Applying Support Vector Machine Algorithm on Multispectral Remotely sensed satellite image for Geospatial Analysis

Fatima Hashim1*, Hayder Dibs2, Hussein Sabah Jaber1

1Surveying Engineering Department, College of Engineering, University of Baghdad, 2Al-Qasim Green University, Water Resources Engineering Faculty, Water Resources Management Engineering Department, Babylon, Iraq.

E-mails: f.saad1412@coeng.uobaghdad.edu.iq; Dr.hayderdibs@wrec.uoqasim.edu.iq ; hussainsabah00@gmail.com

Abstract. In this research support vector machine (SVM) method apply to classify the satellite image and produce land use and land cover (LULC) map. The used data is the multispectral Landsat-8 OLI satellite image with a spatial resolution of (30 x 30)m². However, the Karbala city was the study area. The SVM Applied with the default parameters of Kernel type, gamma in kernel function, penalty parameter and classification probability threshold. The SVM method achieved high accuracy in separating the categories of the study area based on the test samples collected from the study area in the Karbala province, Iraq. The classification training sites were selected based on visual interpretation and Google Earth Program. The image classification carried for six classes of the study area (Urban Area, Vegetation Area, Soil -1, Soil -2, Water Bodies and Roads). The results show a good accuracy of using SVM method based on relying on the capabilities and the precision of each pixel within the categories. The result evaluation was performed using the confusion matrix, the Kappa coefficient and the overall were 0.89 and 90.61% respectively. The SVM method is able to classify the land use and land cover of the study area with good and accurate results.

1. Introduction

One of the most important applications of remote sensing is the classification of land use and land cover, it has been used in many wide applications, the most important of which are urban planning, environmental change studies and classification of natural resources. So, the used classifier must have the ability to classify pixels into categories for the study area [1-4]. When choosing the classification method, consideration must be given to the classification speed, accuracy and practical application, there are several methods of classifications including the artificial neural network (ANN) and the maximum likelihood (ML) [5],[6]. These methods may not be without flaws despite their quality, for example ML requires a large training area and assumes that the data is normally distributed, as well as ANN method, it is exposed to the problems of the local minima line [7]. Many scientists have found methods and techniques to develop classification
methods for LULC mapping, among those methods that have taken great interest is SVM [8]-[10]. The SVM method is characterized by its ability to search for the super plane using a minimal training area and is characterized by fast processing time, it does not require assumption of data type and is able to solve the problem of misclassification because it develops effective decision boundaries [11]. This is done by finding the optimal separation between the superlative layers, depending on the training cases, the support vectors that are present on the edges of the training, excluding other cases [12-15]. Figure 1 below shows the basis for how SVM works.

Figure 1. Basic work of SVM [23]

Figure 1 above shows the simplified method for the basis of SVM work in class separation, where the training data is divided into two categories and as shown in Figure (1) in blue and green colors, where the super-optimal level separates the data set into the two categories with a specific margin width and as shown in the bold slash, and this process is done by computer programming and works SVM, provided that the distance between the classes and the boundary is greater to the nearest training points to obtain good results. The data are considered linearly separable as shown by equations 1,2 and 3 below [16].

\[ W \cdot X_i - b = 1 \]  
\[ W \cdot X_i - b = 0 \]  
\[ W \cdot X_i - b = -1 \]

Where \( W \) = represents a vertical vector on the hyperboloid plane, \( X_i \) = Dimensional space \( i \)  
\( b = \text{bias function} \)
The effectiveness of SVM was proven by Maulik and Debasis, (2017) [17], they conducted research using SVM and ML, and the results showed the superiority of SVM. Deiimai et al. (2014)[18] explained that SVM was able to produce reliable cover by conducting a study of the core of Malaysia, where two types of classification algorithms were used, SVM and ML, to classify the above region by comparing the classification results for both algorithms SVM gave results that outperformed ML with overall accuracy of 91.67 % and a kappa coefficient of 0.86. Also SVM was used by Ali and Jaber (2020) [19] to classify the land cover maps to monitor the changes that occur in the wet areas, the SVM gave good results in classifying the study area, Lake Razazah, Iraq, and the study area was divided into five categories: deep water, shallow water, agricultural lands, saline soils, and arid lands. Abbas and Jaber (2020) [20] conducted a study to compare three classification algorithms, SVM, ML, and Minimum distance (MD) to select the best classifier for a part of Al Hilla, Iraq. A comparison of classification results found that SVM achieved the highest classifier accuracy.

The aim of this research is to classify the satellite image of Landsat-8 satellite in order to estimate the LULC thematic map of the study area using the SVM method and the selected training and testing sites

2. Methodology and Material
In this research we used a multi-spectral image obtained from the Landsat 8 satellite, with a spatial resolution of 30 x 30 meters, obtained on July 13, 2020 freely from the USGS Agency. The image contains 11 bands. Only three bands were used because they are sufficient for the purpose of this research. The study area was Karbala city, it located in Karbala province, Iraq. It is located between of longitude 44°02′- 44 40° E and latitude of 32°37′- 33 31° N. The Karbala city has area about of 5,034 km² and population about of 1.219 million in 2018.

In this research, it is started with (1) the pre-processing steps; layer stacking, image sub-setting, (2) for processing steps, it was applied the geometric and radiometric corrections to prepare the image for further processing and analysis. The geometric correction was done by control points based on the corrected images, (3) the post-processing step, it started by dividing the study area into six categories; urban area, vegetation area, soil-1, soil-2, water bodies, and roads. The study samples were chosen by visual tracking. The SVM algorithm was applied with the default parameters of Kernel type, gamma in kernel function, penalty parameter and classification probability threshold, then (4) the confusion matrix was adopted to perform the accuracy assessment of the classified image, Figure (2) shows the methodology’s flowchart of the study area, and Figure (3) show the location of study area.
3. Results and Discussion
After performing all the steps of the pre-processing and processing stages, the image became ready to conduct further processing stage as shown in Figure 4. In this step the multispectral Landsat-8 satellite image was classified into six classes to estimate the LU/LC thematic map of the study area. The training sites were selected for each single class of the six categories. Then the supervised pixel based classifier was employed to classify the satellite image using the SVM method. Table (1) shows the selected training sites from the Landsat-8 satellite image to involve into image classification for each single class of the six classes. Figure (5) below represents the classified image.
Figure 4. The study area corrected satellite image

Table 1: Collected training sites from the Landsat-8 satellite image

| Class         | Color | Pixels | Polygons     |
|---------------|-------|--------|--------------|
| Urban Area    | Red   | 4.848  | 16/4.848     |
| Vegetation    | Green | 23.275 | 14/23.275    |
| Water bodies  | Blue  | 6.625  | 6/6.625      |
| Soil-1        | Yellow| 19.077 | 10/19.077    |
| Soil-2        | White | 8.650  | 09/8.650     |
| Roads         | Black | 1.571  | By points    |
In the next stage, the testing sites were selected to perform the evaluation of the produced results. The testing sites represent the real pixels and also it were randomly selected in order to evaluate the accuracy of the image classification throughout a comparison between the classified pixels with the corresponding testing sites (real pixels) by adopted the confusion matrix [21],[9].

The results were presented in different terms of producer accuracy, the Kappa coefficient and overall accuracy. All of the producer accuracy, the Kappa coefficient and overall accuracy can be calculated throughout using the equations (4, 5 and 6) below [12]:

![Figure 5. classifier image (MS) by SVM mothed](image)
Producer accuracy = $\frac{C_{a}a}{C^a}$ ……………………………………………………………………………..(4)

where, $C_{a}a$ equal to element at position ath row and ath column, $C^a = \text{column sums}$. However, the overall accuracy represents the total percentage of the pixels correctly classified, it is computed as the equation (2):

$$\text{Overall accuracy} = \frac{\sum_{i=1}^{Q} C_{a}a}{Q} \times 100\%.$$ ……………………………………………………………………(5)

where, the (Q) represents the pixels total number, and the (U) represents the classes total number.

The Kappa coefficient is computed using the equation (3):

$$K = \frac{\sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r} x_{i+} x_{+i}}{N^2 - \sum_{i=1}^{r} x_{i+} x_{+i}}.$$ ……………………………………………………………………..(6)

where $r$, represent no.of row in matrix, $x_{ii} = \text{the measuerment no. in row i and colum i}$, $x_{i+}, x_{+i}=\text{marginal totals for row i and colum i}$, $N=\text{total no. of observation}$ [22].

The outcomes of the confusion matrix present in Table (2) & (3).

| Class          | Urban Area | Vegetation | Water bodies | Soil-1 | Soil-2 | Roads |
|----------------|------------|------------|--------------|--------|--------|-------|
| Urban Area     | 7508       | 232        | 0            | 0      | 0      | 1411  |
| Vegetation     | 154        | 29778      | 666          | 198    | 0      | 122   |
| Water bodies   | 4          | 226        | 7733         | 56     | 0      | 0     |
| Soil-1         | 9          | 0          | 0            | 30543  | 16     | 13    |
| Soil-2         | 0          | 0          | 8            | 0      | 16623  | 0     |
| Roads          | 99         | 22         | 0            | 7      | 0      | 218   |
| Total          | 7774       | 30258      | 8407         | 30804  | 16639  | 1764  |

Table (2) represents the confusion matrix for pixels where the pixels representing samples for each category of the study area are observed as follows: (Urban area = 7774), (Vegetation = 30258), (water bodies = 8407), (Soil-1 =30804), (Soil-2 =16639), (Roads =1764).

| Class        | Ground Truth (percent) |
|--------------|------------------------|
| Urban Area   | 96.58                  |
| Vegetation   | 0.77                   |
| Water bodies | 0.00                   |
| Soil-1       | 0.00                   |
| Soil-2       | 0.00                   |
| Roads        | 0.00                   |
| Total (%)    | 79.99                  |
|              | 09.57                  |
|              | 01.98                  |
|              | 09.38                  |
|              | 09.38                  |

Table (3) represents the confusion matrix for pixels where the pixels representing samples for each category of the study area are observed as follows: (Urban area = 7774), (Vegetation = 30258), (water bodies = 8407), (Soil-1 =30804), (Soil-2 =16639), (Roads =1764).
Table (3) represents the percent confusion matrix, which represents the accuracy of the categories that are classified, for example: Soil-2, represents the highest percentage (99.99 %), and this indicates that it is a feature that gives different spectral properties that cannot be misclassified, while the Road represents the lowest percentage (12.36 %) and that its features are mixed with the features of other classes, which led to a misclassification. The rest of the classes got good accuracy.

Figure (6) & (7) show the result of each producer and user accuracies of each class of the LULC thematic map.

![Figure 6. The calculated producer accuracies of SVM method](image1)

![Figure 7. The outcomes of user accuracies of SVM method](image2)
From Figure (6) and (7) it seems the classes of water bodies and soil-1 have the highest values of producer accuracies, they were 91.89 and 90.9 respectively. The Roads class was indicated the lowest value of the producer accuracy with about of 80.36. However, Figure (7) reveals that the classes of vegetation area and water bodies have the highest values of the user accuracy. On other side, the urban area shows values about 88.05, it represents the lowest value of the user accuracies in this research. After calculating the parameters evaluating the accuracy of the classification of the SVM approach, it was obtained a high overall accuracy about 90.61% with kappa coefficient about of 0.89, it is proved throughout the classification results its ability to classify covers reliably and well and can be used by the subsequent research of the analyzes and planning studies.

4. Conclusions
In this research the SVM method was tested to estimate the LULC map in the city of Kerbala as the study area. It locates in Karbala province, Iraq. The adopted dataset was the multispectral Landsat-8 OLI satellite image with a spatial resolution of (30 x 30)m². The SVM Applied with the default parameters of kernel type, gamma in kernel function, penalty parameter and classification probability threshold. After the radiometric and geometric correction performed the classification training sites and testing sites were selected based on visual interpretation and Google Earth in order to involve both of these sites in the image classification and result verification. The image of the study area was classified into six classes (Urban Area, Vegetation Area, Soil -1, Soil -2, Water Bodies and Roads). The results show a good accuracy of using SVM method based on relying on the capabilities and the precision of each pixel within the categories. The result evaluation was performed using the confusion matrix, the Kappa coefficient and the overall accuracy were 0.89 and 90.61% respectively. The classes of water bodies and soil-1 have the highest values of producer accuracies, they were 91.89 and 90.9 respectively. However, the roads class was indicated the lowest value of the producer accuracy with about of 80.36. In addition, user accuracy reveals that the classes of vegetation area and water bodies have the highest values. On other side, the urban area shows values about 88.05, it represents the lowest value of the user accuracies in this research. The SVM method is able to classify the LULC of the study area with good and accurate results.

Disclosure statement
No potential conflict of interest was reported by the authors.

Reference

[1] Hayder Dibs, Mohammed Oludare Idrees, Vahideh Saeidi and shattri Mansor. (2016). Automatic Keypoints Extraction from UAV Image with Refine and Improved Scale Invariant Features Transform (RI-SIFT). International Journal of Geoinformatics, 12 (3)
[2] Hayder Dibs, Ahmed Al-Janabi, and Chandima Gomes. "Easy To Use Remote Sensing and GIS Analysis for Landslide Risk Assessment." Journal of University of Babylon for Engineering Sciences 26.1 (2018): 42-54.
[3] Borra, S., Dey, N., & Thambi, R. (2019). Satellite Image Analysis: Clustering and Classification. 259
[4] Hayder Dibs, Shattri Mansor, Noordin Ahmad, Biswajeet Pradhan, and Nadhir Al-Ansari. "Automatic Fast and Robust Technique to Refine Extracted SIFT Key Points for Remote Sensing Images." Journal of Civil Engineering and Architecture 14, no. 6 (2020b): 339-350.
[5] Hashim Ali Hasab, Hayder Dibs, Abdulameer Sulaiman Dawood, Wuroid Hasan Hadi, Hussain M. Hussain, and Nadhir Al-Ansari. "Monitoring and assessment of salinity and chemicals in agricultural lands by a remote sensing technique and soil moisture with chemical index models." Geosciences 10, no. 6 (2020a): 207.
[6] Hayder Dibs, Hashim Ali Hasab, Jawad K. Al-Rifaie, and Nadhir Al-Ansari. "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 Pollution 231, no. 9 (2020a): 1-15.

[7] Abburu, S., & Golla, S. B. (2015). Satellite image classification methods and techniques: A review. International journal of computer applications, 119(8).

[8] Mountrakis G, Im J and Ogole C 2011 Support vector machines in remote sensing: A review ISPRS J. of Photogrammetry and Remote Sensing 66 3 247-

[9] Hayder Dibs, Shattri Mansor, Noordin Ahmad, and Nadhir Al-Ansari. "Simulate New Near Equatorial Satellite System by a Novel Multi-Fields and Purposes Remote Sensing Goniometer." Engineering 12, no. 6 (2020c): 325-346.

[10] Aysar Jameel Abdulkadhum Aljanbi, Hayder Dibs, and Bashar H. Alyasery. "Interpolation and Statistical Analysis for Evaluation of Global Earth Gravity Models Based on GPS and Orthometric Heights in the Middle of Iraq." Iraqi Journal of Science (2020): 1823-1830.

[11] Dibs, H., & Al-Hedny, S. (2019). Detection Wetland Dehydration Extent with Multi-Temporal Remotely Sensed Data Using Remote Sensing Analysis and GIS Techniques. International Journal of Civil Engineering and Technology, 10, 143-154.

[12] Bahari, N. I. S., Ahmad, A., & Aboobaider, B. M. (2014, June). Application of support vector machine for classification of multispectral data. In IOP Conference Series: Earth and Environmental Science (Vol. 20, No. 1, p. 012038). IOP Publishing

[13] Bouaziz, M., Eisold, S., & Guermazi, E. (2017). Semiautomatic approach for land cover classification: a remote sensing study for arid climate in southeastern

[14] Marapareddy, R., Aanstoos, J. V., & Younan, N. H. (2017). Accuracy Analysis Comparison of Supervised Classification Methods for Anomaly Detection on Levees Using SAR Imagery. Electronics, 6(4), 83.

[15] Hashim Ali Hasab, Hussain A. Jawad, Hayder Dibs, Hussain Musa Hussain, and Nadhir Al-Ansari. "Evaluation of water quality parameters in marshes zone southern of Iraq based on remote sensing and GIS techniques." Water, Air, & Soil Pollution 231, no. 4 (2020b): 1-11.

[16] TAATI, A., SARMADIAN, F., MOUSAIVI, A., POUR, C. T. H., & SHAHIR, A. H. E. (2015). Land use classification using support vector machine and maximum likelihood algorithms by Landsat 5 TM images. Walailak Journal of Science and Technology (WJST), 12(8), 681-687

[17] Maulik, Ujjwal, and Debasis Chakraborty. "Remote Sensing Image Classification: A survey of support-vector-machine-based advanced techniques." IEEE Geoscience and Remote Sensing Magazine 5, no. 1 (2017): 33-52.

[18] Deilmai, B. R., Ahmad, B. B., & Zabihi, H. (2014, June). Comparison of two classification methods (MLC and SVM) to extract land use and land cover in Johor Malaysia. In IOP conference series: Earth and environmental science (Vol. 20, No. 1, p. 012052). IOP Publishing.

[19] Ali, A. H., & Jaber, H. S. (2020). MONITORING DEGRADATION OF WETLAND AREAS USING SATELLITE IMAGERY AND GEOGRAPHIC INFORMATION SYSTEM TECHNIQUES. The Iraqi Journal of Agricultural Science, 51(5), 1474-1485.

[20] Abbas, Z., & Jaber, H. S. (2020, March). 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.

[21] Ahmad A and Quegan S 2013 Comparative analysis of supervised and unsupervised classification on multispectral data Applied Mathematical Sciences 7 74 3681–3694

[22] Cheruto, M. C., Kauti, M. K., Kisangau, D. P., & Kariuki, P. C. (2016). Assessment of land use and land cover change using GIS and remote sensing techniques: a case study of Makueni County, Kenya.

[23] https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm