

# REMOTELY SENSED IMAGE FAST CLASSIFICATION AND SMART THEMATIC MAP PRODUCTION

Emanuele Alcaras <sup>1</sup>, Pier Paolo Amoroso <sup>1</sup>, Claudio Parente <sup>2\*</sup>, Giuseppina Prezioso <sup>2</sup>

<sup>1</sup> International PhD Programme “Environment, Resources and Sustainable Development”, Department of Science and Technology, Parthenope University of Naples, Centro Direzionale, Isola C4, (80143) Naples, Italy - emanuele.alcaras@studenti.uniparthenope.it; pierpaolo.amoroso@studenti.uniparthenope.it

<sup>2</sup> DIST - Department of Science and Technology, Parthenope University of Naples, Centro Direzionale, Isola C4, (80143) Naples, Italy - claudio.parente@uniparthenope.it; pina.prezioso@uniparthenope.it

**KEY WORDS:** Thematic Map, Landsat 8 OLI, GIS, Image Classification, Smart Applications, Vegetation Index.

## ABSTRACT:

Apps available for Smartphone, as well as software for GNSS/GIS devices, permit to easily mapping the localization and shape of an area by acquiring the vertices coordinates of its contour. This option is useful for remote sensing classification, supporting the detection of representative sample sites of a known cover type to use for algorithm training or to test classification results. This article aims to analyse the possibility to produce smart maps from remotely sensed image classification in rapid way: the attention is focalized on different methods that are compared to identify fast and accurate procedure for producing up-to-date and reliable maps. Landsat 8 OLI multispectral images of northern Sicily (Italy) are submitted to various classification algorithms to distinguish water, bare soil and vegetation. The resulting map is useful for many purposes: appropriately inserted in a larger database aimed at representing the situation in a space-time evolutionary scenario, it is suitable whenever you want to capture the variation induced in a scene, e.g. burnt areas identification, vegetated areas definition for tourist-recreational purposes, etc. Particularly, pixel-based classification approaches are preferred, and experiments are carried out using unsupervised (k-means), vegetation index (NDVI, Normalized Difference Vegetation Index), supervised (minimum distance, maximum likelihood) methods. Using test sites, confusion matrix is built for each method, and quality indices are calculated to compare the results. Experiments demonstrate that NDVI submitted to k-means algorithm allows the best performance for distinguishing not only vegetation areas but also water bodies and bare soils. The resulting thematic map is converted for web publishing.

## 1. INTRODUCTION

Remote sensing allows to acquire a lot of information about an object or phenomenon without making physical contact with the object: the use of satellite or aircraft-based sensor technologies lets to detect and classify objects on Earth, including the surface, the atmosphere and the oceans (Thenkabail, 2018).

However, in some situations, to associate remote sensing techniques with in situ observations is useful. In the classification phase, for example, survey operations are suitable to identify sample areas representative of a known type of coverage to be used for training the algorithms (training sites) or to test the results of the classification (test sites). In situ sampling data as input is required in order to add value to physical imaging remote sensing observations, e.g. in the case of forest health assessment which needs hyperlinks with biotic and abiotic factors (Pause et al., 2016). The mapping of sample areas, i.e. the determination of their shape and spatial localization, can be carried out by acquiring the coordinates of the vertices of its boundary. Among the various methods, these coordinates can be obtained with GPS survey; this can be done in different ways, i.e. static, kinematic, Real Time Kinematic (RTK) and Network Real Time Kinematic (NRTK), and with various types of receivers, from geodetic to low cost ones, characterized by different levels of accuracy.

The advent of modern smartphones with the Android operating system has significantly changed this scenario. In fact, these mobile devices, with their ability to receive GNSS signals, can currently be considered as real positioning devices (Tomaštik et al., 2021).

Nowadays smartphones include sensors that allows them to compute the user position. Inside the device there are different

sensors that are useful for the scope, i.e. accelerometers, gyroscopes and magnetometers, teslameter, proximity sensors, barometer, and GPS/GNSS chipset (Dabove et al., 2017). Smartphones are considered the most widespread device that provides the position of the user within few meters accuracy in Line of Site (LOS) using a low-cost GPS receiver, and indoor using low-cost Inertial Measurement Unit (IMU) based on Micro Electro-Mechanical System (MEMS) containing three accelerometers and three gyroscopes (Al-Hamad and El-Sheimy, 2014). Smartphones also include high resolution digital camera to capture multiple images that are suitable for generating point clouds and 3D models. Since the availability of different sensors in one device, smartphones are considered competitive with the traditional mobile mapping platforms. Regarding this aspect, the main advantage of a smartphone is the cost-effectiveness, the biggest disadvantage is related to the accuracy of the acquired data that can ensure the devices that compose them. Different studies are available in literature on the possibility to enhance the position accuracy acquired by smartphone (Robustelli et al., 2019) as well as the Exterior Orientation Parameters of the smartphone digital camera (Alsubaie et al., 2017).

Using smartphone to test results of remotely sensed image classification requires to display on it the produced thematic map. In addition, global internet further renders billions of ordinary people loyal users of maps and geographic information (Liu and Li, 2019). The fast diffusion of smartphones and their rapidly increased use for different scopes, i.e. navigation, tourism, exploration, etc., require smart maps designed to provide the users with fast access to information. Smart maps can be viewed on smartphones as well as on tablets and PC, making them useful for most situations. Smart maps covering

\* Corresponding Author

different themes can be produced also using remotely sensed images for updating the information and representing the dynamic nature of the surrounding world.

This paper aims to analyse the possibility to produce smart maps in rapid way from remotely sensed images: Landsat 8 OLI satellite data are considered for promptly mapping at medium scale land cover distinguishing the three principal components of the earth surface: water, soil and vegetation.

The article is organized as follow: section 2 presents an introduction of study area and dataset adopted for this research; section 3 describes the methods applied in our experiments in order to classify the images and develop a smart thematic map; section 4 provides the results of the classification and the discussion in relation to them; section 5 reports an overview of the outcomes and further considerations.

## 2. STUDY AREA AND DATASET

The study area chosen for this article covers a surface of 5,400 Km<sup>2</sup> (60 Km x 90 Km) located in the north of Sicily (Italy) as shown in Figure 1.



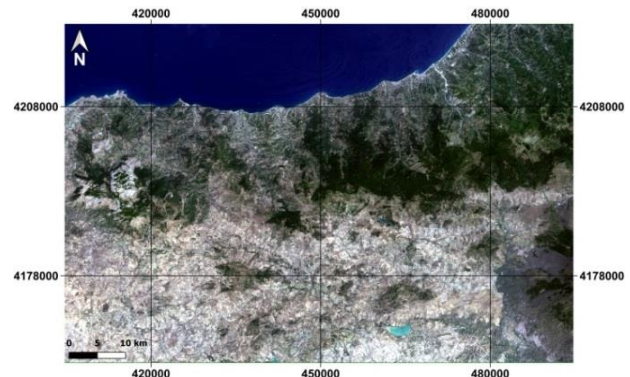
**Figure 1.** Geolocalization of the study area in the north of Sicily (red rectangle), in equirectangular projection and WGS84 geographic coordinates.

The considered area comprises many of the main characteristics of the Island in terms of land cover. In fact, in the south-eastern area is the west hillside of the Etna volcano, which has both dense and scattered vegetation and bare soil covered with lava (Harris et al., 2011). The south/south-western zone of the study area has variety of coverage, in fact there are small wooded areas, cultivated areas that pass from a good state of health in the easternmost areas to senescence (or probably burned) in the western part, also presenting arid areas, and finally, pools of water of natural and artificial origin (Giordano et al., 2002). The northern part collects the stretch of coast that connects the cities of Cefalù and Capo D'Orlando and that overlooks the Tyrrhenian Sea. Finally, the area just south of the coast has dense vegetation, especially in the eastern part.

The study area extends between the following UTM-WGS84 (zone 33 N) plane coordinates:  $E_1 = 404,715$  m;  $E_2 = 494,715$  m;  $N_1 = 4,162,605$  m;  $N_2 = 4,222,605$  m, and it is shown in Figure 2, in RGB true colour composition of the Landsat 8 OLI imagery used for this article.

The Landsat 8 dataset used for this study was acquired on 01/08/2020 at 09:42 GMT. The file was distributed as a L1TP (Level 1 Terrain Product) image data, which is radiometrically

and geometrically corrected by using a Digital Elevation Model to correct parallax errors due to local topographic relief (U. S. Geological Survey, 2020). A clip containing 3000 x 2000 pixels (90 x 60 Km) has been considered for this analysis.



**Figure 2.** RGB true colour composition of the Landsat 8 OLI images, in UTM-WGS84 plane coordinate.

Various studies support the use of Landsat 8 OLI for land cover classification (Jia et al., 2014) (Li et al., 2015) (Deng et al., 2019) making them one of the most adopted in this field also for the possibility of downloading them for free.

Born from a collaboration between NASA and the United States Geological Survey (USGS), Landsat 8 satellite carries the Operational Land Imager (OLI), and the Thermal Infrared Sensor (TIRS) (U. S. Geological Survey, 2013).

Landsat 8 OLI imagery comprises 8 multispectral bands (MS) and one panchromatic band (PAN), as reported in Table 1.

Bands	Wavelength ( $\mu\text{m}$ )	Resolution (m)
Band 1 - Coastal	0.435 - 0.451	30
Band 2 - Blue	0.452 - 0.512	30
Band 3 - Green	0.533 - 0.590	30
Band 4 - Red	0.636 - 0.673	30
Band 5 - Near Infrared	0.851 - 0.879	30
Band 6 - SWIR 1	1.566 - 1.651	30
Band 7 - SWIR 2	2.107 - 2.294	30
Band 8 - Panchromatic	0.503 - 0.676	15
Band 9 - Cirrus	1.363 - 1.384	30

**Table 1.** Main characteristics of Landsat 8 OLI images.

## 3. METHODS

As mentioned earlier, the purpose of this article is to produce a smart thematic map in rapid way from Landsat OLI imagery and to compare different classification methods to provide this map. Here are the steps followed in this study for the production of a smart map:

- Conversion of BV values into reflectance at the top of the atmosphere (TOA) and atmospheric correction;
- Classification of images through the use of different methods of classification (Unsupervised classification, supervised classification Normalized Difference Vegetation Index (NDVI));
- Production of thematic maps;
- Evaluation of the level of thematic accuracy of the produced maps and choice of the most accurate thematic map;
- Production of the Smart Map for web usage.

All experiments are carried out using QGIS 3.16.5, a free and open-source software that includes also SAGA-GIS tools and QGIS2Web plugin for web map production.

### 3.1 Top Of Atmosphere reflectance (TOA) and atmospheric corrections

Landsat Collections namely “Level-1 data products” consist of quantized and calibrated scaled Digital Numbers (DN) provided in 16-bit unsigned integer format. Initial DNs need to be transformed in TOA (Top of Atmosphere) reflectance, the solar radiation incident on the satellite sensor, using formula by USGS (U. S. Geological Survey, 2015) that requires specific parameters comprised in the metadata file included with the dataset (i.e. Band-specific multiplicative rescaling factor, Band-specific additive rescaling factor, Local sun elevation angle). Remotely sensed images at the TOA do not represent the effective reflectance at the Earth surface, also called Bottom-of-Atmosphere (BOA) reflectance, since they include the effect of the interactions of the electro-magnetic radiations (EMR) with the atmosphere that generate scattering, absorption and emission (Campbell and Wynne, 2011). In other terms, what a sensor “sees” (i.e. records) from a satellite is an image disturbed by the atmosphere: satellite sensors measure the radiance in the upper layer of the atmosphere, here the radiance values are different than those measured on the earth's surface (White et al., 2004). Various approaches are available in literature for atmospheric corrections of TOA images to obtain BOA, including empirical methods (Wang and Gastellu-Etchegorry, 2021), such as the dark-object subtraction (DOS) model (Chavez, 1988) and the COST model (Chavez, 1996). In this study TOA is calculated and DOS model applied for atmospheric corrections.

### 3.2 Unsupervised Classification

Classifying the myriad features in satellite images into meaningful categories or classes is the primary purpose of the remote sensing: the image becomes a thematic map capable to represent the spatial variation of one or a small number of geographic distributions e.g. land use, vegetation types, geology, etc. In a broad sense, image classification is defined as the process of categorizing all pixels in an image or raw remotely sensed satellite data to obtain a given set of labels or land cover themes (Lillesand and Keifer, 1994). There are different methods for remote sensing image classification, such as supervised and unsupervised techniques. Generally, they show different levels of accuracy after accuracy assessment was conducted (Hasmadi et al., 2009).

In unsupervised classification the assignment of classes to pixels is carried out without the necessity for external data but completely automatically (Al-Doski et al., 2013). The first step is to group pixels into “clusters” based on their properties; the second step is to manually assign land cover classes to clusters by interpreting the results (cluster analysis) (Chaovalit and Zhou, 2005). Different methods for unsupervised classification are described in literature and available in GIS and Remote sensing software. In the next subsection the main characteristics of K-means are described since it is the unsupervised classification method applied in this study using SAGA-GIS tool.

#### 3.2.1 K-Means

K-means is a numerical, unsupervised, non-deterministic, iterative method usable for image classification (Na et al., 2010). This algorithm divides dataset into K number of clusters by standard Euclidean distance (Kaur et al., 2010); it is an iterative method that needs to cyclically define the centre of the clusters (Xie et al., 2011). In fact, the K-means algorithm detects  $k$  number of centroids, and then assigns every data point to the nearest cluster (Kanungo et al., 2002). The *means* in the

K-means refers to averaging of the data; that is, finding the centroid (Hora et al., 2020)

Supposing that the target object is a  $n$ -dimensional real vector  $x$ ,  $\mu_i$  indicates the mean vector of cluster  $C_i$ , the objective is to find the minimum of the function:

$$E = \sum_{i=1}^k \sum_{x \in C_i} |x - \mu_i|^2 \quad (1)$$

The distance between  $x$  and  $\mu_i$  is Euclidean distance. Being  $n$  the number of the  $j$ -components of  $x$  (and  $\mu$ ), named  $\mu^*$  the mean value of the Cluster  $C^*$ , the Euclidean distance  $d(x, \mu^*)$  can be obtained using the formula (Nazeer and Sebastian, 2009):

$$d(x, \mu^*) = \left[ \sum_{j=1}^n (x_j - \mu_j^*)^2 \right]^{\frac{1}{2}} \quad (2)$$

The process of k-means algorithm can be summarised as reported below (Fahim et al., 2006).

Input:

Number of chosen clusters,  $k$ , and a database  $X = \{x_1, x_2, \dots, x_N\}$  containing  $N$  data objects.

Output:

A set of  $k$  clusters.

Steps:

- 1) Casually select  $k$  data objects from dataset  $X$  as initial cluster centres  $\mu_i$ .
- 2) Calculate the distance between each data object  $x_p$  ( $1 \leq p \leq N$ ) and all  $k$  cluster centres  $\mu_i$  ( $1 \leq i \leq k$ ) and assign data object  $x_p$  to the nearest cluster.
- 3) For each cluster  $i$  ( $1 \leq i \leq k$ ), recalculate the cluster centre.
- 4) repeat step 2 and 3 until no changing in the centre of clusters

The process stops creating and optimizing clusters when either:

- The centroids have stabilized — there is no change in their values because the clustering has been successful.
- The defined number of iterations has been achieved.

In this approach, classes are determined statistically by assigning pixels to the nearest cluster mean based on all available bands (De Amorim and Mirkin, 2012).

### 3.3 Supervised Classification

Supervised classification is based on the idea of dividing the spectral field into regions associable with land cover classes of interest for a particular application (Richards, 2013). It also needs some prior knowledge of samples of the image classes, named in-sites: the image process requires to develop a statistical characterization of the reflectance for each information class (signature analysis) (Sisodia et al., 2014). The image is then classified by examining the reflectance for each pixel and assigning it to the class for which there is the greatest similarity (Schowengerdt, 2012).

Supervised classifiers require an accurate choice of representative training sites for a correct classification (Liu, 2005). In particular, those should satisfy some constraints such as:

- to describes all the land cover classes;
- to include a sufficient number of significant pixels for class for a certain estimate of the classifier parameters;
- to contain-information that completely describes the variability between the different classes due to the

non-stationary behaviour of the spectral signature (Muñoz-Marí, et al., 2007).

Particularly, in this study for supervised classification training sites are detected by means of visual analysis on RGB composition, as well as consulting higher resolution image acquired in the same period.

Naturally, the quality of a supervised classification depends on the quality of the training sites (Palaniswami et al., 2006). Supervised classification includes various methods: the main characteristics of Minimum Distance (MD) and Maximum likelihood (ML) are resumed in the next two sub-sections since they are applied in this study using SAGA-GIS tools.

### 3.3.1 Minimum distance

In the supervised classification with the Minimum Distance (Wacker and Landgrebe, 1972; Mather and Tso, 2016), a pixel is associated to a certain class based on the calculation of the minimum spectral distance between the measurement vector for the considered pixel and the average vector of each sample (Abburu and Golla, 2015). MD algorithm calculates the Euclidean distance between the two vector and the pixel is then assigned to the class having the minimum spectral distance (Perumal and Bhaskaran, 2010). This algorithm is easy to perform and simple to process (Sabins, 2007).

### 3.3.2 Maximum likelihood

The Maximum Likelihood Classification uses training sites to estimate means and variances of the classes, which are used to evaluate probabilities that a pixel is assigned to a determinate class (Sisodia et al., 2014). This method is based on Bayesian probability theory, and, as reported in literature, it is considered as the most powerful classification methods if accurate training sites are provided (Perumal and Bhaskaran, 2010).

## 3.4 Vegetation Indices

Vegetation indices are used in remote sensing to identify vegetation and its health status. They are generally used for the analysis of productivity characteristics, identification of biomass, identification of vegetation cover, etc. (Huete, 2012). Thanks to the vegetation indices it is possible to obtain thematic maps of the land cover (Lyon et al., 1998).

Among the various vegetation indices, we find the Ratio Vegetation Index (RVI) (Jordan, 1969), widely used for the estimation and monitoring of green biomass. This index is very sensitive to vegetation and has a good correlation with plant biomass. Another index used is the Difference Vegetation Index (DVI) (Richardson and Weigand, 1977), which is very sensitive to soil variations and applicable to monitoring the ecological environment of vegetation. We also have the Perpendicular Vegetation Index (PVI) (Richardson and Weigand, 1977), characterized by a lower sensitivity to atmospheric effects and mainly used for the inversion of the parameters of the surface vegetation. Instead, it is sensitive to the brightness and reflectivity of the ground.

The most used among the methods of this group is undoubtedly the Normalized Difference Vegetation Index (NDVI) (Rouse et al., 1974), to which the following subparagraph is dedicated since it is applied in this study.

### 3.4.1 Normalized Difference Vegetation Index

Even if NDVI is usually adopted for the identification of vegetation, however it allows to easily distinguish also two

other classes, namely bare soil and water. It is calculated as normalized ratio between the red and near infrared bands (Rouse et al., 1974), and for Landsat 8 OLI it can be written as follow:

$$NDVI = \frac{Band\ 5 - Band4}{Band\ 5 + Band4} \quad (3)$$

NDVI values are variable in the range [-1, +1].

Because of its simple formula, in this article NDVI is achieved by means of Raster Calculator available in QGIS (QGIS, Raster Calculator). However, NDVI does not produce a thematic map showing three classes, rather it is a further layer or a synthetic band that needs the application of classification criterion for distinguishing categorically in it pixels of water, bare soil and vegetation (Alcaras et al. 2019).

## 3.5 Thematic accuracy assessment

To establish the thematic accuracy of the obtained results, test sites are chosen, distinct from the training sites used in supervised classification, but, like the latter, representative of the individual classes. The tests enable to know the number of pixels correctly classified in each thematic map. For our applications, test sites are chosen in a similar way of training ones. Considering the test sites, confusion Matrix (Story and Congalton, 1986) is used as the quantitative method to evaluate the image classification accuracy: it is a simple cross-tabulation of the predicted and actual class labels for the selected sites (Foody, 2004). Analyzing only the pixels included in the test sites, a confusion matrix is realized in this study for each method as a table showing correspondence between the classification result and ground truth data which, similarly to what was done for the training sites, are derived from the visual analysis of both the pan image and the multispectral images.

The quantitative analysis of the thematic accuracy of each classified image is resumed by the values of the traditional indices named Producer Accuracy (PA), User Accuracy (UA) and Overall Accuracy (OA) (Liu et al., 2007).

## 3.6 Smart Maps production for web application

Operations to publish GIS data to the web and make them accessible to other users are commonly called web mapping (Haklay et al., 2008). The resulting products are also known as “smart maps” since they are suitable for smart applications on tablet and smartphone. We can distinguish smart map from smartphone mapping that is the activity based on smartphone used as a tool for surveying an area: shape and location of different objects are detected and archived by the coordinates of the vertices defining their features. Actually, several mapping applications available for most smartphones allow the user to carry out an accurate survey. However, accuracy levels have only recently gone high. In fact, the first smartphones equipped with GPS were produced in 1999 but, for about twenty years, they have provided the fully processed position without any possibility of modifying the process. Since the raw code and phase measurements were not available, data collected by a smartphone could not be handled by the post-processing software and therefore, under the best conditions, the positioning accuracy could reach three meters, while under adverse conditions due to multipath, it degraded to tens of meters (Robustelli et al., 2019).

For several years, the satellite receivers of most smartphones were single frequencies (L1) receivers and this did not allow to apply neither the most common differencing methods, i.e. double or triple differences, nor to combine different observations (Dabove et al., 2020). In May 2018 appeared the



world's first dual-frequency multi-constellation GNSS smartphone produced by Xiaomi (Xiaomi Mi8), capable of receiving L1 / E1 / and L5 / E5 signals from GPS, Galileo, Beidou and GLONASS satellites (Robustelli et al., 2019). The availability of two frequencies allows to increase the accuracy and reliability of positioning by reducing the multipath error and eliminating the influence of ionospheric refraction, this has paved the way for the application of measurement and calculation methods previously reserved to high-precision geodetic applications (Skorupa, 2019).

The use of smartphones with dual frequency receivers and special apps allows you to acquire, with high levels of accuracy, data for mapping as well as to support remote sensing for the identification of training sites and test sites. However, the production of the map or its update are carried out on a dedicated device, typically PC, by using software ad hoc, specifically web mapping enables the user to transform this map into a product available online, even for smartphones and tablets. Creating a web map is a very different process than creating one in a GIS and requires specific competence as web programmers; nevertheless, there are tools to easily translate your work in GIS to web maps (Gandhi, 2019).

One of these tools is QGIS2Web plugin which is useful to create an interactive web map by turning QGIS layers into HTML, JavaScript, and CSS (Cascading Style Sheets) files that describe how elements are displayed on the screen files.

Starting from the selected thematic map, QGIS2Web generates an index.html file and three folders with associated Java-script and CSS files. The resulting map can be published online using a web hosting account and just uploading the *index.html* file and the three associated folders to the main public html directory: the map is shown in the browser every time the website is accessed (Upadhyay, 2018).

In this study we convert the produced thematic maps in the above mentioned format files to render them suitable for web publishing. In this way we can use them also for testing the achieved level of thematic accuracy in the case of in situ observation.

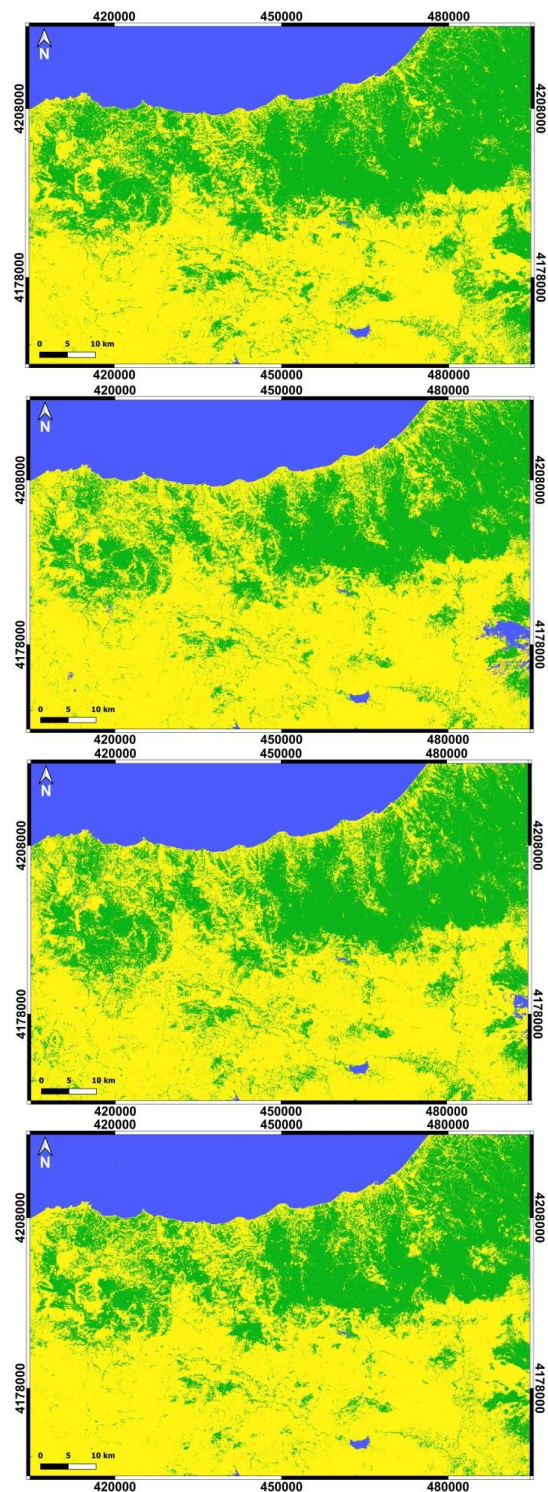
#### 4. RESULTS AND DISCUSSION

In this study, Minimum Distance, Maximum Likelihood and K-means classification methods are applied to all Landsat 8 OLI images or to a selection of them as well as to NDVI; particularly the following combination are used:

- NDVI with Minimum Distance (NDVI MD);
- NDVI with Maximum Likelihood (NDVI ML);
- NDVI with K-Means (NDVI KM);
- RED & NIR with Minimum Distance (RN MD);
- RED & NIR with Maximum Likelihood (RN ML);
- RED & NIR with K-Means (RN KM);
- RGB & NIR with Minimum Distance (RGBN MD);
- RGB & NIR with Maximum Likelihood (RGBN ML);
- RGB & NIR with K-Means (RGBN KM);
- All 8 MS bands with Minimum Distance (8MS MD);
- All 8 MS bands with Maximum Likelihood (8MS ML);
- All 8 MS bands with K-Means (8MS KM).

Supervised classification is applied in SAGA-GIS (SAGA, Module Supervised Classification for Grids) to achieve Minimum Distance and Maximum Likelihood classifications, as well as Unsupervised classification to achieve K-Means clustering (SAGA, Tool K-Means clustering for Grids). The outputs are classified in three classes: water, bare soil and vegetation.

A visual comparison of some classification results, specifically those obtained from NDVI KM, RN KM, RGBN MD and 8MS ML, is shown in Figure 3.



**Figure 3.** Thematic Maps in 3 classes (water, soil, vegetation), obtained by (from higher to lower) NDVI KM, RN KM, RGBN MD and 8MS ML.

The values obtained for the thematic accuracy indices (PA, UA and OA) are shown in Table 2.

Methods	Classes	Accuracy Indices		
		PA	UA	OA
NDVI MD	Water	1	1	0.9923
	Soil	1	0.9774	
	Vegetation	0.9769	1	
NDVI ML	Water	1	0.9997	0.9903
	Soil	0.9997	0.9720	
	Vegetation	0.9711	1	
NDVI KM	Water	1	1	<b>0.9962</b>
	Soil	1	0.9887	
	Vegetation	0.9886	1	
RN MD	Water	1	0.9898	0.8990
	Soil	0.8186	0.8711	
	Vegetation	0.8783	0.8365	
RN ML	Water	0.9903	1	0.9830
	Soil	1	0.9516	
	Vegetation	0.9589	1	
RN KM	Water	1	0.6772	0.8134
	Soil	0.5244	0.8628	
	Vegetation	0.9157	1	
RGBN MD	Water	0.9969	0.9850	0.9439
	Soil	0.9392	0.8979	
	Vegetation	0.8957	0.9509	
RGBN ML	Water	0.9997	1	0.9830
	Soil	1	0.9514	
	Vegetation	0.9491	1	
RGBN KM	Water	1	0.6621	0.8291
	Soil	0.4913	0.9948	
	Vegetation	0.996	1	
8MS MD	Water	1	0.9966	0.9870
	Soil	1	0.9655	
	Vegetation	0.9609	1	
8MS ML	Water	0.9937	1	0.9907
	Soil	1	0.9728	
	Vegetation	0.9783	1	
8MS KM	Water	1	0.9867	0.8269
	Soil	0.4847	0.9953	
	Vegetation	0.996	0.6641	

**Table 2.** Thematic accuracy indices applied to the classification methods.

For each of these indices, which values are comprising between 0 and 1, the closer is the value to 1, the better the result.

NDVI KM is the most performing method, providing the highest OA value (0.9962) and optimal PA and UA values, since they are equal to 1, except in the case of vegetation PA (0.9886) and soil UA (0.9887).

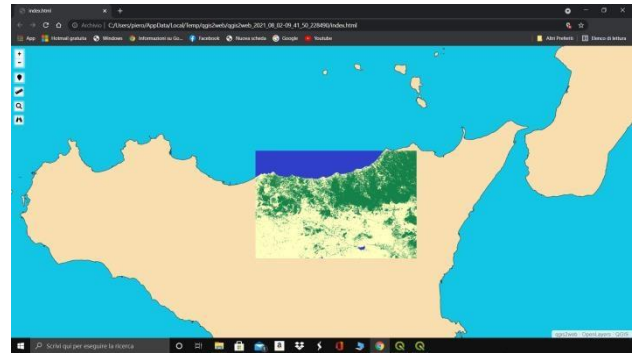
The high level of information of this vegetation index is testified also by the results obtained with NDVI MD and NDVI ML. The good performance of k-means is unique in combination with NDVI since in the case of RN, RGBN and 8MS this clustering algorithm provides the poorest outputs, while maximum likelihood provides the best ones.

Furthermore, we want to highlight that RN KM provides the worst output in terms of OA (0.8134) and RN MD is the only minimum distance application who goes under the value of 0.9. These results permit to make another consideration: even if NDVI is obtained by a combination of the Red and NIR bands, the results provided by it are totally different from those provided by classification applied directly to Red and NIR. This is even more evident when comparing the results obtained by

applying K-Means: NDVI KM and RN KM are respectively the best and the worst results and both are not influenced by human involvement in taking into consideration the training sites.

It also seems that the more bands are used to classify, with the same method, the better are the achieved results, with the exception of NDVI that is a powerful synthetic band derived by two highly unrelated bands.

The resulting thematic map from NDVI KM in web format, produced by using QGIS2Web plugin, is shown in Figure 3:



**Figure 4.** Visualization of thematic map from NDVI KM in web format in overlap with the Italian Region vector file.

## 5. CONCLUSIONS

This study testifies that web thematic maps focused to distinguish the earth surface in water, bare soil and vegetation can be easily produced in automatic way using few MS bands of satellite images, specifically only two bands of Landsat 8 OLI imagery, which are Red and NIR used for NDVI production.

Our experiments demonstrate that the use of k-means, which does not require the human supervision, allows to disaggregate NDVI results and quickly assign labels to the three named classes keeping very high level of thematic accuracy. These results seem to be in contradiction with the fact that supervised classification methods are usually better performing than unsupervised classification ones. Nevertheless, unsupervised classification is useful for distinguish few, uncomplicated and wide-ranging land cover classes such as water, bare soil and vegetation. Furthermore, unsupervised classification may reduce the operator errors in training sites identification. Finally, k-means applied to the NDVI layer that is capable to highlight the different reflectance of the water, bare soil and vegetation in the Red and NIR bands, allow to assign the class label to the pixel in a right way.

The publication on web of the produced thematic map, a difficult task that usually requires specific skills different from those that characterize the traditional map maker, is possible in rapid way by using specific tools such as QGIS2Web plugin.

## REFERENCES

- Abburu, S., Golla, S.B., 2015. Satellite image classification methods and techniques: A review. *International journal of computer applications*, 119(8). doi.org/10.5120/21088-3779.
- Al-Doski, J., Mansori, S. B., Shafri, H. Z. M., 2013. Image classification in remote sensing. *Journal of Environment and Earth Science*, 3(10), 141-148. doi.org/10.1007/s10661-020-08612-8.
- Al-Hamad, A. El-Sheimy, N., 2014. Smartphones based mobile mapping systems. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* doi.org/doi.org/10.5194/isprsarchives-XL-5-29-2014.

- Alcaras, E., Errico, A., Falchi, U., Parente, C., Vallario, A. 2019. Coastline Extraction from Optical Satellite Imagery and Accuracy Evaluation. In: *International Workshop on R3 in Geomatics: Research, Results and Review* 336-349. Springer, Cham. doi.org/10.1007/978-3-030-62800-0\_26.
- Alsubaie, N. M., Youssef, A. A., El-Sheimy, N., 2017. Improving the accuracy of direct geo-referencing of smartphone-based mobile mapping systems using relative orientation and scene geometric constraints. *Sensors*, 17(10), 2237. doi.org/10.3390/s17102237.
- Campbell, J. B., Wynne, R. H. 2011: *Introduction to remote sensing*. Guilford Press, New York, USA.
- Chaovalit, P., Zhou, L., 2005. Movie review mining: A comparison between supervised and unsupervised classification approaches. In: *Proceedings of the 38th annual Hawaii international conference on system sciences*, 112c-112c), IEEE. doi.org/10.1109/HICSS.2005.445.
- Chavez, P.S., Jr., 1988. An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data. *Remote Sensing of Environment*, 24 (3), 459-479. doi.org/10.1016/0034-4257(88)90019-3.
- Chavez, P.S., Jr., 1996. Image-based atmospheric corrections—revisited and improved. *Photogrammetric Engineering and Remote Sensing*, 62(9), 1025-1036.
- Dabove, P., Di Pietra, V., Lingua, A. M., 2017. Positioning techniques with smartphone technology: Performances and methodologies in outdoor and indoor scenarios. In: *Smartphones from an Applied Research Perspective, InTech*, 163-183. dx.doi.org/10.5772/intechopen.69679.
- Dabove, P., Di Pietra, V., Piras, M., 2020. GNSS Positioning Using Mobile Devices with the Android Operating System. *ISPRS International Journal of Geo-Information*, 9(4), 220. doi.org/10.3390/ijgi9040220.
- De Amorim, R.C., Mirkin, B., 2012. Minkowski metric, feature weighting and anomalous cluster initializing in K-Means clustering. *Pattern Recognition*, 45(3), 1061-1075. doi.org/10.1016/j.patcog.2011.08.012.
- Deng, Z., Zhu, X., He, Q., Tang, L., 2019. Land use/land cover classification using time series Landsat 8 images in a heavily urbanized area. *Advances in Space Research*, 63(7), 2144-2154. doi.org/10.1016/j.asr.2018.12.005.
- Fahim, A.M., Salem A.M., Torkey, F.A., 2006. An efficient enhanced k-means clustering algorithm. *Journal of Zhejiang University Science A*, 10, 1626-1633. doi.org/10.1631/jzus.2006.A1626.
- Foody, G.M., 2004. Thematic map comparison. *Photogrammetric Engineering and Remote Sensing*, 70(5), 627-633. doi.org/10.14358/PERS.70.5.627.
- Gandhi, U., 2019. Web Mapping with QGIS2Web (QGIS3), QGIS Tutorials and Tips, Web Mapping with QGIS2Web — QGIS Tutorials and Tips. [qgistutorials.com/en/docs/3/web\\_mapping\\_with\\_qgis2web.html](http://qgistutorials.com/en/docs/3/web_mapping_with_qgis2web.html) (21 July 2021).
- Giordano, L., Giordano, F., Grauso, S., Iannetta, M., Sciortino, M., Rossi, L., Bonati, G., 2002. Identification of areas sensitive to desertification in Sicily Region. ENEA, Centro Ricerche Casaccia, Via Anguillarese, 301, 00060.
- Haklay, M., Singleton, A., Parker, C., 2008. Web mapping 2.0: The neogeography of the GeoWeb. *Geography Compass*, 2(6), 2011-2039. doi.org/10.1111/j.1749-8198.2008.00167.x
- Harris, A. J., Favalli, M., Wright, R., Garbeil, H., 2011. Hazard assessment at Mount Etna using a hybrid lava flow inundation model and satellite-based land classification. *Natural Hazards*, 58(3), 1001-1027. doi.org/10.1007/s11069-010-9709-0
- Hasmadi, M., Pakhriazad, H. Z., Shahrin, M. F., 2009. Evaluating supervised and unsupervised techniques for land cover mapping using remote sensing data. *Geografika: Malaysian Journal of Society and Space*, 5(1), 1-10.
- Hora, A., Bari, A., Rawat, S., 2020. Machine Learning Approaches to Uncover Terrorism Network in India. In: *2020 International Conference for Emerging Technology (INCET)*, 1-6, IEEE. doi.org/10.1109/INCET49848.2020.9154015
- Huete, A.R., 2012. Vegetation indices, remote sensing and forest monitoring. *Geography Compass*, 6(9), 513-532. doi.org/10.1111/j.1749-8198.2012.00507.x
- Jia, K., Wei, X., Gu, X., Yao, Y., Xie, X., Li, B., 2014. Land cover classification using Landsat 8 operational land imager data in Beijing, China. *Geocarto international*, 29(8), 941-951. doi.org/10.1080/10106049.2014.894586
- Jordan, C.F., 1969. Derivation of leaf-area index from quality of light on the forest floor. *Ecology*, 50(4), 663–666.
- Kanungo, T., Mount, D.M., Netanyahu, N.S., Piatko, C.D., Silverman, R., Wu, A.Y., 2002. An efficient k-means clustering algorithm: Analysis and implementation. *IEEE transactions on pattern analysis and machine intelligence*, 24(7), 881-892. doi.org/10.1109/TPAMI.2002.1017616
- Kaur, D., Kaur, A., Gulati, S., Aggarwal, M., 2010. A clustering algorithm for software fault prediction. In: *International Conference on Computer and Communication Technology (ICCCCT)*, IEEE, 603-607. doi.org/10.1109/ICCCCT.2010.5640474
- Li, E., Du, P., Samat, A., Xia, J., Che, M., 2015. An automatic approach for urban land-cover classification from Landsat-8 OLI data. *International Journal of Remote Sensing*, 36(24), 5983-6007. doi.org/10.1080/01431161.2015.1109726
- Lillesand, T. M., Keifer, R. W., 1994: *Remote Sensing and Image Interpretation*. John Wiley & Sons. USA.
- Liu, C., Frazier, P., Kumar, L., 2007. Comparative assessment of the measures of thematic classification accuracy. *Remote sensing of environment*, 107(4), 606-616. doi.org/10.1016/j.rse.2006.10.010
- Liu, X., 2005. Supervised classification and unsupervised classification. ATS 670 Class Project.
- Liu, X. L., Li, C. M., 2019. Consideration and exploration on the demand changes of surveying and mapping geographic information. *Int. Arch. Photogramm. Remote Sens. Spatial Inf.*

- Sci., XLII-4/W16, 375–378. doi.org/10.5194/isprs-archives-XLII-4-W16-375-2019
- Lyon, J.G., Yuan, D., Lunetta, R.S., Elvidge, C.D., 1998. A change detection experiment using vegetation indices. *Photogrammetric engineering and remote sensing*, 64(2), 143.
- Mather, P., Tso, B., 2016: *Classification methods for remotely sensed data*. CRC press.
- Muñoz-Marí, J., Bruzzone, L., Camps-Valls, G., 2007. A support vector domain description approach to supervised classification of remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 45(8), 2683-2692. doi.org/10.1109/TGRS.2007.897425
- Na, S., Xumin, L., Yong, G., 2010. Research on k-means clustering algorithm: An improved k-means clustering algorithm. In: *Third International Symposium on intelligent information technology and security informatics*, 63-67, IEEE. doi.org/10.1109/IITSI.2010.74
- Nazeer, K.A.A., Sebastian, M.P., 2009. Improving the Accuracy and Efficiency of the k-means Clustering Algorithm. In: *Proceeding of the World Congress on Engineering*, 1, London, July 2009.
- Palaniswami, C., Upadhyay, A.K., Maheswarappa, H.P., 2006. Spectral mixture analysis for subpixel classification of coconut. *Current Science*, 91(12), 1706 -1711.
- Pause, M., Schweitzer, C., Rosenthal, M., Keuck, V., Bumberger, J., Dietrich, P., Heurich, M, Jung, A., Lausch, A. 2016. In situ/remote sensing integration to assess forest health—A review. *Remote Sensing*, 8(6), 471. doi.org/10.3390/rs8060471
- Perumal, K., Bhaskaran, R., 2010. Supervised classification performance of multispectral images. *Journal of computing*, 2(2), 124-129.
- QGIS, QGIS User Guide – Manual (QGIS 3.16), 2021. docs.qgis.org/3.16/en/docs/user\_manual/index.html (31 July 2021).
- Richards J.A., 2013. Supervised Classification Techniques, in: *Remote Sensing Digital Image Analysis*. Springer, Berlin, Heidelberg. doi.org/10.1007/978-3-642-30062-2\_8
- Richardson, A.J., Weigand, C., 1977. Distinguishing vegetation from soil background information, *Photogrammetric Engineering and Remote Sensing*, 43(12), 1541-1552.
- Robustelli, U., Baiocchi, V., Pugliano, G., 2019. Assessment of dual frequency GNSS observations from a Xiaomi Mi 8 Android smartphone and positioning performance analysis. *Electronics*, 8(1), 91. doi.org/10.3390/electronics8010091
- Rouse, J. W., Haas, R. H., Schell, J. A., Deering, D. W., 1974. Monitoring vegetation systems in the Great Plains with ERTS. NASA special publication, 351(1974), 309.
- Sabins, F.F., 2007: *Remote sensing: principles and applications*. Waveland Press.
- Schowengerdt, R. A., 2012: *Techniques for image processing and classifications in remote sensing*. Academic Press.
- Sisodia, P.S., Tiwari, V., Kumar, A., 2014. Analysis of supervised maximum likelihood classification for remote sensing image. In: *International conference on recent advances and innovations in engineering (ICRAIE-2014)*, IEEE, 1-4. doi.org/10.1109/ICRAIE.2014.6909319
- Skorupa, B., 2019. The problem of GNSS positioning with measurements recorded using Android mobile devices. *Budownictwo i Architektura*, 18(3), 51-62. doi.org/10.35784/bud-arch.738
- Story, M., Congalton, R.G., 1986. Accuracy assessment: a user's perspective. *Photogrammetric Engineering and remote sensing*, 52(3), 397-399.
- Thenkabail, P., 2018: *Remote Sensing Handbook-Three Volume Set*. CRC Press.
- Tomašík, J., Chudá, J., Tunák, D., Chudý, F., Kardoš, M., 2021: Advances in smartphone positioning in forests: dual-frequency receivers and raw GNSS data. *Forestry: An International Journal of Forest Research*, 94 (2), 292–310. doi.org/10.1093/forestry/cpaa032
- Upadhyay, A., 2018. QGIS2Web – Create Web Maps, QGIS2Web - Create Web Maps - (igismap.com). igismap.com/qgis2web-create-web-maps/ (21 July 2021).
- U. S. Geological Survey, 2013. Landsat Missions - Landsat 8. usgs.gov/core-science-systems/nli/landsat/landsat-8?qt-science\_support\_page\_related\_con=0#qt-science\_support\_page\_related\_con (12 July 2021).
- U. S. Geological Survey, 2019. Landsat 8 (L8) Data Users Handbook?, Version 5. 0, November 2019. EROS, Sioux Falls, South Dakota, 62-63. usgs.gov/media/files/landsat-8-data-users-handbook (31 July 2021).
- U. S. Geological Survey, 2020. Landsat 8-9 OLI/TIRS Collection 2 Level 1 Data Format Control Book, Version 6.0, September 2020. https://prd-wret.s3.us-west-2.amazonaws.com/assets/palladium/production/atoms/files/LSD S-1822\_Landsat8-9-OLI-TIRS-C2-L1-DFCB-v6.pdf (5 October 2021).
- Wang, Y., Gastellu-Etchegorry, J. P., 2021. Accurate and fast simulation of remote sensing images at top of atmosphere with DART-Lux. *Remote Sensing of Environment*, 256, 112311. doi.org/10.1016/j.rse.2021.112311.
- Wacker, A.G., Landgrebe, D.A., 1972. Minimum distance classification in remote sensing. Technical Reports, Laboratory for Application of Remote Sensing, Purdue University.
- White, H. P., Khurshid, K. S., Hitchcock, R., Neville, R., Sun, L., Champagne, C. M., Staenz, K., 2004. From At-Sensor observation to At-Surface reflectance-calibration steps for earth observation hyperspectral sensors. In: *IGARSS 2004, IEEE International Geoscience and Remote Sensing Symposium*, 5, 3241-3244. doi.org/10.1109/IGARSS.2004.1370392.
- Xie, J., Jiang, S., Xie, W., Gao, X., 2011. An Efficient Global K-means Clustering Algorithm. *J. Comput.*, 6(2), 271-279. doi.org/10.4304/jcp.6.2.271-279.