LULC Classification Performance of Supervised and Unsupervised Algorithms on UAV-Orthomosaics

Mirko Saponaro[1,2] and Eufemia Tarantino[1(✉)]

[1] Polytechnic University of Bari, Via Orabona 4, 70125 Bari, Italy
eufemia.tarantino@poliba.it
[2] Agenzia Regionale Strategica Per Lo Sviluppo Ecosostenibile del Territorio (ASSET) - Regione Puglia, Via Gentile, 52 - 70126 Bari, Italy

Abstract. Classification algorithms involve automatically categorising the response functions recorded in any image, i.e., the light reflected and recorded in the pixels by the sensors, as "virtual" representations of real-world objects in classes. These algorithms are finding increasing use in Remote Sensing (RS) applications and the scientific community is deeply engaged in investigating their performance. The substantial increase in spatial resolution that has been introduced through the onset of Unmanned Aerial Vehicle (UAV) platforms, even if equipped with low-cost sensors, has made classes recognition perform well for even the smallest scenario properties. Despite these advantages, it is emerged the need to address their limitations and to analyse their impacts in the data post-processing stages. It is necessary to validate the processing procedure so as to fully define their comparability and, in some cases, their interchangeability with other RS platforms. The aim of this research is to create a validated pixel-based classification procedure that will subsequently result in an automated method that is simple to apply, especially for end-users with limited knowledge of spatial data processing. Supervised and unsupervised approaches to testing performance were deemed more effective by adopting photogrammetric products based on UAV-acquisitions. This has mainly been accomplished by using statistical classifiers for Land Use/Land Cover (LULC) recognition based on several reflectance values across wavebands that compose an image and vegetation indexes in the visible bands. For these tests, the processing chains concerning classifications with supervised Random Forest (RF) and unsupervised K-Means Clustering algorithms were adopted.

Keywords: UAV · Vegetation extraction · Random forest · K-means · SNAP

1 Introduction

Several key developments achieved in the last decade show enormous potential in generating characterising information, in a radically new way, based merely on images and their pixels [1]. Machine learning from these images essentially translates into automatically categorising pixels based on their values and transforming them into a form of final

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022
O. Gervasi et al. (Eds.): ICCSA 2022 Workshops, LNCS 13379, pp. 311–326, 2022.
https://doi.org/10.1007/978-3-031-10545-6_22

component or group of objects in the raster [2]. In other words, classification involves categorising the response functions recorded in the images, i.e., the light reflected from the Earth's surface and recorded in the pixels by the sensors, as "virtual" representations of real-world objects. Basic machine learning proposes a essentially different methodology by relying on programmed algorithms to classify input image data into an output map based on the statistical reflectance properties of their composite pixels. In the enhanced conceptual viewpoint of machine learning, the data and the requested output are delivered to a learning algorithm, called learner, which then generates the algorithm that converts one into the other. In a more complex and developed stage, Deep Neural Networks (DNNs) are composed of several layers between the input and output layers that collectively characterize the correct mathematical manipulation that generates the output from the input through a series of convolutions [3, 4].

These algorithms are finding increasing use in Remote Sensing (RS) applications and the scientific community is deeply engaged in investigating their performance. Even today, the variability and abundance of natural features represented as raster images and the quite coarse resolution of satellite images yield a challenge to pattern recognition algorithms. However, the significant increase in spatial resolution that has been led through the operation of Unmanned Aerial Vehicle (UAV) platforms at much lower altitudes, reducing the size of ground pixel projections by three orders of magnitude, has made machine learning perform well for even the smallest scenario properties. Even in the case of dynamic environments, i.e., subject to natural or anthropogenic change, machine learning techniques have enhanced the identification of features of interest and also tracked how they evolve over time [5].

In the case of feature detection from RS images, one of the key challenges is to agree on the most appropriate approach to reliably recognise real-world objects from a large number of pixels. So far, this has mainly been accomplished by using statistical classifiers that distinguish features or Land Use/Land Cover (LULC) based on several reflectance values across different wavebands that compose an image or by employing predefined rule sets to classify logically segmented objects from an image [6–9]. A first distinction is therefore made between pixel-based and object-based approaches to image analysis [10]. In a further distinction, a supervised approach is found within which, given an input image and a predefined training set of categories, a detection algorithm can identify all instances of pixels and/or objects that fall into these categories in an image. Secondly, quantitative classification methods applied to digital images include unsupervised pixel-based clustering [11]. In general, unsupervised clustering is a useful first step in digital image examination to reveal patterns and possible distinguishing features for purpose in more structured classification methods. Supervised classification delivers higher overall classification accuracies, commonly never less than 55%. Obviously, variations in flight altitude, illumination, shadows, partial constrictions, low sun angles and weather variability will decrease image quality and thus both visual interpretation and digital classifier performance [11].

The aim of this research is to test the performance of classification procedures that could subsequently be translated into an automated and easy-to-apply method, especially for end users with limited knowledge of spatial data processing and that is low cost in terms of the tools required.

Given the considerations discussed above, a pixel-based approach to testing performance was deemed more effective by adopting photogrammetric products based on UAV acquisitions.

In most cases, UAVs are supplied with cheap cameras able of acquiring only the visible RGB bands. These consumer cameras often come up with the problem of not being radiometrically calibrated [12, 13]. In order to offer RS data with a quantitative value, it is crucial to calibrate them both geometrically and radiometrically [14].

Despite the advantages of using UAVs and low-cost cameras, it is also necessary to address their limitations and to analyse their impacts in the data post-processing stages [15]. It is necessary to validate the processing procedure so as to fully define their comparability and, in some cases, their interchangeability with other RS platforms [16].

With a view to the democratisation of processes, accessible to end-users with a high degree of comparability, interchangeability and repeatability, the results of pixel-based classifications were tested and analysed. UAV-based orthomosaics in the radiometrically calibrated visible bands and vegetation maps in the Triangular Greenness Index (TGI) and Red–Green Ratio Index (IRG) were exported from QGIS software and uploaded to the open-source Sentinel Application Platform (SNAP) developed by Brockmann Consult, Skywatch, Sensar and CS and provided free of charge by ESA/ESRIN. In this way, Saponaro et al. [17] examined the statistical comebacks regarding the performance of vegetation indices in terms of separability between vegetated and non-vegetated areas. The impact of the most significant indices in the application of pixel-based classification algorithms was presented but a robust analysis between supervised and unsupervised classification was not assessed by the authors. Starting from the results achieved in [17], in this work the processing chains concerning classifications with supervised Random Forest (RF) [18] and unsupervised K-Means Clustering [19] algorithms were implemented.

At the end of each classification procedure, sets of test samples must be considered to verify the accuracy and precision of the classification. Confusion matrices were extracted from which the robustness and accuracy of the process can be deduced.

2 Methods

2.1 Focus Areas and Dataset Acquisition

Two study areas with different context were selected in order to demonstrate the versatility and non-specificity of the processing chain and the results that can be obtained. The areas were surveyed using different UAV sensors/cameras, using a different georeferencing strategy and at different Above Ground Level (AGL) altitudes, thus different Ground Samples Distance (GSD) values. A preview of these areas is presented in Fig. 1. In detail, in case study (a) an extra-urban environment of Grottole in the province of Matera (Italy) was selected (Fig. 1a). In this scenario, high vegetation, bare ground and, above all, a viaduct are visible. In the second case study, on the other hand, an abandoned archaeological area, called Punta Penna because of the promontory over the sea on which it stands, in the city of Bari (Italy), was considered (Fig. 1b). Unlike the first case, the presence of water is assessed in this scenario.

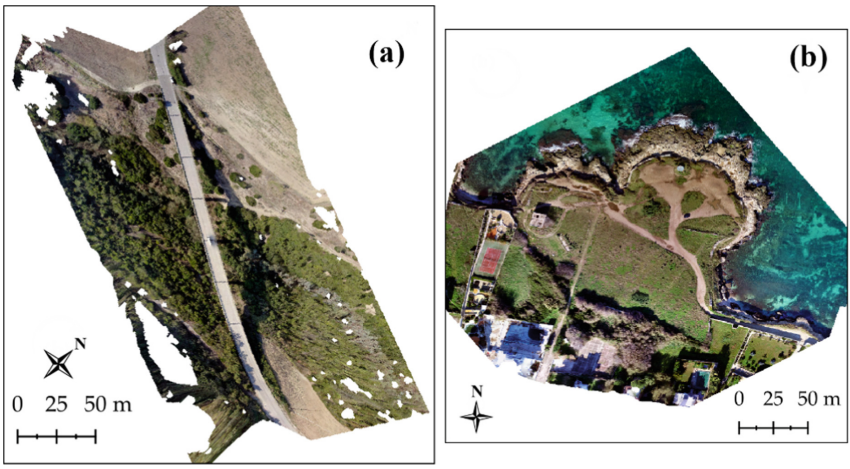

Fig. 1. Case study: (a) an out-of-town viaduct in Grottole (MT), Italy, (b) an abandoned archaeological area in Bari (BA), Italy. The icon to the left of the reader identifies the north orientation of the areas.

Table 1 highlights the assets and technologies used for each dataset.

For the first dataset, a Global Navigation Satellite System (GNSS) survey campaign of 11 Ground Control Points (GCPs) was carried out, acquired in network Real-Time Kinematic (nRTK) mode with an average accuracy of 2 cm along the three axes, in order to accomplish an Indirect Georeferencing (IG) of the photogrammetric products. For the second case study, a Direct Georeferencing (DG) was chosen, using image geo-tags determined with a low-performance GNSS receiver (average accuracy of 3 m).

Table 1. Overview of surveyed scenarios adopted technologies and acquired datasets.

	Case study (a)	Case study (b)
Location	Grottole (MT), Italy	Bari (BA), Italy
Equipment	DJI Mavic 2 Zoom RGB f-4.386, Model FC2204	DJI Inspire 1 v.2 ZenMuse X3 RGB f-3.61, Model FC350
Images	287 images (4000 × 3000 pix)	87 images (4000 × 3000 pix)
AGL/GSD	30 [m]/1.3 [cm/pix]	90 [m]/3.9 [cm/pix]
Georeferencing strategy	IG with 11 GCPs in nRTK	DG with low-cost GNSS receiver

The processing of the collected datasets was based on the workflow proposed in [20–22]. The different parameterisations for each dataset were detailed in Saponaro et al. [17]. Agisoft Metashape software (v.1.4.1, Agisoft LLC -St. Petersburg, Russia) was used during the work on Intel(R) Core (TM) i7-3970X CPU 3.50 GHz hardware,

with 16 GB of RAM and an NVIDIA GeForce GTX 650 graphics card to return the photogrammetric products.

Four orthomosaics were exported for each scenario examined: beginning from the highest resolution and then doubling, tripling, and quadrupling the former. Additional down-sampling would not rationalize the use of UAV technology [23]. Specifically, the Agisoft Metashape software made it possible to export the orthomosaics at the selected resolution, resampling each time by bilinear interpolation (Table 2).

Table 2. Summary of spatial resolution down sampling in terms of GSD values performed at the orthomosaics of the two scenarios under study

	Case study (b)	Case study (c)
Orthomosaic spatial resolution	{1} 0,017 [m/pix]	{1} 0,036 [m/pix]
{1} min.res	{2} 0,035 [m/pix]	{2} 0,071 [m/pix]
{2} min.res.x2	{3} 0,052 [m/pix]	{3} 0,107 [m/pix]
{3} min.res x3	{4} 0,070 [m/pix]	{4} 0,142 [m/pix]
{4} min.res.x4		

The orthomosaics were imported into the open-source software QGIS (3.16.5 'Hannover') [24]. Following the procedures adopted in [25], an Empirical Line Method (ELM) was applied in order to produce a radiometric calibration of the orthomosaics. Several vegetative indices in the visible bands were evaluated and their performance was analysed in terms of their ability to distinguish between vegetative and non-vegetative classes. As described in [17], the most effective vegetative indices in terms of performance were Triangular Greenness Index (TGI) (Eq. 1) and Red–Green Ratio Index (IRG) (Eq. 2), calculated as follows:

$$\mathrm{TGI} = 0.5 * [(\lambda_R - \lambda_B)(\rho_R - \rho_G) - (\lambda_R - \lambda_G)(\rho_R - \rho_B)] \quad (1)$$

$$\mathrm{IRG} = \rho_R - \rho_G \quad (2)$$

2.2 Pixel-Based Classification

The scientific literature presents a multitude of classification methodologies and the choice, in this research work, fell on pixel-based image analysis algorithms. Pixel-based image classifications are based on the simple hypothesis that the spectral response patterns of individual pixels and the textures linked with the classes of interest can be statistically categorized into a suitable set of informative classes. In object-based image analysis, in fact, the analyst is faced with an unavoidable challenge: the determination of the segmentation parameters, in particular the segmentation scale, is often very complex and, if not supported by certain expertise, is bypassed by a default parameterization or passed in overparameterization [26, 27].

In view of the classification procedures' automation, these algorithms can potentially be improved by adding data from any NIR and/or Red-Edge bands to the simple RGB

bands. In other cases, one could combine data from the digital orthomosaic with height data from a DEM. In this research, however, information obtained from the calculation of visible-band vegetation indices were tested. UAV-based orthomosaics in the radiometrically calibrated visible bands and vegetation maps in the TGI and IRG indices were exported from QGIS software and uploaded to the open-source Sentinel Application Platform (SNAP) developed by Brockmann Consult, Skywatch, Sensar and CS and provided free of charge by ESA/ESRIN. SNAP provides several supervised and unsupervised raster image classification algorithms. For this tests, the processing chains concerning classifications with supervised Random Forest (RF) [18] and unsupervised K-Means Clustering [19] algorithms were adopted.

Supervised Classification. A rapid and objective tool for feature selection is a decision tree, as it is a non-parametric statistical method that is not impacted by outliers and correlations, can uncover interactions between variables and is also an outstanding data reduction tool. In a decision tree, a set of data is consecutively subdivided into increasingly homogeneous subsets until terminal nodes are established. However, a single decision tree is not performance-adequate to undertake highly complex feature classification and multi-class dimension prediction.

The Random Forest, on the other hand, is a tree-based machine learning algorithm that utilizes the power of multiple decision trees to get predictions.

This decision trees classifier is really a pixel-based classification method that then performs multi-stage classifications using a series of binary decisions to separate pixels into different classes. Each decision divides the pixels of an image into two classes on the basis of an expression. This large number of relatively unrelated decision trees essentially acts as a committee to overcome the individual constituent models. Each individual tree in the random forest makes a class prediction and so the class with the most votes joins the model prediction. In all that, the low correlation among the models is the key to the proficiency of the whole. In this sense, the trees safeguard each other from their individual errors, assuming they are not all consistently in the same direction. To ensure that RF works, the predictions and thereby the errors committed by individual trees must have low correlations with each other.

In order to do so, however, the algorithm requires training areas to be plotted so that there are real signals of the features being sought and so that models generated using these features perform better than random hypotheses screened at some nodes. In addition, the algorithm makes use of the Bootstrap aggregation method. In practice, Decision trees are very responsive to the data they are trained on: small changes to the training set can yield significantly different tree compositions. However, the random forest capitalises on this advantage by allowing each individual tree to be randomly sampled from the dataset with replacements, producing different trees. This process is commonly known as bagging. This forces even greater variation between trees in the model and eventually leads to less correlation between trees and greater diversity.

De Castro et al. [28] recommended that the Random Forest (RF) machine learning algorithm remained the best for classification automation, as it involves far fewer classification parameters than similar machine learning techniques.

In general, these classification methods are entitled supervised because they employ conventional polygons designed by an expert on the basis of ground truth identification.

Precisely, the signature area or training area is created using an area of interest (AOI) through which class signatures are collected and then applied as seeds for the extraction of the signature used in the classification. It tends to be the case that more than one training area will have to be considered for each class and then merged to overcome inhomogeneities between the AOIs selected as sample data for the respective class. This thus establishes the statistical basis for the software to recognise existing classes in a raster file. The time invested in performing this step is then spent on training the algorithm, i.e., calibrating it to work automatically in recognising the various features.

The automatic but supervised pixel-based classification method has a specific disadvantage regarding the time efficiency of the algorithm, as the time required for its completion is strongly influenced by the separate polygons created and the spectral thresholds of the bands and/or indices used, which in turn often depend on the noise level in the UAV images.

Once all of the above QGIS products from the two scenarios and in the four spatial resolutions have been imported into the SNAP software, the AOI collection was organised in such a way as to consider it balanced by classes, following the recommendations of [29] and [30]. Four types of classes were identified, labelled as asphalt, bare soil, vegetation, and water. As a consequence of the low flight height, UAV images capture a relatively small footprint on the ground, but the number of pixels in the image is comparatively high due to the high resolution. Therefore, the analysis of UAV images is even more powerful than that of conventional aerial and satellite images. Therefore, 10 AOIs per class were plotted, attempting to collect within them the spectral variability and texture of each class. The methodology used to create a training dataset for the automatic methods guarantees that the training data always covers the same percentage of area for each class, which assures the usability and robustness of the algorithm.

Unsupervised Classification. It is considered an iterative procedure with no need for prior information [30]. In general, these methods are considered as clustering algorithms, which calculate distance functions and group similar pixel values into a spectral class. In the final step, the generated categories are recognised and labelled according to the colour composition of the image and expert observations. finally, these categories can be divided into the desired classes. However, the output classes possibly do not correspond to any of the classes of interest, as they account for ambiguous spectral classes. Effectively, clustering involves dividing a large dataset into a multiplicity of data clusters, which reveal certain characteristics of each subset. This is done by estimating similarity or closeness on the basis of the distance measurement method in order to find a structure in an unnamed collection of data. Thus, a cluster is an assortment of objects that are similar to each other and are "dissimilar" to objects belonging to other clusters.

One of the most characteristic and commonly used clustering algorithms for performing unsupervised classifications of satellite-based rasters is the so-called K-means Algorithm. Tang et al. [31] positively combined a K-means method with a machine learning algorithm for weed detection. Vrindts et al. [32] applied revised K-means clustering to establish management zone gradation using soil and crop data. Indeed, from the cluster analysis the properties of the dataset and the target variable can be determined, and these come in handy when determining how to measure the similarity distance.

As described in [19], the algorithm requires as input the number of k centroids for each cluster and a database containing n data items, representing the set of all instances analysed by the algorithm. The centroids are imaginary or real points at the centre of a cluster from which the Euclidean distances to other data will be evaluated. The algorithm iteratively evaluates the distances of these data from the centroids and each is reassigned to the cluster to which the object is most similar according to the average value of the objects in the cluster. As the average value is updated, iteratively re-assignments occur until no more changes occur, i.e., a point of convergence is reached where no more cluster changes occur.

The k-means algorithm has the advantage of being quite fast, as few calculations and consequently, little computer processing time is required to calculate the distances between the data and the centroids at each iteration. On the other hand, k-means has a couple of disadvantages. Firstly, it is necessary to select how many k groups to show. This is not always trivial as it is not always possible to do this, especially for problems of higher complexity. In addition, K-means also starts with a random choice of centroids and therefore may produce different clustering results on different sequences of the algorithm. Consequently, the outcomes may not be repeatable and lack consistency. In the SNAP software, the unsupervised K-Means algorithm classifications were started. A value of the parameter K, i.e., the number of clusters, was set to 5 and the default number of iterations to 30. At the end of each processing, the clusters generated were manually labelled according to the class to which they could be identified.

2.3 Confusion Matrix

In order to verify and validate the classification results, 30 pins for each prediction class were distributed in each investigated scenario to perform the bootstrapping method. These pins represent ground truths digitally identified by the orthomosaics themselves and assigned to a certain class. Once the classification processes are complete, the predictions of the classes in the pins can be compared with the previously assigned classes and confusion matrices can be constructed.

After the algorithms have been efficiently applied, the classification results are assessed against the accuracy of the actual data. The bootstrapping method was adopted to allow a formal statistical comparison of the accuracy of the classification approaches adopted.

Adopting the guidance given in [33], the following metrics are computed to quantify the accuracy of the extracted features:

- Error Rate (ERR) is calculated as the number of all incorrect predictions divided by the total number of the dataset. The best error rate is 0 and the worst is 1.

$$ERR = \frac{FP + FN}{TN + FP + FN + TP} \tag{3}$$

- Accuracy indicates the accuracy of the model as the name implies. Therefore, the best accuracy is 1, while the worst is 0.

$$Accuracy = \frac{TP + TN}{TN + FP + FN + TP} \tag{4}$$

- F-Score: is a weighted harmonic average of the Precision and Recall metrics such that the best score is 1 and the worst is 0.

$$F - Score = \frac{2 * Recall * Precision}{Recall + Precision} \tag{5}$$

where Precision is the ability of a classifier not to label a positive instance that is negative. For each class, it is defined as the ratio of true positives to the sum of true and false positives.

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

Recall, also called sensitivity, is the ability of a classifier to find all positive instances. For each class, it is defined as the ratio between true positives and the sum of true positives and false negatives:

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

For this purpose, True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN) are estimated for the vegetation, asphalt and bare soil classes based on ground truth data. TP signifies the correctly classified objects for the given class among all obtained objects, FP is the falsely classified objects among the extracted objects, and FN stands the missed or not extracted objects. Finally, TN denotes the objects correctly classified among all extracted objects but not within the class in question.

Confusion matrices are thus constructed for each spatial resolution of each scenario for three cases of classification, both supervised and unsupervised: using only the visible (RGB) bands, adding to these the vegetation map of the TGI index and, finally, adding to the RGB bands the information of the IRG index.

3 Results and Discussion

Figures 2 and 3 present the results of the classifications for each scenario and at each spatial resolution, obtained with the RF and K-Means algorithms, respectively. In particular, at each resolution the responses of the classification algorithms are presented when assisted by the vegetation indices in the visible bands, i.e., the TGI and IRG.

A qualitative inspection comparing the corresponding classifications reveals the difficulties of correct clustering by K-Means algorithms compared to RF. In scenario (a), if on the one hand the identification of vegetation and bare ground are quite reliable, in the unsupervised classification the difficulty in distinguishing asphalt from other classes and vice versa stands out. The same algorithm exhibits the same problems in the second study area. In this scenario, in fact, the most critical issues are found in the clustering of water, which is often included in other classes.

After performing and qualitative inspecting the supervised and unsupervised classification procedures using the RF and K-Means algorithms, respectively, the validation

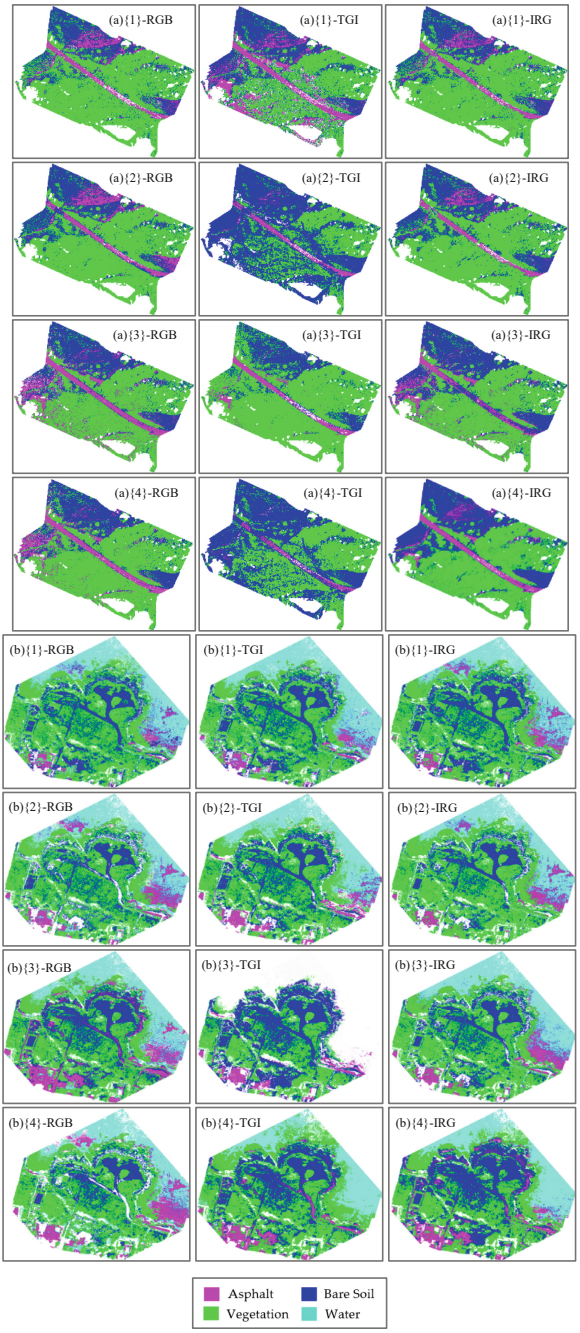

Fig. 2. RF maps for the cases analysed. In the first column, in descending order by spatial resolution, the classifications obtained using only the RGB bands. In the second by adding the TGI map, in the third by adding the IRG map.

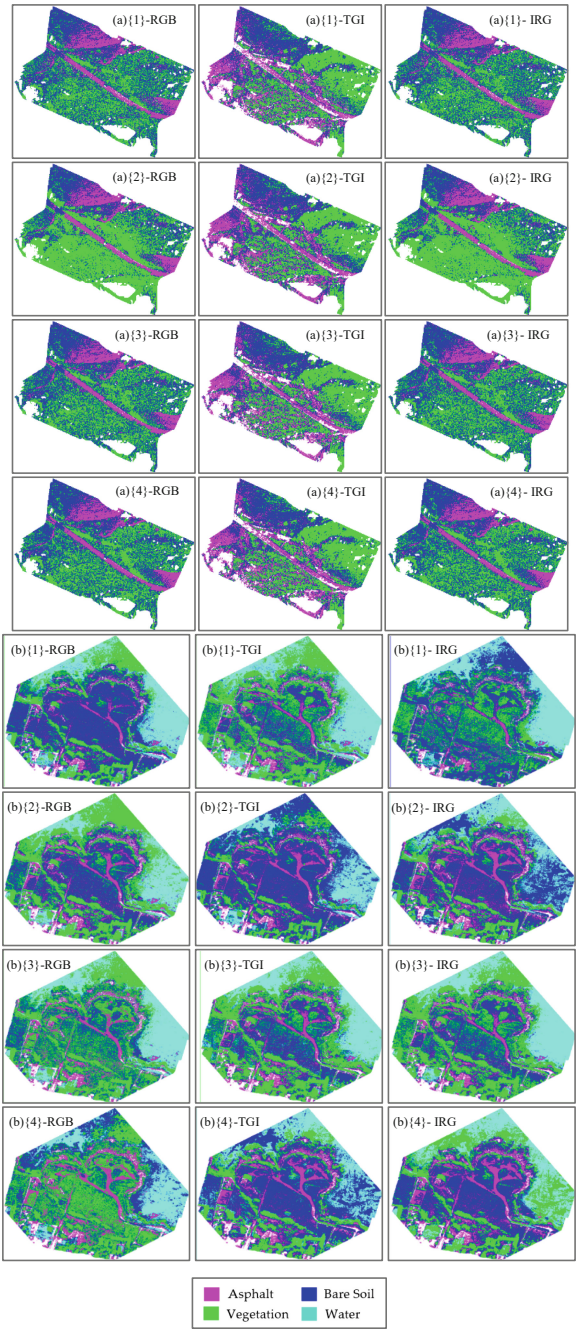

Fig. 3. K-Means algorithm maps for the cases analysed. In the first column, in descending order by spatial resolution, the classifications obtained using only the RGB bands. In the second by adding the TGI map, in the third by adding the IRG map.

metrics were extracted. By comparing the labelling assigned and the labelling predicted by the software in the 90 pins placed, for each scenario, at each resolution and for each classification mode (RGB bands, addition of the TGI band, addition of the IRG band), confusion matrices were drawn and from these the relevant metrics were derived. In Tables 3 and 4, the metrics were summarised in the mean values of the parameters calculated for each class, i.e., vegetation, asphalt and bare soil.

In general, a coarse radiometric calibration such as ELM does not permit a clear distinction to be made between and in the image elements. This results in a diminished capability to distinguish classes such as asphalt and bare ground, which in several cases may have spectral similarities (e.g., as in the case of heavily degraded asphalt). In fact, as can be seen from Figs. 2 and 3, in all the cases analysed, irrespective of the classification methods used, the greatest diatribe emerges precisely between the two classes asphalt and bare ground.

Table 3. Summary of metrics deduced from confusion matrices in cases of supervised classification

		RF [RGB]				RF [RGB+TGI]				RF [RGB+IRG]			
		{1}	{2}	{3}	{4}	{1}	{2}	{3}	{4}	{1}	{2}	{3}	{4}
(a)	Error rate	0.14	0.17	0.05	0.08	0.08	0.07	0.10	0.03	0.11	0.13	0.09	0.02
	Accuracy	0.86	0.83	0.95	0.92	0.92	0.93	0.90	0.97	0.89	0.87	0.91	0.98
	F-score	0.80	0.75	0.92	0.88	0.89	0.88	0.85	0.95	0.83	0.81	0.85	0.97
(b)	Error rate	0.25	0.13	0.10	0.13	0.22	0.12	0.14	0.16	0.32	0.17	0.15	0.11
	Accuracy	0.75	0.87	0.90	0.87	0.78	0.88	0.86	0.84	0.68	0.83	0.85	0.89
	F-score	0.62	0.82	0.86	0.81	0.67	0.83	0.80	0.78	0.58	0.76	0.75	0.84

Table 4. Summary of metrics deduced from confusion matrices in cases of unsupervised classification

		K-Means [RGB]				K-Means [RGB+TGI]				K-Means [RGB+IRG]			
		{1}	{2}	{3}	{4}	{1}	{2}	{3}	{4}	{1}	{2}	{3}	{4}
(a)	Error Rate	0.32	0.25	0.27	0.27	0.11	0.08	0.12	0.08	0.32	0.25	0.27	0.25
	Accuracy	0.68	0.75	0.73	0.73	0.89	0.92	0.88	0.92	0.68	0.75	0.73	0.75
	F-score	0.57	0.66	0.63	0.65	0.82	0.87	0.82	0.88	0.57	0.66	0.63	0.66
(b)	Error Rate	0.32	0.47	0.44	0.47	0.28	0.45	0.45	0.50	0.25	0.32	0.37	0.55
	Accuracy	0.68	0.53	0.56	0.53	0.72	0.55	0.55	0.50	0.75	0.68	0.63	0.45
	F-score	0.58	0.40	0.44	0.40	0.63	0.42	0.44	0.37	0.61	0.57	0.53	0.32

Regarding the classification with the RF algorithms, the metrics obtained from the analysis of the results are presented in Table 3. Considering the Error Rate values obtained

in the two scenarios, it immediately emerges how in the second context the presence of water generates ambiguity and therefore higher values of this metric. In fact, Saponaro et al. [17] had already guessed that the algorithms are often affected in the resolution {1} due to the presence of noise and distortions in the pixels. However, the values obtained can be considered acceptable in all RF classification modes, as also ascertained by the Accuracy values averaging over 0.8. Among the classification modes, those in which the TGI vegetation index is adopted for resolutions {1} and {2} and the simple mode with the RGB bands in the other two resolution solutions prove to be effective. Looking at the F-score values, these are functional in every solution in scenario (a), in particular in the [RGB+IRG] case in resolution {4} where a value of 0.97 is even reached. This combination in fact presents the best metrics. In the second scenario instead, as already discussed, a lowering of the F-score values was recorded in the resolution {1}. In the remaining part of the cases the values can be considered acceptable. In general, from the analysed values it emerges how an efficiency of the classifications occurs at resolutions reduced with respect to the original, within which the presence of noise generates ambiguity.

On the other hand, looking at the F-score values in the cases of unsupervised classification in Table 4, one finds a drastic reduction of the general values of about 30% with respect to those obtained by supervised classification. This is also confirmed by the remaining part of the calculated metrics Accuracy and Precision. Among the analysed values, with the exception of case (a) [RGB+TGI] where the values are very functional, both scenarios where the classification used IRG maps had a slight improvement. This shows how the presence of vegetative indices improves the ability to separate classes, so using unsupervised modes requires more incisive radiometric information such as that retrievable from the NIR bands. Even in this type of classification, comparing the trends of the F-score values with respect to the spatial resolutions at which they are calculated, it is not possible to extrapolate a certain regularity.

In summary, from Figs. 2, 3 and Tables 3, 4, the best-fitting results in the case of supervised classifications result for the scenarios: (a), spatial resolution {4}, classification mode with vegetative index IRG; (b), spatial resolution {3}, basic classification mode with RGB bands.

In the unsupervised modes, however, they result for the scenarios: (a) spatial resolution {4}, classification mode with vegetative index TGI; (b), spatial resolution {1}, classification mode with vegetative index TGI. This again shows how the various datasets react unevenly to the algorithms used.

4 Conclusion

Classification algorithms involve the automatic categorisation of response functions recorded in pixels by sensors in order to identify real-world objects into predefined classes. Higher spatial resolutions, versatility of use and lower survey costs that have emerged from the use of UAVs have made class recognition perform well for even the smallest of scenery properties. Despite these advantages, the need has emerged to address their limitations and to analyse their impact in the data post-processing phases. A need has arisen to validate processing procedures in order to fully define the comparability

of the products and, in some cases, their interchangeability with others obtainable from different RS techniques. The objective of this research was therefore to analyse the performance of two pixel-based classification procedures that could subsequently be translated into an automated and easy-to-apply method, especially for end users with limited knowledge of spatial data processing. A supervised classification approach using Random Forest algorithms and an unsupervised one using K-Means algorithms were tested. These are defined as statistical classifiers useful for Land Use/Land Cover (LULC) recognition and classifications were based on different reflectance values across the radiometrically calibrated wavebands that make up the UAV-based orthomosaics and vegetation indices in the visible bands.

The following observations emerged from the analysis of the results:

- As hypothesised, the supervised RF classifications returned very comfortable metrics compared to the unsupervised modes. In general, a 30% reduction in the values of the analysed metrics was found in the latter classifications.
- Especially for the unsupervised classifications, the presence of vegetative indices noticeably improves their algorithm separation capabilities. In these cases, it would be preferable to use bands in the NIR for a better automatic extraction of the vegetation.
- There is no regularity between the separation capabilities of the classification algorithms and the spatial resolution of the orthomosaics. The efficiency is very much influenced by the context and any similarities between the reflectances of the various classes, e.g. vegetation and water or bare ground and asphalt. It turns out, however, that reduced resolutions compared to the original one optimise the classification capabilities as there is a reduction of noise in the images.

In this work it is demonstrated how the orthomosaics produced by UAV acquisitions equipped with non-metric and low-cost RGB sensors can generate very valuable information for vegetation extraction. Further investigations will have to be undertaken in order to assess how additional ancillary data, such as a digital elevation model (DEM) also based on UAV acquisitions, can improve the chain of processes discussed in this work.

References

1. Hamylton, S.M., et al.: Evaluating techniques for mapping island vegetation from unmanned aerial vehicle (UAV) images: pixel classification, visual interpretation and machine learning approaches. Int. J. Appl. Earth Obs. Geoinf. **89**, 102085 (2020)
2. Alif, A.A., Shukanya, I.F., Afee, T.N.: Crop prediction based on geographical and climatic data using machine learning and deep learning. BRAC University (2018)
3. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**, 436–444 (2015)
4. Novelli, A., Tarantino, E., Caradonna, G., Apollonio, C., Balacco, G., Piccinni, F.: Improving the ANN classification accuracy of Landsat data through spectral indices and linear transformations (PCA and TCT) aimed at LU/LC monitoring of a river basin. In: Gervasi, O., et al. (eds.) ICCSA 2016. LNCS, vol. 9787, pp. 420–432. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-42108-7_32

5. Capolupo, A., Saponaro, M., Fratino, U., Tarantino, E.: Detection of spatio-temporal changes of vegetation in coastal areas subjected to soil erosion issue. Aquatic Ecosyst. Health Manage. 1–8 (2020)
6. Xie, Y., Sha, Z., Yu, M.: Remote sensing imagery in vegetation mapping: a review. J. Plant Ecol. **1**, 9–23 (2008)
7. Capolupo, A., Monterisi, C., Saponaro, M., Tarantino, E.: Multi-temporal analysis of land cover changes using Landsat data through Google Earth Engine platform. In: SPIE (ed.) Eighth International Conference on Remote Sensing and Geoinformation of the Environment (RSCy2020), vol. 11524 pp. 447–458, August 2020
8. Tarantino, E., Novelli, A., Aquilino, M., Figorito, B., Fratino, U.: Comparing the MLC and JavaNNS approaches in classifying multi-temporal LANDSAT satellite imagery over an ephemeral river area. Int. J. Agricult. Environ. Inf. Syst. (IJAEIS) **6**, 83–102 (2015)
9. Capolupo, A., Monterisi, C., Caporusso, G., Tarantino, E.: Extracting land cover data using GEE: a review of the classification indices. In: Gervasi, O., et al. (eds.) ICCSA 2020. LNCS, vol. 12252, pp. 782–796. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58811-3_56
10. Sarzana, T., Maltese, A., Capolupo, A., Tarantino, E.: Post-processing of pixel and object-based land cover classifications of very high spatial resolution images. In: Gervasi, O., et al. (eds.) ICCSA 2020. LNCS, vol. 12252, pp. 797–812. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58811-3_57
11. Keyport, R.N., Oommen, T., Martha, T.R., Sajinkumar, K., Gierke, J.S.: A comparative analysis of pixel-and object-based detection of landslides from very high-resolution images. Int. J. Appl. Earth Obs. Geoinf. **64**, 1–11 (2018)
12. Haghighattalab, A., et al.: Application of unmanned aerial systems for high throughput phenotyping of large wheat breeding nurseries. Plant Methods **12**, 35 (2016)
13. Wang, C., Myint, S.W.: A simplified empirical line method of radiometric calibration for small unmanned aircraft systems-based remote sensing. IEEE J. Sel. Top. Appl. Earth Obser. Remote Sens. **8**, 1876–1885 (2015)
14. Pompilio, L., et al.: Application of the empirical line method (ELM) to calibrate the airborne Daedalus-CZCS scanner. Eur. J. Remote Sens. **51**, 33–46 (2018)
15. Saponaro, M., Capolupo, A., Caporusso, G., Tarantino, E.: Influence of co-alignment procedures on the co-registration accuracy of multi-epoch SFM points clouds. Int. Arch. Photogrammetry, Remote Sens. Spat. Inf. Sci. **43**, 231–238 (2021)
16. Saponaro, M., Capolupo, A., Caporusso, G., Borgogno Mondino, E., Tarantino, E.: Predicting the Accuracy of Photogrammetric 3D Reconstruction from Camera Calibration Parameters Through a Multivariate Statistical Approach. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLIII-B2–2020, pp. 479–486 (2020)
17. Saponaro, M., Agapiou, A., Hadjimitsis, D.G., Tarantino, E.: Influence of spatial resolution for vegetation indices' extraction using visible bands from unmanned aerial vehicles' orthomosaics datasets. Remote Sens. **13**, 3238 (2021)
18. Breiman, L.: Random forests. Mach. Learn. **45**, 5–32 (2001)
19. Jung, Y.G., Kang, M.S., Heo, J.: Clustering performance comparison using K-means and expectation maximization algorithms. Biotechnol. Biotechnol. Equip. **28**, S44–S48 (2014)
20. Capolupo, A., Saponaro, M., Borgogno Mondino, E., Tarantino, E.: Combining interior orientation variables to predict the accuracy of RPAS-SFM 3D models. Remote Sens. **12**, 2674 (2020)
21. James, M.R., et al.: Guidelines on the use of structure-from-motion photogrammetry in geomorphic research. Earth Surf. Proc. Land. **44**, 2081–2084 (2019)

22. Saponaro, M., Turso, A., Tarantino, E.: Parallel development of comparable photogrammetric workflows based on UAV data inside SW platforms. In: Gervasi, O., et al. (eds.) ICCSA 2020. LNCS, vol. 12252, pp. 693–708. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58811-3_50
23. Smith, M.W., Carrivick, J., Quincey, D.: Structure from motion photogrammetry in physical geography. Prog. Phys. Geogr. **40**, 247–275 (2016)
24. https://www.qgis.org/
25. Agapiou, A.: Vegetation extraction using visible-bands from openly licensed unmanned aerial vehicle imagery. Drones **4**, 27 (2020)
26. Safonova, A., Guirado, E., Maglinets, Y., Alcaraz-Segura, D., Tabik, S.: Olive tree biovolume from UAV multi-resolution image segmentation with mask R-CNN. Sensors **21**, 1617 (2021)
27. Shukla, A., Jain, K.: Automatic extraction of urban land information from unmanned aerial vehicle (UAV) data. Earth Sci. Inf. **13**(4), 1225–1236 (2020). https://doi.org/10.1007/s12145-020-00498-x
28. De Castro, A.I., Torres-Sánchez, J., Peña, J.M., Jiménez-Brenes, F.M., Csillik, O., López-Granados, F.: An automatic random forest-OBIA algorithm for early weed mapping between and within crop rows using UAV imagery. Remote Sens. **10**, 285 (2018)
29. Belgiu, M., Stein, A.: Spatiotemporal image fusion in remote sensing. Remote Sens. **11**, 818 (2019)
30. Gašparović, M., Zrinjski, M., Barković, Đ, Radočaj, D.: An automatic method for weed mapping in oat fields based on UAV imagery. Comput. Electron. Agric. **173**, 105385 (2020)
31. Tang, J., Wang, D., Zhang, Z., He, L., Xin, J., Xu, Y.: Weed identification based on K-means feature learning combined with convolutional neural network. Comput. Electron. Agric. **135**, 63–70 (2017)
32. Vrindts, E., et al.: Management zones based on correlation between soil compaction, yield and crop data. Biosys. Eng. **92**, 419–428 (2005)
33. Hossin, M., Sulaiman, M.N.: A review on evaluation metrics for data classification evaluations. Int. J. Data Min. Knowl. Manage. Process **5**, 1 (2015)