Evaluation of the Accuracy of Machine Learning Classifiers and Spectral Indices in Land Cover Classification

Luma Khalid Jasim
Department of Surveying, College of Engineering, University of Baghdad, Baghdad, Iraq
luma.k@coeng.uobaghdad.edu.iq (corresponding author)

Raghad Hadi Hasan
Department of Surveying, College of Engineering, University of Baghdad, Baghdad, Iraq
raghad.h@coeng.uobaghdad.edu.iq

Omar Ali Ibrahim
Department of Surveying, College of Engineering, University of Baghdad, Baghdad, Iraq
omar.a@coeng.uobaghdad.edu.iq

Received: 31 January 2025 | Revised: 19 February 2025 and 10 March 2025 | Accepted: 13 March 2025

 DOI: https://doi.org/10.48084/etasr.10406

ABSTRACT

Population growth and economic and industrial development coupled have significantly accelerated the rate of Land Use and Land Cover (LULC) changes, particularly in developing countries, so finding optimum ways to observe these change has become a pressing issue. Quantification evaluation of these changes is crucial to comprehend and oversee land management conversion, therefore, it is necessary to evaluate the accuracy of various algorithms for LULC classification to determine the most effective classifier for Earth observation applications. The performance of Maximum Likelihood (ML), Support Vector Machines (SVM), Random Forest (RF), and K-Nearest Neighbors (KNN) was examined in this study, based on Sentinel 2A satellite images. The accuracy of those classifiers was evaluated using the Kappa Coefficient and normalized difference index-based verification. The findings indicate that all classifiers exhibit high accuracy levels with variations. The RF algorithm had the highest Kappa coefficient of 0.90, while the KNN algorithm the lowest of 0.76. The accuracy values for RF, SVM, ML, and KNN were 93.1%, 91.2%, 86.2%, and 82.5%, respectively. Results from this study using index-based LULC show that the RF classifier outperforms the others. The results of this study can be used in monitoring LULC change tasks.

Keywords-machine-learning; LULC;kappa coefficient; accuracy; normalized difference indices

I. INTRODUCTION

Understanding Land Use/Land Cover (LULC) alteration is essential in various fields that utilize Earth observations, such as urban and regional planning [1], environmental vulnerability and impact evaluation [2], natural disaster prevention, hazard monitoring, and salinity and soil erosion assessment [3]. To grasp and steer the transformation of landscapes effectively, we can rely on the quantitative evaluation and forecasting of LULC changes [4]. To track the shifts and patterns in LULC, satellite images have proven invaluable alongside traditional ground-based mapping techniques. In recent times, there has been a noticeable uptick in the use of machine learning algorithms to analyze and classify LULC-related images [5]. Different machine learning methods show varying levels of accuracy while the detail and clarity of sensor and image data along with timing, area coverage, equipment, and software used for processing play a significant role in determining LULC accurately [6]. These methods have become increasingly popular in the field of remote sensing especially when it comes to classifying different types of LULC. As a result, there has been a significant increase in research focused on modeling LULC using a wide array of machine-learning algorithms [7, 8]. It has been shown that Artificial Neural Networks (ANNs), Support Vector Machines (SVMs), and Random Forest (RFs) tend to outperform the conventional classifiers in terms of accuracy [9]. In the realm of classifying LULC it is commonly seen that SVM and RF stand out as top choices. Yet one must bear in mind that the precision of such classifications is deeply affected by aspects such as the traits of the sensors used and the quality of image data which encompasses both the spatial and temporal clarity along with the tools and equipment used for processing [10, 11]. When it comes to classifying LULC with satellite observations of medium and low resolution there are challenges related to both spectral and spatial aspects which can impact the overall accuracy [12]. In an effort to overcome these challenges and obtain LULC images with greater precision, experts in the field have turned to the use of machine-learning algorithms [13, 14].

The goal of this research is to evaluate the accuracy of distinct machine-learning algorithms in LULC map extraction. Four machine learning approaches were applied with the aim to identify the optimum technique in generating LULC maps considering the accuracy statistic.

II. STUDY AREA

Karbala Province is located in the central part of Iraq, it is bordered by Anbar to the North and West, Najaf province in the South, and Babylon in the East, about 105 km from the capital Baghdad. The area of Karbala about 4874.74 km² extended between 32°36′00′N - 44°01′00′E. Karbala province was chosen as the study area due to its environmental diversity.

III. MATERIALS AND METHODS

A. Data Source and Field Data Collection

The necessary satellite datasets for this study were obtained from the ESA Hub website (https://scihub.copernicus.eu/) ensuring they were cloud-free and acquired on October 5, 2022. The satellite images consist of three scenes; the bands 2, 3, 4, 8, and 11 were used for analysis, with 10 and 20-m spatial resolution. In examining the data, we chose 160 spots based on what we saw in the field and turned to Google Earth Pro to get a closer look at places we couldn't easily reach. The spots were picked out by image analysis and by on-site visits. In every class, 40 examples were gathered. Within a square that spanned 10 by 10 m for each class, the locations were set and noted down with the help of GPS technology making sure they were not off by more than 2 m side to side. Using a random sampling technique, we were able to gather enough examples for LULC classes [15].

B. Classifiers

SVMs, Maximum Likelihood (ML), RF, and KNN were considered for LULC classification. We chose the classifiers according to whether they are appropriate for land cover classification, the range of algorithmic approaches, implement ability and interpretability, past performance, and availability of resources and tools. Training data and testing samples were used to assess the classification accuracy. Deference land cover classes like water, urban regions, barren land, and vegetation were recognized in training. Based on their spectral responses in the existing image, different indexes have been employed to determine the optimal machine-learning algorithm for mapping LULC.

C. Accuracy Evaluation

Accuracy was considered to assess the classification outcomes and compare the results with indices. Accuracy evaluation is an integral fragment of an image classification procedure. It depends on algorithmic training techniques [16]. Per-pixel land cover classification needs two distinct datasets: one for training the classifiers and another for testing. The amount of sample spots should be at a minimum 10 times the image bands utilized [17] and every sample should include as many pixels as possible. Each sample site is positioned in a center of uniform segments of identified land cover classes to prevent pixels from being mixed. The Kappa coefficient and overall accuracy were considered as performance metrics. The Kappa statistic measures the equivalence between control samples and signature samples [18-19]. It is a multiple-variable statistical method used in accuracy evaluation to assess whether one error matrix significantly differs from another [20]. Kappa analysis presumes a multinomial model of sampling, which is fully satisfied only by simple random sampling. Authors in [21] proposed the kappa coefficient as a best practice. The classification accuracy considering the kappa value is: (0.41–0.60) moderate, (0.61– 0.80) high, and (≥0.80) very high [22]. Indexes were used to estimate and pick the finest machine-learning method for LULC mapping. For this purpose, three indicators derived from satellite data were calculated: Normalized Difference Water Index (NDWI), Normalized Difference Built-up Index (NDBI), and Normalized Difference Vegetation Index (NDVI). NDWI describes water coverage, NDBI built-up or unbuilt areas, and NDVI vegetative or non-vegetative areas. Validation was carried out using ground sampling points [23] of the area and the percentage of pixels extracted were correctly calculated for each indicator to assess accuracy (Table I).

TABLE I. ACCURACY AND NORMALIZED DIFFERENCE INDICES VALUES

Indices	Correct classifications	Incorrect classifications	Sum of ground points	Overall accuracy %	Water area (km²)	Build-up area (km²)	Vegetation area (km²)
NDWI	39	1	40	97.5	288.171		
NDBI	33	7	40	82.5		860.540	
NDVI	37	3	40	92.5			1056.867

IV. RESULTS AND DISUSSION

ArcGIS Pro software was used to distinguish four different land cover classes (water, built-up, barren land, and vegetation) and SVM, ML, RF, and KNN classifiers were utilized. Table II presents the outcomes from the classified image validation method, utilizing the cross-tabulation matrix and kappa coefficient, overall accuracy, producer's accuracy, and user's accuracy. Figure 1 shows the results graphically. Based on the confusion matrix, which was utilized to contrast the classification approaches, the overall accuracy and kappa coefficient indicated that the RF method outperformed SVM, ML, and KNN in terms of overall accuracy, while the KNN was found to be the least accurate.

The areas of NDWI, NDBI, and NDVI were calculated as shown in Figure 2 and were compared with the corresponding areas of water bodies, built-up areas, and vegetation land on LULC maps using the four classifiers as illustrated in Table III.

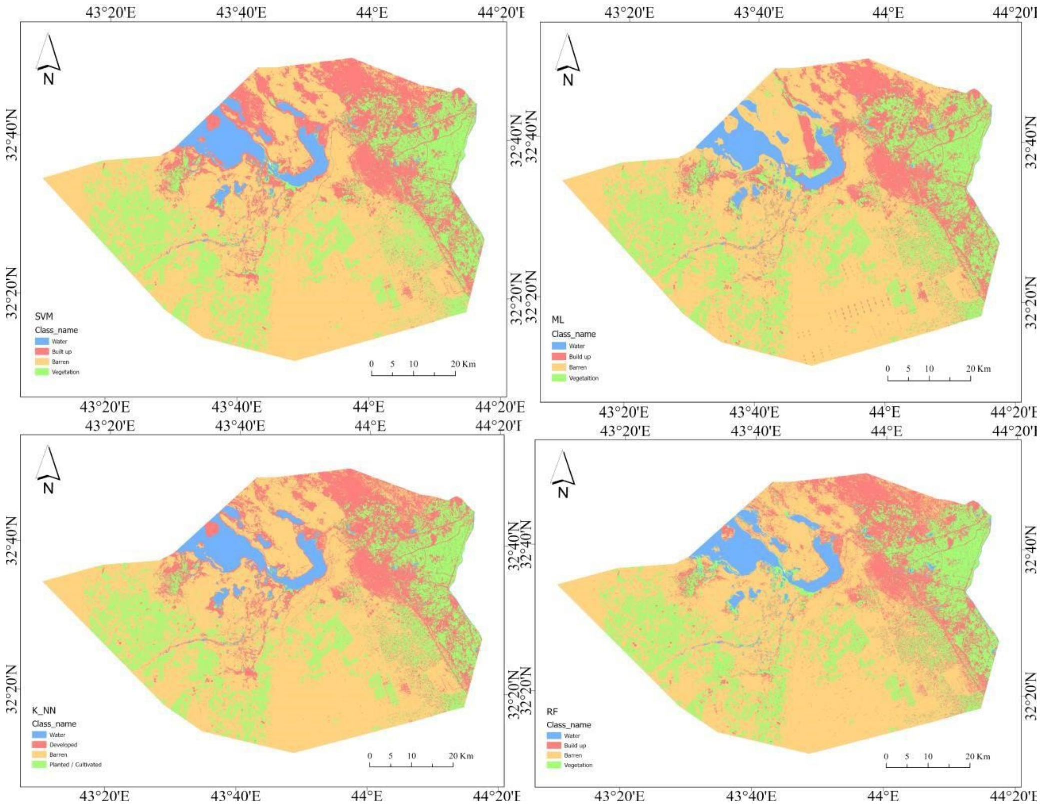

Fig. 1. Land use classification results via SVM, ML, k-NN, and RF.

TABLE II. CONFUSION MATRIX, ACCURACY AND KAPP COEFFICIENT VALUES

	Class Value	**Water**	**Built-up**	**Barren land**	**Vegetation**	**Total**	**U-Accuracy**	**Kappa**
	Water	39	0	1	0	40	0.975	0
	Built-up	0	32	6	2	40	0.8	0
SVM	**Barren land**	1	1	37	1	40	0.925	0
	Vegetation	1	0	1	38	40	0.95	0
	Total	41	33	45	41	160	0	0
	P-Accuracy	0.951	0.9696	0.8222	0.9268	0	0.9125	0
	Kappa	0	0	0	0	0	0	0.8833
	Class Value	**Water**	**Built-up**	**Barren land**	**Vegetation**	**Total**	**U-Accuracy**	**Kappa**
	Water	39	0	1	0	40	0.975	0
	Built-up	0	24	13	3	40	0.6	0
ML	**Barren land**	0	2	37	1	40	0.925	0
	Vegetation	0	0	2	38	40	0.95	0
	Total	39	26	53	42	160	0	0
	P-Accuracy	1	0.923	0.698	0.904	0	0.8625	0
	Kappa	0	0	0	0	0	0	0.8166
	Class Value	**Water**	**Built-up**	**Barren land**	**Vegetation**	**Total**	**U-Accuracy**	**Kappa**
	Water	39	0	0	1	40	0.975	0
	Built-up	0	34	1	5	40	0.85	0
RF	**Barren land**	1	1	38	0	40	0.95	0
	Vegetation	0	0	2	38	40	0.95	0
	Total	40	35	41	44	160	0	0
	P-Accuracy	0.975	0.9714	0.9268	0.8636	0	0.9312	0
	Kappa	0	0	0	0	0	0	0.9083
	Class Value	**Water**	**Built-up**	**Barren land**	**Vegetation**	**Total**	**U-Accuracy**	**Kappa**
	Water	39	0	0	1	40	0.975	0
	Built-up	0	21	12	7	40	0.525	0
KNN	**Barren land**	1	4	35	0	40	0.875	0
	Vegetation	0	0	3	37	40	0.925	0
	Total	40	25	50	45	160	0	0
	P-Accuracy	0.975	0.84	0.7	0.822	0	0.8250	0
	Kappa	0	0	0	0	0	0	0.7666

TABLE III. AREA IN EACH CLASS AND PERCENTAGE BASED ON CLASSIFICATION METHODS

Class-Name	SVM		ML		RF		KNN	
	Area (km²)	%	Area (km²)	%	Area (km²)	%	Area (km²)	%
Water	250.25	5	245.64	5	285.50	6	254.8	6
Built-up	984.46	20	873.84	18	862.75	18	946.605	19
Barren land	2586.09	53	2642.21	54	2662.10	55	2691.823	55
Vegetation	1053.94	22	1113.05	23	1064.39	21	981.512	20
Total	4874.74	100	4874.74	100	4874.74	100	4874.74	100

TABLE IV. AREA OF DIFFERENT CLASSES COMPUTED BY THE SPECTRAL INDICES AND THE MACHINE LEARNING ALGORITHMS

	Algorithms Area (km^2)					Indices Area (km^2)
Class Name	**SVM**	**ML**	**RF**	**K_NN**	**Index**	
Water	250.25	245.635	285.499	254.8	NDWI	288.171
Built-up	984.455	873.838	862.754	946.605	NDBI	860.540
Vegetation	1053.941	1113.052	1064.366	981.512	NDVI	1056.867
Total	2288.646	2232.525	2212.619	2182.917	Total	2205.578

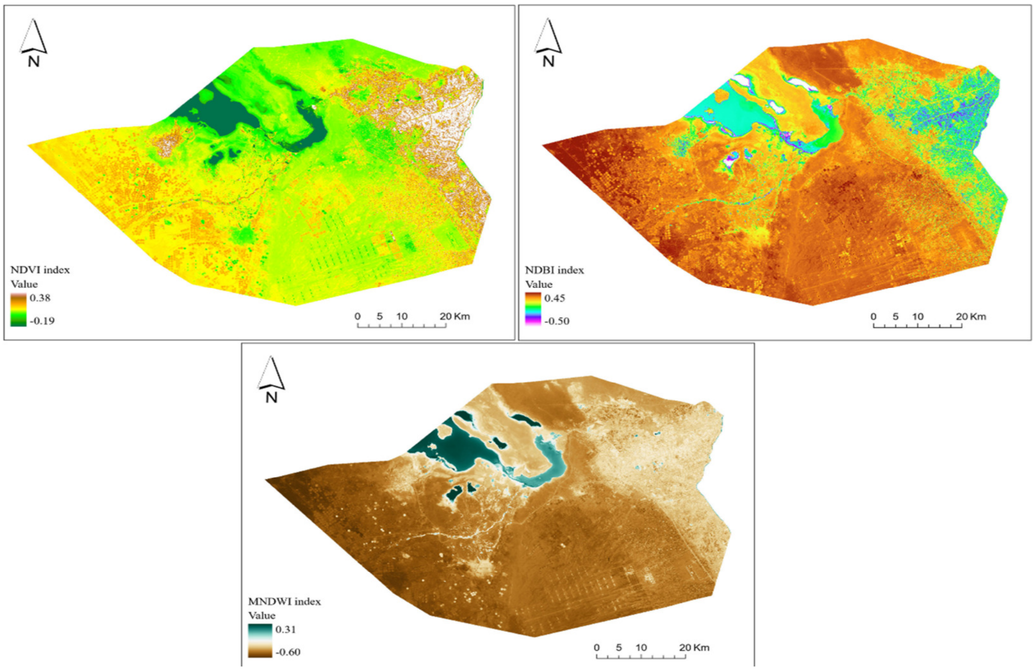

Fig. 2. NDWI, NDBI, and NDVI results in ArcGIS Pro.

Table III shows the land cover area result for each category and its percentage based on the considered classification methods. The results demonstrate that the RF classifier outperforms the others, as the total area of the three spectral indices shows a strong correlation with the areas of the three LULC classes. A similar observation is noted for the NDBI-based built-up area. Therefore, the RF classifier is considered the most suitable for preparing LULC maps in the current study area. Table IV shows the total accuracy and area of the indicator individually, indicating the difference between the results of the classification of algorithms and the accuracy of the spectral indicators and the area of each of them.

V. CONCLUSION

Choosing the most suitable technique is crucial for achieving high classification accuracy. This study aimed to evaluate the effectiveness of four classification algorithms for land cover extraction in Karbala province. By comparing class-specific accuracy, it is evident that the best results for studying the vegetation class are obtained using the SVM and ML methods. Conversely, the SVM method is further appropriate for isolating the water class. RF and SVM are effective for researching the barren class. Lastly, the RF and SVM methods are more appropriate for extracting built-up areas.

This research provides the optimal classification methods for land cover maps that can be useful to support land management decision-making and policy formulation strategies in the Karbala city. In addition, the results can be applied in environmental monitoring tasks due to high accuracy.

REFERENCES

[1] N. Hashem and P. Balakrishnan, "Change analysis of land use/land cover and modelling urban growth in Greater Doha, Qatar," *Annals of*

GIS, vol. 21, no. 3, pp. 233–247, Jul. 2015, https://doi.org/10.1080/19475683.2014.992369.

[2] Y.-A. Liou, A. K. Nguyen, and M.-H. Li, "Assessing spatiotemporal eco-environmental vulnerability by Landsat data," *Ecological Indicators*, vol. 80, pp. 52–65, Sep. 2017, https://doi.org/10.1016/j.ecolind.2017.04.055.

[3] S. Talukdar *et al.*, "Land-Use Land-Cover Classification by Machine Learning Classifiers for Satellite Observations—A Review," *Remote Sensing*, vol. 12, no. 7, Jan. 2020, Art. no. 1135, https://doi.org/10.3390/rs12071135.

[4] J.-F. Mas, R. Lemoine-Rodriguez, R. Gonzalez-Lopez, J. Lopez-Sanchez, A. Pina-Garduno, and E. Herrera-Flores, "Land use/land cover change detection combining automatic processing and visual interpretation," *European Journal of Remote Sensing*, vol. 50, no. 1, pp. 626–635, Jan. 2017, https://doi.org/10.1080/22797254.2017.1387505.

[5] A. E. Maxwell, T. A. Warner, and F. Fang, "Implementation of machine-learning classification in remote sensing: an applied review," *International Journal of Remote Sensing*, vol. 39, no. 9, pp. 2784–2817, May 2018, https://doi.org/10.1080/01431161.2018.1433343.

[6] J. S. Deng, K. Wang, Y. H. Deng, and G. J. Qi, "PCA-based land-use change detection and analysis using multitemporal and multisensor satellite data," *International Journal of Remote Sensing*, vol. 29, no. 16, pp. 4823–4838, Aug. 2008, https://doi.org/10.1080/01431160801950162.

[7] S. Talukdar, P. Singha, Shahfahad, S. Mahato, B. Praveen, and A. Rahman, "Dynamics of ecosystem services (ESs) in response to land use land cover (LU/LC) changes in the lower Gangetic plain of India," *Ecological Indicators*, vol. 112, May 2020, Art. no. 106121, https://doi.org/10.1016/j.ecolind.2020.106121.

[8] A. Jamali, "Evaluation and comparison of eight machine learning models in land use/land cover mapping using Landsat 8 OLI: a case study of the northern region of Iran," *SN Applied Sciences*, vol. 1, no. 11, Oct. 2019, Art. no. 1448, https://doi.org/10.1007/s42452-019-1527-8.

[9] M. Carranza-Garcia, J. Garcia-Gutierrez, and J. C. Riquelme, "A Framework for Evaluating Land Use and Land Cover Classification Using Convolutional Neural Networks," *Remote Sensing*, vol. 11, no. 3, Jan. 2019, Art. no. 274, https://doi.org/10.3390/rs11030274.

[10] D. L. Civco, "Artificial neural networks for land-cover classification and mapping," *International journal of geographical information systems*, vol. 7, no. 2, pp. 173–186, Mar. 1993, https://doi.org/10.1080/02693799308901949.

[11] G. Mountrakis, J. Im, and C. Ogole, "Support vector machines in remote sensing: A review," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 66, no. 3, pp. 247–259, May 2011, https://doi.org/10.1016/j.isprsjprs.2010.11.001.

[12] S. Pal and S. Talukdar, "Assessing the role of hydrological modifications on land use/land cover dynamics in Punarbhaba river basin of Indo-Bangladesh," *Environment, Development and Sustainability*, vol. 22, no. 1, pp. 363–382, Jan. 2020, https://doi.org/10.1007/s10668-018-0205-0.

[13] D. Dutta, A. Rahman, S. K. Paul, and A. Kundu, "Changing pattern of urban landscape and its effect on land surface temperature in and around Delhi," *Environmental Monitoring and Assessment*, vol. 191, no. 9, Aug. 2019, Art. no. 551, https://doi.org/10.1007/s10661-019-7645-3.

[14] S. Pal and Sk. Ziaul, "Detection of land use and land cover change and land surface temperature in English Bazar urban centre," *The Egyptian Journal of Remote Sensing and Space Sciences*, vol. 20, no. 1, pp. 125–145, Jun. 2017, https://doi.org/10.1016/j.ejrs.2016.11.003.

[15] G. Omer, O. Mutanga, E. M. Abdel-Rahman, and E. Adam, "Exploring the utility of the additional WorldView-2 bands and support vector machines in mapping land use/land cover in a fragmented ecosystem, South Africa," *South African Journal of Geomatics*, vol. 4, no. 4, pp. 414–433, Nov. 2015, https://doi.org/10.4314/sajg.v4i4.5.

[16] R. Khatami, G. Mountrakis, and S. V. Stehman, "A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: General guidelines for practitioners and future research," *Remote Sensing of Environment*, vol. 177, pp. 89–100, May 2016, https://doi.org/10.1016/j.rse.2016.02.028.

[17] M.-H. Park and M. K. Stenstrom, "Classifying environmentally significant urban land uses with satellite imagery," *Journal of Environmental Management*, vol. 86, no. 1, pp. 181–192, Jan. 2008, https://doi.org/10.1016/j.jenvman.2006.12.010.

[18] M. J. Cracknell and A. M. Reading, "Geological mapping using remote sensing data: A comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information," *Computers & Geosciences*, vol. 63, pp. 22–33, Feb. 2014, https://doi.org/10.1016/j.cageo.2013.10.008.

[19] A. J. Viera and J. M. Garrett, "Understanding interobserver agreement: the kappa statistic," *Family medicine*, vol. 37, no. 5, pp. 360–363, May 2005.

[20] Y. M. Bishop, S. E. Fienberg, and P. W. Holland, *Discrete Multivariate Analysis: Theory and Practice*. New York, NY, USA: Springer, 2007.

[21] G. H. Rosenfield and K. Fitzpatrick-Lins, "A coefficient of agreement as a measure of thematic classification accuracy.," *Photogrammetric Engineering and Remote Sensing*, vol. 52, no. 2, pp. 223–227, 1986.

[22] N. Kranjcic, D. Medak, R. Zupan, and M. Rezo, "Support Vector Machine Accuracy Assessment for Extracting Green Urban Areas in Towns," *Remote Sensing*, vol. 11, no. 6, Jan. 2019, Art. no. 655, https://doi.org/10.3390/rs11060655.

[23] A. M. Ali, "Making Different Topographic Maps with the Surfer Software Package," *Engineering, Technology & Applied Science Research*, vol. 14, no. 1, pp. 12556–12560, Feb. 2024, https://doi.org/10.48084/etasr.6525.