Supervised Classification Accuracy Assessment Using Remote Sensing and Geographic Information System

Article *in* TEM Journal · February 2024

DOI: 10.18421/TEM131-41

3 authors, including:

Khalid H. Abbas Al-Aarajy
University of Baghdad

Ahmed Asaad Zaeen
University of Baghdad

Supervised Classification Accuracy Assessment Using Remote Sensing and Geographic Information System

Khalid H. Abbas Al-Aarajy [1], Ahmed A. Zaeen [2], Khaleel I. Abood [2]

[1] *Remote Sensing and Geographic Information System Dept, College of Science, University of Baghdad, Iraq*
[2] *Remote Sensing Unit, College of Science, University of Baghdad, Iraq*

***Abstract* – Assessing the accuracy of classification algorithms is paramount as it provides insights into reliability and effectiveness in solving real-world problems. Accuracy examination is essential in any remote sensing-based classification practice, given that classification maps consistently include misclassified pixels and classification misconceptions. In this study, two imaginary satellites for Duhok province, Iraq, were captured at regular intervals, and the photos were analyzed using spatial analysis tools to provide supervised classifications. Some processes were conducted to enhance the categorization, like smoothing. The classification results indicate that Duhok province is divided into four classes: vegetation cover, buildings, water bodies, and bare lands. During 2013-2022, vegetation cover increased from 63% in 2013 to 66% in 2022; buildings roughly increased by 1% to 3% yearly; water bodies showed a decrease of 2% to 1%; the amount of unoccupied land showed a decrease from 34% to 30%. Therefore, the classification accuracy was assessed using the approach of comparison with field data; the classification accuracy was about 85%.**

***Keywords* – Accuracy, classification, ground truth data, remote sensing, satellite image.**

DOI: 10.18421/TEM131-41
https://doi.org/10.18421/TEM131-41

Corresponding author: Ahmed Asaad Zaeen,
Remote Sensing Unit, College of Science, University of Baghdad, Iraq
Email: ahmed.asaad@sc.uobaghdad.edu.iq

Received: 02 October 2023.
Revised: 19 December 2023.
Accepted: 17 January 2024.
Published: 27 February 2024.

1. Introduction

The United States Geological Survey has defined spatial data accuracy as the results' closeness of statements, computations, or assessments to the true values or the values assumed to be true [1]. Any supervised classification is not considered complete until its accuracy is assessed [2]. Identifying and eliminating classification errors causes suffering from less attention regarding this issue. So, this article defines a specific error source that may occur, requiring independent pixel samples from each category [3]. The reference condition, the most suitable available assessment of the ground condition, plays a necessary role in precision assessment and area estimation [4]. Precision assessment is a decisive step in processing remote sensing data [5], where digital elevation models (three-dimensional representations of the earth's surface) are the primary sources of height information that are greatly applied in many fields [6]. With advanced digital satellite remote sensing approaches, the necessity of performing an accurate assessment has received continued interest. That does not mean precision estimation is inconsequential for traditional remote sensing techniques. However, due to the complexity of digital classification, it is necessary to evaluate the reliability of the outcomes [7]. Remote sensing can detect and monitor Earth's surface features employing satellite images with various radiometric, spatial, spectral, and temporal resolutions [8].

The shift in land use and land cover (LULC) impacts the city's geophysical state and causes environmental disturbances like climatic and LST changes [9]. These procedures should use current approaches, such as GIS techniques and remote sensing, to encourage investigating the current state of cities and the issues they suffer from. GIS has developed tremendously and has been exploited in many domains [10] and [11]. The article aims to assess and improve the image classification approach precision.

2. Methodology

The methodology applied to the study area includes several steps, the most important of which is collecting data from Landsat 6 and Landsat 9 and then performing the necessary treatments on that data to arrive at the best classification of land use in the study area for the period between 2002 and 2022, and then using a modern method to measure the accuracy of the classification.

2.1. Study Area

The research area is located in northern Iraq, Kurdistan region, Duhok province, between latitudes 37°00′00″ N and 37°07′30″ N and longitudes 42°27′30″ E and 42°47′30″ E [12], and 585 m above the sea level [13]. It has a population of over 340,900 (2018). It is located on the northern side of Iraq, with a total area of 10,955.91 km^2. Amedi, Simele, Duhok Center, Zakho, and other cities are all part of Duhok province. Two chains of mountains embrace Duhok province; these mountains confer a linear shape and a special landscape on the city. The province is located northwest of Iraq and the western part of the Kurdistan region, about 470 km north of Baghdad and 430–450 m above sea level. Along with five other districts, Sumeal, Zakho, Amedy, Sheikhan, and Akre, Duhok province is administered by the province. Duhok covers 10715 km2 at latitude 36 north and longitude 43 east. Two rivers pass through the city: the Duhok River and the smaller and seasonal Heshkarow River. Both rivers meet southwest of the city, and the water is primarily employed for irrigation, which helps preserve the nearby green areas (Fig 1).

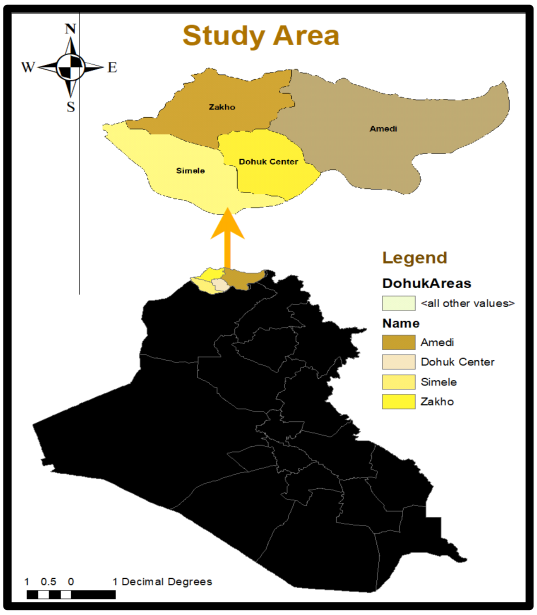


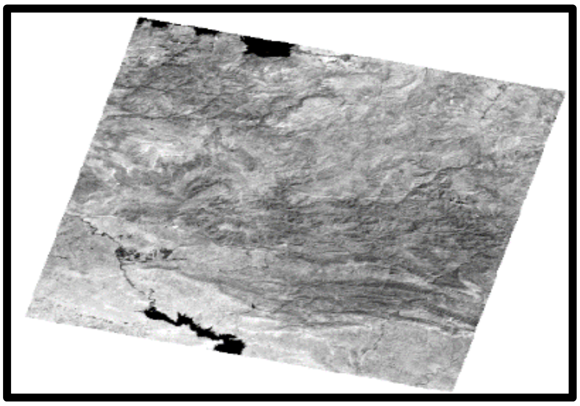
Figure 1. The location of Duhok province in Iraq

2.2. Data (Satellite Image)

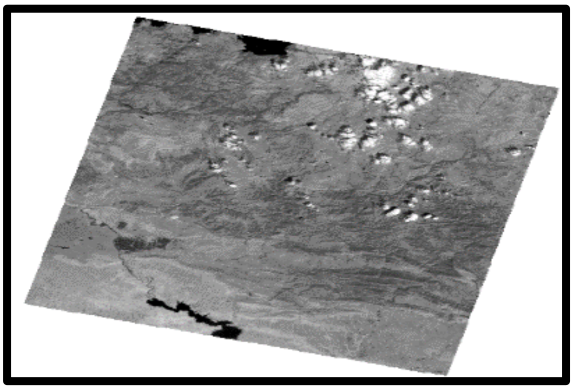
Satellite images obtained from Landsat-7 and 9 were used to study the land cover. A super classification using the geographic information system (GIS) program was conducted to classify the study area. Smoothing operations were conducted on the outputs; classification accuracy was assessed using suitable evaluation methods.

3. Steps of Work

First, satellite imagery was used for 2013 and 2022 obtained from Landsat-8, Fig. 2 (a and b).



(a)

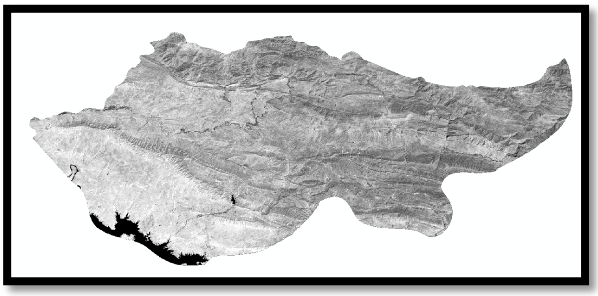


(b)

Figure 2. Satellite imagery in (a) 2013, (b) 2022

3.1. Clipping

In GIS, "clip" refers to overlaying a polygon over one or more target features (layers) and extracting the target feature data inside the clip polygon's defined area. In other words, the first polygon is forced to take on the second polygon's borders. The initial polygon feature no longer includes any of the remaining space. Utilizing the clipping tool can help take a portion of the dataset and maintain the necessary features for the study area; it allows for extracting the features of the study area, Fig 3 (a and b).



(a)

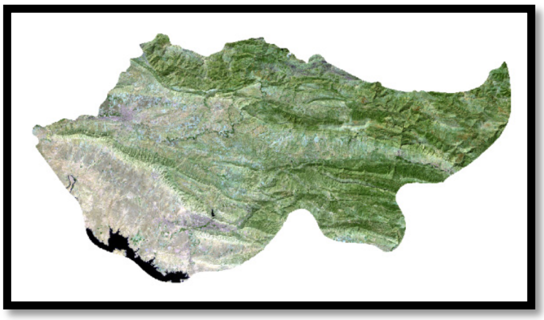
(b)

Figure 3. Clipping satellite image in:(a) 2013, (b) 2022

3.2. Composite

It creates a new raster dataset by combining several existing ones or during composition; this tool was used to integrate numerous bands into a single raster; in this classification, bands 7, 6, and 4 were used.

(a)



(b)

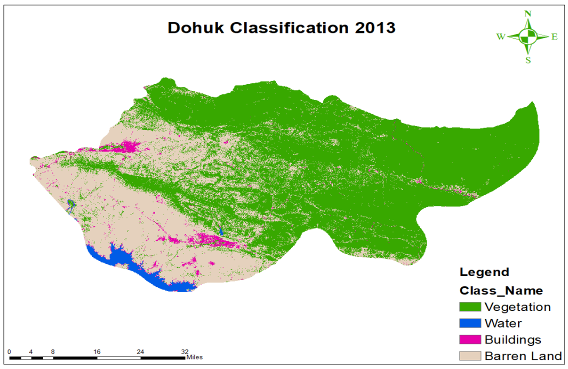
Figure 4. Composition bands of images in (a) 2013, (b) 2022

A collection of feature templates with special construction tools comprised a composite template.

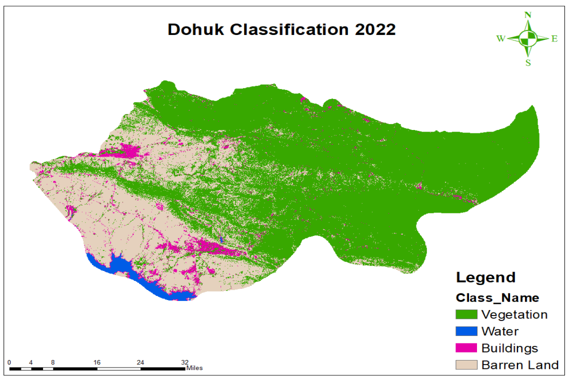
A composite template tool was used to create a new composite template. The composite template's formation on the Create tab was determined by its name and display symbol [14]. To identify the different types of land in the new Raster during this study, a mosaic of bands 4, 6, and 7 was created from each satellite image, whether obtained in 2013 or 2022, Fig. 4 (a and b).

3.3. Classification

Image classification categorizes each pixel in a remotely sensed image to a land cover and land use system. The general correctness of the categorized image was evaluated by comparing the classifications made for each pixel to the precise land cover conditions discovered from the associated ground truth data. Errors of omission were measured by the producer's accuracy, which estimates how well real-world land cover categories can be identified. Comparing the classifications for each pixel to the precise land cover conditions identified from the associated ground truth data allowed for assessing the overall accuracy of the categorized image. The producer's accuracy, which serves as a standard for recognizing real-world land cover categories, was used to measure errors of omission [15], [16], [17]. The super-classification was conducted on two satellite images of the same area on the 1st of August, 2013, and the 2nd of August, 2022, Fig. 5 (a and b).



(a)



(b)

Figure 5. (a) Image classification in (a) 2013, (b) 2022.

As a result, there are four classes of classification, each with their pixel count, multiplied by the number of square centimeters of the satellite image, then divided by a million to get the area of each classification in square kilometers, and the percentages are the time each class takes in the entirety of study area classification, Table 1.

Table 1. The measurement of Duhok in 2013 and 2022

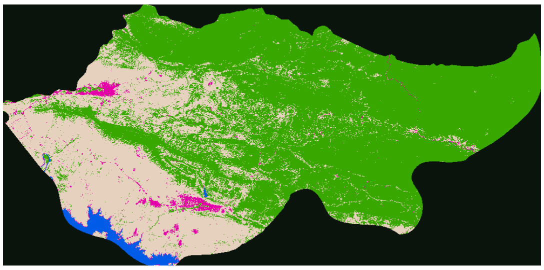
Class Name	Counted 2013	Area 2013 km^2	Percentage% 2013	Counted 2022	Area 2022 km^2	Percentage% 2022
Vegetation	4802016	4,322	63	5022931	4,521	66
Building	107232	97	1	236976	213	3
Water	127441	115	2	109092	98	1
Barren Land	2598438	2,339	34	2266256	2,040	30
Total	7635127	6,872	100	7635255	6,872	100

3.4. Smoothing

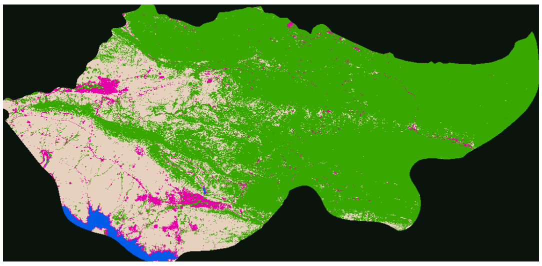
The smoothing technique involves recognizing the fundamental arrangement of information, such as patterns [18], [19]. Numerous types of research investigated the application of smoothing techniques on time-series information to diminish noise. Refining a feature's look via smoothing can be important when modifying features. The user may outline the tolerances for use while smoothing features with the tool. Smoothing lets the user use more than one iteration of the generalized and clean operations. Moreover, obtaining smoothing requires some steps such as:

3.4.1. Majority Filter

There are several restrictions and concerns with digital classification results; for instance, pixel noise can influence the spatial precision and quality of land use land cover (LULC) information. In the post-classification stage, LULC can be produced utilizing a spatial filter to decrease noise and acquire more promising outcomes [20], [21]. The majority filter re-samples the cells in a raster according to the majority value of the neighboring cells. Before re-sampling, two states must stand: first, the number of neighbors with identical values must be considerable enough to be the majority, or at least half the neighbors must have the same value (majority or half). Second, minimize the error of the spatial model in raster with the spatial connection, four-direction approach, or eight-direction approach. If the two conditions cannot stand, the re-sampling cannot be processed, and the raster values will remain, as shown in Fig. 6 (a and b).



(a)

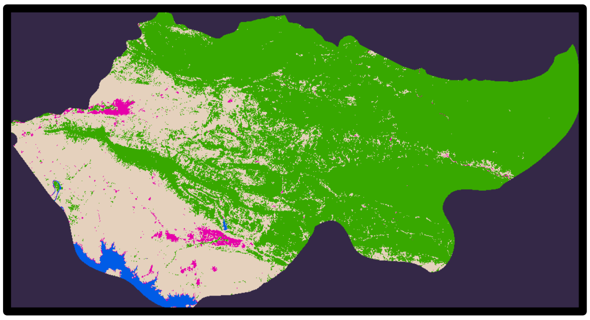


(b)

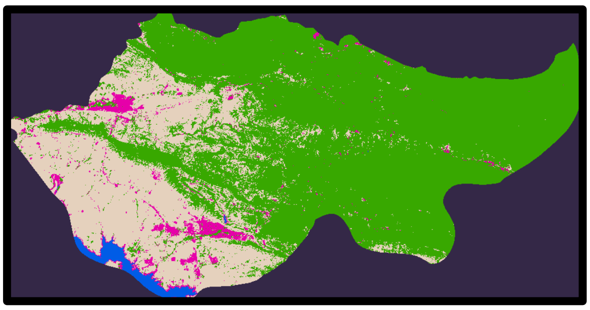
Figure 6. Majority filter classification :(a) 2013, (b) 2022

3.4.2. Boundary Clean

The Boundary Clean tool generalizes or simplifies rasters by smoothing the boundaries between zones. It applies an expand-and-shrink approach to evaluate how each cell operates with its immediate neighbors. The boundary clean and majority filter tools generalize along the edges of zones in a raster. The edges were smoothed by either expanding and shrinking boundaries between the zones or substituting cells with the majority value within their nearest neighborhood, Fig. 7 (a and b).



(a)

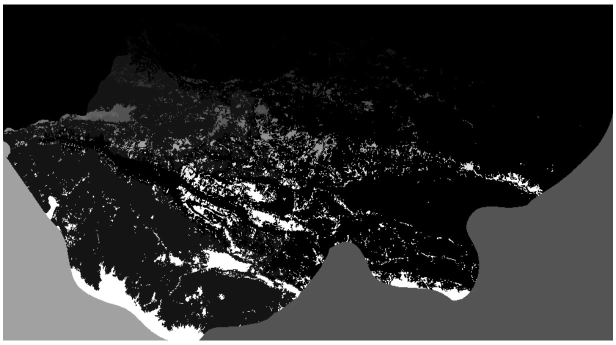


(b)

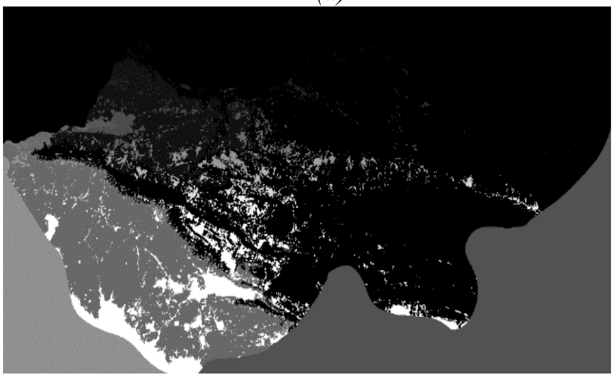
Figure 7. Boundary Cleaned by applying after Majority Filter for (a) 2013, (b) 2022

3.4.3. Region Group

For each cell in the outcome, the identity of the associated region to which that cell belongs was recorded. A remarkable number was assigned to each region. The first region scanned received the value of one, the second one received the value of two, and so on, until all regions were assigned a value. The scan was moved from left to right, top to bottom. The values assigned to the output zones were based on when they were encountered in the scanning operation, Fig. 8 (a and b).



(a)

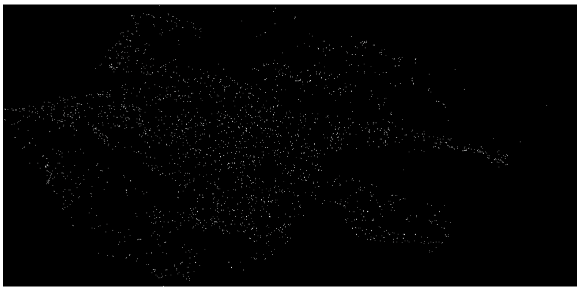


(b)

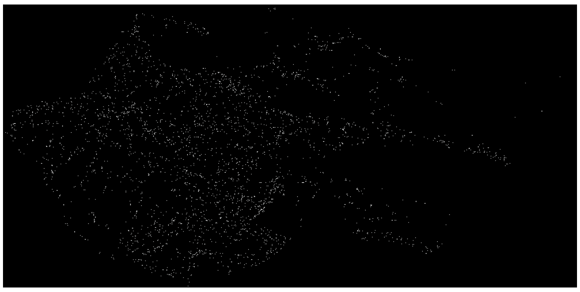
Figure 8. Region groups in (a) 2013, (b) 2022

3.4.4. Set Null

Sets the identified pixels to no data based on the specified measures. It returns no data if a conditional evaluation is true (1) and returns the value specified in the false raster if a conditional evaluation is false (0). This criterion was specified by the logical math function output, which will be the input raster. A raster function within the math logic class must follow the set null function. The result from the logical function is a Boolean raster (1 and 0). Operating the set null function means that all values of 1 will be set to no data, and all values of 0 will be set to false raster, Fig 9. (a and b).



(a)

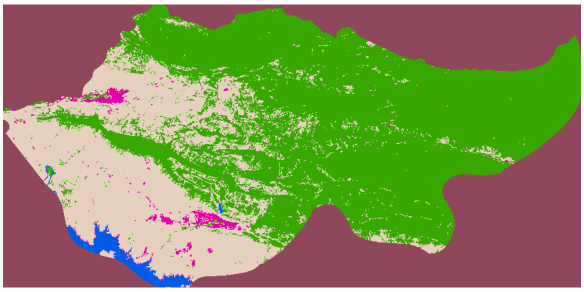


(b)

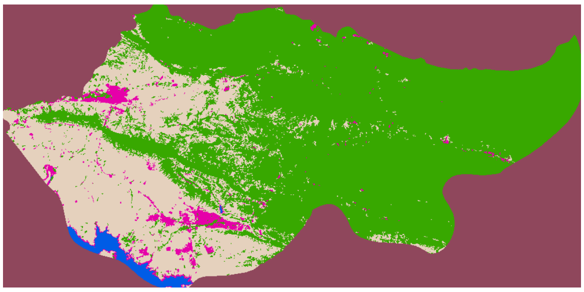
Figure 9. Set Null for Duhok in (a) 2013, (b) 2022

3.4.5. Nibble

The Nibble tool allowed selected raster areas to be assigned the nearest neighbor value. It helped edit raster areas where the data is known to be incorrect. Nibble allowed selected raster areas to be assigned to their nearest neighbor value. It can substitute a few individual cells with the values directly nearby. In larger mask areas, more significant swaths of cells can be replaced, as shown in Fig 10 (a and b).



(a)



(b)

Figure 10. Nibble Smoothing for the classification in (a) 2013, (b) 2022

4. Accuracy Assessment

Accuracy assessments are an essential part of any classification project. It compares the classified image to another accurate source, such as ground truth data that can be collected in the field; nevertheless, it is time-consuming and costly. Ground truth data can be derived from analysing high-resolution imagery, GIS data layers, or existing classified imagers. One hundred twenty ground truth points were collected from random places in Duhok maps to assess and estimate the classification accuracy. Every 30 points were for 1 class, making four classes with 120 ground truthing points. These points would have to be compared to the same points but in the classification to see the endpoint result for the accuracy of the classification. In order to assess the degree of matching between the Raster and Google Earth, a sample of 120 points was divided into 30 points for each land category collected. The location and type of each point were correctly determined on Google Earth and then compared in Fig. 12.

Table 2. The true point of each class

Class	RASTER VALUE	Barren Land	Buildings	Vegetation	Water	Total Points with Inaccuracy	Multiply
Vegetation	1	2	1	30	9	42	1.4
Water	2	0	0	0	17	17	0.57
Buildings	3	0	27	0	1	28	0.93
Barren Land	4	28	2	0	3	33	1.1
		30	30	30	30	120	4
Total Accuracy	102						
Overall Accuracy	0.85						
Total Inaccuracy	18						
Overall Inaccuracy	0.15						

The table shows the four classes applied for Duhok: Vegetation, Water, Buildings, and Barren Land. Since having pin-pointed 30 ground truthing points to be compared, as mentioned earlier, there was an inaccuracy in the table; for example, vegetation has 30 points in the correct vegetation area, which means all 30 points for vegetation are correct. However, there were 17 correct water points: nine in the vegetation region, one in the buildings region, and 3 in the barren land region. As for buildings, there were 27 correct points in the correct region, one in the vegetation region and two in barren land.

This miss-interpolation usually comes because the color of the pixels of the buildings is very close or similar to the color of the barren land or the dust, for example. There were 28 correct points for Barren land and only two in the vegetation region. The right side of the table shows the number of total points considered; for example, 42 points were considered as vegetation, which is inaccurate since only 30 points were manually registered, but 12 were inaccurate; the total of points is 120.

Moreover, the far right shows the multiplication of each class; thirty ground truthing points were registered. Vegetation counts as 42; 12 was added from the inaccuracy of the classes, so the multiplication was 1.4. Water has only 17, which were reduced by 13, so the multiplication of it would be calculated as 0.567. In buildings with 28 correct points, two were reduced to inaccuracy, and multiplication was 0.933. Barren land is 33 with the addition of three inaccurate points; the multiplication of it would be 1.1. Finally, the sum of all points would produce 120 points.

There were only 18 incorrect points. Subtracting 18 from 120 would give 102 correct and accurate total points. To calculate the accuracy, the number of accurate points was divided by the total points required, thus being 102/120, which would equal 85% of the classification's accuracy with the actual satellite image.

5. Accuracy Assessment Using ArcMap Automatic Tools

An enhanced and more robust evaluation method was pursued after conducting a manual accuracy assessment involving the meticulous selection of random data points. This involved leveraging the capabilities of ArcMap tools to attain increased reliability and precision in the outcomes. A second accuracy assessment technique, characterized by a higher degree of automation, was employed in this endeavor. The "create accuracy assessment points" tool within ArcMap facilitated the generation of 120 accuracy assessment points. These points were intelligently distributed, with 30 allocated to each of the four classes under study.

After the acquisition and meticulous comparison of accuracy points against established ground truth data, the analysis transitioned to the "compute confusion matrix" tool. This sophisticated tool allowed for identifying and quantifying class-specific values, culminating in deriving an overarching accuracy measure. The resultant accuracy, calculated utilizing the confusion matrix methodology, revealed a notable level of precision, registering at 86.67%.

This achievement was substantiated through an in-depth assessment, as depicted in Table 3, affirming the robustness and effectiveness of the applied accuracy assessment techniques.

Table 3. Confusion matrix tool results

Class name	vegetation	water	buildings	Barren land	Total	U_Accuracy	Kappa
vegetation	27	0	1	2	30	0.9	0
Water	0	30	0	0	30	1	0
Buildings	3	2	19	6	30	0.63	0
Barren land	2	0	0	28	30	0.93	0
Total	32	32	20	36	120	0	0
P-Accuracy	0.84	0.94	0.95	0.78	0	0.87	0
Kappa	0	0	0	0	0	0	0.82

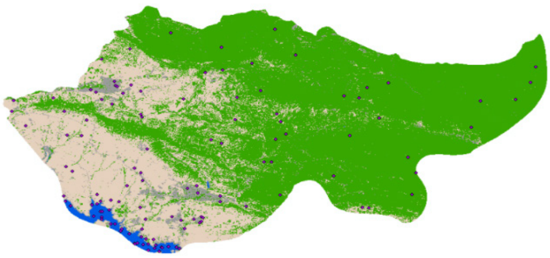


Figure 11. Accuracy assessment points using ArcMap tools

Comparing the accuracy assessment results obtained from both methods revealed a significant improvement in reliability and precision. The computed confusion matrix was employed to quantify the classification performance. The outcomes are presented in Table 3, illustrating the classification results for each class, total points, user and producer accuracy, and the Kappa coefficient. The evaluation demonstrates notable performance in certain classes. Notably, the "water" class achieved a perfect accuracy of 100%, while the "vegetation" class displayed a high accuracy of 90%. In the case of the "buildings" class, the accuracy was relatively lower at 63.33%, and the "barren land" class exhibited an impressive accuracy of 93.33%. When considering the overall accuracy, the classification achieved an accuracy of 86.67%, indicating a robust agreement level between the classified data and the ground truth.

Furthermore, the Kappa coefficient, which quantifies the agreement beyond chance, yielded a value of 0.8222, signifying a substantial level of agreement.

This comprehensive accuracy assessment, combining both manual and automated approaches, underscores the efficacy and reliability of the classification results for delineating the four distinct classes: "vegetation," "water," "buildings," and "barren land." ArcMap's tools facilitated a rigorous evaluation, contributing to the enhanced accuracy and credibility of the study's outcomes.

6. Conclusion

In this study, we embarked on a comprehensive accuracy assessment of supervised classification using remote sensing and GIS techniques. The assessment was crucial in evaluating the reliability and precision of classification results, ensuring their applicability to real-world scenarios. We better understood the classification's effectiveness through manual and automated accuracy assessment methods. Our research centered on the Duhok province in Iraq, utilizing imagery captured by imaginary satellites at different intervals. Employing ArcGIS software, we conducted supervised classifications and applied spatial analysis tools to enhance the accuracy of our categorizations. The study identified four main classes within Duhok province: vegetation cover, buildings, water bodies, and barren lands. Over the study period from 2013 to 2022, changes in land cover proportions were observed, including an increase in vegetation cover and slight fluctuations in other categories. Our accuracy assessment began with manual validation through ground truth data, involving the meticulous selection of 120 points distributed across the four classes. This approach provided valuable insights into the classification's reliability. Subsequently, we harnessed ArcMap's automatic tools, generating another 120 accuracy assessment points. The "compute confusion matrix" tool allowed us to quantify the accuracy and reliability of each class, leading to an overall accuracy of 86.67%. The Kappa coefficient further validated our results, indicating a substantial level of agreement beyond chance.

In summary, our study exemplifies the significance of accuracy assessment in supervised classification. The integration of manual and automated approaches not only validated the classification results but also enhanced their credibility. The combination of remote sensing, GIS, and advanced ArcMap tools provided a robust framework for accuracy assessment, contributing to the accurate delineation of land cover classes in Duhok province. This research underscores the vital role of accuracy assessment in ensuring the validity and applicability of classification outcomes in real-world scenarios.

References:

[1]. Ouchra, H. A. F. S. A., Belangour, A., & Erraissi, A. L. L. A. E. (2023). Machine learning algorithms for satellite image classification using Google Earth Engine and Landsat satellite data: Morocco case study. *IEEE Access*.

[2]. Ippolito, M., Ferguson, J., & Jenson, F. (2021). Improving facies prediction by combining supervised and unsupervised learning methods. *Journal of Petroleum Science and Engineering*, *200*, 108300.

[3]. Van Thinh, T., Duong, P. C., Nasahara, K. N., & Tadono, T. (2019). How does land use/land cover map's accuracy depend on number of classification classes?. *Sola*, *15*, 28-31. Doi: 10.2151/sola.2019-006.

[4]. Stehman, S. V., Fonte, C. C., Foody, G. M., & See, L. (2018). Using volunteered geographic information (VGI) in design-based statistical inference for area estimation and accuracy assessment of land cover. *Remote sensing of environment*, *212*, 47-59.

[5]. Macarringue, L. S., Bolfe, É. L., & Pereira, P. R. M. (2022). Developments in land use and land cover classification techniques in remote sensing: A review. *Journal of Geographic Information System*, *14*(1), 1-28. Doi: 10.4236/jgis.2022.141001

[6]. Apeh, O. I., Uzodinma, V. N., Ebinne, E. S., Moka, E. C., & Onah, E. U. (2019). Accuracy Assessment of Alos W3d30, Aster Gdem and Srtm30 Dem: A Case Study of Nigeria, West Africa. *Journal of Geographic Information System*, *11*(2), 111-123.

[7]. Feizizadeh, B., Darabi, S., Blaschke, T., & Lakes, T. (2022). QADI as a new method and alternative to kappa for accuracy assessment of remote sensing-based image classification. *Sensors*, *22*(12), 4506.

[8]. Abbas, Z., & Jaber, H. S. (2020). Accuracy assessment of supervised classification methods for extraction land use maps using remote sensing and GIS techniques. In *IOP Conference Series: Materials Science and Engineering*, *745*(1), 012166. Iop Publishing.

[9]. Al-Aarajy, K. H. A., & Hassan, Z. H. M. (2023). Land Surface Temperature investigation of Babylon city between (2002-2022) using Remote Sensing and GIS Technique. *Iraqi Journal of Science*.

[10]. Khalid, A., & Ahmed, Z. (2023). Spectrum Analyzing X-ray Data Image (FITS) Using Ds9 Program. *Journal of Survey in Fisheries Sciences*, *10*, 6512-6521.

[11]. Abdullah, S. A., Abbas, K. H., Harif, A. H., Alaa, R. F., & Hiba, S. (2023). Sustainable Urban Distribution of Educational Institutions and Population Density in Baghdad City Using Remote Sensing Techniques. In *IOP Conference Series: Earth and Environmental Science*, *1202*(1), 012015. IOP Publishing.

[12]. Mzuri, R. T., Omar, A. A., & Mustafa, Y. T. (2022). Spatiotemporal Analysis of Land Surface Temperature and Vegetation Changes in Duhok District, Kurdistan Region, Iraq. *The Iraqi Geological Journal, 55,* 68-81. Doi: 10.46717/igj.55.2C.6ms-2022-08-19.

[13]. Rahimi, I., Ahmad, I. H., & Azeez, S. N. (2020). Urban Growth Mapping Using Remote Sensing and GIS Techniques, case study: Darbandikhan, Kurdistan Region, Iraq. *Tikrit Journal for Agricultural Sciences 19*(4), 50-59.

[14]. Holben, B. N. (1986). Characteristics of maximum-value composite images from temporal AVHRR data. *International journal of remote sensing*, *7*(11), 1417-1434.

[15]. Saralioglu, E., & Gungor, O. (2019). Use of crowdsourcing in evaluating post-classification accuracy. *European Journal of Remote Sensing*, *52*, 137-147.

[16]. Khanal, N., Matin, M. A., Uddin, K., Poortinga, A., Chishtie, F., Tenneson, K., & Saah, D. (2020). A comparison of three temporal smoothing algorithms to improve land cover classification: a case study from NEPAL. *Remote Sensing*, *12*(18), 2888.

[17]. Jensen, J. R. (2005). *Digital image processing: a remote sensing perspective.* Upper Saddle River, NJ: sPrentice Hall.

[18]. Wand, M.P. and Jones, M.C. (1995) *Kernel Smoothing*. Chapman & Hall, London.

[19]. Lam, M. W. (2018). One-match-ahead forecasting in two-team sports with stacked Bayesian regressions. *Journal of Artificial Intelligence and Soft Computing Research*, *8*(3), 159-171.

[20]. Mishra, V. N., Prasad, R., Rai, P. K., Vishwakarma, A. K., & Arora, A. (2019). Performance evaluation of textural features in improving land use/land cover classification accuracy of heterogeneous landscape using multi-sensor remote sensing data. *Earth Science Informatics*, *12*, 71-86.

[21]. Shrestha, M., Mitra, C., Rahman, M., & Marzen, L. (2022). Mapping and predicting land cover changes of small and medium size cities in Alabama using machine learning techniques. *Remote Sensing*, *15*(1), 106.