

Article

Harmonized Landsat 8 and Sentinel-2 Time Series Data to Detect Irrigated Areas: An Application in Southern Italy

Salvatore Falanga Bolognesi ¹, Edoardo Pasolli ^{2,*}, Oscar Rosario Belfiore ²,
Carlo De Michele ¹ and Guido D'Urso ²

¹ Ariespace s.r.l., Centro Direzionale IS. A3, 80143 Naples, Italy; salvatore.falanga@ariespace.com (S.F.B.); carlo.demichele@ariespace.com (C.D.M.)

² Department of Agricultural Sciences, University of Naples Federico II, Via Università 100, 80055 Portici, Naples, Italy; oscarrosario.belfiore@unina.it (O.R.B.); durso@unina.it (G.D.)

* Correspondence: edoardo.pasolli@unina.it; Tel.: +39-0812539397

Received: 3 March 2020; Accepted: 14 April 2020; Published: 17 April 2020



Abstract: Lack of accurate and up-to-date data associated with irrigated areas and related irrigation amounts is hampering the full implementation and compliance of the Water Framework Directive (WFD). In this paper, we describe the framework that we developed and implemented within the DIANA project to map the actual extent of irrigated areas in the Campania region (Southern Italy) during the 2018 irrigation season. For this purpose, we considered 202 images from the Harmonized Landsat Sentinel-2 (HLS) products (57 images from Landsat 8 and 145 images from Sentinel-2). Such data were preprocessed in order to extract a multitemporal Normalized Difference Vegetation Index (NDVI) map, which was then smoothed through a gap-filling algorithm. We further integrated data coming from high-resolution (4 km) global satellite precipitation Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN)-Cloud Classification System (CCS) products. We collected an extensive ground truth in the field represented by 2992 data points coming from three main thematic classes: bare soil and rainfed (class 0), herbaceous (class 1), and tree crop (class 2). This information was exploited to generate irrigated area maps by adopting a machine learning classification approach. We compared six different types of classifiers through a cross-validation approach and found that, in general, random forests, support vector machines, and boosted decision trees exhibited the best performances in terms of classification accuracy and robustness to different tested scenarios. We found an overall accuracy close to 90% in discriminating among the three thematic classes, which highlighted promising capabilities in the detection of irrigated areas from HLS products.

Keywords: Harmonized Landsat 8 and Sentinel-2; irrigated areas; machine learning; classification

1. Introduction

Availability of data about irrigated areas is of extreme importance in many different applications including management of water resources [1], modeling of water exchange between atmosphere and land surface [2], and impact of climate change on irrigation water supplies [3]. Despite the great effort conducted in the literature in this direction, there is a general lack of accurate and up-to-date maps about irrigated areas and related irrigation amounts, which is hampering the full implementation and compliance of the Water Framework Directive (WFD). Current monitoring practices rely on acquiring periodical surveys meant to give pictures at national scales. However, they are rather imprecise when regional or local scale resolutions need to be achieved, as the case of management of water resources in

hydrographic basins. At this regard, different actions have been conducted at both local and national levels. An example is represented by the specific actions adopted by the Italian Ministry of Agriculture (Decree 31/07/2015) aimed at monitoring irrigation areas and volumes on a regular basis in order to improve the compliance to the WFD.

Multiple studies have been conducted in the literature to map irrigated areas using different types of sensors, a task that has been frequently obtained through the development and application of machine learning classification approaches. As example, random forests (RFs) were applied to Landsat Thematic Mapper (TM)/Landsat Enhanced Thematic Mapper (ETM+) in conjunction with Moderate Resolution Imaging Spectroradiometer (MODIS) monthly acquisitions to map irrigated areas and summer crops in the Murray–Darling basin [4]. A threshold-dependent decision tree algorithm was used to map irrigated areas in Afghanistan from 2000 to 2013 with 16-day composites of the Normalized Difference Vegetation Index (NDVI) derived from MODIS [5]. A vegetation phenological-based method was developed to differentiate soil water irrigated areas from surface water irrigated areas using 8-day time series from 250-m Moderate Resolution Imaging Spectroradiometer (MODIS) data in the Krishna River basin (India) during the 2000–2001 time span [6]. Combining geospatial irrigation statistics with remotely sensed parameters describing vegetation growth conditions was adopted to identify irrigated areas in agricultural regions with a 250-m spatial resolution in the United States in 2002 [7]. Landsat ETM+ images and MODIS time series data were used to map multiple land use/land cover (LULC) classes including irrigated agricultural areas in Ghana [8]. Discrimination between irrigated and not irrigated crops was done by extracting two metrics from evapotranspiration fluxes and vegetation indices using empirical distribution functions [9].

In the very few last years, growing attention has been given to the Harmonized Landsat 8 OLI and Sentinel-2 MSI (HLS) project, which has been launched with the aim at harmonizing data coming from these two popular satellites [10,11]. Recent methodological advancements have allowed scientists to use this type of product in a growing number of research areas. As example, an HLS dataset was used to remotely control landscape dynamics in Central US grasslands, which improved the ability to track grassland dynamics in season including areas vulnerable to cloud contamination [12]. HLS data was also used to accurately classify land-surface phenology, while excluding imagery from either sensors dramatically reduced the ability to monitor dryland environments [13]. Assessment of the winter wheat yield at regional scale was done using HLS data in [14]. Field level phenologies were captured through a new method optimized for HLS data and able to extract time series of consistent reflectance compositions [15]. HLS datasets were also used to estimate the number and timing of mowing events in Central European grasslands [16], detect crop intensities [17], map land surface phenologies at continental scale with 30-m resolution [18], and compute seasonal vegetation index dynamics for time series analysis [19]. Despite the growing interest in HLS products, there is the lack of papers in the literature focused on their use for detection and characterization of irrigated areas.

In this paper, we present the framework that we developed, implemented, and validated to map the actual extent of irrigated areas in the Campania region (Southern Italy) during the 2018 irrigation season. This was done by considering a machine learning classification-based approach based on time series images coming from HLS data. The procedure described in this paper extended a preliminary analysis that was conducted within the DIANA Project (<http://diana-h2020.eu/>) [20] aimed at mapping irrigated areas within the extent of the Land Reclamation Consortium located in Campania region using Sentinel-2 data, and that therefore was extended here to the mapping of the entire Campania region through HLS data products.

HLS Project

Multispectral Earth Observation data derived from Landsat 8 NASA's Satellite and Sentinel-2 ESA's Satellite represent a precious information source for their large spectral range, from the visible to the short-wave infrared, and the high revisit time. They are characterized by similar spectral and spatial characteristics, which can result in a relevant increase in the number of observations when used together.

The HLS project [10,11] is meant to harmonize data coming from both satellites programs to ease their combined use. The program aims at removing the biases between the two sensors in terms of different spectral band ranges and view geometries with the final goal to obtain a global land surface coverage at 30 m spatial resolution with a time gap of 2–3 days.

The workflow for the HLS data is schematized in [11]. The process that leads to the final HLS outputs starts by applying the same atmospheric correction algorithm to both Landsat 8 OLI (L1T) and Sentinel-2 MSI (L1C) products. Then, Landsat 8 data are geographically divided in accordance with Sentinel-2 data tiling, and both are geometrically resampled using the Automated Registration and Orthorectification Package (AROP) [21]. This step arises from the fact that the view angles between sensors can change for a given ground target, and therefore the angle effect (bidirectional reflectance distribution function, or BRDF) needs to be normalized. The last step consists in applying the band pass adjustment (based on a linear fit between equivalent spectral bands) to the Sentinel-2 data only using the OLI spectral band passes as reference. The resulting outputs are the L30 (OLI NBAR 30m), the S10 (MSI SR 10m), and the S30 (MSI NBAR 30m) data products, but only the first and third types of products were considered in this work.

2. Materials and Methods

2.1. Dataset

The study area is the Campania region (Southern Italy), monitored through the 2018 irrigation season (from April to October).

For this purpose, we considered the HLS data (version 1.4) that resulted in four distinct tiles (T33TVE, T33TVF, T33TWE, and T33TWF) (Figure 1). We collected an initial number of 650 images that spanned the four tiles in the January–December time span of year 2018. This number was reduced to 210 images by removing those images that were partially or entirely covered by clouds or that overlapped only marginally (<20%) with the area of interest. We further removed 12 images in which both Sentinel-2 and Landsat 8 images were acquired at the same date, and the Sentinel-2 data was kept in this case. We finally added one more image for each tile from the end of year 2017 in order to prevent issues in applying the gap-filling algorithm due to cloud-covered data as we will describe later. This resulted in a final number of 202 images, with 141 of them coming from Sentinel-2 and 57 from Landsat 8. The inclusion of Landsat 8 data allowed us for more than 40% increase in the number of images available in our study. Such images were spread across the four tiles as follows: T33TVE (N = 42), T33TVF (N = 74), T33TWE (N = 45), and T33TWF (N = 41). The whole data selection process is summarized in Table 1. The mean effective revisit time spanned was 4.9 and 8.9 for T33TVF and T33TWE, respectively (Table 1), which highlighted a temporal resolution that was appropriate for the application involved in this study.

Table 1. Summary of the Harmonized Landsat Sentinel-2 (HLS) data selection process. (a) Number of initial images spanning year 2018. (b) Number of images after removing data covered by clouds or marginally overlapping (<20%) with the area of interest. (c) Number of final images used in the analysis in addition with the revisit time (in days) for each tile.

	Satellite/Tile	T33TVE	T33TVF	T33TWE	T33TWF	Total
(a)	L30	49	44	42	44	179
	S30	118	117	121	115	471
	Total	167	161	163	159	650
(b)	L30	11	16	17	13	57
	S30	30	57	27	27	141
	Total	41	73	44	40	198
(c)	L30	11	16	17	13	57
	S30	31	58	28	28	145
	Total	42	74	45	41	202
	Revisit time	8.7	4.9	8.1	8.9	–

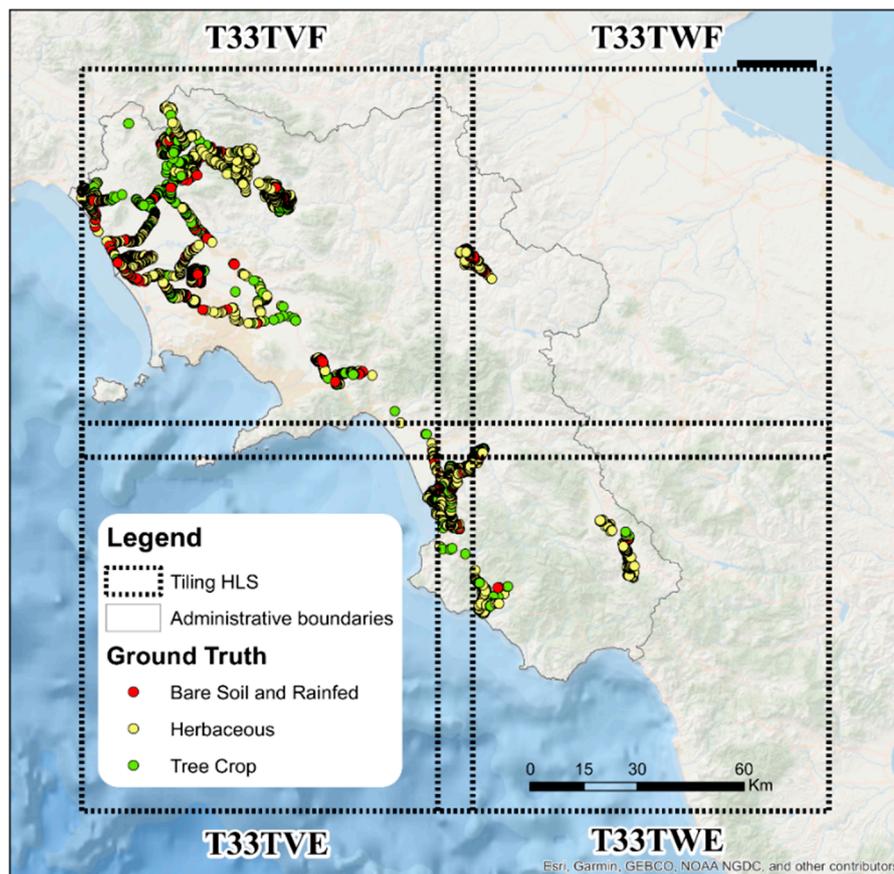


Figure 1. The four tiles on Campania region from Harmonized Landsat Sentinel-2 (HLS) products that we considered in this work. Colored points represent ground truth samples that will be further described in Figures 7 and 8.

2.2. Satellite Data Collection and Preprocessing

The full set of bands and associated wavelengths acquired by Sentinel-2 and Landsat 8 are listed in [11]. We considered here the red (0.64–0.67 μm) and NIR narrow (0.85–0.88 μm) bands, which were used to compute the NDVI [22]. The NDVI can be used as a good indicator for irrigated areas since it represents the amount of green biomass that varies in response to changes in vegetation conditions [23].

The raw NDVI values were processed through a gap-filling algorithm in order to overcome the presence of missing data generated from cloud covers. For this purpose, we first considered the Quality Assessment (QA) layer derived from the cloud masks [11] (Table 2). From this, we extracted integer values (Table 3), and derived the Binary Mask to discriminate between Land (1) and no-Land (0) classes, with the latter one dominated by the information on the cloud cover.

Table 2. Description of the bits in the one-byte Quality Assessment layer for the three products.

Bit Number	QA Description	Bit Combination (Description)
7–6	Aerosol quality	00 (Climatology), 01 (Low), 10 (Average), 11 (High)
5	Water	0 (No), 1 (Yes)
4	Snow/ice	0 (No), 1 (Yes)
3	Cloud shadow	0 (No), 1 (Yes)
2	Adjacent cloud	0 (No), 1 (Yes)
1	Cloud	0 (No), 1 (Yes)
0	Cirrus	0 (No), 1 (Yes)

Table 3. Integer values selected from the Quality Assessment (QA) layer in order to derive the no-Land class associated with the binary mask.

Integer Value	Bit7	Bit6	Bit5	Bit4	Bit3	Bit2	Bit1	Bit0
0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	1	0	0
64	0	1	0	0	0	0	0	0
68	0	1	0	0	0	1	0	0
128	1	0	0	0	0	0	0	0
132	1	0	0	0	0	1	0	0
192	1	1	0	0	0	0	0	0
196	1	1	0	0	0	1	0	0

At this point, the NDVI value of pixels belonging to the no-Land class was estimated through the *smoothing* and *gap-filling* process based on the Whittaker smoother (WS) [24] function available in the MODIS package [25] into the R environment [26]. In the WS filtering process, we set the main parameters empirically as follows: (i) the number of iteration = 1, which implied to perform two running cycles; (ii) the lambda value (i.e., the smoothing factor) = 10, which guaranteed to preserve the temporal variability of specific phenological patterns; and (iii) the number of days = 5. Finally, in order to avoid artifacts in the NDVI values resulting from undetected cloud and poor atmospheric conditions, we forced their values to be constrained in a range defined on the cloud-free multitemporal NDVI series. An example of the derived product is shown in Figure 2.

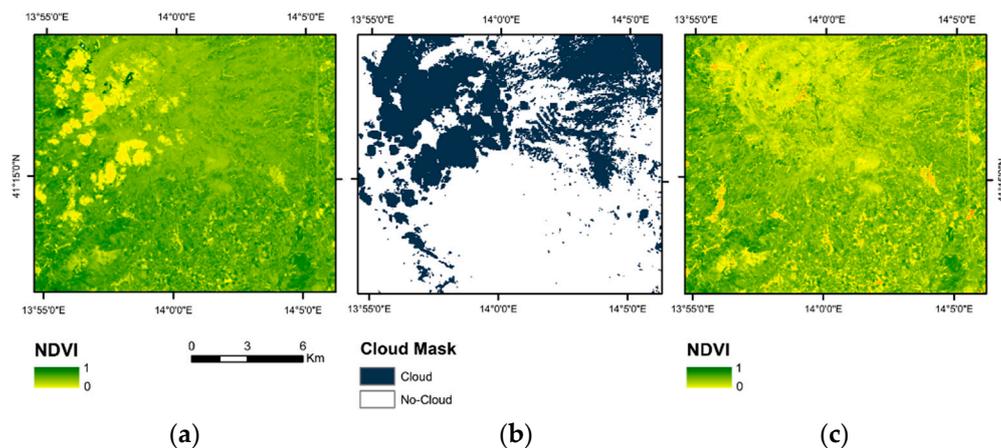


Figure 2. Example of derived product referred to the area of interest (tile T33TVF) from Landsat 8 acquired on the day-of-the-year 48 of year 2018: (a) raw Normalized Difference Vegetation Index (NDVI) data, (b) Cloud Mask, and (c) filtered NDVI.

At this stage, we obtained a multitemporal, filtered, and cloud-free NDVI series data with a 5-day time step. This resulted in a total of 75 observations (features) per pixel spanning the entire year 2018, which were able to preserve crop phenologies and limit differences in terms of revisit time existing across the four tiles covering the Campania region. An example of the NDVI time series product is shown in Figure 3 for different crop types. In such an example, we can observe the distinctive NDVI patterns that distinguish the herbaceous (two cases in panels a and b) from the tree class (one case in panel c). Of relevance is also the importance of the WS algorithm. The raw NDVI decreases when clouds are present with an intensity proportional to the density of the cloudy body. On the other hand, it increases in shadowed areas created by clouds as the values in the red band have larger decreases than in the NIR band. Therefore, the application of the WS algorithm is relevant to smooth such inconsistencies that may occur in the data due to the presence of clouds.

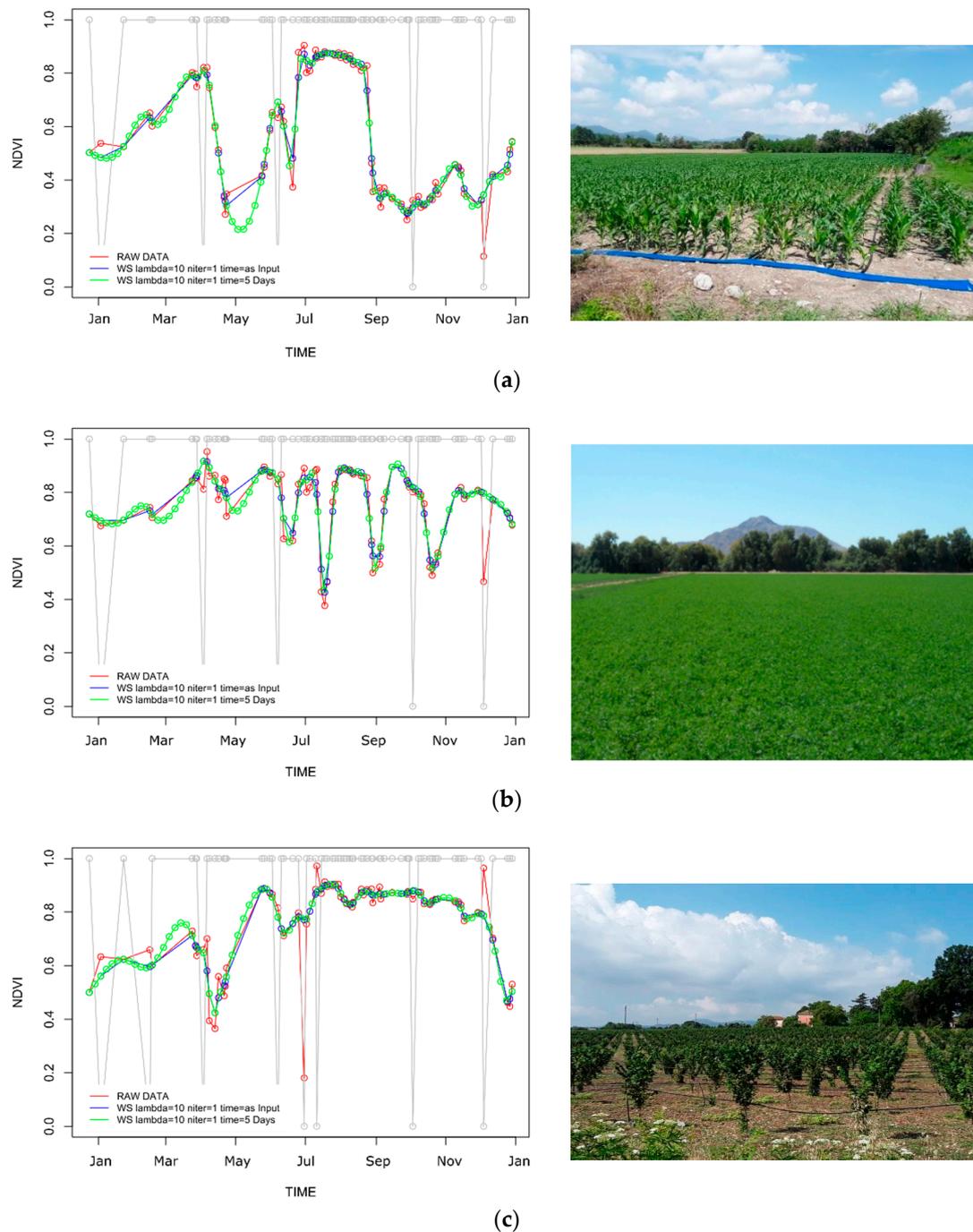


Figure 3. NDVI (left panel) and picture (right panel) for different crop types: (a) herbaceous (maize), (b) herbaceous (alfalfa), and (c) tree crop. In the NDVI panel, the red line represents the raw NDVI data, the blue line shows the NDVI after applying smoothing and gap-filling based on the by Whittaker smoother (WS) algorithm on the original dates (i.e., using the acquisition dates with variable time sampling), and the green line depicts the NDVI by running WS with a constant sampling of 5 days. The WS algorithm was run by applying two filtering iterations (niter = 1) in order to limit the effect of undetected clouds and poor atmospheric conditions. The smoothing parameter (lambda) was limited to 10 in order to preserve the fidelity of the input data in the final resulting curve, which was particularly relevant for the alfalfa herbaceous class as shown in panel (b). The gray line indicates the presence (value = 0) or absence (value = 1) of clouds for a specific date.

As final step, we applied a regional-scale spatial aggregation algorithm. The four tiles were merged together into a single mosaic layer using GDAL [27] for each time step (Figure 4).

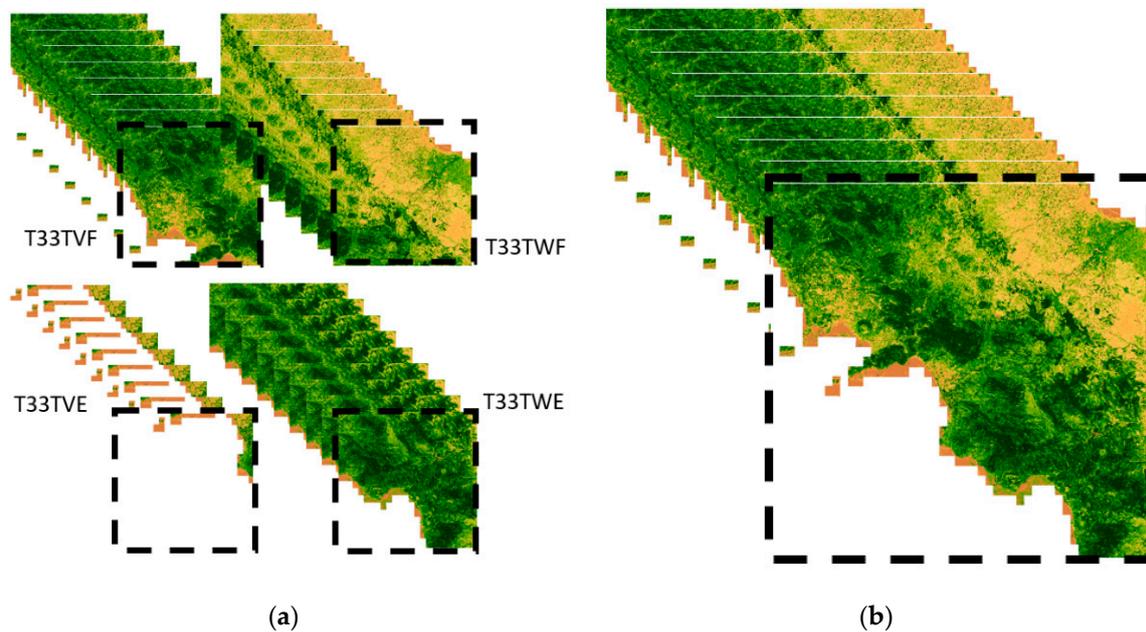


Figure 4. (a) The different tiles and images involved in the spatial aggregation process. (b) The final product after the spatial aggregation algorithm composed by a total of 75 images/observations spanning the year 2018.

2.3. Satellite Precipitation Data Collection and Preprocessing

The NDVI product described in the previous section was integrated with satellite precipitation data in order to ease the discrimination between irrigated and not irrigated areas. For this purpose, we considered the Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN)-Cloud Classification System (PERSIANN-CCS), a real-time global high-resolution ($0.04^\circ \times 0.04^\circ$ or $4 \text{ km} \times 4 \text{ km}$) satellite precipitation product developed by the Center for Hydrometeorology and Remote Sensing (CHRS) at the University of California, Irvine (UCI) that reports the cumulative daily rainfall [28]. The raw PERSIANN-CCS data was processed to be consistent with the NDVI data. More specifically, we first performed a temporal aggregation with a time interval of 5 days, and then cropped and reprojected it in accordance with the HLS grid (merging of tiles T33TVF, T33TWF, T33TVE, and T33TWE) derived as described in the previous sections. An example of the PERSIANN-CSS product is shown in Figure 5.

We further reported an example of the NDVI time series product overlapped with the accumulated rainfall data (Figure 6). Such example highlighted the improved capabilities in discriminating between rainfed and irrigated areas when the two data types were integrated. When the NDVI trend was in phase with the accumulated rainfall, the area was considered as “not irrigated or rainfed” as the crop growth was strictly correlated to the rainfall events (Figure 6a). On the other hand, when the NDVI trend was not synchronized with the accumulated rainfall, the area was considered as “irrigated” as the crop growth resulted independent from the accumulated rainfall (Figure 6b).

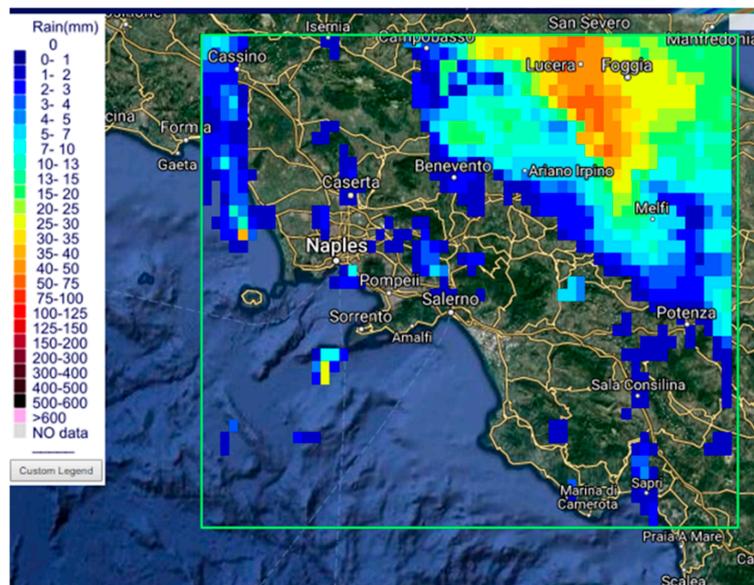


Figure 5. The Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN)-Cloud Classification System (CCS) product acquired on 1 January 2018 and covering the same area associated with the HLS data shown in Figure 4.

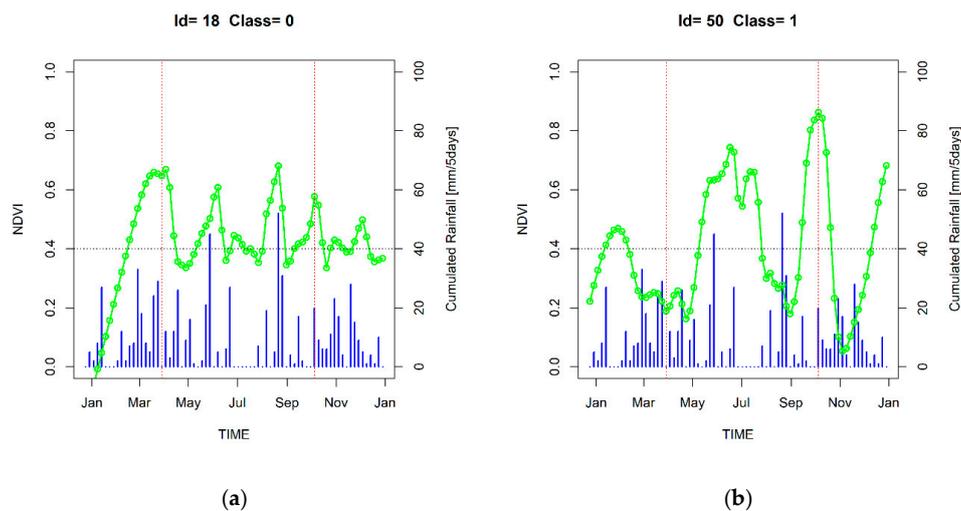


Figure 6. NDVI time series and accumulated rainfall data at pixel scale. Example of (a) class 0 (rainfed and bare soil) in phase and (b) class 1 (herbaceous) not in phase.

2.4. Ground Truth Acquisition

An extensive ground truth was collected in the study area by field inspection using the app mapitGIS [29]. More specifically, we collected a total of 2992 samples spanning the three thematic classes: bare soil and rainfed (class 0, N = 336), herbaceous (class 1, N = 1897), and tree crop (class 2, N = 759). The acquired points are shown in Figure 7 and more detailed in Table 4.

Table 4. Number of reference data acquired per tile and per class.

Class/Tile	T33TVE	T33TVF	T33TWE	T33TWF	Total
0	39	233	33	31	336
1	378	1179	244	96	1897
2	50	626	72	11	759
Total	467	2038	349	138	2992

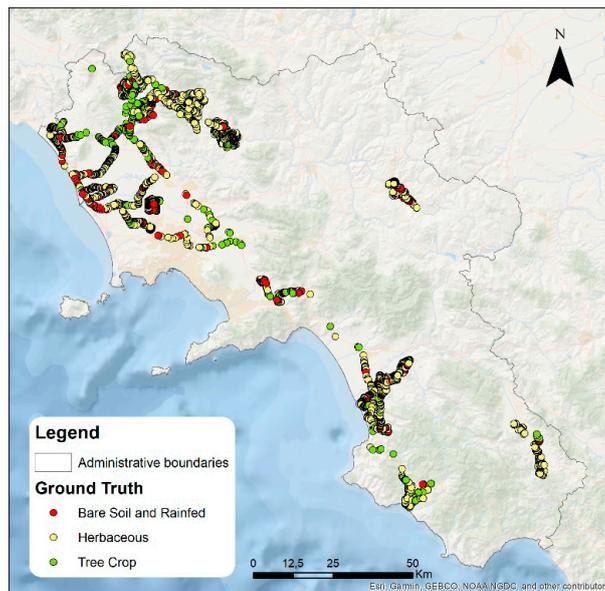


Figure 7. The 2992 ground truth samples collected over the study area across the three thematic classes: bare soil and rainfed (class 0), herbaceous (class 1), and tree crop (class 2).

We note that we acquired one single label/sample per parcel at best in order to limit the spatial correlation among the collected labeled samples. This a relevant aspect to take into consideration in order to avoid an overestimation of the classification accuracies. An example is represented in Figure 8, in which, we show some of the collected samples acquired in an agricultural area in northern Campania and spanning multiple parcels.

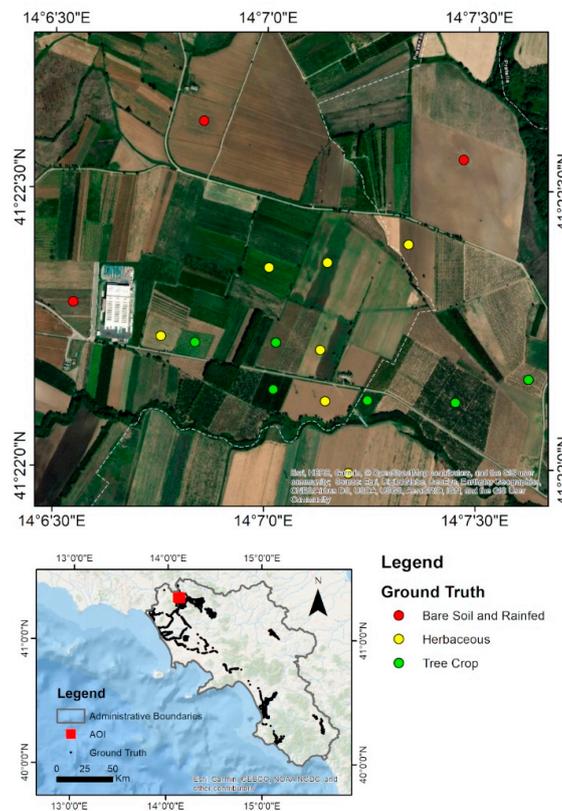


Figure 8. Example of ground truth samples acquired in an agricultural area in northern Campania and spanning multiple parcels.

2.5. Machine Learning-Based Supervised Classification

We employed a machine learning-based supervised classification approach to discriminate among the three classes from the time series NDVI and satellite precipitation data. For this purpose, we considered and compared six classifiers that are commonly used in satellite image classification applications. The classifiers are summarized in Table 5 and comprises random forests (RF), support vector machines (SVM), single decision trees (DT), boosted decision trees (DT), artificial neural networks (ANN), and k-nearest neighbors (k-NN). More specifically, SVMs aim at finding the hyperplane that maximizes the distance among classes and using a limited number of samples (called support vectors) that lie close to the separation margin [30]. Transformation from a linear to a nonlinear problem is done by mapping samples from the original to an high-dimensional feature space through a kernel function (such as the radial basis function (RBF) used in this work [31,32]), an operation that is known as kernel trick. DT classifier is based on a sequential decision process [33]. Starting from the root, a generic feature is evaluated, and one of the two branches is selected. This procedure is repeated for every branch until a final leaf is reached. In case of classification, the leaf values usually correspond to the target classes. RFs are a set of decision trees built on random samples [34]. In this case, a different criterion is used to split nodes; for each tree, a random subset of features is selected in order to find the best splits. As a result, there is the generation of multiple weak trees, whose predictions are combined through a fusion strategy. A common fusion strategy relies on the majority vote rule, although more advanced algorithms based on weighting the posterior probabilities can also be exploited. In addition, RFs provide the relative importance of each feature which can be exploited to understand which features contribute more in the creation of the classification model. Boosted DTs are based on an ensemble of classifiers in which the training set is updated in an iterative way in order to focus more on that samples that tend to be misclassified [35]. ANNs are based on a collection of nodes (called artificial neurons) which mimic the model of neurons in a biological brain [36]. The network is usually composed by an input layer that receives the input features, one or more hidden layers, and an output layer that gives the final predictions. A generic node receives n input channels, which are weighted, summed up, and sent to the output after being processed by an activation function. Finally, k-NN departs from the previous methods since the class of each unknown sample is predicted directly from the training set without really building a classification model [37]. The label is assigned by applying a majority rule on the k labels associated with the k closest training samples.

Table 5. The six algorithms for supervised classification with the related R package used and compared on the collected dataset.

Algorithm	R Package	Reference
Random Forests (RF)	Ranger	[38]
Support Vector Machines (SVM)	Kernlab	[39]
Single Decision Trees (DT)	Rpart	[40]
Boosted Decision Trees (Boosted DT)	C50	[41]
Artificial Neural Networks (ANN)	Nnet	[42]
K-Nearest Neighbors (k-NN)	Caret	[43]

3. Results and Discussion

3.1. Experimental Setting for Irrigated Area Classification Assessment

We first conducted a sensitivity analysis devoted to compare the six classification algorithms introduced in Section 2.5. We used such different supervised classifiers that are widely used in remote sensing image classification applications and evaluated their effectiveness to deal with the classification problem specifically addressed in this paper.

We also compared four preprocessing procedures in order to evaluate the sensitivity of the classification approach to different training data selection approaches (Table 6). More specifically, in the

scenario 1, the classifier was applied to the original training data and the initial number of input features ($d = 150$, which included both NDVI and precipitation data). The scenario 2 implied a classification performed on a balanced training dataset, in which random oversampling was applied to have the same number of training samples per class. Scenario 3 implied a RF-based recursive feature elimination process that reduced the number of features used as input. Finally, scenario 4 incorporated both the balanced training dataset and the recursive feature elimination processes described in scenarios 2 and 3. We, therefore, compared a total of 24 classification scenarios associated with 6 types of classifiers and 4 training data selection strategies.

Table 6. The four considered preprocessing scenarios evaluated on the collected dataset.

Scenario	Preprocessing
1	None
2	Balanced training data
3	Feature selection
4	Feature selection + Balanced training data

The entire set of ground truth samples ($N = 2992$, Table 4) was used in the classification process. Two scenarios were explored in order to evaluate the sensitivity of the framework to different sample sizes. In the first one, 25% of the labeled samples were randomly chosen as training set (stratified by class) and the remaining 75% as validation set. In the second case, 75% of the samples were used for training set and 25% for validation set. In both cases, free parameters of the classifiers were estimated automatically on the training data only using a 10-fold cross-validation approach. More specifically, such parameters were estimated for each type of classifier: (i) number of drawn candidate variables in each split (m_{try}) for RF; (ii) cost (C) and sigma for the RBF kernel (σ) for SVM; (iii) pruning parameter (cp) for DT; iv) number of boosting iterations (trials), indication whether the trees should be collapsed into rules (rules), and activation of the feature selection step before model building (winnow) for Boosted DT; (v) number of units in hidden layer (size) and regularization parameter (decay) for ANN; and (vi) number of neighbors for k-NN. In addition, the `tuneLength` parameter was set to 10, while the other parameters were kept to their default values. After estimating the best parameters through cross-validation, the classification model was built on the entire training set and evaluated on the independent validation set. The analysis was done using the free statistical software R [26]. Within R, we used the `caret` package [43], which provides a standard syntax for running a variety of machine learning algorithms, thus simplifying the process of systematic comparison of different algorithms and approaches.

Accuracies of the classification models were evaluated by considering multiple metrics derived from the confusion matrix associated with the validation set. These metrics were useful to compare the classification performances obtained across classifiers and preprocessing scenarios. More specifically, we considered the overall accuracy (OA), producer's accuracy (PA), user's accuracy (UA), and Cohen's kappa coefficient (denoted as Kappa statistic in the manuscript) [44]. For OA, a confidence level of 95% was considered through the normal approximation method altogether with a continuity correction.

A summary of the overall experimental setting and the different types of analyses conducted as described in the previous paragraph is shown in Figure 9.

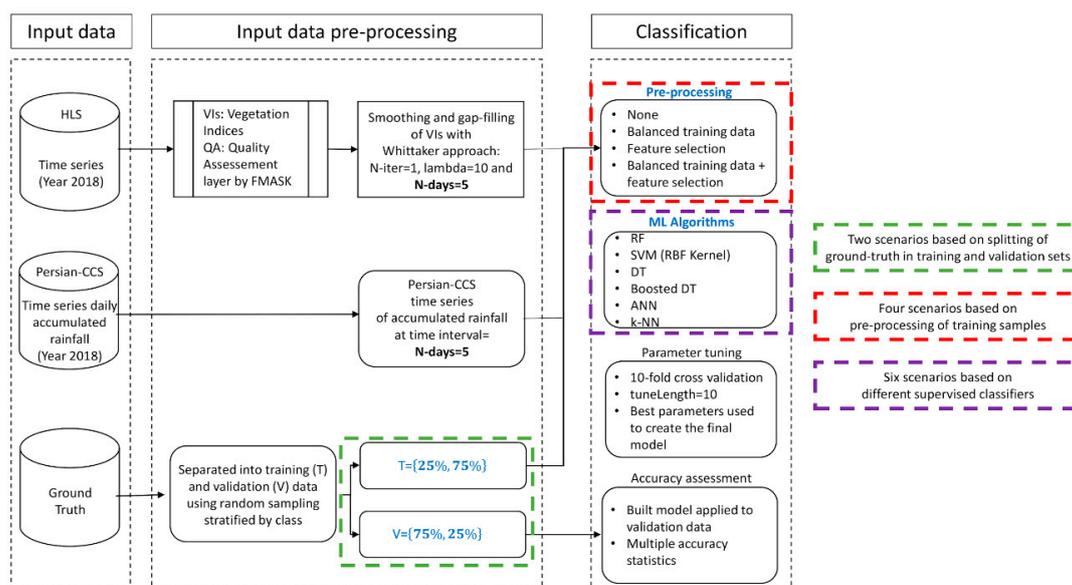


Figure 9. The framework of the overall experimental setting.

3.2. Experimental Results

We compared the six types of classifiers and applied them in conjunction with the four preprocessing scenarios. We first focused on the scenario in which 25% of the labeled samples were used for training. By considering multiple evaluation metrics, we found that RF, SVM, and Boosted DT models exhibited in general better results than the other three classifier types (Table 7). In general, RF and Boosted DT were robust to the different preprocessing steps and comparable results were obtained in the four scenarios (OA ranging from 86.3% to 87.8% for RF and OA ranging from 86.0% to 86.6% for Boosted DT). On the other hand, SVM exhibited a higher variability (OA ranging from 82.7% to 87.8%) and the necessity to employ a proper preprocessing strategy to maximize the performances in terms of classification accuracy. Moreover, RF exhibited the highest accuracies for scenario 1, scenario 2, and scenario 4, whereas SVM gave the best results for scenario 3, which maximized the accuracy at all, both in terms of OA and Kappa statistic (Figure 10). In this case, we obtained an OA = 87.8% and a Kappa statistic equal to 0.770, which highlighted promising capabilities of the proposed approach to detect irrigated areas.

Table 7. OA (%) and Kappa statistic results for each tested classifier and preprocessing scenario. Bold denotes the best accuracies for each preprocessing scenario. The 25% of the labeled samples were used for training and the remaining ones for validation.

Preprocessing	Accuracy Metric	RF	SVM	DT	Boosted DT	ANN	k-NN
None	OA	86.3	84.1	78.5	86.0	81.2	75.8
	Kappa	0.725	0.698	0.596	0.719	0.639	0.535
Balanced training data	OA	87.8	82.7	77.2	86.2	78.7	64.2
	Kappa	0.766	0.699	0.595	0.730	0.604	0.440
Feature selection	OA	86.7	87.8	78.1	86.4	81.7	80.2
	Kappa	0.740	0.770	0.572	0.730	0.665	0.655
Feature selection + Balanced training data	OA	87.5	84.3	77.5	86.6	80.0	71.5
	Kappa	0.763	0.721	0.596	0.744	0.638	0.543

We dissected the best model identified, which corresponded to SVM classifier applied in conjunction with the preprocessing step employed in scenario 3. This was based on a recursive feature selection process that, starting from the original 150 features, found the optimal set and number of variables by maximizing the Kappa statistic on the training set using a cross-validation approach.

This was obtained for a number of features equal to 41 (Kappa statistic = 0.802 and OA = 90.0% in cross-validation), which were therefore selected to build the final classification model (Figure 11).

RF scenario 1						
		Reference			Total	UA
		0	1	2		
Prediction	0	138	41	4	183	0.75
	1	92	1,339	105	1,536	0.87
	2	22	43	460	525	0.88
Total		252	1,423	569	2,244	
PA		0.55	0.94	0.81		
OA		86.3				
Kappa		0.725				

RF scenario 2						
		Reference			Total	UA
		0	1	2		
Prediction	0	166	60	10	236	0.70
	1	57	1,309	63	1,429	0.92
	2	29	54	496	579	0.86
Total		252	1,423	569	2,244	
PA		0.66	0.92	0.87		
OA		87.8				
Kappa		0.766				

SVM scenario 3						
		Reference			Total	UA
		0	1	2		
Prediction	0	176	66	13	255	0.69
	1	38	1,277	40	1,355	0.94
	2	38	80	516	634	0.81
Total		252	1,423	569	2,244	
PA		0.70	0.90	0.91		
OA		87.8				
Kappa		0.770				

RF scenario 4						
		Reference			Total	UA
		0	1	2		
Prediction	0	165	62	13	240	0.69
	1	49	1,288	45	1,382	0.93
	2	38	73	511	622	0.82
Total		252	1,423	569	2,244	
PA		0.65	0.91	0.90		
OA		87.5				
Kappa		0.763				

Figure 10. Confusion matrix, OA (%), PA, UA, and Kappa statistic for the best model associated with each scenario identified in Table 7. RF scenario 1: CI = (0.848, 0.877), McNemar’s Test p -value = 1.579e-12; RF scenario 2: CI = (0.864, 0.892), McNemar’s Test p -value = 0.018; SVM scenario 3: CI = (0.863, 0.891), McNemar’s Test p -value = 3.028e-07; RF scenario 4: CI = (0.861, 0.889), McNemar’s Test p -value = 1.388e-04.

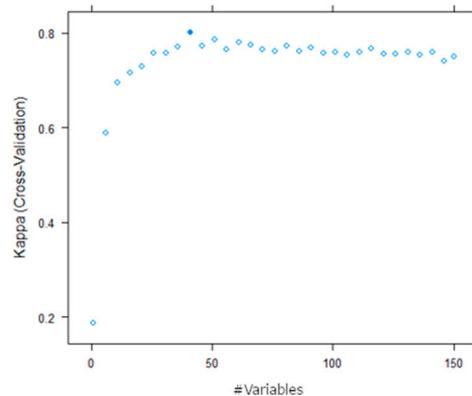


Figure 11. Classification model generated using SVMs as classifier and by applying a recursive feature selection process (scenario 3), which maximized the accuracies on the validation set as shown in Table 7. The plot reports the Kappa statistic obtained in cross-validation in function of the number of variables/features used to build the model. A number of features equal to 41 maximized the Kappa statistic, which is highlighted with the filled circle.

The same analysis was repeated by changing the size of the training set and considering 75% of the samples as training set ($N = 2244$) and the remaining ones ($N = 748$) for validation. In this setting, we confirmed that RF, SVM, and Boosted DT gave better results than the other types of classifiers (Table 8 and Figure 12). Again, RF and Boosted DT resulted quite robust to the preprocessing step in comparison with SVM that required, also in this case, a feature selection process in order to maximize the accuracy. Overall, RF seemed the more robust method, and the highest accuracies were obtained for scenarios 1 and 4. Boosted DT was instead associated with the best models for scenarios 2 and 3, which suggested the effectiveness of this approach when an enough number of training samples was available. In particular, the model related to scenario 3 gave the best accuracy across all classifiers and

scenarios tested with this training sample size. SVM gave slightly worse results, which highlighted the superiority of ensemble models such as RF and Boosted DT when the training set was sufficiently large. On the other hand, we recall that SVM gave the best accuracy when the training set was smaller (Table 7).

Table 8. OA (%) and Kappa statistic results for each tested classifier and preprocessing scenario. Bold denotes the best accuracies for each preprocessing scenario. The 75% of the labeled samples were used for training and the remaining ones for validation.

Preprocessing	Accuracy Metric	RF	SVM	DT	Boosted DT	ANN	k-NN
None	OA	90.5	86.9	81.3	89.8	83.2	80.5
	Kappa	0.814	0.744	0.630	0.802	0.675	0.637
Balanced training data	OA	90.4	87.2	76.2	90.6	76.7	70.5
	Kappa	0.814	0.759	0.593	0.821	0.580	0.525
Feature selection	OA	90.5	88.0	81.3	90.8	83.7	81.3
	Kappa	0.880	0.766	0.630	0.821	0.691	0.653
Feature selection + Balanced training data	OA	90.6	88.2	76.2	90.1	81.4	71.4
	Kappa	0.820	0.777	0.595	0.809	0.653	0.540

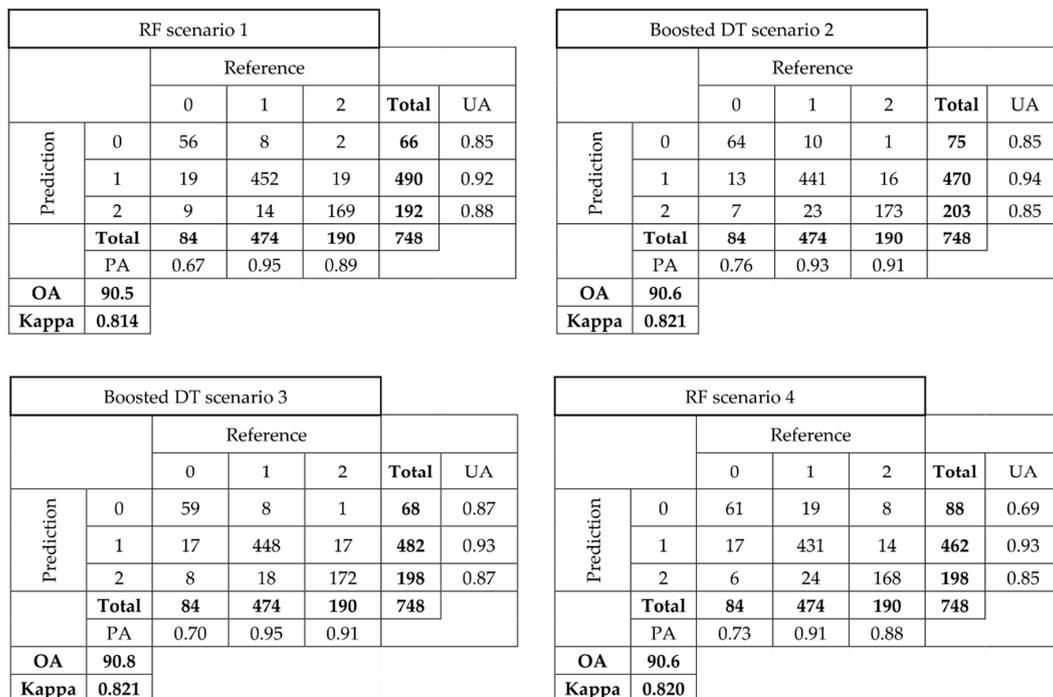


Figure 12. Confusion matrix, OA (%), PA, UA, Kappa statistic, McNemar’s test P-value, and CI for the best model associated with each scenario identified in Table 8. RF scenario 1: CI = (0.882, 0.925), McNemar’s Test *p*-value = 0.021; Boosted DT scenario 2: CI = (0.883, 0.926), McNemar’s Test *p*-value = 0.105; Boosted DT scenario 3: CI = (0.885, 0.923), McNemar’s Test *p*-value = 0.03336; RF scenario 4: CI = (0.861, 0.889), McNemar’s Test *p*-value = 0.434.

We further analyzed the best model obtained with this training size, which resulted to be Boosted DT with the feature selection preprocessing step employed in scenario 3. In this case, a quite large number of variables ($d = 126$) were required to maximize the cross-validation Kappa statistic (Figure 13). We report the feature relevance score estimated by the feature selection process in Figure 13. Interestingly, larger weights were overall attributed to NDVI features with respect to those derived from satellite precipitation data. Among the NDVI features, greater importance was attributed to data spanning the June–October time frame, which was consistent with the relevance of these acquisitions

in discriminating among the different thematic classes. It was also evident the presence of two local maximum located at the end of July and in September, which was again in line with the classification problem involved here. Finally, Figure 14 shows the classification map obtained on the full study area associated with the best classification model.

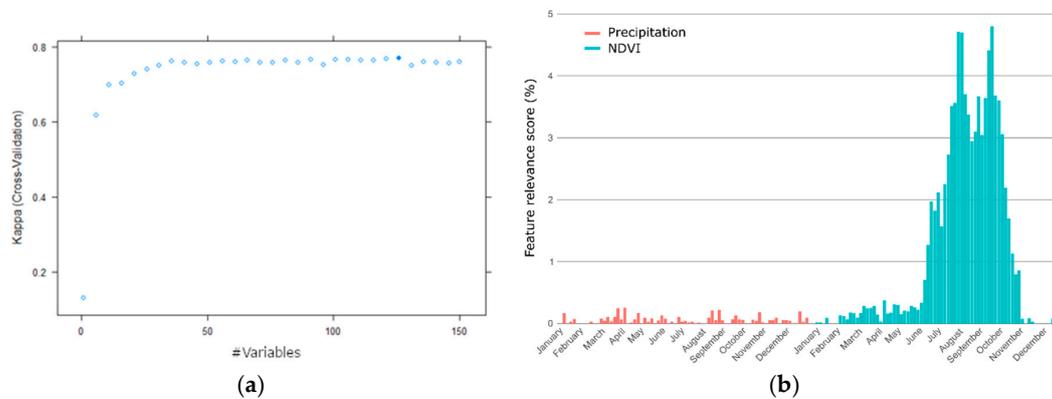


Figure 13. Classification model generated using Boosted DT as classifier and by applying a recursive feature selection process (scenario 3), which maximized the accuracies on the validation set as shown in Table 8. (a) The plot reports the Kappa statistic obtained in cross-validation in function of the number of variables/features used to build the model. A number of features equal to 126 maximized the Kappa statistic, which is highlighted with the filled circle. (b) The feature relevance score. Overall, larger weights were given to NDVI features (mainly in the June–October time frame) rather than satellite precipitation data.

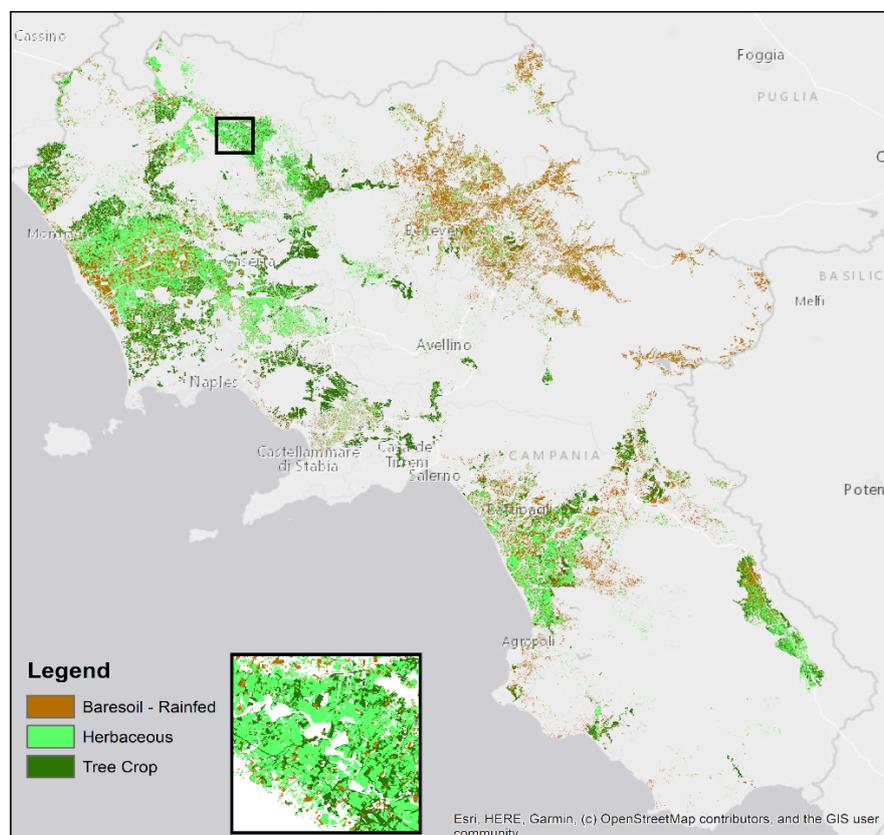


Figure 14. Classification map obtained over the study area in the Campania region (southern Italy) associated with the best model (75% of the labeled samples used for training).

3.3. Cross-Validation with Spatially Separated Folds

The experimental analysis conducted in Section 3.2 was done by considering a random cross-validation approach in which labeled samples were randomly split into training and validation sets. Although this represents a quite common scenario in evaluating and comparing classification accuracies in the remote sensing domain, this may bring to an overestimation of the accuracies as discussed recently [45,46]. More specifically, remote sensing images are characterized by spatial structures that are ignored when performing a traditional cross-validation with the result of underestimating predictive errors. To account for this, we conducted an additional analysis in which we validated the model using a block cross-validation approach in which samples were split strategically rather than randomly. We considered the recently developed blockCV package [47], which was specifically proposed to generate spatially or environmentally separated folds. We considered the spatial blocking strategy (function “spatialBlock”), which aims at building square spatial blocks of a specified size, by setting the parameters as follows: (i) size of the blocks (parameter “theRange”) was set to 10,000, (ii) allocation of blocks to folds (parameter “selection”) was done in a random way in order to find the block-to-fold allocations that achieved most even spread of classes across folds, and (iii) number of folds (parameter “k”) equal to 4 in order to split the labeled samples with the same percentages considered in Section 3.2. We show in Figure 15, the partition of the region in multiple spatially disjoint blocks and their subsequent random assignment to different folds.

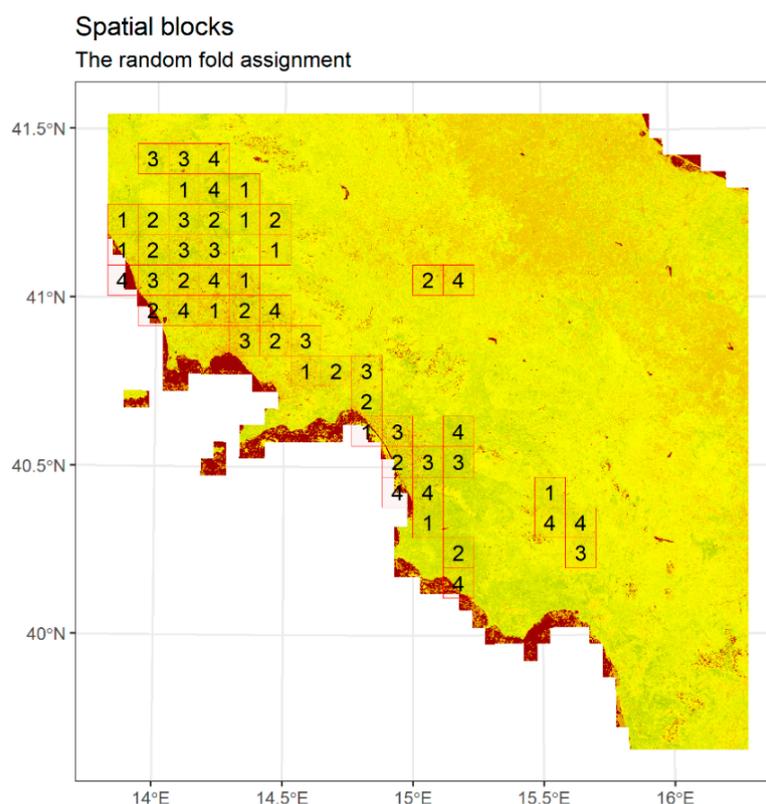


Figure 15. Partition of the region in spatially disjoint blocks (with size of 10,000 m) and their random assignment to four folds. This was accomplished using the R package blockCV [47].

We considered the best scenarios found in the analysis exploited in Section 3.2. In the case that 75% of the labeled samples were used for training, the OA in random cross-validation was maximized using a Boosted DT and was equal to 90.8% (Table 8). Using the same configuration, spatially distinct cross-validation gave an OA equal to 87.2%. Although, as expected, the results shown a decrease of the accuracies with respect to the random cross-validation case, accuracies remained quite satisfactory, which confirmed the validity of our proposed framework. A similar pattern was verified when only

25% of the labeled information was used for training. In this case, OA slightly decreased from 87.8% (Table 8, using SVM in conjunction with feature selection) to 84.2% when moving from the random to the spatially distinct cross-validation.

4. Discussion and Conclusion

In this paper, we have presented the framework that we have developed, implemented, and validated within the DIANA project to map the extent of irrigated areas in the entire Campania region (Southern Italy) during the 2018 irrigation season. The procedure described here is currently employed in the Campania region to assess the annual water volumes that are actually used in the irrigated areas, which it is required to be compliant with the EU Water Framework Directive.

We have considered multiple data types that have spanned the area of interest along the entire year. We have collected 202 images coming from the HLS data product in conjunction with global satellite precipitation PERSIAN-CCS data. We have also acquired an extensive ground truth in the field associated with three thematic classes, namely, bare soil and rainfed, herbaceous, and tree crop. We have compared six machine learning classification algorithms and demonstrated that random forests, support vector machines, and boosted decision trees have given the best results in terms of classification accuracy and robustness to different scenarios. In the best case, the three classes have been distinguished with an overall accuracy close to 90%, which has highlighted the relevance of this recent data product for the detection and characterization of irrigated areas from satellite observations. This has demonstrated promising capabilities in using this set of optical time series, meteorological, and in-situ collected ground truth data to map irrigated areas that may be used in operational applications for water management purposes.

The framework presented here has been specifically developed for the detection of irrigated areas with the goal of being able to distinguish between trees and herbaceous crops without requiring prior knowledge of inter- and intraclass characteristics. While several papers based on machine learning approaches have been proposed in the literature to deal with land cover classification problems, a few works have been specifically devoted to the detection and characterization of irrigated areas. Irrigated and not irrigated croplands were estimated with high accuracy (Kappa statistic > 0.9) by considering the historical evolution of irrigated croplands for the post monsoon (rabi) and summer cropping seasons in the Berambadi watershed of Kabini River basin, southern India. In this case, 30 m spatial resolution Landsat satellite images spanning the 1990–2016 periods were classified with SVM [48]. Landsat imagery (Thematic Mapper, Enhanced Thematic Mapper Plus, and Operational Land Imager) was again processed in conjunction with SVM to quantify the changes in irrigated land areas surrounding the Mogtedo water reservoir, Burkina Faso, between 1987 and 2015. Overall accuracy and Kappa statistic ranged from 94.2% to 95.6% and from 0.92 to 0.94, respectively [49]. Annual maps at 30 m spatial resolution of irrigated corn and soybean areas in southwestern Michigan, United States, were generated in the 2001–2016 timeframe with an OA = 82% by considering Landsat surface reflectance products and RF classification [50]. The relevance in using Sentinel-1 images jointly with Landsat 8 optical imagery was studied and a digital elevation model of the Shuttle Radar Topography Mission (SRTM) was investigated in [51]. The combined use of the different data types in conjunction with RF was able to improve the early classification of irrigated crops from 0.84 to 0.89 (Kappa statistic) in a dataset involving the detection of irrigated maize crops in a temperate zone in South West France. A novel classification-based irrigation mapping procedure that utilized MODIS time series data coupled with ancillary data on climate and agricultural extent was proposed in [52]. It gave moderate Kappa statistic equal to 0.36 and 0.65 for Eastern US and Western US, respectively, using a tree-based method for supervised classification. Finally, a methodology to identify irrigated crops in a semiarid zone in La Mancha, Spain, was proposed using Landsat TM imagery and a multitemporal supervised classification approach [53]. Overall accuracy was equal to 93.1% and 90.2% during 1996 and 1997 growing seasons, respectively.

While we have focused in this paper on mapping areas at regional scale, a similar procedure can be applied to detect and characterize larger regions. Further methodological advancements in this direction can include transfer learning approaches in order to take into account the spatial variability of the class distributions [54] and deep learning strategies to deal with large training set scenarios [55]. Finally, future research lines can be also devoted to the inclusion of additional indices such as the Normalized Difference Water Index (NDWI) or the Modified NDWI. It would be also interesting to introduce into the framework additional information based on the VIS–NIR and SWIR bands, which have been recently exploited by novel algorithms such as the OPTical TRAppezoid Model [56] and that have been used to extract additional indices such as the Soil Moisture Index.

Author Contributions: Conceptualization, S.F.B., E.P., O.R.B., and G.D.; methodology, S.F.B., E.P., and O.R.B.; formal analysis, S.F.B., E.P., and O.R.B.; writing—original draft preparation, S.F.B. and E.P.; writing—review and editing, S.F.B., E.P., O.R.B., C.D.M., and G.D.; and funding acquisition, C.D.M. and G.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was conducted within the DIANA project—“Detection and Integrated Assessment of Non-authorized Water Abstractions using EO”, which was funded from the European Union’s Horizon 2020 research and innovation program under Grant agreement No. 730109.

Acknowledgments: We acknowledge the NASA HLS team under Jeffrey Masek for the access to HLS data.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Vörösmarty, C.J. Global water assessment and potential contributions from Earth Systems Science. *Aquat. Sci.* **2002**, *64*, 328–351. [[CrossRef](#)]
2. Boucher, O.; Myhre, G.; Myhre, A. Direct human influence of irrigation on atmospheric water vapour and climate. *Clim. Dyn.* **2004**, *22*, 597–603. [[CrossRef](#)]
3. Alcamo, J.; Döll, P.; Henrichs, T.; Kaspar, F.; Lehner, B.; Rösch, T.; Siebert, S. Global estimates of water withdrawals and availability under current and future “business-as-usual” conditions. *Hydrol. Sci. J.* **2003**, *48*, 339–348. [[CrossRef](#)]
4. Peña-Arancibia, J.L.; McVicar, T.R.; Paydar, Z.; Li, L.; Guerschman, J.P.; Donohue, R.J.; Dutta, D.; Podger, G.M.; van Dijk, A.I.J.M.; Chiew, F.H.S. Dynamic identification of summer cropping irrigated areas in a large basin experiencing extreme climatic variability. *Remote Sens. Environ.* **2014**, *154*, 139–152. [[CrossRef](#)]
5. Pervez, S.; Budde, M.; Rowland, J. Mapping irrigated areas in Afghanistan over the past decade using MODIS NDVI. *Remote Sens. Environ.* **2014**, *149*, 155–165. [[CrossRef](#)]
6. Gumma, M.K.; Thenkabail, P.S.; Nelson, A. Mapping irrigated areas using MODIS 250 meter time-series data: A study on Krishna river basin (India). *Water* **2011**, *3*, 113–131. [[CrossRef](#)]
7. Brown, J.F.; Pervez, M.S. Merging remote sensing data and national agricultural statistics to model change in irrigated agriculture. *Agric. Syst.* **2014**, *127*, 28–40. [[CrossRef](#)]
8. Gumma, M.K.; Thenkabail, P.S.; Hideto, F.; Nelson, A.; Dheeravath, V.; Busia, D.; Rala, A. Mapping Irrigated Areas of Ghana Using Fusion of 30 m and 250 m Resolution Remote-Sensing Data. *Remote Sens.* **2011**, *3*, 816–835. [[CrossRef](#)]
9. Pun, M.; Mutiibwa, D.; Li, R. Land Use Classification: A surface energy balance and vegetation index application to map and monitor irrigated lands. *Remote Sens.* **2017**, *9*, 1256. [[CrossRef](#)]
10. Claverie, M.; Ju, J.; Masek, J.G.; Dungan, J.L.; Vermote, E.F.; Roger, J.-C.; Skakun, S.V.; Justice, C. The Harmonized Landsat and Sentinel-2 surface reflectance data set. *Remote Sens. Environ.* **2018**, *219*, 145–161. [[CrossRef](#)]
11. Claverie, M.; Masek, J.G.; Ju, J.; Dungan, J.L. *Harmonized Landsat-8 Sentinel-2 (HLS) Product User’s Guide*; National Aeronautics and Space Administration (NASA): Washington, DC, USA, 2017.
12. Zhou, Q.; Rover, J.; Brown, J.; Worstell, B.; Howard, D.; Wu, Z.; Gallant, A.L.; Rundquist, B.; Burke, M. Monitoring landscape dynamics in Central U.S. grasslands with Harmonized Landsat-8 and Sentinel-2 time series data. *Remote Sens.* **2019**, *11*, 328. [[CrossRef](#)]
13. Pastick, N.J.; Wylie, B.K.; Wu, Z. Spatiotemporal analysis of Landsat-8 and Sentinel-2 data to support monitoring of dryland ecosystems. *Remote Sens.* **2018**, *10*, 791. [[CrossRef](#)]

14. Skakun, S.; Franch, B.; Vermote, E.; Roger, J.; Justice, C.; Masek, J.; Murphy, E. Winter wheat yield assessment using Landsat 8 and Sentinel-2 data. In Proceedings of the 2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 5964–5967.
15. Griffiths, P.; Nendel, C.; Hostert, P. Intra-annual reflectance composites from Sentinel-2 and Landsat for national-scale crop and land cover mapping. *Remote Sens. Environ.* **2019**, *220*, 135–151. [[CrossRef](#)]
16. Griffiths, P.; Nendel, C.; Pickert, J.; Hostert, P. Towards national-scale characterization of grassland use intensity from integrated Sentinel-2 and Landsat time series. *Remote Sens. Environ.* **2020**, *238*, 111124. [[CrossRef](#)]
17. Hao, P.-Y.; Tang, H.-J.; Chen, Z.-X.; Yu, L.; Wu, M.-Q. High resolution crop intensity mapping using harmonized Landsat-8 and Sentinel-2 data. *J. Integr. Agric.* **2019**, *18*, 2883–2897. [[CrossRef](#)]
18. Bolton, D.K.; Gray, J.M.; Melaas, E.K.; Moon, M.; Eklundh, L.; Friedl, M.A. Continental-scale land surface phenology from harmonized Landsat 8 and Sentinel-2 imagery. *Remote Sens. Environ.* **2020**, *240*, 111685. [[CrossRef](#)]
19. Jönsson, P.; Cai, Z.; Melaas, E.; Friedl, M.; Eklundh, L. A method for robust estimation of vegetation seasonality from Landsat and Sentinel-2 time series data. *Remote Sens.* **2018**, *10*, 635. [[CrossRef](#)]
20. DIANA. Available online: <http://diana-h2020.eu/en/> (accessed on 14 April 2020).
21. Gao, F.; Masek, J.G.; Wolfe, R.E. Automated registration and orthorectification package for Landsat and Landsat-like data processing. *JARS* **2009**, *3*, 033515.
22. Tucker, C.J. Red and Photographic Infrared Linear Combinations for Monitoring Vegetation. *NASA-TM-79620*, 1 May 1978.
23. Ambika, A.K.; Wardlow, B.; Mishra, V. Remotely sensed high resolution irrigated area mapping in India for 2000 to 2015. *Sci. Data* **2016**, *3*, 160118. [[CrossRef](#)]
24. Eilers, P.H.C. A perfect smoother. *Anal. Chem.* **2003**, *75*, 3631–3636. [[CrossRef](#)]
25. Mattiuzzi, M.; Verbesselt, J.; Stevens, F.; Mosher, S.; Hengl, T.; Klisch, A.; Evans, B.; Lobo, A. MODIS: MODIS Acquisition and Processing Package. 2014. Available online: <http://R-Forge.R-project.org/projects/modis> (accessed on 14 April 2020).
26. R Core Team. *R: A Language and Environment for Statistical Computing*; R Core Team: Vienna, Austria, 2013.
27. GDAL. Available online: <https://gdal.org/> (accessed on 14 April 2020).
28. CHRIS Data Portal. Available online: <https://chrdata.eng.uci.edu/> (accessed on 14 April 2020).
29. mapitGIS. Available online: <https://mapitgis.com/> (accessed on 14 April 2020).
30. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [[CrossRef](#)]
31. Kavzoglu, T.; Colkesen, I. A kernel function analysis for support vector machines for land cover classification. *Int. J. Appl. Earth Obs. Geoinf.* **2009**, *11*, 352–359. [[CrossRef](#)]
32. Huang, C.; Davis, L.S.; Townshend, J.R.G. An assessment of support vector machines for land cover classification. *Int. J. Remote Sens.* **2002**, *23*, 725–749. [[CrossRef](#)]
33. Pal, M.; Mather, P.M. An assessment of the effectiveness of decision tree methods for land cover classification. *Remote Sens. Environ.* **2003**, *86*, 554–565. [[CrossRef](#)]
34. Breiman, L. Random forests. *Mach. Learn.* **2001**, *54*, 5–32. [[CrossRef](#)]
35. Freund, Y.; Schapire, R.E. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.* **1997**, *55*, 119–139. [[CrossRef](#)]
36. Atkinson, P.M.; Tatnall, A.R.L. Introduction neural networks in remote sensing. *Int. J. Remote Sens.* **1997**, *18*, 699–709. [[CrossRef](#)]
37. Altman, N.S. An introduction to kernel and nearest-neighbor non parametric regression. *Am. Stat.* **1992**, *46*, 175–185.
38. Wright, M.N.; Ziegler, A. ranger: A fast implementation of random forests for high dimensional data in C++ and R. *arXiv* **2015**, arXiv:1508.04409. [[CrossRef](#)]
39. Karatzoglou, A.; Smola, A.; Hornik, K. Kernlab: Kernel-based Machine Learning Lab. R Package Version 0.9-29. 2016. Available online: <https://cran.r-project.org/web/packages/kernlab/index.html> (accessed on 14 April 2020).
40. Therneau, T.; Atkinson, B.; Ripley, B. rpart: Recursive Partitioning and Regression Trees. 2015. Available online: <https://cran.r-project.org/web/packages/rpart/index.html> (accessed on 14 April 2020).
41. Kuhn, M.; Weston, S.; Coulter, N.; Quinlan, R. C50: C5.0 decision trees and rule-based models. R Package Version 0.1.0-21. 2014. Available online: <http://CRAN.R-project.org/package=C> (accessed on 14 April 2020).

42. Ripley, B.; Venables, W. nnet: Feed-forward Neural Networks and Multinomial Log-linear Models. R Package Version 7.3-13. 2016. Available online: <https://cran.r-project.org/web/packages/nnet/index.html> (accessed on 14 April 2020).
43. Kuhn, M.; Wing, J.; Weston, S.; Williams, A.; Keefer, C.; Engelhardt, A.; Cooper, T.; Mayer, Z.; Kenkel, B.; R Core Team; et al. caret: Classification and Regression Training. R Package Version 6.0-86. 2016. Available online: <https://cran.r-project.org/web/packages/caret/index.html> (accessed on 14 April 2020).
44. Smeeton, C.N. Early history of the kappa statistic. *Biometrics* **1985**, *41*, 795.
45. Roberts, D.R.; Bahn, V.; Ciuti, S.; Boyce, M.S.; Elith, J.; Guillerá-Arroita, G.; Hauenstein, S.; Lahoz-Monfort, J.J.; Schröder, B.; Thuiller, W.; et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* **2017**, *40*, 913–929. [[CrossRef](#)]
46. Meyer, H.; Reudenbach, C.; Hengl, T.; Katurji, M.; Nauss, T. Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation. *Environ. Model. Softw.* **2018**, *101*, 1–9. [[CrossRef](#)]
47. Valavi, R.; Elith, J.; Lahoz-Monfort, J.J.; Guillerá-Arroita, G. blockCV: An R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Methods Ecol. Evol.* **2019**, *10*, 225–232. [[CrossRef](#)]
48. Sharma, A.K.; Hubert-Moy, L.; Buvaneshwari, S.; Sekhar, M.; Ruiz, L.; Bandyopadhyay, S.; Corgne, S. Irrigation History Estimation Using Multitemporal Landsat Satellite Images: Application to an Intensive Groundwater Irrigated Agricultural Watershed in India. *Remote Sens.* **2018**, *10*, 893. [[CrossRef](#)]
49. Traoré, F.; Bonkougou, J.; Compaoré, J.; Kouadio, L.; Wellens, J.; Hallot, E.; Tychon, B. Using multi-temporal Landsat images and support vector machine to assess the changes in agricultural irrigated areas in the Mogtedo region, Burkina Faso. *Remote Sens.* **2019**, *11*, 1442. [[CrossRef](#)]
50. Xu, T.; Deines, J.M.; Kendall, A.D.; Basso, B.; Hyndman, D.W. Addressing challenges for mapping irrigated fields in subhumid temperate regions by integrating remote sensing and hydroclimatic Data. *Remote Sens.* **2019**, *11*, 370. [[CrossRef](#)]
51. Demarez, V.; Helen, F.; Marais-Sicre, C.; Baup, F. In-season mapping of irrigated crops using Landsat 8 and Sentinel-1 time series. *Remote Sens.* **2019**, *11*, 118. [[CrossRef](#)]
52. Ozdogan, M.; Gutman, G. A new methodology to map irrigated areas using multi-temporal MODIS and ancillary data: An application example in the continental US. *Remote Sens. Environ.* **2008**, *112*, 3520–3537. [[CrossRef](#)]
53. Beltran, C.M.; Belmonte, A.C. Irrigated crop area estimation using Landsat TM imagery in La Mancha, Spain. *Photogramm. Eng. Remote Sens.* **2001**, *67*, 1177–1184.
54. Pan, S.J.; Yang, Q. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [[CrossRef](#)]
55. Zhu, X.X.; Tuia, D.; Mou, L.; Xia, G.; Zhang, L.; Xu, F.; Fraundorfer, F. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 8–36. [[CrossRef](#)]
56. Sadeghi, M.; Babaeian, E.; Tuller, M.; Jones, S.B. The optical trapezoid model: A novel approach to remote sensing of soil moisture applied to Sentinel-2 and Landsat-8 observations. *Remote Sens. Environ.* **2017**, *198*, 52–68. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).