

Article

Spatially Explicit Active Learning for Crop-Type Mapping from Satellite Image Time Series

Beatrice Kaijage, Mariana Belgiu *  and Wietske Bijker 

Faculty of Geo-Information Science and Earth Observation, University of Twente,
7522 NH Enschede, The Netherlands; beatriceanth2006@gmail.com (B.K.); w.bijker@utwente.nl (W.B.)

* Correspondence: m.belgiu@utwente.nl

Abstract: The availability of a sufficient number of annotated samples is one of the main challenges of the supervised methods used to classify crop types from remote sensing images. Creating these samples is time-consuming and costly. Active Learning (AL) offers a solution by streamlining sample annotation, resulting in more efficient training with less effort. Unfortunately, most of the developed AL methods overlook spatial information inherent in remote sensing images. We propose a novel spatially explicit AL that uses the semi-variogram to identify and discard redundant, spatially adjacent samples. It was evaluated using Random Forest (RF) and Sentinel-2 Satellite Image Time Series in two study areas from the Netherlands and Belgium. In the Netherlands, the spatially explicit AL selected 97 samples achieving an overall accuracy of 80%, compared to traditional AL selecting 169 samples with 82% overall accuracy. In Belgium, spatially explicit AL selected 223 samples and obtained 60% overall accuracy, while traditional AL selected 327 samples and obtained an overall accuracy of 63%. We concluded that the developed AL method helped RF achieve a good performance mostly for the classes consisting of individual crops with a relatively distinctive growth pattern such as sugar beets or cereals. Aggregated classes such as ‘fruits and nuts’ posed, however, a challenge.

Keywords: spatial autocorrelation; supervised classification; remote sensing; agriculture; scarce label environments



Citation: Kaijage, B.; Belgiu, M.; Bijker, W. Spatially Explicit Active Learning for Crop-Type Mapping from Satellite Image Time Series. *Sensors* **2024**, *24*, 2108. <https://doi.org/10.3390/s24072108>

Academic Editors: Xianpeng Wang, Fangqing Wen, Jin He, Liangtian Wan and Zhiyuan Zha

Received: 26 February 2024
Revised: 21 March 2024
Accepted: 22 March 2024
Published: 26 March 2024



Copyright: © 2024 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Supervised machine learning has experienced tremendous progress in recent years being increasingly used for classifying crop types from Satellite Image Time Series (SITS) [1,2]. Although deep learning classifiers can successfully identify crops from SITS without requiring the involvement of the image analysts in the selection of the input features [3–8], traditional supervised machine learning remains a viable alternative [9–11]. Machine learning methods learn the characteristics of the target crops from training samples collected through intense field campaigns or expert-based interpretation of very high-resolution satellite images [12]. Unfortunately, annotating a large number of samples is a time-consuming and expensive task [13]. Consequently, different solutions have been proposed to address the challenges associated with the availability of samples: (1) generating labeled crop samples for the target classification years by using samples from previous years [10]; (2) developing (semi-)automatic solutions to label crop samples [13]; (3) using classifiers such as Dynamic Time Warping that require a small number of training samples [14,15]; (4) leveraging the information learned by supervised methods in areas where many labeled samples are available through transfer learning methods [16,17]; and (5) reducing the number of crop samples to be annotated without decreasing the performance of the supervised methods [18]. This means that the training samples remain representative of the statistical distribution of data [19].

Active Learning (AL) is one of the solutions to reduce the number of samples to be annotated by several orders of magnitude. It helps to select the most informative samples to

be labeled from a large unlabeled sample set [20]. Two criteria are commonly used to rank the candidate samples relative to their potential contribution to the classification problem: uncertainty and diversity. The uncertainty criterion aims to identify the samples that pose challenges to the supervised methods. The uncertainty of a sample can be quantified using margin sampling [21], disagreement between a committee of classifiers [22], posterior probabilities of class membership [23], or entropy [24]. Since the samples selected based on uncertainty alone may not be well representative of the classes present in the dataset, the diversity criterion has been introduced in AL studies. Diversity measures ensure class representativeness among selected samples. These measures include density-weighted metrics or clustering-based approaches [18]. Euclidean distance or cosine similarity metrics are the most common similarity measures used to select the most dissimilar samples [25].

AL has been successfully implemented in several remote sensing applications including land cover classification with Support Vector Machine and multispectral images [26], improvement of land cover classification from hyperspectral images using both traditional AL [23] or AL that accounts for spatial information [27], urban land cover-land use classification through object-based image analysis [28]. In terms of application domains, AL has been used for biophysical parameter estimation [29], tree species mapping [30], crop area mapping [31], crop type mapping using satellite image time series [32] and hyperspectral images [33], large-extent cultivated area mapping [31], and poplar plantation mapping at the national level [18]. Recently, AL has been implemented to reduce the annotation efforts required by various deep learning algorithms [34,35].

Previous studies focused mainly on AL heuristics that use spectral data to optimize the collection of samples to be annotated while ignoring the spatial information inherent in remote sensing images. Nonetheless, AL-based selected samples may be spatially contiguous and this might result in a redundant sample set. A limited number of studies accounted for the spatial distribution of samples when implementing AL algorithms. For example, Demir, et al. [36] included topography and road networks to reduce travel time when limited resources are available to annotate training samples. Zhang, Pasolli, and Crawford [33] showed the advantages of both spectral and spatial features extracted from segmentation maps in the proposed multi-view AL method to map crops from hyperspectral images. Pasolli, et al. [37] used the Euclidean distance, Parzen window method, and spatial entropy as spatial criteria to ensure that the selected samples are spatially representative of the entire study area. The authors combined spectral and spatial criteria using nondominated sorting. Calculating spatial autocorrelation between samples within an area is an alternative approach for identifying redundant samples in the space domain. Different measures can be used to quantify the spatial autocorrelation including Moran's I [38], Geary's C ratio [39], or the semi-variogram. Moran's I and Geary's C are global spatial autocorrelation measures, whereas the semi-variogram shows how spatial variation changes as a function of the distance between point location pairs. Stumpf, et al. [40] developed a region-based AL method for landslide mapping that reduces the samples to be annotated to those situated in a few compact spatial batches. The authors measured the spatial autocorrelation of gray values within input images using a semi-variogram analysis to identify the minimum size of sampling windows to capture "spatial variability beyond locally autocorrelated characteristics".

In our study, we propose the semi-variogram to compute the spatial autocorrelation between training samples and discard neighboring samples from the unlabeled but representative sample set selected using an AL method applied for crop type mapping from SITS. The method, referred to as spatially explicit AL, eliminates redundant samples while achieving results comparable to those obtained by the supervised classifier trained with a larger number of randomly selected samples. Our method was implemented using a Random Forest (RF) classifier and tested in two study areas from the Netherlands and Belgium. We compared our results to the traditional AL approach that does not exploit spatial information to select the most representative samples for training.

The remainder of the paper is organized as follows: Section 2 introduces the two study areas, and the Sentinel-2 SITS used to identify the target crop types. Section 3 is dedicated to the adopted methodology where we are introducing the concept of the spatially explicit AL method proposed in our study. Sections 4 and 5 focus on the presentation of the results and their interpretation. The paper ends with conclusions in Section 6.

2. Study Areas and Datasets

Our research was conducted in two study areas with agricultural land use (Figure 1). Study Area 1 (SA1) is located in Noord Beveland, a municipality in the province of Zeeland in the Southwest of The Netherlands. Being a polder resulting from land reclamation, the area yields fertile soils, and large and regularly shaped parcels, making it suitable for agriculture. Study Area 2 (SA2) is situated in the municipality of Kortrijk (Flemish), also known as Courtrai (French/English).

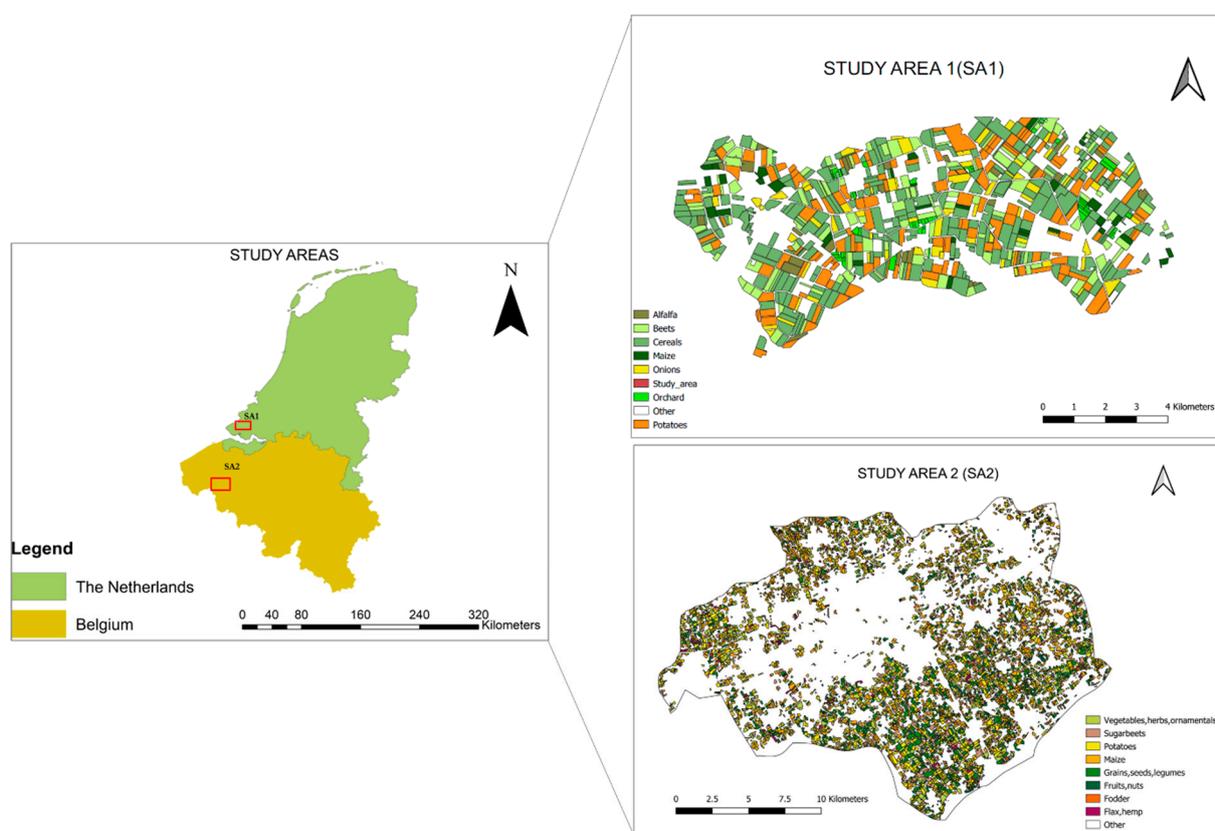


Figure 1. Spatial location of the two investigated study areas: Study Area 1 (SA1) and Study Area 2 (SA2)—including the crop type distribution in these areas.

The area is located in the Flanders Region in the western part of Belgium, along the River Leie (Lys) and the Leie–Scheldt Canal. It is one of the important agricultural areas in the country and contains many farming households. The agriculture parcels in Kortrijk are much smaller than those from Noord Beveland. To implement the proposed method, we used the existing datasets on the cultivated crops in the two selected study areas. A total of 54 crop classes for SA1 and 13 crop classes for SA2 were available. However, we reduced the classes to seven crops for SA1 and eight crop classes for SA2, omitting the classes with less than five parcels. For SA1, the crop data were obtained from the Base Registration Crop Parcels agency in the Netherlands (www.PDOK.nl (accessed on 21 March 2024)), and the parcel boundaries from the Agricultural Area of the Netherlands (AAN). The 1584 parcels available in the investigated study area represent arable land, grassland, natural area, and fallow land. Our research focused solely on the 951 crop parcels and seven crop types:

cereals; potatoes; beets; onions; orchard; maize; and alfalfa (Table 1). The remaining land use/land cover classes were grouped into ‘water’ and ‘other’. For SA2, the crop data were acquired from the Flemish geoportal Geopunt (www.geopunt.be (accessed on 21 March 2024)). The data consist of agricultural parcels associated with cultivated crops. Eight crop classes were used for the research: maize; grains, seeds, and legumes; potatoes; vegetables, herbs, and ornamental plants; sugar beets; fodder; flax and hemp; and fruits and nuts (Table 2). An additional class of ‘other’ areas was added incorporating the remaining land use and land cover classes available in the study area.

Table 1. Total number of crop parcels available in SA1.

Crop Type	Number of Crop Parcels
Cereals	305
Potatoes	157
Beets	117
Onion	68
Orchard	60
Maize	35
Alfalfa	27

Table 2. Total number of crop parcels in SA2.

Crop Type	Number of Crop Parcels
Maize	2989
Grains, seeds, and legumes	1533
Potatoes	1386
Vegetables, herbs, and ornamental plants	1365
Sugar beets	573
Fodder	273
Flax and hemp	96
Fruits and nuts	93

For SA1, we used the crop data from 2019, whereas for SA2 we used the data from 2018. The selection of the year was contingent on the cloud coverage. According to the Royal Meteorological Institute (KNMI) in the Netherlands, 2019 was a very warm, sunny, and quite dry year on average, and a decrease in the rainfall compared to the long-term monthly average was observed in the southwestern part of the country where SA1 is located. On 25 July, the Netherlands had its highest temperature in at least three centuries. This was a time when most of the crops were at their peak growth. Both May and November were unusually cool. These were sowing and harvesting months for maize, potatoes, and onions. In the case of SA2, 2018 was a dry year, with a maximum of 37 °C on the 27 of July and of 36 °C on the 7th of August. The meteorological information for SA1 was obtained from the nearby weather station of Vlissingen via the national meteorological service of KNMI, and the information for Kortrijk from the weather station of Kortrijk via (weer1.com (accessed on 21 March 2024)).

3. Methods

The main methodological steps include (i) image processing; (ii) sample generation for SA1 and SA2. (Note that for our study, we used the existing crop-type database to label the samples selected through AL. In the regions where these data are missing, the users need to label the data through field campaigns.); (iii) AL-based training sample selection considering the spectral domain only; (iv) development of a spatially explicit AL strategy using a semi-variogram; (v) crop type classification, first using a larger number of training samples, then using only samples identified as being relevant by spectral-based AL and, finally, using only training samples generated by spatially explicit AL, and (vi) evaluation of the obtained results (Figure 2).

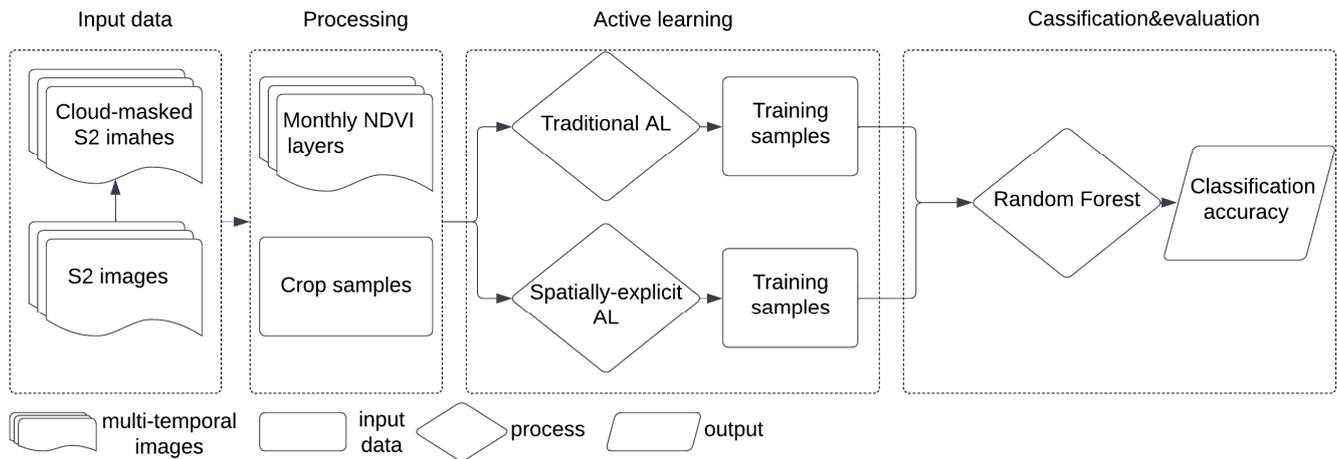


Figure 2. Workflow of the methodology proposed in this study. S2: Sentinel-2 images; NDVI: Normalized Difference Vegetation Index; AL: Active Learning; Traditional AL: AL using solely spectral information; Spatially explicit AL: AL using spatial and spectral information.

3.1. Satellite Image Pre-Processing

The SITS used as input in our research consists of a monthly time series of Sentinel-2 images. The clouds were masked out using the QA60 band. In the next step, the Normalized Difference Vegetation Index (NDVI) for each Sentinel-2 image was computed using bands 8 and 4. The temporal profiles of the target crops representing the average of 50 randomly selected samples in SA1 and 70 randomly selected samples in SA2 are presented in Figures 3 and 4.

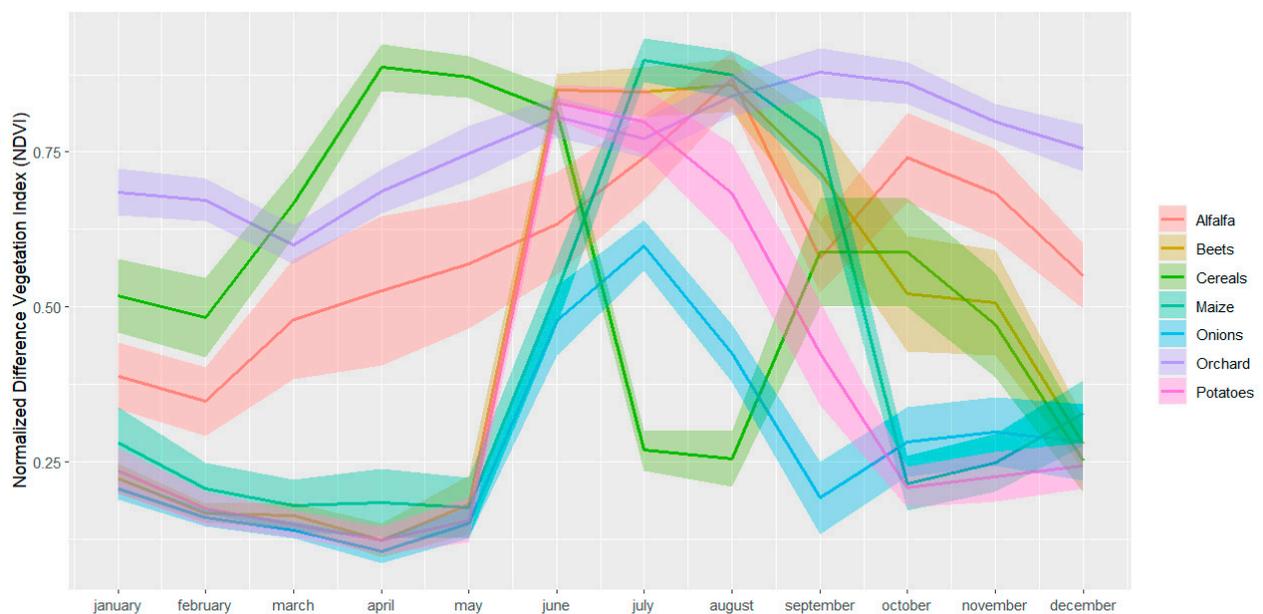


Figure 3. Temporal profiles and standard deviation of the target crops in SA1 calculated using Normalized Difference Vegetation Index (NDVI) Satellite Image Time Series (SITS).

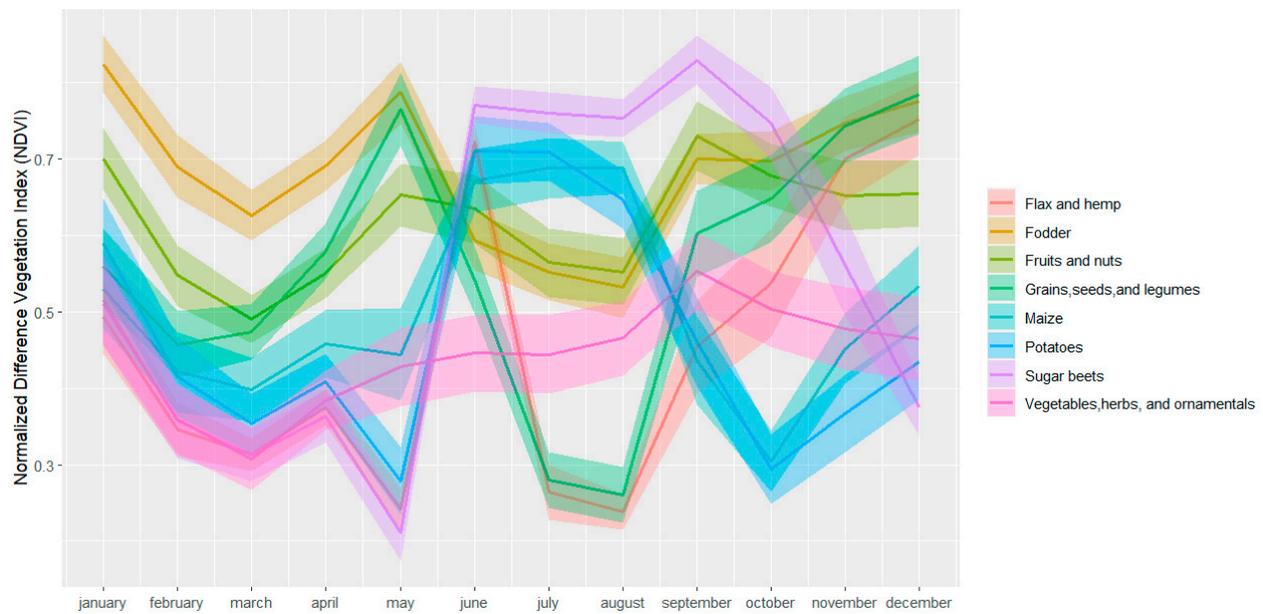


Figure 4. Temporal profiles and standard deviation of the target crops in SA2 calculated using Normalized Difference Vegetation Index (NDVI) Satellite Image Time Series (SITS).

3.2. Training Sample Preparation

Using the crop parcels described in Tables 1 and 2, 630 point-based samples were generated for the investigated target classes in SA1 and 900 point-based samples for the target classes in SA2. For SA1, we selected 70 samples per class using a stratified random sampling approach, whereas for SA2 we applied the same sampling strategies but selected 100 samples per class. The number of samples for the two investigated study areas differs because SA2 has more aggregated classes than SA1 (Figure 3) which increases the inter-class variations as depicted in Figure 4 for ‘vegetables, herbs, and ornamentals’, ‘fruits and nuts’ and ‘flax and hemp’. Consequently, a larger number of samples is required to represent these high variations. The samples were further divided into training (70%) and testing samples (30%). The training and testing sample sets were sampled from different crop parcels to ensure their spatial independence and, in this way, to reduce the risk of overestimating the classification performance [3,41].

3.3. Selection of Training Samples Using Active Learning

The components of an active learner consist of a set of classifiers C trained on a small set of labeled samples L , a query Q , implemented to identify and select the most informative, i.e., uncertain, labels from a set of unlabeled samples U which are not annotated yet and a supervisor S , assigning labels to the retrieved samples [42]. The procedure is initialized by training the selected supervised classifier using the sample set L . In the next step, the query Q identifies and selects the most informative samples from U using a user-defined criterion or several criteria. Lastly, supervisor S annotates the retrieved unlabeled samples that are further added to the L set and used by the supervised classifier for retraining. This is an iterative procedure that continues until it reaches a user-defined stopping criterion that can, for example, be defined based on the classification accuracy [43].

Two AL scenarios were implemented in this research. The first AL scenario uses spectral domain heuristics to query informative samples and is referred to in this paper as spectral-based AL. These heuristics query unlabeled samples using only their characteristics in the feature space. As mentioned in the introduction section, some of these heuristics are posterior probability, least confidence, margin sampling, and entropy [18]. In our work, we propose using Query By Committee (QBC), where a user-defined number of learners are trained from a pool of labeled samples, and each learner, i.e., committee member, votes for the label to which a potential sample belongs. The sample for which

most of the implemented learners of the committee disagree is selected. This method is less computationally intensive than other AL approaches [40]. Since a small number of committee members proved to be efficient in the previous work dedicated to AL [44,45], we used two committee members represented by two RF classifiers [46].

We divided the 350 samples available for SA1 into a labeled sample set L consisting of 40 samples and an unlabeled set U containing 310 samples. For SA2, 40 samples (L) out of 560 available samples were used for initializing the training. In both cases, the L samples were randomly selected. In the next step, the informative unlabeled samples were queried from U and the performance of the committee was assessed with an increment of the number of training samples used to train the model. The prediction accuracy of the committee was assessed by comparing the predicted class label with the known class label of the sample. NDVI values of each unlabeled sample were used to determine its importance using vote entropy (Equation (1)) as a metric for the amount of disagreement between committee members. As expressed by

$$x_{VE}^* = \operatorname{argmax}_x - \sum_i \frac{V_{y_i}}{C} \log \frac{V_{y_i}}{C} \quad (1)$$

where x_{VE}^* = vote entropy for sample x ; y_i = possible labels; V_{y_i} = number of votes received from the committee members; and C = number of members/committee size.

In the second AL scenario, the selection of the informative samples accounted for their characteristics both in the feature space and in spatial information. This implementation is referred to as spatially explicit AL throughout our paper and it accounts for the spatial distribution of the samples assessed by spatial autocorrelation measures. Spatial autocorrelation describes the spatial dependency between objects (or variables) and is an expression of how similarity between objects (or variables) depends on their relative position [47]. It can be quantified by global or local measures or by a semi-variogram. Global measures, such as Moran's I [38] or Geary's C ratio [39] summarize the level of clustering across the entire area of interest in one single statistic without identifying *where* in the area the similarity occurs. For our work, it is crucial to understand *where* the (dis)similarity occurs to select the most relevant, i.e., most informative samples. Therefore, global measures cannot help. Local measures, such as the Local Indicator of Spatial Association (LISA) [48] and local Moran's I [49], indicate the location of clusters explained by the overall global pattern and give a single statistic for each locality. However, the current research needs a measure of spatial autocorrelation at the sample level, on a point-to-point basis. The semi-variogram is a function relating semi-variance to sampling lag, i.e., the distance between samples [50], and can be used as a characterization of the spatial structure of an area. When the semi-variance is plotted against the distance between samples, the semi-variance typically increases until it reaches a plateau. The distance at which the plateau is reached is called the range and the semi-variance at that point is called the sill, i.e., the total variability in the data. The semi-variance when the distance between points is zero is called the nugget and quantifies the randomness of the data [51], accounting for measurement errors and non-spatial variability.

Only sample pairs further apart than the range can be considered spatially uncorrelated. Therefore, only samples further apart than the range were considered in the spatial component for AL. Different semi-variogram models, namely spherical, exponential, and Gaussian models were used to model the semi-variogram and capture the spatial variability of data with distance. The model with the smallest Sum of Squares Error (SSErr), considering both test areas, was chosen as the best-fitting model. Semi-variograms based on NDVI values were estimated for each month for both study areas, using at least 30 pairs of points for each lag distance to obtain a reliable estimate. After several runs, the spherical model gave the smallest SSErr for most of the semi-variograms. For an impression of the varying ranges and nuggets of the fitted models, the values for each month are given in Table 3 for SA1. The range and nugget of the monthly models were used to select the semi-variogram model for the new spatial component in AL. For each month, the semi-

variogram model with the lowest nugget value was chosen. For each study area, from the resulting 12 semi-variograms of the 12 months, the smallest range distance was chosen as the minimum distance for selecting spatially informative samples, i.e., the minimum distance above which samples can be considered uncorrelated. Choosing larger range distances would risk losing potentially informative samples, also this is unfeasible because of area size limitations. For this reason, the chosen ranges were 417.6 m for SA1 and 529 m for SA2.

Table 3. Variation of the ranges and nugget values for the crop type samples cultivated in SA1 calculated using semi-variograms fitted for each month. The range identified for May has been selected to define the minimum distance for selecting spatially informative samples.

Month	Range (m)	Nugget
January	462.3	0.008
February	477.4	0.005
March	611.2	0.029
April	513	0.029
May	417.6	0.025
June	587.3	0.008
July	905.4	0.043
August	721.1	0.050
September	804.9	0.043
October	733.6	0.033
November	735.3	0.028
December	794	0.018

Therefore, all samples with an in-between distance above these ranges were considered spatially informative. The learner first queries an informative sample using the entropy and then assesses the Euclidean distance between the selected sample and the label set. If any of the Euclidean distances between the selected samples and the labeled sample are below the semi-variogram range, the point is discarded. The spatially explicit AL method was implemented to select the most informative samples according to vote entropy and spatial autocorrelation. For both spectral-based and spatially explicit AL, we used a stopping criterion that accounts for the increase in the prediction accuracy with each added sample. Therefore, when the accuracy leveled off, the sample selection procedure stopped.

3.4. Crop Type Classification Using the Random Forest Classifier

Image classification was performed using RF [46]. RF is a very popular classifier in remote sensing due to its high performance [52]. It consists of several internal decision trees built using a randomly selected subset of features and training samples selected randomly through replacement (Breiman, 2001 [46]). This way of selecting training samples and input features minimizes the correlation between the built decision trees [53]. The classifier used about 2/3 of the labeled samples for training, and the remaining 1/3 of the samples, called Out of Bag (OOB) samples, were used to assess the classification accuracy (Breiman, 2001 [46]). RF is mainly sensitive to the number of trees (*ntree*) and the number of selected input features (*mtry*). The *ntree* was set to 1000, whereas *mtry* was defined as being the square root of the total number of input features. Fifty iterations were defined for each RF model. The reported accuracies were averaged across these iterations. An RF-based classification was implemented using the NDVI image stack as an input feature for both study areas. It was first trained using all available annotated samples and then by using training samples generated from the two AL implementations: spectral-based AL and spatially explicit AL. The classification results were assessed using the Kappa coefficient, overall accuracy (OA), User's Accuracy (UA), and Producer's Accuracy (PA) [54–56].

4. Results

4.1. Active Learning Results

In the case of spectral-based AL, the accuracy stopped increasing further after 129 queries, when committee prediction accuracy reached 99.43% for SA1. A total of 169 crop-type samples (40 initial samples plus 129 newly selected samples) were selected through this AL strategy (Table 4). This represents 48% of the entire sample set (350 crop-type samples), excluding water and other classes. For SA2, the accuracy leveled off after 287 queries, with a committee prediction accuracy of 95.2%. In this case, a total of 327 samples were selected (40 initial samples plus 287 newly selected samples) (Table 5). This represents 58% of all available samples excluding the ‘other’ class.

Table 4. Distribution of training samples per class before applying AL and after applying spectral-based AL and spatially explicit AL in SA1.

Crop Type	Total # of Training Samples	# of Training Samples Selected Using Spectral-Based AL	# of Training Samples Using Spatially Explicit AL
Alfalfa	50	22	11
Beets	50	28	16
Cereals	50	15	8
Maize	50	25	15
Onions	50	23	16
Orchard	50	15	7
Potatoes	50	41	24
Total	350	169	97

Table 5. Distribution of training samples per class before applying AL and after applying spectral-based AL and spatially explicit AL in SA2.

Crop Type	Total # of Training Samples	# of Training Samples Selected Using Spectral-Based AL	# of Training Samples Using Spatially Explicit AL
Flax and hemp	70	25	20
Fruits and nuts	70	40	25
Fodder	70	36	25
Grains, seeds, and legumes	70	32	21
Maize	70	56	41
Potatoes	70	58	39
Beets	70	33	30
Vegetables, herbs, and ornamentals	70	47	22
Total	560	327	223

When using spatially explicit AL, the accuracy leveled off after 57 queries with a committee prediction accuracy of 90.9% in SA1. Thus, the labeled sample pool consists of a total of 97 samples (57 newly selected samples plus the 40 initial samples). This represents 28% of the entire available training sample set. For SA2, the accuracy stopped increasing further after 183 queries with a committee prediction accuracy of 82.14%. A total number of 223 informative samples were selected representing 40% of the entire training sample set. Tables 1 and 2 show the sample distribution across target crop types after selecting informative samples using both AL methods. The samples were not well balanced across target crop classes, and classes with high interclass similarity, e.g., potatoes and maize classes from both areas, were given preference in the selection. The distribution of samples per class for both study areas is presented in Tables 4 and 5.

4.2. Image Classification Results Obtained Using Various Training Sample Sets

Four image classification scenarios were tested: Scenario 1—classification using the total number of the available training samples; Scenario 2—classification using training

samples selected by spectral-based AL; Scenario 3—classification using training samples selected by spatially explicit AL; and Scenario 4—classification using a random selection of training samples from the labeled sample set, equal to the number of samples selected by spatially explicit AL. This last experiment was designed to evaluate the benefit of using the developed AL method to select the most informative training samples over selecting the same number of samples randomly.

The accuracies obtained for scenario 1 are presented in Tables 6 and 7. For SA1, the entire 350 crop samples were used for classification obtaining an overall accuracy of 84% (Table 6). For SA2, a total of 560 training crop samples were used and yielded an overall accuracy of 65% (Table 7). The total number of testing samples used for this task was 270.

Table 6. Classification results obtained using different classification scenarios developed for SA1: 1—classification results obtained using all samples; 2—classification results obtained using spectral-based AL; 3—classification results obtained using spatially explicit AL; and 4—classification results obtained using 97 randomly selected samples.

Classification Scenarios	# Samples	Kappa Index	Overall Accuracy (%)
Scenario 1	350	0.82	84
Scenario 2	169	0.79	82
Scenario 3	97	0.78	80
Scenario 4	97	0.66	70

Table 7. Classification results obtained using different classification scenarios developed for SA2: 1—classification results obtained using all samples; 2—classification results obtained using spectral-based AL; 3—classification results obtained using spatially explicit AL; and 4—classification results obtained using 223 randomly selected samples.

Classification Scenarios	# Number	Kappa Index	Overall Accuracy (%)
Scenario 1	560	0.60	65
Scenario 2	327	0.58	63
Scenario 3	223	0.55	60
Scenario 4	223	0.48	54

In the case of the second scenario for SA1, the overall accuracy was 82%. For SA2, there were a total of 327 samples, which gave an overall accuracy of 63%. For the third scenario in SA1, the overall accuracy was 80%. For SA2, 223-AL generated samples were used, which gave an overall accuracy of 60%. The fourth and last classification scenario was dedicated to randomly selecting the same number of samples as obtained by using AL and spectral and spatial components. An overall accuracy of 70% was attained for SA1. For SA2, an overall accuracy of 54% was obtained.

The UA and PA accuracies for all crop-type classes are displayed in Tables 8 and 9 for SA1 and SA2, respectively.

The accuracies were high for SA1. In SA2, on the other hand, there was much more variation in the accuracies: some classes had high PA in all three scenarios (grains, seeds, legumes; and sugar beets), while others had moderate accuracies (flax and hemp; maize; potatoes; and vegetables, herbs, and ornamentals) or even a low producer accuracy (fodder; and fruits and nuts).

The proposed AL method yielded promising results for cereals (UA 90% and PA 95%), alfalfa (UA 86% and PA 95%), and onions (UA 84% and PA 80%), for SA1. The developed method also performed well in aggregated classes like ‘grain, seeds, and legumes’ (UA of 87% and PA of 87%) and the ‘fruits and nuts’ classes (UA of 60% and PA of 30%) as compared to the other approaches. For the ‘vegetables, herbs, and ornamentals’ class, the increase in UA was outweighed by a larger decrease in the PA.

Table 8. Comparison of the User’s (UA) and Producer’s (PA) accuracies of crop types calculated and depicted in % for SA1.

Crop Type	Scenario 1		Scenario 2		Scenario 3	
	UA	PA	UA	PA	UA	PA
Alfalfa	100	95	100	85	86	95
Beets	75	75	65	75	71	75
Cereals	89	85	79	95	90	95
Maize	93	70	93	70	94	75
Onions	94	75	100	75	84	80
Orchard	75	90	87	95	66	95
Potatoes	77	85	68	75	71	75
Water	100	100	100	100	100	100
Other	64	80	76	65	56	30

Table 9. Comparison of UA and PA of crop types in % accuracies for the three scenarios for SA2.

Crop Type	Scenario 1		Scenario 2		Scenario 3	
	UA	PA	UA	PA	UA	PA
Flax and hemp	71	83	74	83	67	87
Fodder	39	30	41	30	38	33
Fruits and nuts	59	33	44	27	60	30
Grains, seeds, legumes	87	87	77	90	87	87
Maize	69	60	67	60	49	63
Potatoes	50	53	51	60	46	60
Sugar beets	63	80	66	77	63	73
Vegetables, herbs, ornamentals	63	67	59	57	75	10
Other	73	90	71	83	65	93

5. Discussion

The overall goal of our work was to assess the potential of a spatially explicit AL strategy for selecting the most informative samples for crop-type mapping from SITS. The goal of this method is to select the smallest number of samples required to achieve good classification results. We showed that our spatially explicit AL strategy provides a high potential for decreasing the time and effort required for sample annotation in the two study areas. Pasolli et al. (2011) [19] have also emphasized that integrating a spatially explicit AL in a single-date image classification task obtained higher accuracy than the AL strategy that solely accounts for spectral information. While Pasolli et al. (2011) [19] used the Euclidean distance between the unlabeled training samples and support vectors (SVs) to select samples that are covered by SVs, we proposed a semi-variogram to discard redundant samples. Stumpf, Lachiche, Malet, Kerle, and Puissant [40] also used the semi-variogram, but with a different aim: to avoid spatially dispersed distribution of landslide samples that would otherwise increase the time and costs for field surveys and/or visual interpretation required to annotate the labels.

Reducing the size of the labeled set makes it more difficult to distinguish the subtle differences between spectrally similar classes. Several classes in SA1 and SA2 showed high inter-class similarities (Figures 3 and 4). For example, in SA1, the NDVI time-series-based crop growth patterns are similar between maize and potatoes. In contrast to these classes, alfalfa, cereals, and orchard show distinct temporal behaviors. In SA2, maize and potatoes also have similar crop growth patterns (see Figure 4) which makes them challenging to distinguish. Previous studies have also highlighted the difficulties of successfully distinguishing crop types with high inter-class similarities like maize and potatoes [13,14]. Sugar beets or ‘grain, seeds, and legumes’, on the other hand, have a more distinct growth pattern and consequently, there is less overlap with the other crops and less confusion and misclassification.

In general, in SA2, accuracies were lower than in SA1 for all classes and in all scenarios. The lower classification accuracy and the smaller reduction in the number of training samples required for SA2 could be because most of the training sample classes are aggregated classes, with high intra-class variation hence causing a challenge in the classification process.

The spatially explicit AL method proposed in our study requires the selection of the range identified by the semi-variogram. The pairs of samples are spatially uncorrelated if the in-between distance is larger than the defined range. The selected samples depend highly on the capability of the implemented semi-variogram to identify the proper range which in turn helps us to select the most informative samples. If, for example, the range is too low, redundant samples might still be present in the sample pool. If the range is too large, the method might discard samples that are informative and essential to achieving good performance. Spatial Simulated Annealing (SSA) [57,58] could be used to optimize the spatial sample design.

Spatial heterogeneity caused by variations in environmental conditions and agricultural management practices might lead to variations in crop diversity and crop representation in the feature space. Therefore, applying the developed AL method to agricultural regions with multiple spatial patterns, e.g., containing a flat area with large parcels and a sloping area with smaller, irregular parcels might be challenging since the samples used to train the machine learning algorithms have to capture the environmental and management conditions that operate at different geographic scales. In areas with multiple spatial patterns, multiple semi-variograms, e.g., per sub-area, need to be generated to capture the spatial variability. In this way, it becomes difficult to select the most suitable range from several ranges generated by the sub-area semi-variograms.

In our work, we used the already annotated samples to test the efficiency of the proposed spatially explicit AL method. Yet, in many areas across the globe, the annotated labels are missing. In this situation, the samples need to be labeled either through visual interpretation, given that very high-resolution images are available, or through intense field campaigns. Our method brings benefits to these scenarios as well since the number of samples and, hence, the time required to annotate them is considerably reduced.

The proposed AL method will benefit future remote sensing-based applications in situations when the researchers can sample only a few locations for time and budget-related restrictions. Contrary to previous studies emphasizing the negative impact of insufficient training samples on the classification results [40], our study revealed the importance of spectral and spatial informativeness of samples in implementing classifiers with high performance.

In addition, this method works well when the training sampling is well-designed. In the case of SA2, almost all reference classes were aggregated classes, which led to low classification accuracy even when the entire dataset was used. This low accuracy was carried on to the developed spatially explicit AL strategy. Despite the overall low accuracy achieved in SA2, the conclusion that less than half of the training samples can be used to achieve similar accuracy if sample selection is based on selecting the spectrally and spatially most informative samples holds for both areas.

Since collecting labeled training samples is an expensive, time-consuming, and laborious task, there are only a few countries where crop-type benchmark datasets are available [59]. Future studies should focus on testing the proposed method in areas where annotated crop samples are limited such as developing countries.

6. Conclusions

This study demonstrated the importance of training sample selection, showing that by selecting spectrally and spatially informative samples, the number of training samples could be reduced to less than half while obtaining similar accuracies. This result was obtained in an area with large, regular parcels and classes consisting of single crops (SA1) as well as for an area with smaller parcels and aggregated classes consisting of several

different crops (SA2). For SA1, the overall classification accuracies were 84% using all available 350 training samples, 82% using 169 samples selected using spectral-based AL, and 80% using 97 samples selected using spatially explicit AL. Selecting 97 samples randomly, without using spectral and/or spatial characteristics, yielded an average accuracy of 70%; therefore, the spatially explicit method gave a higher accuracy for the same number of samples. For SA2, the overall classification accuracies were 65% using all 560 training samples, 63% using 327 samples selected through traditional AL, and 60% using 223 samples selected using spatially explicit AL. Selecting 223 samples randomly resulted in an average accuracy of 54%.

The proposed AL method reveals that accounting for spatial information is an efficient solution to map target crops since it facilitates high accuracy with a low number of samples and, consequently, lower computational resources and time and financial resources for annotation. Future studies could extend the proposed method to different land cover mapping tasks. In addition, further research on the use of semi-variograms could be carried out for sampling design in areas containing sub-areas with different spatial variations caused by varying environmental conditions and management practices. Instead of the semi-variogram, other methods could be used for the spatial sample design, such as SSA.

Author Contributions: B.K.: conceptualization, methodology development, implementation and testing, validation, visualization, writing—original draft; M.B.: conceptualization, supervision, visualization, writing—review and editing, revised the manuscript according to the reviewers' comments; W.B.: conceptualization, supervision, writing—review and editing, revised the manuscript according to the reviewers' comments. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets have been used for this research.

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

References

1. Xu, J.; Zhu, Y.; Zhong, R.; Lin, Z.; Xu, J.; Jiang, H.; Huang, J.; Li, H.; Lin, T. DeepCropMapping: A multi-temporal deep learning approach with improved spatial generalizability for dynamic corn and soybean mapping. *Remote Sens. Environ.* **2020**, *247*, 111946. [[CrossRef](#)]
2. Vuolo, F.; Neuwirth, M.; Immitzer, M.; Atzberger, C.; Ng, W.-T. How much does multi-temporal Sentinel-2 data improve crop type classification? *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *72*, 122–130. [[CrossRef](#)]
3. Nowakowski, A.; Mrziglód, J.; Spiller, D.; Bonifacio, R.; Ferrari, I.; Mathieu, P.P.; Garcia-Herranz, M.; Kim, D.H. Crop type mapping by using transfer learning. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *98*, 102313. [[CrossRef](#)]
4. Rußwurm, M.; Körner, M. Self-attention for raw optical Satellite Time Series Classification. *ISPRS J. Photogramm. Remote Sens.* **2020**, *169*, 421–435. [[CrossRef](#)]
5. Tseng, G.; Kerner, H.; Nakalembe, C.; Becker-Reshef, I. Learning to predict crop type from heterogeneous sparse labels using meta-learning. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Nashville, TN, USA, 19–25 June 2021; pp. 1111–1120.
6. Wang, S.; Waldner, F.; Lobell, D.B. Unlocking Large-Scale Crop Field Delineation in Smallholder Farming Systems with Transfer Learning and Weak Supervision. *Remote Sens.* **2022**, *14*, 5738. [[CrossRef](#)]
7. Xu, J.; Yang, J.; Xiong, X.; Li, H.; Huang, J.; Ting, K.C.; Ying, Y.; Lin, T. Towards interpreting multi-temporal deep learning models in crop mapping. *Remote Sens. Environ.* **2021**, *264*, 112599. [[CrossRef](#)]
8. Wang, Y.; Feng, L.; Zhang, Z.; Tian, F. An unsupervised domain adaptation deep learning method for spatial and temporal transferable crop type mapping using Sentinel-2 imagery. *ISPRS J. Photogramm. Remote Sens.* **2023**, *199*, 102–117. [[CrossRef](#)]
9. Wang, S.; Azzari, G.; Lobell, D.B. Crop type mapping without field-level labels: Random forest transfer and unsupervised clustering techniques. *Remote Sens. Environ.* **2019**, *222*, 303–317. [[CrossRef](#)]
10. Wen, Y.; Li, X.; Mu, H.; Zhong, L.; Chen, H.; Zeng, Y.; Miao, S.; Su, W.; Gong, P.; Li, B.; et al. Mapping corn dynamics using limited but representative samples with adaptive strategies. *ISPRS J. Photogramm. Remote Sens.* **2022**, *190*, 252–266. [[CrossRef](#)]
11. Belgiu, M.; Csillik, O. Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. *Remote Sens. Environ.* **2018**, *204*, 509–523. [[CrossRef](#)]

12. Xia, J.; Yokoya, N.; Adriano, B.; Kanemoto, K. National high-resolution cropland classification of Japan with agricultural census information and multi-temporal multi-modality datasets. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *117*, 103193. [[CrossRef](#)]
13. Belgiu, M.; Bijker, W.; Csillik, O.; Stein, A. Phenology-based sample generation for supervised crop type classification. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *95*, 102264. [[CrossRef](#)]
14. Gella, G.W.; Bijker, W.; Belgiu, M. Mapping crop types in complex farming areas using SAR imagery with dynamic time warping. *ISPRS J. Photogramm. Remote Sens.* **2021**, *175*, 171–183. [[CrossRef](#)]
15. Li, M.; Bijker, W. Vegetable classification in Indonesia using Dynamic Time Warping of Sentinel-1A dual polarization SAR time series. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *78*, 268–280. [[CrossRef](#)]
16. Orynbaikyzy, A.; Gessner, U.; Conrad, C. Spatial Transferability of Random Forest Models for Crop Type Classification Using Sentinel-1 and Sentinel-2. *Remote Sens.* **2022**, *14*, 1493. [[CrossRef](#)]
17. Xun, L.; Zhang, J.; Yao, F.; Cao, D. Improved identification of cotton cultivated areas by applying instance-based transfer learning on the time series of MODIS NDVI. *Catena* **2022**, *213*, 106130. [[CrossRef](#)]
18. Hamrouni, Y.; Paillassa, E.; Chéret, V.; Monteil, C.; Sheeren, D. From local to global: A transfer learning-based approach for mapping poplar plantations at national scale using Sentinel-2. *ISPRS J. Photogramm. Remote Sens.* **2021**, *171*, 76–100. [[CrossRef](#)]
19. Pasolli, E.; Melgani, F.; Tuia, D.; Pacifici, F.; Emery, W.J. Improving active learning methods using spatial information. In Proceedings of the 2011 IEEE International Geoscience and Remote Sensing Symposium, Vancouver, BC, Canada, 24–29 July 2011; IEEE: Vancouver, BC, Canada, 2011; pp. 3923–3926.
20. Sugiyama, M.; Nakajima, S. Pool-based active learning in approximate linear regression. *Mach. Learn.* **2009**, *75*, 249–274. [[CrossRef](#)]
21. Tuia, D.; Ratle, F.; Pacifici, F.; Kanevski, M.F.; Emery, W.J. Active Learning Methods for Remote Sensing Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 2218–2232. [[CrossRef](#)]
22. Di, W.; Crawford, M.M. View generation for multiview maximum disagreement based active learning for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2011**, *50*, 1942–1954. [[CrossRef](#)]
23. Rajan, S.; Ghosh, J.; Crawford, M.M. An active learning approach to hyperspectral data classification. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 1231–1242. [[CrossRef](#)]
24. Li, J.; Bioucas-Dias, J.M.; Plaza, A. Semisupervised hyperspectral image segmentation using multinomial logistic regression with active learning. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 4085–4098. [[CrossRef](#)]
25. Zhong, P.; Gong, Z.; Li, S.; Schönlieb, C.-B. Learning to diversify deep belief networks for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 3516–3530. [[CrossRef](#)]
26. Mitra, P.; Shankar, B.U.; Pal, S.K. Segmentation of multispectral remote sensing images using active support vector machines. *Pattern Recognit. Lett.* **2004**, *25*, 1067–1074. [[CrossRef](#)]
27. Pasolli, E.; Yang, H.L.; Crawford, M.M. Active-metric learning for classification of remotely sensed hyperspectral images. *IEEE Trans. Geosci. Remote Sens.* **2015**, *54*, 1925–1939. [[CrossRef](#)]
28. Ma, L.; Fu, T.; Li, M. Active learning for object-based image classification using predefined training objects. *Int. J. Remote Sens.* **2018**, *39*, 2746–2765. [[CrossRef](#)]
29. Pasolli, E.; Melgani, F.; Alajlan, N.; Bazi, Y. Active learning methods for biophysical parameter estimation. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 4071–4084. [[CrossRef](#)]
30. Malek, S.; Miglietta, F.; Gobakken, T.; Næsset, E.; Gianelle, D.; Dalponte, M. Optimizing field data collection for individual tree attribute predictions using active learning methods. *Remote Sens.* **2019**, *11*, 949. [[CrossRef](#)]
31. Amor, I.B.S.B.; Chehata, N.; Bailly, J.-S.; Farah, I.R.; Lagacherie, P. Parcel-based active learning for large extent cultivated area mapping. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *11*, 79–88. [[CrossRef](#)]
32. Niazmardi, S.; Homayouni, S.; Safari, A. A computationally efficient multi-domain active learning method for crop mapping using satellite image time-series. *Int. J. Remote Sens.* **2019**, *40*, 6383–6394. [[CrossRef](#)]
33. Zhang, Z.; Pasolli, E.; Crawford, M.M. Crop Mapping through an Adaptive Multiview Active Learning Strategy. In Proceedings of the 2019 IEEE International Workshop on Metrology for Agriculture and Forestry (MetroAgriFor), Portici, Italy, 24–26 October 2019; pp. 307–311.
34. Liu, P.; Zhang, H.; Eom, K.B. Active deep learning for classification of hyperspectral images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *10*, 712–724. [[CrossRef](#)]
35. Cao, X.; Yao, J.; Xu, Z.; Meng, D. Hyperspectral image classification with convolutional neural network and active learning. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 4604–4616. [[CrossRef](#)]
36. Demir, B.; Minello, L.; Bruzzone, L. A cost-sensitive active learning technique for the definition of effective training sets for supervised classifiers. In Proceedings of the 2012 IEEE International Geoscience and Remote Sensing Symposium, Munich, Germany, 22–27 July 2012; pp. 1781–1784.
37. Pasolli, E.; Melgani, F.; Tuia, D.; Pacifici, F.; Emery, W.J. SVM active learning approach for image classification using spatial information. *IEEE Trans. Geosci. Remote Sens.* **2013**, *52*, 2217–2233. [[CrossRef](#)]
38. Moran, P.A. Notes on continuous stochastic phenomena. *Biometrika* **1950**, *37*, 17–23. [[CrossRef](#)] [[PubMed](#)]
39. Geary, R.C. The contiguity ratio and statistical mapping. *Inc. Stat.* **1954**, *5*, 115–146. [[CrossRef](#)]
40. Stumpf, A.; Lachiche, N.; Malet, J.-P.; Kerle, N.; Puissant, A. Active learning in the spatial domain for remote sensing image classification. *IEEE Trans. Geosci. Remote Sens.* **2013**, *52*, 2492–2507. [[CrossRef](#)]

41. Ploton, P.; Mortier, F.; Réjou-Méchain, M.; Barbier, N.; Picard, N.; Rossi, V.; Dormann, C.; Cornu, G.; Viennois, G.; Bayol, N.; et al. Spatial validation reveals poor predictive performance of large-scale ecological mapping models. *Nat. Commun.* **2020**, *11*, 4540. [[CrossRef](#)]
42. Li, M.; Sethi, I.K. Confidence-based active learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2006**, *28*, 1251–1261.
43. Vlachos, A. A stopping criterion for active learning. *Comput. Speech Lang.* **2008**, *22*, 295–312. [[CrossRef](#)]
44. Settles, B. *Active Learning Literature Survey*; University of Wisconsin-Madison, Department of Computer Sciences: Madison, MI, USA, 2009.
45. Seung, H.S.; Opper, M.; Sompolinsky, H. Query by Committee. In Proceedings of the Fifth Annual Workshop on Computational Learning Theory, Pittsburgh, PA, USA, 27–29 July 1992; pp. 287–294.
46. Breiman, L. Random Forest. *Mach. Learn.* **2001**, *45*, 31–39.
47. Fotheringham, A.S. “The problem of spatial autocorrelation” and local spatial statistics. *Geogr. Anal.* **2009**, *41*, 398–403. [[CrossRef](#)]
48. Anselin, L. Local indicators of spatial association—LISA. *Geogr. Anal.* **1995**, *27*, 93–115. [[CrossRef](#)]
49. Zhang, C.; Luo, L.; Xu, W.; Ledwith, V. Use of local Moran’s I and GIS to identify pollution hotspots of Pb in urban soils of Galway, Ireland. *Sci. Total Environ.* **2008**, *398*, 212–221. [[CrossRef](#)]
50. Curran, P.J. The semivariogram in remote sensing: An introduction. *Remote Sens. Environ.* **1988**, *24*, 493–507. [[CrossRef](#)]
51. Diggle, P.J.; Tawn, J.A.; Moyeed, R.A. Model-based geostatistics. *J. R. Stat. Soc. Ser. C Appl. Stat.* **1998**, *47*, 299–350. [[CrossRef](#)]
52. Belgiu, M.; Drăguț, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [[CrossRef](#)]
53. Gislason, P.O.; Benediktsson, J.A.; Sveinsson, J.R. Random Forests for land cover classification. *Pattern Recognit. Lett.* **2006**, *27*, 294–300. [[CrossRef](#)]
54. Congalton, R.G. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sens. Environ.* **1991**, *37*, 35–46. [[CrossRef](#)]
55. Foody, G.M. Status of land cover classification accuracy assessment. *Remote Sens. Environ.* **2002**, *80*, 185–201. [[CrossRef](#)]
56. Stehman, S.V.; Czaplewski, R.L. Design and analysis for thematic map accuracy assessment: Fundamental principles. *Remote Sens. Environ.* **1998**, *64*, 331–344. [[CrossRef](#)]
57. Van Groenigen, J.; Stein, A. *Constrained Optimization of Spatial Sampling Using Continuous Simulated Annealing*; Wiley Online Library: Hoboken, NY, USA, 1998; pp. 0047–2425.
58. Shetty, S.; Gupta, P.K.; Belgiu, M.; Srivastav, S.K. Assessing the Effect of Training Sampling Design on the Performance of Machine Learning Classifiers for Land Cover Mapping Using Multi-Temporal Remote Sensing Data and Google Earth Engine. *Remote Sens.* **2021**, *13*, 1433. [[CrossRef](#)]
59. Selea, T. AgriSen-COG, a Multicountry, Multitemporal Large-Scale Sentinel-2 Benchmark Dataset for Crop Mapping Using Deep Learning. *Remote Sens.* **2023**, *15*, 2980. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.