



# Improved Vegetation Cover Classification Using Remote Sensing Images and Spectral Indices: Case Study of Mecheria in Southwestern Algeria

Nezha Farhi<sup>1,2</sup>, Li Shuai<sup>2</sup> and Sarah Kreri<sup>1</sup>

<sup>1</sup> Agence Spatiale Algérienne, Centre des Techniques Spatiales, Arzew, Algeria

<sup>2</sup> Beihang University (Beijing University of Aeronautics and Astronautics), Beijing, China

farhinezha@gmail.com

**Abstract.** Environmental issues like deforestation are major challenges in the context of dry regions. To characterize this topic, we propose a new algorithm based on the unsupervised K-Harmonic Means classification algorithm and vegetation indices (VIs). The purpose is to optimize vegetation cover information mapping using Landsat images. The region of Mecheria in South-Western Algeria, classified as a semi-arid to arid area, is selected for experimentations. Moreover, two dates 1987 and 2019 are considered for a better assessment of the results.

The proposed methodology integrates multiple vegetation indices regarding their ability to extract vegetation covers in dry climate conditions. The classes presenting the highest correlation ratio are then combined in a quick yet ingenious way creating the final vegetation area.

The new combination technique, inspired from clustering ensembles algorithms, shows an average improvement in accuracy of 16.55% and 29.15% respectively for 1987 and 2019 classification results. These values were computed using confusion matrices. An additional assessment is conducted comparing the proposed methodology with established combination techniques using multiple criteria.

**Keywords:** K-Harmonic Means, Vegetation Indices, Automatic Classification, Landsat Images, Indices Correlation.

## 1 Introduction

In Algeria, the term ‘steppe’ is adopted to describe the vegetation in arid and Saharan areas. According to studies by several authors, an important steppe zone is the region of Naâma in South-Western Algeria known to contain a natural species richness that has yet to be valued [1][2][3].

Mecheria is one of the principal cities situated in Naâma that is constantly subject to extensive human pressures [4] and multiple climate issues [5]. The pasture system of Mecheria, often characterized by bad practices, associated with the region semi-arid

© The Author(s) 2024

C. A. Kerrache et al. (eds.), *Proceedings of the International Conference on Emerging Intelligent Systems for Sustainable Development (ICEIS 2024)*, Advances in Intelligent Systems Research 184,

[https://doi.org/10.2991/978-94-6463-496-9\\_7](https://doi.org/10.2991/978-94-6463-496-9_7)

characteristics, are currently pushing the regression of vegetation covers and land degradation [6]. The delimitation of vegetation areas, in particular, is especially useful as it can provide important information regarding several natural disasters such as deforestation, floods, and climatic phenomena [4] [6] [7].

There is a wide availability of methods and techniques for vegetation area extraction in steppe region. Among them, remote sensing based classification approaches are the most popular as they can decrease human error while efficiently reducing costs, and time [8] [9] [10]. However, developing suitable methods with satisfactory performances using limited resources remains one of the biggest challenges. This especially true in case of developing country where ground truth data is mostly unavailable [11] [13].

In fact, most recent state of the art vegetation mapping and monitoring techniques involve advanced classifiers such as random forest [14] [15] and support vector machine [16] [17]. This type of approaches are known as rule-based classifications, and despite their satisfactory performances they are highly time and resources consuming [18]. As a result, the use of completely unsupervised efficient techniques is essential in order to exploit unlabeled open source datasets.

In this context, the K-Means (KM) algorithm is by far the most exploited technique [19] [20]. However, it presents the issue of randomly selecting K initial centers which may lead to unstable results and sometimes large errors.

In [21], a density peak k-means (DPK-means) algorithm is used to classify vegetation and non-vegetation pixels from a normalized difference vegetation index greyscale image with the fitting accuracies of 0.87 and 0.93 respectively. To solve the local optimization problem of the k-means algorithm, density peaks are used as the initial classification center points. Despite good performances, the methodology relies on additional pretreatments where, the data is fitted by a multimodal Gaussian fitting.

In [22], the authors relied on visual interpretation to perform a series of reclassifications in order to overcome KM initialization issue. The purpose of the study is to extract forest vegetation on the basis of their spectral characteristics and estimate the level of forest fire danger.

The K Harmonic Means (KHM) is an improved version of the KM that uses the harmonic averages of Euclidean Distances to overcome the K-Means high dependence on its initial identification of elements, thus our choice. The algorithm is quick and simple to implement and have the advantage of being a robust classifier [23] [24] [25].

Recently, classification algorithms are computed within a Data Integration technique that combines the selected images bands with additional spectral bands such as different color composition and spectral indices [26]. This way of proceeding can significantly improve the classification accuracy without the need of labelled resources [11] [12] [27].

In [26], the authors integrate three types of features including geometry data in order to reach an accuracy of 86.9%. However, the latter relies on high resolution images in conjunction with the Digital Elevation Model that are not always available in practice.

Another common technique to extract LU/LC data is to perform a Majority Voting combining several spectral indices classification results [28]. However, each individual index outcome contains errors that when combined with other indices errors, in case of Majority Voting, can distort the final output map. Data Integration techniques deal with this issue by performing a classification of the multi-feature data, still, accuracy go significantly down as the image size increases [29].

Our work, on the other hand, performs an additional classification in local classes areas of the indices results averaging the number of misclassified elements and effectively dealing with both problems.

Following the lack of resources issue that greatly hinders researches advancement, multiple companies have created interactive online access to free satellite images. The leading technology USGS Earth Explorer with Landsat archives is freely providing a limitless well of information, accessible for multitude regions all around the world [19] [22]. Our experiments are carried out using two Landsat images covering the city of Mecheria over the years 1987 and 2019. The observed vegetation areas within the region mostly corresponds to agricultural zones and reforestations.

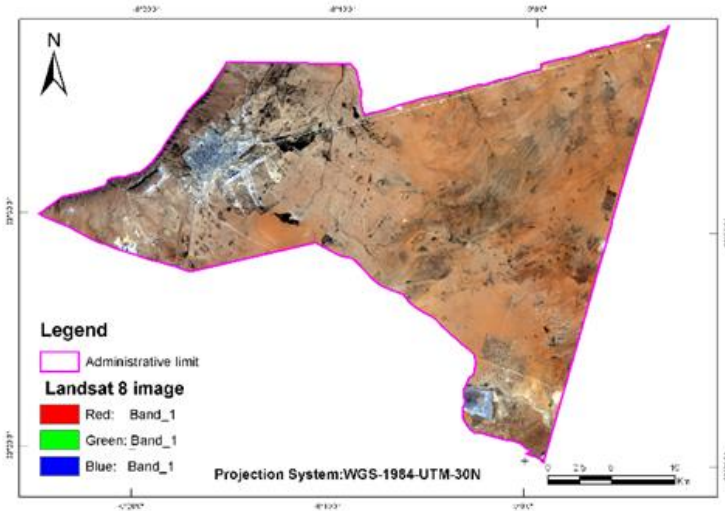
The proposed methodology, aims to explore the ability of multiple VIs combined with remote sensing datasets to improve vegetation cover classification. Considering the natural characteristics of the region of Mecheria, six indices [30] [31] [32] are assessed namely: the Normalized Difference Vegetation Index (NDVI), the Soil-Adjusted Vegetation Index (SAVI), the Modified Soil-Adjusted Vegetation Index (MSAVI), the Optimized Soil-Adjusted Vegetation Index (OSAVI), the Green Soil-Adjusted Vegetation Index (GSAVI) and the Green Optimized Soil-Adjusted Vegetation Index (GOSAVI). The NDVI is by far the most commonly used index in agricultural parcels and vegetation covers detection and monitoring [33]. The SAVI and its multiple versions (MSAVI, OSAVI, GSAVI and GOSAVI), on the other hand, are developed to overcome some of the NDVI limitations especially in less dense areas such as semi-arid regions [34]. Indeed, these indices contain a regulatory parameter, often referred as “L”, which serves the purpose of minimizing the soil effect on the vegetation signal [35]. Therefore, resulting in a more realistic estimation of sparse vegetation covers. In this work, the SAVI is set as the reference index in a Pearson Coefficient [36] based correlation process in conjunction with a surface ratio (equation 3). The proposed merging approach is rooted within unsupervised ensemble clustering techniques. These methods use different procedures to combine several classification outcomes [37]. The final classes retain the useful information from the combined results with an increased precision [37].

Finally, the simplicity of the proposed procedure ensure its effectiveness and flexibility improving the obtained classification results.

## 2 Used Data and Study Area

In the presented work, the region of Mecheria in Algeria with a longitude of  $33^{\circ}.483185$  and a latitude of  $-0^{\circ}.0373201$  was selected as an experimental site.

Our choice is first based on its elements diversity including dense vegetation, sparse vegetation, urban areas, soil and wetlands (5 classes). Secondly, the vegetation landscape of the region suffered visible changes over the years [38]. This allows us to evaluate the performance of the proposed algorithm on a zone characterized by different features and configurations that often get misclassified [12].



**Fig. 1.** Landsat 8 RGB image of Mecheria (2019).

The used images (Table I) are freely available through the U.S. Geological Survey (USGS). Landsat 8 data were acquired using the Operational Land Imager (OLI) and the Thermal Infrared Sensor (TIRS) sensors, while Landsat 5 data were collected using Thematic Mapper sensor.

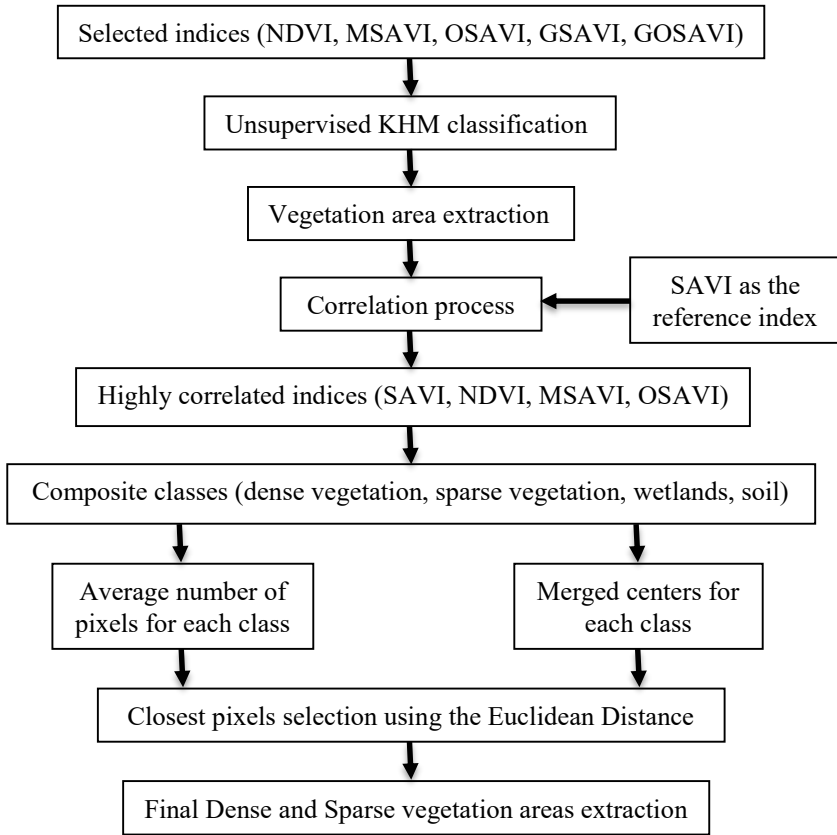
The images are then radiometrically and geometrically corrected and clipped to extract the region of Mecheria.

**Table 1.** Used Landsat satellite images.

Satellite	Date	Size	Resolution
Landsat 8	2019/03/16	7801 x 7941	30 meters
Landsat 5	1987/04/25	8001 x 7648	30 meters

## 3 Methodology

Fig. 2 shows a diagram of the proposed methodology and its different steps.



**Fig. 2.** The proposed methodology diagram.

### 3.1 Selected Indices

Six vegetation indices (Table 2) are selected and assessed based on their classification results. According to studies in [34] [39] [40], the SAVI is especially adapted to extract vegetation under dry climate conditions and is considered our reference index. The ‘L’ parameter of the SAVI is fixed to -0.25 according to experiments in [39] [41].

**Table 2.** Vegetation indices equations and references.

Formula	Ref
$NDVI = \frac{R_{NIR} - R_{Red}}{R_{NIR} + R_{Red}}$	[30]

$$SAVI = \frac{0.75 * (R_{NIR} - R_{Red})}{R_{NIR} + R_{Red} - 0.25} \quad [41]$$

$$MSAVI = \frac{2 * R_{NIR} + 1 - val}{2} \quad [30]$$

$$val = \sqrt{(2 * R_{NIR} + 1)^2 - 8 * (R_{NIR} - R_{Red})}$$

$$OSAVI = \frac{1.16 * (R_{NIR} - R_{Red})}{R_{NIR} + R_{Red} + 0.16} \quad [42]$$

$$GSAVI = \frac{1.5 * (R_{NIR} - R_{Green})}{R_{NIR} + R_{Green} + 0.5} \quad [18]$$

$$GOSAVI = \frac{(R_{NIR} - R_{Green})}{R_{NIR} + R_{Green} + 0.16} \quad [42]$$


---

The selected indices range from [-1 +1], where the closest values to +1 represent dense vegetation and the second closest values represent sparse vegetation covers. In case of Mecheria, like previously cited, these regions mainly corresponds to agricultural and reforestation parcels.

### 3.2 K-Harmonic Means Classification

The selected indices are classified using the KHM unsupervised algorithm [37], which function is given as follows:

$$KHM = \sum_{i=1}^N \frac{K}{\sum_{j=1}^K D_{ij}} \quad (1)$$

Where,  $K$  is the number of clusters empirically fixed by the user and  $D_{ij}$  is the Euclidean Distance between pixels  $x_i$  ( $i=1, \dots, N$ ) and centers  $c_j$  defined by:

$$D_{ij} = \|x_i - c_j\|^2 \quad (2)$$

We initially fixed the number of classes to  $K=5$  corresponding to the visual elements in Mecheria. However, the majority of indices presented a confusion between soil and urban areas that were consequently merged into a single class adjusting the number of clusters to  $K=4$ .

### 3.3 Indices Correlation

Given a set of 10000 randomly extracted pixels from the reference, the correlation process uses a newly proposed measure (CM) combining the Pearson Coefficient (PR) [36] and a surface ratio (SR) to assess the selected indices. The CM between the reference index SAVI and a given index VI is defined as follows:

$$CM = \frac{PR + SR}{2} \quad (3)$$

$$PR = \frac{\sigma_{SAVI,VI}}{\sigma_{SAVI} * \sigma_{VI}} \quad (4)$$

$$SR = \frac{\min(Surface_{SAVI}, Surface_{VI})}{\max(Surface_{SAVI}, Surface_{VI})} \quad (5)$$

Where,

- $\sigma(SAVI, VI)$  is the covariance between indices SAVI and VI
- $\sigma_{SAVI}$  and  $\sigma_{VI}$  are the standard deviations of SAVI and VI respectively.

The CM ranges from [-1 +1], where -1 indicates a negative correlation, +1 indicates a positive correlation and 0 corresponds to nonexistent correlation.

The correlation is computed over vegetation class values, where the rest of the pixels are set to null during this process. We argue that, the selected VIs were initially adapted for vegetation extraction and should be tested for that purpose to maximize their performance.

**Table 3.** Indices correlation values.

	SAVI 1987	SAVI 2019
<b>NDVI</b>	<u>0.9999</u>	<u>0.7120</u>
<b>MSAVI</b>	<u>0.9967</u>	<u>0.6859</u>
<b>OSAVI</b>	<u>0.9999</u>	<u>0.9999</u>
<b>GSAVI</b>	0.5275	0.4847
<b>GOSAVI</b>	0.5277	0.4630

Table 3 shows CM values for all selected indices. Only correlations close to 70% and above are retained [36] corresponding to the MSAVI, OSAVI and NDVI. In case of GSAVI and GOSAVI, the indices presented a large confusion between soil and vegetation classes. However, when using the Pearson Coefficient alone, the metric suggested a high correlation close to +1. This is explained by the SAVI vegetation area being entirely included within the confused vegetation area (with soil region) of the GSAVI and GOSAVI. Integrating the surface ratio within the CM efficiently corrects this issue and returns realistic correlation values.

### 3.4 Composite Classes

In this context, a composite class is formed by combining the surface areas belonging to the same classes across all used indices. The newly obtained combination zone is then extracted from each of the selected indices classifications to form a 4 (4 indices) dimensional feature that we refer to as a composite class. This principle is rooted within clustering ensemble techniques [37] and is illustrated in Fig. 3 for a supposed vegetation class.

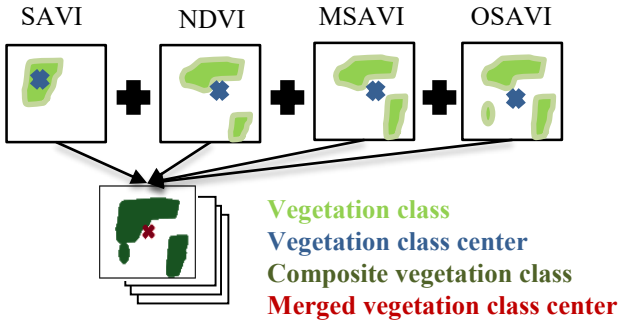


Fig. 3. Vegetation composite class.

There is a total of four composite classes ( $K=4$ ). For each, we compute a merged-center (MC) obtained by averaging the centroids of a given class across all selected indices classifications. The average number of pixels (ANP) of each composition is also extracted for the next step.

### 3.5 Final Classes Extraction

Using the obtained MCs and ANPs from the previous step, we select the final pixels of a specific class by calculating the Euclidean distance over each composite area. The resulting classes consist of the closest pixels to merged-centers with a total corresponding to the ANP of each class.

## 4 Results and Discussions

The evaluation section relies on a careful visual analysis using meticulously selected Regions of Interest to create viable ground truth datasets and compute confusion matrices. The assessment of the proposed methodology is performed in two different schemes: an accuracy assessment and a comparison with the widely used Majority Voting [28] and Data Integration [43]. Majority Voting consists of attributing for each pixel the class that is returned by the majority of VIs classification results. This approach is highly preferred due to its simplicity and satisfactory results [28]. Data Integration, also known as band combination, is another commonly used technique that combines the

VI and spectral bands into a single data. A chosen classification algorithm is then computed on the multiband constructed image. This methodology increases classes separability by introducing higher amounts of spectral information [43]. Our proposed technique, on the other hand, could be viewed as the combination of both approaches with satisfactory results.

#### 4.1 Accuracy Assessment

Accuracy assessment is common practice for rigorous validation based on a robust ground truth dataset. In our case, however, such data is not available. In order to address this issue, we rely on a careful visual analysis of the selected images.

A total of 4 groups (see section 3.2) known as Regions of Interest (ROIs), each corresponding to a given class, are formed. The ground truth datasets are then constructed through a classification process that satisfies an overall accuracy of 97% and 96% respectively for images acquired in 1987 and 2019. These values are calculated by computing confusion matrices between the used multispectral images and the obtained ROIs. In case of the 2019 data, a Sentinel-2A high resolution image (10 meters) was also exploited to further confirm ROIs precision.

Using the generated ground truth, confusion matrices are computed based on a stratified random sampling [42] process to assess the proposed methodology in comparison with the unsupervised KHM classification algorithm. Both approaches are applied on the two selected Landsat multispectral images of 1987 and 2019. Finally, Table 4 reports the average values between user's and producer's accuracies with the showed improvement (the difference between values) for sparse and dense vegetation classes.

In this case, we chose to not report the overall accuracy, despite values being higher than 80%, as our work focusses on vegetation extraction only.

**Table 4.** Accuracy values using confusion matrices.

	1987		2019	
	<i>Sparse</i>	<i>Dense</i>	<i>Sparse</i>	<i>Dense</i>
<b>Proposed methodology</b>	75.1%	71.4%	82.3%	72.9 %
<b>KHM unsupervised classification</b>	58.9%	54.5%	59.4%	37.5%
<b>Difference</b>	16.2%	16.9%	22.9%	35.4%

According to [26] a classification with an accuracy above 70% could be used in practice. From Table 4 we can see that the proposed methodology reported satisfying and practical results. In comparison with the KHM classification of images, there are significant improvements of 16.55% and 29.15% respectively for 1987 and 2019 data. The obtained results are especially interesting, considering that they are achieved by simply introducing a couple spectral indices. In addition, the unsupervised characteristic of the proposed methodology is extremely useful in the case of limited resources such as Landsat images that do not usually provide ground truth labels.

However, the major drawback of the proposed techniques is its high computational time, where several KHM classifications are performed. Though, other advanced classifiers like Random Forest and Support Vector Machine, known for best achieving results, are equally time consuming. Another issue resides in the poor extraction of urban areas. The use of urban spectral indices in combination with vegetation spectral indices presents a good future perspective.

## 4.2 Cluster validity indices

In this section, we perform a comparison of the proposed combination technique with two established methodologies that are Majority Voting and Data Integration. The obtained classification results from each method are evaluated through cluster validity indices [44]. A total of three criteria [44] were selected namely: the Xie-Beni Index (XB), the Davies–Bouldin Index (DB) and the Sum-of-Squares based Index (WB).

Validity indices are often used to evaluate the quality of classification results, where minimum values are associated with better performances [45].

The Majority Voting was applied on the KHM classifications of the four correlated indices (see section 3.3). As for the Data Integration technique, we combined all of Landsat spectral bands with the four correlated indices bands and classified the new data using the KHM algorithm also.

**Table 5.** Cluster validity indices for 1987 classification results.

<b>1987</b>	<b>XB</b>	<b>DB</b>	<b>WB</b>
<b>Proposed Methodology</b>	2.13	2.94	0.35
<b>Majority Voting</b>	2.14	2.94	0.35
<b>Data Integration</b>	2.52	1.52	1.09

**Table 6.** Cluster validity indices for 2019 classification results.

<b>2019</b>	<b>XB</b>	<b>DB</b>	<b>WB</b>
<b>Proposed Methodology</b>	0.13	1.58	0.31
<b>Majority Voting</b>	0.21	1.13	0.38
<b>Data Integration</b>	3.06	2.78	1.0

Table 6 shows the outperformance of the proposed methodology in case of the 2019 data for the majority of indices. The 1987 classification results (Table 5), on the other hand, are almost similar especially in comparison with Majority Voting. This is explained by the fact that the KHM classifications of all indices returned almost identical results. This could also be seen within their correlation values close to +1 ( $\approx 0.99$ ) suggesting similar outcomes.

From both Tables we can observe that the proposed methodology closely matches performances with Majority Voting. However, our technique has the notable advantage of combining a pair of indices (two indices). In comparison, Majority Voting is computed from a minimum of three indices.

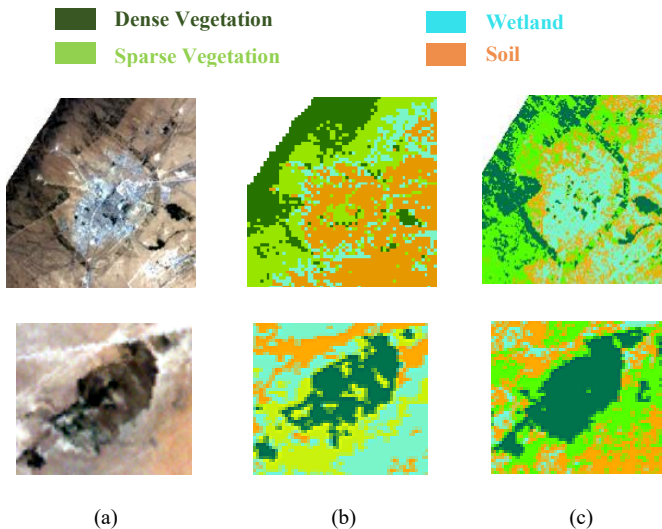
As for the KHM algorithm, it showed satisfactory classification results despite high computational time for both Majority Voting and the proposed method. Finally, Data Integration technique possessed the fastest execution time of the three but with the lowest results.

### 4.3 Visual Interpretation

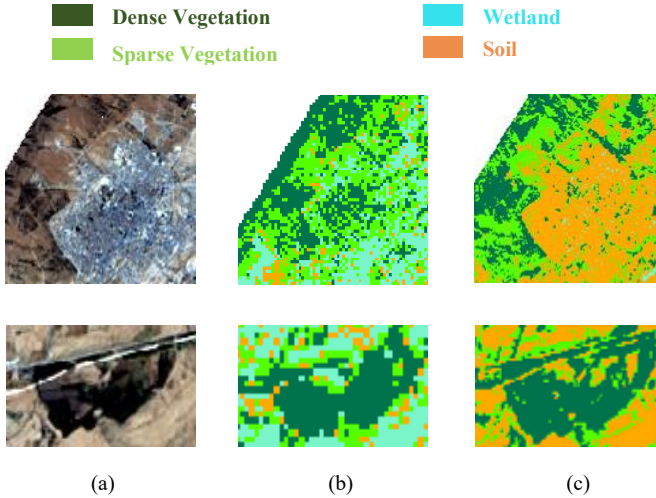
A visual comparison over the obtained results with their corresponding RGB images is shown in Fig .4 and Fig .5. Two vegetation areas covering reforestations and agricultural parcels are selected and compared.

The performances of the proposed methodology are clearly visible especially in reforestation parcels (Top section of Fig .4 and Fig .5) with an increased definition and a better visual quality. In addition, we can observe from the top section of Fig .5 that the KHM classification on the Landsat images included parts of urban structure within the obtained vegetation class further confirming our results.

From Fig. 4 bottom section, we could notice a small confusion between sparse and dense vegetation by the proposed methodology. However, the latter performed well in general and only showed a few areas of error as suggested by the accuracy analysis (see section 4.1).



**Fig. 4.** Visual interpretation of Landsat 5 image (1987). Top: reforestation parcels. Bottom: agricultural parcels. (a) Landsat 5 RGB image. (b) KHM classification of 1987 image. (c) Classification results using the proposed methodology.



**Fig. 5.** Visual interpretation of Landsat 8 image (2019). Top: reforestation parcels. Bottom: agricultural parcels. (a) Landsat 8 RGB image. (b) KHM classification of 2019 image. (c) Classification results using the proposed methodology.

## 5 conclusion

In this study, an automated scheme for an improved classification of vegetation land cover was introduced using Landsat open data. The methodology is rooted within ensemble clustering techniques and relies a simple combination of surfaces and distances to merge the results of several VIs. In the proposed procedure, the KHM algorithm is computed to perform an unsupervised classification of selected indices. The purpose being the use of open source unlabeled Landsat datasets. For this study, the region of Mecheria, in Naâma Algeria, is selected over two dates: 1987 and 2019. The natural characteristic of the city combined with its uncontrolled pasture system, has led to severe land and vegetation covers degradation over the years despite the increase of agricultural areas. This is due to the Algerian government plan to establish new agricultural perimeters in arid region in the 2000s. The phenomenon can be observed through the obtained results with satisfactory performances in terms of accuracy and quality using confusion matrices and cluster validity indices respectively.

The major drawbacks of the proposed work were finally discussed consisting of a high computational time and the inability to extract urban areas. Consequently, it could be interesting to combine vegetation indices with urban spectral indices while using a faster classifier for future perceptive.

## References

1. BENARADJ, A. : Mise en défens et remontée biologique des parcours steppique dans la région de Naama dissémination et multiplication de quelque espèce stéppique. Doctoral dissertation, Université Mohamed Boudiaf des Sciences et de la Technologie-Mohamed Boudiaf d'Oran (2009).
2. Bouzenoune, A. : Etude phytogéographique et phytosociologique des groupements végétaux du Sud oranais (wilaya de Saida). Th. Dot. 3<sup>e</sup> cycle, Alger, Univ. Sci. Tech. Houari BOUMEDIENE (1984).
3. Ozenda, P.: Flora of the Sahara. pp. 625-pp (1977).
4. Haddouche, I., Toutain, B., Saidi, S., & Mederbal, K.: How to reconcile the development of steppe populations and the fight against desertification? Case of the wilaya of Nâama (Algeria). *New Medit* (2008).
5. Khedidja, B., Remaoun, K., & Ahmed, S. S.: GIS and remote sensing-desertification and progression of silting-up in the high plains of Oran in Western Algeria. *Technium Soc. Sci. J.*, 40, 481 (2023).
6. Boucherit, H., Benabdeli, K., & Benaradj, A.: Biological recovery the steppe of Hammada scoparia after enclosure in the region of Naama (Algeria). *Ekológia (Bratislava)*, 36(1), 52-59 (2017).
7. Zhan, Z., Zhang, X., Liu, Y., Sun, X., Pang, C., & Zhao, C.: Vegetation land use/land cover extraction from high-resolution satellite images based on adaptive context inference. *IEEE Access*, 8, 21036-21051 (2020).
8. Sishodia, R. P., Ray, R. L., & Singh, S. K.: Applications of remote sensing in precision agriculture: A review. *Remote sensing*, 12(19), 3136 (2020).
9. Xie, Y., Li, J., Wulan, T., Zheng, Y., & Shen, Z.: Scale dependence of forest fragmentation and its climate sensitivity in a semi-arid mountain: Comparing Landsat, Sentinel and Google Earth data. *Geography and Sustainability*, 5(2), 200-210 (2024).
10. Lemenkova, P.: Exploitation d'images satellitaires Landsat de la région du Cap (Afrique du Sud) pour le calcul et la cartographie d'indices de végétation à l'aide du logiciel GRASS GIS. *Physio-Géo. Géographie physique et environnement*, (Volume 20), 113-129 (2024).
11. Potapov, P., Turubanova, S., & Hansen, M. C.: Regional-scale boreal forest cover and change mapping using Landsat data composites for European Russia. *Remote Sensing of Environment*, 115(2), 548-561 (2011).
12. Li, W., Dong, R., Fu, H., Wang, J., Yu, L., & Gong, P.: Integrating Google Earth imagery with Landsat data to improve 30-m resolution land cover mapping. *Remote Sensing of Environment*, 237, 111563 (2020).
13. Dutra, D. J., Elmiro, M. A. T., & Garcia, R. A.: Comparative analysis of methods applied in vegetation cover delimitation using Landsat 8 images. *Sociedade & Natureza*, 32, 732-744 (2020).
14. Almalki, R., Khaki, M., Saco, P. M., & Rodriguez, J. F.: Monitoring and mapping vegetation cover changes in arid and semi-arid areas using remote sensing technology: a review. *Remote Sensing*, 14(20), 5143 (2022).
15. Liu, H., Zhang, F., Zhang, L., Lin, Y., Wang, S., & Xie, Y.: UNVI-based time series for vegetation discrimination using separability analysis and random forest classification. *Remote Sensing*, 12(3), 529 (2020).
16. Lebrini, Y., Boudhar, A., Hadria, R., Lionbouï, H., Elmansouri, L., Arrach, R., ... & Benabdelouahab, T.: Identifying agricultural systems using SVM classification approach based on phenological metrics in a semi-arid region of Morocco. *Earth Systems and Environment*, 3(2), 277-288 (2019).

17. Valero-Jorge, A., González-De Zayas, R., Matos-Pupo, F., Becerra-González, A. L., & Álvarez-Taboada, F. Mapping and Monitoring of the Invasive Species *Dichrostachys cinerea* (Marabú) in Central Cuba Using Landsat Imagery and Machine Learning (1994–2022). *Remote Sensing*, 16(5), 798 (2024).
18. Bayramoğlu, Z., & Melis, U. Z. A. R.: Performance analysis of rule-based classification and deep learning method for automatic road extraction. *International Journal of Engineering and Geosciences*, 8(1), 83-97 (2023).
19. Mullapudi, A., Vibhute, A. D., Mali, S., & Patil, C. H.: Spatial and seasonal change detection in vegetation cover using time-series landsat satellite images and machine learning methods. *SN Computer Science*, 4(3), 254 (2023).
20. Navarro, R., Wirkus, L., & Dubovyk, O.: Spatio-Temporal Assessment of Olive Orchard Intensification in the Saïss Plain (Morocco) Using k-Means and High-Resolution Satellite Data. *Remote Sensing*, 15(1), 50 (2022).
21. Liu, D. Z., Yang, F. F., & Liu, S. P.: Estimating wheat fractional vegetation cover using a density peak k-means algorithm based on hyperspectral image data. *Journal of Integrative Agriculture*, 20(11), 2880-2891 (2021).
22. Yankovich, K. S., Yankovich, E. P., & Baranovskiy, N. V.: Classification of vegetation to estimate forest fire danger using landsat 8 images: Case study. *Mathematical Problems in Engineering*, 2019 (2019).
23. Amalita, N., Vionanda, D., & Permana, D.: Grouping Level of Poverty Based on District/City in Indonesia Using K-Harmonic Means. *UNP Journal of Statistics and Data Science*, 1(3), 157-163 (2023).
24. Tian, Y., Liu, D., & Qi, H.: K-harmonic means data clustering with differential evolution. In 2009 International Conference on Future BioMedical Information Engineering (FBIE) (pp. 369-372). IEEE (2009, December).
25. Yunistya, D. I., Goejantoro, R., & Amijaya, F. D. T.: The Application Of K-Harmonic Means Method In District/City Grouping. *Jurnal Matematika, Statistika dan Komputasi*, 19(1), 51-64 (2022).
26. Sun, X. F., & Lin, X. G.: Random-forest-ensemble-based classification of high-resolution remote sensing images and nDSM over urban areas. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42, 887-892 (2017).
27. Allbed, A., & Kumar, L.: Soil salinity mapping and monitoring in arid and semi-arid regions using remote sensing technology: a review. *Advances in remote sensing*, 2013 (2013).
28. Yang, X., Qin, Q., Yésou, H., Ledauphin, T., Koehl, M., Grussenmeyer, P., & Zhu, Z.: Monthly estimation of the surface water extent in France at a 10-m resolution using Sentinel-2 data. *Remote Sensing of Environment*, 244, 111803 (2020).
29. Nayini, S. E. Y., Geravand, S., & Maroosi, A.: A novel threshold-based clustering method to solve K-means weaknesses. In 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS) (pp. 47-52). IEEE (2017, August).
30. Das, S., & Singh, T. P.: Correlation analysis between biomass and spectral vegetation indices of forest ecosystem. *Int. J. Eng. Res. Technol*, 1(5), 1-13 (2012).
31. Ren, H., & Zhou, G.: Determination of green aboveground biomass in desert steppe using litter-soil-adjusted vegetation index. *European Journal of Remote Sensing*, 47(1), 611-625 (2014).
32. Sonobe, R., Yamaya, Y., Tani, H., Wang, X., Kobayashi, N., & Mochizuki, K. I.: Crop classification from Sentinel-2-derived vegetation indices using ensemble learning. *Journal of Applied Remote Sensing*, 12(2), 026019-026019 (2018).

33. Alex, E. C., Ramesh, K. V., & Sridevi, H.: Quantification and understanding the observed changes in land cover patterns in Bangalore. *International Journal of Civil Engineering and Technology*, 8(4), 597-603 (2017).
34. Luz, L. R., Giongo, V., Santos, A. M. D., Lopes, R. J. D. C., & Júnior, C. D. L.. Biomass and vegetation index by remote sensing in different caatinga forest areas. *Ciência Rural*, 52, e20201104 (2021).
35. SANTOS, F.: Albedo Seasonal Behavior and vegetation indices in the upper basin Paraíba River. *Revista Brasileira de Geografia Física*. v07, (5), 1015-1027 (2014).
36. Egghe, L., & Leydesdorff, L.: The relation between Pearson's correlation coefficient  $r$  and Salton's cosine measure. *Journal of the American Society for information Science and Technology*, 60(5), 1027-1036 (2009).
37. Mahi, H., Farhi, N., & Labeled, K.: Remotely sensed data clustering using K-harmonic means algorithm and cluster validity index. In *Computer Science and Its Applications: 5th IFIP TC 5 International Conference, CIIA 2015, Saida, Algeria, May 20-21, 2015, Proceedings 5* (pp. 105-116). Springer International Publishing (2015).
38. Bensaïd, A. : SIG ET TÉLÉDÉTECTION POUR L'ÉTUDE DE L'ENSABLEMENT DANS UNE ZONE ARIDE: LE CAS DE LA WILAYA DE NAËMA (ALGÉRIE). Doctoral dissertation, Université Joseph-Fourier-Grenoble I (2006).
39. Ren, H., Zhou, G., & Zhang, F.: Using negative soil adjustment factor in soil-adjusted vegetation index (SAVI) for aboveground living biomass estimation in arid grasslands. *Remote Sensing of Environment*, 209, 439-445 (2018).
40. Fatiha, B., Abdelkader, A., Latifa, H., & Mohamed, E.: Spatio temporal analysis of vegetation by vegetation indices from multi-dates satellite images: Application to a semi arid area in ALGERIA. *Energy Procedia*, 36, 667-675 (2013).
41. Vani, V., & Mandla, V. R.: Comparative study of NDVI and SAVI vegetation indices in Anantapur district semi-arid areas. *Int. J. Civ. Eng. Technol*, 8(4), 559-566 (2017).
42. Capolupo, A., Monterisi, C., & Tarantino, E.: Landsat images classification algorithm (LICA) to automatically extract land cover information in Google Earth Engine environment. *Remote Sensing*, 12(7), 1201 (2020).
43. Abubakar, G. A., Wang, K., Shahtahamssebi, A., Xue, X., Belete, M., Gudo, A. J. A., ... & Gan, M.: Mapping maize fields by using multi-temporal Sentinel-1A and Sentinel-2A images in Makarfi, Northern Nigeria, Africa. *Sustainability*, 12(6), 2539 (2020).
44. Meroufel, H., Mahi, H., & Farhi, N.: Comparative Study between Validity Indices to Obtain the Optimal Cluster. *International Journal of Computer Electrical Engineering*, 9(1), 1-8 (2017).
45. Arbelaitz, O., Gurrutxaga, I., Muguerza, J., Pérez, J. M., & Perona, I.: An extensive comparative study of cluster validity indices. *Pattern recognition*, 46(1), 243-256 (2013).

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

