Roof covering classification using Skysat multispectral imagery

Gaetana Ganci[1, a)], Giuseppe Bilotta[1], Michele Mangiameli[2], Giuseppe Mussumeci[2], and Annalisa Cappello[1, b)]

Author Affiliations
[1]*Istituto Nazionale di Geofisica e Vulcanologia, Osservatorio Etneo, Piazza Roma 2, 95125 Catania, Italy*
[2] *Dept. of Civil Engineering and Architecture, University of Catania, Via Santa Sofia 64, 95125 Catania, Italy*

[a)] Corresponding author: gaetana.ganci@ingv.it
[b)] annalisa.cappello@ingv.it

Abstract. Classification of roof covering in urban areas using aerial imagery is a challenging task. In this work we present a preliminary mapping of roofs using the high-resolution Skysat multispectral images. The classification is performed using a two-stage machine learning approach: the first stage includes a supervised classification for land use, while the second stage includes the classification of terraces and roofs with one or more pitches in those areas previously recognized as edifices. The methodology has been tested to classify the roofs in the north-east part of the Stromboli Island (Sicily, Italy). Our preliminary results are promising and encourage us to pursue further developments as ways to improve accuracy and reliability of the classification.

INTRODUCTION

Satellite images are gaining a continuously growing importance in all the fields of Earth surface monitoring including Land - Oceans - Atmosphere, providing observations at bird-eye and low cost difficult to impossible to obtain from the ground. Apart from classical weather forecasting applications, information coming from satellite sensors are currently used for Land Surface Temperature (LST), high temperature events characterization, such as volcanic eruptions or forest fire, ground deformation measurements, digital surface models productions, land use classification, etc., to monitor the changing environment and support actions to assess and possibly mitigate natural disasters (see e.g. [1-5]).

The rapid increase of low-cost, small satellites (e.g., nano- and microsatellites) has represented a revolution for the regular observation and monitoring of Earth. These satellites, which are cheap to make and cheap to launch, offer greater operational capacity and reduced revisit interval (from hours to days). One of the most successful examples of microsatellites is SkySat. Skysat data is acquired by a constellation of 21 high-resolution satellites owned and operated by Planet Labs, and is available in different product lines [6].

In this work we used a Skysat Collect image acquired on the Island of Stromboli (Sicily, Italy) in May 2021. This imagery was analyzed through a two-stage machine learning approach using Google Earth Engine (GEE), the cloud-based computing platform established by Google to support remote sensing data processing.

The output consists of two maps focused on the north-east area of the Stromboli island. The first map includes four land use types (sea, vegetation, roads and edifices), while the second map provides a classification of the roof coverage among flat, single pitch and double (or more than two) pitch roofs.

MATERIAL AND METHODS

The SkySat Collect product used in this work is an orthorectified composite of sensor and geometrically corrected scenes, acquired on 21 May 2021 at 09:40 UTM above Stromboli Island. The imagery consists of 4 multispectral bands (blue, green, red and NIR-infrared), as well as the panchromatic band, with a spatial resolution of 50 cm. For our analysis, we focused on an area of 2.43 km^2 including the most densely populated part of Stromboli Island (Fig. 1).

FIGURE 1. Skysat image showing the north-east part of Stromboli Island, which was used to test our approach. The inset shows the location of the study area.

In order to perform the roof cover classification, we applied the “Random forest” (RF) algorithm, which is a supervised learning algorithm commonly used to solve regression and classification problems. It uses the so-called “ensemble learning”, a technique that combines many classifiers to provide solutions to complex problems. The RF algorithm consists of many decision trees, whose predictions are aggregated to establish a final outcome.

Different algorithms for random decision forests are available in literature (e.g. [7, 8]). Here we used the RF implementation available through the Classifier package of GEE, whose general workflow consists of five steps:

1. Collection of the training data. This is done by creating n different layers (one per class) as FeatureCollection, using the imagery as guidance;
2. Instantiation of the RF classifier;
3. Training of the RF classifier using the training data;
4. Classification of the image to create a n class image;
5. Accuracy assessment using independent validation data.

In this work, we applied the GEE workflow first for land cover by collecting four training classes (sea, vegetation, paved roads/dirt tracks, and edifices), and then to discriminate the coverage of only the edifices using other three training classes (flat, single pitch, double pitch roofs).

In order to assess the accuracy of the classification, we divided the training samples in two parts and used one for training the classifier (the 70%) and the other for validating the prediction (the 30%). After training the classifier, the whole image is classified and the classified values are compared with the ones of the validation. We thus computed the confusion matrix to evaluate the accuracy of the classification. Moreover, we performed a qualitative validation by visual inspection using the high resolution color satellite data available in GEE.

RESULTS

The outcome of our two-stage machine learning approach is shown in Figure 2. The first stage of the RF algorithm provided a land use classification (Fig. 2a) with four classes: sea (in blue), vegetation (in green), paved roads/dirt tracks (in gray), and edifices (in red). We found that the sea and vegetation cover most of the study area (40% and 39%, respectively), while roads and dirt tracks cover 11% of the area. The part classified as edifices is the smallest one, with only 10% (about 2.5 km^2).

The roof covering classification (Fig. 2b) from the second stage of the RF algorithm returns three classes: terrace (in red), single pitch (in yellow), two or more pitches (in blue). Following this classification, 45% of edifices have a flat covering, 14% have a single pitch roof, and 41% have a double pitch roof or roofs with more than two pitches.

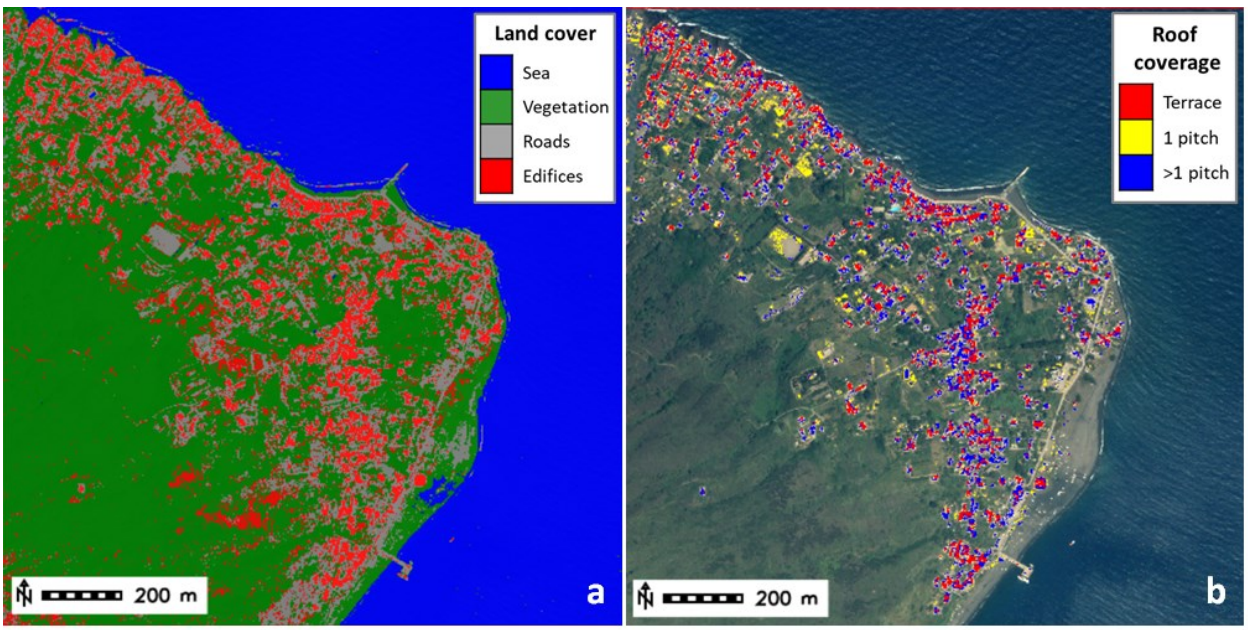

FIGURE 2. Land use (a) and roof covering (b) classification of the north-east part of Stromboli Island.

Using the confusion matrix to perform a quantitative validation of the results, we found an overall accuracy of about 85% for both classifications (0.847 for land use; 0.852 for roof covering). By making sample qualitative comparisons with the high resolution images provided by Google Earth, we also found a good fit.

CONCLUSION

We developed a two-stage machine learning approach based on the Random Forest algorithm to perform a land use classification and a subsequent roof covering classification. We implemented our approach using the Google Earth Engine Platform. For the classification we used a very high resolution Skysat imagery (0.5 m of spatial resolution) acquired on 21 May 2021 on Stromboli Island. We obtained two maps, which were validated using quantitative and qualitative validation.

In this preliminary study, our intention was to explore the potential of Skysat imagery to map the land use and roof covering, as well as to identify possible limitations of using these images. We believe that the achieved results are encouraging and indicate the appropriateness of the method. However further investigations are needed to improve the accuracy and reliability especially of the roof covering.

ACKNOWLEDGMENTS

Skysat imagery was provided by Planet Labs Inc. (https://www.planet.com/products/planet-imagery). The authors acknowledge Google Earth Engine for providing online data processing service for free. This work was supported by the INGV project Pianeta Dinamico (CUP D53J19000170001) funded by MIUR ("Fondo finalizzato al rilancio degli investimenti delle amministrazioni centrali dello Stato e allo sviluppo del Paese," legge 145/2018), Tema 8 – PANACEA 2021.

REFERENCES

1. G. Ganci, A. Cappello, G. Bilotta, A. Herault, V. Zago and C. Del Negro. Mapping Volcanic Deposits of the 2011–2015 Etna Eruptive Events Using Satellite Remote Sensing, 6, 83, doi: 10.3389/feart.2018.00083 (2018).
2. A. Gennaro, A. Candiano, G. Fargione, M. Mangiameli and G. Mussumeci. Multispectral remote sensing for post-dictive analysis of archaeological remains. A case study from Bronte (Sicily). Archaeological Prospection. 1– 13, doi:10.1002/arp.1745 (2019).
3. G. Ganci, A. Cappello, G. Bilotta, C. Corradino and C. Del Negro. *Satellite-Based Reconstruction of the Volcanic Deposits during the December 2015 Etna Eruption.* Data, 4, 120, doi: 10.3390/data4030120 (2019).
4. G. Ganci, A. Cappello, G. Bilotta, C. Del Negro. *How the variety of satellite remote sensing data over volcanoes can assist hazard monitoring efforts: The 2011 eruption of Nabro volcano.* Remote Sensing of Environment 236, 111426, doi: 10.1016/j.rse.2019.111426 (2020).
5. A.V. Kavitha, A. Srikrishna, Ch. Satyanarayana, A Review on Detection of Land Use and Land Cover from an Optical Remote Sensing Image. CHSN 2020 IOP Conf. Series: Materials Science and Engineering 1074. doi:10.1088/1757-899X/1074/1/012002 (2021).
6. Planet Team. https://assets.planet.com/docs/Planet_Combined_Imagery_Product_Specs_letter_screen.pdf (2021). Last accessed July 2021.
7. T.K. Ho, "Random decision forests," Proceedings of 3rd International Conference on Document Analysis and Recognition, pp. 278-282 vol.1, doi: 10.1109/ICDAR.1995.598994 (1995).
8. L. Breiman, Random Forests. *Machine Learning* 45, 5–32. doi: 10.1023/A:1010933404324 (2001).