Land Cover and Crop Classification using Multitemporal Sentinel-2 Images Based on Crops Phenological Cycle

Aleem Khaliq
Department of Electronics
& Telecommunications (DET)
Politecnico di Torino (PoliTO)
Turin, Italy.
aleem.khaliq@polito.it

Leonardo Peroni
Department of Control
& Computer Engineering (DAUIN)
Politecnico di Torino (PoliTO)
Turin, Italy.
s230061@studenti.polito.it

Marcello Chiaberge
Department of Electronics
& Telecommunications (DET)
Politecnico di Torino (PoliTO)
Turin, Italy.
marcello.chiaberge@polito.it

Abstract— Italy is considered one of the developed country in the field of agriculture in Europe. For many reasons, reliable classification of crops and mapping plays an important role in Precision Agriculture (PA). Increasing availability, improving spatial resolution and high revisit time of sentinel-2 satellite become more useful and play an important role in analyses for land cover, crop classification and other remote sensing applications. Most of the Crops with similar spectral characteristics can be distinguished by accumulating spectral information of different phenological stages. In literature, many solutions have been proposed to classify crops using multitemporal images acquired from various satellites equipped with multispectral imagery sensors. However, features and images selection from multispectral, multitemporal images still needs improvement. In this paper, crops phenological cycles are investigated using temporal normalized difference vegetation index (NDVI) patterns and major crop phenological information are used to select best multitemporal images for the input of random forest (RF) classifier to classify land cover and crops.

Keywords— precision agriculture; multispectral image; multitemporal image; Crop Phenology; Land Cover and crop classification;sentinel-2;

I. INTRODUCTION

Nowadays, governments and international agencies demand detailed information about spatial distribution of various crops that improves the decision-making process to manage agricultural practices and needs. Precise information about crop-area identification, early estimation of yield, cultivated and non-cultivated areas are some examples where accurate information is required [l]. One of the method used for the implementation of effective management decisions is agricultural mapping, which enables to get information and statistics on crops. To collect desired data, remote sensing (RS) is one of the popular source that offers suitable approach for land cover information which is economic and feasible [2,3].

In past few decades, RS has been used in many areas in particular for agricultural practices to achieve goals in precision agriculture. Crop identification and classification are the main sources to estimate spatial diversity occupied by various crops. In many studies, imagery from variety of satellites equipped with optical instruments have been used in crop classification [4-6]. However, imagery from optical instrument often contain clouds depending on weather at the time of acquisition, which makes classification process more sophisticated. As sentinel-2 has high revisit time, we can select cloud free images for our study area and can ignore images that contains clouds. To generate agricultural crop maps, classification of remotely sensed images is required for area of interest. Many techniques developed for classification, the selection of a algorithm is yet one of the challenges when doing image classification for crops [7], Various factors need to be considered such as computational resources, algorithm performance and classification accuracy. Algorithms can be categorized by per pixel, subpixel, and per-field and it can be either supervised or un-supervised.

Multispectral and multitemporal remote sensing data has already proved its potential to characterize the vegetation condition over different time spans, and has been widely used to generate crop maps [8]. Time period at which the images are taken, plays vital role in classification. In [9], various spectral bands from multispectral time series images were used for classification. Apart from time series data, various vegetation indices (VIs) derived from multispectral data have also been exploited and used to enhance the information to classify vegetation and non-vegetation area more effectively. However, using all bands information from several sources and vegetation indices of different time series, to form a feature space, involves large amount of data that increases complexity of classifier [10].

In recent years, machine learning based algorithms proposed and adopted for remote sensing applications such as support vector machine, random forest, maximum likely hood, k nearest neighbors and neural networks. In contrast to parametric classifiers, a machine learning approach does not follow data model but instead learns the relationship between the training and the response dataset [14]. The Random Forest algorithm has been widely used and received much attention for good classification results and processing speed.

978-1-5386-6405-6/18/$31.00 ©2018 IEEE

In [11], statistical measures are discussed which are used for assessment of classification accuracy. In which, confusion matrix is the key measure and it contains misclassification pattern without spatial information of overlapping region that leads to provide generally accurate measurement for the user.

In this work, time series of multispectral images from sentinel-2 are acquired. Importance and their impact has also been investigated in terms of features and best images selection for the multitemporal analysis, then best images are used for land and crop classification using RF classifier. Classification accuracy is used to assess classification performance for mono-temporal and multi-temporal. In addition, ground survey of land area of interest was conducted with GPS enabled mobile device to collect samples for the training and validation of classifier.

II. STUDY AREA AND DATASETS

A. Study Area

In this study, land cover and crop classification are addressed for Racconigi, Piedmont region, situated in north part of Italy with central coordinates 44° 48' 26" N, 7° 37' 37" E. The Piedmont region is one of the most fertile plains of Italy.

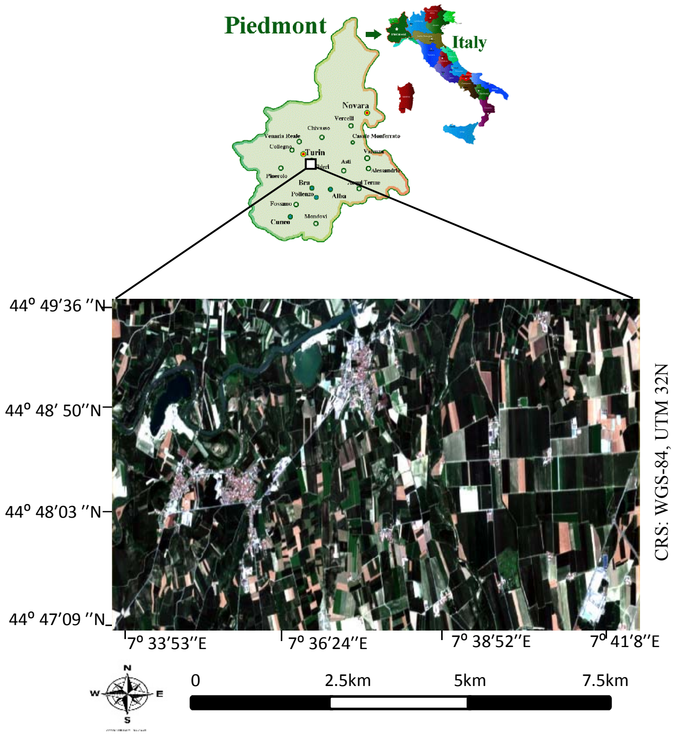


Fig. 1. RGB composite of subset of sentinel-2 images of study area.

Almost 46 Km² area covered in this study which includes lake, urban area, bare soil, maize, double maize, pearl millet, cabbage and few fields of onion.

B. Data Set

Sentinel-2 mission consists of twin polar-orbiting satellite launched by European Space Agency (ESA) in 2015 and are used in various application areas such as land cover change detection, natural disaster monitoring, forest monitoring and most importantly in agricultural monitoring and management [12]. It is equipped with multispectral optical sensors which captures 13 bands of different wavelengths shown in table-1. It has also high revisit time (10days at the equator and 5 days with twin satellites (Sentinel-2A, Sentinel-2B). It has gain more importance due to fact that it possesses various key features such as, free data products available at reasonable spatial resolution (which is 10m for Red, Green, Blue and Near Infrared bands), high revisit time and has very good spectral resolution among other available free data sources.

TABLE I. SENTINEL-2 MULTISPECTRAL BANDS DETAIL

Sentinel-2 Bands	**Central Wavelength (µm)**	**Resolution (m)**
Band 1 (B1) – Coastal	0.443	60
Band 2 (B2)– Blue	0.490	10
Band 3 (B3)– Green	0.560	10
Band 4 (B4)– Red	0.665	10
Band 5 (B5)– Vegetation Red Edge	0.705	20
Band 6 (B6)– Vegetation Red Edge	0.740	20
Band 7 (B7)– Vegetation Red Edge	0.783	20
Band 8 (B8)– NIR	0.842	10
Band 8A (B8A)– Narrow NIR	0.865	20
Band 9 (B9)– Water vapor	0.945	60
Band 10 (B10)– SWIR	1.375	60
Band 11 (B11)– SWIR	1.610	20
Band 12 (B12)– SWIR	2.190	20

Landscape is mostly flat terrain and crop calendar for summer is normally May to September. Most of the land is covered by maize among other types of crops. Sentinel-2 data products were downloaded from scihub which is free and open source [13]. In order to have reliable classification, nine data products/datasets were downloaded with different acquisition dates according with growing season of crops in our area of interest.

III. METHODOLOGY

In Fig. 2, overall methodology applied in this work is presented. Spectral bands from all data products were extracted and atmospheric correction was applied to all bands using SNAP tool box provided by ESA, then NDVI and ratio band R were derived from the bands reflectance, shown in Eq. (1) and Eq. (2).

$$\mathrm{NDVI} = \frac{\rho(B8)-\rho(B4)}{\rho(B8)+\rho(B4)} \quad (1)$$

$$R = \frac{\rho(B3)}{\rho(B2)} \quad (2)$$

Where ρ is the reflectance value for the associated band. As shown in table-1, in one sentinel data product, there are 13 bands having different spatial resolution. In this work, B2, B3, B4, B8, B8A and B12 were used as basic features and NDVI and R as derived features from B2, B3, B4 and B8.

NDVI band was used to establish temporal phenological pattern of different crops and B8, B12 and R bands were used to discriminate built ups from the bare soil and harvested crops. These bands were resampled at 10m resolution by using bi linear interpolation and formed as a stack of 8 bands. In total, nine stacks were formed to assess the mono-temporal and multitemporal classification performance.

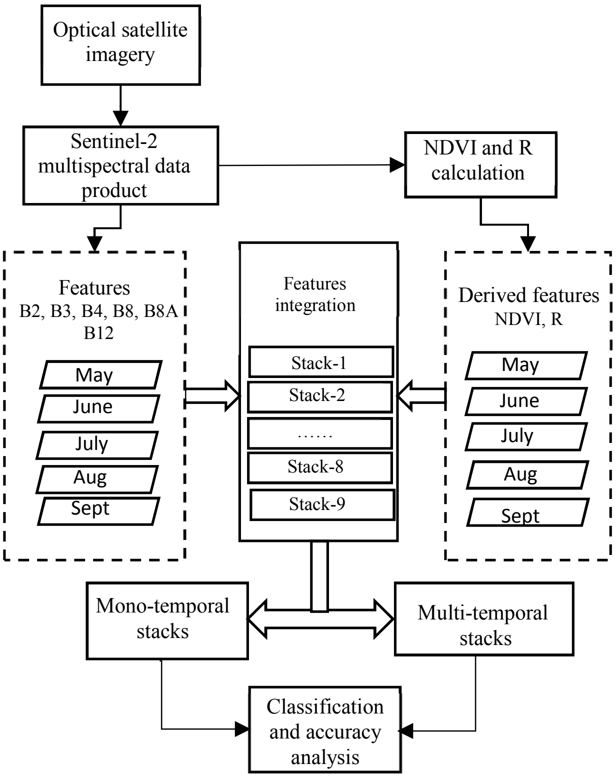
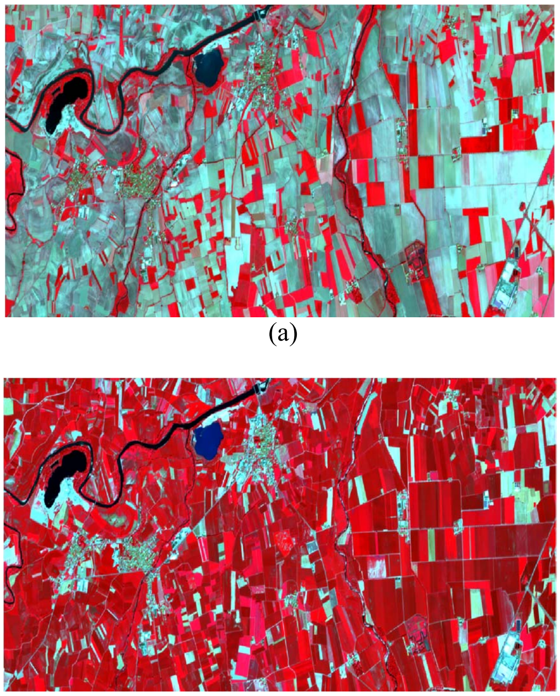


Fig. 2. RF classification of Landover and crops using various bands and vegetation indices as features from MS and MT images.

Random forest classifier package from scikit-learn library in python was used. In RF, primarily two parameters need to be tuned, first one is the number of trees which will be created by randomly selecting samples from the training samples, it was set to 100 to achieve optimum results in terms of accuracy and execution time, and the second is the number of variables used for tree nodes splitting, it was set on auto option. Land survey was conducted to get ground truth samples for Fig. 1. using mobile phone with integrated GPS. Pixel of images that belongs to ground truth data, was used for training and validation of RF algorithm.

From all the nine stacks, one by one stack was used as input variables for RF algorithm to assess the classification of mono temporal images. Over all accuracy (OAA) of mono-temporal stacks are shown in Fig.5. For multi-temporal classification, best two stacks were selected of two months by computing Jeffries–Matusita Distance for separability measure [15], and used as input variables to the classifier.

In Fig. 3, false infrared color representation is shown for 17^{th} May and 26^{th} July images. Red color is assigned to near infrared band, green color is assigned to red band and blue is assigned to green. As red color shows more vegetation, it can be noted that in May most of the crops are in starting phenological phase whereas in late July most of the crops are in closing phenological phase.



(a)

(b)

Fig. 3. False infrared color representation of 17^{th} May (a) and 26^{th} July (b)

IV. RESULTS

As shown in Fig. 4, multitemporal NDVI plots are used to assess the phenological cycle of each crop form 17^{th} May to 17^{th} September. Maize cycle begins in May, has the highest values of NDVI from June to late August, and is harvested in September. Double crop maize, which was considered spring maize during specific time span, reaches its peak growing period in May and July and is harvested in June and September respectively.

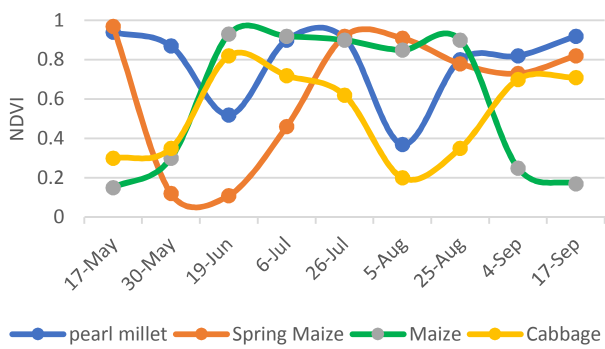


Fig. 4. Temporal NDVI patterns for each crop.

In Fig. 5, mono-temporal and multi-temporal overall accuracy (OAA) defined in Eq. 5. obtained from RF classifier is depicted. In multi-temporal, two stacks of features were taken from 17th May and 26th July based on major crop phenological phases and it gains the highest OOA among all other mono-temporal stacks.

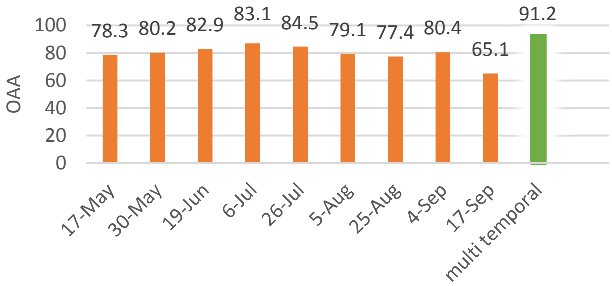


Fig. 5. Mono-temporal and multitemporal (17th May and 26th July) overall accuracy obtained from RF classifier.

Confusion matrix is mostly used to assess the performance of the classifiers in crop classification domain. In table II, confusion matrix is shown for only multi-temporal stack. Six classes were considered during the training and testing the RF classifier. There were also found few field of onions and grassy land and trees but these represents very few pixels in images, therefore these were not assigned as class and ignored.

TABLE II. CONFUSION MATRIX FOR BEST MULTI-TEMPORAL STACK

	Classification							
	Class	Built ups	Maize	Spring Maize	Cabbage	Pearl Millet	Water	PA (%)
Ground Truth	Built ups	280	2	18	1	2	0	92.4
	Maize	7	313	23	13	4	0	86.9
	S- Maize	6	0	160	0	6	0	93
	Cabbage	3	3	0	89	2	0	91.7
	Pearl Millet	9	16	0	8	180	0	83.3
	Water	0	0	0	0	0	188	100
	UA (%)	90.3	96.6	79.6	80.2	92.7	100	
	OOA	91.2%						

UA: User's Accuracy/ Recall, PA: Producers Accuracy/ Precision, OOA: Overall Accuracy, S-Maize: spring maize

The diagonal of the matrix contains the number of correctly classified pixels of each class. The off-diagonal elements contain misclassified pixels i.e. the number of ground truth pixels that ended up in another class during classification. Off-diagonal row elements represent ground truth pixels of a certain class which were excluded from that class during classification and Off-diagonal column elements represent ground truth pixels of other classes that were included in a certain classification class.

$$\mathrm{PA} = \frac{\text{Number of correctly classified pixels for class}}{\text{Total number of ground truth pixels for class}} \qquad (3)$$

$$\mathrm{UA} = \frac{\text{Number of correctly classified pixels}}{\text{Total number of pixels classified for class}} \qquad (4)$$

$$\mathrm{OAA} = \frac{\text{Total number correctly classified pixels of all classes}}{\text{Total number of ground truth pixels of all classes}} \qquad (5)$$

Producer accuracy, user accuracy and overall accuracy are defined in Eq. (3), Eq. (4) and Eq. (5) respectively and calculated in the matrix for each class. Maize achieved 96.6 % UA and stands highest among crops in terms of UA whereas Spring maize achieved 93% PA which is highest among other crops. OAA for the multitemporal stack was calculated 91.2% which is 6.7% higher than the highest OAA achieved by mono-temporal stack. Results suggest that only considering two months of images based on crop phenological information, can be useful in classification of landcover and crops which is quite easy and efficient in terms of algorithm complexities, number of multitemporal images, number of features and classification accuracy as compared with [9,11]. Final classification map is shown in Fig. 6 derived from multi-temporal stacks.

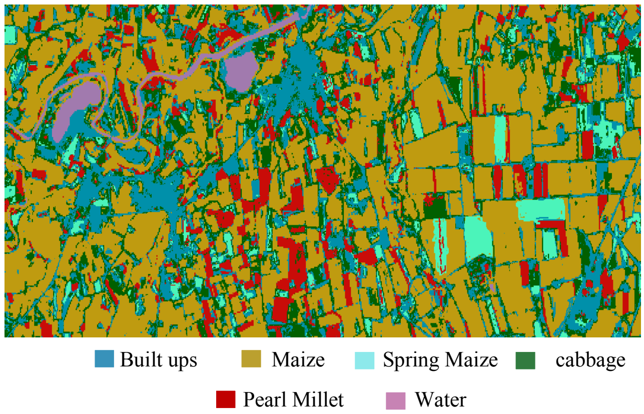


Fig. 6. Final classification map.

V- CONCLUSION

In this paper, open source sentinel-2 multispectral images were used for the land cover and crop classification for the small region in Italy. Crops phenological cycles were studied using temporal NDVI patterns and major crop phenological information and separability measure was used to select best multi-temporal images for the input of RF classifier. Classification with multi-temporal images achieved 91.2% OAA which is 6.7% higher than the best mono-temporal OAA. Results suggest that, considering phenological information about the major crops of study area can be very useful to select multi-temporal images for input of classifier to increase the accuracy and quality of the classification map.

Further, more features can be derived from the sentinel-2 data and may be exploited to increase the classification accuracy.

REFERENCES

[1] Cristina Gómez , Joanne C. White and Michael A. Wulderl, “Optical remotely sensed time series data for land cover classification: A review”, ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 116, pp.55–72, 2016.

[2] R. Khatami, G. Mountrakis, and S. V. Stehman, “A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: General guidelines for practitioners and future research,” Remote Sens. Environ., vol. 177, pp. 89–100, May 2016

[3] Andrew, M.E., Wulder and M.A., Nelson, T.A. “Potential contributions of remote sensing to ecosystem service assessments”. Prog. Phys. Geogr, Vol. 38 (3), pp. 328–353, 2014.

[4] B. D. Wardlow, S. L. Egbert, and J. H. Kastens, “Analysis of time-series MODIS 250 m vegetation index data for crop classification in the U.S. Central Great Plains,” Remote Sens. Environ., vol. 108, no. 3, pp. 290–310, Jun. 2007.

[5] K. Tatsumi, Y. Yamashiki, M. A. C. Torres, and C. L. R. Taipe, “Crop classification of upland fields using Random forest of time-series Landsat 7 ETM+ data,” Comput. Electron. Agriculture, vol. 115, pp. 171–179, 2015.

[6] Martino Pesaresi et al, “Assessment of the Added-Value of Sentinel-2 for Detecting Built-up Areas”, Remote Sens. Vol. 8, pp. 299-,2016.

[7] M. Lavreniuk, N. Kussul, S. Skakun, A. Shelestov, and B. Yailymov, “Regional retrospective high resolution land cover for Ukraine: methodology and results,” In 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 3965–3968, 2015.

[8] N. Matton, G. S. Canto, F. Waldner, S. Valero, D. Morin, J. Inglada, and P. Defourny, “An Automated Method for Annual Cropland Mapping along the Season for Various GloballyDistributed Agrosystems Using High Spatial and Temporal Resolution Time Series,” Remote Sensing, vol. 7, no. 10, pp. 13208-13232, 2015.

[9] Hao, P., Zhan, Y., Wang, L., Niu, Z. and Shakir, M., “Feature selection of time series MODIS data for early crop classification using random forest: A case study in Kansas, USA”. Remote Sens. Vol-7,pp. 5347, 2015.

[10] Petitjean, F., Weber, J., “ Efficient satellite image time series analysis under time warping”. IEEE Geosci. Remote Sens. Lett. 11, pp. 1143–1147, 2014.

[11] Skakun et al,” Efficiency Assessment of Multitemporal C-Band Radarsat-2 Intensity and Landsat-8 Surface Reflectance Satellite Imagery for Crop Classification in Ukraine”, Ieee Journal Of Selected Topics In Applied Earth Observations And Remote Sensing, Vol. 9, NO.8, 2016

[12] Sentinel-2 MSI Technical Guide, (https://sentinel.esa.int/web/sentinel/technical-guides/sentinel-2-msi)

[13] European Space Agency's (ESA)Sentinel Scientific Data Hub (https://scihub.copernicus.eu/).

[14] Belgiu, M., Drăguţ, L., “Random forest in remote sensing: a review of applications and future directions” ISPRS J. Photogramm. Remote Sens. Vol-114, pp. 24‑31,2016.

[15] Adam, E.; Mutanga, O. Spectral discrimination of papyrus vegetation (Cyperus Papyrus L.) in swamp wetlands using field spectrometry. ISPRS J. Photogramm. Remote Sens. 2009, 64, 612–620.