



Hierarchical clustering and stochastic distance for indirect semi-supervised remote sensing image classification

Gabriela Ribeiro Sapucci¹ · Rogério Galante Negri¹

© Springer Nature Switzerland AG 2019

Abstract

Usually, image classification methods have supervised or unsupervised learning paradigms. While unsupervised methods do not need training data, the meanings behind the classified elements are not explicitly known. Conversely, supervised methods are able to provide classification results with an intrinsic meaning, since a labeled dataset is available for training, which may be a limitation in some cases. The semi-supervised learning paradigm, which simultaneously exploits both labeled and unlabeled data, may be an alternative to this dilemma. This work proposes a semi-supervised classification framework through the combination of the Hierarchical Divisive Algorithm and stochastic distance concepts, where the former is adopted to automatically determine clusters in the data and the latter is used to label such clusters in a supervised way. In order to verify the potential of the proposed framework, two case studies about land use and land cover classification were carried out in an Amazonian area using synthetic aperture radar and multispectral data acquired by ALOS PALSAR and LANDSAT-5 TM sensors. Supervised methods based on statistical concepts were also included in these studies as baselines. The results show that when very small training sets are available, the proposed method provides results up to 14.6% and 3.8% more accurate than the baselines with respect to the classification of TM and PALSAR images, respectively.

Keywords Semi-supervised · Indirect model · Stochastic distance · Clustering · Image classification · Remote sensing

1 Introduction

Image classification is one of the most important pattern recognition applications for remote sensing studies. Typical examples of environmental studies supported by image classification and remote sensing data are the monitoring of forest [10] and river [18] areas, agricultural inspection [2, 14], mapping areas affected by natural disasters [7], urban planning [19], and even fish-farming studies [26]. These techniques aim to perform automatic recognition of elements/targets/objects in the scene (i.e., the remote sensing imagery) through a classification function,

which acts as a decision rule applied to the information measured by the sensor.

Several image classification methods have been proposed in the literature. Usually, these methods are categorized in terms of their learning paradigm, which defines how the classification function is modeled. Among the different learning paradigms proposed in the literature, supervised and unsupervised are the most common. In the case of supervised methods, the classification function is modeled based on prior information available in training sets, from which it is possible to estimate the function's parameters. On the other hand, unsupervised methods perform the classification of elements/targets/objects in

The authors thank FAPESP (Grants 2014/14830-8, 2016/06242-4 and 2018/01033-3) for their financial support of this research.

✉ Rogério Galante Negri, rogerio.negri@unesp.br; Gabriela Ribeiro Sapucci, gabrielasapucci@gmail.com | ¹Instituto de Ciência e Tecnologia – ICT, Universidade Estadual Paulista – UNESP, São José dos Campos, SP, Brazil.



SN Applied Sciences (2019) 1:272 | <https://doi.org/10.1007/s42452-019-0278-x>

Received: 13 December 2018 / Accepted: 18 February 2019 / Published online: 27 February 2019

the image based on the similarities found in its spectral behavior.

In terms of results, unsupervised methods are able to produce clusters of similar elements without an assigned meaning. Although it is possible to manually assign a meaning to these clusters, a consistent “group meaning” cannot be guaranteed [22]. Faced with this characteristic, supervised methods tends to be more suitable for producing classification maps, since the classes are previously known through a training set. However, the quality of their results depends on the sufficiency and quality of the training set, which may be a limitation in some cases.

The semi-supervised paradigm combines supervised and unsupervised concepts in order to learn robust classifiers without depending on large training sets. Several semi-supervised models have been proposed in the literature, such as the generative, low-density separators, graph-based, and change of representation models [5]. This last mentioned model is referred to in this paper as the “indirect model.”

The graph-based model is widely adopted in remote sensing applications, especially for hyperspectral image classification, where the high dimensionality of the data makes large training datasets necessary to learn through supervised methods. Exploiting contextual (i.e., the pixel's neighborhood) information and kernel function concepts, [4] proposed the graph-based approach.

Through reformulating support vector machines (SVM) and using a particular kernel function, [15] presented two alternative low-density separators to deal with binary classifiers trained by small training sets. In order to cope with small and noisy training sets, [3] proposed another low-density separator modeled in terms of contextual information. Integrating a Markov random field and the Expectation-Maximization algorithm, [20] proposed a generative semi-supervised method to be applied in urban area classification.

The so-called indirect models are less frequent in the literature when compared with the other mentioned models. A recent example of an indirect semi-supervised model was presented in [24], which comprises an iterative process combining a variant of the Fuzzy C-Means algorithm followed by a classification process with SVM. The manifold notion has been also used to change data representation and help to discover intrinsic information in low-dimensional and noisy datasets [6].

When defined through the straight combination of unsupervised and supervised paradigms, the indirect model stands out as one of the simplest ways to perform semi-supervised learning. In this case, an unsupervised method is first applied to generate clusters of similar elements without semantic meaning assigned

to them. After this, a supervised method is trained using a few labeled data and then applied to assign a class to each cluster. Although appropriate learning is not guaranteed by the supervised method due to the reduced quantity of training data, it is expected that clusters generated by the unsupervised method represent a simplification of the complexity of the original dataset. Since several classification methods have been proposed in the literature, it is easy to conclude that plenty of combinations involving unsupervised and supervised methods may be considered to define an indirect semi-supervised classification framework.

Stochastic distance appears to be a convenient tool with which to compose an indirect model of semi-supervised learning. Formalized based on Shannon's information theory [23], stochastic distances may be understood as similarity measures between probability density models. Stochastic distances have been shown to be useful in several remote sensing applications, such as image segmentation [16], filtering [28], and region-based classification [17, 25]. In this context, a minimum distance classifier equipped with stochastic distances allows a class to be assigned to a cluster based on the similarity between its probability distribution models. To the best of the author's knowledge, no semi-supervised methods defined through the combination of clustering and stochastic distance concepts can be found in the literature.

Faced with the presented motivation, this work proposes an indirect semi-supervised image classification framework based on the combination of clustering and stochastic distance. Two case studies about land use and land cover classification across a region near the Tapajós National Forest, Brazil, are carried out in order to assess the performance of the proposed semi-supervised method. The first case study uses a multispectral image acquired by the LANDSAT-5 TM sensor, while a synthetic aperture radar (SAR) image acquired by the ALOS PAL-SAR sensor is adopted in the second study. Comparisons with methods from the literature are included in this study. Beyond the distinct kinds of images, these case studies also consider different classification scenarios and training sets of various sizes.

The following text is organized as follows: Sect. 2 presents preliminary notations about image classification and learning paradigms via brief discussions regarding data clustering by unsupervised methods and stochastic distances; in Sect. 3, the proposed indirect semi-supervised method is formalized; experiments and results regarding the above-mentioned case studies are presented and discussed in Sect. 4; finally, the conclusions and future directions for this work are shown in Sect. 5.

2 Preliminary concepts

This section briefly discusses basic concepts about image classification (Sect. 2.1), clustering by hierarchical algorithm (Sect. 2.2), and stochastic distance (Sect. 2.3). Such discussions are the basis of the proposed indirect semi-supervised method introduced in Sect. 3.

2.1 Image classification and learning paradigms

Image classification is the application of $F : \mathcal{X} \rightarrow \mathcal{Y}$ on the attribute vector \mathbf{x}_i behind each pixels s_i of an image \mathcal{I} , with support $S \subset \mathbb{N}^2$, in order to assign a class indicator expressed by y_i . The means that how F is obtained depends on the learning paradigm. In the supervised paradigm such modeling uses the information available in a training set $\mathcal{D} = \{(\mathbf{x}_i, y_i) \in \mathcal{X} \times \mathcal{Y} : i = 1, \dots, m\}$ with $m \in \mathbb{N}^*$, where the indicators $y = 1, \dots, c$ define a class in $\Omega = \{\omega_1, \omega_2, \dots, \omega_c\}$ with a semantic meaning assigned. On the other hand, unsupervised methods are not supported by training sets in modeling F . In this case, the learning is based on analogies found when the dataset is analyzed and the results are clusters of similar elements without semantic meaning. Formally, let $\mathcal{U} = \{\mathbf{x}_i : \mathcal{I}(s_i) = \mathbf{x}_i; s_i \in S\}$, the dataset of attribute vectors from an image \mathcal{I} , which builds a function $F : \mathcal{U} \rightarrow \mathcal{G}$ that assigns the elements of \mathcal{U} to subsets $\mathbf{G}_j \subseteq \mathcal{G}$, $j = 1, \dots, h$ of similar elements such that $\bigcup_{j=1}^h \mathbf{G}_j = \mathcal{G}$ and $\bigcap_{j=1}^h \mathbf{G}_j = \emptyset$.

As mentioned, unsupervised methods are not able to assign a semantic meaning through the classification process. It is possible to interpret such results and then manually give a semantic meaning to the identified clusters, but the consistency between these clusters and such meaning is not guaranteed. When the classes are defined a priori, supervised classification is preferred [13]. Supervised methods can produce accurate results since the training set has enough information to model the classifier [4]. This imposition is usually related to large training sets, which is a limitation in some cases, for example, due to the cost and time needed to label training examples [30]. An alternative to the insufficiency of training data may be minimized exploitation of the implicit information of unlabeled data. This alternative motivated the development of the semi-supervised paradigm [5].

Modifications to well-known methods in order to use unlabeled data in the training process [3, 12] use intermediary processes to expand the training set [13, 21] and combine supervised and unsupervised methods [24]. Especially, the latter example is denominated as "indirect semi-supervised learning," which is the focus of this work.

2.2 Hierarchical divisive algorithm

Clustering methods play an important role in exploratory data analysis, especially in cases where there is little if any knowledge about the data [11]. Furthermore, in a "classification point of view," clustering and unsupervised methods are synonymous, since they agree with the definitions of the previous section.

Among several clustering methods proposed in the literature, the Hierarchical Algorithms perform successive mergers or divisions over the initial dataset [9]. In particular, the Hierarchical Divisive Algorithm (HDA) starts with a single set that comprehends all the data involved in the clustering process, which is successively and recursively split into new subsets/clusters. As a joint step, the new "sub-clusters" are determined, for example, by the K -Means algorithm [29]. When the HDA is combined with K -Means and $K = 2$, the hierarchical data clustering comprehends a binary tree, of which the determined clusters are the leaves.

Regarding the HDA, the so-called successive-recursive splitting process is limited by two criteria: minimum cluster size (i.e., the set cardinality) and minimum cluster diameter (i.e., the biggest distance found between two elements inside the cluster). While the minimum cluster size prevents the definition of too-small clusters, the minimum cluster diameter avoids splitting compact clusters and induces the splitting of clusters that are diffuse over the attribute space.

2.3 Stochastic distances and minimum distance rule for data labeling

Stochastic distances come from the information theory formalized in [23]. Such distances can be used to measure the similarity between two sets of information according to the distance between their probability distributions. The Bhattacharyya distance is a classical stochastic distance commonly used in remote sensing applications. If the multivariate Gaussian distribution is adopted to model the information sets, the Bhattacharyya distance is written as:

$$B_G(\mathbf{U}, \mathbf{V}) = \frac{1}{8} (\boldsymbol{\mu}_U - \boldsymbol{\mu}_V)^T \left(\frac{\boldsymbol{\Sigma}_U + \boldsymbol{\Sigma}_V}{2} \right)^{-1} (\boldsymbol{\mu}_U - \boldsymbol{\mu}_V) + \frac{1}{2} \ln \left(\frac{|\boldsymbol{\Sigma}_U + \boldsymbol{\Sigma}_V|}{\sqrt{|\boldsymbol{\Sigma}_U| + |\boldsymbol{\Sigma}_V|}} \right) \quad (1)$$

where $\boldsymbol{\mu}_Z$ and $\boldsymbol{\Sigma}_Z$ are the average vector and covariance matrix estimated for modeling the random variable \mathbf{Z} , with $(\cdot)^T$, $|\cdot|$, and $(\cdot)^{-1}$ denoting respectively the transpose, determinant, and inverse operations.

Among several applications, stochastic distances may be used in a minimum distance rule to classify unlabeled datasets [25]. Supposing a training set $D = \{(x_i, y_i) \in \mathcal{X} \times \mathcal{Y} : i = 1, \dots, m\}$ and an unlabeled set of clusters $\mathcal{G} = \{\mathbf{G}_1, \mathbf{G}_2, \dots, \mathbf{G}_h\}$, as defined in Sect. 2.1, the elements of cluster \mathbf{G}_k are labeled as ω_j according to the following decision rule:

$$(\mathbf{G}_k, \omega_j) \Leftrightarrow j = \arg \min_{j=1, \dots, c} B(f_{\mathbf{G}_k}, f_{\omega_j}), \tag{2}$$

where $f_{\mathbf{G}_k}$ and f_{ω_j} are probability density functions that model the distribution of elements in \mathbf{G}_k and the elements of D assigned to ω_j , respectively. $B(\cdot, \cdot)$ is a stochastic distance like Eq. 1.

3 Clustering and stochastic distance for indirect semi-supervised image classification

Based on the motivations and concepts presented in Sects. 1 and 2, an indirect semi-supervised image classification framework is proposed. A general flowchart regarding such framework is shown in Fig. 1, where three main steps appear: “clustering”, “cluster labeling using stochastic distance,” and “classification result assessment.”

Firstly, an unsupervised classification (referenced here as “clustering”) of the input image is performed. The HDA using the K -Means algorithm, as discussed in Sect. 2.2, is adopted for this purpose. It is worth mentioning that a parameter tuning process is needed in order to define the best configuration for minimum cluster elements and diameter values. This process is guided based on the visual analysis of results.

Once an appropriate unsupervised classification of the input image is achieved, a labeling process based on a minimum stochastic distance classification (MSDC), as presented in Sect. 2.3, is performed on each defined cluster. The Bhattacharyya distance (Eq. 1) is adopted. It is expected that a training set is available to estimate the parameters that model the distribution probabilities of the classes.

Finally, as an additional step, the classification result may be assessed using some accuracy measure.

Given that HDA and MSDC are at the core of previous discussions, the proposed framework is referred to as HDA+MSDC.

4 Results and discussions

In order to check the effectiveness of the proposed HDA+MSDC framework, two case studies were carried out. Given the “statistical nature” of this framework, the classic Maximum Likelihood Classifier (MLC) and Mahalanobis Distance Classifier (MDC) methods were included as baselines in these studies. Further details regarding the MLC and MDC methods are found in [27].

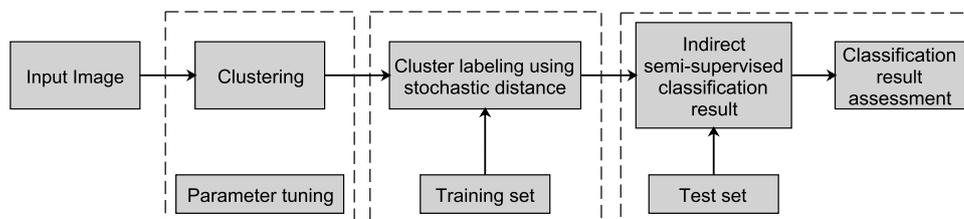
The aforementioned case studies deal with land use and land cover classification in a region near the Tapajós National Forest in the state of Pará, Brazil, using images acquired by ALOS PALSAR and LANDSAT-5 TM sensors. Figure 2 shows the study area’s location. The PALSAR data (Fig. 3a) acquired on March 13, 2009, have HH, HV, and VV intensity polarization in the L-band and 20 m of spatial resolution and cover an area of 729×1100 pixels. The TM image (Fig. 4b) obtained on September 26, 2010, has five multispectral bands with 30 m spatial resolution and corresponds to an area of 650×650 pixels.

PALSAR data are a convenient choice for forest land cover mapping. For a synthetic aperture radar (SAR), atmospheric factors like dense cloud cover do not interfere with information acquisition, and its frequency (L-band) allows better land cover characterization in rainforest regions.

On the other hand, TM data were used to test the proposed framework because they use a different kind of image (i.e., multispectral) than SAR that is frequently adopted in forest monitoring studies [1]. Additionally, TM images are freely available for the study area and near the period when the mentioned fieldwork campaign was carried out.

Based on a fieldwork campaign conducted in the study area in September 2009, different land uses and land cover classes were identified. In such fieldwork, several sites across the study area were visited and their land use/cover annotated. Surveys, geographic localization, and photos

Fig. 1 Flowchart of proposed indirect semi-supervised image classification framework



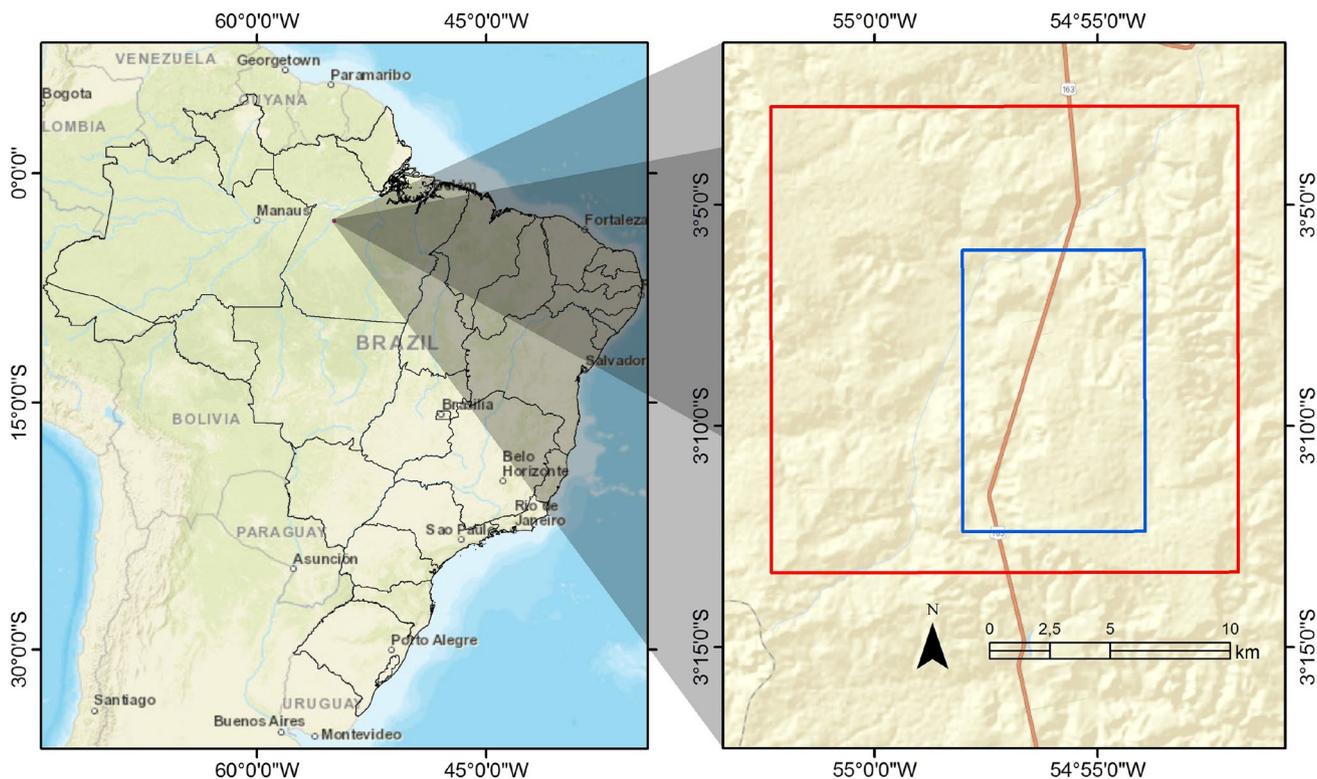


Fig. 2 Study area location. Blue and red rectangles represent the ALOS PALSAR and LANDSAT-5 TM image areas, respectively

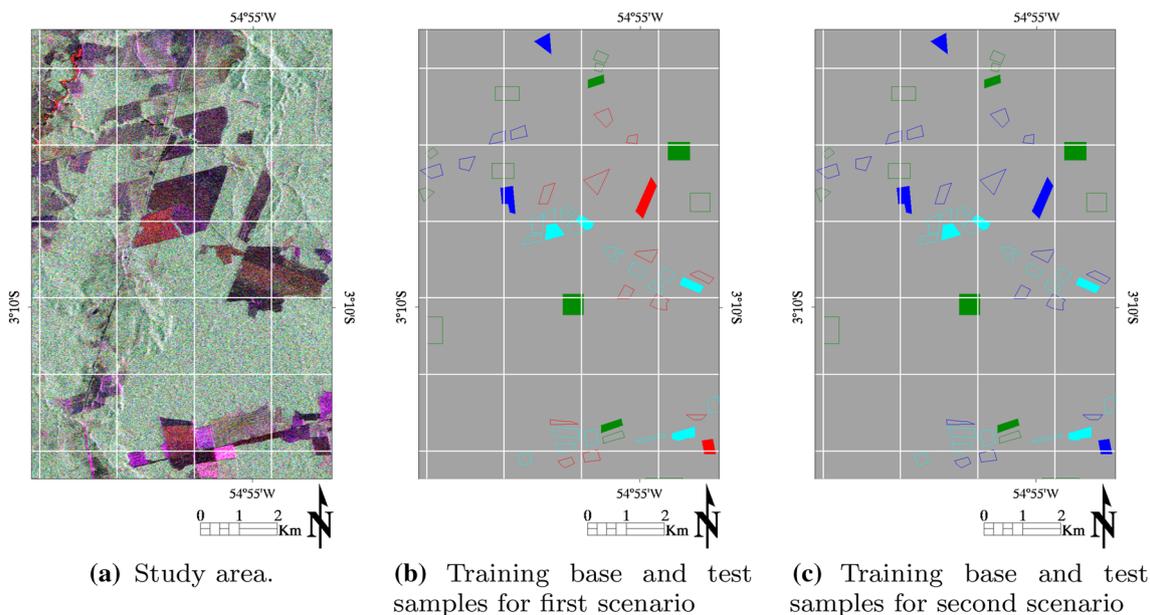


Fig. 3 ALOS PALSAR image in R(HH)G(HV)B(VV) color composition and the spatial distribution of training base (solid) and test set (void) sample polygons for each classification scenario. For class/scenario color legend, see Table 1

were registered in this stage. Using the collected information, several land cover polygons were identified for PALSAR and TM images.

It is worth noting that the fieldwork was carried out six months after the PALSAR image acquisition; then, a visual interpretation was necessary to relate the field

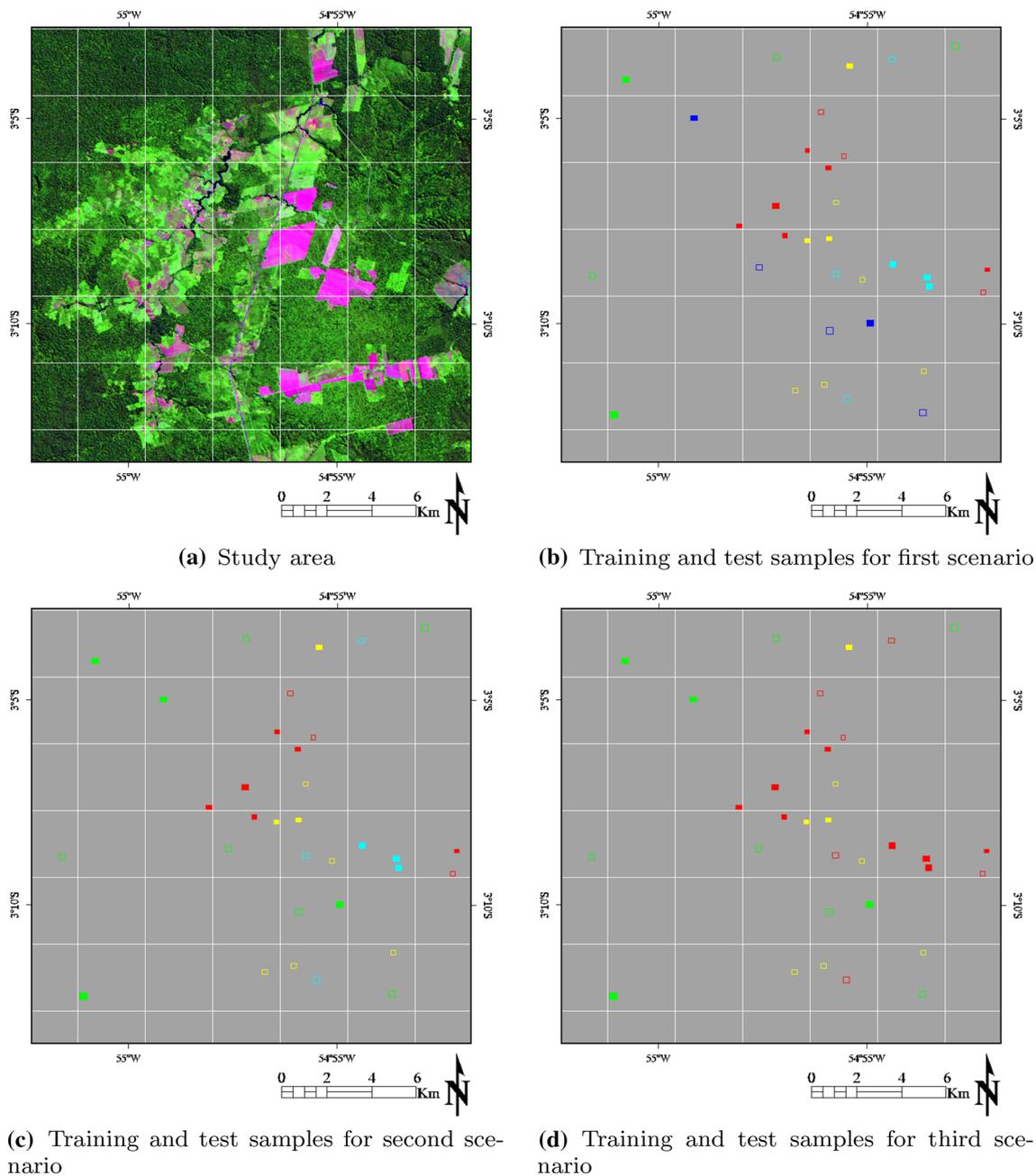


Fig. 4 LANDSAT-5 TM image in R(4)G(3)B(5) color composition and the spatial distribution of training (solid) and test (void) sample polygons for each classification scenario. For class/scenario color legend, see Table 2

observations to the image. Due to the spectral behavior of some land cover classes (especially pasture and agriculture) as well as cloud cover conditions, the TM image adopted in this research was taken approximately one year after the fieldwork campaign. In this case, the fieldwork registers were updated through temporal analysis of LANDSAT images and then used as a basis for identifying land cover polygons across the TM area.

Regarding the LANDSAT-5 TM image, the classes of agriculture (AG), pasture (PS), new regeneration (NR), old regeneration (OR), and forest (FO) were identified. These classes were organized into three classification scenarios. The first scenario is defined by the five mentioned classes. The second scenario arises from merging FO and OR into a single class called high biomass (HB). Finally,

the third scenario merges AG and NR to define the low biomass (LB) class.

Similarly, for the ALOS PALSAR, PS, AG, FO, and bare soil (BS) were considered as primary classes. This slight difference among the classes considered in TM and PALSAR images comes from the different data acquisition (i.e., 2010 and 2009, respectively) and is mainly due to sensor type (i.e., multispectral and SAR). Furthermore, just two scenarios were considered for the PALSAR data. While the first scenario is defined by all primary classes, the second scenario has the PS and SE classes merged into the LB class.

After identifying land cover classes in the study area and collecting the respective samples (as polygons), this information was randomly split into two subsets. In such splitting process, for each land cover class identified, it was imposed that the first subset of polygons should comprise approximately one-third of the pixels, and the remaining polygons should be placed in the second subset. The first subset is referred to as the “training base” and the other as the “test set.” Since the polygons may have different sizes and shapes, the mentioned proportion are not guaranteed.

Posteriorly, the training base subset was adopted to define training sets with 10, 15, 25, and 50 pixels per class. The pixel selection were randomly repeated 10 times, producing 40 training sets. The second subset (comprising approximately two-thirds of all sample polygons) was used to test each classification result from the classification methods learned through the defined training sets.

The use of training sets with varying numbers of pixels per class aims to observe if the performance gain/loss also

follows a non-linear tendency. Furthermore, the numbers of 10, 15, 25, and 50 pixels/class have the objective of defining small training sets that make sense in the context of semi-supervised learning.

Tables 1 and 2 summarize the classes and scenarios of PALSAR and TM images, respectively. Moreover, these tables also present the number of samples available in the training base and test set.

The spatial distribution of the training base and test set samples of the PALSAR image for each scenario are presented in Fig. 3b, c, and those of the TM image in Fig. 4b–d. It is worth observing that when two classes are merged, the respective samples inside the training base and test set are also merged.

The HDA parameters were fine-tuned considering different values for the “minimum cluster size” and “minimum cluster diameter.” The tested values for “minimum cluster diameter” were in {0.1, 0.25, 0.5, 0.75, 0.9} and the “minimum cluster size” in {200, 1K, 10K, 100K}. The best parameter configuration found for “minimum cluster diameter” was 0.1 for the TM image and 0.25 for the PALSAR image. Regarding the “minimum cluster size,” 1K was found to be the best value for both PALSAR and TM images.

The accuracy of the obtained classification results was assessed according to the kappa agreement coefficient [8] computed for the test samples. Furthermore, since 10 training sets with different numbers of pixels per class (i.e., 10, 15, 25 and 50) of each scenario were randomly defined from the training base, a total of 10 classifications were produced for each of these training set’s dimensions and scenarios as well as the respective kappa values. Consequently, in order to obtain a general value to represent the performance of the analyzed methods, with respect to each scenario and training set dimension, the average and standard deviation of kappa values were computed through the 10 individual kappas. Although simple, this process avoids the influence of a particular training set (randomly defined, as previously discussed) on the method assessment.

The above-discussed process is represented by the scheme in Fig. 5. All processing was conducted on a computer with an Intel Core i7 processor and 16 GB of RAM running the Ubuntu-Linux operating system version 14.04.

Table 1 Training base and test set for the ALOS PALSAR image and respective classification scenarios

Primary classes	Training base Pixels/polygons	Test set Pixels/polygons	Scenarios	
			1st	2nd
PS—Pasture	2679/2	3717/4	● PS	● LB
BS—Bare soil	3052/2	10,376/12	● BS	
AG—Agriculture	4591/4	12,464/16	● AG	
FO—Forest	6787/4	11,807/10	● FO	

Table 2 Training base and test set for the LANDSAT-5 TM image and respective classification scenarios

Primary classes	Training base Pixels/polygons	Test set Pixels/polygons	Scenarios		
			1st	2nd	3rd
AG—Agriculture	177/3	252/5	● AG	● AG	● AG
PS—Pasture	147/3	356/6	● PS	● PS	● LB
NR—New regeneration	270/3	216/3	● NR	● NR	
FO—Forest	211/2	262/3	● FO	● HB	● HB
OR—Old regeneration	180/2	262/3	● OR		

The Interactive Data Language (IDL) programming language was used to implement the classification methods.

The performance of the analyzed methods is shown in Figs. 6 and 7. Regarding the results for the PALSAR

image (Fig. 6), HDA+MSDC using small training sets, with 10 and 15 pixels per class, provides more accurate classification results than MLC and MDC. When bigger training sets are considered, and the semi-supervised

Fig. 5 General flowchart of the experiment

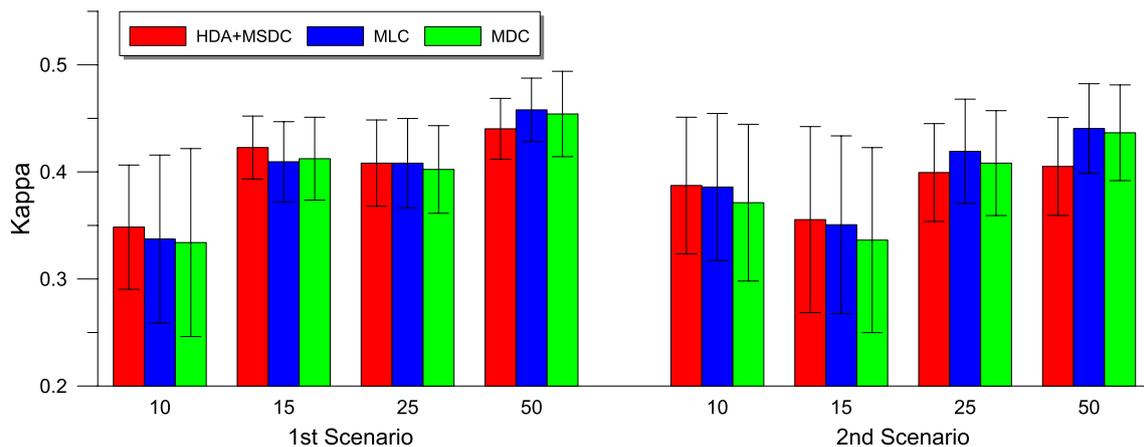
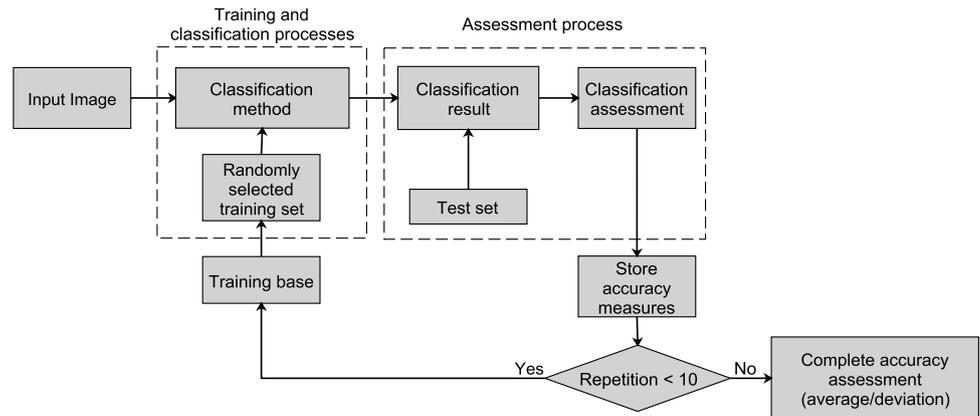


Fig. 6 Performance of analyzed methods for each scenario and training condition in ALOS PALSAR image

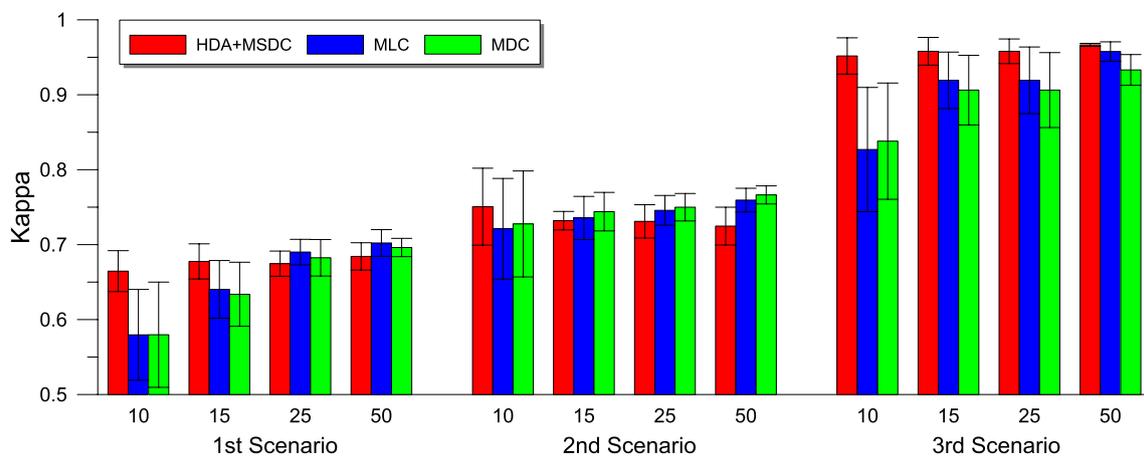


Fig. 7 Performance of analyzed methods for each scenario and training condition in LANDSAT-5 TM image

Table 3 Best performance methods in each scenario and training condition for the ALOS PALSAR image

Pixels per class	ALOS PALSAR	
	Scenarios (Kappa/standard deviation)	
	1st	2nd
10	HDA + MSDC 0.348/0.058	HDA MSDC 0.387/0.064
15	HDA + MSDC 0.423/0.029	HDA + MSDC 0.355/0.087
25	HDA+MSDC 0.408/0.040	MLC 0.419/0.049
50	MLC 0.458/0.030	MLC 0.441/0.042

Table 4 Best performance methods in each scenario and training condition for the LANDSAT-5 TM image

Pixels per class	LANDSAT-5 TM		
	Scenarios (Kappa/standard deviation)		
	1st	2nd	3rd
10	HDA + MSDC 0.665/0.027	HDA + MSDC 0.751/0.051	HDA + MSDC 0.952/0.024
15	HDA + MSDC 0.678/0.023	MDC 0.744/0.026	HDA + MSDC 0.958/0.019
25	MLC 0.690/0.017	MDC 0.750/0.018	HDA + MSDC 0.958/0.016
50	MLC 0.702/0.018	MDC 0.766/0.012	HDA + MSDC 0.967/0.002

motivation is less reasonable, the MLC method appears to be preferred.

Similarly, for the TM image (Fig. 7), in the first and second scenarios, HDA+MSDC tends to provide higher kappa values when small training sets (10 and 15 pixels/class) are adopted. Furthermore, independent of the training set size, the HDA+MSDC achieves better accuracy in the third scenario than the other methods. Additionally, MLC provides better results than MDC in the first and third scenarios. Tables 3 and 4 summarize the better classification results for each image, training set size, and scenario.

Figures 8 and 9 show the classification results obtained by the analyzed methods using the training set with 10 pixels per class. Regarding the classification results of the PALSAR image, HDA+MSDC shows better performance in separating FO than the other methods. However, when the second scenario is analyzed, the proposed and MLC methods provide similar results, although MDC shows lower accuracy in classifying LB and AG classes. Focusing on the first classification scenario of the TM image, it can be noted that the HDA+MSDC method is able to better

distinguish OR, FO, and NR classes. In the second scenario, it is possible to observe HDA+MSDC's ability to better discriminate AG and PS classes. Likewise, in the third scenario, better distinguishing between LB and HB is observed.

Regarding the computational run-time, MLC and MDC were less expensive than HDA+MSDC. While the MLC and MDC methods took about 2–3 minutes, the proposed framework spent 10–12 minutes. Such behavior is explained by the two-stage classification of HDA+MSDC, where firstly the input image is "clusterized" by the HDA algorithm, and after this, each identified cluster is compared to the training classes by means of Bhattacharyya stochastic. In order to compute this distance, the mean vector and the covariance matrix that models the multivariate Gaussian distribution of each cluster should be computed, which involves a time-expensive process.

5 Conclusions

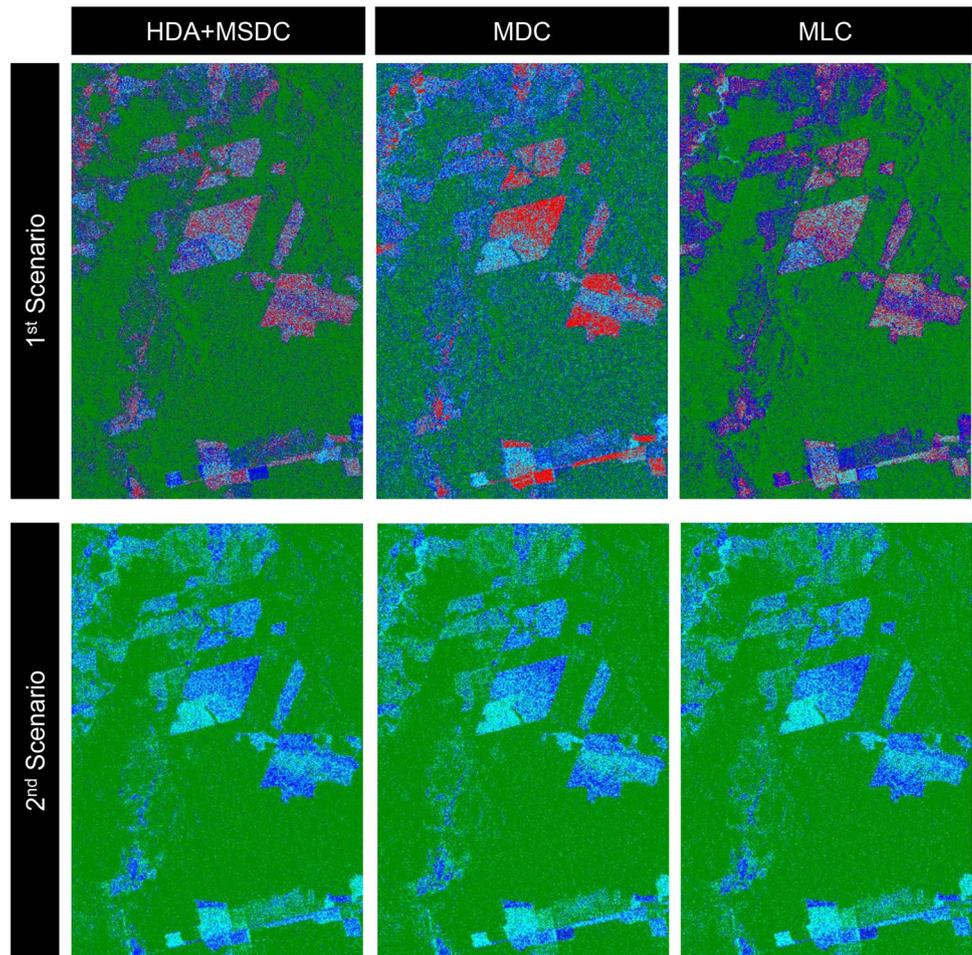
In this paper, an indirect semi-supervised image classification framework, denominated HDA+MSDC, was proposed and compared with the classic MLC and MDC statistical methods. Two case studies about land use and land cover classification in an Amazonian area were conducted using ALOS PALSAR and LANDSAT-5 TM images to compare the analyzed methods. Different classification scenarios were also considered.

The results show that HDA+MSDC has better performance than MLC and MDC, independent of the considered image (i.e., multispectral or SAR) and scenario, when very small training sets (i.e., 10 or 15 labeled pixels/class) are available. Specifically, with respect to the TM image, it was observed that HDA+MSDC achieved results up to 14.6%, 3.2%, and 13.6% more accurate than MLC and MDC in the first, second, and third scenarios, respectively. Regarding the results using PALSAR data, the proposed framework was 3.2% and 3.8% more accurate than the compared methods in reference to the first and second scenarios.

However, it is important to mention that the proposed framework has two parameters that should be adjusted before its use. Consequently, beyond the initial efforts spent tuning the parameters, the possibility of using a sub-optimal parameter configuration that impairs the method performance should also be considered.

An innovative contribution of this work lies in the proposal of an indirect semi-supervised framework based on cluster labeling through stochastic distances. Furthermore, the use of stochastic distances in semi-supervised frameworks was not observed in prior work. Additionally, while

Fig. 8 Classifications for first and second scenarios of ALOS PALSAR image using 10 training pixels/class. Legend of first scenario: ■ PS, ■ BS, ■ AG, and ■ FO—Legend of second scenario: ■ LB, ■ AG, and ■ FO

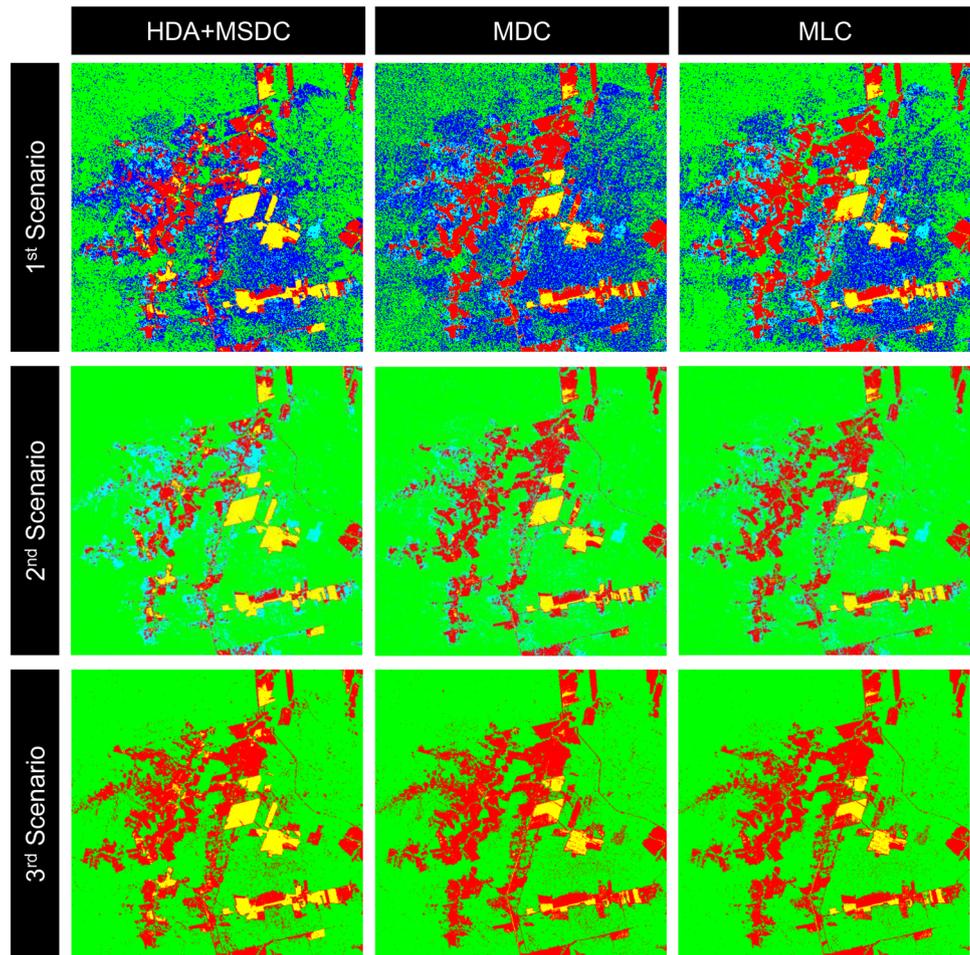


the recent semi-supervised methods come from more complex concepts, the proposed framework has a very simple architecture.

As perspective for future works, the use of different clustering algorithms—for instance, Fuzzy C-Means and Gustafson–Kessel—should be investigated. Additionally, other stochastic distances like the Jeffries–Matusita and Kullback–Leibler distances should also be considered.

Furthermore, the actual parameter tuning process may be carried out using an automatic optimization strategy like Simulated Annealing, with objective function expressed in terms of a clustering assessment index, such as Xie–Beni’s measure. The inclusion of contextual information (i.e., neighborhood behavior) through Markov random field or even filtering techniques may also be considered as a tactic to improve the framework’s accuracy and provide more regularized/smooth results.

Fig. 9 Classifications for first, second, and third scenarios of LANDSAT-5 TM image using 10 training pixels/class. Legend of first scenario: AG, PS, NR, FO, and OR—Legend of second scenario: AG, PS, NR, and HB—Legend of third scenario: AG, LB, and HB



Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

- Banskota A, Kayastha N, Falkowski MJ, Wulder MA, Froese RE, White JC (2014) Forest monitoring using landsat time series data: a review. *Can J Remote Sens* 40(5):362–384. <https://doi.org/10.1080/07038892.2014.987376>
- Brisco B, Brown RJ, Hirose T, McNairn H, Staenz K (1998) Precision agriculture and the role of remote sensing: a review. *Can J Remote Sens* 24(3):315–327. <https://doi.org/10.1080/07038892.1998.10855254>
- Bruzzone L, Persello C (2009) A novel context-sensitive semi-supervised SVM classifier robust to mislabeled training samples. *IEEE Trans Geosci Remote Sens* 47(7):2142–2154. <https://doi.org/10.1109/TGRS.2008.2011983>
- Camps-Valls G, Bandos TVM, Zhou D (2007) Semi-supervised graph-based hyperspectral image classification. *IEEE Trans Geosci Remote Sens* 45:2044–2054
- Chapelle O, Schölkopf B, Zien A (2006) *Semi-supervised learning*. MIT Press, Cambridge
- Chen M, Wang J, Li X, Sun X (2018) Robust semi-supervised manifold learning algorithm for classification. *Math Probl Eng* 2018:8. <https://doi.org/10.1155/2018/2382803>
- Chesnel A, Binet R, Wald L (2007) Object oriented assessment of damage due to natural disaster using very high resolution images. In: *Proceedings of international geoscience and remote sensing symposium*. IEEE, Barcelona, pp 3736–3739
- Congalton RG, Green K (2009) *Assessing the accuracy of remotely sensed data*. CRC Press, Boca Raton
- Everitt BS, Landau S, Leese M, Stahl D (2011) *Cluster analysis*. Wiley, Hoboken. <https://doi.org/10.1002/9780470977811>
- Freitas CC, Soler L, Sant’Anna SJS, Dutra LV, Santos JR, Mura JC, Correia AH (2008) Land use and land cover mapping in the Brazilian Amazon using polarimetric airborne P-band SAR data. *IEEE Trans Geosci Remote Sens* 46(10):2956–2970. <https://doi.org/10.1109/TGRS.2008.2000630>
- Jain AK, Dubes RC (1988) *Algorithms for clustering data*. Prentice-Hall Inc., Upper Saddle River
- Joachims T (1999) Transductive inference for text classification using support vector machines. In: *Proceedings of ICML-99, 16th international conference on machine learning*. Morgan Kaufmann Publishers, San Francisco, US, Bled, SL, pp 200–209
- Kiyasu S, Yamada Y, Miyahara S (2009) Semi-supervised land cover classification of remotely sensed data using two different types of classifiers. In: *Proceedings of international conference on control, automation and systems*. SICE, Fukuoka, pp 4874–4877

14. Liaghat S, Balasundram S (2010) A review: the role of remote sensing in precision agriculture. *Am J Agric Biol Sci* 5:50–55. <https://doi.org/10.3844/ajabssp.2010.50.55>
15. Munoz-Mari J, Bovolo F, Gomez-Chova L, Bruzzone L, Camp-Valls G (2010) Semisupervised one-class support vector machines for classification of remote sensing data. *IEEE Trans Geosci Remote Sens* 48(8):3188–3197. <https://doi.org/10.1109/TGRS.2010.2045764>
16. Nascimento ADC, Horta MM, Frery AC, Cintra RJ (2014) Comparing edge detection methods based on stochastic entropies and distances for polsar imagery. *IEEE J Sel Top Appl Earth Obs Remote Sens* 7(2):648–663. <https://doi.org/10.1109/JSTAR.2013.2266319>
17. Negri RG, Dutra LV, Sant'Anna SJS, Lu D (2016) Examining region-based methods for land cover classification using stochastic distances. *Int J Remote Sens* 37(8):1902–1921. <https://doi.org/10.1080/01431161.2016.1165883>
18. Niedermeier A, Lehner S, Sanden J (2001) Monitoring big river estuaries using SAR images. In: *Proceedings of international geoscience and remote sensing symposium*, vol 4. IEEE, Sydney, pp 1756–1758
19. Nielsen MM (2015) Remote sensing for urban planning and management: the use of window-independent context segmentation to extract urban features in stockholm. *Comput Environ Urban Syst* 52:1–9. <https://doi.org/10.1016/j.compevnurbsys.2015.02.002>
20. Niu X, Ban Y (2012) An adaptive contextual sem algorithm for urban land cover mapping using multitemporal high-resolution polarimetric sar data. *IEEE J Sel Top Appl Earth Obs Remote Sens* 5(4):1129–1139. <https://doi.org/10.1109/JSTARS.2012.2201448>
21. Qi HN, Yang JG, Zhong YW, Deng C (2004) Multi-class svm based remote sensing image classification and its semi-supervised improvement scheme. In: *Proceedings of international conference on machine learning and cybernetics*, vol 5. pp 3146–3151 <https://doi.org/10.1109/ICMLC.2004.1378575>
22. Richards JA, Xiuping J (2006) *Remote sensing digital image analysis: an introduction*, 4th edn. Springer, Berlin
23. Shannon CE (1948) A mathematical theory of communication. *Bell Syst Tech J* 27:379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x#>
24. Shao Z, Zhang L, Zhou X, Ding L (2014) A novel hierarchical semisupervised svm for classification of hyperspectral images. *IEEE Geosci Remote Sens Lett* 11(9):1609–1613. <https://doi.org/10.1109/LGRS.2014.2302034>
25. Silva WB, Freitas CC, Sant'Anna SJS, Frery AC (2013) Classification of segments in polsar imagery by minimum stochastic distances between wishart distributions. *IEEE J Sel Top Appl Earth Obs Remote Sens* 6(3):1263–1273. <https://doi.org/10.1109/JSTAR.2013.2248132>
26. Singh SS, Parida BR (2018) Satellite-based identification of aquaculture farming over coastal areas around bhitarkanika, odisha using a neural network method. In: *Proceedings*, vol 2(7). <https://doi.org/10.3390/ecrs-2-05144>. <http://www.mdpi.com/2504-3900/2/7/331>
27. Theodoridis S, Koutroumbas K (2008) *Pattern recognition*, 4th edn. Academic Press, San Diego
28. Torres L, Sant'Anna SJ, da Freitas CC, Frery AC (2014) Speckle reduction in polarimetric sar imagery with stochastic distances and nonlocal means. *Pattern Recogn* 47(1):141–157. <https://doi.org/10.1016/j.patcog.2013.04.001>
29. Webb AR, Copsey KD (2011) *Statistical pattern recognition*, 3rd edn. Wiley, Hoboken. <https://doi.org/10.1002/9781119952954>
30. Zhu X, Goldberg AB (2009) *Introduction to semi-supervised learning*. Morgan & Claypool Publishers, San Rafael

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.