Bulletin of the Mineral Research and Exploration

http://bulletin.mta.gov.tr

Segmentation and classification algorithms applied to sentinel-2A images for geological mapping: case of the Al Glo'a sheet (1/50000), Morocco

Abdessamad EL ATILLAH[a]*, Mouhssine EL ATILLAH[b], Zine El Abidine EL MORJANI[a], Khalid EL FAZAZY[b] and Mustapha SOUHASSOU[a]

[a]*Ibn Zohr University, Polydisciplinary Faculty of Taroudant, Egerne, Taroundant, Morocco*
[b]*Computer Science, Signals, Automation and Cognitivism Laboratory (LISAC), FSDM, USMBA, Fes, Morocco*

Research Article

Keywords:
Segmentation Algorithms, Classification Algorithms, Sentinel 2A, Geological Mapping, Anti-Atlas.

ABSTRACT

The multispectral image, of Landsat 7 and 8; Aster and Sentinel-2A, has good results in lithological, structural, hydrothermal and mineralogical alteration mapping. Segmentation and image classification are two complementary steps as they are allowed the most important operations in the field of image processing. In this sense, this work aims at evaluating the potential of segmentation and classification algorithms for the generation of surface geological maps, hydrothermal alterations and lineaments. Given the good resolution of the Sentinel 2A (10m), the three images, (11/12; 11/2; 11/8), 12.8.2, main component 1, 2 and 3 (11.12.2), processed by five algorithms (K-means, isodata, watershed, efficient graph-based image segmentation, thresholding) for geological mapping and then mining exploration. The study displayed that 1) Watershed algorithm can be used for topographic and hydraulic studies, it can be very useful in the preparation phase of geological and mining infrastructures; 2) threshold segmentation does not give good results in terms of geological discrimination since it divides each image into two parts; 3) the same thing for the effective thresholding and segmentation of graph-based images; 4) The Isodata and K-means algorithms show good geological discrimination.

Received Date: 25.08.2020
Accepted Date: 19.01.2021

1. Introduction

The new Moroccan Mining Code, Law 33-13, has provided opportunities for exploration and research of a large-area of mineral resources through the introduction of the exploration permit with an area of 2.400 square kilometer maximum and the merger of research license.

The multispectral image (Landsat 7, Landsat 8 and Aster) processing generally have good results in litho-structural mapping. The Sentinel 2A image at 10 m resolution has shown great capacity in mapping areas with mineral potential by assembling lithological, structural and hydrothermal alteration data.

The aim of this work is to evaluate the potential of segmentation and classification algorithms for the generation of surface geological maps, hydrothermal alterations and lineaments. These areas may be the subject of a study on conventional methods of mining research to identify the physicochemical and mining characteristics of the geological formations of the Al Glo'a sheet (1/50000).

Citation Info: El Atillah, A., El Atillah, M., El Morjani, Z.E.A., El Fazazy, K., Souhassou, M. 2021. Segmentation and classification algorithms applied to sentinel-2A images for geological mapping: case of the Al Glo'a sheet (1/50000), Morocco. Bulletin of the Mineral Research and Exploration 166, 113-125. https://doi.org/10.19111/bulletinofmre.864492

**Corresponding author: Abdessamad EL ATILLAH, elatillah@gmail.com*

2. Local Geological Setting (Test Area)

The Al Glo'a sheet (1/50000) is a part of the Precambrian inlier of Bou Azzer El Graara of the Central Anti-Atlas in South East Morocco. This inlier (Figure 1) is a pivotal area whose study has greatly contributed to the improvement of our knowledge of the pan-African chain. The Precambrian outcrops along two contiguous massifs, Bou Azzer and El Graara, following a Variscan antiform structure that runs along the major accident of the Anti-Atlas (Choubert, 1947). Precambrian sites in the Bou Azzer-El Graara inlier can be divided into two large sets, one metamorphic, the other non-metamorphic which is discordant on the first (Oukassou, 2013).

The Tonien, Cryogenian, and Lower to Middle Ediacarean terrains (Precambrian II2 and II3) outcrop in the center of the Al Glo'a sheet (Figure 2), and they are covered elsewhere by a volcano-sedimentary cover of the Upper Ediacaran (Ouarzazate Group) then Lower Palaeozoic sedimentary (Soulaimani, et al., 2013).

3. Material and Method

Segmentation and classification are two complementary steps: segmentation consists of cutting out the elements constituting an image while classification is the identification to which category or class belongs each element or part of the image. In geology, categories or classes are known a priori in the case of supervised classification.

Image segmentation is one of the most important operations in the image processing field. Its purpose is to gather pixels together according to predefined criteria. The pixels are thus grouped into segments, which constitute a partition of the image. It can be for example to separate the objects from the bottom if the number of classes is equal to two, it is also called binarization (Maarir et al., 2014). The objective of this operation is to classify the image into homogeneous areas called segments. Thus, the segmentation of an image makes it possible to find forms or areas having a meaning for the user or for another image processing operation (Dupas, 2009). The methods used are based on contour detection, threshold, histogram, region-based methods and watershed transformation (Khan, 2014). This technique is based on two important approaches: approach regions and approach contours or border; there are other less important approaches: Mumford Shah, deformable models, level sets, and Markov fields (Vialard, 2018). We have several algorithms that deal with this theme.

A segmentation is parameterized by a homogeneity predicate P: R → {true, false} indicating whether a region is homogeneous according to the criterion tested. In order to predicate to be consistent, it is necessary that each subset of a homogeneous region be homogeneous. A segmentation algorithm is thus

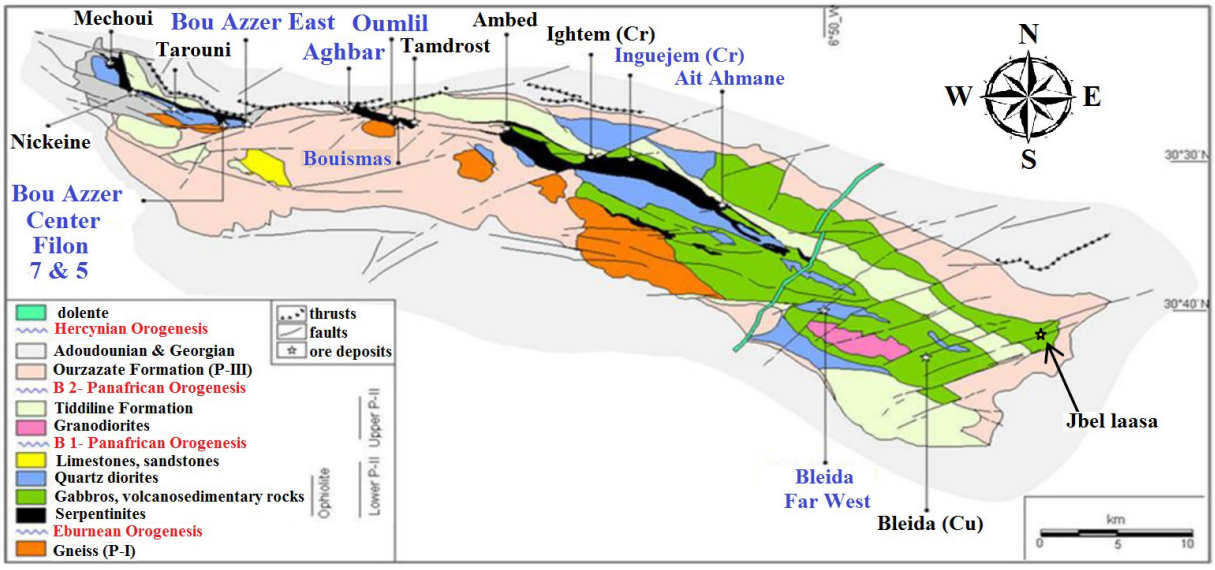


Figure 1- Map diagram of the El Graara - Bou Azzer inlier (modified from Maacha et al., 2014).

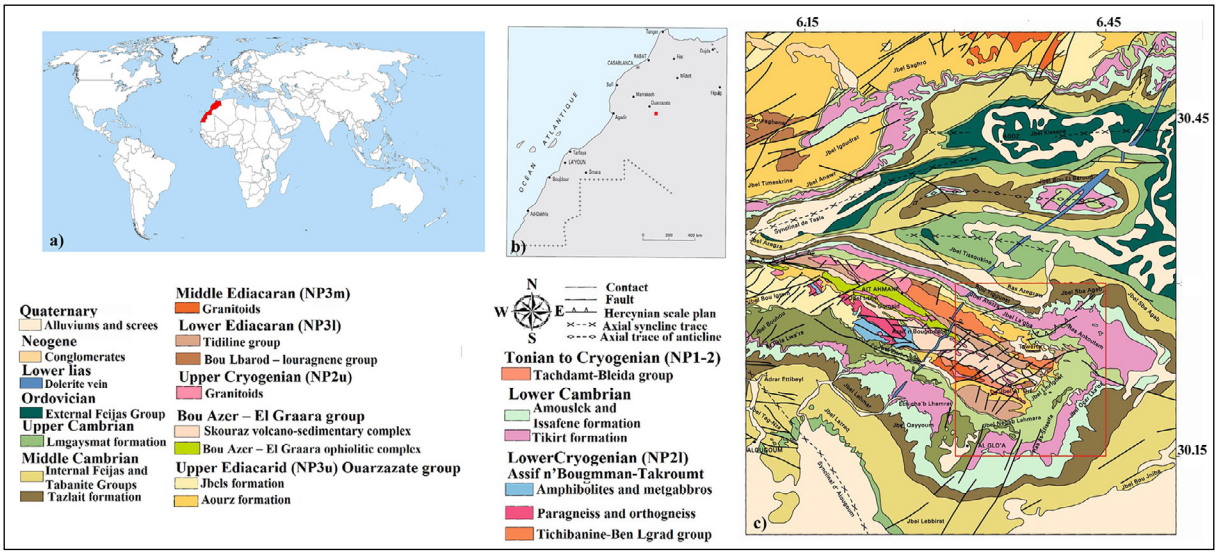


Figure 2- Study area location; a) Morocco's location in the world, b) the location of the sheet of Al Glo'a in Morocco, c) geological map of the study area (Al Glo'a sheet, rectangle in red, 1/500000) (modified from Soulaimani et al., 2013).

parameterized by a predicate P and an image. It returns a partition of the image in homogeneous zones (Pavlidis, Horowitz, and Theodosios, 1976).

Segmentation in homogeneous regions of an image I for a predicate P is a partition R of I such that:

1. $I = \cup_{r \in R} r$

2. $r_i \cap r_j = \varnothing$ for all $r_i, r_j \in R^2, i \neq j$.

3. P(r) = true for all $r \in R$ için.

4. $P\ r_i \cap r_j$ = false for all $r_i, r_j \in R^2$, $i \neq j$, r_i and r_j adjacent.

The first condition ensures that all elements of an image belong to a region and therefore the union of regions represents the entire image. The second condition states that regions do not overlap: there is no intersection between regions. The third condition indicates that each region must be homogeneous according to the criterion defined by the oracle. Finally, the last condition specifies that the fusion of two adjacent regions must not be homogeneous.

Starting from these conditions, numerous segmentation methods exist to solve many different problems. According to (Serra, 2006), there are more than a thousand distinct approaches.

3.1. Watershed

The watershed is a powerful tool for analyzing the topography of an image, since it can delimit a set of areas that form a partition of the original image. In this context, the watershed is generally applied to the contour image, where the strong contours can be seen as ridges that separate fairly flat valleys (homogeneous regions). This transformation is usually applied to the image gradient; so the basins correspond to homogeneous zones of the image (Urtasun, 2017). Like most classic segmentation methods this transformation gives over-segmented images which produces deferent instances on the same object. The Watershed algorithm (Vincent and Soille, 1991) is a method derived from mathematical morphology, which directly extracts closed and skeletonized contours from a contour power image (gradient module image). This algorithm is very interesting to reduce the number of false contours; the basin dynamics associated with a local minimum is defined as the difference in altitude between this minimum and the lowest peak point that must be crossed to arrive in a basin with a minimum lower than the first. The dynamics of a contour arc are defined as being the maximum value of the threshold for which this contour arc remains, when all the basins with a dynamics lower than the threshold are deleted (Ducrot, 2005).

The watershed is a supervised segmentation method based on the growth of regions using the segmented gradient image standard whose pixels with the lowest intensity form the initial watersheds. So the efficiency of this algorithm is related to the gradient of the image that is to say the number of minima obtained. If it is close to many of the objects constituting the original image, the result will be a well-segmented image. On the other hand, the result obtained will be segmented. Watershed segmentation is a region-based growth segmentation technique inspired by watersheds for rivers. The algorithm task on the gradient image representing the image. The gradient image associates with each element of the image a numerical value expressing the gradient between the values of the pixels of the image. The seeds are automatically selected as local minimums of the gradient image (i.e. all neighbors of a seed have a greater or equal gradient value). Each seed determines a new pool. The elements of the image are then attributed to the seed so as to fill the basins. The boundaries between two basins are called ridges.

Algorithm:

Calculate the gradient (or Laplacian) of the image.

Pixels with the lowest intensity form the initial watersheds.

For each intensity level i:

For each pixel group of intensity i:

If adjacent to exactly one existing region:

add these pixels in this region.

If adjacent to several regions simultaneously:

mark as the watershed.

If not, start a new region.

3.2. Segmentation by Thresholding

Thresholding aims to segment an image into several classes using only the histogram (MacQueen, 1967) (Figure 3).

The principle of Segmentation by thresholding is to extract thresholds from the histogram (image / region) and the classification of a pixel p by comparison of I (p) with the thresholds. There are three types of thresholding methods:

• Global threshold: S (x; y) def = S (I (x; y))

• Local threshold: S (x; y) def = S (I (x; y); P (x; y))

• Dynamic thresholding: S (x; y) def = S (I (x; y); P (x; y); x; y)

Here we should demonstrate that we talk about binarization if the image cantains 2 classes. (such as: pixel: (x; y), gray level: I (x; y), local property: P (x; y), and threshold used to classify the pixel (x; y): S (x; y)) (Vialard, 2018).

3.3. Effective Image Segmentation Based on Graphs

Efficient image segmentation based on PEGIS, Python Efficient Graph Based Image Segmentation, graphs made by (Felzenszwalb and Huttenlocher, 2004). This algorithm runs in time almost linearly in the number of edges of the graph and is also fast in practice. An important characteristic of the method is its capacity to retain detail in low variability image zones while ignoring details in zones of high variability. The algorithm deals with the problem of segmenting an image into regions. It defines a predicate to measure the evidence of a border between two regions using a graphic representation of the image. It then develops an efficient segmentation algorithm based on this predicate; and shows that even if this algorithm makes greedy decisions; it produces segmentations satisfying global properties (Felzenszwalb and Huttenlocher, 2004).

3.4. The k-Means Algorithm

The k-means algorithm (or k-means in English) is a method of data partitioning and a combinatorial optimization problem. Given points and an integer

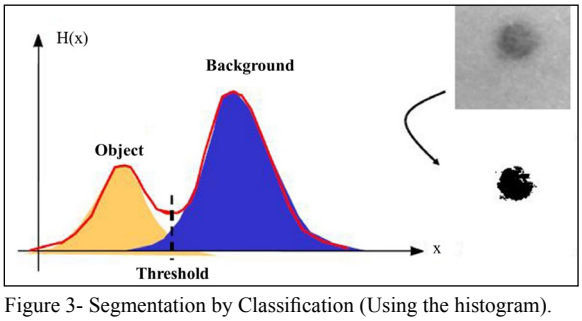


Figure 3- Segmentation by Classification (Using the histogram).

k, the problem is to divide the points into k groups, often called clusters, in order to minimize a certain function. We consider the distance of a point to the average of the points of its cluster; the function to be minimized is the sum of the squares of these distances (simplilearn, 2020), (Fahim, 2006).

K-means is an unsupervised learning algorithm dedicated to non-hierarchical clustering. It is an iterative algorithm that minimizes the sum of the distances between each individual and the centroid. The initial choice of centroids determines the final result.

Admitting a cloud of a set of points, K-Means changes the points of each cluster until the sum cannot decrease. The result is a set of compact and clearly separated clusters, subject to choosing the right value of the number of clusters (Mrmint, 2018).

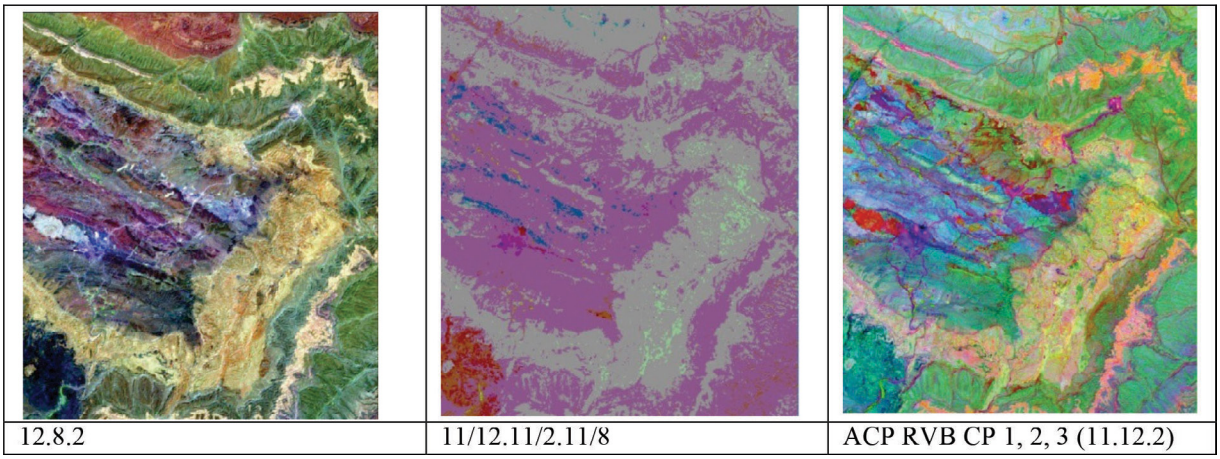
3.5. The Isodata Algorithm

The ISODATA-algorithm is an iterative method that utilizes Euclidean distance as a measure of similarity to group data elements into different classes. To reduce the processing load and increase the throughput, the ISODATA procedure is usually applied to the first principal component derived from the original set of multi-spectral images. The disadvantage of the principal component approach is that it relies entirely on the statistical significance of the spectra, rather than on the uniqueness of the individual spectra (Dhodhi et al., 1999). The Iterative Self-Organizing Data Analysis Technique (ISODATA) (Ball and Hall, 1965) is sometimes used to refine the clusters obtained by a partitioning algorithm. This technique consists in exploding or merging the clusters according to pre-established thresholds. For example, two groups Ci and Cj will be merged if their inter-center distance d (x * i, x * j) is less than a certain threshold α. Likewise, a Ci cluster of inertia greater than a given threshold β will be exploded. In this family of algorithms, the number of expected clusters is not fixed a priori (Cleuziou, 2004).

3.6. MSI Images of the Sentinel 2A

Sentinel 2A's MSI instrument captures images of thirteen spectral bands ranging from visible to mid-infrared over a width of 290 km: The four spectral bands (blue (490 nm), green (560 nm), red (670 nm) nm) and near infrared (850 nm) having a resolution of 10 m, three spectral bands (440, 940 and 1370 nm) are provided with a resolution of 60 m while the resolution of the other six bands is 20 m.

Good practices in this area of use of satellite imagery (including Landsat 7, Landsat 8, Aster) for mineral exploration have proved to be useful and effective. And given the good resolution of the Sentinel 2A, we have opted for an analogy between the bands of these satellites and those of the Sentinel 2A in order to find derivative methods with a great capacity of geological discrimination; this work concludes that the RGBs (Red-Green-Blue) of the following Sentinel 2A images (Figure 4): (11/12; 11/2; 11/8), 12.8.2, Main Component 1,2 and 3 (11.12.2) show a good discrimination.



12.8.2	11/12.11/2.11/8	ACP RVB CP 1, 2, 3 (11.12.2)

Figure 4- RGB Sentinel 2A images.

These three images will be processed by the aforementioned algorithms in order to perform geological mapping with good discrimination. All these processed images listed in this paper have the same scale (1/50000) and the same coordinates as the GLO'A geological map; They are oriented North.

4. Findings and Discussion

4.1. Watershed

In our case where the initial image is composed of three bands, the processing of this image goes through several stages; the first step is to convert the RGB values to grayscale; panchromatic image; forming a weighted sum of R; G and B according to the following relation: (0.2989 * R) +(0.5870 * G) + (0.1140 * B). These are the same weights used by the same function to calculate the Y component coefficients used to calculate the gray scale values which are identical to those used to calculate the luminance after being rounded to 3 decimals; calculation of the luminance using the following formula: (0.299 * R) + (0.587 * G) + (0.114 * B); while the second step aims at detecting the outlines of objects (areas of high values) and local minima (areas of low values). Using markers on the lowest values in the third step will ensure discovery of segmented objects. The combination of the last two stages gives us the final image (team, s.d.). To obtain an efficient segmentation, the disk value (the number of neighbors surrounding a pixel) is chosen correctly; if 3: the algorithm tests the three pixels surrounded by a pixel has a local minimum in the four locations (right, left, up, down), if the density of a pixel is one less than 20, the algorithm will add it to the tested pixel and gives them the same label, et cetera until the end of local minima. The same for the disk value. According to the results obtained for the first image below, the disk value 3 gives a more precise segmentation than the value 5 which means that the image contains several local minima (Figures 5 to 10).

On the other hand, the second image gives a more precise segmentation for the disk value, which means that the image gradient is minimized by the number of local minima relative to the first image.

Disk value 5 shows the weakness of watershed algorithm in the third image which is totally based on the results obtained by the gradient of images. This means that the absence of local minima here forces the algorithm to consider the image as a single object.

The discrimination achieved by this algorithm is not adequate with maps and existing geological documents; but it can be good for topographic image segmentation.

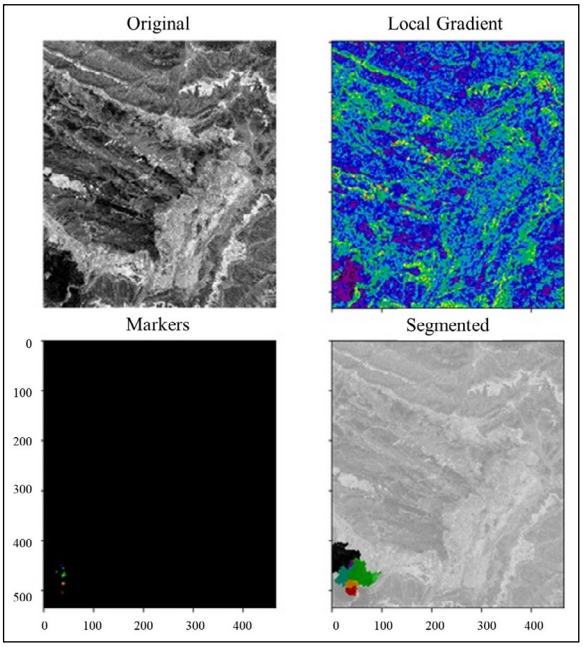


Figure 5- Watershed (RVB 12.8.2) [rank gradient (disk 3) <20].

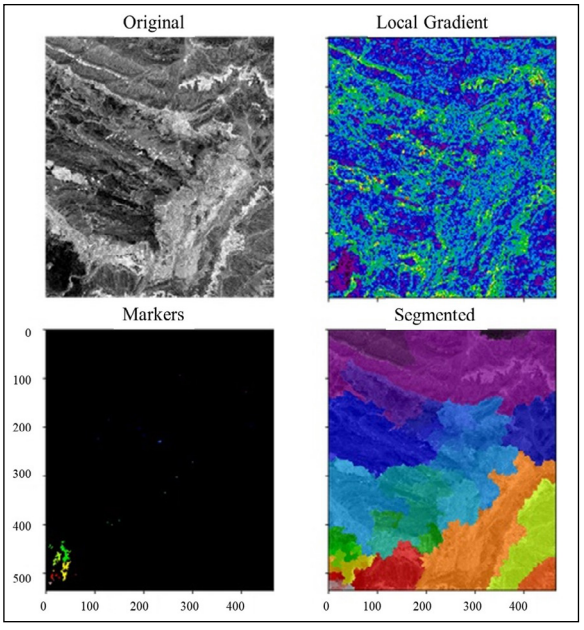


Figure 6- Watershed (RVB 12.8.2) [rank gradient (disk 5) <20].

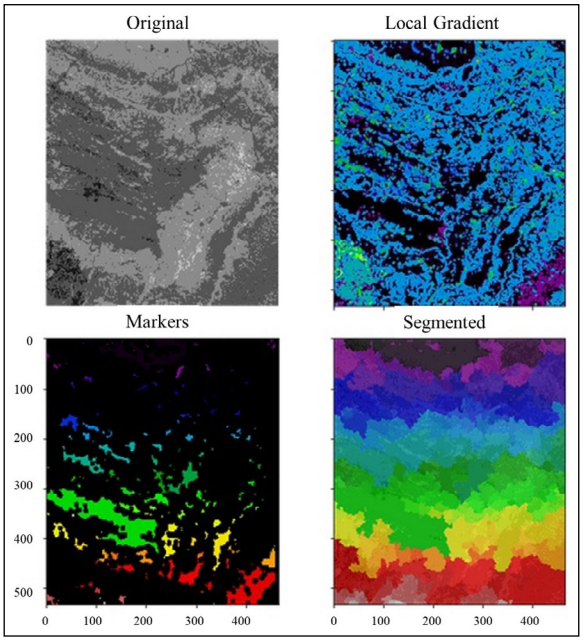


Figure 7- Watershed (RVB 11/12.11/2.11/8) [rank gradient (disk3) <20].

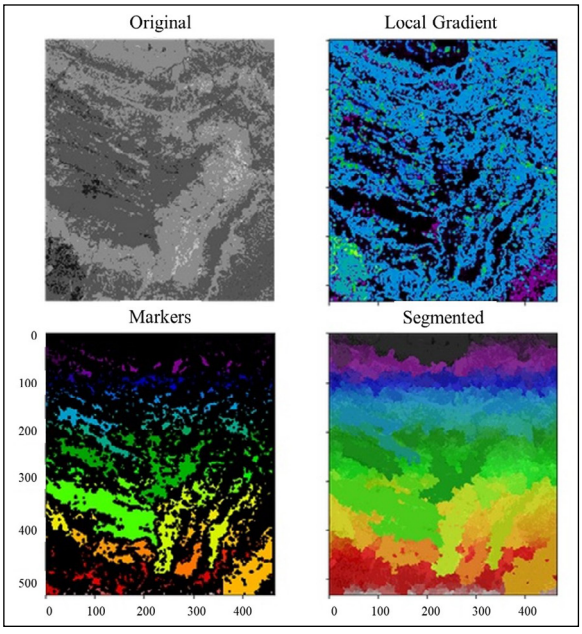


Figure 8- Watershed (RVB 11/12.11/2.11/8) [rank gradient (disk5) <20].

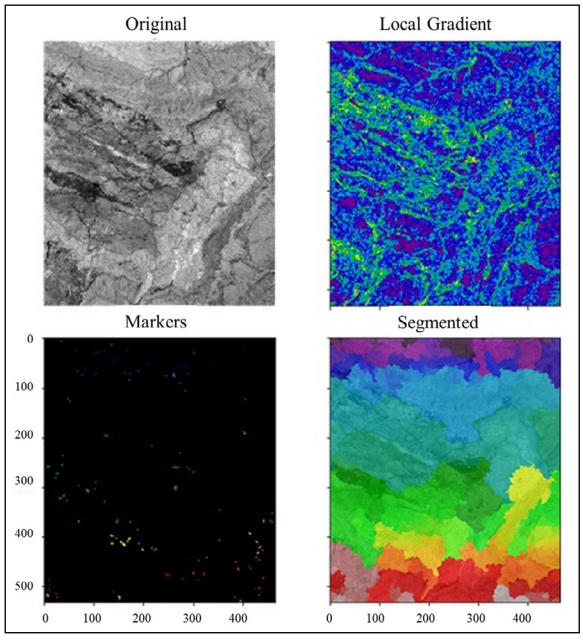


Figure 9- Watershed (RVB CP 1, 2, 3 (11.12.2)) [rank gradient (disk 3)<20].

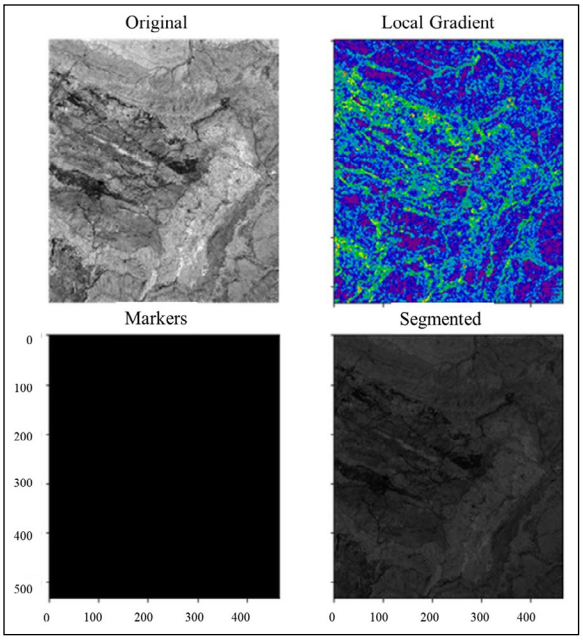


Figure 10- Watershed (RVB CP 1, 2, 3 (11.12.2)) [rank gradient (disk 5) <20].

4.2. Efficient Image Segmentation Based on Graphs

This algorithm produces a segmented image based on a graphical representation (Graph (V, E): the vector V contains the pixels and E contains a set of non-directed edges between pairs of pixels) of original image. The algorithm defines a predicate to measure the evidence of a boundary between two regions, this predicate will be the basis for constructing an efficient algorithm as a result. The problem is that the image segmentation is based on its graph and its predicates which sometimes gives undesirable results like our case because the algorithm considers the image as a homogeneous unit (Huttenlocher, 2004), (Figures 11 to 13).

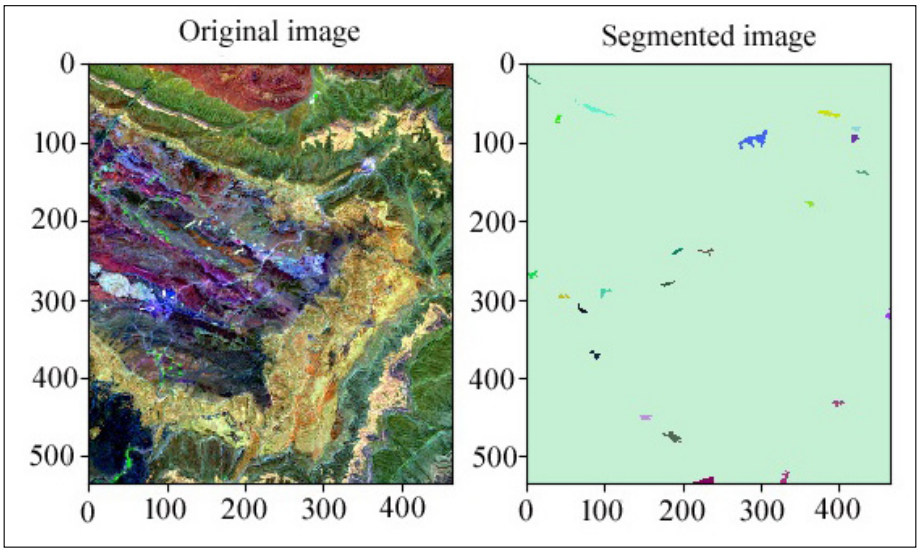


Figure 11- Effective graph-based image segmentation (RGB 12.8.2).

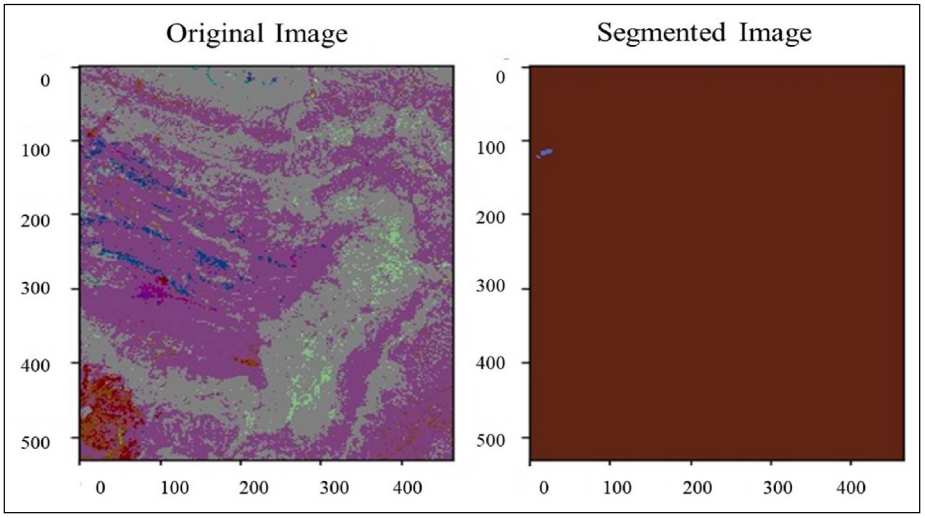


Figure 12- Effective graph-based image segmentation (RGB 11 / 12.11 / 2.11 / 8).

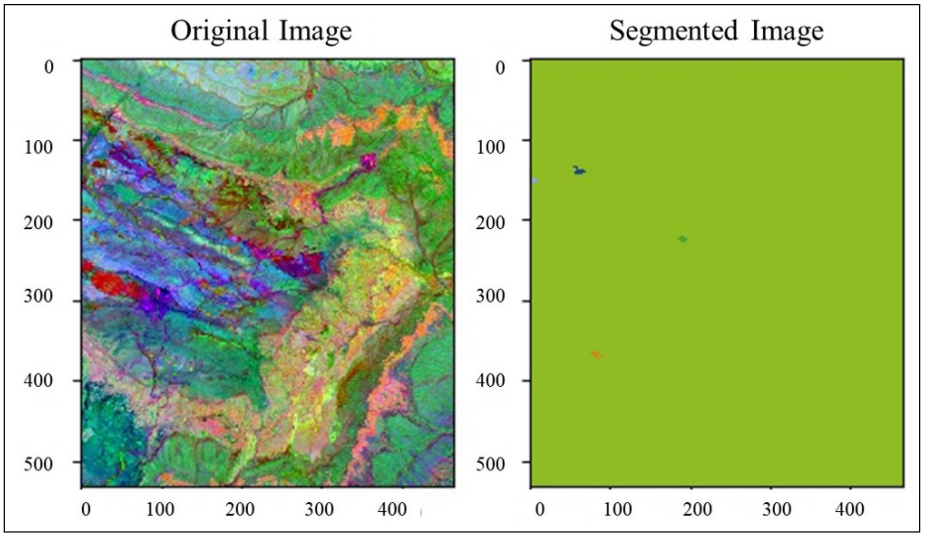


Figure 13- Effective graph-based image segmentation (RGB CP 1, 2, 3 (11.12.2)).

4.3. Segmentation by Thresholding

This algorithm remains the simplest at it transforms the image into grayscale and subdivides the image into two parts according to a threshold calculation via the histogram (number of pixels having the same level of gis) (Figures 14 to 16).

The geological discrimination of the thresholding is not detailed.

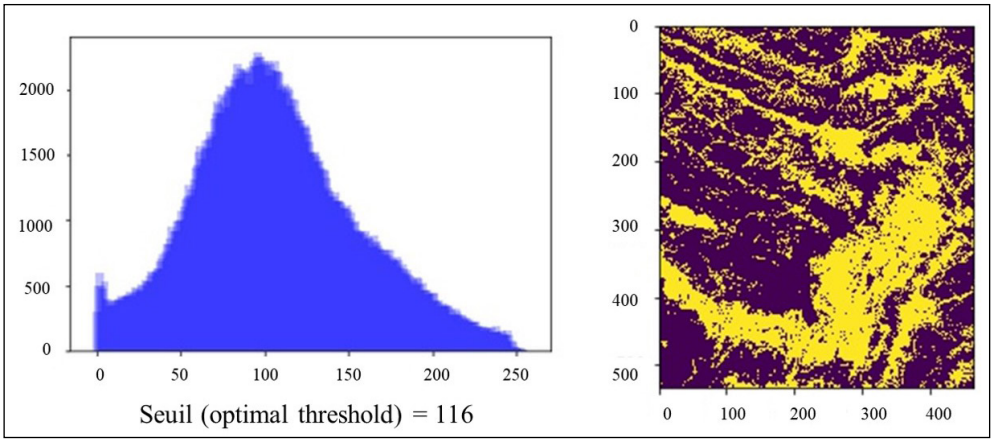


Figure 14- Segmentation by thresholding (RGB 12.8.2).

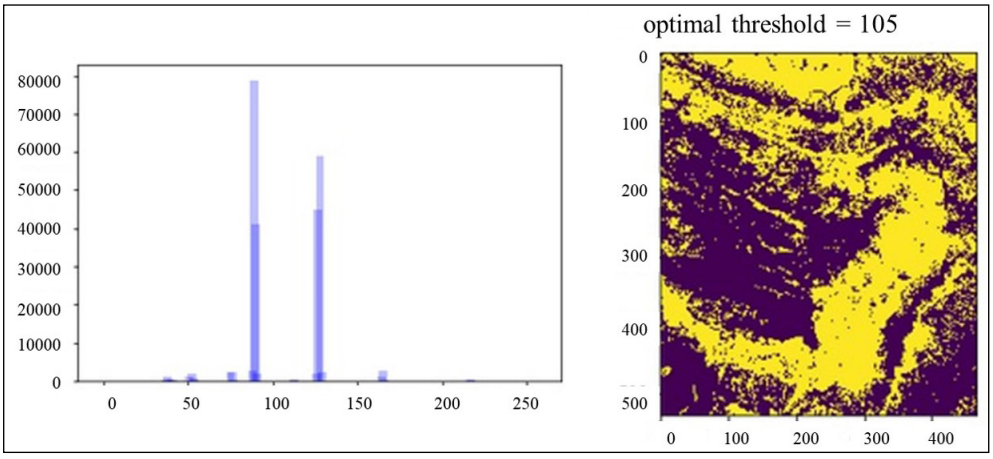


Figure 15- Segmentation by thresholding (RGB 11 / 12.11 / 2.11 / 8).

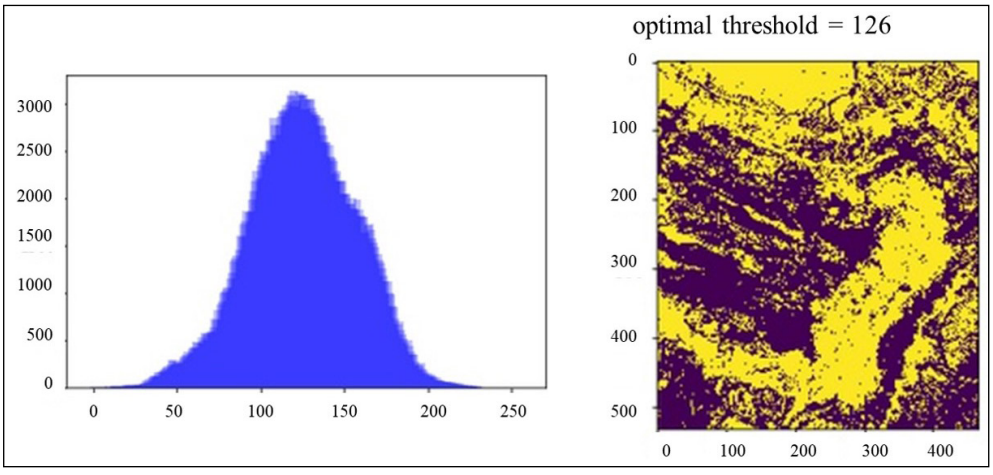


Figure 16- Segmentation by thresholding (RGB CP 1, 2, 3 (11.12.2)).

4.4. The k-Means and Isodata Algorithms

These two k-Means and Isodata algorithms have demonstrated a great capacity for segmentation and geological discrimination. Furthermore, the automatic comparison between the six resulting images below of these two algorithms asserts that each produced unsupervised classification image of the Isodata algorithm is similar to that of K-Means. This comparison takes into consideration a variety of criteria: Layer info, Raster data, map info, info projection, log file, table description, statistics, pyramid layers, XForms (Figures 17 to 22).

We choose in the previous treatment, either by Isodata or K-Means, 18 classes based on the simplified geological map of the Al Glo'a sheet mentioned in Figure 23.

Thanks to Figures 17 to 22 and to the detailed geological map of Al Glo'a (1/50000) we can create geological maps resulting from images processed by Isodata or K-Means. Figure 24 presents a geological map of the unsupervised k-Means RGB 12.8.2 classification with 18 lithological classes, which is characterized by a very good geological discrimination.

K-Means algoritması

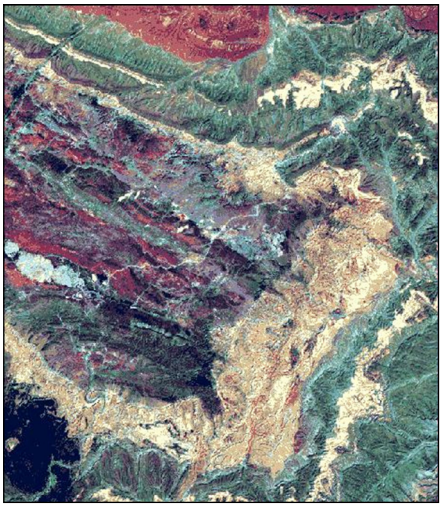


Figure 17- Unsupervised classification by RGB k-Means 12 8 2 (18 classes).

Isodata algoritması

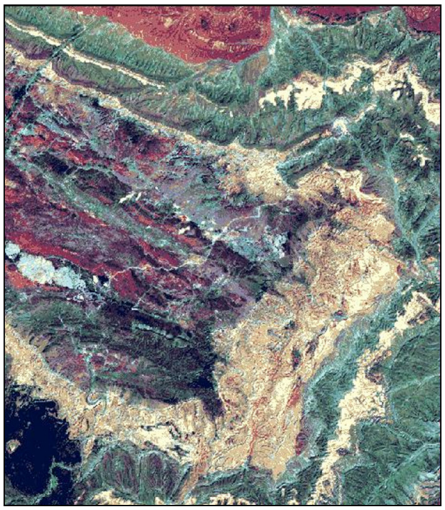


Figure 18- Unsupervised Isodata classification of RGB 12 8 2 (18 classes).

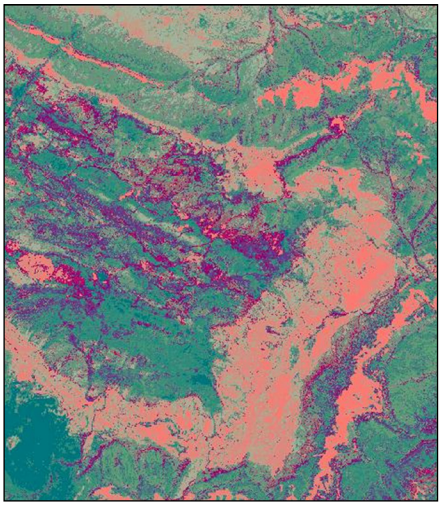


Figure 19- Unsupervised CP RGB 1, 2 and 3 (11 12 2) k-Means Classification (18 classes).

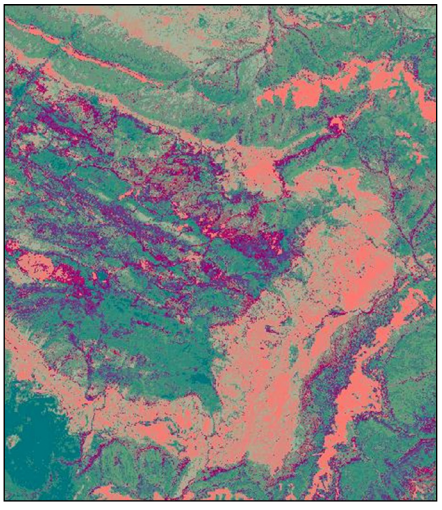


Figure 20- Non-supervised Isodata RGB classification of CP 1, 2 and 3 (11 12 2) (18 classes).

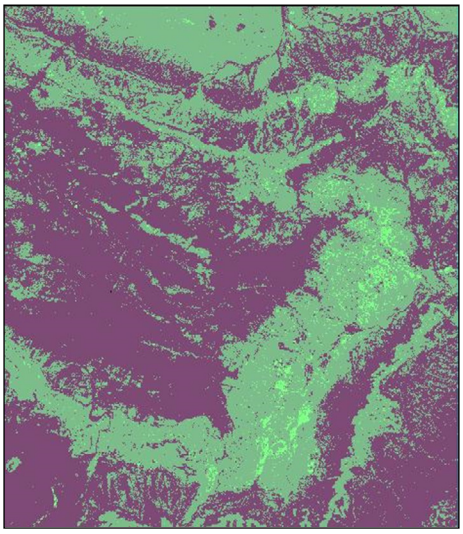


Figure 21- Unsupervised RGB K-Means Classification of Ratios 11/12 11/2 11/8 (18 Classes).

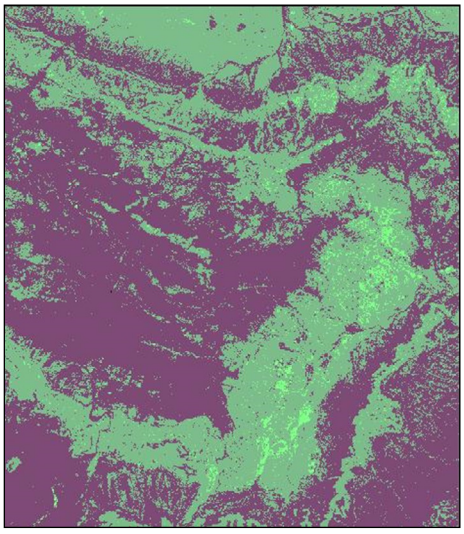


Figure 22-SODATA unsupervised RGB classification of ratios 11/12 11/2 11/8 (18 classes).

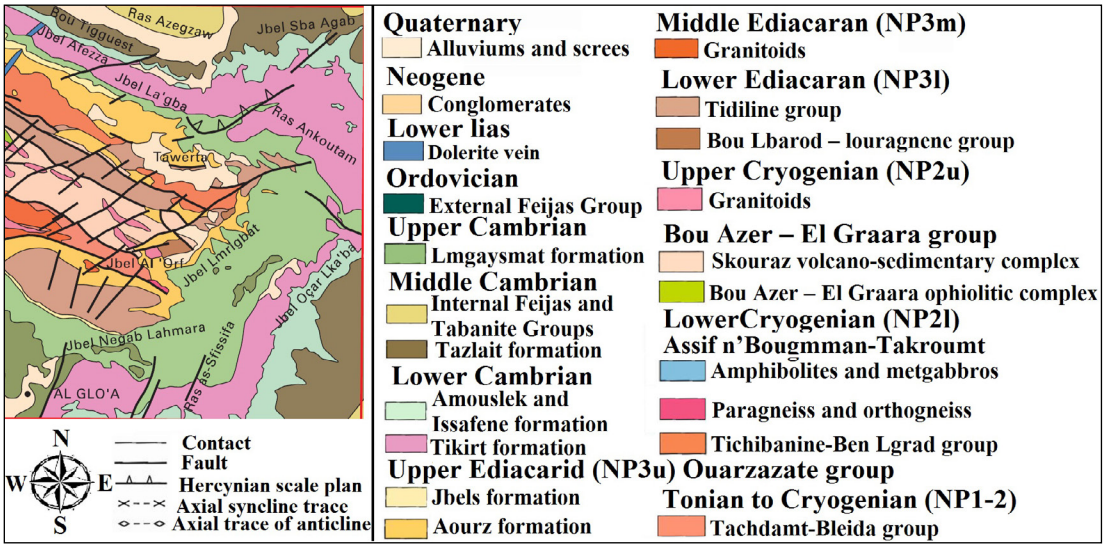


Figure 23- Simplified geological map of the Al Glo'a sheet (1/500000). (edited from Soulaimani, et al., 2013).

5. Results

We can see that threshold segmentation does not give good results in terms of geological discrimination since it divides each image into two parts, especially that the geological map is a variety of geological formations. It is recommended to use the thresholding principle but with several thresholds like the ISODATA type algorithm which is based on the concept of optimal thresholding (we measure the (statistical) separation between two regions from a criterion function).

The contribution of the Watershed algorithm for geological mapping is limited but is of interest from the point of view of topographic and hydraulic studies. It can be very useful in the preparation phase of geological and mining infrastructures.

After knowing: 1) the limited role of segmentation algorithms (the watershed), the effective thresholding and segmentation of graph-based images in this area of mineral exploration; 2) the good potential of unsupervised classification algorithms (Isodata and K-Means), it is recommended to extend the study

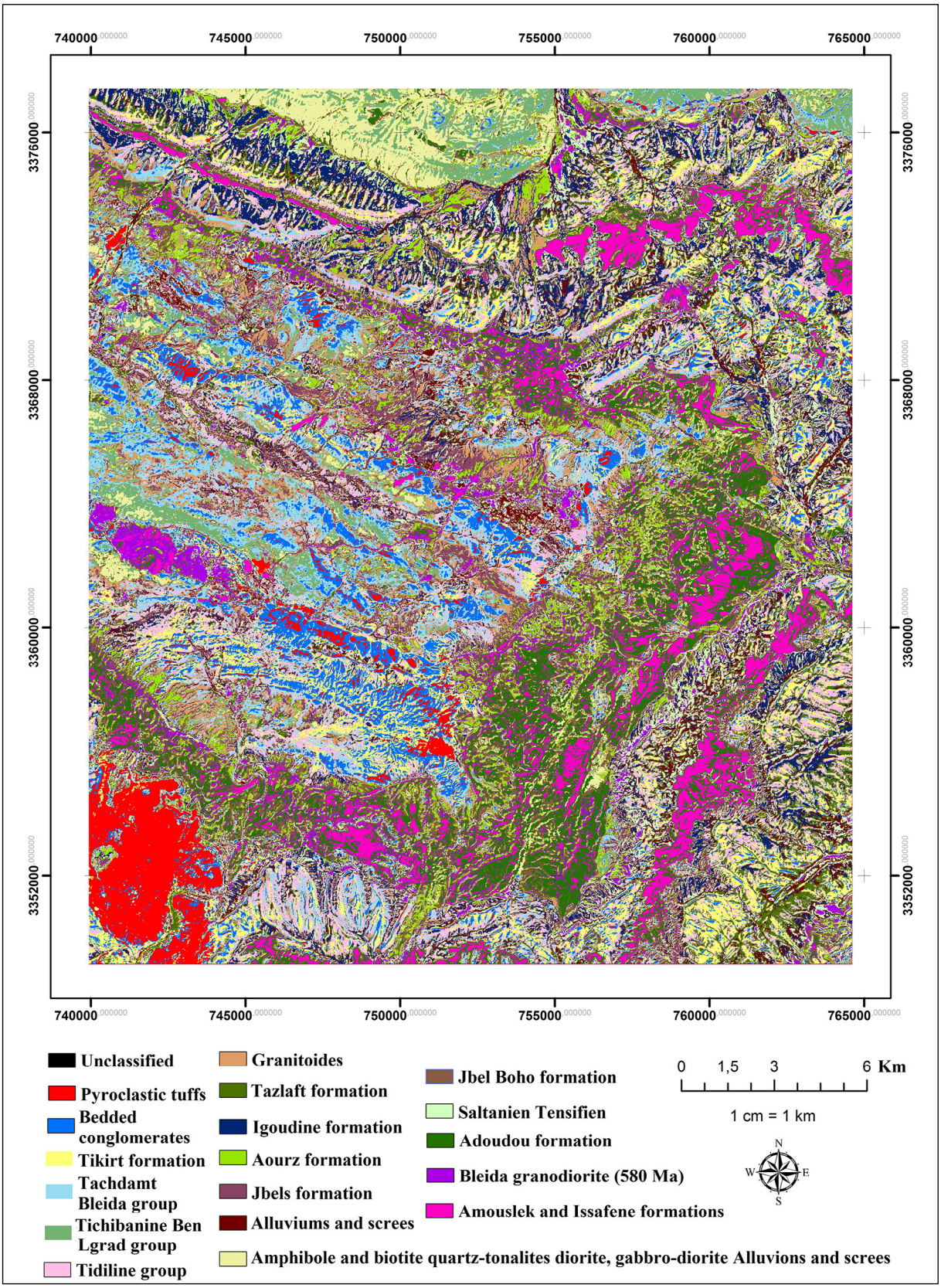


Figure 24- Geological map of the Glo'a for the application of the K-Means to RGB algorithm 12 8 2.

to: 1) supervised classification algorithms (including SVM); 2) deep learning algorithms; and 3) algorithms related to the creation of (structural) lineament maps.

References

Ball, G., Hall, D. 1965. A novel method of data analysis and pattern classification. Rapport technique, Menlo Park, CA, Stanford Research Insitute.

Choubert, G. 1947. L'accident majeur de l'Anti-Atlas. Comptes Rendus de l'Academie des Sciemces 224, 16, 1172-1173.

Cleuziou, G. 2004. Une méthode de classification non-supervisée pour l'apprentissage de règles et la recherche d'information. Université d'Orléans.

Dhodhi, K. M., Saghri, J. A., Ahmad, I., Ul-Mustafa, R. 1999. D-ISODATA: A Distributed Algorithm for Unsupervised Classification of Remotely Sensed Data on Network of Workstations. Journal of Parallel and Distributed Computing, 59(2), 280-301.

Ducrot, D. 2005. Méthodes d'analyse et d'interprétation d'images de télédétection multi-sources : Extraction de caractéristiques du paysage. Habilitation à diriger des recherches, INP TOULOUSE.

Dupas, A. 2009. Opérations et Algorithmes pour la Segmentation Topologique d'Images 3D.

Fahim, A. S. 2006. An efficient enhanced k-means clustering algorithm. J. Zhejiang Univ- Sci, A 7, 1626–1633. https://doi.org/10.1631/jzus.2006.A1626.

Felzenszwalb, P., Huttenlocher, D. 2004. Efficient graph-based image segmentation. International Journal of Computer Vision 59(2).

Huttenlocher, P. F. 2004. Efficient graph-based image segmentation. International Journal of Computer Vision.

Khan, M. 2014. A survey: image segmentation techniques. Lahore. International Journal of Future Computer and Communication, 3.

Maarir, A., Agnaou, I., Bouikhalene, B. 2014. Evaluation de Quelques Méthodes de Segmentation -Application aux Caractères Tifinagh, https://tal.ircam.ma/conference/docs/TICAM2014/1.pdf.

MacQueen, J. 1967. Some methods of classification and analysis of multivariate observations. Berkeley Symposium on Mathematical Statistics and Probability, 81-297.

Maacha, L., Elghorfi, M., Zouhair, M., Sadiqui, O., Soulaimani, A. 2014. Reconsidérations des systèmes métallogéniques de la Boutonnière de Bou Azzer-El Grâara (Anti-Atlas occidental). Maroc.

Mrmint. 2018, 11, 26. algorithme-k-means. Retrieved from mrmint: https://mrmint.fr/algorithme-k-means.

Oukassou, M. 2013. Mouvements verticaux de la bordure nord du craton ouest africain (anti-atlas central, maroc) apport de la thermochronologie basse temperature. thèse de Doctorat, Faculté des Sciences-Aïn Chock, Casablanca.

Pavlidis, S., Horowitz, L., Theodosios. 1976. Picture segmentation by a tree traversal. Journal of ACM, 23 (2), 366-388.

Serra, J. 2006. A lattice approach to image segmentation. Journal of Mathematical Imaging, 24.

Simplilearn. Apprentissage profond. Retrieved from simplilearn:https://www.simplilearn.com/tutorials/machine-learning-tutorial/k-means-clustering-algorithm. 13 July 2020.

Soulaimani, A., Egal, E., Razin, P., Youbi, N., Admou, H., Blein, O., Anzar, C. 2013. Notice explicative de la carte géologique du maroc au 1/50 000 -feuille al glo'a-. Département de lEnergie et des Mines. éditions du service géologique du Maroc.Team, s.-i. d. (n.d.), from scikit-image: http://scikitimage.org/docs/dev/auto_examples/segmentation/plot_marked_watershed.html?fbclid=IwAR2zYJ0VCkLBvsWUJuku-x9sGdzx7pz0W_O5-ho4O5KgPTVfsYl35flDFIc#sphx-glr-auto-examples-segmentation-plot-marked-watershed-py. 29 November 2018.

Bai, M., Urtasun, R. 2017. Deep watershed transform for instance segmentation. Toronto.

Vialard, A. 2018. Segmentation et Analyse d'images (partie 1). LaBRI, Université Bordeaux 1.

Vincent, L., Soille, P. 1991. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. IEEE Transactions on Pattern Analysis and Machine Intelligence 13(6).