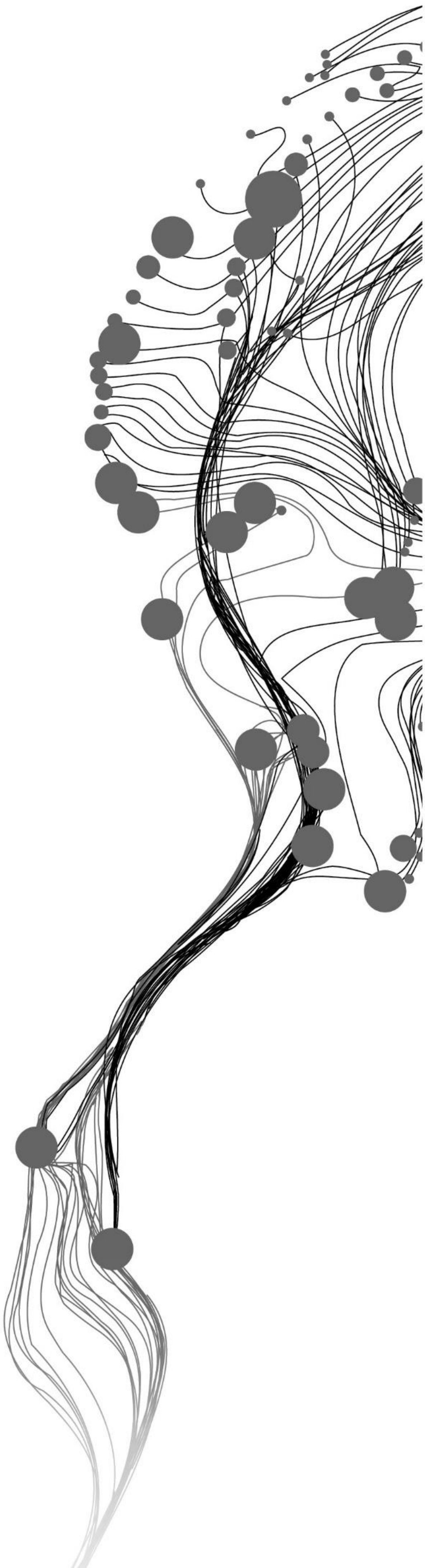
SAFFRON FIELD CLASSIFICATION AND FLOWERING PHENOLOGY DETECTION USING SENTINEL-2 TIME SERIES IN TORBAT-E HEYDARIYEH, IRAN

KEKE DUAN
June 2020

SUPERVISORS:
Dr. A. Vrieling
Dr. R. Darvishzadeh

ADVISOR:
Dr. H. Kaveh



SAFFRON FIELD CLASSIFICATION AND FLOWERING PHENOLOGY DETECTION USING SENTINEL-2 TIME SERIES IN TORBAT-E HEYDARIYEH, IRAN

KEKE DUAN
Enschede, The Netherlands, June 2020

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Spatial Engineering

SUPERVISORS:

Dr. A. Vrieling

Dr. R. Darvishzadeh

ADVISOR:

Dr. H. Kaveh (University of Torbat Heydarieh; Saffron Institute)

THESIS ASSESSMENT BOARD:

Dr.Ir. C.A.J.M. (Kees) de Bie (Chair)

Dr. M. Belgiu (External Examiner, University of Twente)

ABSTRACT

Saffron (*Crocus sativus L.*) is the most expensive spice worldwide and has high medicinal value, which increases the demand of the global saffron market. However, saffron yield reduced in recent years due to climate shifts and the gradual lowering of groundwater tables. Identifying and monitoring the saffron field changes is of exceptional importance for effective agronomic management for local agricultural sectors and farmers. Frequently acquired satellite images with 10m spatial resolution, such as Sentinel-2 (S2) with five days revisit interval, are able to provide more available cloud-free observations and capture more distinct temporal characteristics for effective classification and monitoring. The objectives of this study are to evaluate the utility of S2 time series in accurately mapping saffron field distribution, classifying different age of saffron crops, and detecting saffron phenological behaviour. The study area, Torbat-e Heydariyeh, is a county famous for its saffron cultivation in Khorasan Province, Iran. In-situ data were collected during a two-week field survey in December 2019. To separate saffron from other land covers, first, 252 spectral-temporal features were derived from the S2 images, and Random Forest (RF) was used to select a subset of variables with high importance value to achieve optimal accuracies of saffron field identification. To evaluate the feasibility of discriminating different age saffron, the separability between saffron fields with different age were analysed. RF was then conducted to evaluate the classification accuracy of grouped age classes. Apart from the vegetation period, peak flowering is also a critical phenological stage of saffron that can directly reflect the yield level of saffron fields. The customized Enhanced Blooming Index (EBI_c) aimed to enhance the purpleness of saffron flowers and reduce background noise from soil and green vegetation. Selected spectral features were mostly vegetation indices that incorporate spectral information from red, NIR, and SWIR bands. Two phenological phases were identified important in separating saffron from other crops, which are the rapid green-up stage (January to March) and dormant period (August to September). Pixel-based RF achieved good classification accuracy (overall accuracy of 95.3%, kappa coefficient of 0.93) in discriminating saffron with other crops using multi-temporal S2 imagery. Based on an independent in-situ dataset on saffron fields, 87.4% of the existing fields were correctly classified as being saffron. Five saffron field age groups could be well discriminated based on their spectra-temporal characteristics, i.e. 1st year, 2nd year, 3rd year, 4th-6th year, and 7th-8th year groups showed high separability. The age-based classification result presented an overall accuracy of 86.8% using NDVI time series from December to May. Merged two age groups outperformed individual age classes. The S2 derived peak flowering date agreed well (R^2 = 0.69, RMSE = four days) with surveyed crop calendar data of 46 fields. Overall, this study demonstrated the potential utility of S2 time series data for accurately mapping saffron field distribution, age classification, and flowering phenology detection. These findings provide a basis for further investigation in upscaling the study area in a larger extent, and monitoring changes of saffron distribution and phenology in the spatial and temporal patterns.

Keywords: Sentinel-2 time series, Random Forest, spectral-temporal features, age-based classification, EBI, peak flowering

ACKNOWLEDGEMENTS

First, I would like to express my sincere gratitude to my supervisors, Dr. Anton Vrieling and Dr. Roshanak Darvish, for their great support, guidance, and encouragement throughout my research period. Their valuable advice really helped me a lot to improve my research skills. They provided their support even during non-working hours and weekends. My heartfelt to them for that.

I would also like to thank my advisor Dr. Hamed Kaveh and other staff in the Saffron Institute for their all-out assistance and meticulous care during the fieldwork in Torbat-e Heydariyeh, Iran.

Many thanks to ITC Foundation Scholarship Programme for supporting me to pursue my MSc study at ITC.

My special thanks to my parents and sister for their unconditional support and encouragement throughout the MSc period. Last but not least, I would like to express my thanks to my boyfriend, Rui Xie, for his patience and accompany all the time to give me endless happiness.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1. Background

Saffron (*Crocus sativus L.*) is an autumnal-flowering perennial geophyte whose dried scarlet stigmas are known as the costliest spice and have been dubbed “the red gold” (Basker & Negbi, 1983; Negbi, 1999; Winterhalter & Steaubinger, 2000; Fernandez, 2004). The main reason for its high price is the low productivity and the need for intensive labour for cultivation, harvesting, and processing (Kumar et al., 2009; Ghorbani & Koocheki, 2017). Since time immemorial, saffron has been highly valued for its flavouring, colouring, aromatic capacity, as well as medicinal function for analgesia and sedation (Winterhalter & Steaubinger, 2000; Gresta et al., 2016; Kumar et al., 2009; Lichtfouse, 2013). Recent studies are boosting interests in its latent medicinal value, especially in cytotoxic, antitumor, and anticarcinogenic properties (Abdullaev & Frenkel, 1999; Fernandez, 2004; Gresta et al., 2016). Due to its edible and medicinal value, saffron has attracted considerable new generation consumers, which increases its global market demand (Gresta et al., 2008; Kumar et al., 2009).

Iran is the major producer of saffron products worldwide. It accounts for more than 90% of the world saffron production, and 60% of the total saffron cultivation area (Ghorbani, 2007), most of which is found in its north-eastern Khorasan region (GIAHS, 2018). This is not only due to its high commercial value, but also because of the suitability of local climate conditions for saffron growth. The cultivation of saffron has very low fertilizer and water requirements. The daughter corms can survive inside the soil during the hot summer due to its heat-tolerant characteristics. The low input requirements of saffron also makes it an alternative viable crop for organic and low-input cropping systems, which is able to provide promising production for sustainable agriculture. In the last 30 years, an enormous increase in saffron cultivation area has been registered in Iran. As a result, saffron cultivation has greatly stimulated the development of the local economy and created vast job opportunities for about 400,000 people in the region. These jobs effectively reduce the general depopulation trend of rural Iran by providing sustainable livelihoods (Esmaeilpour & Kardavani, 2011).

Saffron yields show large spatial and interannual variability, which is influenced by field age, agronomic management, and environmental factors. The dried stigma yield can vary from 1.5 to 15 kg/ha; they are relatively low in the first year and increase to the maximum level in the third or fourth year. After that, production declines because of reduced size and reproduction capability of corms, as well as increasing competition between overcrowded corms for water and nutrients (Kumar et al., 2009). Besides, the corm dimension and sowing time also influence flower production. Research showed that larger corm size and earlier sowing time generally have a positive effect on stigma production (De Mastro & Ruta, 1993; Negbi et al., 1989; Gresta et al., 2008). In addition, appropriate crop management (e.g. irrigation, weeding, etc.) is also important for improving saffron production. Irrigation before saffron flowering directly influences the anthesis duration, flower amount, and stigma quality. Due to the short height and narrow leaves of saffron vegetation, weeds are important competitors for contesting nourishment, sunlight, and water (Esmaeilpour & Kardavani, 2011; Ghorbani & Koocheki, 2017). Moreover, the production of saffron is sensitive to climate change. Hosseini et al. (2008) found that saffron yield variation in Iran is associated with temperature and precipitation, and given the expected climatic shifts a further yield decrease is expected. Zahmati et al. (2018) and Molina et al. (2005) stated the appropriate temperature for daughter corm producing and flower initiation is 23 to 27°C during late spring, while flower emergence needs a temperature below 16°C during autumn. Winter chilling with ten days below 8°C is also required for daughter corm development. Predicted drought (Daneshvar et al., 2019; IPCC, 2014) requires more intensive irrigation using groundwater, which will cause gradual lowering of the groundwater table (Motagh et al., 2008). These stresses from environmental changes threaten the sustainable production of saffron and may eventually demand geographic shifts or reduction in cultivation area to locations where the (new) climate condition will be more suitable for saffron growth.

To better understand the impacts of environmental change on saffron cultivation and productivity, identifying and monitoring the saffron field changes is of exceptional importance for effective agronomic management for local agricultural sectors and farmers. However, detailed spatial and temporal saffron cultivation assessments at the landscape scale hardly exist. The first step should devote to improving knowledge on accurate location and area information of saffron cultivation.

In Iran, the documentation of crop types, location, and area, is traditionally performed using ground-based agricultural surveys under the responsibility of two major organizations, i.e. the Ministry of Jihad-e-Agricultural (MOJA) and the Statistical Centre of Iran (SCI) (Mehrdad, 2014). Given the impossibility to cover all farms and fields during these surveys, and the lack of georeferenced field level information, time-efficient accurate methods are needed to better understand crop distribution.

Satellite imagery has become a promising alternative data source for mapping and monitoring cropland at large scales (Wardlow et al., 2007; Xie et al., 2019). Since the last decades, optical satellite images have been widely used for mapping croplands using different classification algorithms. Recent research used image time series to monitor crop growth by providing precise and timely information on the phenological performance and growth status of vegetation. Moderate Resolution Imaging Spectroradiometer (MODIS) satellite has proven to be useful for large-area cropland classification by providing near-daily global coverage of free intermediate resolution (250m) data (Lobell & Asner, 2004; Wardlow et al., 2007; Pittman et al., 2010; Dheeravath et al., 2010). However, it is not suitable for saffron field detection because a large number of saffron fields in Iran cover small areas (0.05-1.0 ha) (Behdani et al., 2009).

The new generation of Landsat and Sentinel-2 (S2) satellites with a fine spatial resolution (10-30m) provide a better option for this purpose. Landsat 8 imagery has a spatial resolution of 30 meters and a revisiting time of 16 days. Its utility for cropland classification has been reported by literatures (Badhwar et al., 1987; Zhong et al., 2014; Turker & Arikan, 2005). The S2 constellation provides satellite images with higher spatial-temporal resolution. Taking advantage of fine-spatial resolution (10-60m) Multi Spectral Instrument (MSI) with 5-day revisit interval, S2 has been increasingly used in cropland mapping and phenology detection studies in recent years (Gómez et al., 2016; Belgiu & Csillik, 2018; Vrieling et al. 2018; Stendardi et al., 2019). Despite advances that have been made with crop mapping at fine spatial detail using 10-30 m optical satellites, at present no precise spatial information is available on saffron cultivation areas and its changes.

Until present, only three saffron studies have been conducted using remote sensing technology (Rahimzadegan & Pourgholam, 2017; Dehghani Bidgoli et al., 2018; Farzadmehr & Bajestani, 2018). They all devoted to mapping saffron distribution with single-date or two dates Landsat 8 imagery at the county level. The classification result showed an overall accuracy ranging between 82% to 95%. However, these studies are reported in Persian and only English abstract is available. Nonetheless, they have suggested the potential of satellite imagery for saffron mapping, but there seems scope for improvement. Firstly, even though Landsat imagery has a relatively high spatial resolution (30m), the classification accuracy still not satisfactory for small fields (overall accuracy of 62% for fields under 0.2 ha, 72% for fields between 0.2 and 0.5 ha) (Dehghani Bidgoli et al., 2018). Secondly, because the selected images (acquired in January and May) correspond to the start and end of the vegetation period of saffron, it is easy to confuse saffron and winter wheat, which have a similar fractional cover of green vegetation for this two dates (Rahimzadegan & Pourgholam, 2017). Besides, the long revisit interval (16 days) of the Landsat 8 satellite lead to an insufficient number of cloud-free observations to capture detailed phenological behaviour, especially during seasons

(January to April) with persistent cloud cover in Iran. However, as this period is exactly the vegetation stage of saffron, more temporal data during this period would help detect distinct spectral-temporal behaviours of saffron. It can be beneficial for improving mapping accuracy and analysing the intra-class spectral variability of saffron.

Taking advantage of the short revisit interval (5 days) of S2 imagery, more temporal data would be available to produce a dense time series. Saffron's particular phenology in combination with the high temporal density of S2 should allow for effective classification and monitoring purposes. To improve the mapping accuracy, enhancing the understanding of the temporal behaviour of saffron phenology and cultivation practices is the first step. It also has great implications for monitoring phenological dynamics as a response to environmental changes. Since saffron crop age and its yield are interdependent, it is necessary to know the distribution and area under different age groups. Meanwhile, monitoring saffron phenological performance, such as the flowering period also provides important information for effective crop management and yield assessment.

1.2. Research objectives

The aim of this study is to evaluate if S2 time series allow for accurate mapping of saffron fields, classification of saffron crop age, and monitoring its phenological behaviour in Torbat-e Heydariyeh county in northeast Iran. Extending from the main research aim, the following objectives [O] and hypotheses [H] are defined:

[O1] To develop and assess a spectral-temporal feature selection method for identifying saffron fields from S2 time series.

[H1-1] Saffron fields display specific temporal variability in spectral reflectance, which relates to its phenology and is significantly distinct from other land covers.

[O2] To map saffron fields for the study area based on selected spectral-temporal features and assess the map accuracy.

[H2-1] A Random Forest (RF) classification model that incorporates key spectral-temporal features of saffron can provide maps with at least 90% overall accuracy.

[O3] To analyse if time series of vegetation indices (VIs) allow differentiating different ages of saffron cultivation.

[H3-1] Similar temporal behaviour but with increasing magnitude of the VI peak values during the vegetation stage is expected to be observed with the increase of saffron field age (1-4 years old) on multi-year time series.

[H3-2] Old saffron fields (6-8 years old) are hard to be distinguished between each other due to a similar vegetation cover.

[O4] To demonstrate the possibility of detecting the peak of saffron flowering from S2 time series and characterize temporal variation of reflectance in relation to management.

[H4-1] The retrieved peak of saffron flowering from S2 time series shows on average no more than 5 days difference compared with the peak flowering date obtained from in-situ survey data.

2. MATERIALS AND METHODS

This part consists of seven sections. The study area, field data collection, and satellite data preprocessing are described in Section 2.1, 2.2, and 2.3, respectively. In Section 2.4, spectral-temporal features of known saffron fields were identified from S2 time series. Subsequently, these features were used as input to a RF classifier in order to map saffron fields in the study area (Section 2.5). Then, it was analysed if S2-derived spectral-temporal information allows to differentiate different ages of saffron fields (Section 2.6). Finally, saffron peak flowering phenology was retrieved from S2 time series and compared with in-situ survey data (Section 2.7).

Image processing, data analysis, result assessment, and visualization were realized in Jupyter notebooks with Python 3.7. Other platforms and tools were also used as auxiliary means, such as QGIS for data preview and mapping, and MATLAB for statistical analysis.

2.1. Study area

The study area is located in the eastern region of Torbat-e Heydariyeh county, Razavi Khorasan Province in northeast Iran (*Figure 1*). It covers an area of 506 km^2. The area has a cold semi-arid steppe climate (BSk) according to the Köppen-Geiger classification (Peel et al., 2007). The annual mean temperature and precipitation are 26 °C and 236 mm respectively. *Figure 2* shows the average monthly weather data which is characterized by a mild winter, rainy spring, and hot dry summer.

The county is famous for its saffron production and is first in Iran in terms of saffron production area. Saffron covers more than 30% of the total planted area of the county due to its characteristics of high returns and low water requirement compared with other crops. Saffron corms are usually lifted from the soil and planted in new fields during real dormancy (in May to the end of June) or pseudo-dormancy (from early July up to August) stages. After planting, corms stay in the same field for five to eight years during which saffron remains productive. The timing of the first irrigation in autumn (i.e. pre-flowering irrigation) is a critical determinant of flowering time and saffron yield (Sepaskhah & Yarami, 2009). This first irrigation should take place at the moment when air temperatures are declining (below 16 °C) in order to optimally replenish soil moisture and reduce evaporation losses. About one week after the pre-flowering irrigation soil ploughing is performed in order to loosen the soil, i.e. to assist the emergence of the flowers (Koocheki & Khajeh-Hosseini, 2019). Flowering starts between two and three weeks after irrigation when air temperatures are around 12 °C, which normally is between mid-October and early November (Alizadeh et al., 2009). When most flowers bloom and have red stigmas, farmers start to harvest the flowers manually, and the most precious part (the stigmas) are subsequently separated and dried. The vegetative stage starts immediately after flowering from the end of November and lasts until late May (Koocheki & Khajeh-Hosseini, 2019). During this stage, the leaves reach maturity and provide necessary supplies for corm development through photosynthesis. Irrigation is performed four to five times during this phase to improve the corm number, yield, and nutrient uptake (KHAZAEI et al., 2013). -In June, the leaves start to senesce, and the daughter corms become dormant to prepare for the new growing season. During the dormant phase when temperatures are high and soils are dry, irrigation is not recommended to reduce mite population and avoid corm infection (Behdani & Fallahi, 2015). Saffron is cultivated in the same fields for five to eight growth seasons. However, one well-known problem of the perennial crop is the reduction of soil fertility and saturated density of corms in long term cultivation, which can ultimately lead to soil overexploitation and dramatic yield reduction (Gresta et al., 2016). To solve this problem, farmers usually leave the field fallow or grow other crops for multiple years. Apart from saffron, other main crops in the county include barley, winter wheat, spring wheat, pistachio, and alfalfa.

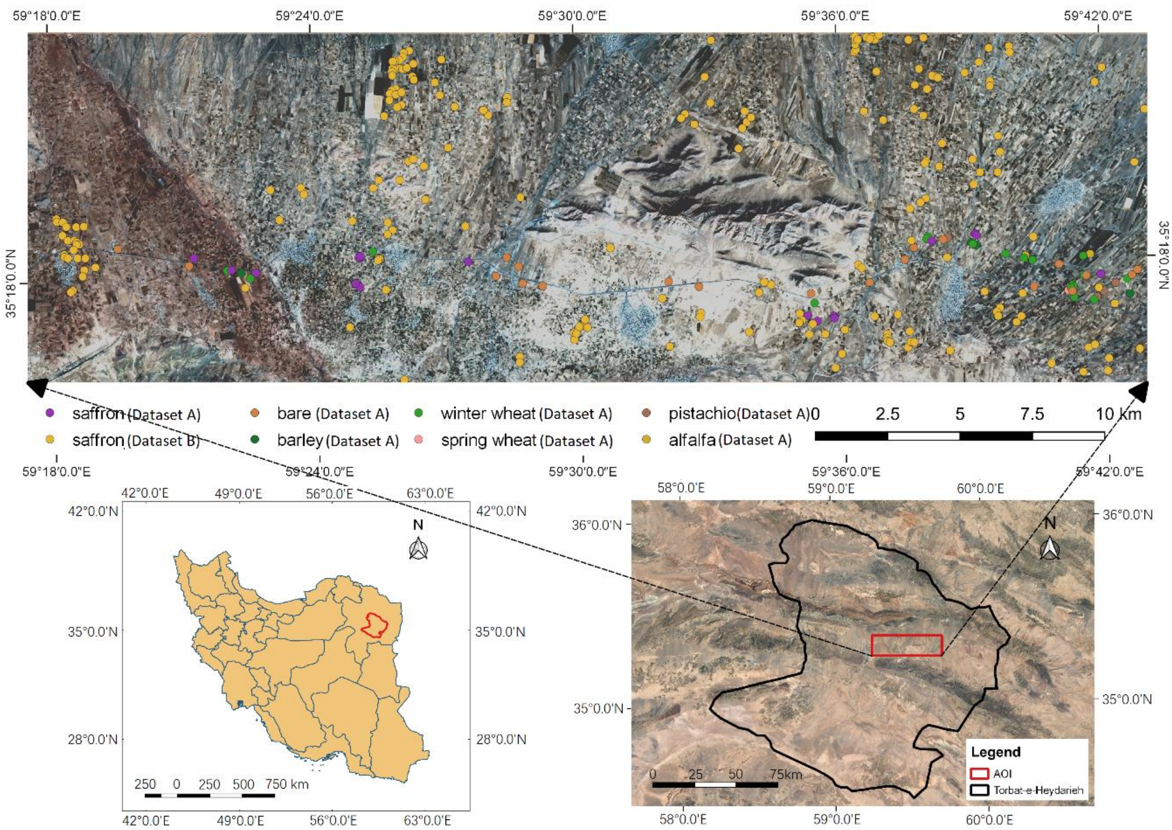


Figure 1. Overview of the study area and the sampled fields (coloured dots). The background shows a Sentinel-2A image acquired on 26th Sep 2018 with a RGB composite of NIR, Red, and Green band.

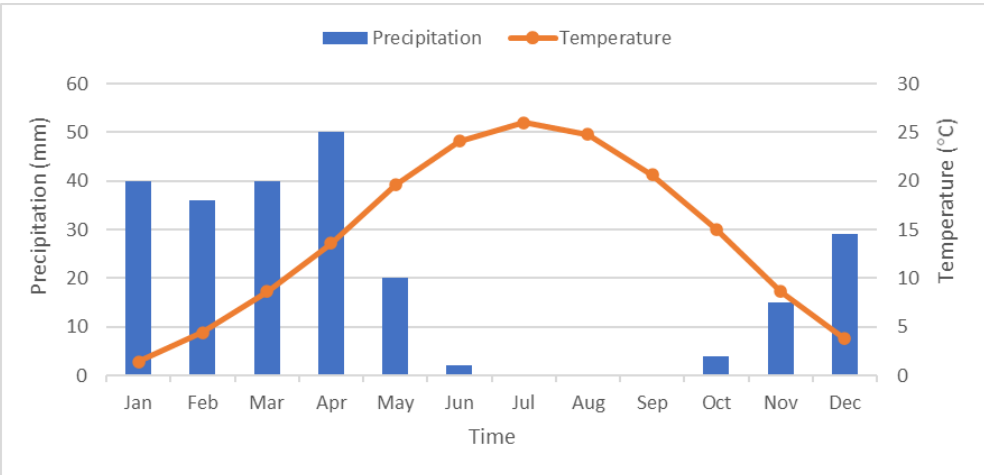


Figure 2. Average monthly temperature and precipitation in Torbat-e Heydariyeh as derived from weather data of the past 30 years (January 1982 -December 2012)

2.2. Field data collection

In-situ data were collected during a two-week field survey that took place between 9 and 23 December 2019, i.e. when saffron was at the vegetative stage). Considering that the study area is located between two mountain chains that run in east-west direction, the survey route was designed along the main road that runs in the same direction and spans a total length of 42 km. The in-situ survey data include parcel location information for fields with saffron and with other crops, saffron field age information, and phenological stages related farming calendar information such as irrigation date, flowering and harvesting period. More details about the survey data are explained below.

2.2.1. Crop field location data

The field location data (a total of 119 fields, named Dataset A) collected through field surveys were used as training and test samples for feature selection and classification (*Spectral-temporal feature selection for saffron identification*). Geographic coordinates were recorded for the corners that delimit the boundaries of each field using a Garmin ETrex 30x GPS device. The EPSG: 32640 (WGS 84/ UTM 40N) coordinate system was used throughout the survey. The dataset includes 30 saffron fields, 19 winter wheat fields, seven spring wheat fields, 11 pistachio fields, four barley fields, four alfalfa fields, and 44 bare soil areas. The area of these field samples varied between 0.7 and 22 ha.

In addition, a field survey conducted by the student Halimeh Eslahi from the University of Torbat Heydarieh between June 2018 and February 2019 provided additional locational information for 213 saffron fields. This dataset (named Dataset B) was used to assess the accuracy of the constructed saffron map (*Saffron field mapping and accuracy assessment*). The distribution of these sample fields is shown in *Figure 1*.

2.2.2. Saffron crop calendar

To better understand the timing of farming practices related to saffron cultivation, a number of farmers were interviewed during the fieldwork. Since few farmers were working in the fields during the survey period, most farmers were contacted prior to the field visit with the help of staff (Dr. Hamed Kaveh and Dr. Ali Salariyan) from the Saffron Institute. In total, 30 farmers were interviewed directly at their saffron fields, resulting in field-specific survey information on crop calendars and practices was collected. The dataset includes information about the 2016 to 2019 saffron growing seasons, and includes irrigation frequency and dates, ploughing date, and peak flowering period (which directly corresponds to harvesting period). It has to be mentioned that during the interview, some farmers cannot remember the exact date of these operations for years prior to 2018. To avoid incorporation of recall error and guarantee data quality, only information for the last two growing seasons (2018-2019) was used in subsequent analyses.

2.2.3. Saffron field photography and age information

Saffron is a perennial plant that is generally cultivated in the same field for five to eight years in the study area (Gresta et al., 2008). To perform age-based classification, planting year information for the 30 saffron fields in Dataset A were also recorded during the farmer interviews. Besides, during the field survey, nadir photos were taken from a fixed height (around 1.3 meters) at multiple positions within each saffron field. These photos are able to reflect the vegetation density of different age saffron fields during late December. *Figure 3* shows some examples of the photos; it shows that the older the saffron field, the higher the vegetation density. This performance was confirmed by farmers during the interviews. Besides, this information also contributes to better understanding the spectral differences between different age saffron fields at a certain time on VI time series.

2.3. Sentinel-2 data acquisition and preprocessing

The S2 mission comprises two polar-orbiting satellites, S2A and S2B, which share the same orbit but phase at 180° (ESA, 2015). They were launched by the European Space Agency (ESA) Copernicus programme on 23 June 2015 and 7 March 2017, respectively. The combination of the two satellites provides a five-day revisit interval for the same location on the earth surface. Areas covered by overlapping orbits have an even shorter revisit time, but this is not the case of Torbat-e Heydariyeh. Each S2 satellite carries the multi spectral instrument (MSI) with 13 spectral bands at 10-60 m spatial resolution.

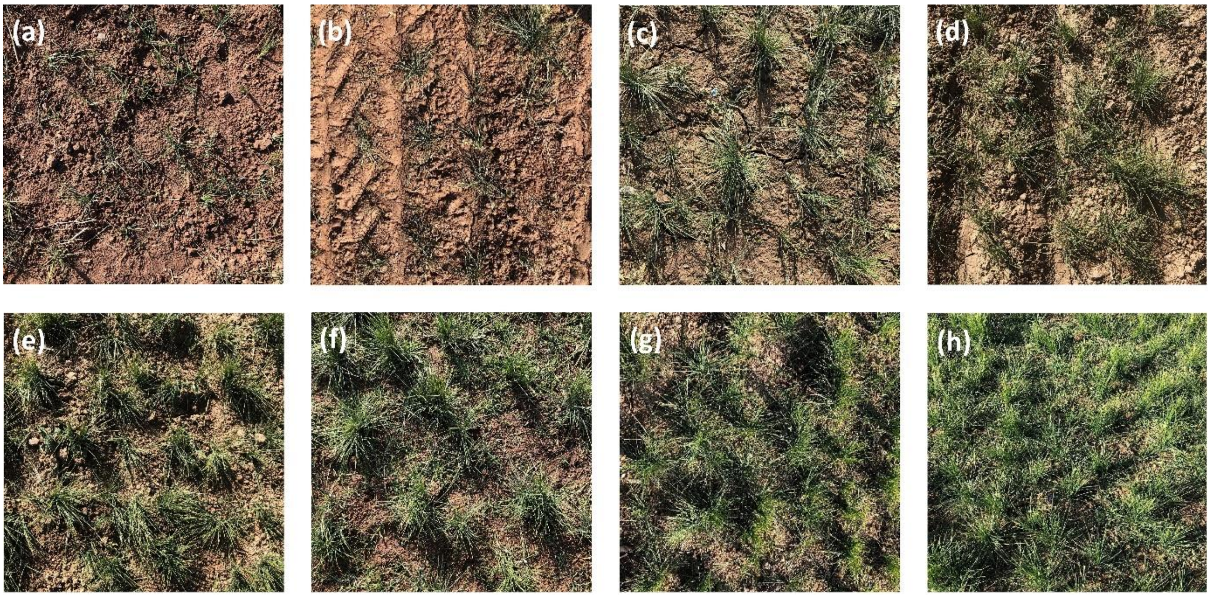


Figure 3. Nadir pictures of saffron fields with different age crops taken from 9th to 15th December 2019. (a)-(h) refers to 1 to 8 years age saffron crop respectively

A set of 259 Level-1C S2 images for tile 40SGE between 20 August 2015 and 20 February 2020 were downloaded from the Copernicus Open Access Hub (*https://scihub.copernicus.eu*) and USGS Earth Explorer website (https://earthexplorer.usgs.gov). This study did not directly download and use the Level-2A Bottom-Of-Atmosphere (BOA) product processed by ESA because it is only available from 15 December 2018 for the test region, and consequently does not match the full-time frame of this study. The Sen2Cor processor (version 2.8) was used for atmospheric correction of each single date Level-1C Top-Of-Atmosphere (TOA) product. One of the outputs of Sen2Cor is the 20m resolution Scene Classification (SCL) layer, which was used to mask out clouds, cloud shadows, snow, and other noises by aggregating seven classes, i.e. no data (0), saturated or defective (1), dark area (2), cloud shadow (3), cloud medium probability (8), cloud high probability (9), and snow (11) with a 20 m buffer. The masked images of 20 m resolution bands (*Table 1*) were resampled to 10 m, and then all these masked images were clipped to the study area extent. Besides, image stacks were made for each band image and prepared to extract spectral time series.

Table 1. A summary of the ten spectral bands in S2 satellite data used in this study

Band name	Central wavelength (nm)	Resolution (m)	Band name	Central wavelength (nm)	Resolution (m)
Band 2	490 (Blue)	10	Band 7	783 (Red Edge)	20
Band 3	560 (Green)	10	Band 8	842 (NIR)	10
Band 4	665 (Red)	10	Band 8A	865 (Red Edge)	20
Band 5	705 (Red Edge)	20	Band 11	1610 (SWIR)	20
Band 6	740 (Red Edge)	20	Band 12	2190 (SWIR)	20

2.4. Spectral-temporal feature selection for saffron identification

To select which spectral-temporal features are typical for saffron fields, a feature selection framework was applied for accurately identifying saffron fields using satellite time series. Firstly, possible spectral-temporal features were extracted from S2 time series. Secondly, the RF feature importance score was used as the basis for selecting a subset of features that achieved optimal accuracies in separating saffron fields from other crops or land cover.

2.4.1. Feature extraction

Feature extraction is the fundamental step of image-based classification to transform satellite images into useful information for a specific purpose, such as crop classification. Previous research has demonstrated that multi spectral time series data allowed for more accurate crop classification as compared with the use of single-date imagery (Chang et al., 2007; Zhong et al., 2014; Arvor et al., 2011). Multitemporal information can capture the seasonal dynamics and phenological stages, which often differ between crop types (Foerster et al., 2012; Pan et al., 2012). Such dynamics can be expressed differently depending on the spectral domain considered (Gitelson et al., 2002; Jiang et al., 2006). S2 MSI provides information in spectral regions that are sensitive to crop characteristics, such as the visible and near-infrared (VNIR) bands, which are related to leaf pigments, and short-wave infrared (SWIR), which is influenced by water content and non-photosynthetic components (Peña-Barragán et al., 2011). Moreover, vegetation indices (VIs) combine information from spectral bands to enhance the vegetation signal and as such are useful in crop classification. They are capable to characterize the crop behaviour during different phenological phases in relation to vegetation status, residue cover, and canopy structures (Bégué et al., 2011; Huete et al., 2002).

In this study, to reduce the feature dataset volume, multiple satellite acquisitions in a single month were summarized to average monthly spectra or VI per pixel (i.e. 12 temporal features per year). The spectral features consist of spectral reflectance of 10 individual bands (*Table 1*) and 11 VIs (*Table 2*), which have been previously used in crop mapping studies to detect the different biochemical and physical properties of specific crops. The 11 VIs were divided into six groups according to their differences in spectral regions used. Group A to F are the combination of spectral regions in visible and NIR, RedEdge and NIR, NIR and SWIR, SWIR and SWIR, visible and visible bands, respectively.

2.4.2. Feature selection

The combination of 21 spectral features for each of the 12 months results per year in a high dimensional dataset of 252 spectral-temporal features, which may increase the computation time with little improvement in classification accuracy (Xie et al., 2019; Hao et al., 2015; Löw et al., 2013; Hu et al., 2019). To remove redundant and irrelevant variables prior to image classification, various feature selection (variable elimination) methods have been developed and applied to remote sensing data (Hao et al., 2015; Hu et al., 2019; Howard & Wylie, 2014; Löw et al., 2013; Loosvelt et al., 2012). In this study, the RF feature importance score was used as the basis for selecting a subset of variables to achieve optimal accuracies of saffron field identification. Pal & Foody (2010) showed that RF feature importance score ranking has competitive performance for feature selection compared with other algorithms. The main advantages of RF compared with other algorithms are its ability in handling high dimensional input variables and its robustness to over-fitting. The final importance of each variable is determined by the permutation importance, which is calculated based on the increase of misclassification rate after permuting a certain feature (Goldstein et al., 2011).

Sample data was generated from the pixels in the 119 surveyed fields (Dataset A). To reduce the influence of heterogeneous pixels around the field edge, a 10 m negative buffer was created for each field parcel. Each sample is a combination of 1) the spectral-temporal information (i.e. 252 features) in the pixel, and b) the known crop type or land cover corresponding to that information. Two free parameters in RF were optimized: the number of trees was set at 500 to allow the convergence of out-of-bag error (OOB error) statistics; the number of features to split the nodes was set as the square root of a total number of input features as commonly recommended to decorrelate the trees (Gislason et al., 2006; Belgiu & Drăgu, 2016). The classifier was repeated for 50 times. Each time 70% of the 3797 pixels in 119 samples were randomly selected as training data for each class, and the remaining 30% (1139 pixels) were used as test data for validation. Eventually, the feature importance was calculated as an average of the 50 runs. The feature importance was assessed using the *Scikit-Learn* package in Python 3.7. Following the ranking of importance scores, the most important features were selected as the optimal subset of features for saffron mapping.

Table 2. A summary of the vegetation indices explored in this study

Group	Spectral region	Vegetation index	VI derived from S2	Commonly related to	Reference
A	Visible-NIR	NDVI	(B8-B4)/(B8+B4)	Vegetation status, canopy structures	(Rouse et al., 1974) (Tucker, 1979)
		EVI-2	2.5*(B8-B4)/(B8+2.4*B4+1)		(Jiang et al., 2008)
B	RedEdge-NIR	RERVI	B8/B6	Biophysical characters of vegetation	(Jasper et al., 2009)
		RENDVI	(B8-B6)/(B8+B6)		(Gitelson & Merzlyak, 1994)
		REVI-2	2.5*(B8-B6)/(B8+2.4*B6+1)		Modification of (Jiang et al., 2008)
C	NIR-SWIR	NDII	(B8-B12)/(B8+B12)	Water content, residue cover	(Hardisky et al., 1983)
		NDWI	(B8-B11)/(B8+B11)		(McFeeters, 1996)
D	Visible-SWIR	NDSVI	(B11-B4)/(B11+B4)	Vegetation status, water content, residue cover	(Qi et al., 2002)
		NDRI	(B4-B12)/(B4+B12)		(Gelder et al. , 2009)
E	SWIR-SWIR	NDTI	(B11-B12)/(B11+B12)	Non-photosynthetic components, residue cover	(Van Deventer et al., 1997)
F	Visible-Visible	VIgreen	(B3-B4)/(B3+B4)	Leaf pigments, vegetation status	(Gitelson et al., 2002)

2.5. Saffron field mapping and accuracy assessment

RF is one of the commonly used classification algorithms that has shown good performance for cropland classification (Tatsumi et al., 2015; Sonobe et al., 2014; Ok et al., 2012; Pal, 2005). RF classifier is an ensemble of decision trees where each tree is constructed on a random subset from the total set of input variables. The final classification results are obtained by taking the majority voted class in the forest. In this study, the selected spectral-temporal features under section 2.4 were used as input variables for the RF classifier. The setting of the RF parameters was kept the same as for the training model setting described in Section 2.4.2.

A k-fold cross-validation was performed as a first assessment of the classification accuracy. It is a resampling procedure that is able to make predictions on all data and result in a less biased estimate of the model skill than a simple sample split method (Kohavi, 1995). The value of k refers to the group number that the samples will be split out, it was set as ten as recommended which generally results in a model skill estimate with low bias and modest variance. Samples for training and testing the RF classifier were similar to the dataset for feature importance calculation in Section 2.4.2 (i.e. Dataset A). The difference is that only the selected features were used as input variables. Among this dataset, 90% of the samples (nine groups) were retained as the training dataset, and the left 10% (remaining one group) were used as validation data for testing the model. The 10-fold cross-validation then consisted of repeating this procedure ten times, each time using a different 10% of the dataset as the test data (and the remaining 90% for training). Finally, the accuracy of this model is the mean value of all evaluation scores. The performance of the RF classification model was evaluated in terms of common statistical measures derived from the confusion matrix, which include the overall accuracy (OA), user's accuracy (UA), producer's accuracy (PA), and kappa coefficient (Foody, 2002).

To analyse the accuracy of the obtained saffron map, a set of new samples were extracted from Dataset B. The Dataset B consists of 213 saffron field parcels and is independent to Dataset A. Pixels which are located in these 213 polygons were used as sample data. This resulted in a total of 3019 pixels which have a label "saffron". The accuracy of saffron map was assessed using Sensitivity (i.e. true positive rate) index which refers to the ratio of truly predicted saffron pixels number to actual saffron pixels number (Altman & Bland, 1994). Specificity (i.e. true negative rate) cannot be calculated since no additional non-saffron fields are available in the study area (Altman & Bland, 1994).

2.6. Assessing the feasibility of classifying saffron fields by age

2.6.1. Spectral separability analysis

The attempt on classifying saffron crop age can be beneficial to analyse the reason for spectral-temporal variability of saffron which may cause disturbance on discriminating saffron with other crops. Moreover, saffron field age information is also important for farmers and decision-makers in crop management and rough yield estimation due to the strong link between crop age and saffron production (Kumar et al., 2009). Saffron age discrimination has not yet been attempted using remote sensing. For a very different tree crop (arecanut), crop age discrimination was attempted with hyperspectral imagery based on the age-dependent distinct spectral behaviour (Bhojaraja et al., 2016). According to the information collected from the field survey in December 2019 (nadir pictures shown in *Figure 3*) and interviews with local saffron experts, saffron fields with different ages usually presents various vegetation density during the vegetation stage (December to May), which is related to the number of daughter corms. In principle, the older the saffron field, the higher the density of green vegetation, although also external factors such as weeding practices and irrigation play a role. *Figure 4* demonstrates this using a 4-year time series of field-level NDVI observations from S2. On the first field (*Figure 4a*) saffron was planted in 2005, and on the second field (*Figure 4b*) in 2011. Despite the smaller number of observations before 2018, the figure clearly shows that in the first four years of cultivation, there is a gradual increase of green vegetation cover, whereas after this it remains more stable.

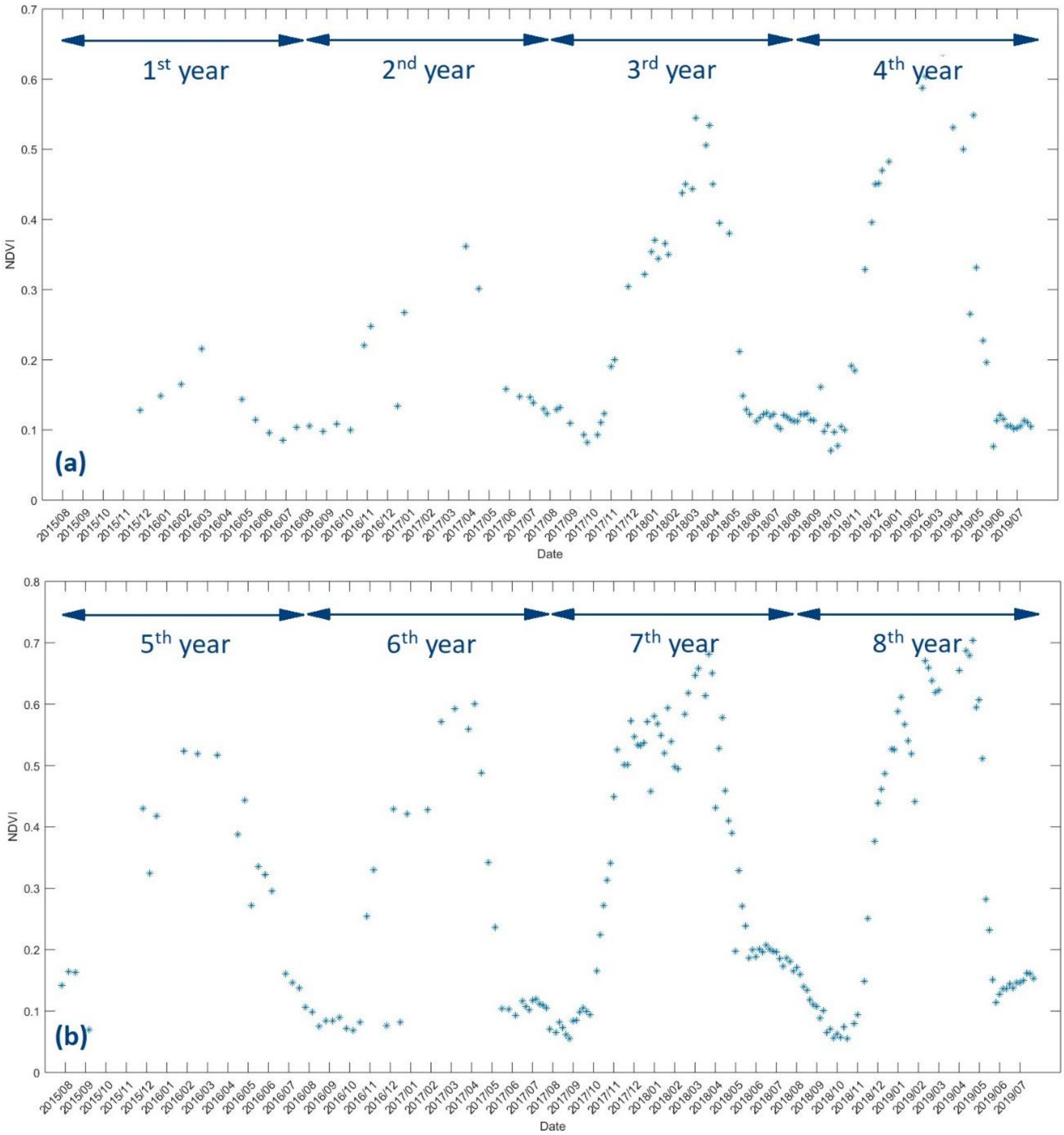


Figure 4. Two examples of NDVI time series extracted from saffron fields which are cultivated by saffron (a) age from 1 to 4 years (b) age from 5 to 8 years from August 2015 to July 2019

Among the VIs used for feature selection (*Table 2*), NDVI is the most commonly used vegetation index, which is associated with vegetation canopy greenness (Myneni et al., 1995). It was applied for analysing the differences in spectral performance of different age saffron crops during the vegetative stage. To analyse if saffron field age can be classified based on NDVI time series, the spectral variability of NDVI within same age fields (intra-class) and separability between different age fields (inter-class) during the vegetation stage (December to May) were quantified by the Jeffries-Matusita (JM) distance. JM distance is a parametric criterion with a value range from 0 to 2, where a large value indicates that two compared classes are more distinct, i.e. that their interclass variability is larger than their within-class variability. Therefore, it is able to verify the distinctness of the NDVI time series between saffron fields with different planting age.

As input to assessing the JM separability between different age groups, for each of the 120 sample fields (30 fields with age information in four growth seasons), we considered the average monthly NDVI between

December and May (i.e. six bands). The overall separability S between each group pairs for the combination of six bands was calculated by *Eq.1*.

$$S = \sqrt{\sum JM_i{}^2} \qquad (Eq.1)$$

2.6.2. Age-based classification and accuracy evaluation

Supervised classification method RF classifier (as mentioned in Section 2.5) with 500 trees was executed in this study. The number of age classes was determined based on the previous analysis. The average monthly NDVI between December and May in four growth seasons (2015-2019) for each of the 30 sample fields were combined and used as input dataset for the training classification model. Cross-validation and accuracy assessment follow the same methods used in Section 2.5.

2.7. Retrieval of peak flowering times from Sentinel-2

Flowering is an essential phenological period for saffron growth that directly reflects its yield. Saffron anthesis generally lasts for four to six weeks, but harvesting is only performed two weeks of this period due to the high labour cost. Usually, the harvesting practice starts when most of the flowers in the field bloom which is so-called peak flowering. Consequently, detecting the peak flowering date can be thought of equal to detect the start of the harvesting period. The method used in this study for detecting the peak flowering date consists of three steps. Firstly, a customized Enhanced Bloom Index (EBI_c) was developed to capture the unique spectral dynamics of saffron fields during the flowering period. Secondly, the EBI time series were smoothed and peaks were detected for saffron field samples with crop calendar data. Thirdly, the consistency and accuracy between detected peak flowering date and surveyed harvest date were analysed at the field level.

2.7.1. Enhanced blooming index customization

To capture the signals of saffron anthesis with optical satellite data, the Enhanced Bloom Index (EBI) was selected and customized as it is able to reflect the unique spectral performance of saffron flower during the flowering period. The EBI was first proposed by Chen et al. (2019) to characterize the blooming timing and intensity of almond flowers. Chen et al. (2019) demonstrated that the EBI can enhance the signals of flowers and reduce the background noise from soil and green vegetation. It was calculated using equation: $EBI = Brightness/(Greenness * Soil\ Signature) = R + G + B/\big(G/B * (R - B + \varepsilon)\big)$. The equation is built on the differences between in-situ spectral measurements of almond flower, green vegetation, and soil in the field. Since the almond flowers are normally white or pink, i.e. different from the purple colour of saffron flowers. The equation for calculating EBI needs to be customized based on saffron field spectral profiles to make it suitable for this study.

Since there are no available in-situ spectral measurements for saffron fields, the averaged spectral reflectance of saffron fields at different crop stages (bare soil, flowering, and green vegetation) were extracted from S2 data. First, the time windows that can reflect the three growing stages were selected. According to saffron phenological stages defined by the extended international Biologische Bundesanstalt, Bundessortenamt und Chemische Industrie (BBCH) scale (Lopez Corcoles et al., 2015; Yasmin & Nehvi, 2018) and interview information from local farmers about saffron growth stages and time, saffron fields are bare from July to September (BBCH 01 to 09), peak flowering period usually happens from mid-October to late November (BBCH 63 to 67), and vegetation has maximum greenness from February to April (BBCH 17 to 19).

For all saffron field samples, their average spectral value was calculated in different wavelength bands (visible R, G, B, and NIR) for these three stages (i.e. soil, flower, green vegetation) within the selected time windows

(*Figure 5*). Green leaves (green colour) have a high reflectance in NIR as compared to the visible bands where it shows a small peak in the green band. During the saffron flowering period, a similar spectral pattern can be observed as for bare soil, but with overall a lower reflectance. This is likely because the soil signal remains dominant given the low density of saffron flowers. Moreover, since irrigation and ploughing practices are usually operated around half a month and one week before flowering respectively, the soil becomes wetter, darker and rougher and consequently displays a lower reflectance. Considering the principle of EBI which aims to enhance the spectral signature of bloom while weakening background spectral signals from the soil and green leaves, above analysis on the performance of saffron field reflectance during bare soil and vegetation stage can be used to define the soil and greenness signature. However, the S-2 reflectance on a single band during flowering period cannot be linked to the spectral signature of saffron flower due to the disturbance of soil background.

To explore the spectral signature of saffron flower, its purple colour can be considered as a unique characteristic of saffron flower which performs differently on visible bands compared with green leaves and black soil. Chittka & Waser (1997) studied the spectral reflectance curves of coloured flowers and indicated that purple flowers typically have a relatively higher reflectance in blue and red bands compared with the green band (*Figure 6*). Therefore, the equation of customized EBI (EBI_c) can be formulated as *Eq.2*. $(0.5(R+B)-G)/(0.5(R+B)+G)$ in the numerator is used as the *purpleness* index to represent the overall higher reflectivity of saffron flowers. $(G-R)/(G+R)$ which is the equation for VIgreen (i.e., *greenness*), it is added to the denominator to reduce the impact of green leaves. NIR is introduced as another multiplicative term in the denominator to reduce the *background signature* influences from both soil and vegetation. The ε was set to 1 if the reflectance data ranging from 0 to 1, and set to 256 when applied to data between 0 and 255 (Chen et al., 2019).

$$EBI_c = \frac{Purpleness}{Greenness*background\ signature} = \frac{(0.5(R+B)-G)/(0.5(R+B)+G)}{((G-R)/(G+R)+\varepsilon)*NIR} \quad (Eq.2)$$

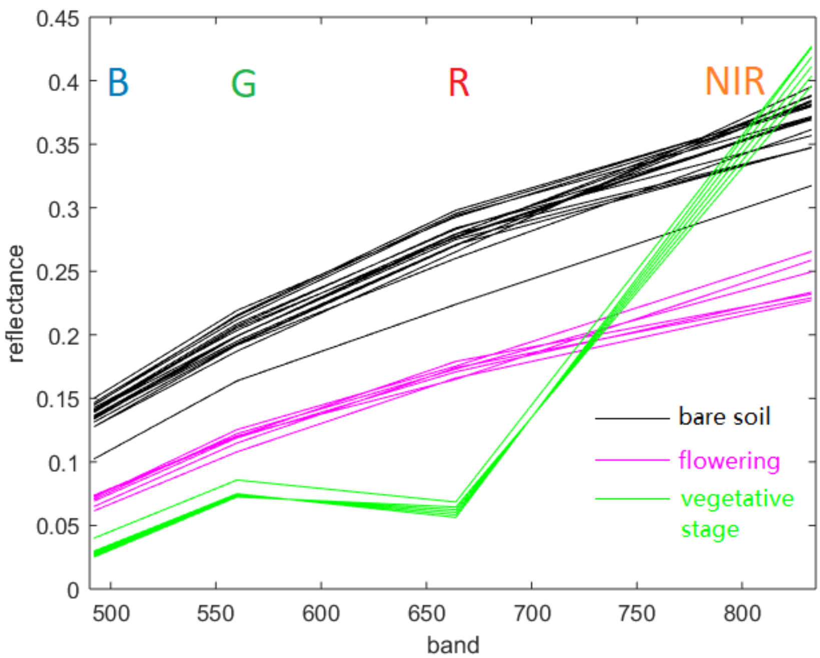


Figure 5. S2 reflectance in red, green, blue, and near-infrared bands during the expected time periods of bare soil presence, flowering, and vegetative stages. Each broken line refers to the average reflectance of saffron fields for one S-2 image.

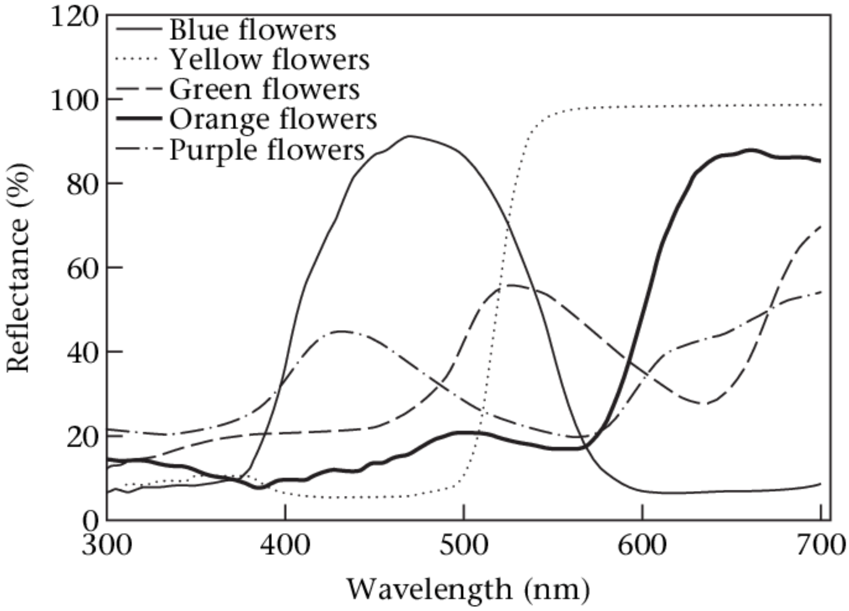


Figure 6. Spectral reflectance for flowers of different colour in the visible part of the electromagnetic spectrum (Chittka & Waser, 1997)

2.7.2. Peak flowering date detection from smoothed EBI time series

EBI time series were extracted based on *Eq.3* from preprocessed cloud-free S2 band stacks of the 2018 and 2019 growing seasons, and were spatially averaged for each saffron field with a 10 m negative buffer. Since the time-series data are spatial-temporally discontinuous, it is hard to observe a clear pattern (e.g. peak) from the time series. A number of mathematical filters have been developed in recent years to smoothen time series. The most commonly used methods can be divided into two groups: 1) smoothing in the frequency domain such as Fourier-based fitting methods (Sellers et al., 1994; Geerken, 2009); 2) smoothing in the time domain such as asymmetric Gaussian function fitting methods (Jönsson & Eklundh, 2002), Savitzky-Golay filtering (Chen et al., 2004), double logistic models (Beck et al., 2006), and the Whittaker smoother (Eilers, 2003). Studies compared the capability and reliability of different smoothing algorithms and indicated that their application resulted in small temporal differences (less than one week) when retrieving phenology (Atkinson et al., 2012; Beck et al., 2006; Hird & McDermid, 2009). In this study, the Savitzky-Golay filter was used to smooth EBI time series and executed using *the SciPy* package in Python 3.7. It is a convolution process which fits sub-sets of adjacent data with a low-degree (two-degree was used in this study) polynomial by the method of linear least squares. The smoothing window size was set to 11 and a three times iteration was used to make data approach the upper EBI envelope to best fit the EBI variations during the flowering period.

2.7.3. Comparison with phenology survey data

To assess the relationship between satellite derived peak EBI_c and surveyed start dates of harvesting, the coefficient of determination (r^2) and root mean squared error (RMSE) between both were calculated. For the surveyed 30 saffron fields, 46 crop calendar data are available in total, consisting of 16 fields with reliable in-situ data in 2018 and 30 fields with reliable data in 2019.

3. RESULTS

3.1. Spectral-temporal features for discriminating saffron from other crops

Figure 7 presents the temporal profiles of different VIs and spectral bands for saffron (black curves) and other crops (coloured curves) from July 2018 to June 2019. These temporal profiles were calculated by per-class averaging of the corresponding VI or spectral values of pixels within the 119 sample field parcels. *Figure 7* shows that the spectral differences between saffron and other classes vary over time. For example, NDVI from July to October presents low value for most crops apart from alfalfa. The NDVI value of saffron increases in November, which indicates the start of vegetative stage, and reaches the highest value (above 0.6) around March, while the NDVI of other crops such as barley and winter wheat remain low until February. During summer (from July to early-September), several spectral bands such as visible (band 2, 3, 4) and SWIR (band 11, 12) perform relatively higher reflectance for saffron as compared with other crops. This is because saffron fields were dry and bare soil during summer while other crop fields such as alfalfa and pistachio were covered by vegetation. After that, these band reflectance of saffron fields decline along with the first irrigation in September, which increases the water content of soil and promotes the development of plants. These different behaviours indicate the importance of selecting unique spectral-temporal features for distinguishing saffron fields from other crops.

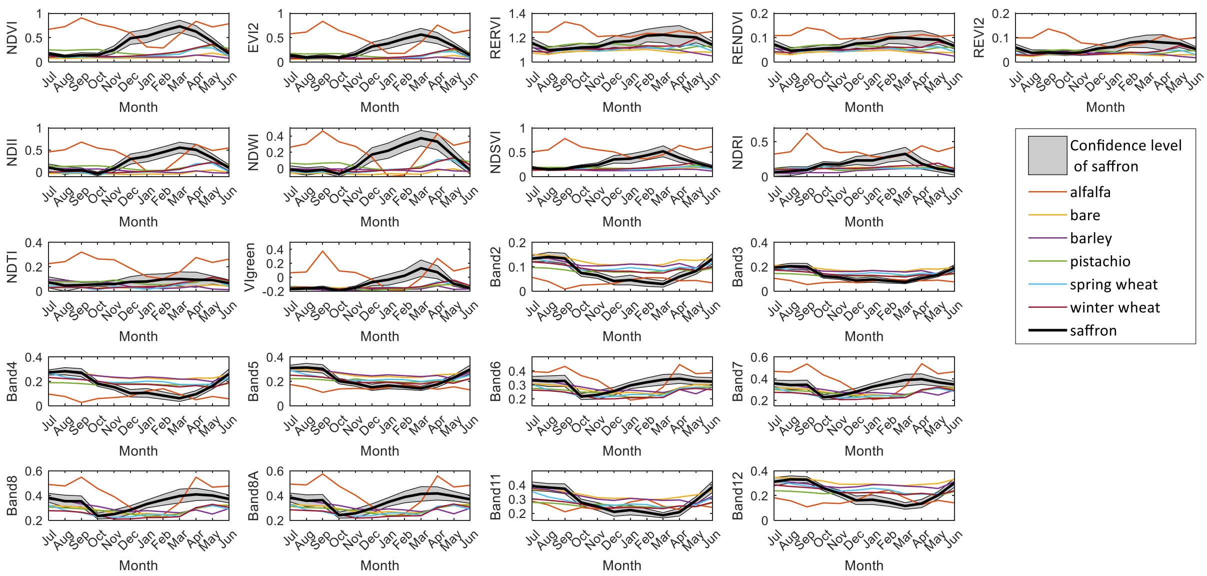


Figure 7. Temporal profile of 21 spectral features between July 2018 and June 2019 for different crops averaged from 119 sample fields

Figure 8 displays the average importance score of each spectral-temporal feature for separating saffron from the other classes. The horizontal axis of the chart represents the VIs and spectral band reflectance, the vertical axis refers to the time scale in month. The colour bar presents the importance value of each feature, with the yellower colours referring to a higher importance score, which means that the feature contributes more to discriminating saffron from other classes. The result shows the highest importance scores are in February for NDVI, EVI2, and NDII. The relative average importance rank and its corresponding value of each feature in each run can be found from *Figure A1* in the Appendix. With the average ranking result of each feature, the influence of the number of features on classification accuracy was assessed and shown in *Figure 9*. The OA of classification increased with the number of input features until a saturation point reached at 19 features (95.5%).

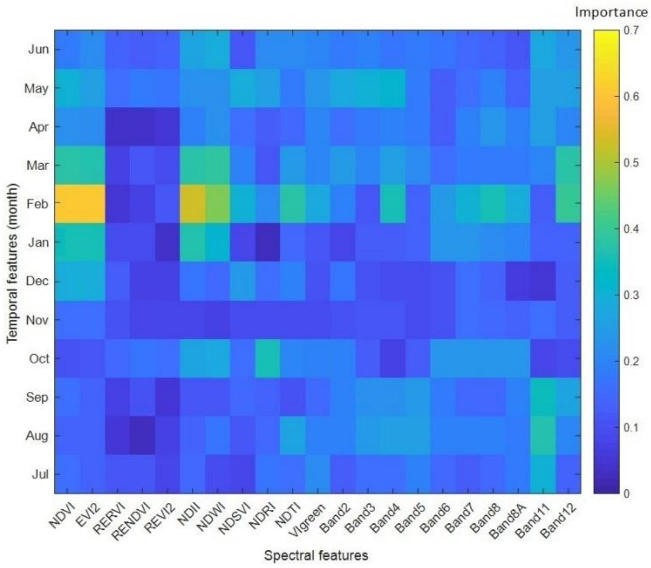


Figure 8. Importance value of spectral-temporal features. The horizontal and vertical axes of the chart represent the VIs or spectral band and time scale, respectively. The value in each grid cell represents the importance value of the corresponding feature.

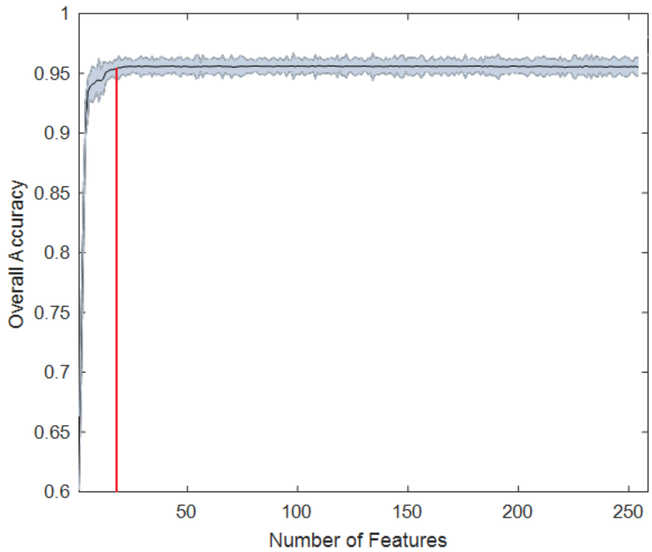


Figure 9. Overall accuracy (OA) of identifying saffron from other classes using different input features suggested by the importance score. The black curve shows the average OA, shadow area represents 95% confidence interval of OA. Red line shows the saturation point and its corresponding feature number.

Among the first 19 important features (panel (a) in *Figure 10*), VIs account for 13, while the remaining six features are spectral bands. Moreover, the result shows all these 19 most important features are related to spectral regions in Red, NIR, and SWIR bands. Among the 6 VIs groups, group A (i.e. the combination of Visible and NIR) and C (i.e. the combination of NIR and SWIR) dominate a large portion of the 13 VIs features, which accounts for six and four features, respectively. Most temporal features are concentrated in spring (from January to March), and a few features have a high importance value in August and September.

The Pearson correlation coefficient between the first 19 important features were calculated and displayed in *panel (b) Figure 10*. It shows that NDVI, EVI2, NDII, NDWI, and NDTI from January to March were significantly correlated. The correlation among the same VIs groups were stronger than others since they are calculated from the same combination of spectral regions and present similar time series behaviours.

Band4_Feb, band12_Feb, and band12_Mar were negatively correlated with VIs in the same time periods. For both VIs and spectral bands, their behaviours in the same month show a higher correlation than features in different months.

Considering the high correlation of features within the same groups and same time windows, finally, NDVI_Jan, NDVI_Feb, NDVI_Mar, Band12_Feb, Band12_Mar, Band11_Aug, Band11_Sep were selected as the optimal set of input features for saffron classification. *Figure 11* displays the time series of selected spectral features and their time windows (i.e., temporal features).

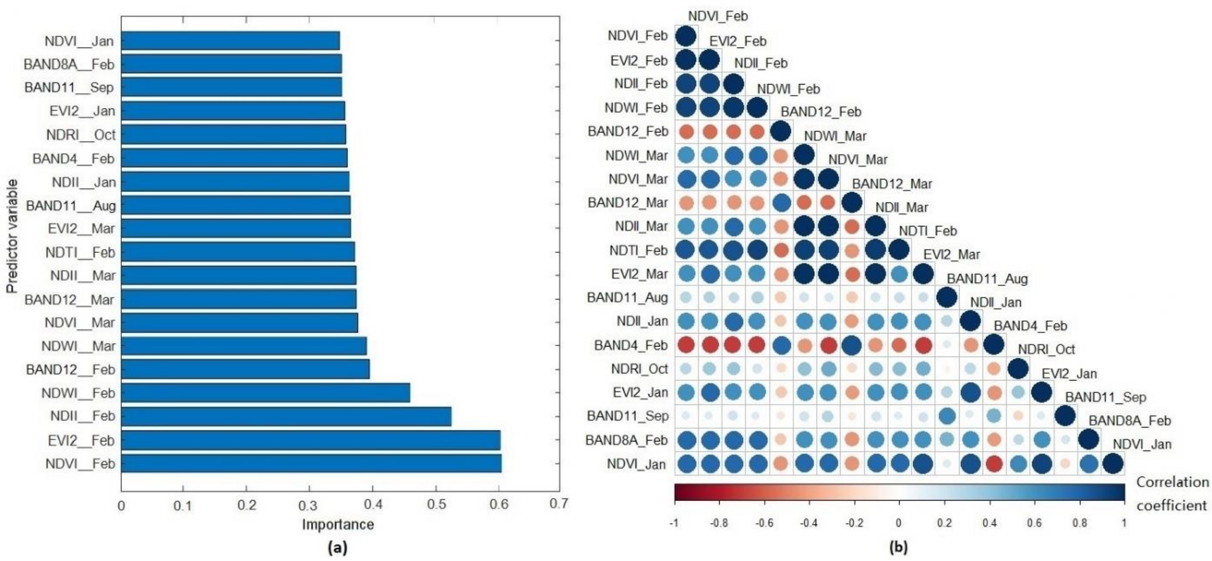


Figure 10. Importance value of the first 19 most important features and their Pearson correlation coefficient between each other. Panel (a) presents the average importance value of these 19 features in 50 runs of RF-based classification. Panel (b) shows the correlation matrix with p-value < 0.001. Positive correlations are displayed in blue and negative correlations in red colour. Colour intensity and the size of the circle are proportional to the correlation coefficients.

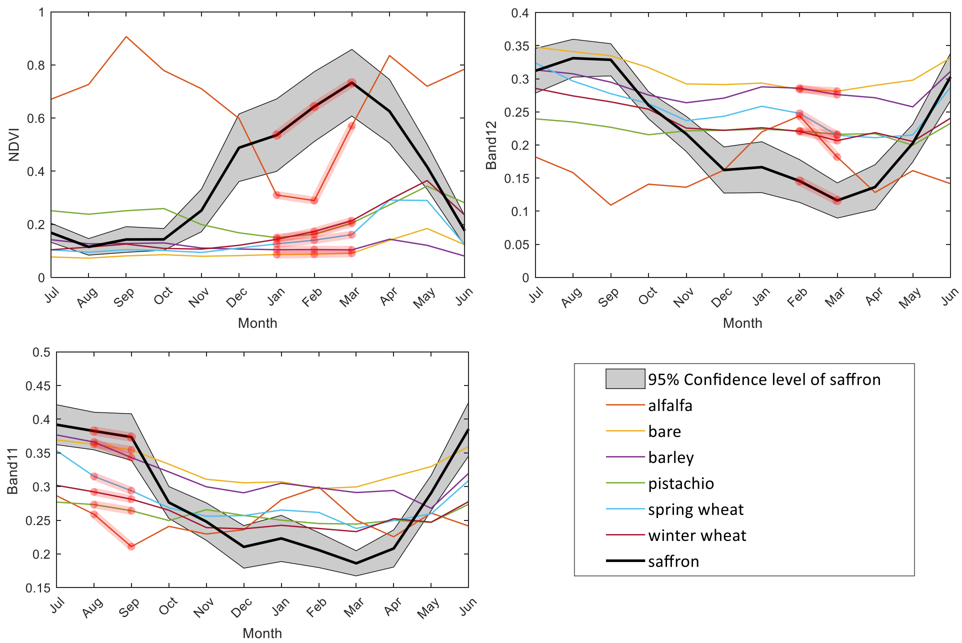


Figure 11. Time series of selected spectral features. Red shadows refer to the spectral reflectance within time windows (i.e. temporal features) which has relatively high importance value for saffron identification.

3.2. Accuracy assessment of saffron cultivation map

The application of RF to the selected spectral-temporal features results in a good classification of saffron fields (*Table 3*). The classification model has an OA of 95.3% and a kappa coefficient of 0.93. UA and PA achieve 95.6% and 92.9% for saffron and 97.7% and 98.6% for non-saffron classes, respectively. It suggests that the saffron class has relatively high errors of omission (i.e. 20 saffron pixels classified as non-saffron). The misclassification of saffron was mostly caused by the first-year age saffron field (not shown). It is hard to be distinguished from winter crops such as barley and winter wheat. This is due to the low density of greenness for young saffron vegetation, and the emergence of weeds in March, which is also the moment when winter crops start to grow.

Figure 12 shows the map of classified saffron cultivation area in the whole study area. The saffron map has an accuracy of 87.4% when calculate using sensitivity index and an independent dataset (i.e., 2637 pixels were correctly classified as saffron among the 3019 saffron pixel samples derived from dataset B) (not shown). The green polygons in *Figure 12* displays highlight the incorrect classified saffron fields. The underestimation of saffron was mostly caused by pixels near to parcel edge (*Figure 13 a*), which typically have an integrated spectral response from multiple crops or land cover types. Consequently, some small saffron fields (< 0.2 ha) were omitted due to the heterogeneity of pixels with 10m resolution. The results present saffron fields larger than one ha has an accuracy (sensitivity) of 92.3%, while the accuracy is only 73.1% for small fields (< 0.2 ha).

Table 3. Confusion matrix of the crop classification result

	Reference	Saffron			Non-saffron				UA (%)		Commission
Classified		Saffron	Alfalfa	Bare	Barley	Pistachio	Spring wheat	Winter wheat			error (%)
Saffron	Saffron	263	0	5	2	0	0	5	95.6	**95.6**	4.4
Non-saffron	Alfalfa	0	37	0	0	0	0	0	100.0		
	Bare	3	0	425	0	2	1	2	98.2		
	Barley	5	0	0	32	0	0	1	84.2		
	Pistachio	0	0	1	0	102	1	1	97.1	**97.7**	2.3
	Spring wheat	0	0	0	0	4	65	1	92.9		
	Winter wheat	12	0	7	1	1	0	161	88.5		
		92.9	100.0	96.6	91.4	93.6	97.0	94.2			
	PA (%)	**92.9**				**98.6**					
	Omission error (%)	7.1				1.4					

OA: **95.3%**
Kappa coefficient: **0.93**

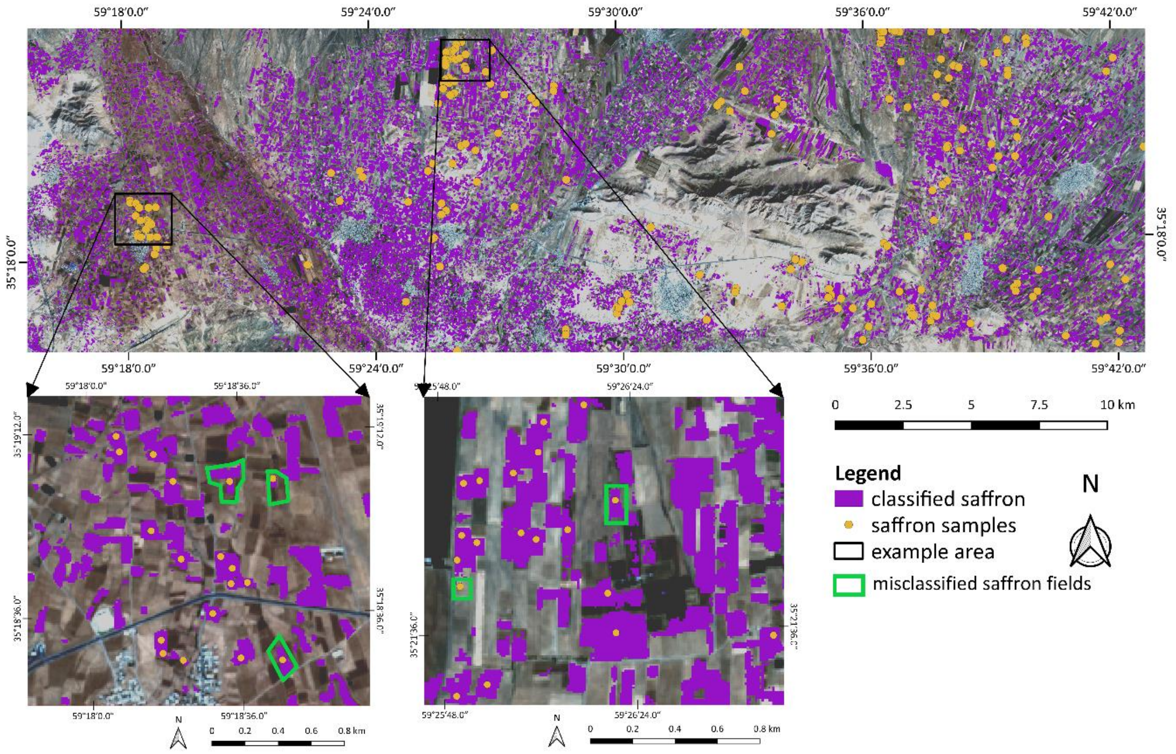


*Figure 12. Saffron cultivation map in the study area. The background shows a S2A image acquired on 26th Sep 2018 with an RGB composite of NIR, Red, and Green band.*Feasibility evaluation for age-based saffron field classification

3.3. Feasibility evaluation for age-based saffron field classification

The average NDVI time series for different age of saffron fields and their associated standard deviation are shown in *Figure 13*. As illustrated in *Section 2.6.1*, the NDVI value of the first four years of saffron plants mainly increases with age, while for the older fields (4th to 6th year), it remains more stable. The 7th and 8th year saffron fields have the highest NDVI during the vegetative stage compared with other age fields. These results are consistent with the performance of green vegetation density investigated during the fieldwork (*Figure 3*).

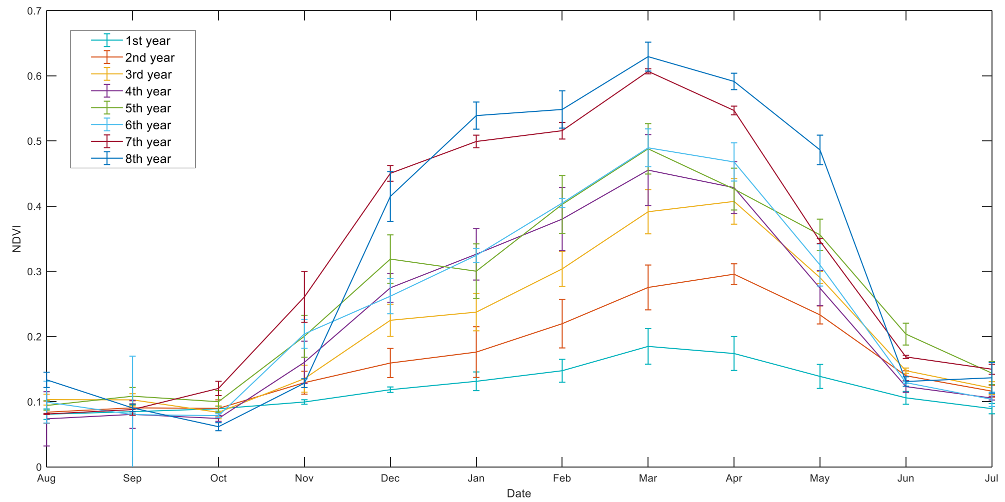


Figure 13. Average NDVI time series for different age of saffron fields. Error bars indicate the standard deviation of NDVI value in each month for the same age of field samples

Figure 14 presents the pairwise separability metrics (i.e., JM distances) of NDVI time series between each age saffron field and the seven other age fields. These charts map the separability between different age saffron fields across the six months from December to May using the 30 saffron field samples which have age information in four growth seasons (2015-2019). The yellow colours indicate a high separability between two classes in the corresponding month. The global separability (S_G) between each pair of classes (ages) are shown in *Figure 15*. It suggests that, in general, the S_G between saffron fields with smaller age differences are lower than those with larger age differences. But some strange results also exist in the figure, for example, 1st versus 3rd year has better separability as compared to 1st versus 4th year, 6th versus 8th year performs higher separability than 4th or 5th versus 8th year. Between the 4th to 6th year saffron, the S_G is relatively lower, indicating the difficulty of distinguishing these classes from NDVI time series. Therefore, saffron fields ages from the 4th to 6th year were merged as one group for later classification. Besides, due to the fewer samples (only 1 field) for 8th year saffron field, the 7th and 8th year fields were also combined as the same group.

The age-based classification was implemented using the RF algorithm, the six-monthly values of NDVI from December to May were used as input variables. *Table 4* shows the accuracy of the classification result. The OA and kappa coefficient reach 86.8% and 0.81, respectively. Among these age groups, the 1st year saffron fields had the best performance (UA of 93.1% and PA of 90.0%) for its discrimination from other classes. The 3rd year saffron fields had considerably higher errors of commission, i.e. 21 pixels of 4th to 6th year saffron were incorrectly classified as 3rd year plants. Age group 4th to 6th year and 7th to 8th year performed better classification accuracy (UA of 88.7% and PA of 85.9% for 4th to 6th year class, UA of 88.7% and PA of 87.9% for 7th to 8th year class) as compared with classification result for individual age crop (*Table 4*).

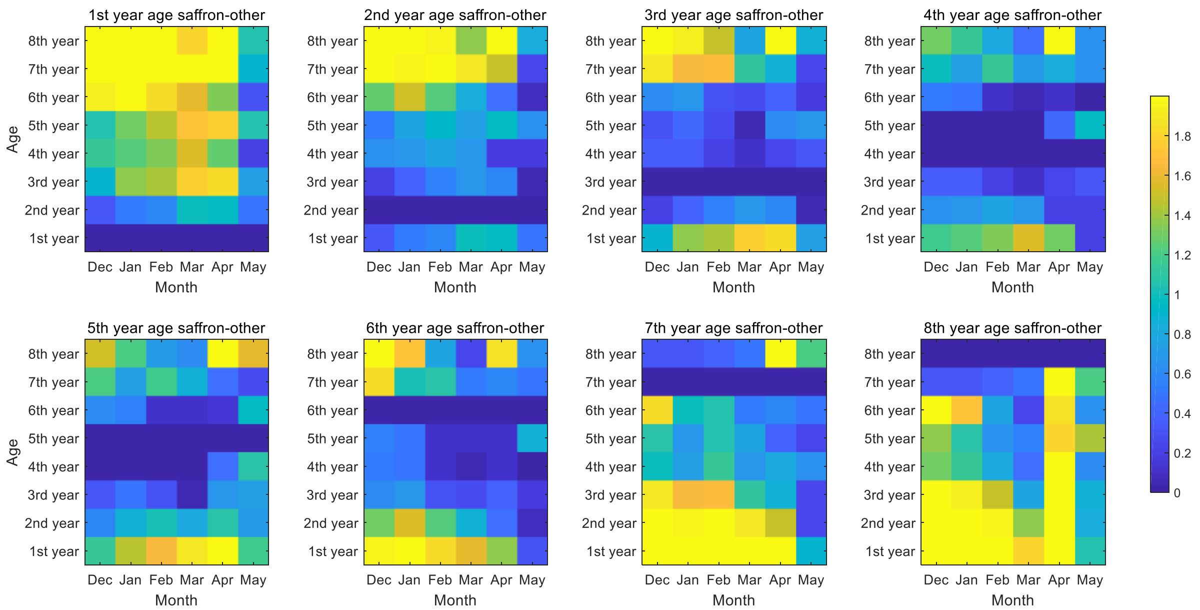


Figure 14. Separability metrics (JM distance values) for all saffron age pair comparisons in each month between December and May.

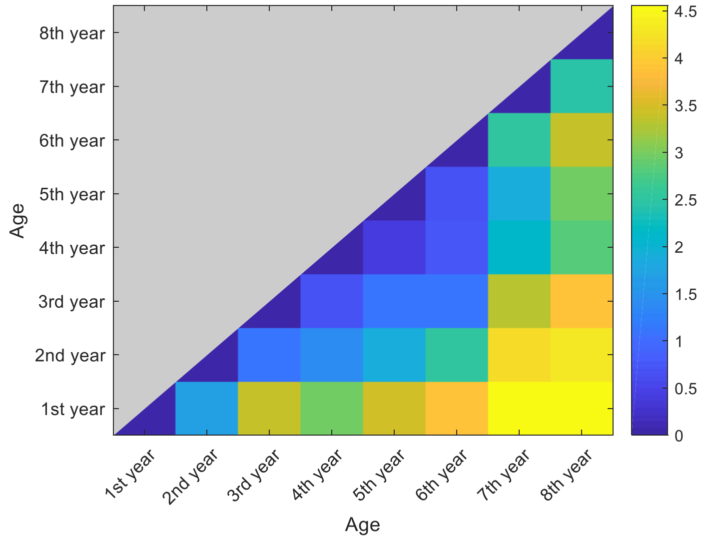


Figure 15. Global separability of NDVI time series during December and May for each pair of saffron age comparison.

Table 4. Age-based classification results for five grouped classes and eight individual classes.

	8 classes (before merged)				5 classes (after merged)				
Age	PA (%)	UA (%)	OA (%)	Kappa	Age	PA (%)	UA (%)	OA (%)	Kappa
1st year	89.9	93.1	75.4	0.72	1st year	90.0	93.1	86.8	0.81
2nd year	83.3	83.2			2nd year	83.1	83.3		
3rd year	81.5	78.1			3rd year	82.6	77.9		
4th year	61.3	62.5			4th-6th year	85.9	88.7		
5th year	66.7	68.4							
6th year	72.1	70.4							
7th year	81.2	78.8			7th-8th year	87.9	83.7		
8th year	85.3	84.9							

3.4. Detection of peak flowering and comparison with crop calendar data

The smoothed EBI_c time series are illustrated in *Figure 16*. The black curve presents averaged EBI_c time series extracted from all 30 saffron field samples in 2018-2020 growth season. It shows a clear peak during the expected flowering period (i.e. late October to late November). *Figure 17* displays EBI_c profiles for eight saffron fields for which surveyed crop calendar information was collected. This included the date of flowering irrigation, ploughing, and harvesting, which is corresponded to the duration of peak flowering. It shows a strong agreement between the start date of harvesting and the time of maximum EBI, which both are found between October to late November.

The peak of flowering date was extracted from EBI_c time series in 2018-2020 growth seasons for each of the 30 saffron field parcels from EBI_c time series and compared to *in-situ* observations (*Figure 18*). The determination of correlation (R^2) between S2-derived peak flowering date, and the surveyed start date of harvesting is 0.69, the difference (RMSE) is around four days.

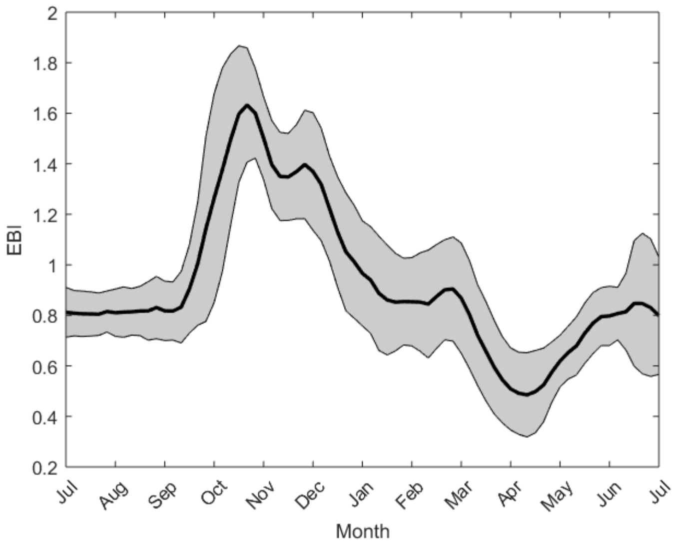


Figure 16. Smoothed EBI_c *time series of 30 saffron sample fields. Black curve refers to the average* EBI_c *time series, gray shadow presents the 95% confidence interval of all extracted time series.*

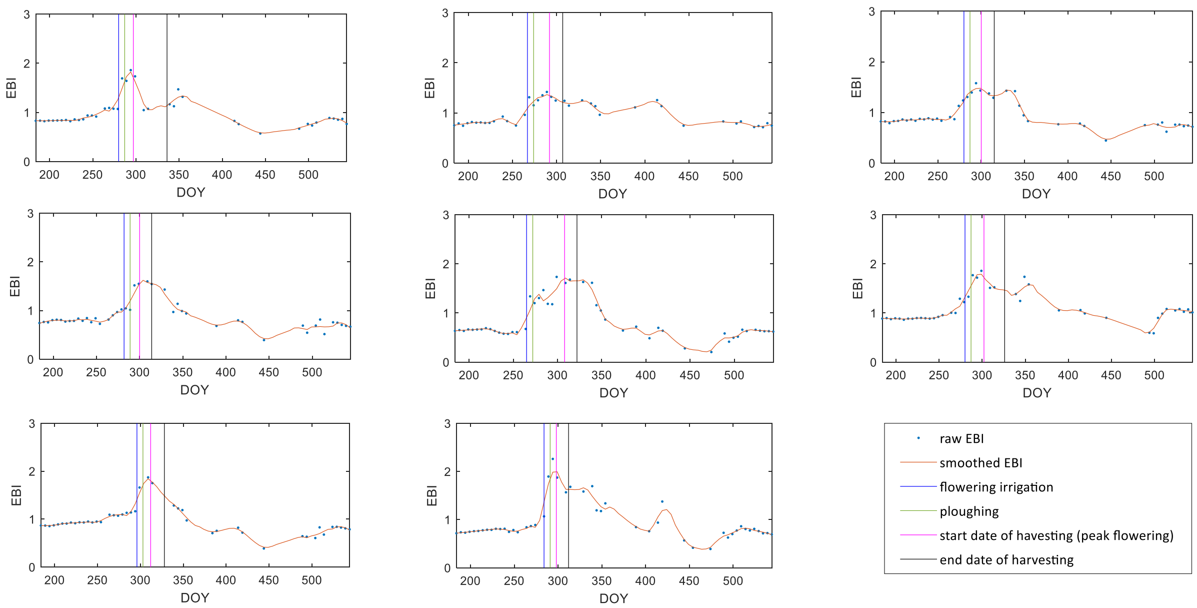


Figure 17. Examples of smoothed EBI_c *time series in relation to surveyed crop calendar for eight saffron fields*

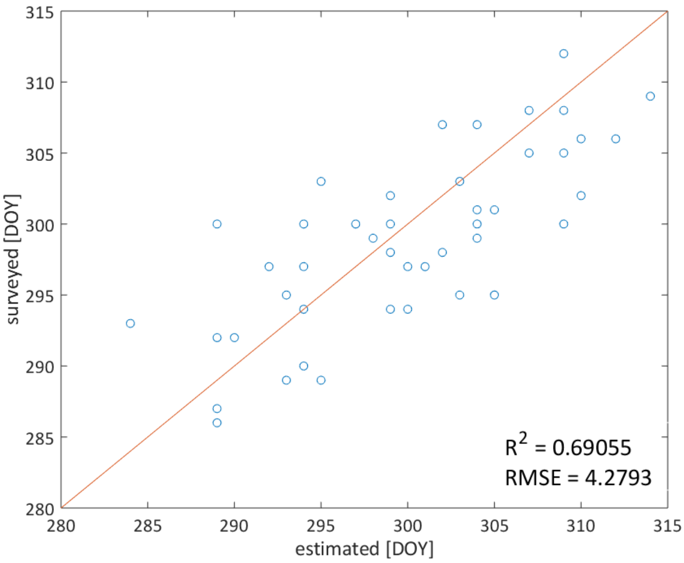


Figure 18. The coefficient of determination (R^2) and difference (RMSE) between estimated peak flowering date and surveyed start date of harvesting from 46 samples (consist of 16 field data in 2018-2019 and 30 field data in 2019-2020)

4. DISCUSSION

This study evaluated the performance of S2 time series in improving saffron mapping accuracy, conducting age-based classification, and detecting saffron phenological behaviours. The use of multi-temporal S2 imagery resulted in a high classification accuracy for identifying and mapping saffron fields in Torbat-e Heydariyeh. Input variables of the RF classifier are spectral-temporal features that were selected based on RF importance ranks and displayed distinct phenological behaviours for saffron as compared with other crops. The study also showed that NDVI during the vegetation stage allows for discriminating different ages of fields, given that age affects the density of the green cover. EBI_c was designed to capture the unique spectral performance of saffron anthesis stage. The S2-derived EBI_c time series show great potential to detect the peak flowering date of saffron and agreed well with the surveyed crop calendar data. The main findings of saffron classification, age discrimination, and flowering phenology detection were discussed in Section 4.1, 4.2, 4.3, respectively. Section 4.4 analysed the limitation of this study and provided recommendations for improvement. The application of the outputs in agricultural management for different stakeholders was explained in Section 4.5.

4.1. Sentinel-2 time series for saffron classification

The RF importance rank was used to infer which spectral properties of the saffron crop in different seasons contribute more to identify saffron fields. The top important spectral features mainly across VNIR and SWIR spectral regions, which reflect the biochemical and biophysical properties of vegetation, such as photosynthetic pigment absorption, structural photon scattering, and water content. These spectral features (NDVI, NDWI, NDRI) has proven to be efficient in crop classification (Peña-Barragán et al., 2011; Hao et al., 2015). Moreover, two phenological phases were identified important in separating saffron from other crop classes, which are the rapid green-up stage (January to March) and dormant period (August to September). The first phase is the rapid green-up stage of saffron from January to March when winter crops are just beginning to growth and summer crops were harvested and fallow. The second importance stage for identifying saffron cultivation is the dry summer in August and September when saffron corms are dormant underground. During this period, saffron fields are bare with very low soil moisture and have a high spectral reflectance in SWIR bands until the first irrigation is carried out in late September. This contrasts with spring crops that are harvested and covered by residues in summer.

The classification result demonstrated the utility of multi-temporal S2 imagery in distinguishing saffron from other crops. Compared with previous research that used single date Landsat 8 imagery to map saffron fields (Dehghani Bidgoli et al., 2018), the OA of classification improved by 13.3% (from 82.0% to 95.3%). This mainly benefits from the high temporal resolution of S2 satellite imagery. Multi-temporal images are effective in reflecting the phenological changes and differences of crops (Wardlow et al., 2007; Zhong et al., 2014; Arvor et al., 2011). The higher temporal resolution provides more cloud-free data as compared to Landsat-8 imagery and allows to select suitable time windows for capturing distinct temporal behaviours of target crops. Besides, the mapping results showed the classification accuracy for small saffron fields improved as compared to the study by Dehghani Bidgoli et al. (2018). The accuracy for saffron fields increased from 62.0% to 73.1% for fields smaller than 0.2 ha, and from 72.0% to 78.3% for fields between 0.2 ha to 0.5 ha. This improvement probably benefits from the higher spatial resolution of S2 imegery (10m) in comparison with Landsat 8 (30m).

4.2. Feasibility of age-based classification

The work on age-based classification is the first attempt to classify the age of perennial crops using multi spectral time series data. This study indicated the feasibility of NDVI time series for saffron age classification. With increasing age of saffron fields, NDVI levels during the vegetation stage (December to June) increased. NDVI temporal patterns of different age saffron crops have revealed that five age groups can be separated. Particularly, young saffron fields (i.e., the first-year crop), showed a high separability and classification accuracy compared with older fields. This finding provided a new insight for further improving the accuracy of saffron identification by treating the first-year saffron as a separate class, which was shown to be easily confused with other winter crops in Section 3.2. The difficulties in distinguishing saffron fields with ages from four to six years are likely due to their similar vegetation cover, which is consistent with the performance of green vegetation density investigated during the fieldwork (*Figure 3*). This performance could be explained as the increasing competition between overcrowded corms for water and nutrients reduced the reproduction capability of corms (Kumar et al., 2009).

4.3. Estimation of peak flowering time

The customized EBI (EBI_c) time series indicate the potential to detect peak flowering of saffron at the parcel level. The peak flowering date is identified as the time of the main peak (local maximum) in EBI_c time series. The flowering phenology retrieval mechanism and spectral time series performance of index are similar to the study by d'Andrimont et al. (2020). Comparing the S2-derived peak flowering date with 46 surveyed crop calendar data, the result (*Figure 18*) shows a four-day difference (RMSE) which is below the nominal revisit interval of S2 sensors (5 days).

In the EBI_c time series, after peak flowering (first peak), the signal becomes more fluctuated. This is likely related to irrigation practices. According to the knowledge I got from interviews with local farmers, after the flowering irrigation, farmers usually conducted four- or five- more times irrigation in late November, late January, late February, and mid-May. These practices increase the soil moisture, which usually results in lower spectral reflectance. This signal will mainly influence the element of "background signature" in the EBI_c equation, which introduces "NIR" as a multiplicative term in the denominator. Therefore, EBI_c may present a small increase after each irrigation.

Similarly, the rapid increase of EBI_c before the first main peak could be explained not only by the growing purpleness of saffron flowers, but also the darker and wetter soil as a consequence of flowering irrigation and ploughing practices. Then, the first (main) peak appears at the moment when the index start to decrease, this behaviour is related to the harvesting practices which decrease the saffron purpleness.

Besides, it is observed that, for some saffron fields, a second small peak followed the first (maximum) extreme in the EBI_c time series. Comparing with surveyed crop calendar data, it usually happened after the end of harvesting period. This is because the harvesting period refers to the peak blooming period instead of the whole flowering stage. Saffron anthesis generally lasts for four to six weeks, but harvesting is only performed two weeks of this period due to the high labour cost. Usually, the same fields could be harvested several times due to the newly bloomed flowers every day. Even after the end of harvest, new saffron flowers could also bloom if the temperature is good (12-18°C by day and 4-8°C at night) (Molina et al., 2005).

4.4. Potential limitations and further improvements

RF was used in this study for classification purposes because previous studies demonstrated its effectiveness in cropland classification (Tatsumi et al., 2015; Sonobe et al., 2014; Ok et al., 2012; Pal, 2005). Nevertheless, the RF classifier requires a large training dataset, particularly if high inter-class variability exists (Rodriguez-Galiano et al., 2012). The saffron classification result shows that saffron has high spectral variability due to age influences (especially the first year saffron), which causes a disturbance in discriminating with winter crops. One of the solutions could be separating the first year saffron as a single class (as mentioned in Section 4.2) to improve the identification accuracy of other age saffron fields. This indicates the needs of more training samples for the first year saffron. However, the field survey cannot always provide enough data to represent all the variability in a class. To improve the classification accuracy, other classifiers, such as time-weighted dynamic time warping (TWDTW), can be considered as an alternative method in the case of class with high spectral variability and few training samples. TWDTW proved as an efficient solution in handling these challenges due to its advantages of balancing between time series shape matching and temporal alignment (Maus et al., 2016), and low sensitivity to the intra-class variability and size of the training dataset (Belgiu & Csillik, 2018).

The high accuracy of the saffron classification suggests the possibility to scale this method to a larger region. Iran accounts for more than 90% of the global saffron production (Ghorbani, 2007), mapping saffron fields, especially the key cultivation region (i.e., the north-eastern Khorasan region in Iran), can provide geographically explicit information and contribute to the sustainable management of saffron cultivation. For a larger area, the exists of other crops that may have similar phenological behaviours and saffron cultivated under different management practices could influence the transferability of the classification model (Jin et al., 2018; Juel et al., 2015). Therefore, apart from the study area (Torbat-e Heydariyeh county), more tests could be conducted by transferring this classification model in other counties which have different crop composition and climate conditions in the Khorasan region. Considering the high memory requirements and time cost for processing a large amount of S2 images, this process can be executed on Google Earth Engine (GEE, Gorelick, et al., 2017), which is a cloud-based platform that allows for access and processing of large earth observation datasets. GEE has shown to be effective for large-scale crop mapping and monitoring applications by reported researches (Lemoine & Leo, 2015; Tian et al., 2019).

To increase the size of saffron field age samples, this study combined age data in four growth seasons (from July 2015 to June 2019) to expand the dataset and improve the reliability of age classification results. But this approach ignores the annual spectral variation, which could be influenced by different weather conditions and (timing of) agricultural practices between years. If more reference data available, the age-based classification could be conducted for each year's data. It provides a more operational type of monitoring to analyse the sensitivity of classification results in inter-annual changes. By comparing the classification result of the same patch in different years, we can better understand the impact of spatial-temporal variation on classification accuracy. Moreover, mapping saffron age in each year based on the trained classification model also provides an opportunity to deliver accurate age maps for any particular year. For example, if a field was classified as saffron in the 4th to 6th year according to its spectral signature in 2020, then we can go back to previous years to check the classification results and calculate the exact age of this patch in 2020.

This study demonstrated the utility of S2 time series in saffron classification and phenology detection. However, optical imagery is susceptible to clouds, which causes missing and irregular data especially during the rainy season (from January to April), which is a crucial period of saffron vegetation growth in northeast Iran. The absence and irregularity of images in a long period could influence the quality of generated monthly data and reduce the classification accuracy. Compared with optical imagery, microwave radiation is capable of penetrating clouds with negligible attenuation (Richards, 2005), which can ensure continuous measurements over the growing stage of saffron. To minimize the cloud occlusion problem of optical images, the synergistic use of radar and optical information has become a research hotspot. Earlier studies have successfully improved crop classification accuracy (Skakun et al., 2016; Dong et al., 2013; Van Tricht et al., 2018) and phenology monitoring (De Bernardis et al., 2016) using a combination of optical and Synthetic aperture radar (SAR) images. Different from optical imagery which reflects vegetation biophysical characteristics, SAR imagery provides new insights into vegetation structure, surface roughness, and soil moisture (Veloso et al., 2017). Therefore, analysing the spectral-temporal performance of saffron fields in backscatter coefficients and their ratio time series may detect more phenological features and agricultural activities in combined with the biomass and soil changes. Further study could investigate the performance of data fusion of radar and optical time series in saffron mapping, age-based classification, and phenological monitoring.

4.5. Application in future agricultural management

The saffron cultivation map generated in this study offers geographically explicit information on saffron distribution. It is typically useful for transparent and effective management of saffron fields. Traditionally, the cultivation area information of saffron relied on field surveys by agricultural sectors, which is laborious and time-consuming. The Satellite-derived saffron maps significantly improved the efficiency of data collection and provide timely updated information in a large scale. Moreover, having maps can also contribute to regional land use planning. For example, the distribution of water reservoirs influences the accessibility of water for saffron irrigation. The spatial information of saffron fields can be used for local government to properly allocate and manage the use of groundwater to maximize water efficiency. Besides, considering that counterfeit saffron stigma has become a serious problem in global saffron markets, identifying the authenticity of saffron farms stated by businesses is of importance for consumers. Satellite-based saffron field identification provides an opportunity for better supervising and management of the saffron market.

The age information is one of the most important internal factors which is correlated with saffron field productivity (Kumar et al., 2009). It could be useful for agricultural sectors to estimate and predict the approximate regional yield level in the next years. Flowering phenology is able to reveal the saffron production which mainly refers to the harvested flower (or stigma) weights at a parcel level. Detecting the changes of peak flowering in spatial and temporal extent is of importance to understand the influence of environmental factors (e.g. drought) on the field production. This information help decision-makers to select suitable cultivation area which is more preferred for sustainable growth of saffron. It is important to note that the further implementation of the proposed recommendations requires the corporation between farmers and agricultural sectors. The in-situ data provided by farmers could be used as complementary data and to validate satellite-derived information.

5. CONCLUSIONS

This study demonstrated the potential of S2 time series in accurately mapping saffron cultivation area, classifying different age of saffron fields, and estimating peak flowering time.

RF importance ranks were used to select an optimal set of spectral-temporal features for discriminating saffron and other crops. The feature selection result showed most important spectral features are VIs across spectral regions in Red, NIR, or SWIR bands. The temporal features are mainly concentrated in the rapid green-up period of saffron (from January to March), while other crops are in fallow or the start of vegetation stage during that period. Besides, early autumn (August and September) metrics are also important and display distinct phenological behaviours as compared with other crops. Considering the high correlation between features, finally, NDVI from January to March, SWIR from August to September (band 11) and February to March (band 12) were selected as the optimal set of input features for the RF classifier. RF resulted in good classification results in discriminating saffron with other crops using multi-temporal S2 imagery. The classification model had an OA of 95.3% (kappa coefficient of 0.93). Based on an independent in-situ dataset on saffron fields, 87.4% of the existing fields were correctly classified as being saffron.

Global separability analysis of different age saffron fields revealed that five age groups can be distinguished with clear spectral separability on NDVI time series, which are 1st, 2nd, 3rd, 4th-6th, and 7th-8th year saffron. The classification result presented an OA of 86.8% (Kappa coefficient of 0.81) for the five age groups using NDVI time series from December to May. Young saffron fields especially the first-year crop showed a high classification accuracy (PA of 90.0%, UA of 93.1%) compared with older fields. Merged two class groups showed significantly better classification result (UA of 88.7% and PA of 85.9% for 4th to 6th year class, UA of 88.7% and PA of 87.9% for 7th to 8th year class) than classifying each age individually (UA varied between 62.5% and 70.4% for 4th to 6th year class).

The customized EBI (EBI_c) time series indicated the potential of detecting peak flowering of saffron at the parcel level. The S2-derived peak flowering date showed a close correlation (R^2 = 0.69) with surveyed crop calendar data based on 46 fields, and reached four days difference (RMSE), which is shorter than the nominal revisit time interval of S2 sensors.

The findings of this study provided insights for accurately mapping and monitoring saffron fields distribution, age, and phenology changes using S2 time series data. These outputs provide a basis for future research on production forecast and estimation and can further contributed to better agronomic management for decision-makers.

LIST OF REFERENCES

Abdullaev, F. I., & Frenkel, G. D. (1999). Saffron in biological and medical research. *Saffron: Crocus Sativus L. Harwood Academic Publishers, Australia*, 103–114.

Alizadeh, A., Sayari, N., Ahmadian, J., & Mohamadian, A. (2009). Study for zoning the most appropriate time of irrigation of saffron (Crocus sativus) in Khorasan Razavi, north and southern provinces. *J. Water Soil.*, *23*, 109–118.

Altman, D. G., & Bland, J. M. (1994). Statistics Notes: Diagnostic tests 1: Sensitivity and specificity. *BMJ*. https://doi.org/10.1136/bmj.308.6943.1552

Arvor, D., Jonathan, M., Meirelles, M. S. P., Dubreuil, V., & Durieux, L. (2011). Classification of MODIS EVI time series for crop mapping in the state of Mato Grosso, Brazil. *International Journal of Remote Sensing*. https://doi.org/10.1080/01431161.2010.531783

Atkinson, P. M., Jeganathan, C., Dash, J., & Atzberger, C. (2012). Inter-comparison of four models for smoothing satellite sensor time-series data to estimate vegetation phenology. *Remote Sensing of Environment*, *123*, 400–417. https://doi.org/10.1016/j.rse.2012.04.001

Badhwar, G. D., Gargantini, C. E., & Redondo, F. V. (1987). Landsat classification of Argentina summer crops. *Remote Sensing of Environment*. https://doi.org/10.1016/0034-4257(87)90010-1

Basker, D., & Negbi, M. (1983). Uses of saffron. *Economic Botany*, *37*(2), 228–236. https://doi.org/10.1007/BF02858789

Beck, P. S. A., Atzberger, C., Høgda, K. A., Johansen, B., & Skidmore, A. K. (2006). Improved monitoring of vegetation dynamics at very high latitudes: A new method using MODIS NDVI. *Remote Sensing of Environment*, *100*(3), 321–334.

Bégué, A., Vintrou, E., Ruelland, D., Claden, M., & Dessay, N. (2011). Can a 25-year trend in Soudano-Sahelian vegetation dynamics be interpreted in terms of land use change? A remote sensing approach. *Global Environmental Change*. https://doi.org/10.1016/j.gloenvcha.2011.02.002

Behdani, M. A., & Fallahi, H. R. (2015). Saffron: Technical knowledge based onresearch approaches. University of Birjand Press.[In Persian]

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*(September 2017), 509–523. https://doi.org/10.1016/j.rse.2017.10.005

Belgiu, M., & Drăgu, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*. https://doi.org/10.1016/j.isprsjprs.2016.01.011

Bhojaraja, B. E., Shetty, A., Nagaraj, M. K., & Manju, P. (2016). Age-based classification of arecanut crops: a case study of Channagiri, Karnataka, India. *Geocarto International*. https://doi.org/10.1080/10106049.2015.1094528

Chang, J., Hansen, M. C., Pittman, K., Carroll, M., & DiMiceli, C. (2007). Corn and soybean mapping in the United States using MODIS time-series data sets. *Agronomy Journal*. https://doi.org/10.2134/agronj2007.0170

Chen, B., Jin, Y., & Brown, P. (2019). An enhanced bloom index for quantifying floral phenology using multi-scale remote sensing observations. *ISPRS Journal of Photogrammetry and Remote Sensing*. https://doi.org/10.1016/j.isprsjprs.2019.08.006

Chen, J., Jönsson, P., Tamura, M., Gu, Z., Matsushita, B., & Eklundh, L. (2004). A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky–Golay filter. *Remote Sensing of Environment*, *91*(3), 332–344. https://doi.org/https://doi.org/10.1016/j.rse.2004.03.014

Chittka, L., & Waser, N. M. (1997). Why red flowers are not invisible to bees. *Israel Journal of Plant Sciences*. https://doi.org/10.1080/07929978.1997.10676682

d'Andrimont, R., Taymans, M., Lemoine, G., Ceglar, A., Yordanov, M., & van der Velde, M. (2020). Detecting flowering phenology in oil seed rape parcels with Sentinel-1 and -2 time series. *Remote Sensing of Environment*, *239*(September 2019), 111660. https://doi.org/10.1016/j.rse.2020.111660

De Bernardis, C., Vicente-Guijalba, F., Martinez-Marin, T., & Lopez-Sanchez, J. M. (2016). Contribution to real-time estimation of crop phenological states in a dynamical framework based on NDVI time series: data fusion with SAR and temperature. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. https://doi.org/10.1109/JSTARS.2016.2539498

De Mastro, G., & Ruta, C. (1993). Relation between corm size and saffron (*Crocus Sativus L.*) flowering. *Acta Horticulturae*. https://doi.org/10.17660/actahortic.1993.344.58

Dehghani Bidgoli, R., Koohbanani, H. R., & bashiri, mehdi. (2018). Preparation of map for lands under saffron cultivation using timely plant's indicator based agronomic calendar (Case study: Darbeghazi Village in Neyshabur province). *Journal of Saffron Research*, *6*(1), 103–113. https://doi.org/10.22077/jsr.2018.1050.1045 [In Persian]

Dheeravath, V., Thenkabail, P. S., Chandrakantha, G., Noojipady, P., Reddy, G. P. O., Biradar, C. M., … Velpuri, M. (2010). Irrigated areas of India derived using MODIS 500 m time series for the years 2001-2003. *ISPRS Journal of Photogrammetry and Remote Sensing*, *65*(1), 42–59. https://doi.org/10.1016/j.isprsjprs.2009.08.004

Dong, J., Xiao, X., Chen, B., Torbick, N., Jin, C., Zhang, G., & Biradar, C. (2013). Mapping deciduous rubber plantations through integration of PALSAR and multi-temporal Landsat imagery. *Remote Sensing of Environment*. https://doi.org/10.1016/j.rse.2013.03.014

Eilers, P. H. C. (2003). A perfect smoother. *Analytical Chemistry*, *75*(14), 3631–3636.

ESA. (2015). *ESA's Optical High-Resolution Mission for GMES Operational Services*.

Esmaeilpour, A. M., & Kardavani, P. (2011). Saffron (Crocus sativus) potentials for sustainable rural development : A case study of Balavelayat village in Kashmar , North Eastern Iran, *6*(13), 3149–3160. https://doi.org/10.5897/AJAR11.212

Farzadmehr, J., & Bajestani, K. T. (2018). Capability of Landsat 8 satellite images to estimate the area under cultivation of saffron (case study: city of Torbat Heydarieh), *6*(1), 49–60. https://doi.org/10.22048/jsat.2017.48518.1194 [In Persian]

Fernandez, J. A. (2004). Biology, biotechnology and biomedicine of saffron. *Recent Res. Dev. Plant Sci.*, *2*, 127–159.

Foerster, S., Kaden, K., Foerster, M., & Itzerott, S. (2012). Crop type mapping using spectral-temporal profiles and phenological information. *Computers and Electronics in Agriculture*. https://doi.org/10.1016/j.compag.2012.07.015

Foody, G. M. (2002). Status of land cover classification accuracy assessment. *Remote Sensing of Environment*. https://doi.org/10.1016/S0034-4257(01)00295-4

Geerken, R. A. (2009). An algorithm to classify and monitor seasonal variations in vegetation phenologies and their inter-annual change. *ISPRS Journal of Photogrammetry and Remote Sensing*, *64*(4), 422–431.

Gelder, B. K., Kaleita, A. L., & Cruse, R. M. (2009). Estimating mean field residue cover on midwestern soils using satellite imagery. *Agronomy Journal*. https://doi.org/10.2134/agronj2007.0249

Ghorbani, M. (2007). The economics of saffron in Iran. *Acta Horticulturae*, *739*, 321–331. https://doi.org/10.17660/ActaHortic.2007.739.42

Ghorbani, R., & Koocheki, A. (2017). Sustainable Cultivation of Saffron in Iran. In E. Lichtfouse (Ed.), *Sustainable Agriculture Reviews* (pp. 169–203). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-58679-3_6

GIAHS. (2018). *GonabasSaffron_IranGIAHSProposal* (Vol. 23). https://doi.org/10.1177/1758998318809574

Gislason, P. O., Benediktsson, J. A., & Sveinsson, J. R. (2006). Random forests for land cover classification. In *Pattern Recognition Letters*. https://doi.org/10.1016/j.patrec.2005.08.011

Gitelson, A. A., Kaufman, Y. J., Stark, R., & Rundquist, D. (2002). Novel algorithms for remote estimation of vegetation fraction. *Remote Sensing of Environment*. https://doi.org/10.1016/S0034-4257(01)00289-9

Gitelson, A., & Merzlyak, M. N. (1994). Spectral reflectance changes associated with autumn senescence of aesculus hippocastanum L. and acer platanoides L. Leaves. Spectral features and relation to chlorophyll estimation. *Journal of Plant Physiology*. https://doi.org/10.1016/S0176-1617(11)81633-0

Goldstein, B. A., Polley, E. C., & Briggs, F. B. S. (2011). Random forests for genetic association studies. *Statistical Applications in Genetics and Molecular Biology*. https://doi.org/10.2202/1544-6115.1691

Gómez, C., White, J. C., & Wulder, M. A. (2016). Optical remotely sensed time series data for land cover classification: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, *116*, 55–72. https://doi.org/10.1016/j.isprsjprs.2016.03.008

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*. https://doi.org/10.1016/j.rse.2017.06.031

Gresta, F., Lombardo, M., Siracusa, L., & Ruberto, G. (2008). Saffron, an alternative crop for sustainable agricultural systems. A review. *Agronomy for Sustainable Development*. https://doi.org/10.1051/agro:2007030

Gresta, Fabio, Santonoceto, C., & Avola, G. (2016). Crop rotation as an effective strategy for saffron (Crocus sativus L.) cultivation. *Scientia Horticulturae*, *211*, 34–39. https://doi.org/10.1016/j.scienta.2016.08.007

Hao, P., Zhan, Y., Wang, L., Niu, Z., & Shakir, M. (2015). Feature selection of time series MODIS data for early crop classification using random forest: A case study in Kansas, USA. *Remote Sensing*, *7*(5), 5347–5369. https://doi.org/10.3390/rs70505347

Hardisky, M. A., Klemas, V., & Smart, R. M. (1983). The influence of soil salinity, growth form, and leaf moisture on the spectral radiance of Spartina alterniflora canopies. *Photogrammetric Engineering & Remote Sensing*.

Hird, J. N., & McDermid, G. J. (2009). Noise reduction of NDVI time series: An empirical comparison of selected techniques. *Remote Sensing of Environment*, *113*(1), 248–258.

Hosseini, M., Mollafilabi, A., & Nassiri, M. (2008). Spatial and temporal patterns in saffron (Crocus Sativus L.) yield of Khorasan Province and their relationship with long term weather variation. *Iranian journal of field crops research*, *6*(1), 79–88. [In Persian]

Howard, D. M., & Wylie, B. K. (2014). Annual crop type classification of the US great plains for 2000 to 2011. *Photogrammetric Engineering and Remote Sensing*. https:d//doi.org/10.14358/PERS.80.6.537-549

Hu, Q., Sulla-Menashe, D., Xu, B., Yin, H., Tang, H., Yang, P., & Wu, W. (2019). A phenology-based spectral and temporal feature selection method for crop mapping from satellite time series. *International Journal of Applied Earth Observation and Geoinformation*, *80*(May), 218–229. https://doi.org/10.1016/j.jag.2019.04.014

Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., & Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*. https://doi.org/10.1016/S0034-4257(02)00096-2

IPCC. (2014). Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. *Ipcc*.

Jasper, J., Reusch, S., & Link, A. (2009). Active sensing of the N status of wheat using optimized wavelength combination: Impact of seed rate, variety and growth stage. In *Precision Agriculture 2009 - Papers Presented at the 7th European Conference on Precision Agriculture, ECPA 2009*.

Jiang, Z., Huete, A. R., Chen, J., Chen, Y., Li, J., Yan, G., & Zhang, X. (2006). Analysis of NDVI and scaled difference vegetation index retrievals of vegetation fraction. *Remote Sensing of Environment*. https://doi.org/10.1016/j.rse.2006.01.003

Jiang, Z., Huete, A. R., Didan, K., & Miura, T. (2008). Development of a two-band enhanced vegetation index without a blue band. *Remote Sensing of Environment*. https://doi.org/10.1016/j.rse.2008.06.006

Jin, S., Su, Y., Gao, S., Hu, T., Liu, J., & Guo, Q. (2018). The transferability of Random Forest in canopy height estimation from multi-source remote sensing data. *Remote Sensing*. https://doi.org/10.3390/rs10081183

Jonsson, P., & Eklundh, L. (2002). Seasonality extraction by function fitting to time-series of satellite sensor data. *IEEE Transactions on Geoscience and Remote Sensing*, *40*(8), 1824–1832.

Juel, A., Groom, G. B., Svenning, J. C., & Ejrnæs, R. (2015). Spatial application of Random Forest models for fine-scale coastal vegetation classification using object based analysis of aerial orthophoto and DEM data. *International Journal of Applied Earth Observation and Geoinformation*. https://doi.org/10.1016/j.jag.2015.05.008

Khazaei, M., Monfared, M., Kamgar, H. A. L. I. A., & Sepaskhah, A. (2013). The trend of change for weight and number of saffron corms as affected by irrigation frequency and method in different years. *Journal of saffron research*. [In Persian]

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *International Joint Conference of Artificial Intelligence*.

Koocheki, A., & Khajeh-Hosseini, M. (2019). *Saffron: Science, Technology and Health. Woodhead Publishing Series in Food Science, Technology and Nutrition*

Kumar, R., Singh, V., Devi, K., Sharma, M., Singh, M. K., & Ahuja, P. S. (2009). State of art of saffron (Crocus sativus L.) agronomy: A comprehensive review. *Food Reviews International* (Vol. 25). https://doi.org/10.1080/87559120802458503

Lemoine, G., & Leo, O. (2015). Crop mapping applications at scale: Using Google Earth Engine to enable global crop area and status monitoring using free and open data sources. In *International Geoscience and Remote Sensing Symposium (IGARSS)*. https://doi.org/10.1109/IGARSS.2015.7326063

Lichtfouse, E. (2013). *Sustainable Agriculture Reviews* (Vol. 12). https://doi.org/10.1007/978-94-007-5961-9

Lobell, D. B., & Asner, G. P. (2004). Cropland distributions from temporal unmixing of MODIS data. *Remote Sensing of Environment*, *93*(3), 412–422. https://doi.org/10.1016/j.rse.2004.08.002

Loosvelt, L., Peters, J., Skriver, H., De Baets, B., & Verhoest, N. E. C. (2012). Impact of reducing polarimetric SAR input on the uncertainty of crop classifications based on the random forests algorithm. *IEEE Transactions on Geoscience and Remote Sensing*. https://doi.org/10.1109/TGRS.2012.2189012

Lopez Corcoles, H., Brasa Ramos, A., Montero García, F., Romero Valverde, M., & Montero Riquelme, F. (2015). Phenological growth stages of saffron plant (*Crocus sativus L.*) according to the BBCH scale. *Spanish Journal of Agricultural Research*, *13*(3). https://doi.org/10.5424/sjar/2015133-7340

Löw, F., Michel, U., Dech, S., & Conrad, C. (2013). Impact of feature selection on the accuracy and spatial uncertainty of per-field crop classification using Support Vector Machines. *ISPRS Journal of Photogrammetry and Remote Sensing*, *85*, 102–119. https://doi.org/10.1016/j.isprsjprs.2013.08.007

M.A.Behdani, A.Koocheki, P.Rezvani, & AL-Ahmadi, M. J. (2009). Agro-ecological zoning and potential yield of saffron in Khorasan-Iran. *Journal of Biological Sciences*. https://doi.org/10.3923/jbs.2008.298.305

Mansouri Daneshvar, M. R., Ebrahimi, M., & Nejadsoleymani, H. (2019). An overview of climate change in Iran: facts and statistics. *Environmental Systems Research*, *8*(1). https://doi.org/10.1186/s40068-019-0135-3

Maus, V., CĂmara, G., Cartaxo, R., Sanchez, A., Ramos, F. M., & De Queiroz, G. R. (2016). A time-weighted dynamic time warping method for land-use and land-cover Mapping. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. https://doi.org/10.1109/JSTARS.2016.2517118

McFeeters, S. K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing*. https://doi.org/10.1080/01431169608948714

Mehrdad, N. A. (2014). The agricultural survey improvement program in Islamic Republic of Iran., (No.2014/28), 181–191.

Molina, R. V, Valero, M., & Navarro, Y. (2005). Temperature effects on flower formation in saffron (Crocus sativus L .), *103*, 361–379. https://doi.org/10.1016/j.scienta.2004.06.005

Motagh, M., Walter, T. R., Sharifi, M. A., Fielding, E., Schenk, A., Anderssohn, J., & Zschau, J. (2008). Land subsidence in Iran caused by widespread water reservoir overexploitation. *Geophysical Research Letters*. https://doi.org/10.1029/2008GL033814

Myneni, R. B., Hall, F. G., Sellers, P. J., & Marshak, A. L. (1995). Interpretation of spectral vegetation indexes. *IEEE Transactions on Geoscience and Remote Sensing*. https://doi.org/10.1109/36.377948

Negbi, M. (1999). Saffron cultivation: past, present and future prospects. In *Saffron Crocus sativus L.* (1st Editio, p. 148). London. https://doi.org/https://doi.org/10.1201/9780203303665

Negbi, M., Dagan, B., Dror, A., & Basker, D. (1989). Growth, flowering, vegetative reproduction, and dormancy in the saffron crocus (Crocus sativus l). *Israel Journal of Botany*. https://doi.org/10.1080/0021213X.1989.10677116

Ok, A. O., Akar, O., & Gungor, O. (2012). Evaluation of random forest method for agricultural crop classification. *European Journal of Remote Sensing*, *45*(1), 421–432. https://doi.org/10.5721/EuJRS20124535

Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, *26*(1), 217–222. https://doi.org/10.1080/01431160412331269698

Pal, Mahesh, & Foody, G. M. (2010). Feature selection for classification of hyperspectral data by SVM. *IEEE Transactions on Geoscience and Remote Sensing*, *48*(5), 2297–2307. https://doi.org/10.1109/TGRS.2009.2039484

Pan, Y., Li, L., Zhang, J., Liang, S., Zhu, X., & Sulla-Menashe, D. (2012). Winter wheat area estimation from MODIS-EVI time series data using the Crop Proportion Phenology Index. *Remote Sensing of Environment*. https://doi.org/10.1016/j.rse.2011.10.011

Peel, M. C., Finlayson, B. L., & McMahon, T. A. (2007). Updated world map of the Köppen-Geiger climate classification. *Hydrology and Earth System Sciences*, *11*(5), 1633–1644. https://doi.org/10.5194/hess-11-1633-2007

Peña-Barragán, J. M., Ngugi, M. K., Plant, R. E., & Six, J. (2011). Object-based crop identification using multiple vegetation indices, textural features and crop phenology. *Remote Sensing of Environment*, *115*(6), 1301–1316. https://doi.org/10.1016/j.rse.2011.01.009

Pittman, K., Hansen, M. C., Becker-Reshef, I., Potapov, P. V., & Justice, C. O. (2010). Estimating global cropland extent with multi-year MODIS data. *Remote Sensing*, *2*(7), 1844–1863. https://doi.org/10.3390/rs2071844

Qi, J., Marsett, R., Heilman, P., Biedenbender, S., Moran, S., Goodrich, D., & Weltz, M. (2002). RANGES improves satellite-based information and land cover assessments in Southwest United States. *Eos*. https://doi.org/10.1029/2002EO000411

Rahimzadegan, M., & Pourgholam, M. (2017). Identification of the area under cultivation of saffron using Landsat-8 temporal satellite images (Case study: Torbat Heydarieh). *Journal of RS and GIS for natural resources*, *7*(4). [In Persian]

Richards, M. A. (2005). Introduction to radar system. In *Fundamentals of Radar Signal Processing* (pp. 45–47). New York, USA: Tata McGraw-Hill Education.

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*. https://doi.org/10.1016/j.isprsjprs.2011.11.002

Rouse Jr., J. W., Haas, R. H., Schell, J. A., & Deering, D. W. (1974). Monitoring vegetation systems in the great plains with erts. In *NASA SP-351, 3rd ERTS-1 Symposium*.

Sellers, P. J., Tucker, C. J., Collatz, G. J., Los, S. O., Justice, C. O., Dazlich, D. A., & Randall, D. A. (1994). A global 1 by 1 NDVI data set for climate studies. Part 2: The generation of global fields of terrestrial biophysical parameters from the NDVI. *International Journal of Remote Sensing*, *15*(17), 3519–3545.

Sepaskhah, A. R., & Yarami, N. (2009). Interaction effects of irrigation regime and salinity on flower yield and growth of saffron. *Journal of Horticultural Science and Biotechnology*. https://doi.org/10.1080/14620316.2009.11512507

Skakun, S., Kussul, N., Shelestov, A. Y., Lavreniuk, M., & Kussul, O. (2016). 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*. https://doi.org/10.1109/JSTARS.2015.2454297

Sonobe, R., Tani, H., Wang, X., Kobayashi, N., & Shimamura, H. (2014). Random forest classification of crop type using multi-temporal TerraSAR-X dual-polarimetric data. *Remote Sensing Letters*, *5*(2), 157–164. https://doi.org/10.1080/2150704X.2014.889863

Stendardi, L., Karlsen, S. R., Niedrist, G., Gerdol, R., Zebisch, M., Rossi, M., & Notarnicola, C. (2019). Exploiting time series of Sentinel-1 and Sentinel-2 imagery to detect meadow phenology in mountain regions. *Remote Sensing*, *11*(5), 1–24. https://doi.org/10.3390/rs11050542

Tatsumi, K., Yamashiki, Y., Canales Torres, M. A., & Taipe, C. L. R. (2015). Crop classification of upland fields using Random forest of time-series Landsat 7 ETM+ data. *Computers and Electronics in Agriculture*, *115*, 171–179. https://doi.org/10.1016/j.compag.2015.05.001

Tian, H., Huang, N., Niu, Z., Qin, Y., Pei, J., & Wang, J. (2019). Mapping winter crops in China with multi-source satellite imagery and phenology-based algorithm. *Remote Sensing*, *11*(7), 1–23. https://doi.org/10.3390/rs11070820

Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*. https://doi.org/10.1016/0034-4257(79)90013-0

Turker, M., & Arikan, M. (2005). Sequential masking classification of multi-temporal Landsat7 ETM+ images for field-based crop mapping in Karacabey, Turkey. *International Journal of Remote Sensing*. https://doi.org/10.1080/01431160500166391

Van Deventer, A. P., Ward, A. D., Gowda, P. M., & Lyon, J. G. (1997). Using thematic mapper data to identify contrasting soil plains and tillage practices. *Photogrammetric Engineering and Remote Sensing*.

Van Tricht, K., Gobin, A., Gilliams, S., & Piccard, I. (2018). Synergistic use of radar sentinel-1 and optical sentinel-2 imagery for crop mapping: A case study for Belgium. *Remote Sensing*, *10*(10), 1–22. https://doi.org/10.3390/rs10101642

Veloso, A., Mermoz, S., Bouvet, A., Le Toan, T., Planells, M., Dejoux, J. F., & Ceschia, E. (2017). Understanding the temporal behavior of crops using Sentinel-1 and Sentinel-2-like data for agricultural applications. *Remote Sensing of Environment*, *199*, 415–426. https://doi.org/10.1016/j.rse.2017.07.015

Vrieling, A., Meroni, M., Darvishzadeh, R., Skidmore, A. K., Wang, T., Zurita-Milla, R., … Paganini, M. (2018). Vegetation phenology from Sentinel-2 and field cameras for a Dutch barrier island. *Remote Sensing of Environment*, *215*(November 2017), 517–529. https://doi.org/10.1016/j.rse.2018.03.014

Wardlow, B. D., Egbert, S. L., & Kastens, J. H. (2007). Analysis of time-series MODIS 250 m vegetation index data for crop classification in the U.S. Central Great Plains. *Remote Sensing of Environment*, *108*(3), 290–310. https://doi.org/10.1016/j.rse.2006.11.021

Winterhalter, P., & Straubinger, M. (2000). Saffron—renewed interest in an ancient spice. *Food Reviews International*, *16*(1), 39–59. https://doi.org/10.1081/FRI-100100281

Xie, Q., Dash, J., Huete, A., Jiang, A., Yin, G., Ding, Y., … Huang, W. (2019). Retrieval of crop biophysical parameters from Sentinel-2 remote sensing imagery. *International Journal of Applied Earth Observation and Geoinformation*, *80*(May), 187–195. https://doi.org/10.1016/j.jag.2019.04.019

Yasmin, S., & Nehvi, F. A. (2018). Phenological Growth Stages of Saffron (*Crocus sativus L.*) under Temperate Conditions of Jammu & Kashmir-India. *International Journal of Current Microbiology and Applied Sciences*. https://doi.org/10.20546/ijcmas.2018.704.428

Zahmati, R., Shekari, H. A., & Fotokian, M. H. (2018). Growth and development of saffron (*Crocus sativus L* .) in response to temperature pre-treatment and environmental conditions, *7*(1), 47–50.

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*. https://doi.org/10.1016/j.rse.2013.08.023

APPENDIX

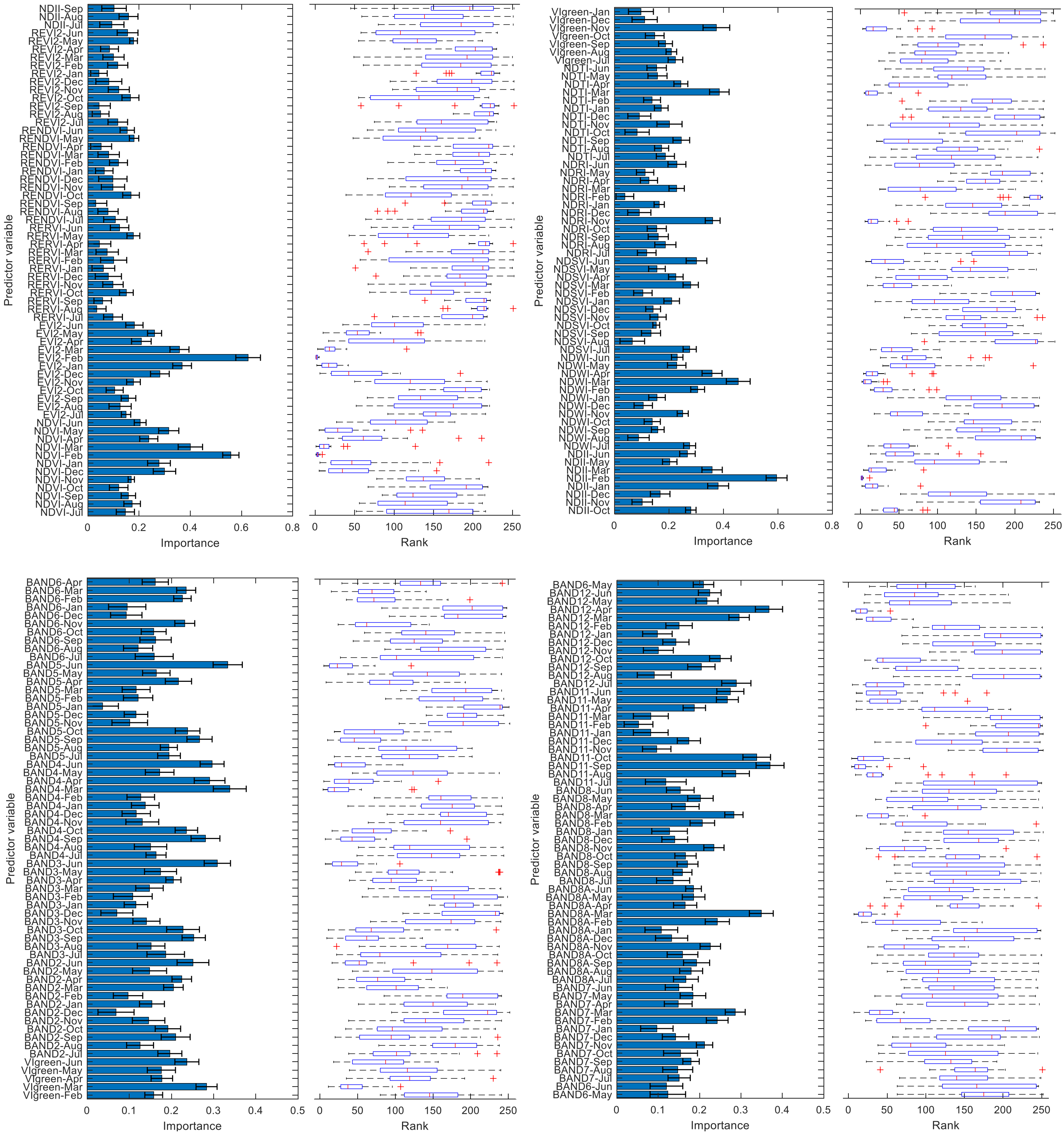


Figure A1. The importance values and ranks of 252 spectral-temporal features in the 50 runs of RF-based saffron field classification. The blue bar shows the average importance score which scaled from 0 to 1, and the error bar presents its corresponding standard deviation in the 50 runs. The box plot reflects the 50 times ranks for each feature.