Earth Observation Semantic Data Mining: Latent Dirichlet Allocation-Based Approach

Reza Mohammadi Asiyabi, *Member, IEEE*, and Mihai Datcu, *Fellow, IEEE*

***Abstract*—Recent advances in remote sensing technology have provided (very) high spatial resolution Earth Observation data with abundant latent semantic information. Conventional data processing algorithms are not capable of extracting the latent semantic information form these data and harness their full potential. As a result, semantic information discovery methods, based on data mining techniques, such as latent Dirichlet allocation and bag of visual words models, can discover the latent information. Despite their crucial rule, there are only a few studies in the field of semantic data mining for remote sensing applications. This article is focused on this shortage. Three different scenarios are used to evaluate the semantic information discovery in various remote sensing applications, including both optical and synthetic aperture radar (SAR) data with different spatial resolutions. In the first scenario, semantic discovery method correlated the semantic perception of the user and machine to correct and enhance the user defined Ground Truth map in very high-resolution RGB data. The potential of the semantic discovery is evaluated for wildfire affected area detection in Sentinel-2 data in the second scenario. Finally, in the third scenario, the semantic discovery method is utilized to detect the misclassifications as well as the patches with ambiguous or multiple semantic labels in a Sentinel-1 SAR patch-based benchmark dataset to enhance the robustness and accuracy of the annotation in the dataset. Our results in these three scenarios demonstrated the capability of the data-mining-based semantic information discovery methods for various remote sensing.**

***Index Terms*—Bag of visual words (BOVW), benchmark dataset, data mining, Earth Observation (EO) semantics, Ground Truth (GT) map, latent Dirichlet allocation (LDA), latent semantic information discovery.**

I. Introduction

EXTRACTING semantic and meaningful information is one of the most important applications of remote sensing that provides the necessary information for various exercises. Several articles have suggested different methods for semantic information discovery and Earth Observation (EO) image classification, based on the spectral information of the multispectral EO images [1]–[4], amplitude and phase information of the backscattered radar signal in synthetic aperture radar (SAR) images [4]–[6], range point cloud information of LIDAR data [7], [8], and other EO data, using many different algorithms.

Despite immense advancements in land cover classification with EO images, semantically meaningful and comprehensive information extraction is still a challenge. Unsupervised classification methods suffer from semantically meaningless classes which can cause many semantic classes mixing in the classified map, as well as not detecting the desired target semantic classes [9]. On the other hand, supervised methods have the advantage of providing semantically meaningful classes and give the full control to the analyst over the classification process [9]. However, supervised methods require Ground Truth (GT) data for training. Creating the GT map is not straightforward and cause several challenges [10], [11]. Collecting GT data in remote locations is costly and laborious. Besides, the quality of the GT map, defined by different users/experts, is variable based on the user's experience, perspective, knowledge, and target application. But the GT map influences the classification results, considerably [10], [11]. Moreover, high-resolution and very-high-resolution EO images contain abundant latent semantic information about the land cover, which might get neglected through conventional information extraction techniques. As a result, robust data mining methods are necessary for latent semantic information discovery in high- and very-high-resolution EO images.

By definition, data mining is "the discovery of interesting, unexpected, or valuable structures in large datasets" [12]. It is an interdisciplinary subfield of statistics, artificial intelligence, and machine learning with an overall goal to discover latent patterns and structures in the dataset for further use. Because of the digitization and massive volume of the available data in various fields, data mining techniques have attracted many interests for different applications [13]. In the field of remote sensing, hundreds of terabytes of EO data are being acquired by various sensors with different characteristics every day. Considering the huge archive of the EO data as well, an efficient approach for handling and processing this data source is important. However, data mining in remote sensing has not been exploited thoroughly and a hug improvement gap exists in this field.

In the present article, two well-known data mining techniques, latent Dirichlet allocation (LDA) and bag of visual words (BOVW) models, are utilized for exploiting different EO data and extracting latent semantic information from them. LDA is a generative probabilistic model which has been proposed by

Manuscript received September 30, 2021; revised January 6, 2022 and February 26, 2022; accepted March 8, 2022. Date of publication March 15, 2022; date of current version April 6, 2022. This work was supported by the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie under Grant 860370. *(Corresponding author: Reza Mohammadi Asiyabi.)*

Reza Mohammadi Asiyabi is with the Center for Spatial Information, University POLITEHNICA of Bucharest, 060042 Bucuresti, Romania (e-mail: reza.mohammadi@upb.ro).

Mihai Datcu is with the German Aerospace Center (DLR), Remote Sensing Technology Institute, EO Data Science, 82234 Oberpfaffenhofen, Weßling, Germany, and also with the Center for Spatial Information, University POLITEHNICA of Bucharest, 060042 Bucuresti, Romania (e-mail: mihai.datcu@dlr.de).

Digital Object Identifier 10.1109/JSTARS.2022.3159277

Blei *et al.* [14] for text modeling. In image domain, LDA models each image as a mixture of latent topics from a Dirichlet distribution. LDA uses BOVW representation of the image for this purpose and considers the visual words as topic representors. LDA represents each image with a topic probability vector through a generative procedure. LDA is completely unsupervised and do not require any reference data, but provides an interpretable intermediate representation of the data. In addition, using LDA fulfill many conditions of the explainable machine learning models, which is getting more attention in the artificial intelligence fields [15].

Several articles have been carried out on semantic analysis of EO images with LDA. For instance, Liénou *et al.* [16] utilized LDA model for semantic annotation of satellite images with the semantic concepts defined by the user. In this article, maximum-likelihood classifier is used to assign each image patch to the semantic concepts based on the LDA topic probabilities. Du *et al.* [17] suggested using LDA for change detection in EO images. In this article, BOVW and LDA are used for mid-level feature dimension reduction and representing two multitemporal images in the topic space. Later, the multivariate alteration detection method is applied to detect the semantic changes between the topic representations of two multitemporal images. Zhong *et al.* [18] used LDA in a multifeature fusion approach to extract semantic information for scene classification and semantic interpretation of high-resolution remote sensing images. Li *et al.* [19] employed LDA to combine different land cover products and achieve a harmonized land cover map. Regionalized class co-occurrences in different land cover maps are used in the LDA to generate the harmonized land cover label for each pixel through statistically characterizing land attributes from the class co-occurrences. In another article, Bahmanyar *et al.* [20] used multimodal LDA for land cover classification with SAR and multispectral satellite images. In this article, the BOVW histograms of the SAR and multispectral image patches are computed and simultaneously fed into the multimodal LDA, which results in a joint latent data model of the land cover topics. Karmakar *et al.* [15] used LDA for discovering the latent structure of SAR data for content classification. The explainability, feature independence, and unsupervision of LDA model is discussed in this article. Furthermore, Espinoza-Molina *et al.* [21] used LDA for vegetation dynamics representation and change detection in EO image time series. In this article, the latent structures of the normalized difference vegetation index (NDVI) patches of the area are extracted and the divergence between the latent structures is considered as a measure of change. In addition, Lienou *et al.* [16] used LDA model for the annotation of large satellite images, using semantic concepts defined by the user. The simple features such as mean and standard deviation for the LDA-image representation are used in the maximum-likelihood method to classify the large unseen satellite images. Furthermore, Karmakar and Datcu [22] utilized BOVW and LDA models for representing EO images in word- and topic-level and applied text-based search algorithms for fast semantic search in EO images.

The main objective of this article is to evaluate the capabilities of the data mining latent semantic analysis methods based on LDA and BOVW models for latent semantic information discovery in EO images with different scenarios (i.e., different EO data, target purposes, data processing algorithms, and applications). Therefore, three different scenarios with three different EO image datasets were employed for semantic information discovery of remote sensing images. As the first scenario, kernel-based BOVW and LDA are applied on very-high-resolution (30 cm) multispectral (three RGB bands) EO image for semantic information discovery and enhancing the user-defined GT map in order to achieve a better and more semantically comprehensive classified map. Later, a similar procedure is employed with coarser spatial resolution (10 m) multispectral (three RGB and one infrared bands) images to detect wildfire affected areas. And finally, in the last scenario, patch-based BOVW and LDA are employed on initially annotated SAR patches to detect misclassifications and errors in the initial classification results. The contributions of this article can be summarized as follows:

1) The present article is not focused on classification, but the main novelty and purpose is to adhere to the lack of semantic data mining researches in remote sensing and to demonstrate the applicability and value of data mining techniques for latent semantic information discovery in different contexts from EO images.
2) Many semantic classes in the EO images are hidden and each user will have a biased interpretation of the scene, based on their perception, semantic understanding, and the target application. However, data mining techniques, such as LDA, make the hidden information evident.
3) The benefits of two well-known data mining techniques, LDA and BOVW models, for latent semantic information discovery from EO images are comprehensively discussed in various circumstances.

In Section II, a brief and basic theoretical background of the models used in this article, BOVW and LDA, are given. The datasets used for experimental verification and the three experimental scenarios with different EO data, algorithms, goals, and target applications, as well as the obtained results and discussions are explained in Section III. Finally, Section IV summarizes the main points and draws a conclusion based on the findings of the present article.

II. Theoretical Background

In this section, the basic theoretical background of the BOVW and LDA models are briefly explained.

A. Bag of Visual Words (BOVW) Model

BOVW is a well-known mid-level representation model which is originally inspired from bag of words model in text mining [23]–[25]. Bag of words model represents each document with a histogram, corresponding to the frequency of each keyword in the document [26], [27]. The resulting histograms can be utilized for various applications, including document categorization and subject detection.

Similarly, BOVW model represents each image with the histogram of occurrences of the visual words in the image [24], [25]. After low-level feature extraction from the EO image,

visual words have to be constructed in the BOVW model. The most frequently used technique for visual words construction is clustering techniques, especially well-known and simple clustering algorithm, K-means [28], [29]. Alternatively, random visual dictionary can be employed for visual words construction [30].

After constructing the visual dictionary, each pixel in the image assigns to the closest visual word in the low-level feature space. Later, a quantization technique represents the image with a histogram, corresponding to the frequency of each visual word in the image.

Usually, BOVW model is applied in the patch level for patch-based categorization [24], [28], [31]. However, kernel-based BOVW can provide pixel-based representation of the image [32]. In the kernel-based BOVW, an arbitrary weighted kernel is used and the BOVW histogram of the area covered by the kernel is assigned to the central pixel in the kernel. After sliding the kernel window over the image and repeating the procedure for all the pixels in the image, each pixel will be represented by a separate BOVW histogram [32]. These histograms can be utilized for categorization and classification of the image.

B. Latent Dirichlet Allocation (LDA) Model

In text mining, LDA considers each document as a composition of different words that determine the topic of that document. Another assumption of LDA is that each topic is represented by several words. With these assumptions, LDA maps the documents to the topics through assigning the words of the document to the topics in a Dirichlet distribution. Particularly, LDA assumes that each document is a distribution of topics and each topic is a distribution of words in Dirichlet distribution [14], [16], [22].

In order to apply LDA in image domain, an analogy between the text and image domain terminologies is necessary. A word in text can be the equivalent of a visual word (i.e., from BOVW model) in the image, which itself corresponds to a segment or a window of pixels. Additionally, a text document can be analogous to the image [16], [22].

LDA assumes that the image I is a mixture of N visual words w_n and starts with randomly assigning the topics to each word and improves this assignment through approximate inference algorithms such as Gibbs sampling [33] and variational expectation maximization [16]. Probability of each topic in each image in the LDA is calculated through the generative process as the following steps:

1) If K is the number of the topics, choose a K-dimensional Dirichlet random variable $\theta_i \sim \text{Dirichlet}(\alpha)$.
2) For each visual word w_n, do the following:
   a) Choose a random topic $z_n \sim \text{Multinomial}(\theta)$.
   b) Choose a word w_n from $p(w_n|z_n, \beta)$, a multinomial probability conditioned on the topic z_n.

And the likelihood of the image with the LDA model is computed as

$$p\left(w|\alpha,\beta\right) = \int p(\theta|\alpha)\left(\prod_{n\,=\,1}^{N}\sum_{z_n} p\left(z_{n|\theta}\right) p(w_n|z_n,\beta)\right) d\theta. \tag{1}$$

In the abovementioned procedure, α and β are Dirichlet distributions that determine the distribution of the topics in the image and visual words in the topic, respectively. The goal of the learning procedure is to find the α and β parameters which maximize the likelihood for all the images in the training dataset.

The reasons behind choosing the LDA model in this article are as follows:

1) LDA is completely unsupervised and do not require any reference data. Supervised methods require semantically annotated reference data which is not available in many practical cases. Besides, the available GT maps are usually biased by the creator's perception, semantic understanding, and target application [34], and will affect the results of the information discovery method. Unsupervised methods help with the unavailability and correctness of the GT maps.
2) LDA provides an interpretable intermediate semantic representation of the data in the form of topic maps, which gives an educated guess and initial information about the study area.
3) Despite the undeniable success of machine learning and deep learning methods, many of these algorithms are usually considered as a black box without any explanation of how the model functions and what are the reasons behind the decisions of the model. The demand for the explainable machine learning models are increasing and some studies claim explainability as a prerequisite for the scientific value of the model's outcomes [15], [35]. It has been showed in [15] that the data mining with LDA satisfies three main aspects of the explainable machine learning as defined in [35], including transparency, interpretability, and explainability.

III. Experimental Scenarios

In this article, three different scenarios have been employed to evaluate the potential of the data mining semantic analysis based on the BOVW and LDA models for latent semantic information discovery from EO images. First, this section describes the datasets used in this article, and later, demonstrates the experimental results in each scenario.

A. Dataset

In order to consider various remote sensing contexts, three different EO datasets with diverse characteristics are employed in this article. The first dataset is a subset of the United States Geological Survey (USGS) very-high-resolution aerial imagery. The aerial image is acquired in RGB bands over San Francisco Bay, USA, in September 2008, with the spatial resolution of 0.3 m [36]. This dataset is used in this article as the very-high-resolution (submeter) RGB EO imagery. Very detailed observations such as individual trees, cars, and buildings, as well as their shadows complicate the processing of this type of data.

Sentinel-2 is a multispectral spaceborne imagery mission as a part of the Copernicus program, which measures the Earth surface with 13 different spectral bands with 10, 20, and 60 m

TABLE I
SAR DATA SPECIFICATIONS

Feature	Chicago Scene	Houston Scene	Sao Paulo Scene
Product ID	S1A_S1_SLC__1SDH_20210518T121201_037942_047A61_DE5A	S1B_S3_SLC__1SDH_20210505T001823_026762_033263_16AD	S1A_S3_SLC__1SDH_20210516T213615_037918_0479A0_E1C4
Acquisition Date	2021/05/18	2021/05/05	2021/05/16
Orbit Pass	Descending	Ascending	Ascending
Beam	S1	S3	S3
Mission	Sentinel-1A	Sentinel-1B	Sentinel-1A
Acquisition Mode	SM	SM	SM
Product	Level-1 SLC	Level-1 SLC	Level-1 SLC
Polarization	HH, HV	HH, HV	HH, HV
Subset size (pixels)	54900 ×21000	46600×18800	46700×18600

SM = StripMap, SLC = single look complex.

spatial resolutions. Moreover, Sentinel-2 data are publicly available and can be accessed free of charge through the Copernicus datahub.[1] In comparison to the first dataset, Sentinel-2 data have coarser spatial resolution and consequently, small targets (e.g., individual trees or cars and their shadow) are not observable, however, more spectral bands provide more information and enable us to extract more semantic details about the land cover of the area. The wildfire in Çınarpınar forest unit, Andırın, Kahramanmaraş in Turkey, which occurred in October 2019 and affected about 47.43 ha Pinus brutia forest [37] is selected as the case study. Three subsets, covering the wildfire area, acquired in July 2019 (before the wildfire), October 2019 (a few days after the wildfire), and April 2020 (several months after the wildfire), are used. Four spectral bands with 10-m resolution (three RGB and one NIR bands) were selected to be used in this article.

SAR systems provide informative EO images with unique characteristics and different nature from the optical data. As a result of the nonvisual nature of the SAR images, visual inspection for semantic interpretation of the data will not provide detailed and accurate information. An SAR dataset is used as the third dataset in this article to evaluate the performance of the semantic data mining techniques for semantic discovery in SAR EO images. Sentinel-1 is an SAR imagery mission as a part of the Copernicus program. Sentinel-1 measures the Earth surface with C-band frequency and up to 5 m × 5 m spatial resolution based on the operation mode [38]. Besides, Sentinel-1 SAR data are also accessible, free of charge, through the Copernicus datahub. In this article, three Sentinel-1 SAR images acquired in high-resolution StripMap (SM) mode over Chicago and Houston in the USA and Sao Paulo in Brazil are used. Table I summarizes the features of the SAR scenes used in this article. Sentinel application platform [39] is used for extracting the intensity information from the single look complex (SLC) SAR images. Later, the SAR scenes are divided into 289 760 nonoverlapping patches of 100 × 100 pixels. The patch size is chosen considering the semantic integrity, and meaningfulness (i.e., based on the semantic classes), and to be in accordance with the other well-known Sentinel-1 SAR datasets such as [40]. No further preprocessing (including multilooking or slant range to ground range conversion) is applied on the SAR patches and the intensity values are used directly. As a result, each 100×100 pixel SAR patch covers approximately 170 m × 430 m or 250 m × 360 m (range × azimuth, based on the acquisition beam of the scene) on the ground [41]. A total of 1274 patches are annotated manually, using the visual inspection via Google Earth, in seven semantic classes including agriculture (AG), forest (FR), high density urban areas (HD), high rise buildings (HR), low density urban areas (LD), industrial regions (IR), and water regions (WR).

[1][Online]. Available: https://scihub.copernicus.eu/

B. Scenario 1: Submeter Resolution Land Cover Mapping

The main objective of the first scenario is to demonstrate the capability of the semantic data mining techniques for identifying the neglected semantic classes and correcting the user-defined GT map in submeter resolution EO data to correlate the semantic perception of the user and the machine. For this purpose, the first dataset, USGS RGB image with 0.3-m spatial resolution is used. In the first step of this experiment, five semantic classes including buildings, roads, vegetation, water, and shadow are visually identified in the subset and the user-defined GT map is created. Kernel-based BOVW histograms of the scene are computed according to the procedure explained in the Section II-A [32]. In the BOVW model, K-means clustering algorithm is used to construct the visual words and the number of the visual words set to 20, considering the size of the image and histograms. Besides, a 15×15 weighted kernel is used to increase the influence of the closer pixels. The constructed histograms are fed into the well-known support vector machine (SVM) for classification and the five-class classified map is produced. The RGB image with the user-defined GT regions for five semantic classes and the resulting five-class classified map are illustrated in Fig. 1.

In the next step, the constructed pixelwise BOVW histograms are fed into the LDA model with different number of the topics (i.e., 5, 8, 10, and 20 topics) and the topic maps are produced. LDA is a probabilistic topic model and the output of the LDA will be a vector of T probability for each pixel or segment, where T is the number of the topics. The topic with the highest probability is considered as the topic of the pixel to produce the topic maps. Fig. 2 illustrates the topic maps with different number of the topics.

A comprehensive comparison between the semantic classes identified by the user in the five-class classification and the LDA topic maps has been carried out to identify the neglected semantic classes. The first topic map is produced with five topics (i.e., the same as the number of the semantic classes in the user-defined GT map) to compare the differences between the semantic perception of the user and the machine. In this topic map, the constructed areas (i.e., buildings and roads) are

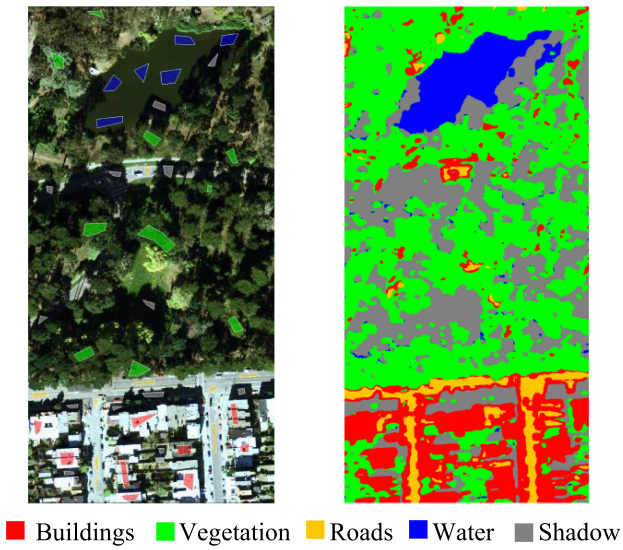


Fig. 1. USGS RGB aerial image with labeled GT regions and classified map with five semantic classes.

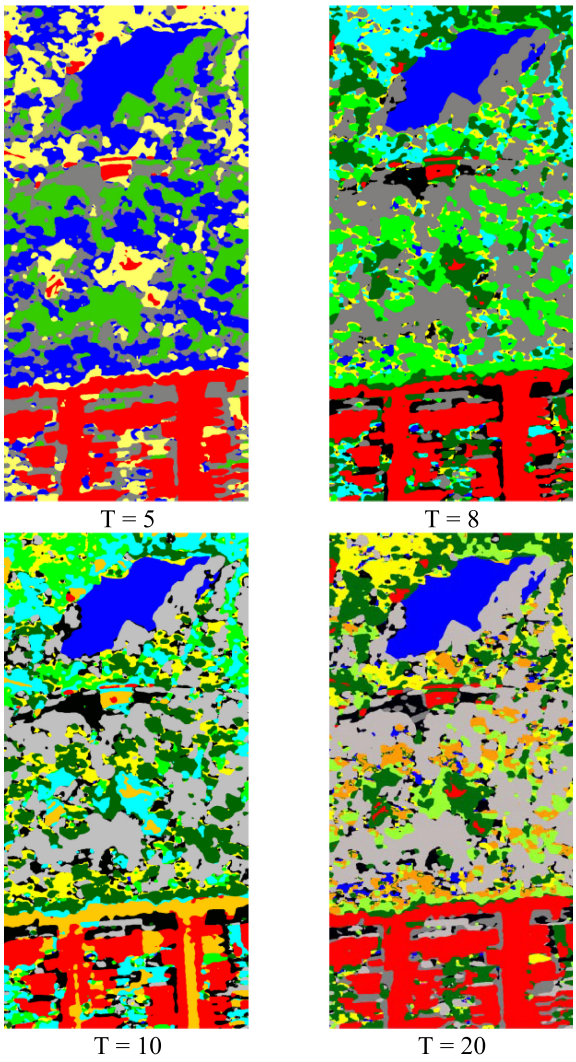


Fig. 2. LDA topic maps with various topic numbers. More latent semantic structures are visible in comparison to the five-class classified map.

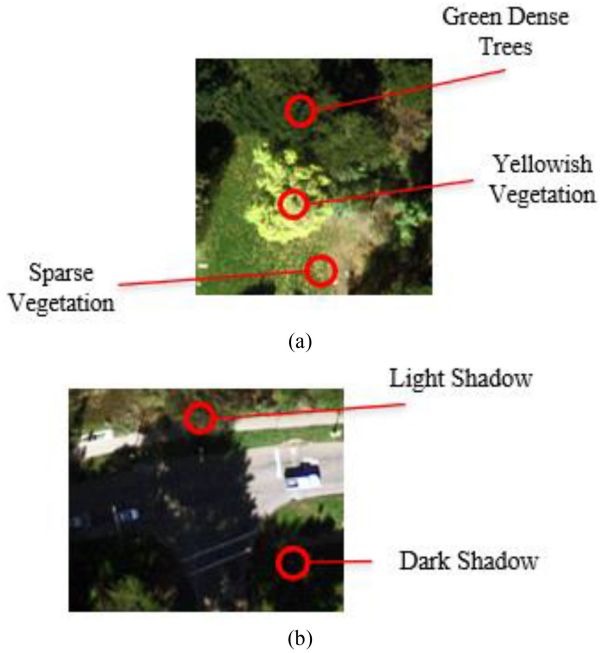


Fig. 3. Examples of the new identified semantic classes through the visual interpretation of the LDA topic maps for (a) vegetation and (b) shadow classes.

represented by one topic. Water and some dark vegetation areas are represented by the same topic and the three remaining topics are for the shadow areas and two different vegetation covers. The huge differences between the user-defined semantic classes and the topics, identified by the LDA model, demonstrate the distinct perception of the user and machine learning methods of the semantic information in the image. As a result, using semantic information discovery methods to correlate these distinct understandings is necessary for robust data processing in various remote sensing applications.

With eight topics, water is almost isolated and is less mixed with dark green vegetations. Constructed areas are still in the same topic, but the shadow areas are separated into two different topics, based on the darkness of their background. When the number of the topics is increased to 10, roads and buildings are separated, but there are many small redundant segments, representing different vegetation covers, which have decreased the smoothness and quality of the topic map. Increasing number of the topics to 20 does not help with the better representation of the scene; besides, it increases the small redundant vegetation segments.

According to the applied visual semantic analysis on the topic maps, the five-class user-defined GT is corrected and an eight-class GT is produced. The resulting GT map is semantically more meaningful and more comprehensive. Fig. 3 illustrates an example of the identified semantic classes through the semantic information discovery. In the corrected GT map, the vegetation class is divided into three classes, including green dense trees, yellowish vegetation, and sparse vegetation. The green dense trees class includes the dense trees with dark green leaves, the yellowish vegetation class includes the trees and other plants with light green leaves, and the sparse vegetation class includes

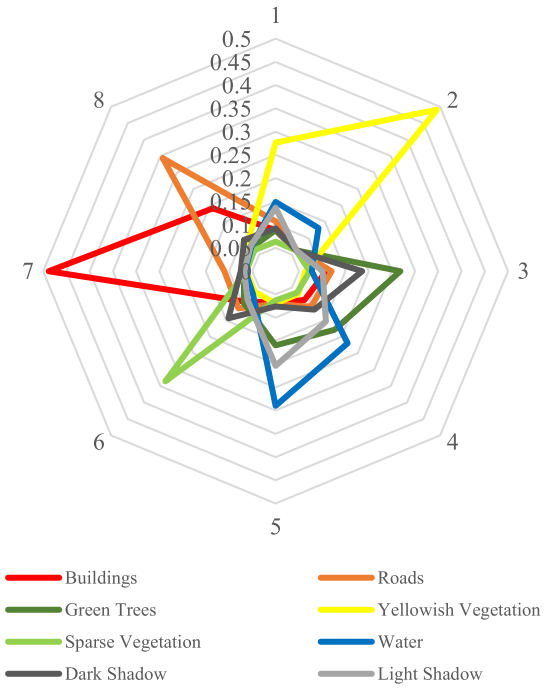


Fig. 4. Radar chart for LDA mean probabilities with eight topics for each semantic class in the corrected GT.

the sparse grass and isolated small bushes and shrubs. Moreover, the shadow class is divided into two dark and light shadow classes. The dark shadow class represents the shadows in closed areas (e.g., between trees or buildings), where there is no other source of light, whereas the light shadow class represents the shadows in open areas (e.g., roads) that are illuminated by diffused and indirect light sources.

In order to demonstrate the different LDA representations for each semantic class, a radar chart is used. Fig. 4 represents the mean probability for each topic in each semantic class for the LDA with eight topics (same as the number of the semantic classes in the corrected GT). In this radar chart, each axis represents a topic and the chart of each semantic class is represented with different color. Fig. 4 manifests the similarities as well as the evident differences between different semantic classes. For instance, the chart of the two shadow classes are very similar because of the similar nature of the classes, but have apparent differences, as the dark shadow class has the highest probability in topic 3, while the light shadow class has the highest probability in topic 5. Some other semantic classes are more distinctive. For example, the yellowish vegetation class has a very high probability in topic 2 and buildings class has a high probability in topic 7.

Finally, the BOVW histograms are used in the SVM classifier with the corrected GT map to produce the eight-class classified map. The new classified map includes the semantic classes of buildings (B), roads (R), green trees (G), yellowish vegetation (Y), sparse vegetation (S), water (W), dark shadow (D), and light shadow (L). Fig. 5 represents the RGB image with the GT regions for eight semantic classes and the eight-class classified map.

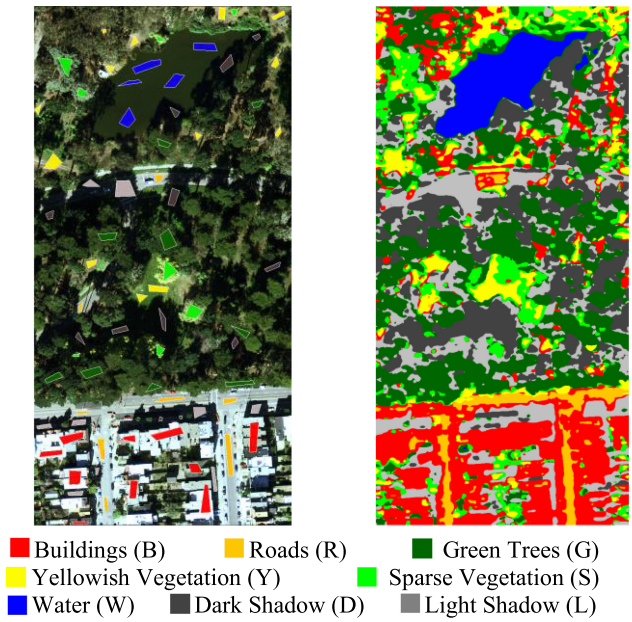


Fig. 5. USGS RGB aerial image with the corrected GT labels and classified map with eight semantic classes.

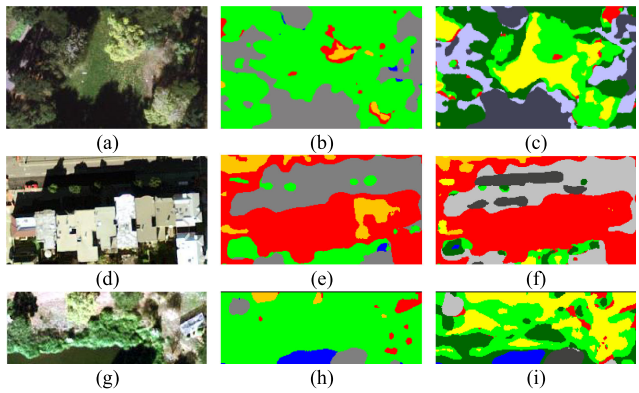


Fig. 6. Example of the enhanced classification results in the eight-class classification with the corrected GT map: (a), (d), and (g) RGB images, (b), (e), and (h) five-class classified maps, (c), (f), and (i) eight-class classified maps.

The classified map with the corrected semantic classes provides more information regarding the various types of vegetation in the scene. Besides, the separation between the dark and light shadow can be utilized for retrieving the information in the shadow covered areas. These results demonstrated the capability of the data mining semantic analysis to correlate the semantic perception of the user and the machine to enhance the data processing algorithms.

Moreover, comparing the five-class and eight-class classified maps, it can be indicated that a more semantically comprehensive GT map not only can help to identify the neglected semantic classes and achieve a semantically more meaningful classified map, but also can enhance the general performance of the classifier and decrease the misclassification. Fig. 6 shows three examples of the five- and eight-class classified maps as well as their corresponding RGB images. For instance, in the first row of the Fig. 6, the eight-class classified map not only detected

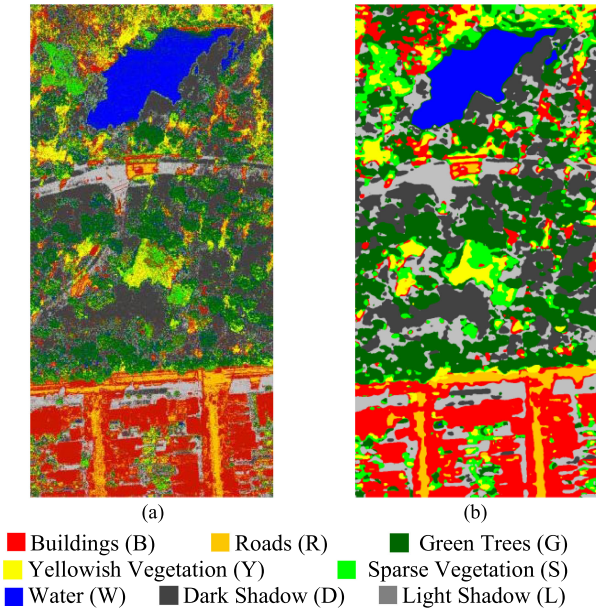


Fig. 7. Eight-class pixelwise classification using SVM classifier (a) without and (b) with the kernel-based BOVW representation.

different types of the vegetation cover, but also decreased the misclassification between the constructed areas and vegetation covered areas. Similarly, in the second row, the eight-class classified map not only separated the shadows with different backgrounds, but also decreased the misclassification between two constructed areas (i.e., buildings and roads).

The main purpose of this article is semantic discovery and detecting the neglected semantic information in the EO data. As a result, well-known classification quantitative measures cannot comprehensively evaluate the performance of the model. For instance, comparing the classification accuracies of the two classified maps is not possible because of the different semantic classes of the classified maps, and the classification accuracy comparison will not evaluate the capability of the model for semantic discovery. However, comparing the overall accuracy and the false-positive rate of the classified maps demonstrate that the semantic data mining has increased the overall accuracy of the classification by about 3% and the false-positive classified samples are decreased by about 2% for each class (average) in the eight-class classified map.

Moreover, comparing the classification algorithms is not the main purpose of this article, but a comparison between the classified maps, with and without using the kernel-based BOVW representation, is carried out. The classification results are illustrated in Fig. 7. Both of these classified maps are pixelwise and have been classified through the same procedure; however, using the kernel-based BOVW representation improved the classification results, noticeably.

As a conclusion, the first scenario demonstrated the huge differences between the semantic understanding of the user and the machine (i.e., data processing algorithms), which necessitates the data mining latent semantic information discovery models

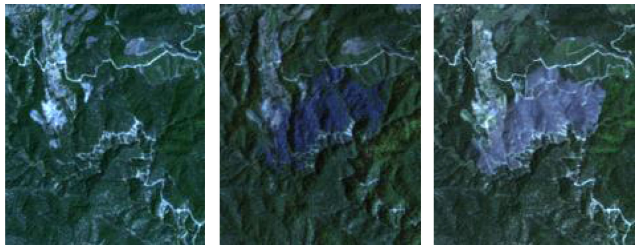


Fig. 8. Sentinel-2 RGB images of the study area in the Çınarpınar forest unit, Andırın, Kahramanmaraş, Turkey (a) before, (b) few days after, and (c) several months after the wildfire.

for remote sensing applications. The main purpose of the first scenario was to evaluate the potential of the semantic analysis for identifying neglected semantic classes and correcting the user-defined GT map in submeter resolution EO data to correlate the semantic perception of the user and the machine. The obtained results demonstrated that the semantic analysis of the EO images can be used for GT map correction, which will enhance the classified map and achieve a semantically more comprehensive classification.

C. Scenario 2: Sentinel-2-Based Forest Fire Monitoring

The main objective of the second scenario is to evaluate the potential of the data mining semantic analysis to detect particular phenomena (e.g., natural disaster) in EO images with coarser spatial resolution and publicly available EO data. As described in Section III-A, Sentinel-2 EO images of the wildfire incident in the Çınarpınar forest unit, Andırın, Kahramanmaraş, in Turkey, in October 2019, are used in this scenario. Fig. 8 illustrates the RGB images of the subsets. Two images for after the incident are chosen to evaluate the possibility of detecting the affected areas a few days after the wildfire (when there are evidences and marks of the incident), as well as the affected area after removing the wildfire marks (when the affected area is transformed to the unvegetated area).

In the first experiment, only three spectral bands in the visible portion of the electromagnetic spectrum (RGB bands) are used. In the first step, the kernel-based BOVW histograms of the subsets are created [32]. The same parameters as the previous scenario are used in the BOVW model. The LDA is trained with all of the histograms from all three scenes and the topic maps with different number of the topics are created. LDA model with five topics resulted in better topic maps and will be used in the rest of the experiments in this scenario.

Fig. 9 demonstrates the LDA topic maps with five topics, using only the RGB spectral bands, for the three scenes. The wildfire area, a few days after the incident [see Fig. 8(b)], is represented with a separate topic [see Fig. 9(b)]. This topic is not present in the topic maps before and several months after the incident and only represents the wildfire affected area. In the topic map of the scene several months after the wildfire, LDA represented the affected areas and some other unvegetated areas by one topic (same topic as the unvegetated areas before the incident).

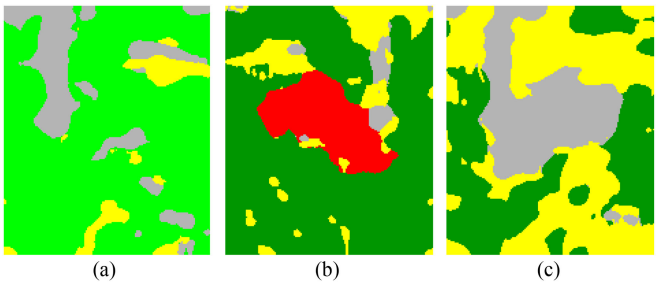


Fig. 9. Topic maps with five topics, using only the RGB spectral bands for (a) before, (b) few days after, and (c) several months after the wildfire.

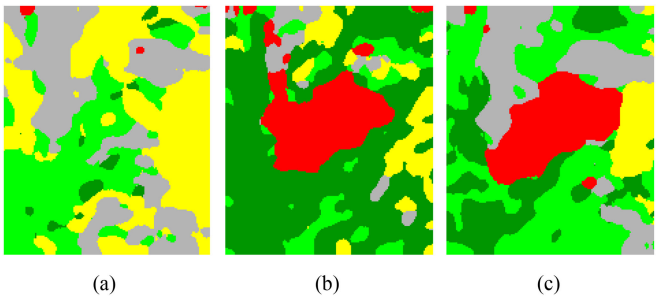


Fig. 10. Topic maps with five topics, using the RGB and NIR spectral bands for (a) before, (b) few days after, and (c) several months after the wildfire.

The wildfire has destroyed the vegetation cover of the affected area and the LDA correctly categorized the affected area into the same topic as the unvegetated areas. However, LDA with only RGB spectral bands was not successful to detect the wildfire affected area several months after the incident from other unvegetated areas. The obtained results show the capability of the semantic analysis for detecting and categorizing the wildfire shortly after the incident, as well as the affected areas. Moreover, the differences in the other areas between the topic maps are due to the different imaging times. Apparently, the vegetation cover varies during the different seasons, i.e., July, October, and April.

In the next experiment, the NIR band of the Sentinel-2 is added to evaluate the effect of the NIR spectral imagery. Fig. 10 represents the topic maps with five topics for this experiment. The LDA model has categorized the affected areas by the wildfire into the same topic for the scenes a few days and several months after the incident [Fig. 10(b) and (c)]. As a conclusion, adding the NIR band has enabled the model to distinguish the wildfire affected area and the unvegetated area even several months after the incident, which is very difficult to detect even through the visual inspection by the user. Besides, the vegetated areas are categorized into more topics in comparison to the previous experiment with only RGB spectral bands, which indicates the effect of the NIR band in the vegetation categorization.

In the last experiment of this scenario, the NDVI is calculated and used alongside the RGB and NIR spectral bands for topic maps preparation. NDVI is a well-known vegetation index which is calculated according to (2), using the red and NIR spectral bands. NDVI is a useful index which measures the greenness and density of the vegetation [37]. Very low values of NDVI (<0.1) represent unvegetated areas such as rocks and bare soil,

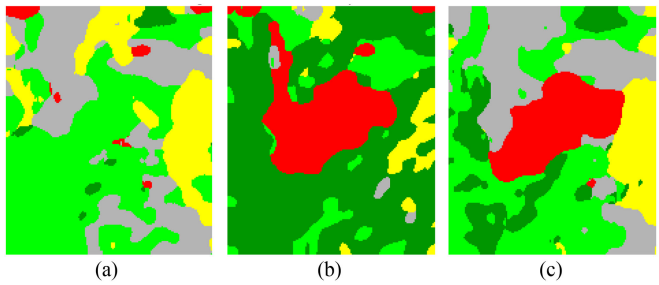


Fig. 11. Topic maps with five topics, using the RGB and NIR spectral bands and NDVI for (a) before, (b) few days after, and (c) several months after the wildfire.

sparse vegetation areas usually have moderate NDVI value ($0.1 < \text{NDVI} < 0.5$), and higher NDVI ($0.5 < \text{NDVI} < 1$) is the indication of higher density green vegetation [37]

$$\text{NDVI} = \frac{(\text{NIR} - \text{Red})}{(\text{NIR} + \text{Red})}. \tag{2}$$

Fig. 11 represents the topic maps with five topics, using RGB and NIR spectral bands as well as the NDVI. Comparing the obtained results in Figs. 10 and 11, adding the NDVI did not help with the better detection of the wildfire in the LDA model, however, slight differences in the categorization of the vegetated area is visible, which is the effect of the NDVI information on vegetation density measurement.

As a conclusion in this scenario, using only the semantic analysis technique can detect the affected areas a few days after the incident, using only the RGB spectral bands. In the image after several months, the semantic analysis technique correctly categorized the affected areas in the same category as the unvegetated areas, since the wildfire eliminated the vegetation cover [visible in the RGB image in Fig. 8(c)]. However, adding the NIR or NDVI bands has enabled the semantic analysis method to detect the wildfire affected area, even in the image several months after the incident. The semantic analysis method has been able to extract latent semantic information which is even difficult to detect by the user via visual inspection, such as wildfire affected area after transforming to the unvegetated regions [Figs. 8(c) and 10(c)]. Therefore, the semantic analysis is useful to discover the relevant spectral bands for better target detection in the remote sensing applications. Additionally, this experiment illustrated that utilizing the relevant spectral bands is important in semantic analysis.

D. Scenario 3: Sentinel-1 Patch-Based Annotation Analysis

SAR systems provide informative EO images with unique characteristics. Due to the different nature of the SAR imagery systems, in many case studies, the conventional machine learning algorithms are unable to achieve comparable results as they obtain with the optical data. Despite the remarkable advances in the SAR data processing techniques, semantic information extraction from SAR data is still a challenge [42]. Moreover, deep learning methods have provided immense opportunities

TABLE II
NUMBER OF THE CLASSIFIED PATCHES IN EACH SEMANTIC CLASS

Semantic Class	Chicago Scene
Agriculture (AG)	61,980
Forest (FR)	112,192
High Density Urban Area (HD)	8,875
High Rise Buildings (HR)	1,745
Low Density Urban Area (LD)	42,049
Industrial Regions (IR)	20,713
Water Regions (WR)	42,206
Total Number of the Patches	**289,760**

for various computer vision applications, and due to their remarkable performance they have attracted wide attention in the remote sensing data processing community [43], [44]. However, due to the various challenges, the huge potential of deep learning methods for SAR data processing remains unexploited [44]. One of the main challenges in this field is the scarcity of semantically annotated SAR data, whereas deep learning methods require high number of the annotated training data [44].

The main purpose of the third scenario is to utilize the data mining semantic analysis for enhancing the annotation quality of the SAR benchmark datasets. As a result, the performance of the semantic analysis for enhancing the semantic categorization of SAR image patches is evaluated in this scenario. Three SAR scenes acquired by Sentinel-1 (Section III-A) are used in this article. In the first step, the mean and standard deviation of the Gabor features in six directions ($1/6\pi$, $1/3\pi$, $1/2\pi$, $2/3\pi$, $5/6\pi$, and $1/\pi$) and four filter sizes (3×3, 7×7, 11×11, and 19×19) are computed for each patch and used as the low-level features [45]. SVM classifier, with histogram intersection kernel, is utilized to classify the SAR patches into seven semantic classes using the Gabor features. Histogram intersection kernel is selected through trial-and-error between different well-known kernels for SVM classifier. Fig. 12 represents the HH polarization SAR images and initially classified scenes. The classification overall accuracy reaches over 80% and the number of the patches classified in each of the semantic classes are shown in Table II.

Several obvious classification errors are noticeable in the classified patches. In the next step, BOVW- and LDA-based semantic analysis is used in this article to identify and remove the classification errors and enhance the annotation of the patches. For this purpose, patch-based BOVW histograms are constructed for each of the patches. Similar to the previous scenarios, 20 visual words are constructed by k-means algorithm. However, due to the patch-based processing in this scenario, patch-based BOVW is used, instead of the kernel-based model (Section II-A and [24], [28], [31], [32]). The BOVW histogram of the training patches, which are used in the initial training stage of the SVM classifier (1274 patches), is used to train the LDA model with seven topics. Number of the topics is chosen with several trials and comparing the perplexity and coherence of the topic models, as well as

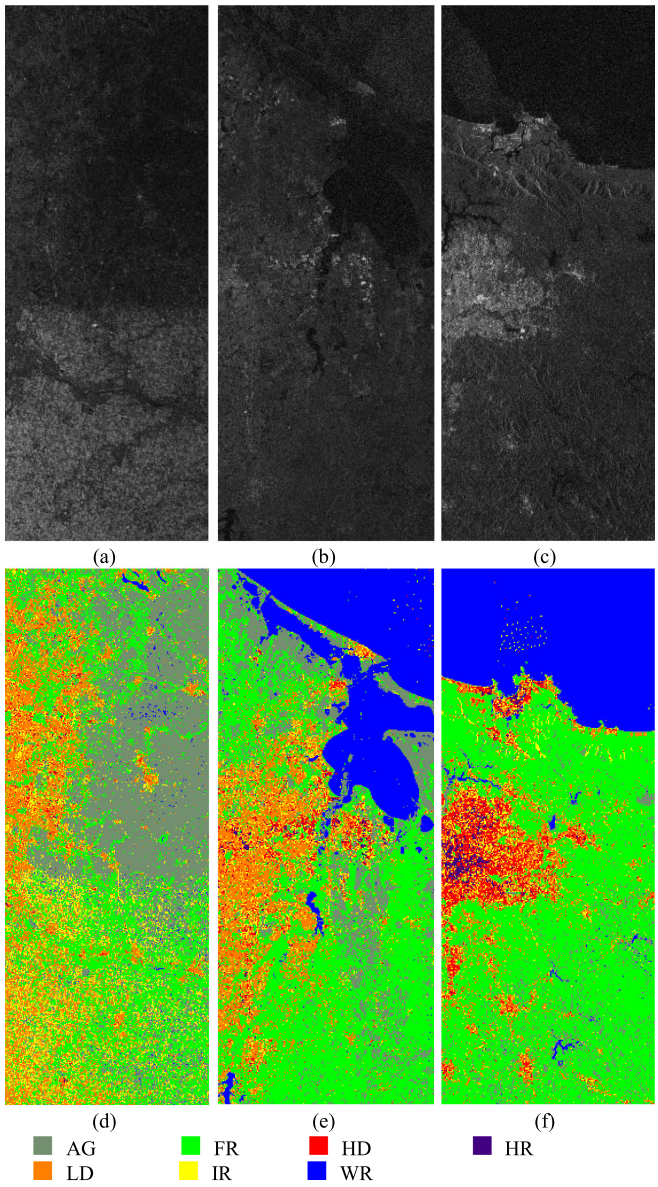


Fig. 12. HH Polarization SAR images and initially classified maps with SVM classifier and Gabor features for (a) and (d) Chicago, (b) and (e) Houston, and (c) and (f) Sao Paulo scenes.

considering number of the semantic labels in annotation of the SAR patches. Table III demonstrates the number and percentage of the patches in each semantic class for each LDA topic. Zero patches are categorized in topics 3 and 4, which make them redundant, however, reducing the number of the topics did not remove redundant topics (e.g., there were still two redundant topics with five topics) and increased the perplexity of the topic model. Furthermore, seven semantic classes were defined in the initial annotation of the dataset. As a result, seven topics have been chosen to carry out the experiments in this scenario.

Correct and accurate semantic class labels of the SAR patches are not available, as a result, several patches of each semantic class in each topic is inspected visually, as well as comparing with the Google Earth image of the area. For this purpose, several random patches from each semantic class and in each LDA topic

TABLE III
NUMBER AND PERCENTAGE OF THE CLASSIFIED PATCHES IN EACH SEMANTIC CLASS

	1	2	3	4	5	6	7
AG	2653 4.3%	23146 37.3%	0 0%	0 0%	8056 13%	9067 14.6%	19058 30.7%
FR	423 0.4%	2667 2.4%	0 0%	0 0%	19369 17.3%	121 0.1%	89612 79.9%
HD	43 0.5%	170 1.9%	0 0%	0 0%	7884 88.8%	18 0.2%	760 8.6%
HR	7 0.4%	10 0.6%	0 0%	0 0%	1670 95.7%	2 0.1%	56 3.2%
LD	192 0.9%	989 4.8%	0 0%	0 0%	15638 75.5%	51 0.2%	3843 18.6%
IR	31 0.1%	1225 2.9%	0 0%	0 0%	21545 51.1%	46 0.1%	19293 45.8%
WR	41325 97.9%	100 0.2%	0 0%	0 0%	282 0.7%	397 0.9%	102 0.2%

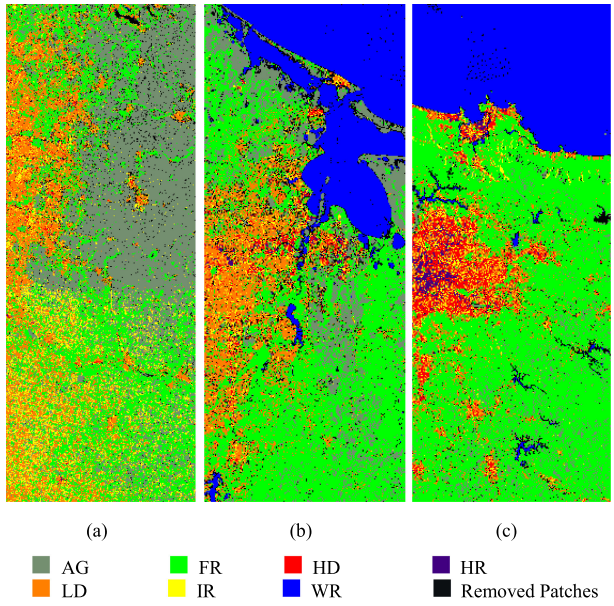


Fig. 13. Classified maps with SVM classifier and Gabor features after the semantic analysis which removed the misclassified patches and patches with ambiguous and mixed semantic labels for (a) Chicago, (b) Houston, and (c) Sao Paulo scenes.

are reviewed (e.g., at least 20 patch per topic for the topics with more than 20 patches) to identify the topics with correct semantic label in each class. The patches categorized in the topics with incorrect semantic label will be removed from the dataset. Table IV illustrates HH (above) and HV (below) polarizations of two typical example patches for each of the topics in each semantic class. The topics that considered incorrect in each class and are removed from the dataset are shown with red color in Table IV, and the other topics with correct semantic labels are shown with green color.

For instance, topic 1 for almost all of the classes represents the patches with water regions that are mixed with land, harbor, or ships. As a result, the patches categorized in the topic 1 will be removed for all of the classes, except for the WR class. The other topics in the WR class represent the patches containing mostly agriculture, bare soil, and some other mixed patches that will be removed. For the AG class, after removing the patches from the first topic, the remaining topics mostly represent agricultural fields with some variations in the crop, density, etc. that should be maintained in the AG class. Topic numbers 2 and 6 in the FR class consist of mixed patches, mostly with agricultural field, and should be removed, but the patches in the topics 5 and 7 will remain in the FR class.

The semantic labels from the constructed areas are the most complex patches with the most frequent errors. The only topic with correct semantic labels for the HD patches is topic 5. The other topics in this class are mixed patched, mostly with low density and industrial areas. Similarly, in the HR class, the patches categorized in topics 2 and 5 will remain and the other topics will be removed as they include several patches from industrial fields as well as other urban areas. LD class includes higher number of the patches in the SAR scenes in this article and after this visual semantic analysis, the patches represented by the topics 5 and 7 remain in this semantic class. The patches from the IR class, which are categorized in the topic numbers 2 and 5 are most correctly labeled patches of this class and will remain with this semantic label, whereas the other patches of this class are mostly representing mixed classes such as different vegetation covers, roads, and farm houses in the agricultural fields which should be removed.

As a result of the abovementioned procedure, number of the patches will be reduced and the size of the dataset will be decreased by about 9.7%. However, this procedure enhances the dataset in terms of less incorrectly classified patches, as well as less mixed patches consisting of multiple semantic classes. Fig. 13 represents the patches in the scenes where the removed patches are represented by the black color. Apparently, the removed patches mostly consist of the mixed and ambiguous patches such as water patches with ships, shallow water areas with the effect of the sediment or above water surface vegetation, beach and harbor, edge patches, roads, and other mixed patches with ambiguous semantic label, as well as the misclassified patches as a result of the confusion between water surface and flat agricultural fields, and between various constructed areas.

After removing the patches categorized in wrong LDA topics, another visual review has been carried out among the remaining patches. The results illustrated that there are still misclassified patches, however, the errors in the classification, as demonstrated in Fig. 13, decreased dramatically and many mixed patches with ambiguous or multiple semantic labels are removed.

The accuracy and false-positive rate of the classification is computed for before and after the semantic analysis using the manually annotated patches. The overall accuracy is

TABLE IV
EXAMPLE PATCHES FOR EACH TOPIC FROM EACH SEMANTIC CLASS

In each example, the above image is the HH and the below image is the HV polarization.
Green and red color represents the topics that remained or removed, respectively, in each class after the semantic analysis.

improved by about 2%. The semantic class classification accuracy improvement was more prominent for water class as many mixed patches are removed from this class. However, the agriculture class was the only class that the classification accuracy is decreased (about 1%) after the semantic analysis. This is because of the high diversity of the agricultural patches (different crops, different crop growth level, and different sowing direction). Moreover, the average false-positive rate is also decreased by about 2% for each class after the semantic analysis.

Several articles have been dedicated to develop annotated benchmark dataset for machine learning and deep learning models in remote sensing and many of them have faced the problem of mixed patches and patches with ambiguous semantic labels [40], [46]–[48]. The explained semantic information discovery method can help to enhance the developed benchmark remote sensing datasets and resolve the abovementioned challenges.

IV. Conclusion

Discovering latent semantic information from (very) high-resolution EO images is necessary to harness the capabilities of the advanced remote sensing technologies. There is a huge improvement area in the field of semantic data mining for various remote sensing applications in the literature. Data-mining-based semantic information discovery techniques including LDA and BOVW models are employed in this article with different remote sensing datasets to extract the latent semantic information from EO images for various applications.

Utilizing kernel-based BOVW representation and LDA topic model has enabled us to correct and enhance the user-defined GT map and identify the neglected semantic classes in very-high-resolution (0.3 m) USGS aerial imagery with RGB optical bands. The corrected GT map resulted in a more semantically meaningful and comprehensive classified map as well as less misclassification errors.

Additionally, RGB and NIR spectral bands of the Sentinel-2 optical imagery with coarser spatial resolution (10 m) are used in the semantic information discovery technique to detect the affected areas in the wildfire, a few days and several months after the incident. The results demonstrated the capability of the semantic discovery method to detect various phenomena (e.g., wildfire affected area) in EO optical imagery. Additionally, this scenario demonstrated the capability of the data mining semantic analysis to detect the relevant spectral bands with more informative data for the target application.

Furthermore, three scenes from the Sentinel-1 SAR imagery with SM imaging mode are divided into 100×100-pixel patches and a few patches are annotated manually into seven semantic classes via visual inspection and Google Earth images. Gabor texture features and the well-known SVM classifier are used for the initial annotation of the patches. Later, the data mining semantic information discovery method is utilized to clean the annotated dataset and remove the misclassified patches, as well as removing the patches with ambiguous or multiple semantic labels.

The conducted experiments in this article demonstrated the following remarks:

1) The main objective of this article is focused on the semantic data mining for various remote sensing applications and to demonstrate the latent semantic information discovery method with various EO images in different remote sensing case studies.
2) Conventional machine learning methods are not capable of extracting the latent semantic information from advanced remote sensing imagery systems with abundant semantic information.
3) There is a huge difference between the semantic perception of the user and the machine, and data mining semantic analysis can be used to correlate these semantic understandings to enhance the classification results.
4) Well-known data mining techniques are capable of extracting meaningful semantic information from various EO images and enhancing the results of the remote sensing practices in different contexts.
5) Land cover classification in very-high-resolution EO images can be semantically enhanced in terms of a more comprehensive classified map and less classification errors through semantic discovery methods.
6) Different natural phenomena, including wildfire affected areas, can be detected accurately and practically in EO images through semantic analysis methods, even in the situations that are difficult to detect through visual inspection by the user.
7) Data mining latent semantic information discovery techniques are capable of enhancing the annotated benchmark datasets by means of detecting the mixed patches with ambiguous semantic labels and the misclassified patches in an annotated dataset.

The experiments carried out in this article demonstrated the immense capabilities of semantic data mining techniques for various remote sensing contexts. However, more inspections are necessary to evaluate the competence of the semantic analysis methods in future articles. The performance of the models should be investigated with larger scale datasets to evaluate the generalizability of the method for generating high-quality training EO data for machine learning algorithms. Additionally, more reliable quantitative evaluations are necessary to assess the effect of the semantic data mining on the classification results. Experiments with well-known EO classification datasets with reliable GT maps for this purpose should be carried out in the future articles.

Acknowledgment

The authors would like to thank USGS and European Space Agency for making the data, used in this article, publicly available.

References

[1] F. Xu and B. Somers, "Unmixing-based Sentinel-2 downscaling for urban land cover mapping," *ISPRS J. Photogramm. Remote Sens.*, vol. 171, pp. 133–154, 2021.

[2] T. Zhang, J. Su, Z. Xu, Y. Luo, and J. Li, "Sentinel-2 satellite imagery for urban land cover classification by optimized random forest classifier," *Appl. Sci.*, vol. 11, no. 2, 2021, Art. no. 543.

[3] C. Liu, R. Tao, W. Li, M. Zhang, W. Sun, and Q. Du, "Joint classification of hyperspectral and multispectral images for mapping coastal wetlands," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 982–996, 2020.

[4] A. Ghorbanian, S. Zaghian, R. M. Asiyabi, M. Amani, A. Mohammadzadeh, and S. Jamali, "Mangrove ecosystem mapping using sentinel-1 and sentinel-2 satellite images and random forest algorithm in Google Earth engine," *Remote Sens.*, vol. 13, no. 13, 2021, Art. no. 2565.

[5] S. Liu, Z. Zhou, H. Ding, Y. Zhong, and Q. Shi, "Crop mapping using sentinel full-year dual-polarized SAR data and a CPU-optimized convolutional neural network with two sampling strategies," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 7017–7031, 2021, doi: 10.1109/JSTARS.2021.3094973.

[6] A. G. Mullissa, C. Persello, and A. Stein, "PolSARNet: A deep fully convolutional network for polarimetric SAR image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 12, pp. 5300–5309, Dec. 2019.

[7] N. Li, O. Kähler, and N. Pfeifer, "A comparison of deep learning methods for airborne lidar point clouds classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 6467–6486, 2021, doi: 10.1109/JSTARS.2021.3091389.

[8] W. Y. Yan, A. Shaker, and N. El-Ashmawy, "Urban land cover classification using airborne LiDAR data: A review," *Remote Sens. Environ.*, vol. 158, pp. 295–310, 2015.

[9] D. I. M. Enderle and R. C. Weih Jr, "Integrating supervised and unsupervised classification methods to develop a more accurate land cover classification," *J. Arkansas Acad. Sci.*, vol. 59, no. 1, pp. 65–73, 2005.

[10] A. Baraldi, L. Bruzzone, and P. Blonda, "Quality assessment of classification and cluster maps without Ground Truth knowledge," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 857–873, Apr. 2005.

[11] M. K. Reif *et al.*, "Ground Truth sampling to support remote sensing research and development submersed aquatic vegetation species discrimination using an airborne hyperspectral/lidar system," Dredging Oper. Environ. Res. Progr., USA, Eng. Res. Develop. Center, USA, Series ERDC, Tech. Note TN-DOER-E30, 2012. [Online]. Available: http://hdl.handle.net/11681/8627

[12] D. J. Hand, "Principles of data mining," *Drug Saf.*, vol. 30, no. 7, pp. 621–622, 2007.

[13] M. K. Keleş, "An overview: The impact of data mining applications on various sectors," *Tehnički Glasnik*, vol. 11, no. 3, pp. 128–132, 2017.

[14] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, 2003.

[15] C. Karmakar, C. O. Dumitru, G. Schwarz, and M. Datcu, "Feature-free explainable data mining in SAR images using latent Dirichlet allocation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 676–689, 2020, doi: 10.1109/JSTARS.2020.3039012.

[16] M. Lienou, H. Maitre, and M. Datcu, "Semantic annotation of satellite images using latent Dirichlet allocation," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 1, pp. 28–32, Jan. 2010.

[17] B. Du, Y. Wang, C. Wu, and L. Zhang, "Unsupervised scene change detection via latent Dirichlet allocation and multivariate alteration detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 12, pp. 4676–4689, Dec. 2018.

[18] Y. Zhong, Q. Zhu, and L. Zhang, "Scene classification based on the multifeature fusion probabilistic topic model for high spatial resolution remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 11, pp. 6207–6222, Nov. 2015.

[19] Z. Li, J. C. White, M. A. Wulder, T. Hermosilla, A. M. Davidson, and A. J. Comber, "Land cover harmonization using latent Dirichlet allocation," *Int. J. Geogr. Inf. Sci.*, vol. 35, no. 2, pp. 348–374, 2021.

[20] R. Bahmanyar, D. Espinoza-Molina, and M. Datcu, "Multisensor earth observation image classification based on a multimodal latent Dirichlet allocation model," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 3, pp. 459–463, Mar. 2018.

[21] D. Espinoza-Molina, R. Bahmanyar, M. Datcu, R. Díaz-Delgado, and J. Bustamante, "Land-cover evolution class analysis in image time series of landsat and sentinel-2 based on latent Dirichlet allocation," in *Proc. 9th Int. Workshop Anal. Multitemporal Remote Sens. Images*, 2017, pp. 1–4.

[22] C. Karmakar and M. Datcu, "A fast search system for remote sensing imagery based on bag of visual words and latent Dirichlet allocation," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2020, pp. 6918–6921.

[23] J. Feng, L. C. Jiao, X. Zhang, and D. Yang, "Bag-of-visual-words based on clonal selection algorithm for SAR image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 4, pp. 691–695, Jul. 2011.

[24] X. Li, L. Zhang, L. Wang, and X. Wan, "Effects of BOW model with affinity propagation and spatial pyramid matching on polarimetric SAR image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 7, pp. 3314–3322, Jul. 2017.

[25] R. M. Asiyabi, M. R. Sahebi, and A. Ghorbanian, "Segment-based bag of visual words model for urban land cover mapping using polarimetric SAR data," *Adv. Space Res.*, 2021, doi: 10.1016/j.asr.2021.10.042.

[26] R. Mohammadi, M. R. Sahebi, M. Omati, and M. Vahidi, "Synthetic aperture radar remote sensing classification using the bag of visual words model to land cover studies," *Int. J. Geol. Environ. Eng.*, vol. 12, no. 9, pp. 588–591, 2018.

[27] C. He, X. Liu, C. Kang, D. Chen, and M. Liao, "Attribute learning for SAR image classification," *ISPRS Int. J. Geo-Inf.*, vol. 6, no. 4, 2017, Art. no. 111.

[28] J. Feng, L. C. Jiao, X. Zhang, and R. Niu, "An effective bag-of-visual-words framework for SAR image classification," in *Proc. Remote Sens. Image Process., Geogr. Inf. Syst., Other Appl.*, 2011, Art. no. 800606.

[29] M. Amrani, S. Chaib, I. Omara, and F. Jiang, "Bag-of-visual-words based feature extraction for SAR target classification," in *Proc. 9th Int. Conf. Digit. Image Process.*, 2017, Art. no. 104201J.

[30] S. Cui, G. Schwarz, and M. Datcu, "Remote sensing image classification: No features, no clustering," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 11, pp. 5158–5170, Nov. 2015.

[31] R. Bahmanyar, S. Cui, and M. Datcu, "A comparative study of bag-of-words and bag-of-topics models of EO image patches," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 6, pp. 1357–1361, Jun. 2015.

[32] R. M. Asiyabi and M. Datcu, "Earth observation image semantics: Latent Dirichlet allocation based information discovery," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2021, pp. 2620–2623.

[33] T. L. Griffiths and M. Steyvers, "A probabilistic approach to semantic representation," in *Proc. 24th Annu. Conf. Cogn. Sci. Soc.*, vol. 24, no. 24, 2002.

[34] R. Bahmanyar, A. M. M. de Oca, and M. Datcu, "The semantic gap: An exploration of user and computer perspectives in earth observation images," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 10, pp. 2046–2050, Oct. 2015.

[35] R. Roscher, B. Bohn, M. F. Duarte, and J. Garcke, "Explainable machine learning for scientific insights and discoveries," *IEEE Access*, vol. 8, pp. 42200–42216, 2020.

[36] "USGS high resolution orthiomagery (Entity ID:1289214_10SEG445790)," U.S. Geological Survey. 2008. Accessed Sep. 30, 2020. [Online]. Available: https://earthexplorer.usgs.gov/

[37] T. Dindaroglu, E. Babur, T. Yakupoglu, J. Rodrigo-Comino, and A. Cerdà, "Evaluation of geomorphometric characteristics and soil properties after a wildfire using sentinel-2 MSI imagery for future fire-safe forest," *Fire Saf. J.*, vol. 122, 2021, Art. no. 103318.

[38] D. Geudtner, R. Torres, P. Snoeij, M. Davidson, and B. Rommen, "Sentinel-1 system capabilities and applications," in *Proc. IEEE Geosci. Remote Sens. Symp.*, 2014, pp. 1457–1460.

[39] "SNAP—ESA Sentinel application platform v8.0," Oct. 2020. [Online]. Available: http://step.esa.int

[40] J. Zhao, Z. Zhang, W. Yao, M. Datcu, H. Xiong, and W. Yu, "OpenSARUrban: A sentinel-1 SAR image dataset for urban interpretation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 187–203, 2020, doi: 10.1109/JSTARS.2019.2954850.

[41] "Stripmap SLC—Sentinel-1 SAR technical guide—Sentinel Online," Accessed: Feb. 24, 2022. [Online]. Available: https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-1-sar/products-algorithms/level-1/single-look-complex/stripmap

[42] A. Mehra, N. Jain, and H. S. Srivastava, "A novel approach to use semantic segmentation based deep learning networks to classify multi-temporal SAR data," *Geocarto Int.*, vol. 37, pp. 163–178, 2020.

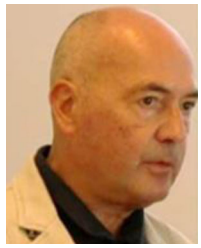
[43] G. Cheng, X. Xie, J. Han, L. Guo, and G.-S. Xia, "Remote sensing image scene classification meets deep learning: Challenges, methods, benchmarks, and opportunities," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 3735–3756, 2020, doi: 10.1109/JSTARS.2020.3005403.

[44] X. X. Zhu *et al.*, "Deep Learning Meets SAR: Concepts, models, pitfalls, and perspectives," *IEEE Geosci. Rem. Sens. Mag.*, vol. 9, no. 4, pp. 143–172, Dec. 2021, doi: 10.1109/MGRS.2020.3046356.

[45] J.-K. Kamarainen, V. Kyrki, and H. Kalviainen, "Invariance properties of Gabor filter-based features—Overview and applications," *IEEE Trans. Image Process.*, vol. 15, no. 5, pp. 1088–1099, May 2006.
[46] G.-S. Xia *et al.*, "DOTA: A large-scale dataset for object detection in aerial images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 3974–3983.
[47] C. O. Dumitru, G. Schwarz, and M. Datcu, "SAR image land cover datasets for classification benchmarking of temporal changes," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 5, pp. 1571–1592, May 2018.
[48] Y. Wang, C. Wang, H. Zhang, Y. Dong, and S. Wei, "A SAR dataset of ship detection for deep learning under complex backgrounds," *Remote Sens.*, vol. 11, no. 7, 2019, Art. no. 765.

Reza Mohammadi Asiyabi (Member, IEEE) received the M.Sc. degree in geodesy and geomatics engineering remote sensing from K. N. Toosi University of Technology, Tehran, Iran, in 2018 with a thesis on enhancement of SAR data classification using Bag of Visual Words representation model. He is currently working toward the Ph.D. degree in the field of uncertainty aware and complex-valued deep learning models for SAR data classification in presence of adversarial samples, at the University POLITEHNICA of Bucharest (UPB), Bucharest, Romania.

He is currently an Early Stage Researcher within the EU Marie Skłodowska-Curie Innovative Training Network (ITN), MENELAOS-NT and a Ph.D. researcher at the Research Center for Spatial Information (CEOSpaceTech) of the Faculty of Electronics Engineering, Telecommunications, and Information Technology, UPB. His research interests include semantic information discovery and classification of SAR data, using data mining techniques and complex-valued deep learning models with a focus on complex-valued and uncertainty aware deep networks.



Mihai Datcu (Fellow, IEEE) received the M.S. and Ph.D. degrees in electronics and telecommunications from the University POLITEHNICA of Bucharest (UPB), Bucharest, Romania, in 1978 and 1986, respectively, and the Habilitation a Diriger Des Recherches degree in computer science from the University Louis Pasteur, Strasbourg, France, in 1999.

Since 1981, he has been a Professor with the Department of Applied Electronics and Information Engineering, Faculty of Electronics, Telecommunications and Information Technology, UPB. Since 1993, he has been a Scientist with the German Aerospace Center (DLR), Weßling, Germany. He has held a Visiting Professor appointments with the University of Oviedo, Oviedo, Spain, the University Louis Pasteur and the International Space University, both in Strasbourg, France, the University of Siegen, Siegen, Germany, the University of Innsbruck, Innsbruck, Austria, University of Alcala, Alcala, Spain, the University Tor Vergata, Rome, Italy, University of Trento, Italy, Unicamp, Campinas, Brazil, China Academy of Science, Shenyang, China, Universidad Pontificia de Salamanca, campus de Madrid, Spain, University of Camerino, Camerino, Italy, and the Swiss Center for Scientific Computing, Manno, Switzerland. From 1992 to 2002, he had an Invited Professor Assignment with the Swiss Federal Institute of Technology (ETH Zurich), Zurich, Switzerland. Since 2001, he has been initiating and leading the Competence Center on Information Extraction and Image Understanding for Earth Observation, Paris Institute of Technology, ParisTech, Paris, France, a collaboration of DLR with the French Space Agency (CNES). He has been a Professor holder of the DLRCNES Chair with ParisTech. He has initiated the European frame of projects for image information mining and is involved in research programs for information extraction, data mining and knowledge discovery, and data science with the European Space Agency (ESA), NASA, and in a variety of national and European projects. He is the Director of the Research Center for Spatial Information, UPB. He is a Senior Scientist and the Data Intelligence and Knowledge Discovery Research Group Leader with the Remote Sensing Technology Institute, DLR, Köln, Germany, and delegate in the DLR-ONERA Joint Virtual Center for artificial intelligence (AI) in Aerospace. His research interests include explainable and physics aware AI, smart radar sensors design, and quantum machine learning with applications in Earth Observation.

Prof. Datcu is a member of the ESA Working Group Big Data from Space and Visiting Professor withe ESA's Lab. He was the recipient of the National Order of Merit with the rank of Knight for outstanding international research results, awarded by the President of Romania, in 2008, and the Romanian Academy Prize Traian Vuia for the development of the system for automated analysis of digital images analysis system and his activity in image processing in 1987 and of the Chaire d'excellence internationale Blaise Pascal 2017 for international recognition in the field of data science in Earth observation. He was a Coorganizer for international conferences and workshops and the Guest Editor for a special issues on AI and Big Data of the IEEE and other journals. He is the Representative of Romanian in the Earth Observation Program Board.