4th Intercontinental Geoinformation Days

igd.mersin.edu.tr

Crop mapping using Sentinel-1 and Sentinel-2 images and random forest algorithm

Ali Shamsoddini *1, Bahar Asadi 1

[1]Tarbiat Modares University, faculty of Humanities, Department of Remote Sensing and GIS, Tehran, Iran

Keywords
Optical and radar image fusion
Crop mapping
Red-Edge Band
Random Forest
Remote Sensing

Abstract

Crop mapping can provide valuable information for agricultural land management and crop estimation. This study investigated the spectral bands of Sentinel-2 time series, their vegetation indices, Sentinel-1 VV and VH time series of the radar backscatter coefficient, and the VV/VH ratio. This study explored the importance of the red-edge wavelengths of Sentinel-2 imagery for crop mapping using the random forest (RF) method. Therefore, the 2019 time series of Sentinel-1 and 2 images for the growing season in northwest Ardabil, Iran, were retrieved from the Google Earth Engine. After pre-processing, these images were segmented using the multi-scale method, and then the spectral features of optical imagery and the radar backscatter coefficient were extracted for each segment. To examine the importance and role of red-edge wavelengths, in addition to the three red-edge bands, visible and infrared wavelengths, and plant indices derived from these bands, red-edge indices were also factored in as input features. Overall, nine scenarios were simulated using different inputs and combinations. In each scenario, key features were identified using RF feature selection and introduced as inputs for the RF algorithm for an object-oriented classification. The research results showed that the addition of red-edge bands and the derived indices increased the accuracy of crop type mapping. The best result was obtained for a combination of optical and radar images with an overall accuracy of 87.59% and a kappa coefficient of 85.40%.

1. Introduction

The area of land used for cultivating different crops in a certain period is called the cropping pattern. Meanwhile, the agricultural production level is a decisive factor that influences the agricultural economy (Aduvukha et al., 2021).

Aside from the significance of temporal information, considering the unique phenological properties of different crops, it is crucial to identify and employ suitable data for specific dates in a time series in order to obtain more accurate results. Therefore, it is worthwhile to assess the relative importance of Sentinel-2 spectral bands for crop classification and particularly the performance of data from red-edge bands, which are missing in conventional sensors such as Landsat and SPOT and are less often studied. Indices used based on red-edge features are influenced by leaf structure, leaf chlorophyll, and leaf canopy cover and can acceptably distinguish crops (Zhong et al., 2014). Although the spectral reflectance of crops is affected by the vegetation cover status (e.g., chlorophyll content, pigmentation, canopy, and leaf water content) and provides valuable information for crop mapping (Zhang et al., 2017), plant structure information can also prove critical in this regard. When it comes to crops with similar phenological cycles, spectral information alone can hardly be enough for crop classification. Radar images, such as Sentinel-1 data, contain information about the vegetation structure, and so combined optical and radar images can complement each other and help to improve the results (Sun et al., 2020). Some studies confirm the high capability of radar images, especially cross-polarized imagery, in the separation of agricultural products (Ban et al., 2003; Skriver et al., 2011). Meanwhile, considering the differences in plant growth stages and spectral features of different crops during the growth period, various combinations of optical time series images and radar back-scattering coefficients were tested to assess the performance of different indices in crop mapping Several similar studies have used multi-time Sentinel-2 and Sentinel-1 images, whereas few have investigated the spectral features of the three red-edge bands of this sensor's optical images and their vegetation indices.

* **Corresponding Author**

*(ali.shamsoddini@modares.ac.ir) ORCID ID 0000-0002-4559-7563
(bahar.66asadi@gmail.com) ORCID ID 0000-0002-6355-7418

Cite this study

Shamsoddini, A., & Asadi, B. (2022). Crop mapping using Sentinel-1 and Sentinel-2 images and random forest algorithm. 4th Intercontinental Geoinformation Days (IGD), 103-107, Tabriz, Iran

Therefore, this study also aimed to assess the performance of Sentinel red-edge bands in crop classification through different approaches and combinations of optical and radar images.

2. Case Study and Research Data

The study area is located at 38.13 to 38.23 N and 48.16 to 48.9 E in northwestern Iran (Figure 1). Monocropping is dominant across the study area due to the region's climatic conditions, yet adequate precipitation makes it suitable for growing a variety of crops, including cereals, legumes, potatoes, and sugar beet.

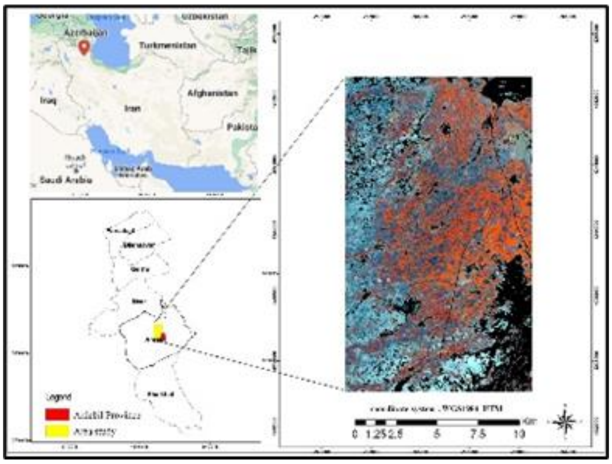


Figure 1. Location of the study area and its false-color image from Sentinel-2

The research data consisted of the region's crop type in 2019 collected through ground monitoring. These included alfalfa/sainfoin, wheat, barley, beans, corn, broad bean, Linaceae (flax), potatoes, and sugar beet, which make up the primary produce of the Ardabil plain. The fields' location and crop type were acquired through land survey and GPS measurements.

3. Methodology

The research flowchart is presented in Figure 3. After image collection and preprocessing (e.g., radiometric corrections and speckle noise reduction in radar images), the necessary features were extracted from the imagery. More detailed descriptions of each step are provided below.

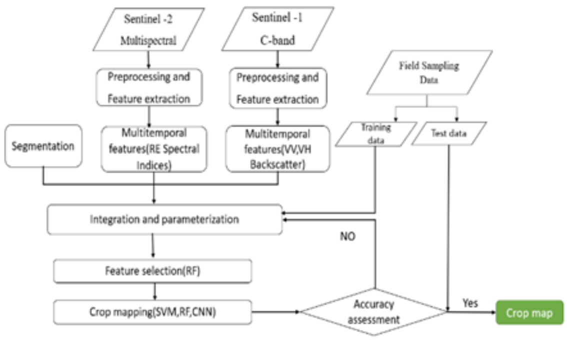


Figure 2. Conceptual diagram and research flowchart

This study used time series images of Sentinel-2 and Sentinel-1. Considering the features of different time series bands, 10 m and 20 m bands (specifically, bands 2–7, 8 and A8, and also 11 and 12) were selected here and the pixel size of the 20 m bands was downscaled from 20 m to 10 m. The crop phenological dates were obtained from the crop calendar and NDVI time-series curves to determine the suitable research period, and accordingly, L2A time series images for a two-month period (from June 18 to September 18) were retrieved for the study area using the Google Earth Engine (i.e., Sen2Cor atmospherically-corrected L2A images in GEE). The Sentinel-1 Ground Range Detected (GRD) time series images were retrieved from May 7 to September 28 using the GEE platform. Then the images were preprocessed, including conversion of images into sigma0 images in the logarithmic scale (dB) and reduction of speckle effects using the Enhanced Lee Filter.

3.1. Image Segmentation

Image segmentation means partitioning a digital image into multiple distinct regions, that is, grouping the image elements using specific homogeneity criteria (Meinel et al., 2001). The first stage in object-based classification is segmenting the image into meaningful parts, which was performed here using the popular multi-scale segmentation method (Salehi et al., 2017). Therefore, segmentation was done using the time series (10, 20 m) images, the first four (PCA1-PCA4) components of the time series of sentinel-2 image bands (showing the regional changes), and NDVI.

3.2. Feature Extraction

In the second stage, vegetation indices, spectral features, and the radar backscattering coefficient ratio were extracted (Table 1). A total of 27 vegetation indices based on visible, infrared, and red bands were calculated and the time series (10, 20 m) of optical images were selected as input features for classification. Features obtained from SAR images include the VV and VH back-scattering time series, the VV/VH back-scattering ratio, and the Normalized Ratio Procedure between Bands (NRPB) estimated using the VV and VH backscatter (Filgueiras et al., 2019).

3.3. Random Forest Algorithm

The random forest (RF) algorithm is a collection of decision trees that make output predictions by combining outcomes from all decision trees. To classify an input vector, it is submitted as an input to each of the trees in the forest, and the classification is then determined based on the class label with the majority vote (Rodriguez et al., 2012). The parameters that need to be optimized in this classifier are the number of trees and the number of features used in each tree. The former was obtained through cross-validation and out-of-the-box data analysis and the latter through the square root of the number of features (Belgiu et al., 2018).

Due to the large size of input features, the RF feature selection method was used to reduce the dimensionality of the input data (Shamsoddini et al., 2013).

Table 1. Research features

Sentinel-2 time series features	Spectral band time series (B2, B3, B4, B5, B6, B7, B8, B8A, B11, and B12) Vegetation indices, including, NDVI (Tucker, 1979b), ReNDVI (Fernández et al., 2016), MSRre (Chen et al., 1996), MCARI (Daughtry et al., 2000), IRECI (Frampton et al., 2013), TVI (Khosravi et al., 2018), Ndre1 (Gitelson et al., 1994), NDSAVI (Qi et al., 2002), NDRI (Gelder et al., 2009), NDWI (Sun et al., 2020) (192 features in total)
SAR time series features	1. VV and VH radar backscattering coefficient 2. VV/VH ratio, NRPB (Filgueiras et al., 2019) (48 features in total)

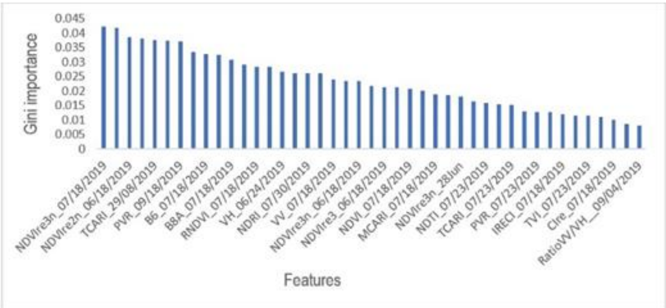
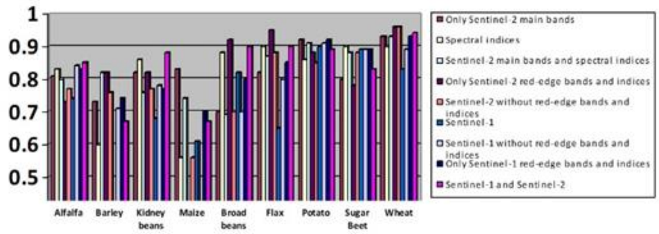


Figure 3. Feature importance results of combined Sentinel 1 and 2 imagery using RF

Table 2. Object-oriented classification results using different input data combination scenarios



3.4. Assessment of classification accuracy

The k-fold cross-validation method was used to prevent feature over-fitting in machine learning methods and optimize the RF algorithm parameters. Aiming to assess the performance of the classifier, the field data was split into training sample (70%) and testing sample (30%) sets, and different indices such as overall kappa accuracy, user accuracy, producer accuracy, and F1-Score were used.

4. Results and Discussion

Based on the Gini index (Figure 3), the most important features in crop classification using a combination of optical and radar imagery were the red-edge vegetation indices NDVIre3n (July 18), IRECI (Sept. 18), and NDVIre2n (June 18), respectively. These results demonstrate the importance of red-edge bands and the vegetation indices obtained from them in crop mapping, with the most effective bands being red-edge (especially bands B6 and B7), B8 near-infrared, and B8A near-infrared wavelengths. The first red-edge B5 band was more affected by the leave chlorophyll content than the other two red-edge bands. Many crops have a high chlorophyll content in the middle and peak stages of their growth and fruiting. Green and red bands are saturated in areas with high chlorophyll content, which limits their application in corp classification, whereas, Sentinel-2 6B band, with more sensitivity to chlorophyll concentration, can provide the necessary information (Gitelson et al., 1996).

According to Table 2, the best result was obtained for the combination of all the features obtained from optical and radar images through the object-oriented classification with an overall accuracy of 87% and a kappa coefficient of 85%. The results showed that RF-based feature selection in all different combinations of features extracted from Sentinel images increased the accuracy by 2-3% compared to the scenario of using all the features. Random forest is highly effective in reducing the effect of data redundancy in a time series and enhancing the classification accuracy with minimal image input (Nitz et al., 2015). This method enjoys a high training and classification speed and can effectively classify high-dimensional features (Joelsson et al., 2005). The advantages of the RF algorithm in this study included higher training speed, no overfitting, and compatibility with high-dimensional features (Kang et al., 2021). According to Table 2, classification using only the red-edge bands and their vegetation indices can acceptably separate crops; followed by the near-infrared (NIR) and short-wave infrared (SWIR) bands (Zahang et al., 2020). Moreover, using the red-edge bands along with other Sentinel-1 and Sentinel-2 bands improved the classification accuracy. Overall, red-edge bands performed better than visible wavelengths, mainly due to their higher capability to obtain plant biochemical information (Zahang et al., 2020). In Scenario 4, classification using red-edge bands and their vegetation indices was more accurate for crops such as barley, wheat, broad beans, kidney beans, and flax.

The selection of spectral bands and their vegetation indices influences the accuracy of crop classification (Orynbaikyzy et al., 2019). The accuracy of wheat, beet, potato, and broad bean classification was 083%, 89%, 90%, and 82%, respectively. For other classes, product classification based only on Sentinel-1 images was not adequately successful, likely because: (1) the radar back-scattering intensity is different for different crops based on their canopy cover, however, the radar incidence angle and soil surface moisture can affect the radar back-scattering, which makes it difficult to differentiate between crops with a similar canopy and appearance (Orynbaikyzy et al., 2019). (2) The speckle in radar images increases the inter-class diversity and decreases the separability of different classes, even though speckle is highly dependent on the surface properties and is not actually noise. On the other hand, speckles can affect the statistics and the distribution of pixel values in the feature space by increasing the variance of each class and the covariance between classes. This increases the separability between classes and possibly the classification error (Tavakoli, 2011). (3) The incidence angle leaves little room for differentiating certain crop types. Therefore, the limited viewing angle and radar data orbits make using only radar data insufficient for crop classification, especially when there are various crop types (Brisco, 1998). The higher the radar viewing angle, the higher the accuracy of crop separation (Brisco, 1998). According to the results, vegetation indices outperformed the main bands in crop classification. Specifically, the NDVIre3n, IRECI, and NDVIre2n indices proved highly effective. The data extracted from Sentinel-1 images alone did not provide enough accuracy, and so the classification accuracy was increased (by 6%) after including three red-edge bands and their vegetation index (Ndre, Cire, and MSRre). In this scenario, alfalfa, maize, and potato plants were separated more accurately than in other scenarios. The combined use of Sentinel-1 and 2 imageries in this study showed that optical images alone can accurately map crops. Considering that combining optical and radar images increases the data dimensionality, which hinders machine learning, this should be done in specific circumstances and based on the type of crops under study (Orynbaikyzy et al., 2019).

5. Conclusion

This study examined the performance of a random forest classifier with nine different scenarios in crop classification using input features extracted from Sentinel-1 and 2 imageries. The research results can be summarized as follows:

- Separation of crops with only optimal features from the RF classifier provided better results than classification using all features.
- In this study, indices based on red-edge, near-infrared, narrow-band near-infrared, and short-wave infrared (SWIR) wavelengths were more accurate. Overall, the inclusion of red-edge wavelengths increased the crop classification accuracy.
- The combined use of the Sentinel-1 and Sentinel-2 time series data had the best performance in crop separation.
- Red-edge vegetation indices showed optimal performance in identifying crops such as wheat, barley, broad beans, kidney beans, and flax.
- The research findings can inform decisions concerning the selection of spectral bands for achieving higher accuracy in crop classification.

References

Aduvukha, G. R., Abdel-Rahman, E. M., Sichangi, A. W., Makokha, G. O., Landmann, T., Mudereri, B. T., ... & Dubois, T. (2021). Cropping Pattern Mapping in an Agro-Natural Heterogeneous Landscape Using Sentinel-2 and Sentinel-1 Satellite D atasets. Agriculture, 11(6), 530.

Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.

Bégué, A., Arvor, D., Bellón, B., Betbeder, J., Abelleyra, D. D., Ferraz, R. P. D., ... & Verón, S. R. Remote Sensing and Cropping Practices: A Review. Remote Sens, 2018, 10 (1), 99.

Belgiu, M., & Csillik, O. (2018). Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. Remote sensing of environment, 204, 509-523.

Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. Remote sensing of environment, 37(1), 35-46.

Chong, L. U. O., LIU, H. J., LU, L. P., LIU, Z. R., KONG, F. C., & ZHANG, X. L. (2021). Monthly composites from Sentinel-1 and Sentinel-2 images for regional major crop mapping with Google Earth Engine. Journal of Integrative Agriculture, 20(7), 1944-1957.

Drăguţ, L., Tiede, D., & Levick, S. R. (2010). ESP: a tool to estimate scale parameter for multiresolution image segmentation of remotely sensed data. International Journal of Geographical Information Science, 24(6), 859-871.

Filgueiras, R., Mantovani, E. C., Althoff, D., Fernandes Filho, E. I., & Cunha, F. F. D. (2019). Crop NDVI monitoring based on sentinel 1. Remote Sensing, 11(12), 1441.

Haghshenas, N., Shamsoddini, A., & Aghighi, H. (2021). Urban Tree Canopy Mapping Using Object Oriented Classification and Machine Learning Algorithms. Iranian Journal of Remote Sensing & GIS, 13(1), 17-32.

Gitelson, A. A., Kaufman, Y. J., & Merzlyak, M. N. (1996). Use of a green channel in remote sensing of global vegetation from EOS-MODIS. Remote sensing of Environment, 58(3), 289-298.James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning (Vol. 112, p. 18). New York: springer

Joelsson, S. R., Benediktsson, J. A., & Sveinsson, J. R. (2005, July). Random forest classifiers for hyperspectral data. In Proceedings. 2005 IEEE International Geoscience and Remote Sensing Symposium, 2005. IGARSS'05. (Vol. 1, pp. 4-pp). IEEE.

Kang, Y., Meng, Q., Liu, M., Zou, Y., & Wang, X. (2021). Crop Classification Based on Red Edge Features Analysis of GF-6 WFV Data. Sensors, 21(13), 4328

Nitze, I., Barrett, B., & Cawkwell, F. (2015). Temporal optimisation of image acquisition for land cover classification with Random Forest and MODIS time-series. International Journal of Applied Earth Observation and Geoinformation, 34, 136-146.

Meinel, G., Neubert, M., & Reder, J. (2001). The potential use of very high resolution satellite data for urban areas—First experiences with IKONOS data, their classification and application in urban planning and environmental monitoring. Regensburger Geographische Schriften, 35, 196-205.

Orynbaikyzy, A., Gessner, U., & Conrad, C. (2019). Crop type classification using a combination of optical and radar remote sensing data: a review. international journal of remote sensing, 40(17), 6553-6595

Pal, M., & Mather, P. M. (2004). Assessment of the effectiveness of support vector machines for hyperspectral data. Future generation computer systems, 20(7), 1215-1225.

Shamsoddini, A., J. C. Trinder, and R. Turner. 2013. "Non-Linear Methods for Inferring Lidar Metrics Using SPOT-5 Textural Data." ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences II-5/W2: 259–264. doi:10.5194/isprsannals-II-5-W2-259-2013.

Salehi, B., Daneshfar, B., & Davidson, A. M. (2017). Accurate crop-type classification using multi-temporal optical and multi-polarization SAR data in an object-based image analysis framework. International Journal of Remote Sensing, 38(14), 4130-4155.

Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. ISPRS Journal of Photogrammetry and Remote Sensing, 67, 93-104.

Zhong, L., Gong, P., & Biging, G. S. (2014). Efficient corn and soybean mapping with temporal extendability: A multi-year experiment using Landsat imagery. Remote Sensing of Environment, 140, 1-13.

Zhang, T., Su, J., Liu, C., Chen, W. H., Liu, H., & Liu, G. (2017, September). Band selection in Sentinel-2 satellite for agriculture applications. In 2017 23rd international conference on automation and computing (ICAC) (pp. 1-6). IEEE.

Zhang, H., Kang, J., Xu, X., & Zhang, L. (2020). Accessing the temporal and spectral features in crop type mapping using multi-temporal Sentinel-2 imagery: A case study of Yi'an County, Heilongjiang province, China. Computers and Electronics in Agriculture, 176, 105618.