AL-Bahir
Journal for Engineering and Pure Sciences
BJEPS

Manuscript 1122

Classification of Land Cover in Sentinel-2 Imagery Using Machine Learning Models

Ehsan Ali Al-Zubaidi

Mohammed Ridha Hammoodi

Ahmed Naser Alzurfi

ORIGINAL STUDY

Classification of Land Cover in Sentinel-2 Imagery Using Machine Learning Models

Ehsan A. Al-Zubaidi [a,*], Mohammed R. Hammoodi [b], Ahmed N. Alzurfi [c]

[a] Department of Environmental Planning, Faculty of Physical, Planning, University of Kufa, Najaf, Iraq
[b] Computer Science Department, Faculty of Computer Science and Mathematics, University of Kufa, Najaf, Iraq
[c] Department of Electrical Engineering, Faculty of Engineering, University of Kufa, Najaf, Iraq

Abstract

Remote sensing data of medium resolution are commonly used to classify land cover, and machine learning (ML) models have taken on a central aspect in the necessary data analysis. Ordinarily, land cover is coded on a pixel basis on the basis of Digital Number (DN) values, which in turn are computed across several spectral bands. This paper is concerned with land cover mapping in Mosul, Iraq, based on satellite images captured by Sentinel-2. Two platforms featuring unsupervised classification algorithms were used, Google Earth Engine and ArcMap, making it possible to use K-means and X-means in Google Earth Engine and ISO Cluster in ArcMap. The comparative analysis of these algorithms allowed development of four significant land cover classes: built-up areas, vegetation, water bodies, and bare land. The results of the classification for the Mosul study area also indicated that using X-means gave an accuracy of 76.34, while using ISO Cluster gave 73.61, and K-means gave 80.14. The highest classification accuracy was achieved by the use of K-means, suggesting that it is the best algorithm to use in this region.

Keywords: Remote sensing, Image classification, Unsupervised learning, Spectral bands, Land cover

1. Introduction

Data on land use and land cover classifications derived from remote sensing systems is crucial for the evaluation of social and environmental applications, and this is now commonly achieved by using machine learning [1]. Modern technical breakthroughs have accelerated the development of pattern recognition systems, leading to the creation of decision rules that tend to the ever-more complicated [2]. Active data collection by multiple spaceborne satellites has been used to improve comprehension of Earth's atmospheric conditions, as well as to drive coastline research, ice surveys, and geological studies. Successful results have been achieved in these areas through the use of many novel approaches to remote sensing applications [3]. To determine the physical characteristics of the environment and detect features on Earth's surface, remote sensing as a scientific discipline makes use of electromagnetic radiation (EMR) to facilitate the analysis of spatial, temporal, and spectral fingerprints without the requirement of physical contact. By doing away with reference classification data, the unsupervised classification method makes spectral remote monitoring of land much more feasible, and the procedure has become crucial for deciphering satellite images. This classification approach groups or stratifies pixels based on similar spectral features as a way to analyse and monitor forests, agricultural zones, urban development, and environmental conditions.

Satellite image classification is an essential technique in remote sensing that employs machine learning and image processing to classify the various features of satellite data [4]. This includes classifying pixels into various land cover categories, such as forested areas, bodies of water, cities, and

Received 21 November 2025; revised 8 January 2026; accepted 10 January 2026.
Available online 27 March 2026

* Corresponding author.
E-mail addresses: Ihsana.kareem@uokufa.edu.iq (E.A. Al-Zubaidi), mridha.hamoodi@uokufa.edu.iq (M.R. Hammoodi), ahmedn.hussein@uokufa.edu.iq (A.N. Alzurfi).

https://doi.org/10.55810/2313-0083.1122

farmland, utilising the data across a range of images [5]. Such classification is beneficial in many areas, including agriculture, urban planning, and monitoring environmental change, and a great deal of potentially relevant data has been uncovered relating to both these and other domains. Data analysis and validation still require human intervention, however, which restricts the use of the technique to situations with a high degree of precision requirement [6].

Three unsupervised classification algorithms were implemented in this study to perform such work, to assess the feasibility of their use in situations that merit less human input. The K-means algorithm is one of the most effective unsupervised learning techniques for addressing clustering problems, having become one of the most renowned clustering algorithms. Stuart Lloyd, an American statistician, initiated its development in 1957 while employed at Bell Labs, though his initial paper on the topic was not published until 1982. Subsequent to that publication, numerous researchers have enhanced and broadened the algorithm, thereby expanding it to a broader spectrum of disciplines. In contexts such as image analysis, customer segmentation in marketing, or the analysis of geographical areas in satellite images, the K-means algorithm is designed to divide a set of data into K distinct clusters based on the similarity of each data point to other such points, thereby facilitating the grouping of data according to distinctive characteristics [7].

The main contributions of the current work include the evaluation of the performance of unsupervised machine learning techniques (K-means, X-means, ISODATA) with respect to classifying Sentinel-2 imagery for Mosul, Iraq, a region that remains relatively understudied in terms of remote sensing research. The challenges of a close spectral band and the 10-m pixel resolution of Sentinel-2 data make accurate classification particularly difficult, and this study addresses this issue by systematically comparing the clustering algorithms within Google Earth Engine and ArcMap environments across a large validation dataset.

2. Literature review

The classification of land cover has been the focus of a number of scientific projects that have incorporated a broad variety of machine-learning methods. Table 1 offers a comparison of related research.

Li et al. [8] examined multiple supervised and unsupervised image classification approaches, including ISO and k-means algorithms, in the context of Heze City, China. The authors concluded that ISO performed better than k-means in terms of clustering accuracy for satellite images. Nevertheless, they acknowledged that supervised classification methods exhibit higher overall accuracy. Dibs et al. [9] developed an effective approach for land-use and land-cover mapping in Baghdad, Iraq, using Landsat OLI images. Greatest likelihood classification produced the highest accuracy in that case, at 98.9 %. Nearby, Masini and Lasaponara [10] studied the use of satellite remote sensing to detect and quantify archaeological pillage in Syria, particularly at Tell Sheikh Hamad. They utilised an automatic feature extraction method (ALFEA) that successfully identified both previous and recent looting with high accuracy by utilising texture analysis and Google Earth images.

Highly related to the current study, Amal Muhammad Saleh [11] utilised Geographic Information Systems (GIS) to examine changes in land use and land cover in the Mosul District, Iraq. That survey revealed that urbanisation was increasing, especially after the violence in 2014. In IV and Tanya Jain [12], Al-Saydiya and Al-Hurriya, two neighborhoods in Baghdad City, were evaluated and categorised based on changes in land use and land cover across Landsat 8 satellite pictures in 2020. Green spaces made up a reduced percentage of Al-Hurriya, while in Al-Saydiya there was a noticeable decline in arable land. Environmental management and future city development in both areas thus profitted from the observed outcomes. Anul Haq [13] largely evaluated the changes in urban land-cover in Baghdad, Iraq, from 2003 to 2021. Landsat pictures were examined using maximum likelihood classification, which indicated a decrease of 16.4 % in urban areas and a reduction of 5.4 % in agricultural regions. For those working to decrease the detrimental effects of development on ecosystems, the study underlined the significance of balanced urban expansion plans.

Ouchra et al. [14] offered a Google Earth Engine and Landsat data-driven unsupervised land use categorisation technique for the Casablanca region. This identified k-means as fairly effective and LVQ as poor, with the two algorithms' efficacy in clustering different types of land cover used as a comparitor. Han and Lee [15] sought to identify satellite images using unsupervised models. To achieve this, they introduced the Inter-Image K-means Clustering Algorithm (IIkMC), an updated form of the k-means clustering algorithm. In order to guarantee consistent categorisation across several satellite images while substantially lowering calculation

Table 1. Comparative analysis of existing land-cover classification studie.

Author	Study Area	Method	Accuracy	Limitations
X. C. Li, L. L. Liu, and L. K. Huang [8].	Heze City, Shandong Province, China	- Supervised: Maximum Likelihood, Minimum Distance Unsupervised: ISODATA, K-means	Maximum Likelihood: 82.33 % (highest), Other methods: lower accuracy, ISODATA better than K-means.	Accuracy varies significantly by algorithm, Limited to Landsat imagery resolution,
H. Dibs, H. A. Hasab, J. K. Al-Rifaie, and N. Al-Ansari [9].	Baghdad, Iraq	Integration and fusion of multispectral Landsat OLI images using advanced classification techniques (including supervised classification and image fusion approaches)	Reported overall accuracy exceeded 85 %	Dependence on Landsat OLI spatial resolution (30 m), which restricts fine-scale urban detail.
N. Masini and R. Lasaponara [10].	Syria – archaeological sites affected by recent and past looting	(multi-temporal analysis of high-resolution imagery)	————————	Detection limited by spatial resolution of available satellite imagery.
F. W. A. Amal Muhammad Saleh [11].	Mosul District, Iraq	Remote sensing and GIS techniques applied to Landsat imagery	The study achieved reliable classification accuracy using supervised methods, with overall accuracy reported above 80 %	Dependence on medium-resolution Landsat imagery (30 m), limiting fine-scale urban detail
I. L. IV, M. J. K. Tanya Jain, Shubham Pandey, Shivam Mishra, Shubham Yadav [12]	India – using ISRO LISS IV satellite imagery	Application of machine learning algorithms (including supervised classification approaches such as Random Forest, Decision Trees, and SVM)	The study reported high classification accuracy, with Random Forest and SVM outperforming traditional methods. Overall accuracy exceeded 85–90	Dependence on the spatial resolution of LISS IV imagery (5.8 m), which, while detailed, still faces challenges in highly heterogeneous urban areas.
M. Anul Haq [13]	Case study using Planet Scope nanosatellite imagery (high-resolution, daily coverage)	Application of machine learning algorithms (including Support Vector Machines, Random Forest, and Decision Trees) to classify PlanetScope multispectral data.	Reported high classification accuracy, with Random Forest and SVM outperforming traditional classifiers. Overall accuracy exceeded 85–90 %,	Dependence on Planet Scope's spectral and spatial resolution (3–5 m), which, while detailed, may still face challenges in highly heterogeneous urban areas.
H. Ouchra, A. Belangour, and A. Erraissi [14].	Casablanca, Morocco	Landsat imagery processed on Google Earth Engine (GEE).	Random Forest achieved the highest classification accuracy (≈90 %),	Dependence on Landsat imagery (30 m resolution), limiting fine-scale urban detail.
S. Han and J. Lee [15].	series of satellite images (multi-temporal datasets)	Development of a parallelized inter-image k-means clustering algorithm for unsupervised classification of satellite image series.	Reported overall accuracy exceeded 80–85 %	As an unsupervised method, accuracy depends heavily on spectral separability and cluster initialization.

(*continued on next page*)

Table 1. (continued)

Author	Study Area	Method	Accuracy	Limitations
E. K. Hassan, H. M. Saeed, and A. H. T. Al-Ghrairi [16].	Australia – regions affected by wildfires	The study used supervised classification methods within GIS and remote sensing frameworks to quantify burned areas and vegetation loss.	he classification achieved high accuracy (≈80–85 %)	Dependence on medium-resolution satellite imagery, which may miss fine-scale fire impacts.
M. G. Mohammed [17].	Erbil City, Iraq	Application of GIS and remote sensing techniques	Reported overall classification accuracy exceeded 80 %,	Dependence on medium-resolution Landsat imagery (30 m), limiting detection of fine-scale urban changes.
W. H. H. Suhail Tuama Khazaal and Musa Habib Al-Shammari [18].	Karbala City, Iraq	Application of GIS techniques with satellite imagery to classify and assess land use/land cover (LULC).	Reported overall classification accuracy exceeded 80 %, with GIS	Dependence on medium-resolution satellite imagery, limiting detection of fine-scale urban changes.
K. Chomani and S. Pshdari [19].	Kurdistan Region, Iraq	Comparative evaluation of different classification algorithms (including Maximum Likelihood, Support Vector Machine, Random Forest, and Decision Tree)	The study found that machine learning classifiers (Random Forest and SVM) achieved higher accuracy (≈85–90 %) compared to traditional statistical methods such as Maximum Likelihood.	Dependence on medium-resolution satellite imagery, limiting detection of fine-scale urban and agricultural changes.
A. V. Memon, N. V. Shah, Y. S. Patel, and T. Parangi [20].	Case study using Sentinel-2 imagery	Comparative evaluation of classification algorithms (Random Forest, Support Vector Machine, Decision Tree, etc.) combined with dimensionality reduction techniques (such as PCA) to improve accuracy and efficiency in LULC mapping.	Dimensionality reduction improved computational efficiency and reduced redundancy while maintaining accuracy levels above 85 –90 %.	Dependence on Sentinel-2 spatial resolution (10–20 m), which may miss fine-scale urban changes.
S. A. K. Saad and E. A. Al-Zubaidi [21].	Erbil City, Iraq	Application of unsupervised classification methods (e.g., K-means, ISODATA) on Sentinel-2 satellite imagery	The unsupervised methods achieved moderate classification accuracy (≈70–75 %),	Lack of ground truth data reduced validation robustness.
M. Sharma, C. J. Kumar, and A. Deka [22].	Case study using Sentinel-2 satellite imagery	Comparative analysis of unsupervised clustering techniques (including K-means, ISODATA, and hierarchical clustering)	Reported moderate to high accuracy (≈75–85 %),	Dependence on Sentinel-2 resolution (10–20 m), limiting fine-scale urban detail.

time, the algorithm was fine-tuned utilising parallel processing. The results revealed considerable enhancements in classification accuracy and speed when tested using images from the Planet Scope mission.

Elsewhere, Hassan et al. [16] employed K-Means clustering and remote sensing approaches to categorise the land cover in Wollemi National Park, Australia. Trees, rivers, bare earth, structures with trees, and buildings without trees were the five land cover groups identified in the study, which focused on wildfire damage and its environmental implications. Mohammed [17] examined land cover changes in Erbil City, Kurdistan, between 2003 and 2020 using GIS and remote sensing. There was a dramatic drop in shrublands and desert terrain as urbanisation grew in that period, most notably in agricultural and built-up districts. Overall, the categorisation accuracy found within that study reached up to 90 %.

Suhail Tuama Khazaal [18] studied changes in land cover in Karbala city from 2000 to 2023 using data from Landsat 5 and 9. Areas of water and soil decreased, whereas built-up areas and vegetation expanded dramatically, according to that study's supervised categorisation methodology. The land cover categorisation used attained an overall accuracy of roughly 90 %, with a Kappa coefficient of 0.9.

Chomani and Pshdari [19] worked within Ranya, Sulaymaniyah, to explore various machine learning algorithms, including SVM, Maximum Likelihood, and Random Tree, with the goal of mapping land usage and land cover. The Random Tree approach with Sentinel-2 offered the best accuracy, although support vector machines were the most successful for deleting building footprints according to that study.

To improve the accuracy of land cover mapping using Sentinel-2 data, Memon et al. [20] evaluated alternative categorisation algorithms and dimensionality reduction procedures. This underlined the necessity of algorithm calibration and spectral band selection for unsupervised classification. Saad and Al-Zubaidi [21] used unsupervised categorisation of Sentinel-2 data to explore the topic of urban mapping in Erbil, Iraq. The clustering algorithms were thus reviewed and difficulties relating to precision and spectrum confusion addressed. More specifically, Sharma et al. [22] examined three approaches, K-means, hierarchical clustering, and DBSCAN, using Sentinel-2 data to categorise land cover. That study discussed and contrasted the various strategies for vegetation delineation and described their advantages and disadvantages.

3. Materials and methods

The current study adopted a different approach to most previous investigations, seeking to obtain a substantial improvement in classification accuracy by applying a large spectrum of unsupervised machine learning methods across six bands of Sentinel-2 images. The study thus utilised medium-resolution multispectral data sets with overlapping spectral bands in an attempt to address the problem of erroneous unsupervised satellite image classifications.

3.1. Materials

3.1.1. Study area

The geographic coordinates of Mosul, Iraq, are often given as 36.34°N, 43.12°E, though the latitude range is actually 36.20 to 36.45°N and the longitude range is 42.90 to 43.30°E. At an elevation of roughly 220 meters, Nineveh Governorate stretches over nearly 10,000 km^2, while the city is roughly 180 km^2. The annual precipitation in the area is fairly low at roughly 250 to 400 mm, with the bulk falling between November and April; there is also a dry period from May to October. The city's hydrology and agriculture are both influenced by the Tigris River, which passes through the city. Environmental and urban studies support Mosul as being well-positioned in terms of location and temperature.

The WGS84 UTM Zone 37 N projection is crucial for precision mapping and geographical analysis in the Mosul region. Northern Iraq, home to the city of Mosul, features a diversified environment that blends desert, farming, and urban districts. Land classification in Mosul has thus, of necessity, become a lengthy and careful procedure owing to the multiplicity of land uses [5,23]. To better comprehend the variations in land usage throughout the region, Mosul's land can be categorised into three fundamental groups: vegetative regions, residential areas, and desert areas.

The underlying situation regarding this Iraqi setting is, however, that the availability of reliable reference data is low, necessitating an emphasis on the most vital land cover categories that may directly affect environmental monitoring and resource management. The four classes most commonly used are thus Built-up, Vegetation, Bare Land, and Water Bodies, as most policy-relevant land cover types in Iraq. These classes are also vital in the analysis of urban expansion, assessment of agricultural productivity, monitoring of water resources, and desertification. All four classes thus

needed to feature in this study to ensure that the outputs could both be accurately interpreted and directly applied to national priorities.

Sentinel-2, one of the satellites in the Copernicus programme, managed by the European Space Agency, supplied the multispectral data utilised in this analysis. The data it produces contains both geographical and temporal resolutions. The complexity of extracting data on land coverage and classification from Sentinel-2 data is reduced, as its outputs encompass multiple spectral bands, extending from the visible to the near-infrared.

Among the numerous potential applications for the data acquired by satellites are the sectors of environmental monitoring, agriculture, disaster response, and climate change. In this work, Sentinel-2 Level 2 surface reflectance data, which is extensively adopted in practically all disciplines of environmental science [24], was thus used. This data included the adjusted ambient reflectance of S2A_MSIL2A T38SMF sensors. Sentinel-2 monitors 13 bands, including the visible (B2, B3, B4), the near-infrared (B8), and shortwave infrared (B11, B12) [25]. The resolution of the resulting geographical dataset was 20 m. A saturation mask of the pixels, as well as a water mask, snow mask, cloud mask, and shadow mask were developed. The gathered data bands were then integrated into overlapping "scenes" that covered an area of roughly 170 km^2 × 183 km^2 using a standard reference grid. Due to its geographical and environmental relevance in Iraq, this research focused on the city of Mosul [26]. Fig. 1 displays a map of the research region utilised in this investigation.

3.1.2. Classification of the reference data

In the Iraqi context, there are very few sources of authoritative and current reference data regarding land cover and remote sensing applications. To address this limitation, reference data based on supervised classification of high-resolution satellite imagery was created with the addition of expert information regarding local land use characteristics. This was subject to repeated refinement, graphical checks, and cross checks with other data such as Google Earth, local reports, and field knowledge, where available. While this solution could not provide true ground data in the exact meaning of the term, it was a viable solution to the existing state of data scarcity.

The accuracy measure in this study was based on reference data derived from verifiable satellite imagery; as far as accuracy is concerned, this approach is well accepted in the literature. The data was assembled in accordance with standardised protocols in order to achieve consistency and reduce the possibility of bias.

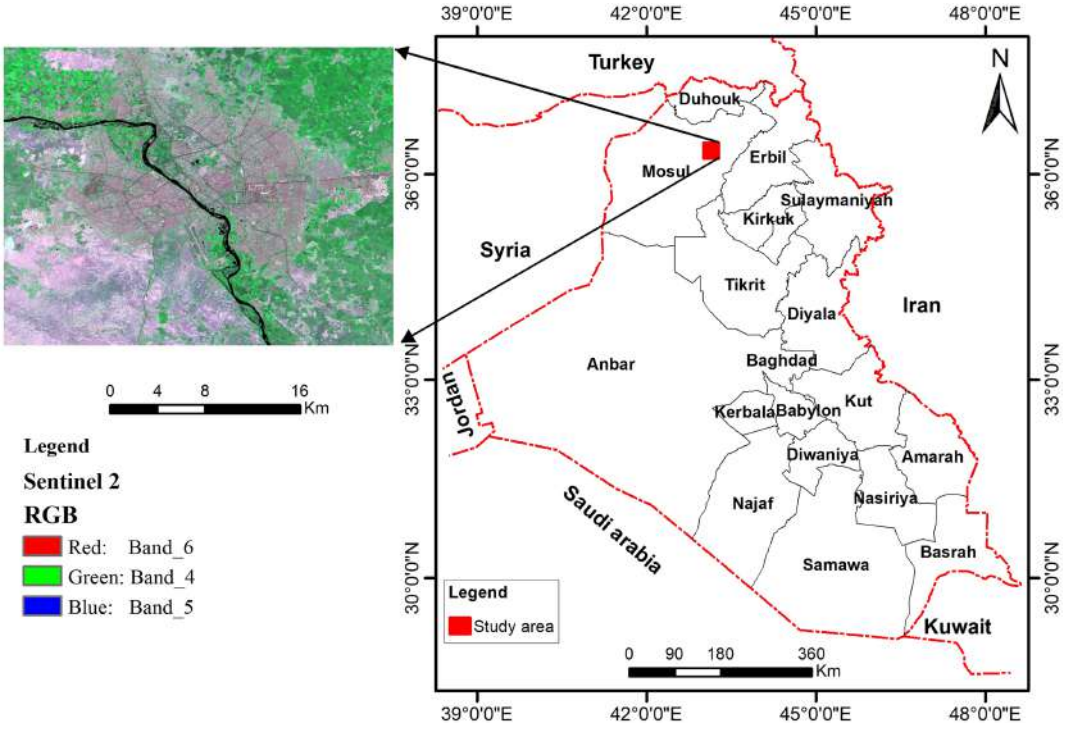

Fig. 1. Location map of the study area in Mosul, Iraq.

Some limitations remained, however. In particular, while the reference data offered good coverage in the majority of the study area, though certain regions were sampled less frequently due to accessibility factors. In addition, the absolute accuracy must be assumed to be influenced by minor time differences between the time of collecting the reference data and those noted in the primary dataset. These constraints were considered during analysis, and methodological modifications were made in order to counteract their possible effects, however.

3.1.3. Supervised learning as ground truth

The outcomes of supervised classification may be used as a proxy for reference data in the absence of direct field measurements or official datasets. The rationale behind the use of such supervised techniques is that such approaches generate results that approximate the actual land cover distribution in the real-world cue to the models being trained on closely controlled samples verified using expert interpretation. This therefore offers a more reliable comparator for unsupervised classification. Such an approach is both practical and necessary in those areas, such as Iraq, where logistical, political, and infrastructural factors prevent systematic field surveys.

There are certain implications to this, however: when ground truth takes the form of supervised results, the strengths and weaknesses of the unsupervised models are best identified by the latter's ability to reproduce known patterns. This approach thus offers a specific methodological testing framework for data-poor settings.

Nevertheless, in terms of limitations, the outputs under supervision must be acknowledged as being prone to classification errors, sampling bias, and overreliance on training data quality. They therefore cannot represent any absolute truth, and they may impose a threat of circular validation in which the unsupervised results are being compared only to potentially imperfect supervised maps.

3.2. Methods

This work sought to develop a technique for classifying satellite images using unsupervised learning approaches. The three primary stages of the proposed technique are represented in Fig. 2, with each stage being represented by a sequence of consecutive processes.

The choice to not to use any supervised approaches was not meant to ignore the possibility of applying supervised approaches, being instead based on the nature of the data: it was not possible to consistently and reliably label all data points with the same label across the entire sample. In such cases, controlled procedures can be biased and the results misleading: ground truth examples are usually scarce or uneven in remote sensing applications in such cases, and this can impair the generalisation performance of supervised classifiers. Comparatively, unsupervised methods present a more neutral topology in order to elicit inherent data structures, better fitting the aims of this study.

3.2.1. Data preparation stage

This work relied on retrieving and processing Sentinel-2 satellite data through the Google Earth Engine (GEE) platform by means of cloud-based resources. The recall period was set from January 4, 2024, to May 30, 2024. Images were filtered based on cloud ratio to guarantee data quality, with inclusion requiring a cloud ratio of less than 1 %. Table 2 shows the quantity of pixels prepared for training for each class. GEE offers several tools and functions to prepare multispectral data for classification. Loading and filtering imagery was done across image collections after the Image Collection () command was used to import datasets, while for temporal filtering, filterDate () was used to select images by date. In terms of spatial filtering, filterBounds () was then used to trim regions, while cloud removal was facilitated by the use of dataset-specific QA bands or algorithms such as Sentinel-2's SCL band or Landsat's pixel_qa. Compositing and mosaicking were done by applying the Median/Mean Composite to reduce each collection to a single image.

A cloud ratio filter was also used to reduce atmospheric and cloud effects. On choosing the number of bands, six bands, identified as B2, B3, B4, B8, B11, and B12, which represent Blue, Green, Red, NIR, SWIR, and SWIR, respectively, were chosen as appropriate for the search, from the total of 12 bands available on the Sentinel-2 satellite. After selecting these bands, a singular raster was created featuring all of them. The bands were chosen due to complementary nature of the spectral information they offer, which aided in distinguishing between vegetation, water, soil, and built-up areas. Such combinations are common in land cover research, as they assure the reproducibility of results.

3.2.2. Classification stage

Both K-means and X-means algorithms were implemented using the Google Earth Engine platform. The process of identifying satellite images

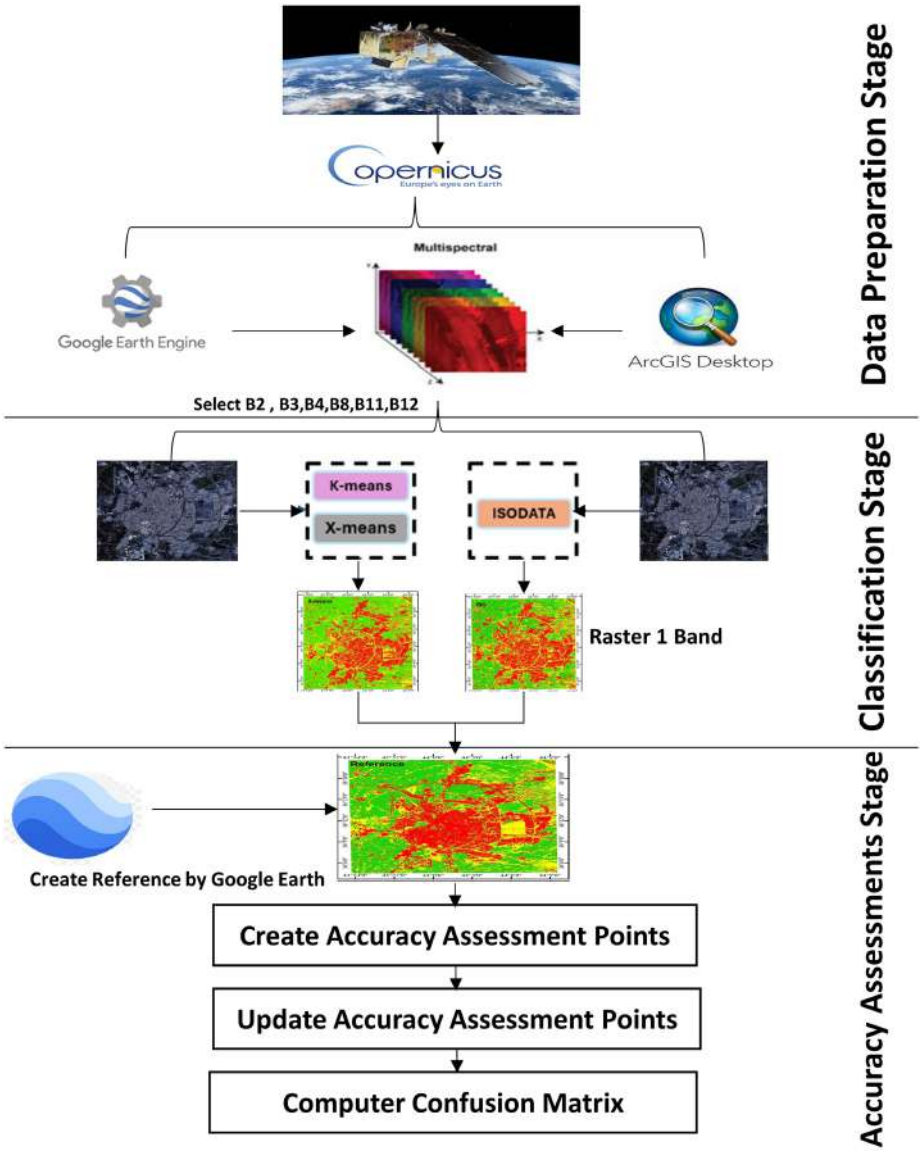

Fig. 2. Workflow of the proposed methodology for land-cover classification.

using Google Earth Engine is thus explained below. In this work, the number of clusters was determined according to the types of the targeted areas in each study region in terms of Built-up, Water, Vegetation, and Bare land. The process began by determining initial random locations for cluster centres in each case. The algorithm then determined the Euclidean distance between each pixel and the centres of the current clusters, with each pixel then assigned to the cluster closest to its centre. After the pixels were dispersed to the clusters, the coordinates of the cluster centres were modified to match the average values of all pixels assigned to their cluster. This process continued iteratively until the clusters reached a level of stability such that the cluster centres did not change at the end of a given iteration. The results emerged in the form of classed maps, which enabled the examination of distinct areas depending on the spectral properties of the employed bands. This approach offers an excellent strategy for supporting comprehension of the distribution of land in the

Table 2. Number of pixels per class in reference and classified raster.

Reference Data		ISO	k-Means	X-Means
Water	25581	28246	29353	27626
Bulit-up	557729	668040	399777	540047
Vegetation	731529	629266	945735	399735
Bare land	678116	667412	61353	1025556
Total	1992955	498241	359054.5	498241

study area and classifying it appropriately. This also allowed for the analysis of regions that share similar qualities, based on assigning them to distinct classifications across Built-up, Vegetation, Bare Land, and Water Bodies. The mathematical formulations for the necessary clustering algorithms K-Means, X-Means, and ISO Data are offered below:

K-Means Clustering: The K-Means algorithm aims to minimise the following objective function (also called the within-cluster sum of squares) [27]:

$$J = \sum_{i=1}^{k} \sum_{x \in Ci} \| x - \mu i \|^2 \quad (1)$$

where k = number of clusters, Ci = set of points in cluster I, μi = centroid of cluster i (mean of points in Ci), x = data point, and $\| x - \mu i \|^2$ = squared Euclidean distance between point x and centroid μi.

X-Means Clustering: X-Means is an extension of K-Means that automatically determines the optimal number of clusters. It is thus similar to K-Means, while incorporating a Bayesian Information Criterion (BIC) to determine the relevant k [28].

$$\mathrm{BIC} = \log L - \frac{P}{2} \log n \quad (2)$$

where L = log-likelihood of the data, p = number of parameters (k*d, where d is dimension), and n = number of data points.

ISO Data Clustering: the ISODATA algorithm is an extension of K-means with additional adaptive cluster splitting and merging:

Merge Criterion

$$d(i, j) < \varnothing C \text{ and } \| Ci \| < \varnothing N \text{ or } Cj < \varnothing N \quad (3)$$

Split Criterion

$$i > \varnothing S \ (\text{where } i = Mxa(i1 \ldots id)) \quad (4)$$

Cluster Centre Update

$$i = \frac{1}{Ci} \sum_{zj \in Ci} xj \quad (5)$$

where

i j: Cluster centres for clusters.

Ci Cj : Sets of data points assigned to clusters.

Ci : Number of samples in cluster.

$d(i, j)$: Distance between cluster centres (Euclidean).

$\varnothing C$: Threshold distance for merging clusters.

$\varnothing N$: Minimum cluster size threshold.

i: Standard deviation of cluster.

id: Standard deviation of cluster along dimension.

$\varnothing S$: Variance threshold for splitting clusters.

3.2.3. Accuracy assessment

Each technique was assessed for accuracy through the comparison of its classification results with reference data; such assessment of classification accuracy thus relies on reference data specific to each class. The precision assessment of remote sensing classification using reference data within GEE took the form of reference data evaluation. Such information is commonly necessary for specific projects, and it may include the data sources or codes, such as Landsat, Sentinel-2, and MODIS, as these are required to conduct model calibration and result verification. The documentation on GEE or the websites of the organisations supplying this data (such as NASA or the European Space Agency) must thus be used as references for the data. Open-source datasets, which can also serve as reference data, are commonly made available by government organisations and university institutions in various developed nations that specialize in remote sensing.

The reference data sources used in this article are shown in Fig. 3, and these facilitated the development of supervised classification of the study area and a mapping programme, based on the Google Earth platform. As supervised classification is commonly accepted as the best method for allowing comparison of remote sensing results with unsupervised classification results for the same study area, 3000 points were evenly distributed across four classes (Built-up, Vegetation, Bare land, and Water body) in ArcMap in order to produce accurate results.

Accuracy matrices or confusion matrices offer an organised test of the effectiveness of a classification model, dissecting the performance of the model across numerous land cover groupings. A confusion matrix is a tabular form of representation of results from comparing the predicted labels and the actual reference labels, with the rows and columns illustrating the various land cover categories. Here, the term True Positives (TP) is used to refer to the count of pixels classified correctly to a particular class; False Positives (FP) refers to those pixels that have been assigned to that particular class incorrectly; False Negatives (FN) refers to the pixels that have been assigned to another particular class incorrectly, while True Negatives (TN) are those pixels that are correctly omitted from the class in question. Such measures enable the determination of accuracy measures regarding the producer, user, and overall scores, as well as the kappa coefficient which can be taken as a complete measurement of the reliability of classification. Moreover, the F-score is especially practical where the class

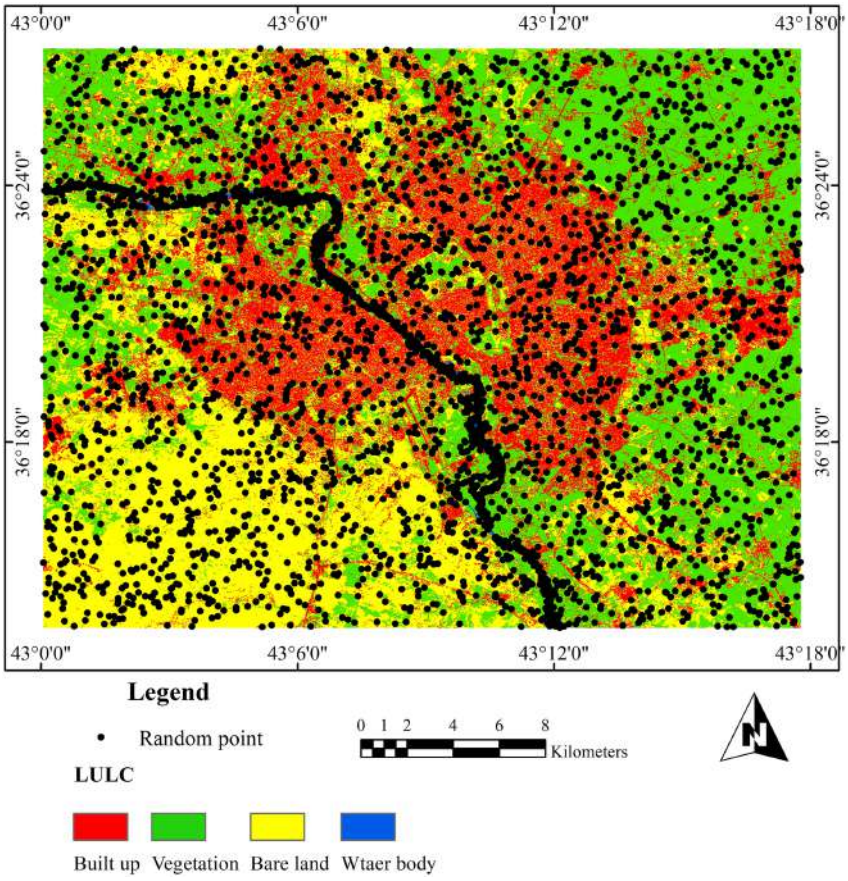

Fig. 3. Random point allocation within the study area (3000 Samples).

distributions are uneven, such as assessing differences between water bodies and non-water pixels in Sentinel-2 images, when the precision and recall should be balanced to promote the hard detection of a minority group. This framework guarantees that classification performance is judged similarly regardless of all the categories of land cover instead of being restricted to two-way situations [29].

By integrating sensitivity to false positives and false negatives, the F1-Score offers a comprehensive aid to the assessment of the model's efficacy across unbalanced groups. The Cohen's Kappa formula may be expressed in the conventional two-dimensional confusion matrix used in machine learning, with the OA and Kappa are calculated using the following formulae [30–33]:

$$\text{Overall Accurcy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

$$\text{kappa} = \frac{Po \parallel Pe}{1 \parallel Pe} \tag{7}$$

$$Po = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

$$Pe = P(+1) + P(\parallel 1) \tag{9}$$

$$P(+1) = \frac{TP + FN}{TP + TN + FP + FN} \times \frac{TP + FP}{TP + TN + FP + FN} \tag{10}$$

$$P(\parallel 1) = \frac{FP + TN}{TP + TN + FP + FN} \times \frac{FN + TN}{TP + TN + FP + FN} \tag{11}$$

$$PA = \frac{TP}{TP + FN} \tag{12}$$

$$UA = \frac{TP}{TP + FP} \tag{13}$$

$$F1 \parallel \text{score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{14}$$

4. Results and discussion

4.1. Qualitative analysis

Fig. 4 displays the classification maps of the research region using all the techniques utilised as

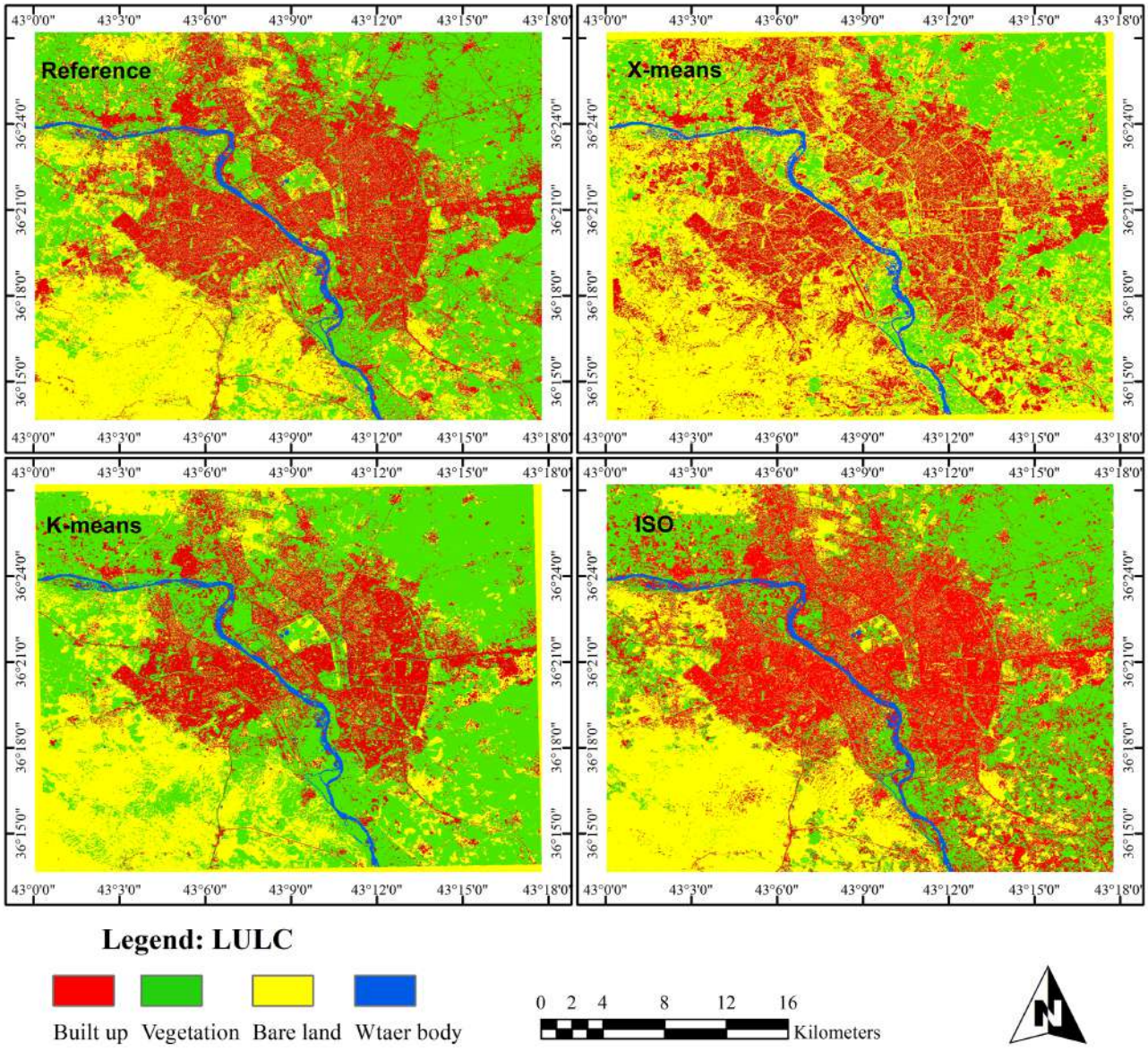

Fig. 4. Comparative land-cover maps: K-means, X-means, ISODATA vs. Reference data.

compared to the reference picture. The results of the unsupervised classification of the Mosul region using the ISODATA algorithm in ArcMap and the K-means and X-means algorithms in Google Earth Engine demonstrated pronounced variations in terms of the efficiency and accuracy of classification into the four classes of Buit-up, Vegetation, Bare land, and Water. An outline of the most noteworthy differences is thus presented below:

K-means method: The K-means method could discriminate between the four classes with an 80.14 % classification accuracy, as it was able to discern between green and arid plains, as well as identifying the reasonably distinct appearance of urban blocks and structures. It thus offers a suitable alternative for categorising land with a high degree of accuracy and reasonable time efficiency, due to its average speed.

X-means algorithm. In contrast to the K-means algorithm, the X-means method demonstrated more adaptability in the categorisation of the classes; this is facilitated by it deciding on the appropriate number of classes on its own, which aided the attainment of reasonably accurate discrimination of green areas. While it offered a broader view of the spectral classes across land types, X-means had a slightly lower accuracy of roughly 76.34 % as compared to K-means, which might require future investigation. Due to the extra calculations required to establish the classes initially, its pace was decent,

yet it lagged behind that of K-means. ISODATA Algorithm: With an accuracy of around 73.61 %, the ISODATA algorithm in ArcMap performed less effectively than either of the Google Earth Engine algorithms.

While spectral indices such as NDVI and NDBI are popular because of their simplicity and easily available, they are relatively simple tools that might not reflect the full complexity of land-cover variability in heterogeneous landscapes. In contrast, machine learning methods, including Deep learning, Support Vector Machines, and Random Forest, provide more advanced functionality by utilising a wider range of spectral and spatial data from remote sensing devices. They can thus model non-linear relationships, cope with high-dimensional data, and give strong classification outputs despite the occurrence of noise or the lack of ground truth examples. Introduction of artificial intelligence into remote sensing processes thus supports considerable methodological innovation, leading to more precise, scalable, and transferable classification results than historic index-based processes.

4.2. Quantitative analysis

The results of the research allow the addition of add confidence to the conclusions via statistical proof. Tables 3–7 show the confusion matrices for all three classes, which are similarly distributed and contain between 1000 and 5000 points. Table 8 clarifies the links between the scattered points, F1 score, kappa, and accuracy category, while Fig. 5 presents a comparison chart of the algorithms utilised in terms of accuracy.

Table 7 of the confusion matrix shows a significant error in terms of the confusion of bare land pixels with water classifications. This mismatch is mainly found in transitional areas where the soil surface resembles shallow or turbid water, a phenomenon based on both mixed-pixel effects and neighbour effects. The changing moisture conditions between seasons also added to this confusion, especially in regions where temporary water storage changed spectral measurements. While the general performance of the K-means algorithm was good, this weakness highlights its lower discriminatory capabilities between spectrally similar classes. To mitigate this, class-specific accuracy measurements are required: the relevance of K-means is dependent on the application scenario, and this may thus involve the need for auxiliary data or post-classification amendments to enhance accuracy.

The dependence on imagery obtained in the wet season in Mosul (January to May 2024) also lent itself to confusion of spectral classes between built-up areas and bare land. Damp soil surfaces during the season tend to reflect in a similar manner to concrete or shallow water, resulting in misidentification within transition areas. This seasonal effect contributed to the errors observed in Table 7, highlighting the need to take into account temporal conditions when interpreting results.

Three unsupervised classification algorithms—K-means, X-means, and ISODATA—were evaluated using Sentinel-2 data. While X-means and ISODATA are intended to enhance traditional K-means by automatically estimating cluster numbers or enabling cluster splitting and merging, their

Table 3. Classification validation results based on 1000 random points.

Class	Water	Vegetation	Build-up	Bare land
Water	165	2	21	4
Vegetation	34	232	68	15
Build-up	51	15	159	3
Bare land	0	1	2	227

Table 4. Classification validation results based on 2000 random points.

Class	Water	Vegetation	Build-up	Bare land
Water	282	9	39	4
Vegetation	71	447	161	29
Build-up	146	42	299	9
Bare land	1	2	1	457

Table 5. Classification validation results based on 3000 random points.

Class	Water	Vegetation	Build-up	Bare land
Water	478	7	71	4
Vegetation	100	687	216	56
Build-up	170	56	463	9
Bare land	2	0	0	680

Table 6. Classification validation results based on 4000 random points.

Class	Water	Vegetation	Build-up	Bare land
Water	589	16	79	7
Vegetation	164	910	268	81
Build-up	241	73	647	17
Bare land	6	1	1	895

Table 7. Classification validation results based on 5000 random points.

Class	Water	Vegetation	Build-up	Bare land
Water	748	25	109	6
Vegetation	194	1147	375	84
Build-up	301	76	762	21
Bare land	1249	2	2	1137

Table 8. Validation results of classification algorithms using accuracy, Kappa, and F1 score.

Random Point	Accuracy			Kappa			F1-score		
	X-means	ISO	K-means	X-means	ISO	K-means	X-means	ISO	K-means
1000	78.38 %	74.90 %	79.80 %	71.17 %	66.53 %	73.07 %	78.36 %	75.38 %	80.15 %
2000	74.29 %	73.60 %	79.95 %	65.72 %	64.80 %	73.27 %	74.22 %	74.25 %	80.25 %
3000	76.96 %	72.83 %	80.47 %	69.28 %	63.78 %	73.96 %	76.89 %	73.78 %	80.77 %
4000	76.12 %	73.36 %	80.00 %	68.16 %	64.49 %	73.33 %	76.11 %	74.08 %	76.11 %
5000	75.96 %	73.37 %	80.50 %	67.94 %	64.50 %	74.01 %	75.86 %	74.19 %	80.83 %
AVERAGE	**76.34 %**	**73.61 %**	**80.14 %**	**68.45 %**	**64.82 %**	**73.52 %**	**76.29 %**	**74.33 %**	**79.62 %**

Values in bold refer to the final average score of the class, which summarizes the overall performance results.

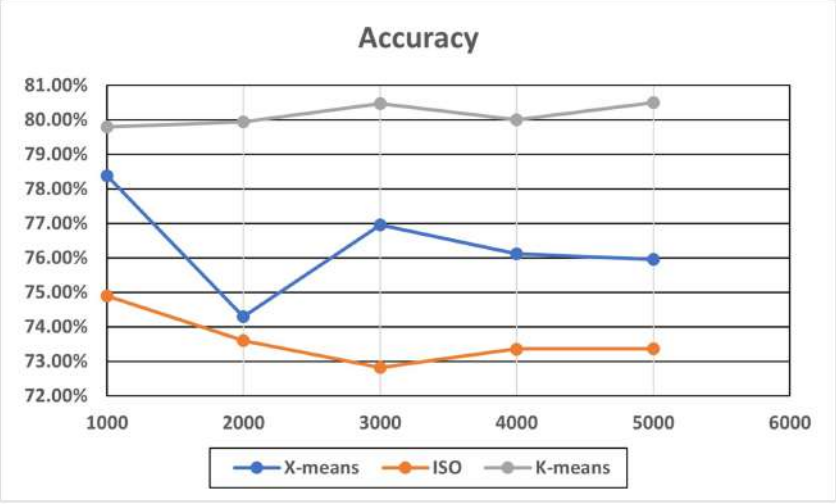

Fig. 5. Comparative overall accuracy across K-means, X-means, and ISODATA.

performance in this context was unstable. X-means exhibited high sensitivity to initialization and frequently generated inconsistent cluster counts. ISODATA required adjustment of multiple parameters, such as thresholds for merging and splitting, which increased variability and reduced reproducibility.

In contrast, K-means demonstrated superior overall performance and achieved higher Kappa values. Its simplicity and robustness enabled effective capture of spectral variability in the study area without extensive parameter tuning or increased risk of overfitting. This stability is especially valuable in situations with limited ground truth data, where reproducibility and methodological clarity are essential. These results align with recent studies on Sentinel-2 imagery (e.g., [M. G. Mohammed, 2023]; [W. H. H. Suhail Khazaal, 2024]), which have identified K-means as a reliable baseline for unsupervised classification in heterogeneous landscapes.

In addition to confirming the utility of K-means, this research systematically benchmarks advanced clustering techniques against a widely used baseline. Although X-means and ISODATA provide theoretical flexibility, their practical limitations related to sensitivity and parameter dependence constrain their operational applicability. Given these constraints, future research could investigate hybrid approaches that integrate the stability of K-means with adaptive features from advanced algorithms to enhance both accuracy and reproducibility. Overall, the findings reinforce K-means as a feasible and efficient standard model for unsupervised classification of Sentinel-2 data.

5. Conclusion

The study was based in Mosul, Iraq, which is a relatively important area regionally that has nevertheless not been well explored by remote sensing. The research took advantage of the high-quality features of Google Earth Engine (GEE) and desktop tools such as ArcMap, as well as a substantial validation dataset, to strengthen the validity of its conclusions. Remote sensing analysis across various applications is thus improved by the results of this work.

K-means was found to be the best clustering algorithm when compared to X-means and ISODATA

in the Mosul case study, with a general accuracy of 80.14 %. Due to its ease of use, computational efficiency, and strong clustering capabilities, K-means is the current model of choice for practical use in the sphere of remote sensing in the real world, especially in the context of urban monitoring and land-use classification, where answers need to be delivered in good time to drive action.

An examination of errors in class-level attribution indicated that most of the mistakes occurred in terms of distinguishing between agricultural land and barren land, an issue caused by spectral confusion, as these have similar reflectance structures in arid areas of Iraq. Agricultural fields may be misclassified through having spectral signatures resembling those bare soil due to seasonal variability and sparse vegetation cover, which increases overlap. Some confusion also existed in terms of comparing urban settlements with barren land, especially at the periphery settlements where building materials have the same spectral characteristics as the soil. Water bodies were the least commonly confused, though in rare cases, small reservoirs were confused with agricultural fields due to adjoining effects and mixed pixels.

Future studies must also incorporate hybrid clustering models, spectral indices, and supervised machine learning algorithms to increase the accuracy of classification. Generalisability can be enhanced by expanding studies to use multi-temporal data from Sentinel-2 and by the application of uncertainty analysis and models to additional regions of Iraq to enable the development of more actionable information on urban surveillance and environmental management.

Source of Funding

This work was not supported by any funding.

Conflict of Interest

No conflicts of interest are disclosed by the authors.

Ethical Approval

This study did not involve human participants, animals, or sensitive data. Therefore, ethical approval was not required.

Data Availability

This published article contains all of the data produced or examined during this investigation.

Author Contributions

Ehsan Ali Al-Zubaidi; Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing — original draft.

Mohammed Ridha Hammoodi; Conceptualization, Data curation, Investigation, Methodology, Supervision, Validation, Writing — review & editing.

Ahmed Naser Alzurfi; Formal analysis, Methodology, Software, Supervision, Validation, Visualization, Writing — original draft, Writing — review & editing.

References

[1] Li A, Fan M, Qin G, Xu Y, Wang H. Comparative analysis of machine learning algorithms in automatic identification and extraction of water boundaries. Appl Sci 2021;11(21). https://doi.org/10.3390/app112110062.

[2] Basheer S, Wang X, Farooque AA, Nawaz RA, Liu K, Adekanmbi T, et al. Comparison of land use land cover classifiers using different satellite imagery and machine learning techniques. Remote Sens 2022;14(19). https://doi.org/10.3390/rs14194978.

[3] Swetanisha S, Panda AR, Behera DK. Land use/land cover classification using machine learning models. Int J Electr Comput Eng 2022;12(2). https://doi.org/10.11591/ijece.v12i2.pp2040-2046.

[4] Athiyah U, Rustad S, Soeleman MA, Muljono, Akrom M. Effectiveness of K-means clustering-based outlier handling techniques in colonic condition endoscopy image classification using CNN. In: 2024 IEEE international conference on communication, networks and satellite (COMNETSAT); 2024. p. 276–82. https://doi.org/10.1109/comnetsat63286.2024.10862507. Mataram, Indonesia.

[5] El-Omairi MA, El Garouani A. A review on advancements in lithological mapping utilizing machine learning algorithms and remote sensing data. Heliyon Sep 2023;9(9): e20168. https://doi.org/10.1016/j.heliyon.2023.e20168.

[6] Mantilla L. Classification of satellite images using Rp fuzzy c means for unsupervised classification algorithm. In: 2019 IEEE Colombian conference on applications in computational intelligence (ColCACI); 2019. p. 1–5. https://doi.org/10.1109/colcaci.2019.8781988. Barranquilla, Colombia.

[7] Tibshirani R, Walther G, Hastie T. K-Means clustering and related algorithms. Technical report. Stanford University; 2004. https://doi.org/10.17760/d20291355.

[8] Li XC, Liu LL, Huang LK. Comparison of several remote sensing image classification methods based on Envi. Int Arch Photogram Rem Sens Spatial Inf Sci 2020;XLII-3/W10: 605–11. https://doi.org/10.5194/isprs-archives-xlii-3-w10-605-2020.

[9] Dibs H, Hasab HA, Al-Rifaie JK, Al-Ansari N. An optimal approach for land-use/land-cover mapping by integration and fusion of multispectral landsat oli images: case study in Baghdad, Iraq. Water, Air, Soil Pollut 2020;231(9). https://doi.org/10.1007/s11270-020-04846-x.

[10] Masini N, Lasaponara R. Recent and past archaeological looting by satellite remote sensing: approach and application in Syria. In: Remote sensing for archaeology and cultural landscapes. Springer Remote Sensing/Photogrammetry; 2020. p. 123–37. https://doi.org/10.1007/978-3-030-10979-0_8. ch. Chapter 8.

[11] Amal Muhammad Saleh FWA. Analyzing land use/land cover change using remote sensing and GIS in Mosul District, Iraq. Annals Romanian Soc Cell Biol 2021;25(3): 4342–59. https://doi.org/10.21275/v5i2.nov161379.

[12] Jain T, Khan MJ, Pandey S, Mishra S, Yadav S. Satellite image classification and analysis using machine learning with ISRO LISS IV. https://doi.org/10.26706/ijceae.2.2.20210404; 2022.
[13] Anul Haq M. Planetscope nanosatellites image classification using machine learning. Comput Syst Sci Eng 2022;42(3):1031–46. https://doi.org/10.32604/csse.2022.023221.
[14] Ouchra H, Belangour A, Erraissi A. Comparison of machine learning methods for satellite image classification: a case study of Casablanca using landsat imagery and google earth engine. J Environ Earth Sci 2023;5(2):118–34. https://doi.org/10.30564/jees.v5i2.5928.
[15] Han S, Lee J. Parallelized inter-image k-means clustering algorithm for unsupervised classification of series of satellite images. Remote Sens 2023;16(1). https://doi.org/10.3390/rs16010102.
[16] Hassan EK, Saeed HM, Al-Ghrairi AHT. Classification and measurement of land cover of wildfires in Australia using remote sensing. Iraqi J Sci 2022:420–30. https://doi.org/10.24996/ijs.2022.63.1.38.
[17] Mohammed MG. Land use land cover changes detection of Erbil City using GIS and remote sensing. Polytechnic Journal 2023;13(1). https://doi.org/10.59341/2707-7799.1729.
[18] Khazaal ST, Hussan WH, AL-Shammari MH. Land use/Land cover assessment for Karbala City by using geographic information systems (GIS) technique. https://doi.org/10.65115/7zq3sp06; 2024.
[19] Chomani K, Pshdari S. Evaluation of different classification algorithms for land use land cover mapping. Kurdistan Journal of Applied Research 2024;9(2):13–22. https://doi.org/10.55277/researchhub.2vexo15g.1.
[20] Memon AV, Shah NV, Patel YS, Parangi T. Enhancing land use/land cover analysis with sentinel-2 bands: comparative evaluation of classification algorithms and dimensionality reduction for improved accuracy assessment. Nat Environ Pollut Technol 2025;24(2). https://doi.org/10.46488/nept.2025.v24i02.b4264.
[21] Saad SAK, Al-Zubaidi EA. Classification of satellite Sentinel-2 imagery using unsupervised methods: a case study of Erbil. In: AIP Conf Proc, 3318. AIP Publishing LLC; 2025. p. 030028. https://doi.org/10.1063/5.0286607. 1.
[22] Sharma M, Kumar CJ, Deka A. Land cover classification: a comparative analysis of clustering techniques using Sentinel-2 data. Int J Sustain Agric Manag Inform 2021;7(4):321–42. https://doi.org/10.1504/ijsami.2021.122008.
[23] Rahman A, Abdullah HM, Tanzir MT, Hossain MJ, Khan BM, Miah MG, et al. Performance of different machine learning algorithms on satellite image classification in rural and urban setup. Remote Sens Appl: Society and Environment 2020;20. https://doi.org/10.1016/j.rsase.2020.100410.
[24] Zhang P, Hu X, Ban Y, Nascetti A, Gong M. Assessing Sentinel-2, Sentinel-1, and ALOS-2 PALSAR-2 data for large-scale wildfire-burned area mapping: insights from the 2017–2019 Canada Wildfires. Remote Sens 2024;16(3). https://doi.org/10.3390/rs16030556.
[25] Marlina D. Land cover classification in sentinel-2 image Kuningan District with NDVI and algorithm random forest. https://doi.org/10.30998/string.v7i1.12948; 2022.
[26] Al-Zubaidi EA, Al-Sulttani AH, Rabee F. Sand dunes spectral index determination using machine learning model: case study of Baiji Sand Dunes Field Northern Iraq. Iraqi Geological Journal 2022;55(1F):102–21. https://doi.org/10.46717/igj.55.1f.9ms-2022-06-24.
[27] MacQueen J. Some methods for classification and analysis of multivariate observations. In: Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, Volume 1: Statistics, 5. University of California press; 1967. p. 281–98. https://doi.org/10.1002/0471271357.ch9.
[28] Pelleg D, Moore A. X-means: extending K-means with efficient estimation of the number of clusters. In: ICML'00; 2000. p. 727–34. https://doi.org/10.7717/peerj-cs.2516/fig-1. Citeseer.
[29] Congalton RG. Accuracy assessment and validation of remotely sensed and other spatial information. Int J Wildland Fire 2001;10(4). https://doi.org/10.1071/wf01031.
[30] Congalton RG, Green K. Assessing the accuracy of remotely sensed data: principles and practices. CRC press; 2019. https://doi.org/10.1201/9781420055139-12.
[31] Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960;20(1):37–46. https://doi.org/10.1177/001316446002000010.
[32] Foody GM. Status of land cover classification accuracy assessment. Rem Sens Environ 2002;80(1):185–201. https://doi.org/10.1016/s0034-4257(01)00295-4.
[33] Sokolova M, Lapalme G. A systematic analysis of performance measures for classification tasks. Inf Process Manag 2009;45(4):427–37. https://doi.org/10.1016/j.ipm.2009.03.002.