

Evaluate the Accuracy of Supervisor Classification for Al-Shatrah Image, Using Random Points by Remote Sensing and GIS

Ehsan S. Jassim and Ban Abd Abbas

Abstract--- Remote sensing and GIS techniques are one of the very important tools for producing land-use maps and land cover through a process called image classification. This study examines the evaluation of the accuracy of the supervisor classification of the land cover for land use using Google Earth case of Shatrah city in Iraq for the year 2017, where a Landsat-8 OLI_TIRS image was used and analyzed using Arc GIS 10.6. After classifying the landcover/ landuse types, 100 Random Points were created in Arc GIS and convert random points to KML to open in Google Earth. The value of each random point in Google Earth has been validated to assess accuracy. This research includes two parts (1) Landuse / Landcover (LULC) classification and (2) accuracy evaluation. the supervised classification was performed. The major classified LULC were uncovered agricultural (66.6%), water (1.6%), urban areas (8.5%), Dense agricultural (9.9%), and barren lands (13.4%). The results indicate that the overall accuracy of the rating was 78% and the Kappa Coefficient (K) of 0.73. The kappa coefficient is classified as acceptable and therefore the categorized image was found to be suitable for further research. This study provides an essential source of information where planners and decision makers can use it for sustainable environmental planning.

Keywords--- Landuse Landcover, Remote Sensing, GIS, Classification.

I. INTRODUCTION

As one of the main driving forces of global environmental change, land use and land cover change are a key factor in sustainable development. Land-use changes and land cover affect a wide range of environmental and landscape features including water quality, land resources, atmosphere, ecosystem processes, and the climate system itself through greenhouse gas flows and the effects of surface white. For decades, remote sensing using high-precision multi-angle satellite data has been widely used to obtain information about land cover such as the level of forest and wetland degradation, the rate of urbanization, the intensity of agricultural activities, and other human-induced changes [1]

Remote sensing and GIS techniques are a major means of surveying and controlling land resources and learning about their distribution and characteristics, including preparing development plans and programs. Because it is an important source of data as it provides information efficiently and effectively, traditional methods cannot provide and interpret it. [2]

Ehsan S. Jassim, Astronomy and Space Department, College of Science, University of Baghdad.
Ban Abd Abbas, Astronomy and Space Department, College of Science, University of Baghdad.

GIS provides a flexible environment and a powerful tool in spatial information analysis. The integration of remote sensing (RS) and (GIS) is an important tool for studying the changes in patterns and dynamics of the land cover in order to obtain fast, economical, reliable and accurate results and what we are witnessing today in the information age The huge amount of data needs efficient devices to deal With it, accurate and based on scientific analysis leading to optimum and sustainable land resource use. [3]

In classification process, the supervised classification using the Maximum Probability Method, which is also used in this study, has been used extensively in remote sensing applications. The importance of research appears in assessing the efficacy of remote sensing techniques and GIS in the classification of space images when the user is identified Or researcher for classification criteria[4]

Image classification is an important part of remote sensing and image analysis. In some cases, the classification itself may be the subject of the analysis. As this research, the land-use classification from remote-sensed data results in a map like image as the final product of the analysis. [5]

Step-by-step survey, inventory and classification of the land-use planning, in its assessment and comparison of alternatives and in selecting the optimum and sustainable land use with a view to achieving development. Means Human activity associated with a specific piece of land uses the pattern of features (the type of phenomenon) that covers the Earth[5]

The term image classification refers to a computer program that performs a specific image classification procedure.

The analyst should choose the classification method that would best perform a specific task. Currently, it is not possible to determine the best classification for all cases because the characteristics of each image differ according to the circumstances of each study. Therefore, it is imperative that each analyst understand alternative image classification strategies in order to be prepared to choose the classifier best suited to the task in question. [6]

Nowadays, there are different image classification procedures used for different purposes by different researchers.

These technologies are characterized by two main methods, as supervised and non-supervised classifications, in addition, the supervised classification has various sub-classification methods which are called parallelepiped, maximum probability, minimum distances and Fisher's classified methods. These methods are called steel classifier. [6]

Study area

The study area is town al-Shatrah a part of Dhi Qar Governorate in the south of Iraq northeast of Nasiriyah. The study area is bounded by the coordinate (from 45° 50' E to 46° 40' E) longitude and (from 31° 00' N to 31° 40' N) latitude in zone 38N according to UTM projected coordinate system. It covers an area of 384km², it had a population estimated 254,000 In 2009. The selected area as in figure (1)

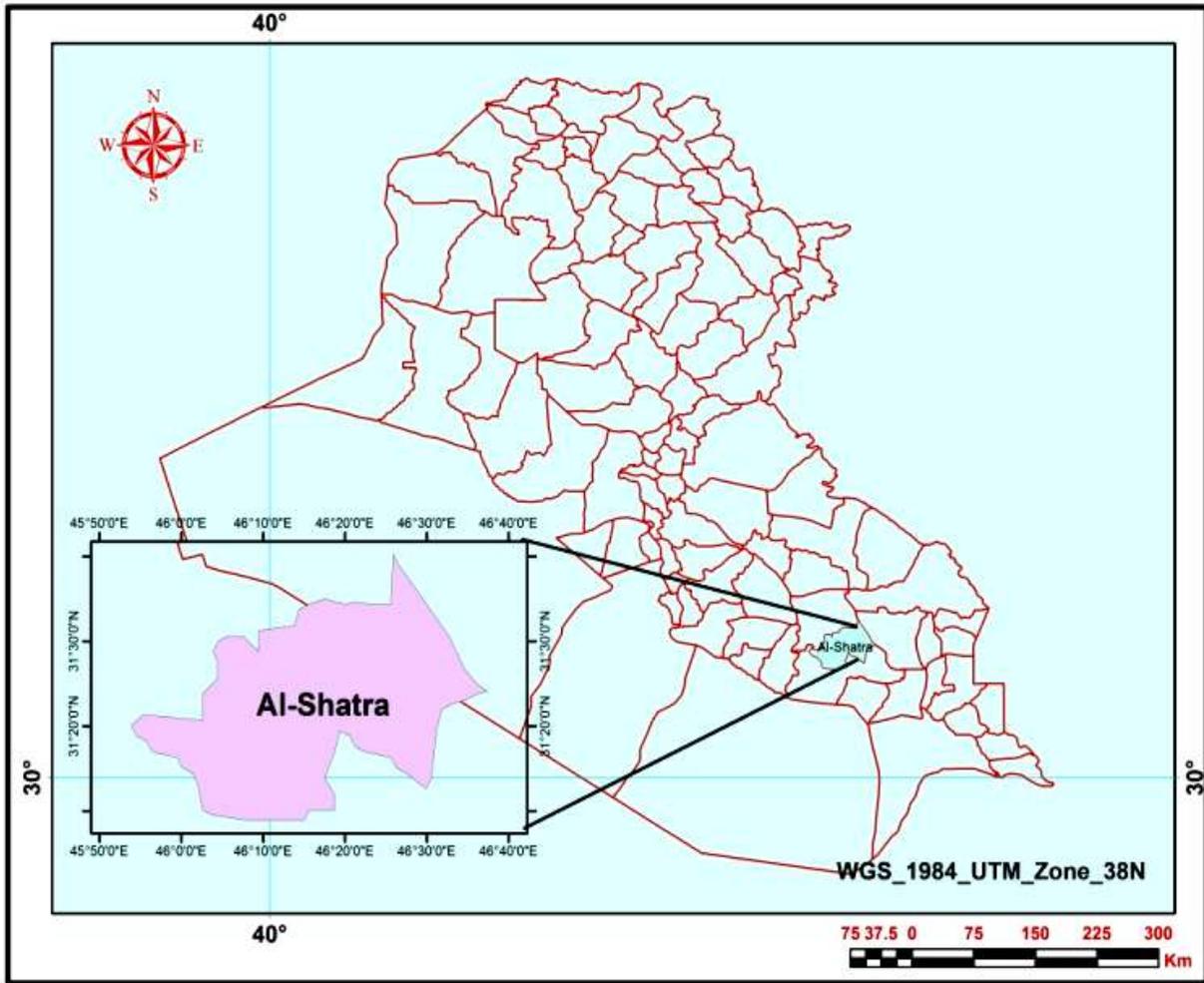


Figure 1: Which illustrates the Iraq map with selection window representing the study area

Study methodology

The image processing was performed in three stages:

1. Image Pre-processing
2. Image Classification
3. Accuracy Assessment.

These applications were performed using ArcGIS 10.6 program

II. MATERIALS AND METHODS

The data used for the search are varied and are therefore represented

1. Landsat 8 image covers the study area
2. Arc Map 10.6 program
3. Google Earth Pro program
4. Shape file for the study area

2. Image classification

In order to take advantage of the many digital data available from satellite imagery, it must be processed in a manner suitable for the end user. For many projects, this treatment includes land classification in different usage functions.

There are two main categories that can be used to achieve this result and are called supervised and uncensored classification techniques. In the supervised land classification, the individual processing the image instructs the image-processing program to help him decide how to classify certain features. This is done by using a vector layer containing training polygons. In a non-supervised classification, the program processes most of it on its own, resulting in more usage categories than the user is interested in. This is the point where the user has to make decisions about which categories can be grouped together in a single land use category. [7]

In either case, additional image processing can be used to help determine the best method for a particular situation. It should not be forgotten that maps are simple attempts to represent what is actually in the world and are never completely accurate[5], in this study the classification of the image, the supervised classification is used, and in this classification the maximum likelihood method is used, and this method is based on the variance and difference in the classification of the impersonal pixel. [6]

2.1 Types of Images classifications

Classification study of visual digital data patterns that appear geographical features depending on the characteristics of spectral reflections and emissions, where data is used in a multi-spectrum classification process, because the pattern is determined by the spectral classification of each cell, this technique is one of the best methods to transform data into visual information. There are two classifications are applied to images. [8]

2.1.1 Unsupervised Classification

This ranking is based on the classification data visible first, and it compiled in groups spectral natural or clusters existing in the scene, on the basis of the convergence of data and digital similarity, (Digital Number (DN)) and known varieties resulting from this method Spectral Classes) then determines the analyst visual identity lid Earth to these groups spectral, and by comparing the data with data classified visual ground reference[8]. As in Figure (2)

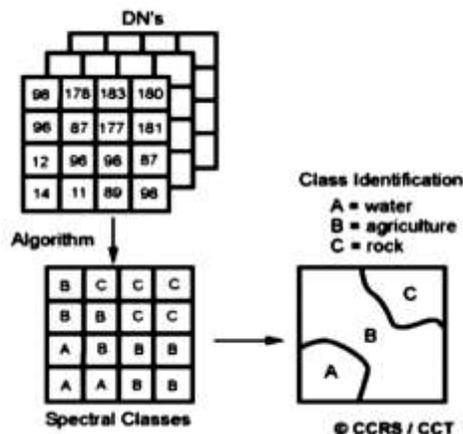


Figure 2: Unsupervised Classification [9]

In Unsupervised Classification When the use of algorithms to assemble the image units with similar spectral characteristics in specific communities (clusters). These gatherings a spectral class did not know the identity of each group of them [9]

2.1.2 Supervised Classification

This method relies on the choice of sampling sites homogeneous within the visual relying on maps and fields of study, be representative of the known pattern of land cover called "Training area so as to put Interpretation key so that the program sensor comparison between data cells and information taken from the field training area and then the process classification based on the similarity metric radio cells, as shown in Figure (3)

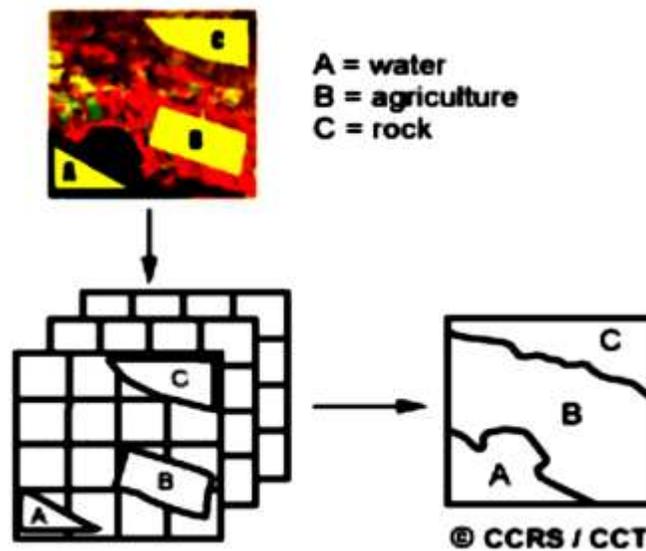


Figure 3: Supervised Classification. [9]

In supervised classification the user relies on own pattern recognition skills and a priori knowledge of the data to help the system determine the statistical criteria (signatures) for data classification. To select reliable samples, the user should apply information (either spatial or spectral) about the pixels that they want to classify. such as a land cover type, may be known through field observations that acquire knowledge about the study area from first-hand observation, analysis of aerial photography. [7]

Field data are considered to be the most accurate (correct) data available about the area of study but should be collected at the same time as the remotely sensed data, so that the data correspond as much as possible. However, all field data may not be completely accurate because of observation errors, instrument inaccuracies, and human shortcomings [6]

GPS receivers are useful tools for performing ground-based studies and collecting training packages. Training samples are groups of pixels that represent what is known as a distinguishable pattern, or potential category. The system will calculate statistics from the sample pixels to create a border signature for the category. [8],figure (4) summarizes three basic steps involved in conducting a supervised model classification.

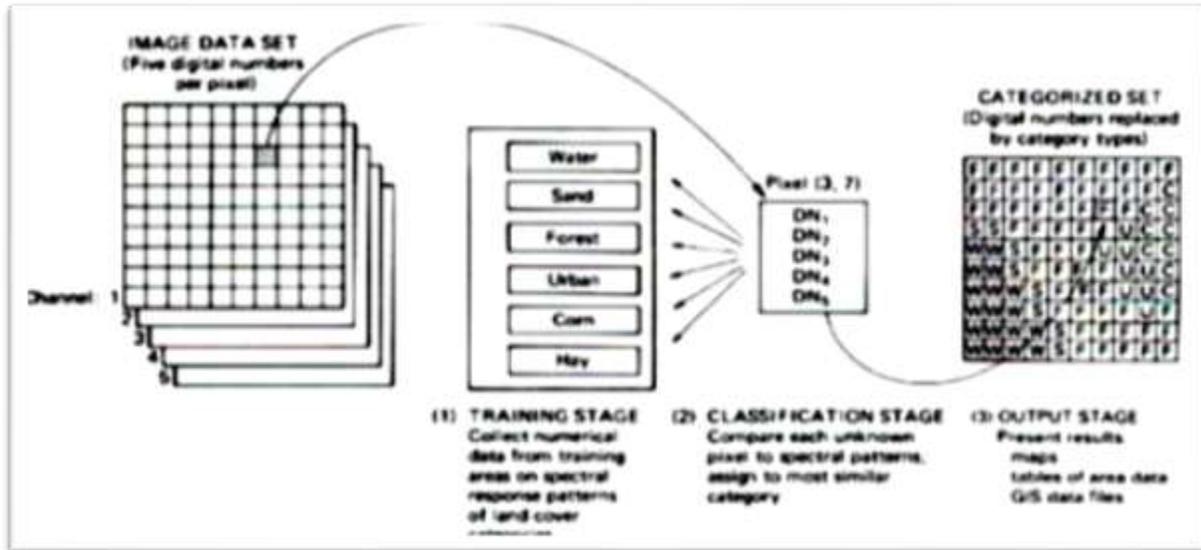


Figure 4: Basic steps in supervised classification [10]

Standard land use and ground cover standards using land classifications according to Anderson et al. (1976) Published by the US Geological Survey (USGS) (for use with satellite images. Symbols are organized into four hierarchical levels. At the top of the pyramid there is the first level and includes:

1. Urban and built lands
2. Agricultural lands
3. Brush or transitional between open and forest
4. Forest
5. Water
6. Wetlands
7. The barren land
8. Tundra
9. Permanent Snow And Ice [11]

2.2 Training Samples

There were two different types of training samples collected for the use of supervised classifications. Both sets were represented by vector layers, first – by polygons, the other – by points. Training polygons were used for Maximum Likelihood Classification. They were created using standard procedure through ArcMap 10.6 Image Classification tool by drawing them over the areas of each of distinguished land cover. In order to locate and identify each land cover class properly. [12]

III. IMAGE ENHANCEMENT

It is the process of improving data in space visibility by increasing the differences between the spectral rows that are expressed in color differences, so that this visual is more visible and visual interpretation.

The removal of distortion precedes the processes of improvement within the visible space, as it is distinguished by the difference in the level of gray color values in them, i.e. the variation in the brightness values, which is the ratio between the highest value of the level of gray to the lowest value within the visible. [5]

The contrast ratio can be considered a reliable focal point in analyzing the visual properties, because a large percentage of this contrast means more ease in the interpretation of the visualization, so space visuals that do not have a ratio of contrast need to be improved and developed to obtain more information, and for that Some technical methods are used.

Contrast (improvement) treatment techniques extend the range of luminosity values in the input visible, for confinement Of the total range due to the luminosity affinity of these values in the narrow cell section. As the brightness density values are pulled to the farthest part of the visual, in other words, they are stretched And expanded to include the largest extent as we will see later[7]

Among the most important techniques used in improved treatment processes:

1. linear Contrast Stretch
2. Non-Linear Contrast Enhancement

In this technique, the histogram is equalized From the original visible histogram rows, this is done by the number of cells forming per row According to the histogram aggregate it works by grouping the adjacent gray values and setting each Among them are within specific groups, as this results in a new spectral level. So it's gray values (Luminosity) in the improved visualization is less than in the original visualization (due to the operation of cells Agglomeration [6]

Upon receiving the satellite data from the sensors in their raw state (meaning that they contain a lot of distortion), the visualization is subject to various computer-assisted processes, techniques and corrections, to visualize it to obtain good results. [13]

IV. DIGITAL IMAGE PROCESSING

4.1 Radiometric Correction

The radiometric correction includes the reorganization of all digital units within the visual, so that the relationship is linear between them and the radiative or reflective values in all digital visible units. [10]

4.2 Geometric Correction

The geometrical correction, when displacing the visible component cells, reveals their true geographical locations, meaning that the cells inside the visible take less or higher coordinates than their real coordinates, and they lose the engineering relations between them.

Distortions are handled in two ways: the first is concerned with systemic errors due to data capture devices and Earth rotation, where the treatment is carried out by applying a mathematical model, Second method is concerned with irregular errors, which are corrected by analyzing the ground control points for the specific cells in the input video, represented for specific places on the surface of the earth with a reference source as a topographic map of the depicted area. [11]

4.3 Radiometric and Geometric Calibration in landsat8

Data from the two sensors, the Operational Land Imager (OLI) and the Thermal Infrared Sensor (TIRS), are calibrated to better than 5 percent uncertainty in terms of Top of Atmosphere (TOA) reflectance or absolute spectral radiance and have an absolute geodetic accuracy better than 65 meters (m) circular error at 90 percent confidence (CE 90)..[17]

The $\rho\lambda'$ is not true TOA Reflectance, because it does not contain a correction for the solar elevation angle. This correction factor is left out of the Level 1 scaling at the users' request; some users are content with the scene-center solar elevation angle in the metadata, while others prefer to calculate their own per-pixel solar elevation angle across the entire scene. Once a solar elevation angle is chosen, the conversion to true TOA Reflectance is as follows in eq. (1):

$$p\gamma = \frac{p'\gamma}{\cos(\theta_{sz})} = \frac{p'\gamma}{\sin(\theta_{se})} \quad (1)$$

$p'\gamma$ = TOA planetary reflectance

θ_{se} = Local sun elevation angle; the scene center sun elevation angle in degrees is provided in the metadata, θ_{sz} = Local solar zenith angle; $\theta_{sz} = 90^\circ - \theta_{se}$. [14]

V. ACCURACY ASSESSMENT

Accuracy assessment or validation is an important step in processing remote sensing data. It determines the value of the resulting data information to the user. Productive use of geographical indications is possible only if the quality of the data is known. [14]

The overall accuracy of the categorized image compares how each pixel is categorized against the specific ground cover conditions obtained from its corresponding fact data. Product accuracy measures errors of omission, and is a measure of the quality of ground cover types in the real world. It measures the accuracy of a user's faults, which are for example, classified pixel data that matches the ground cover type of its corresponding physical location. [15]

The error matrix and the kappa parameter are a standard method for assessing image resolution accuracy. Moreover An error matrix was used in several land classification studies and was a key component of this research. [14]

5.1 Cohen's Kappa

Cohen's original kappa coefficient. Suppose we have two raters, A and B, who have classified independently the same group of N units into one of k categories that were defined in advance. Suppose the data are summarized in the square contingency table $P = p_{ij}$, where p_{ij} denotes the relative frequency (proportion) of units that were classified into category as table (1)

Table 1: Pairwise Classifications of Units Into Three Categories. [18]

Raster B				
Raster A	Category 1	Category 2	Category 3	total
Category 1	P 11	P12	P13	P1+
Category 2	P 21	P22	P23	P2+
Category 3	P 31	P32	P33	P3+
Total	P +1	P+2	P+3	1

The kappa coefficient is a function of two quantities: the observed percent agreement

$$p_o = \sum_{i=1}^k (p_{ij}) \quad (2)$$

Which is the proportion of units on which both raters agree, and the expected percent agreement

$$p_e = \sum_{i=1}^k (p_i + p_{+i}) \quad (3)$$

Which is the value of the observed percent agreement under statistical independence of the classifications. The observed percent agreement is generally considered artificially high. It is often assumed that it overestimates the actual agreement since some agreement may simply occur due to chance. The kappa coefficient is given by

$$k = \frac{p_o - p_e}{1 - p_e} \quad (4)$$

Coefficient (4) corrects for agreement due to chance by subtracting (3) from (2). To ensure that the maximum value of the coefficient is 1, the difference $P_o - P_e$ is divided by its maximum value $1 - P_e$. Thus, Cohen's kappa is defined as a measure of agreement beyond chance compared with the maximum possible beyond chance agreement. The value of kappa usually lies between 0 and 1. It has value 1 if there is perfect agreement between the raters (i.e., $P_o = 1$) and value 0 if the observed percent agreement is equal to the expected percent agreement (i.e., $P_o = P_e$). [15]

VI. RESULTS AND DISCUSSION

Several operations to perform before classifying the image.

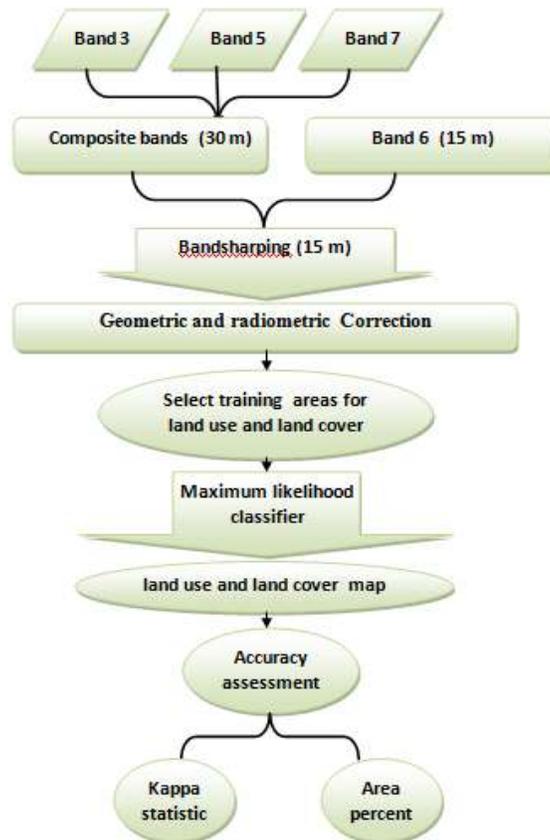


Figure 5: Work flow

These operations include download Landsat8 image covers the study area and the work of composite bands between (B7 + B5 + B3) to obtain a (RGB) image with spatial resolution (30m), and to obtain (15m) spatial resolution make a bandsharpening with band 6, after which the study area was clip from the image (RGB) using the shapefile of study area, After that, the image should be corrected in both geometric and radiometric correction, the figure (5) show flowchart.

The result image as in figure (6)

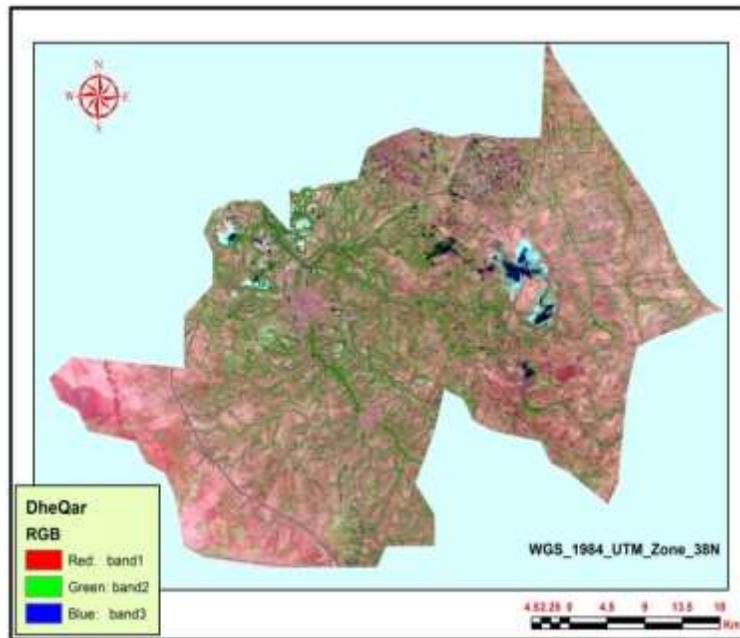


Figure 6: Show study area image after radiometric and geometric correction

The next step is supervisor classification by take training samples as figure (7)

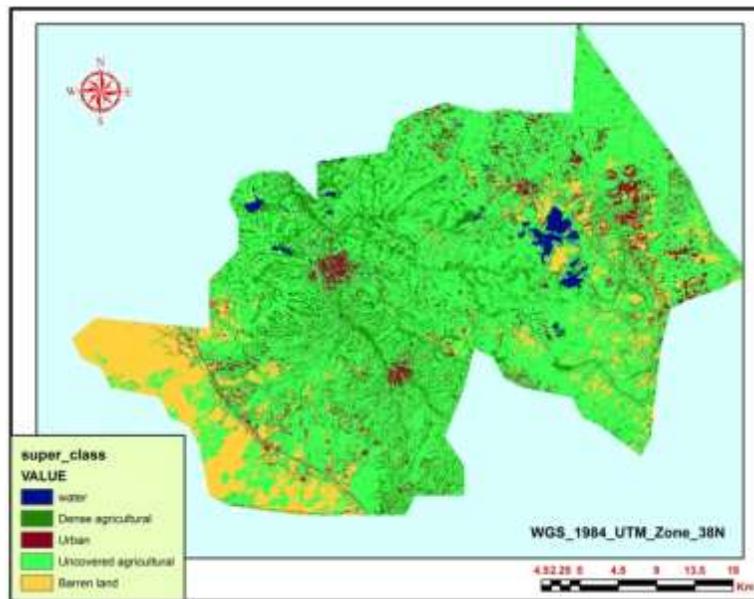


Figure 7: The land use land cover map

The land use land cover classification of the area for 2017 from OLI_IRS satellite image (table 2) showed that the majority of the study area is covered by uncovered land with 1293 (km²), contributes 66.6% of the total area. barren land 260 km² contributes 13.4% and dense agriculture land cover an aerial size of 192 km² (9.9 %) While urban land cover more than 164 km² in 8.5% and at least water covered about 31 km² in 1.6%,

Table 2: LU/LC classes, their corresponding areas for 2017

LULC	COUNT	AREA (km ²)	Area present %
water	137579	30.955275	1.6
dense agre.	853630	192.06675	9.9
urban	731742	164.64195	8.5
un_covered	5747881	1293.2732	66.6
barren	1157598	260.45955	13.4
total	8628430	1941.39675	100

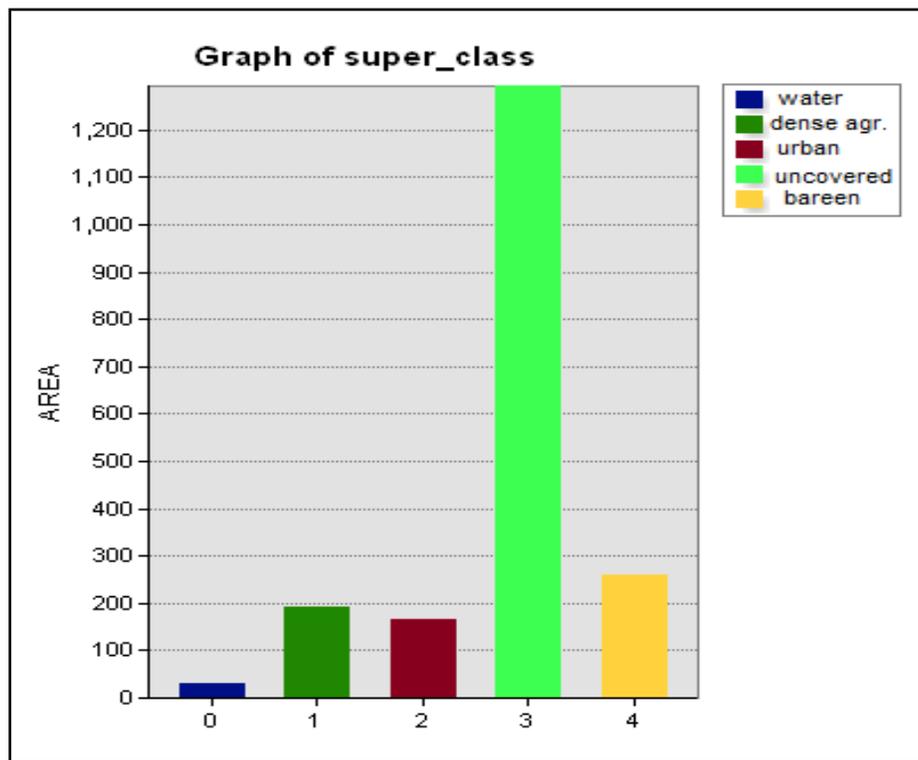


Figure 8: LU/LC classes their corresponding areas for 2017

Accuracy Assessment of Classifications

Google Earth represents a powerful and attractive source of positional data that can be used for investigation and preliminary studies with suitable accuracy and low cost. Since Images from Google Earth with high spatial resolution are free for public and can be used directly in land use land cover mapping in small geographical extend. A study which was conducted by Abineh and Teferie in 2015, and the result of accuracy assessment of land use land cover with the help of Google Earth was more than 77% which is acceptable After image is classified, generating a

set of 100 random points was done in ArcMap (Toolbox >>> Data Management Tools >>> Feature Class >>> Create Random Points >>> create extract values to points. Then the value of each random points were identifying from Google Earth image. As figures (9)

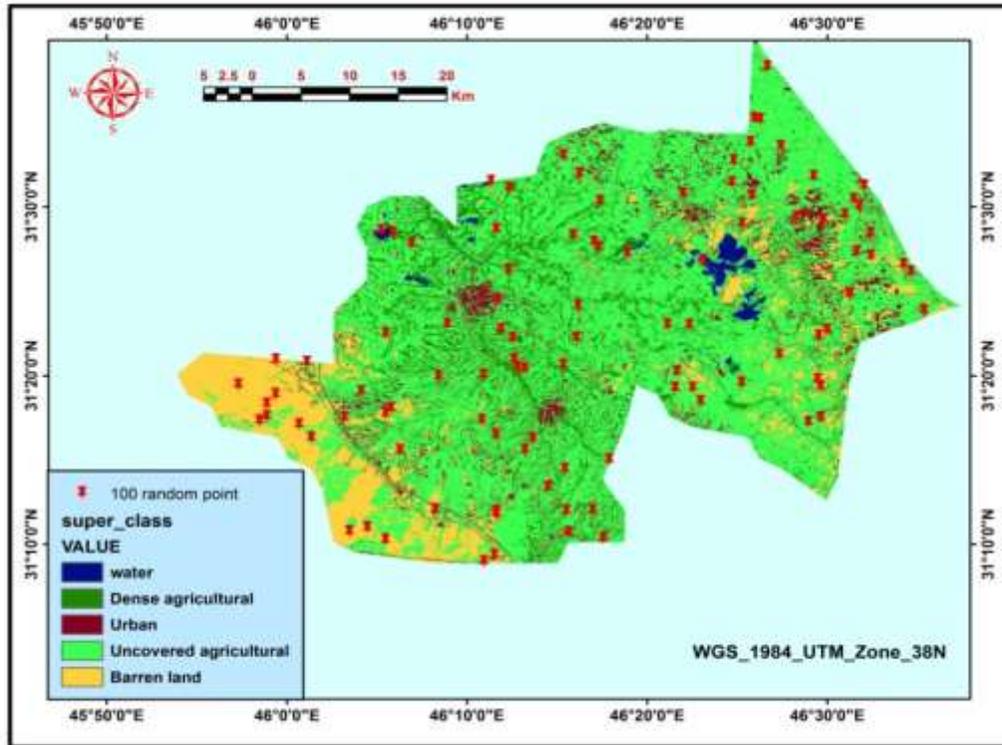


Figure 9: Projected 100 random point

Convert 100 random point from shapefile format to KML format, to open in Google earth program as in figure (10)



Figure 10: 100 random point in KML format in Google earth program

Then by using visual interpretation for all point projection and compared between true point and can get the table (3)

Table 3: Accuracy assessment of land use land cover

S.No	Classified	water	Dense agricult.	uncovered	barren	urban	total	correct samples
1	water	10	1	2	0	0	14	10
2	agriculture	0	23	2	0	2	27	23
3	uncovered	0	2	22	1	2	27	22
4	barren land	0	0	2	17	2	21	16
5	urban	0	1	2	1	7	11	7
	total	10	28	30	19	13	100	78

From table (3) shows the relationship between ground truth data and the corresponding classified data obtained through error matrix report. The overall classification accuracy = No. of correct points/total number of points = (78/100)=0.78=78%, to calculate kappa statistical by using equation ()

$$K = \frac{N \sum_{i=1}^r x_{ii} - \sum_{i=1}^r (x_{i+} \times x_{+i})}{N^2 - \sum_{i=1}^r (x_{i+} \times x_{+i})}$$

$$Kappa = \frac{100(10+23+22+17+7)-[(14 \times 10)+(27 \times 28)+(27 \times 30)+(21 \times 19)+(11 \times 13)]}{(100^2)-[(14 \times 10)+(27 \times 28)+(27 \times 30)+(21 \times 19)+(11 \times 13)]} \times 100\% = 73\%$$

VII. CONCLUSIONS

1. Remote sensing is very important for producing land use / land cover maps that can be done through a method called image classification. This process has revolutionized over the past years
2. Although there are many highly efficient programs in the classification of images, but the user experience and efficiency still have a significant impact on the extent of image classification
3. Landat image grading in order to obtain accurate and reliable information from LULC remains a challenge dependent on several factors such as the quality of the chosen images, the nature and complexity of the landscape, methods and techniques used to process the images and classification methods. The accelerated use of remote sensing data and technologies has made the geospatial process faster and stronger, although the increased complexity also increases the probability of error.
4. The aim of this paper was to extract the land use - land cover map (LULC) of the study area using remote sensing techniques and geographic information systems and also to conduct an assessment of the accuracy of the controlled classification in order to assess the success of the classification.
5. The supervised classification process used to classify the land cover shows five main maps of the land cover in the study area.
6. The supervised classification was performed. The image is categorized into five chapters: Dense agricultural (192 km²) with area percent (9.9%), water (30 km²) with area percent (1.6%), urban areas (164 km²) with area percent (8.5%), uncovered agricultural (1293 km²) with area percent (66.6%), and barren lands (260 km²) with area percent (13.4%). The uncovered farm was the predominant type of Land use classified, which covered about 66.6% of the total study
7. In addition, the classified image must be evaluated to ensure its accuracy, before the same image can be used as input for any applications. The Individual Accuracy Measurement Tool for Accuracy Assessment is

useful for assessing model performance in relation to a specific category / category of particular interest to the study. In this study, a precision evaluation was performed using the error matrix. The study had a comprehensive classification accuracy of 78% and a Kappa coefficient of 0.73. The kappa coefficient is classified as large and therefore the categorized image was found to be suitable for further research.

REFERENCES

- [1] Dr. Abdul Razzak T.Ziboon, 2013, Utilization of Remote Sensing Data and GIS Applications for Determination of the Land Cover Change in Karbala Governorate, *Eng. &tech. journal*, vol.31, part(A), No.15.
- [2] Sophia S. Rwanga, 2017, Accuracy Assessment of Land Use/Land Cover Classification Using Remote Sensing and GIS, *International Journal of Geosciences*, 8, 611-622.
- [3] Tammy Parece, Remote Sensing in an ArcMap Environment, 19. Classification of a Landsat Image (Supervised), <http://www.virginiaview.net/education.html>,
- [4] Leonardo Paolini, 2006, Radiometric correction effects in Landsat multi-date/multi-sensor change detection studies, *International Journal of Remote Sensing* Vol. 27, No. 4, February, 685–704
- [5] Jonas Puzinas, 2017, oLand cover classification using satellite imagery and LiDAR, Masters of Geoinformatics 4th Semester, Aalborg University Copenhagen, A.C. Meyers Vænge 15, 2450 Copenhagen SV.
- [6] FM Mutuku, 2009, 1A supervised land cover classification of a western Kenya lowland endemic for human malaria: associations of land cover with larval Anopheles habitats, *International Journal of Health Geographics*, doi:10.1186/1476-072X-8-19,
- [7] Joan-Cristian Padro, 2018, Radiometric Correction of Landsat-8 and Sentinel-2A Scenes Using Drone Imagery in Synergy with Field Spectroradiometry, *Remote Sens.*, 10, 1687; doi:10.3390/rs10111687.), No.15, 2 *Eng. &Tech. Journal.*, 31, rnal,
- [8] Alexandra De Raadt, 2019, o Kappa Coefficients for Missing Data, *Educational and Psychological Measurement*, Vol. 79(3) 558–576.
- [9] Thomas R Nichols, 2011, Putting the Kappa Statistic to Use, Copyright © John Wiley & Sons, Ltd., 1 *Qual Assur J* 2010; 13, 57–61, DOI: 10.1002/qaj.481
- [10] M.V.K. Sivakumar, 2004, 3Satellite Remote Sensing and GIS Applications in Agricultural Meteorology, World Meteorological Organization, 7bis, Avenue de la Paix, 1211 Geneva 2 Switzerland,
- [11] Sabah S. Al-Jnied *et al* 2013 Assessment of Land Consumption Rate of the Main Island of the Kingdom of Bahrain 1986-2012 using GIS and Remote Sensing, *AGJSR* 31 (4): 276-285
- [12] Jean Carletta, 1996, Assessing agreement on classification tasks: the kappa statistic, *University of Edinburgh, Computational Linguistics* Volume 22, Number 2.
- [13] Merchant, 2009, "Integrating Remote Sensing and Geographic Information Systems". *Papers in Natural Resources*. 216
- [14] Mousa A. Ahmed, 2014, Integration Remote Sensing and GIS Techniques to Evaluate Land Use-Land Cover of Baghdad Region and Nearby Areas, *Iraqi Journal of Science.*, Vol 55, No.1, pp:184-192
- [15] Landsat 8 surface reflectance code (LASRC) product guide, Department of the Interior, U.S. *Geological Survey*, LSDS-1368 Version 2.0
- [16] Landsat 8 (L8) Data Users Handbook, Department of the Interior, U.S. Geological Survey, LSDS-1574 Version 5.0
- [17] Abineh Tilahun, 2015, Accuracy Assessment of Land Use Land Cover Classification using Google Earth, *American Journal of Environmental Protection*, 2015; 4(4): 193-198.
- [18] Nawal K. Ghazali, Accuracy Assessment of Land Use and Land Cover Classification of AL-Najaf Province, Iraq, *journal of kufa – physics*, Vol.9, No.1 (2017) *t (A)*, No.15, 201313 *Eng. &*