*, pp. 745–752, Springer, Singapore, 2021.

[30] W. Chen, X. Li, H. He, and L. Wang, "A review of fine-scale land use and land cover classification in open-pit mining areas by remote sensing techniques," *Remote Sensing*, vol. 10, no. 2, p. 15, 2017.

[31] A. N. Soni, "Spatical context-based satellite image classification-review," *International Journal of Scientific Research and Engineering Development*, vol. 2, no. 6, pp. 861–868, 2019.

[32] R. Qiu, W. Xu, J. Zhang, and K. Staenz, "Modeling and simulating industrial land-use evolution in Shanghai, China," *Journal of Geographical Systems*, vol. 20, no. 1, pp. 57–83, 2018.

[33] L. Wang, J. Yan, L. Mu, and L. Huang, "Knowledge Discovery from Remote Sensing Images: A Review," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 10, Article ID e1371, 2020.

[34] V. S. Sahithi, J. Nehru, I. V. M. Krishna, and M. V. S. S. Giridhar, "Hyperspectral data classification algorithms for Delineation of LULC Classes," in *Proceedings of the International Conference on Industry 4.0*, May 2021.

[35] R. Girma, C. Fürst, and A. Moges, "Land use land cover change modeling by integrating artificial-neural-network with cellular Automata-Markov chain model in Gidabo River basin, Main Ethiopian Rift," *Environmental Challenges*, vol. 6, Article ID 100419, 2021.

[36] S. Talukdar, K. U. Eibek, S. Akhter, S. Ziaul, A. R. M. Towfiqul Islam, and J. Mallick, "Modeling fragmentation probability of land-use and land-cover using the bagging, random forest and random subspace in the Teesta River Basin, Bangladesh," *Ecological Indicators*, vol. 126, Article ID 107612, 2021.

[37] D. Verma and A. Jana, "LULC classification methodology based on simple Convolutional Neural Network to map complex urban forms at finer scale: Evidence from Mumbai," 2019, https://arxiv.org/abs/1909.09774.

[38] H. Yuan, C. Van Der Wiele, and S. Khorram, "An automated artificial neural network system for land use/land cover classification from Landsat TM imagery," *Remote Sensing*, vol. 1, no. 3, pp. 243–265, 2009.

[39] S. S. Rwanga and J. M. Ndambuki, "Accuracy assessment of land use/land cover classification using remote sensing and GIS," *International Journal of Geosciences*, vol. 8, no. 4, pp. 611–622, 2017.

[40] A. Sarica, A. Cerasa, and A. Quattrone, "Random forest algorithm for the classification of neuroimaging data in Alzheimer's disease: a systematic review," *Frontiers in aging neuroscience*, vol. 9, 2017.

[41] https://www.google.co.in/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwjarPzL0_r1AhUuzIsBHWkIC3UQFnoECA4QAw&url=https%3A%2F%2Fwww.sciencedirect.com%2Ftopics%2Fengineering%2Frandomorest&usg=AOvVaw3_WWAq4GBRTh7MlFCL_RRO.

[42] I. Ahmad, M. Basheri, M. J. Iqbal, and A. Rahim, "Performance comparison of support vector machine, random forest, and extreme learning machine for intrusion detection," *IEEE Access*, vol. 6, Article ID 33789, 2018.

[43] J. Sharma, R. Prasad, V. N. Mishra, V. P. Yadav, and R. Bala, "Land use and land cover classification of multispectral landsat-8 satellite imagery using discrete wavelet transform," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLII-5, pp. 703–706, 2018.

[44] https://yemen-nic.info/gover/amanaa/brife/The total area was 15,284.84 square kilometers.

[45] M. Al-shalabi, B. Pradhan, L. Billa, S. Mansor, and O. F. Althuwaynee, "Manifestation of remote sensing data in modeling urban sprawl using the SLEUTH model and brute force calibration: a case study of Sana'a city, Yemen," *Journal of the Indian Society of Remote Sensing*, vol. 41, no. 2, pp. 405–416, 2013.

[46] Thekerneltrip, "Computational complexity of machine learning algorithms," 2018, https://www.thekerneltrip.com/machine/learning/computational-complexity-learning-algorithms.

[47] F. Ramdani, B. Setiawan, A. Rusydi, and M. Furqon, "An artificial neural network approach to predict the future land use land cover of great Malang Region, Indonesia," 2021.