Multispectral Data Analysis for Semantic Assessment—A SNAP Framework for Sentinel-2 Use Case Scenarios

Alexandru Cosmin Grivei, Iulia Coca Neagoe, Florin Andrei Georgescu, Andreea Griparis, Corina Vaduva, Zoltan Bartalis, and Mihai Datcu, *Fellow, IEEE*

***Abstract*—Sentinel-2 satellites provide systematic global coverage of land surfaces, measuring physical properties within 13 spectral intervals at a temporal resolution of five days. Computer-based data analysis is highly required to extract similarity by processing and to assist human understanding and semantic annotation in support of mapping Earth's surface. This article proposes a data mining concept that uses advanced data visualization and explainable features to enhance relevant aspects in the Sentinel-2 data and enable semantic analysis. There is a two-stage process. At first, spectral, texture, and physical parameters related features are extracted from the data and included in a learning process that models the data content according to statistical similarities. In parallel, the second processing stage maximizes the data impact on the human visual system to help image understanding and interpretation. Target classes are subject to exploratory visual analysis, such that both visual and latent characteristics are revealed to the user. The concept is further implemented as Sentinel-2 dedicated data analysis (DAS-Tool) plugin for the Sentinel Application Platform and deployed as an open-source tool empowering the Earth observation community with fast and reliable results. Accommodating multiple solutions for each processing phase, the plugin enables flexibility in information extraction and knowledge discovery that will bring the best accuracy in mapping applications. For demonstration purposes, the authors focus on a detailed benchmark against reference data (ground truth) for the Southern region of Romania, then use the selected algorithms in a forest fires scenario analysis for the Sydney region in Australia. The processing involves full-size Sentinel-2 images.**

***Index Terms*—Exploratory visual analysis, multispectral data content representation, semantic annotation, Sentinel application platform (SNAP), Sentinel-2 data analysis.**

Manuscript received March 31, 2020; revised June 17, 2020 and July 3, 2020; accepted July 20, 2020. Date of publication July 30, 2020; date of current version August 14, 2020. This work was supported by the framework of the "Multispectral Data Analysis Toolbox for SNAP ESA's SentiNel Application Platform" project, funded by the European Space Agency (ESA) under Grant 4000121870/17/I-SBo. *(Corresponding author: Alexandru Cosmin Grivei.)*

Alexandru Cosmin Grivei is with the Research Center for Spatial Information, University Politehnica of Bucharest, 061071 Bucharest, Romania, and also with the Centre of Excellence in Communications and Information Technologies, Military Technical Academy "Ferdinand I", 050141 Bucharest, Romania (e-mail: alex.grivei@ymail.com).

Iulia Coca Neagoe, Florin Andrei Georgescu, Andreea Griparis, and Corina Vaduva are with the Research Center for Spatial Information, University Politehnica of Bucharest, 061071 Bucharest, Romania (e-mail: neagoe.iulia@yahoo.com; florin.andreig@gmail.com; andreea.griparis@gmail.com; corina.vaduva@gmail.com).

Zoltan Bartalis is with the EO Directorate of the European Space Agency ESA ESRIN, 00044 Frascati, Italy (e-mail: Zoltan.Bartalis@esa.int).

Mihai Datcu is with the Research Center for Spatial Information, University Politehnica of Bucharest, 061071 Bucharest, Romania, and also with the Remote Sensing Technology Institute, German Aerospace Center, 82234 Oberpfaffenhofen, Germany (e-mail: mihai.datcu@dlr.de).

Digital Object Identifier 10.1109/JSTARS.2020.3013091

I. Introduction

EARTH observation (EO) plays a pivotal role in understanding the environment with all the land cover transformations. Expanding and innovating remote sensing technologies are of utmost importance in monitoring the surrounding processes and activities. Imaging sensors have been designed and developed such that they record specific aspects of Earth land cover based on the spatial, spectral, or radiometric resolutions, which can help estimate geophysical parameters and land processes.

Copernicus, the flagship program of the European Union for EO, is expanding fast, doubling its footprint on Earth's surface coverage through the Sentinel satellite constellation and supplying approximately a fifth of the world's EO data. In 2017 alone, the volume of downloaded EO products integrated into daily activities to improve decision making grew by over 133%. The percentage of companies using Copernicus data increased with 11% in Europe and 6% globally [1]. In this frame, over 9 million Sentinel products have been acquired and 160.000 users have been registered. The growing interest in EO products turns into a Big Data basis, raising huge challenges for information extraction, and value-adding.

Part of the Copernicus Program, the twin satellites of the Sentinel-2 (S2) mission (Sentinel-2A and Sentinel-2B) target terrestrial observations in support of services like vegetation, soil and water cover, inland waterways and coastal areas assessment, land use and change detection mapping, disaster relief support or climate change monitoring. Sentinel-2 ensures the continuity for Landsat and Spot observations and improves the data availability. With a systematic global coverage and a five-days temporal resolution, this mission raises great interest due to the augmented volume of medium resolution imagery it provides and its ability to measure the radiation reflected by the Earth surface in 13 different intervals of the electromagnetic spectrum.

Sentinel-2 data are expected to strengthen the routine generation of generic land-cover and land-use maps, but also to ease monitoring processes of hotspots—areas of interest that are prone to specific environmental challenges and problems. They

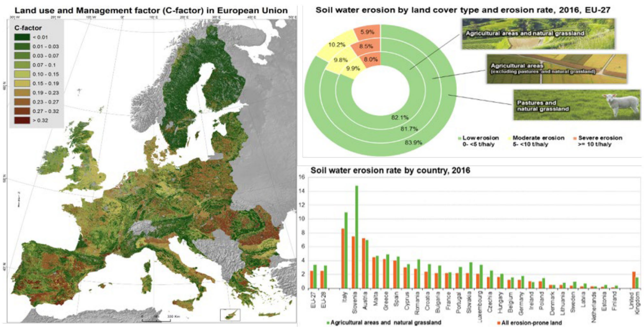


Fig. 1. Example of EO derived products: Regional/global thematic maps, charts and analytics (Eurostat: Agri-environmental indicator-soil erosion).

can contribute to the mapping of geophysical variables like leaf coverage, leaf chlorophyll content, moisture, or burnt matter. To this aim, raw data needs to be transformed into actionable intelligence and become measurable. Based on the intended type of application, the extracted information must be wrapped into value-added products integrating accurate land use, land cover, thematic maps, and derived analytics on the scene content (see Fig. 1).

Effective EO data exploitation lies from data mining for information extraction to machine learning for automatic knowledge discovery and data analytics for the final examination and semantic interpretation. In terms of multispectral data exploitation, the literature shows significant progress on theoretical algorithms (as presented in Section II.A), but not many of them were developed or accustomed to the particularities of Sentinel-2 data, despite their promising accuracy of results and they act in an isolated manner.

The traditional approach considers that the data content must be first described through its main characteristics, wrapped up as feature vectors with a low level of understanding. Discovering similarities between these features are ultimately leading toward extracting structures and contextual meaning from the data. The goal is to define those rules hiding within unstructured data warehouses and extract knowledge. Human inference has been for a long time the key factor to bridge the semantic gap between the real meaning and numerical analysis provided through computation [2]. Even if traditional approaches are rather fast and easy to apply, they are data-driven and applicable at a limited scale.

The deep learning (DL) paradigm is making the scientific community gradually shift toward self-learning methods. Once defined, they can adapt their rules to each scenario, deal with big data collections, and provide extremely high accuracy given a sufficient training dataset. The setup comes with specific know-how and significant computational resources as a tradeoff [3].

As the EO data volume is growing, the need for computer-based analysis is highly required to assist human understanding and allow users to harness the data potential. The trend is to develop exploitation tools that encourage communication and the dissemination of results among end-users, scientific communities, and developers. Current solutions in the field provide basic statistics, processing, and graphical interfaces for data visualization. They also include a collection of routines that can be linked and embedded in a workflow to extract knowledge and accelerate the implementation of envisaged applications. A very well-known open-source system is QGIS. Its toolbox, Orfeo [4], integrates a set of algorithms that process multispectral and radar images, being available through the command line. Google Engine [5] is an alternative solution for data analysis as it includes a large image database and predefined methods, but also the possibility to add new algorithms for data processing. A community platform was created to increase Sentinel data exploitation and accelerate further scientific development. Named the Sentinel Application Platform (SNAP), it reunites dedicated Toolboxes to offer the most complex platform for all Sentinel missions, including Sentinel-2 [6]. Despite being EO data processing tools, they have limited capability in dealing with the semantic exploitation of Sentinel-2 data, and specific know-how is needed to complete a data analysis workflow.

Considering SNAP and the Sentinel-2 Toolbox as the support system infrastructure, this article introduces a knowledge discovery paradigm based on explainable features for the semantic annotation of Sentinel-2 data. A preliminary assessment and benchmark of the most relevant state-of-the-art algorithms for data content representation allowed the authors to select a set of complementary features to describe the data. The focus was on spectral and texture characteristics of structures, invariant and reliable properties, and physical parameters. The added value chain aims for simplicity in the calculation (a key aspect in the context of big EO data) and continues with traditional data mining and machine learning algorithms. The goal is to discover similarities and classify the extracted features of objects and reveal high-level semantics. To increase precision up to the values obtainable with DL methods, the proposed concept integrates an exploratory visual analysis approach to bridge the semantic gap and help the user select relevant training dataset. This can be performed in two ways: by feature selection (FS)—which will result in the selection of the most relevant three bands, and by feature transformation (FT)—which will generate a new set of 3 bands by combining all 13 bands of Sentinel-2 image. The best data visualization will be automatically computed, making the user acknowledge relevant features outside the visible range. As a result, further user-based data representation, especially the classification process, will consider both visual and latent characteristics, ensuring high accuracy with minimum human interaction.

Ultimately, the proposed concept is implemented as an extension for SNAP—a multispectral data analysis tool that benefits of intuitive graphical tools and full data processing capability. Named DAS-Tool, it follows an open-source approach to spread across the EO community and serves in land cover applications as well as support in further scientific development through knowledge discovery, machine learning, and semantic annotation.

The rest of this article is structured as follows. The proposed concept is detailed in Section II, described from both conceptual and architectural outlook. A review of existing methods and algorithms concluding with the integrated solutions are also presented in this section. Section III encompasses the results obtained during a benchmark analysis for a Sentinel-2 image acquired over the south region of Romania and a damage assessment use case scenario for the Australian fires in early 2020. Further potential and drawbacks of the DL perspective is resumed in Section IV, leading toward conclusions in Section V. Part of a technological project funded by the European Space Agency, the outcome of this article defines a best practice for Sentinel-2 data analysis that is openly accessible through the SNAP platform.

II. Exploratory Sentinel-2 Data Analysis

A. Existing Methods and Algorithms for EO Data Analysis

1) Feature Extraction: Image information mining considers the data content based on unique numerical patterns representing dominant characteristics like coarseness, contrast, color distribution, or directionality. In reality, these patterns can be associated with physical parameters of the Earth's surface. Taking into account the complexity and informational diversity of the multispectral images, different methods have been proposed for feature extraction (FE) and content description for both pixel and patch-based multispectral image analyses. There is a proliferation of solutions in the literature, but the authors will only pay attention to the most representative algorithms that have already been proved successful for EO data content representation in view of scene classification.

Spectral features. The most important advantages of the features derived from the spectral values of the image are the simplicity of extracting color information from images and the power of representing visual content. An extension of the color histogram, the spectral histogram uses the distribution of the spectral values for image search and retrieval applications with reasonable accuracy [21]. Further on, spectral indexes [22] were derived as all the combinations of surface reflectance from multiple spectral bands to highlight specific features that indicate the presence of vegetation, water, mud, ice, geologic coverage, etc. Their computation implies simple algebraic formulations, like sums, differences, and ratios applied between different band combinations. For features invariant to the scene illumination, the authors of [23] suggest the use of a polar coordinates transformation (a derivative of the MPEG7 standard) of the spectral values instead of rectangular coordinates space. Applied to multispectral EO images, it helps in the classification of regions with powerful shadow and cloud coverages.

Texture features. A textural descriptor incorporates data concerning the structural arrangement of surfaces and their relationship to the surrounding environment [24]. Since the texture is given by the repetitive visual elements like color, shape, or shadow it can be described by statistic measures. Among the numerous descriptors proposed for texture analysis, the most frequent are Tamura features, Wold features, Markov Random Fields, Gabor Features, and Wavelet Features [25].

In the EO field, the texture descriptors are based on the statistical properties or structure of the texture. The cooccurrence matrix is the basis for the texture descriptor in [26], while in [27], the Gabor features are obtained by filtering the first principal component image with Gabor kernels at four different scales and four different directions. In [28], the Wavelet features are used for object-based retrieval in EO archives. A texture descriptor with good results for both optical and SAR data is based on the Gauss–random Markov field [29]. The texture structure is exploited in [30] employing local binary patterns [31] and local edge patterns. From the three texture descriptors proposed by MPEG7 standard, homogenous texture descriptor, and edge histogram are used for image retrieval [32], [33].

Contextual features. Currently evolving texture analysis and local FE techniques have led the way to mixed feature methods that are joining texture and spectral features in the same descriptor, and also to bag of words (BoW)-based methods, which are relying on learning dictionaries of visual words. In the remote sensing community, this technique has been recently introduced for image annotation, object classification, target detection, and land use classification, and it has already proven its discrimination power in image classification [33]. In the BoW framework, there are several ways to generate a visual codebook, but k-means-based approaches are preferred. The BoW framework computes local features for each patch, then a k-words dictionary is formed collecting k randomly selected local features. Finally, each patch is represented by a k-dimensional descriptor with only one nonzero element, the one related to its closest word. An assessment of several patch-based approaches for FE is presented in [33]. The study in [35] tackles the problem of choosing the optimum number of classes that can be extracted and optimum patch size for Sentinel-2 data analysis.

2) Feature Classification: Feature vectors provide low-level characteristics with very low capability in representing the semantic content, but they provide the key properties assisting further data modeling at a reduced computational burn. The next stage of the data description is the identification of similarities within the achieved feature space and class label assignment to a set of measurements (feature vector) [36]. This will result in the EO image classification, where a class label corresponds to a land cover type like agriculture, forest, urban, water, etc. To assign labels, a decision criterion must be designated to discriminate the land cover types. The semantic gap between the content description provided by low-level features and the high-level semantic content of the image can be confined by a series of learning techniques that are grouping features using similarity. This is a feature classification process resulting in classes of objects with semantic meaning. Depending on the user's involvement in the learning process, the algorithms can be organized in supervised, semisupervised, and unsupervised algorithms.

Supervised classification. The learning process is user-driven and requires his knowledge to discover similarities within the data. The information is passed by the user in the shape of training samples, previously labeled by the user. Consequently, each training example is represented by its feature vector related to the label whose classes its content belongs to. It is important that the training set to be as comprehensive as possible because the prediction function, based on which the system will classify new data, depends on it. The most common approaches are instance-based learning, decision trees, and support vector machines (SVMs). The representative solution for the first category, *k*-nearest neighbor classifier computes the distance between the unclassified points to all the labeled training data and performs class assignment based on the metric used to measure the distance [37]. With a training set of feature vectors and their corresponding class labels at hand, the decision trees identify the hierarchy of the feature vectors attributes based on which the classification is done. The attributes hierarchy is decided with respect to the information gained by their selection [38]. The category of SVMs relies on a kernel-based technique aiming to identify the optimal separating hyperplane that maximizes the distance between distinct classes' margins [39]. They require a small training dataset and present high accuracy when applied for EO data analysis. A survey of these algorithms is presented in [40] revealing a widespread use of SVM-based algorithms as active SVMs, semisupervised SVMs, and combined SVMs with other approaches. The kernel function selection is determining the accuracy of the similarities between dataset samples. SVMs have been experimentally shown to work under certain constraints, such as linearity, balanced dataset, and near Gaussian-distributed data.

Semi-supervised classification exploits the statistical models of the data to relieve the knowledge transfer from the user. The learning process will use both, labeled (input from the user) and unlabeled (derived from the image) data, requiring less human effort than supervised algorithms for comparable accuracy [41]. Among the state of the art such methods, the latent dirichlet allocation (LDA) allows for the highest flexibility when it comes to use case applications. It is a generative probabilistic model, first introduced for text modeling, which provides very good results for EO image analysis [42], [43].

Unsupervised classification is a self-organized learning process that looks for similarities within the data with no prior knowledge and minimum human interaction. Their main goal is to identify hidden patterns in unlabeled data based on the relationships between data points. The algorithm discovers correlations within the set and groups its elements into clusters so that the points in the same cluster are similar but dissimilar to other clusters. Not being user-dependent, this type of process is normally preferred for the analysis of unknown data. However, the classification results are inferior to those achieved by supervised algorithms. The prevailing algorithm in this category is *k*-means [44], which assigns each data point to the cluster whose center is closer. This algorithm is also used in this study to find the optimal number of classes that can be distinguished in the feature space.

3) Visual Data Analysis: Visual exploration of multidimensional remote sensing images is very important for bridging the semantic gap in data analysis. The images need to be reduced to only three bands before displayed to the user. FS and FT algorithms are usually applied to enable data visualization. Both approaches decrease the input data dimensionality, sustaining though different composition of the feature subset.

FS algorithms identify a data subset formed by the most relevant features of the input data based on one of the following three types of processes that measure scores of relevance for each individual feature. Filtering relies on selection criteria like battacharyya or Jeffries–Matusita distances to identify the subset that best separates the dataset classes, but also on statistical measures, like mutual information to define the dependency between two variables selecting the ones most dissimilar to each other and most relevant to the dataset. Wrapper methods focus on maximizing the performance of classifiers defining the relevance score. Sequential selection algorithms and heuristic search algorithms [7], Sequential forward floating search [8], and backward elimination correlated with a Gaussian process regression [9] are the most common algorithms. Finally, the embedded methods consider the classification error rate or criteria like mRMR to evaluate the importance of the feature subset for the model. The authors of [9] used mRMR as "spectral band selector that automatically enhances visualization of target classes for image analysis and photo-interpretation." All the user needs to do is to select the image patch labeled as target class and the system will automatically measure the mutual information between the target class and images spectral bands. Based on the mRMR algorithm, the bands are ranked according to their capability to visually separate the target class from other classes.

The FS methods preserve only a reduced set of features, leading to the loss of information from the rest of the data. On the other hand, FT methods exploit the entire feature set X with dimensionality D in order to fulfill the dataset mapping to a space Y with dimensionality d $(d < D)$. The low-dimensional projection is a combination of the original features according to an explicit transformation. Depending on the nature of the transformation, FT methods can be linear or nonlinear. The first category includes algorithms like principal component analysis (PCA) [10] and derivatives (probabilistic PCA [12], Kernel PCA [13]) centered on the covariance matrix and its eigenvalues, linear discriminant analysis—projecting data to increase the distance between means of the classes while decreasing the variance within each class [14], factor analysis (FA)—maximizing intraclass correlation by minimizing interclass correlation [15]. From the nonlinear transformations, diffusion maps [16], *t*-distributed stochastic neighbor embedding [17], Sammon mapping [18], and autoencoders (AE) [19] are the most common methods. They envisage multilayers transformation based on various rules to weight down the number of features by preserving the relevant information. Specific results for remote sensing data visualization are presented in [20].

FS and FT are a matter of removing the redundant data in the feature space and preserve the most relevant information for visualization. Additionally, these approaches are fitting the purpose of big EO data analysis as they map the existing features into a lower feature space. They can improve learning performance, create better models, lower computational complexity, and reduce the storage requirements. However, linking the new set of features to a physical meaning is difficult without further analysis.

B. Proposed Concept for Sentinel-2 Data Analysis

The correct understanding of the data is critical for meaningful analysis in EO. By targeting the proper features, the user can focus content modeling toward the intended application. To this aim, the present article introduces a data mining concept

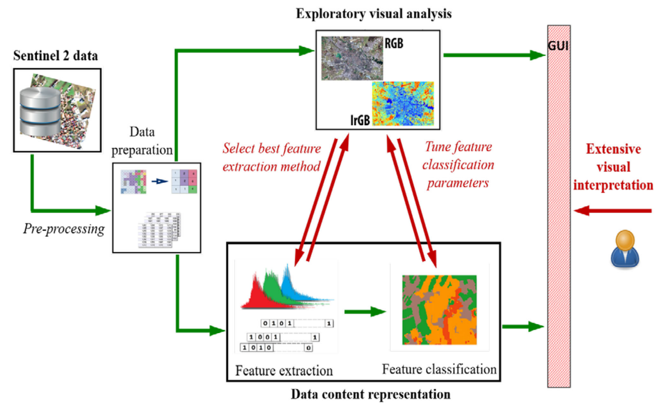


Fig. 2. Methodology for exploratory Sentinel-2 data analysis: Combining visual characteristics (exploratory visual analysis) with latent features (data content representation) in view of extensive visual interpretation. Green arrows—computational workflow; Red arrows—user role.

that uses advanced data visualization and explainable features to enhance relevant aspects in the Sentinel-2 data and enable semantic analysis.

Built to exploit both understandable, perceivable characteristics of the analyzed scene and numerical features hidden inside the data code, the proposed concept integrates all the general data analysis methods (as presented in Section II.A) into a single methodology entailing for two processing stages (see Fig. 2): exploratory visual analysis for data visualization and data content representation for semantic annotation. They act as independent processes, yet, using the result obtained in the first stage as visual support for the second stage, will allow for better data understanding. The green arrows indicate the computational workflow and dependencies, while the red arrows highlight the user role and how he should exploit the two processing stages (the result of the exploratory visual analysis will facilitate parameter setup for the data content representation and results interpretation). Nevertheless, all bands of a Sentinel-2 image must be first brought to a common spatial resolution.

The first stage in the methodology focuses on data understanding. The existence of 13 spectral bands confers Sentinel-2 a complexity that is too high for a user alone to interpret through simple visual analysis. Human perception is only able to analyze and comprehend the information exposed through visualization procedures, which are limited to the use of just three spectral bands.

To make the most of the tri-channel display, the multispectral information to be revealed during the process will be the output of an exploratory visual analysis capable to underpin predominant data features for the scene content. The discrimination between ground structures is determined by their spectral signature. Therefore, advanced data visualization will help the user to perceive certain aspects that are not always reflected in the visible part of the spectrum. Each of the semantic classes (e.g., water, forest, and urban) distinguishable in the Sentinel-2 data has a specific spectral signature, which may not be focused on the visible spectral bands, so the "true color" image representation may hide important details. Data visualization includes 2 approaches: FS and FT. The first process avoids inconsistent data representation provided usually by the bands in the visible range and identifies a data subset of the most relevant spectral bands according to a specific application. The second approach exploits the entire feature set encapsulated in the Sentinel-2 data and applies a low-dimensional transformation based on the original features to project the full dataset into three bands.

For the second stage, a data content representation will focus on the identification of relevant scene characteristics and further grouping by means of classification methods. A compact process interconnects FE and feature classification in such a way to best describe the Sentinel-2 data content characteristics. FE is usually employed to extract and infer knowledge about patterns that are hidden inside the image, offering insights about the scene and advanced content description. Specific classes in the image share characteristics like coarseness, contrast, color distribution, or directionality, which will make FE methods sensitive to spectral, texture, and shape information. The informational content will be encapsulated into multi-dimensional feature vectors (mathematical representations of the image properties). Yet, feature vectors are a low-level semantic representation. To reach actionable information, more compact structures must be identified, as a combination between proportions of different feature vectors. This is subject to a classification task, where a set of patterns are assigned into a group so that, according to some similarity metrics, the elements in the same cluster are more similar to each other than to those in other clusters. To this extent, similar extracted features are meaningfully grouped together and generate a classification map.

The two stages of the proposed methodology complement each other. While exploratory visual analysis is addressing visual properties of structures to stress features beyond visible, data content representation is focusing more on extracting numerical patterns, which will lead to similarity by data processing and results not always corresponding to the user perception on the scene. Latent features are revealed, with no evident meaning what so ever. As such, an extensive visual interpretation is required, where, for each semantic class, a different data visualization can be computed to explain the classification result through all land cover, land use, geographical variables, and land transformations. Data visualization help the user understand the similarities and the resulting classes. In the situation, where several algorithms for data content representation are available, data visualization will support the selection of those result best fitting the application objective and the data semantic annotation.

Abstractly, the proposed methodology for Sentinel-2 data analysis could accommodate a wide list of data visualization and content representation algorithms. As already presented, the research community generated over the years a multitude of mathematical solutions capable to handle the data and make use of the contained information. The motivation behind the authors' selection for the algorithms is the result of an extended benchmark and experience gain over the past years.

For the *exploratory visual analysis* processing, there were considered two complementary approaches dealing with both observable features—*mRMR* [10]—and latent characteristics—*FA* [15]. The spectral bands ranking provided through mRMR criteria has been proven efficient and effective in multispectral data-based applications as it maximizes the response in the human visual system. It increases the saliency of the target class and highlights the bands enabling the higher contrast among neighboring objects to enhance cognitive response to visual physical characteristics. The FA method statistically models the joint variation of observed features and extracts linear combinations of hidden features causing such variability. By dealing with interdependencies between observed variables, the resulting information is less contaminated with error variance than alternative methods for dimensionality reduction, such as the PCA algorithms.

Data content representation aims at semantic annotation by means of explainable features. The proposed methodology opens a double opportunity for data analysis: local processing, at pixel level, and contextual processing, at patch level (after preliminary dividing the scene into a grid of patches).

The selection of FE algorithms derived from their ability to describe meaningful features from multispectral EO imagery (mainly Sentinel-2 data) with high accuracy and performance in information extraction.

At pixel level, the most important element to consider in data interpretation and analysis is the tone expressed as a shade of grey and color (hue, saturation, and value) [45]. Tone is actually the result of measurements performed on the radiation reflected by the Earth's surface. Basic math between spectral bands is describing physical parameters in terms of indexes for vegetation, water, burned area, etc. A generalized approach to determine such parameters is formalized by the spectral Indexes FE algorithm [22], included in the proposed methodology. From the color perspective, illumination invariance has been proven more effective in feature separation. An extension of the MPEG-7 standard for multispectral data representation leads to the polar coordination (PC) based FE [23], which overcomes the limitations of data analysis in case of shadows and light clouds. This is the second algorithm considered in this article.

At the patch level, the spectral features are enriched with spatial dependencies. Combined, spectral, and spatial information can generate signatures encapsulating complex features sharing similar patterns for categories of semantic objects. There are numerous solutions to model the high diversity of spectral information within an image patch and many algorithms were successfully exploited in the field of EO. A selection of five algorithms selection was incorporated for contextual FE. Spectral histogram (Hist) and Gabor features (Gabor) [33] are probably the most common FE algorithms with very good results in multispectral data exploitation depicting spectral and, respectively, texture properties. Mean and dispersion (MeDiB) of the patch form a robust and fast computing feature vector, very useful in the context of big EO data [43]. Bag of spectral indexes (BoWSI) and bag of polar coordinates (BoWPC) [33] use the BoW approach to define visual codebooks based on the physical parameters of the scene.

Further *feature classification* makes use of learning processes to extract similarity from the feature space and defined semantics. The pixel-based analysis is designed for automatic modeling and the LDA model will reveal hidden topics in an unsupervised manner [42]. Local features are difficult to perceive, therefore, user intervention brings no significant value. However, the contextual analysis is centered on features that can be visually distinguished and understood. The user plays an important role in knowledge extraction. SVM and kNN [33] are the selected classifiers, as they deliver increased accuracy with a small training dataset. *K*-means clustering algorithm is adapted to work on both pixel and patch-based approaches, aiming to showcase computational similarities between the extracted features.

C. DAS-Tool—A Plugin for Sentinel-2 Toolbox in SNAP

Further attention has been paid to build a standardized methodology for Sentinel-2 data analysis (DAS), extendable to multispectral imagery in general. The proposed concept has been implemented as an operational tool (DAS-Tool) and installed as a plugin for the SNAP, [6]. SNAP is a multisession scientific platform released by the European Space Agency (ESA) as a free, open-source tool to support the exploitation of ERS-ENVISAT, Sentinel -1, -2, and -3 and a range of national and third party missions. Some of the platform feature highlights are common architecture for all toolboxes, including DAS-Tool: fast image display and navigation, layer management, band arithmetic, and interactive visualizations of three-dimensional (3-D) globe (the WorldWind visualization, [46]). All these benefits transform this platform into a very useful tool suited for everyone.

SNAP underlies sensor-specific toolboxes and is based on the legacy of ESA BEAM, which incorporates many years of evolution and improvements in terms of application programming interfaces (API) design, processing framework development, and in terms of common EO data model, also called product data model.

Usually, it is a good practice to separate the data and functionality from the Graphical User Interface (GUI) code and this is also the case here. The GUI and non-GUI code are placed in different projects named SNAP desktop and SNAP engine. The first one contains the interface with the user and the second includes the functionality. The Engine is independent but the GUI is dependent on the first one. On top of these two components, the user can add any combination of toolboxes as add-ons.

Being a logical and physical software configuration item with a unique name, the module is the main component of the SNAP architecture. The module comprises both resources and source code, has a version and it may be interdependent with other modules, meaning that it may reuse provided resources and functions or use their application programming interfaces (API). In SNAP every module is a plug-in and it can be dynamically loaded and unloaded at runtime using the dedicated plug-in management. The installation of new extension modules or the update of the installed ones can be performed by the users through a local module file.

The integration of DAS-Tool in SNAP has a double purpose envisaged: enrich SNAP capabilities, on one hand, allow the proposed tool to access a large pool of processes, functionalities, and GUI, on the other hand, all for the ease of Sentinel-2 data

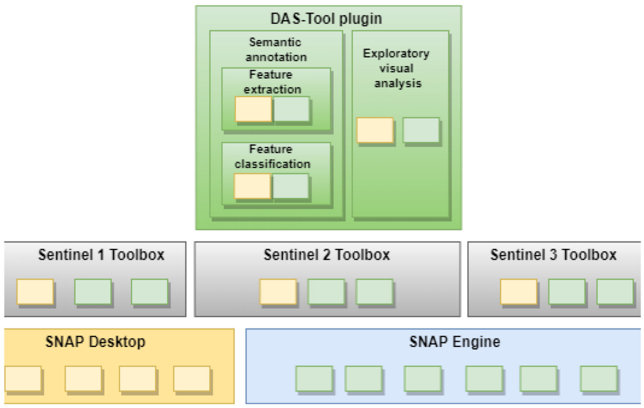


Fig. 3. DAS-Tool modules and their dependencies on the Sentinel-2 Toolbox and SNAP main framework, composed of SNAP Desktop (for GUI components), and SNAP Engine (functionalities for products).

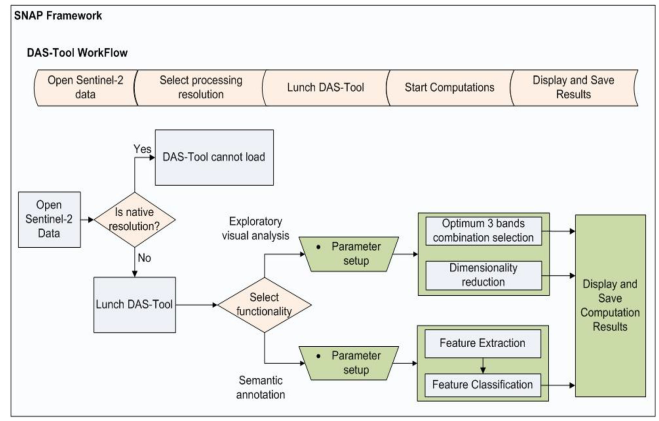


Fig. 4. DAS-Tool Workflow—The overall processing chain goes outside the functionality of DAS-Tool, stressing SNAP basic components to cover all necessary actions from data loading and preparing through actual data processing, up to results display and saving.

understanding and analysis. The architecture of the DAS-Tool plugin follows the methodology in Fig. 2 and comprises two main modules, *Exploratory visual analysis* and *Semantic annotation based on data content description*. The second module is also divided into two submodules: *FE* and *Feature classification*. The plugin depends on the Sentinel-2 Toolbox, as shown in Fig. 3.

The development of the plugin was done using IntelliJ IDEA, the predominant programming language being Java. Both main modules are independent, being structured following the good practice explained above.

The general workflow follows few general steps: load data into product explorer, perform basic data preparation (i.e., resampling), visual inspection through exploratory analysis, extract image features, and define groups of similar patterns that can depict a semantic meaning to the user.

The GUI of SNAP guides the user for the entire process, allowing different interactions with the system focusing on the selection of parameter values for the algorithms applied, saving the results. Using the SNAP GUI the user may access the DAS-Tool GUI from where he can navigate into one of the two main modules and their functionalities. The *exploratory visual analysis* module enables the identification of a visual representation to fit the image content by means of the two approaches. The *semantic annotation based on data content* description allows the user to control the algorithms for both *features extraction* and *feature classification* actions in a way that the resulting data include actionable information in the context of the intended application. All of the processing is based on Sentinel-2 data. Fig. 4 resumes the main steps of the DAS-Tool workflow. SNAP is automatically indicating if the bands of Sentinel-2 data do not have the same spatial resolution and automatically opens the resampling functionality. Following the dedicated process to bring all bands to a native resolution, Sentinel-2 data will be ready to enter the DAS-Tool processing. If bands are not brought to the same spatial resolution, DAS-Tool will remain inactive.

Being a plugin to SNAP, a platform already known in terms of user interface, DAS-Tool has a very big advantage regarding logic and usability. The new functionalities are naturally integrated with the old ones and also offer a very intuitive usage flow. The modules are independent of each other, but both of them allow the ingestion of human knowledge through human interaction. The integration of multiple algorithms gives flexibility to the process.

The results obtained using the plugin do not require special resources or special knowledge and do not take too much time to be displayed. The hardware requirements are in accordance with the ones of SNAP.

Fig. 5 shows the path to DAS-Tool functionalities into SNAP interface (the A snapshot), the graphical elements involved in parameter setup for the exploratory visual analysis (the B snapshot) and the GUI flow to follow to perform semantic annotation through FE and feature classification (the C snapshots).

III. Results

A. Experimental Setup

For demonstration purposes, the proposed framework is deployed as the core of a procedure dedicated to the analysis of Sentinel-2 data properties. To present the overall capabilities of the DAS-Tool plugin, we present two use case scenarios. The first use case presents a situation, where significant information about the analyzed area is available. We benchmarked all integrated FE and feature classification algorithms against reference data. Our goal is to identify the best performing combination of algorithms for a given class (forest) and the overall best performing duo for an entire Sentinel-2 scene. For the benchmark, we selected a cloud-free scene from august 19, 2019, covering southern Romania, between the cities of Bucharest, Ploieşti, Buzău, Slobozia, and the Danube, presented in Fig. 6.

Experiments will be continued with the second use case depicting a damage assessment scenario (recent fires in Australia, at the end of 2019, Fig. 13) for which there is no ground truth. Considering no prior information about the affected area, the second use case will import the combination of the algorithms

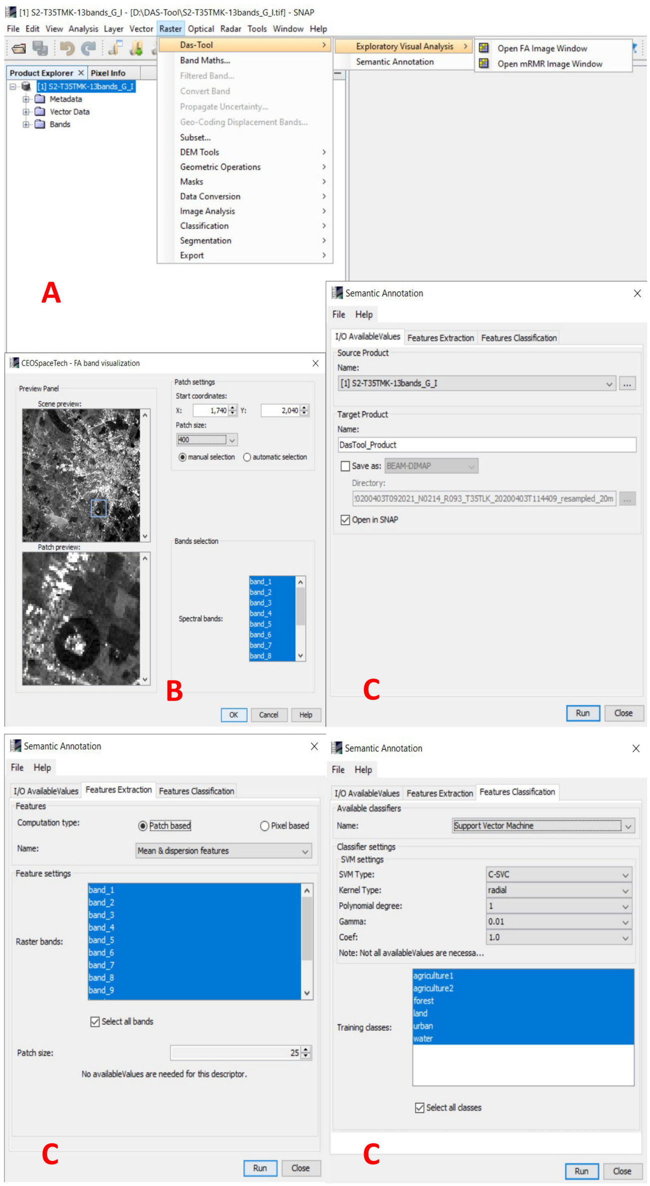


Fig. 5. (A) SNAP GUI indicating the path for the DAS-Tool interface; (B) DAS-Tool GUI for exploratory visual analysis (the FA approach); (C) DAS-Tool GUI components for the semantic annotation—the steps to analyze EO data through feature extraction and feature classification.

Fig. 6. Sentinel-2 scene from August 19, 2019, covering southern Romania, between Bucharest, Ploieşti, Buzău, Slobozia cities and the Danube.

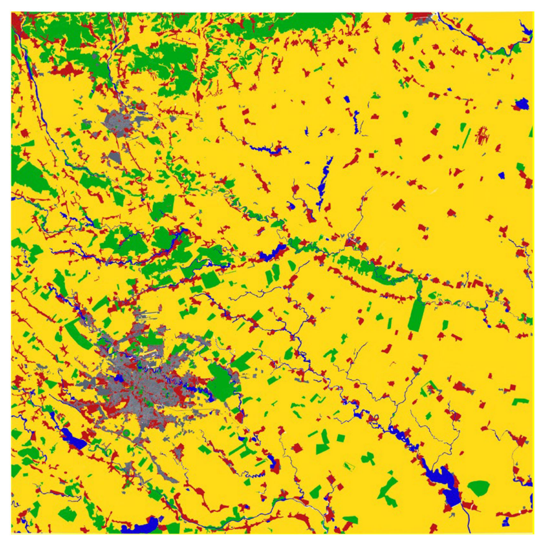


Fig. 7. Ground truth for Fig. 6: forests (green), agriculture (yellow), water (blue), low density urban (red), high density urban (grey).

selected during the benchmark in the first use case. The overall processing workflow is presented in the rest of this section.

To support the evaluation of data content representation, we constructed the ground truth for the scene in Fig. 6 using Urban Atlas data from the Copernicus Land Monitoring Service [47] to precisely map the main urban areas covered by our test scene. The ground truth from all the other regions from our scene was built from land cover data from GEOFABRIK [48]. The resulted land cover map, built at pixel level is presented in Fig. 7 and contains the following land cover classes: forests (green), agriculture (yellow), water bodies (blue), low-density urban areas (red), and high-density urban areas (grey).

We resampled all Sentinel-2 bands to the same spatial resolution of 10 m, increasing the size of our test image to the level of big data. To assess the performance of the integrated DAS-Tool algorithms, we mainly focused on patch-based land cover classifications—where the scene is divided into patches, for each patch the features are extracted and then classified. Contextual information is thus considered, reducing also the size of the data and lowering the computational burn, as the analyzed scene is a full Sentinel-2 acquisition of 10 980 $\times$ 10 980 pixels. The process is applied for two subscenarios (two patch sizes): 25 $\times$ 25pixels (demonstrated to enable best structure discrimination for Sentinel-2, [35]), and 60 $\times$ 60pixels (representing the average-sized patch to encompasses a semantic class for medium size multispectral image resolution, [2]). The process includes MeDiB, Hist, Gabor, BoWSI, BoWPC (FE algorithms), SVM, kNN, K-means (classification methods).

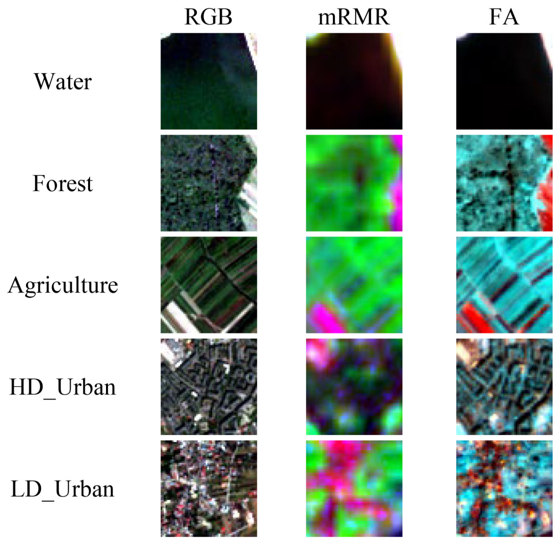


Fig. 8. Training samples (60 × 60 pixels patches): water, forest, agriculture, high density urban (HD_Urban), low density urban (LD_Urban). The columns from left to right are the RGB, mRMR, FA visualization modes.

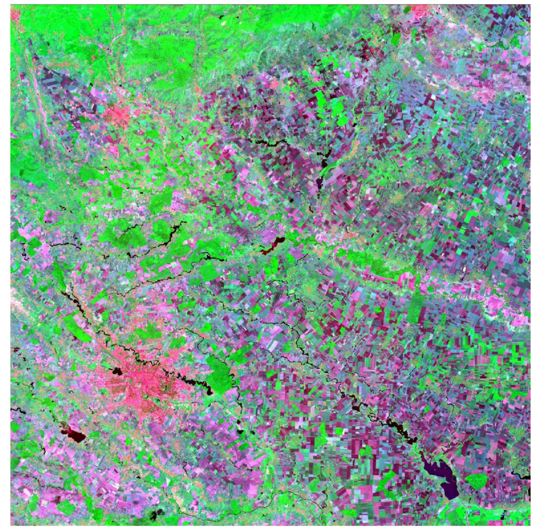


Fig. 9. mRMR representation (Relevant bands selection) for image in Fig. 6.—bands 1, bands 9, and bands 5.

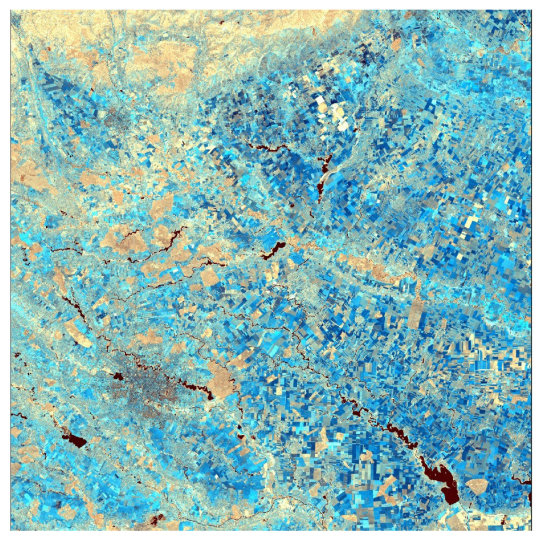


Fig. 10. FA representation for image in Fig. 6.

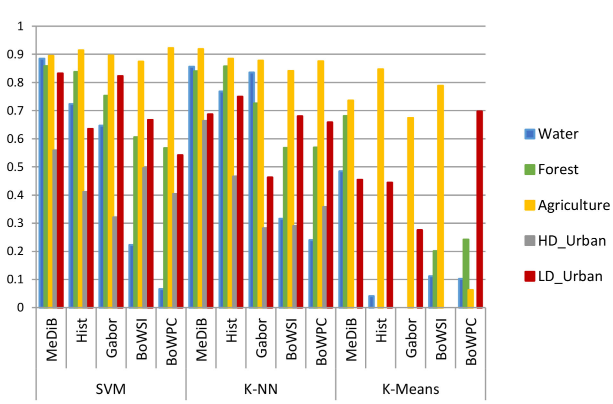


Fig. 11. F1 score computed for different combinations of feature extraction algorithms (MeDiB, Hist, Gabor, BoWSI, BoWPC) and feature classification algorithms (SVM, *K*-NN, *K*-Means) applied on the Sentinel-2 image divided in patches with a size of 25 × 25 pixels.

Upon comparing the obtained classification results, we determine the FE—feature classification algorithms combination generating the highest accuracy for the class of forest (the central class for the intended application) and also that leads to the highest average accuracy (over all classes). We apply these two algorithm combinations on two other scenes from west of Sydney, Australia: the first is before the large vegetation fires (November 06, 2019) and the second is acquired during the fires (December 31, 2019). In this case, no reference data were available, particularly in the context of a dynamic land transformation, and the classification was built on the annotations of observable land cover classes and with the help of the integrated visual enhancement algorithms, mRMR and FA.

B. Prospect of Both Visual and Latent Characteristics

In order to build the training set for benchmarking the integrated supervised algorithms and FE algorithms, we employed FA and mRMR algorithms, to enhance data visualization and minimize the semantic gap between human perception (a trichannel perspective, centered on the RGB bands) and numeric data (13 bands, including also radiation outside the visible spectrum). For each of the five land cover classes, we selected representative samples (see Fig. 8), which are easier to identify using the two visualization modes (see Figs. 9 and 10).

C. DAS-Tool—the Semantic Assessment

The overall performances for each combination of benchmarked algorithms are presented in Figs. 11 and 12 as F1-scores.

The interpretation should go, however, beyond the values obtained for the validation measures. For instance, the proposed methodology is intended to integrate human feedback to obtain the best class separation (through supervised feature classification), but, at the same time, it allows data content analysis with no prior information about data or the process itself (when

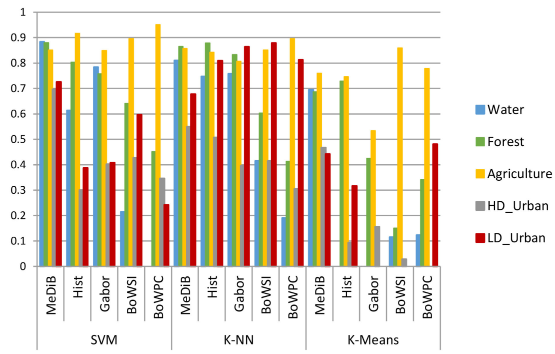


Fig. 12. F1 score computed for different combinations of feature extraction algorithms (MeDiB, Hist, Gabor, BoWSI, BoWPC) and feature classification algorithms (SVM, *K*-NN, *K*-Means) applied on the Sentinel-2 image divided in patches with a size of 60 × 60 pixels.

using kmeans or LDA, combined with a set of default parameters experimentally determined). Both scenarios (25 × 25 pixels and 60 × 60 pixels patch size) show increased accuracy for the supervised analysis, yet all results show an average F1 of 0.5861 for patches of 25 × 25 pixels and of 0.5855 for patches of 60 × 60 pixels.

After we analyzed all the possible combinations for patch-based land cover classifications, we resume the best scores as follows

1) Best F1, averaged over all classes: 0.8074 for MeDiB and SVM, patches of 60 × 60 pixels, 0.8056 for MeDiB and SVM, patches of 25 × 25 pixels, 0.7929 for MeDiB and kNN, patches of 25 × 25 pixels.
2) Best F1 score for the forest class: 0.8843 for MeDiB and SVM, patches of 25 × 25 pixels, 0.8786 for Hist and kNN, patches of 60 × 60 pixels, 0.8396 for Hist and kNN, patches of 25 × 25 pixels.

Expanding to the forest fire assessment near Sydney, Australia scenario, we employed the MeDiB FE algorithm combined with the SVM classifier, applied on 25 × 25 pixels patches. For the training set, we manually selected representative areas from each of the two analyzed scenes (before and after fire event).

Fig. 14 illustrates the training dataset (manually selected using polygon drawings) for the scene from December 31, 2019, where examples were given for the following semantic classes: water bodies, agriculture, high-density urban areas, low-density urban areas, forests, burned areas, and smoke. For the scene before the forest fires, from November 06, 2019 (see Fig. 13), the burned areas and smoke classes were omitted. The results for semantic land cover classification are presented in Fig. 15 for the pre fires scene, whereas the results for the post fires scene are shown in Fig. 16.

Due to the presence of smoke during the fires, most of the classes show variations compared to the prior scene, but the overall impact of the large scale forest fires can be observed and tracked. For performance assessment, we compare the results with the information provided by the My Fire Watch platform [62], presented in Fig. 17. Fires from 2019 and 2020 are represented as grey and black shapes, respectively. The grey areas largely represent the burned area land cover class, while the black shapes cover the smoke. As the post fires Sentinel-2

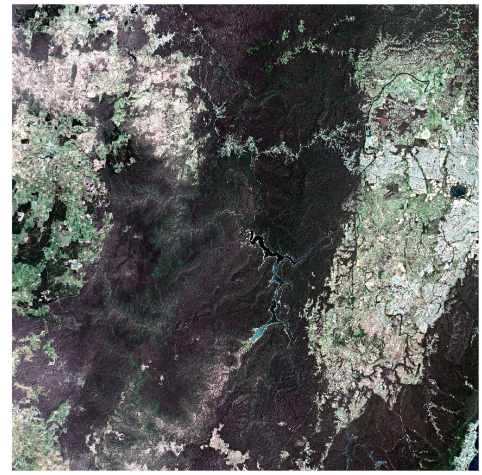


Fig. 13. Sentinel-2 scene from West of Sydney, acquired on November 6, 2019, before the large scale forest fires.

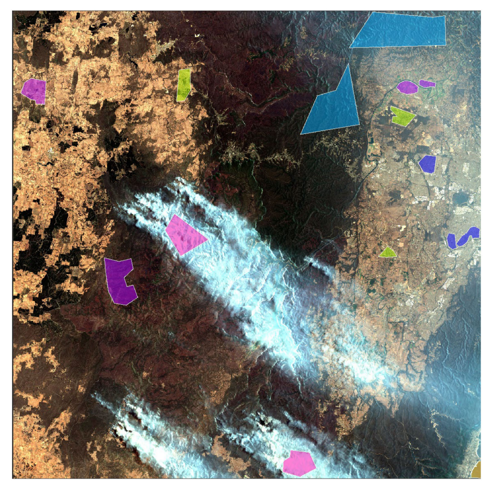


Fig. 14. Sentinel-2 scene from West of Sydney, acquired on December 31, 2019 during the large scale forest fires. Training dataset are marked using polygon regions: water bodies (orange), agriculture (light purple), high density urban areas (dark blue), low density urban areas (light green), forests (light blue), burned areas (purple) and smoke (light pink).

scene used in our analysis is dated, December 31, 2019, we can estimate that representations in Figs. 16 and 17 mostly cover the same land cover classes.

To conclude the forest fires scenario, Fig. 18 presents a FA data visualization mode for the image acquired postevent to help the user better understand the extent of the fires. The features computed are revealing part of the structures covered by smoke. Visually, they look similar to the burned area.

IV. Deep Learning—A Future Perspective

Data analysis, and machine learning in general, is a multiple-stage process. The corresponding algorithms can be classified as shallow and deep learners. In the EO community, traditional (implemented in this article and the most common) approaches, such as SVM and ensemble classifiers are successful "shallow learners" used for image classification and change detection.

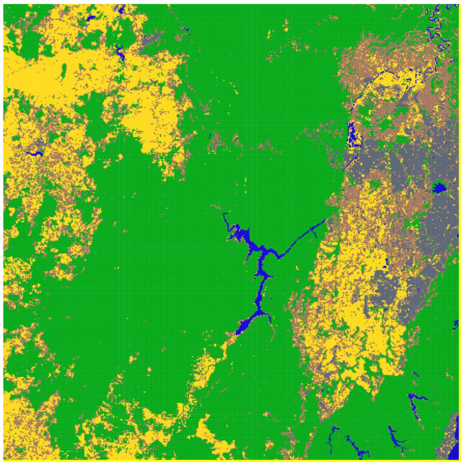


Fig. 15. Classified Sentinel-2 scene from West of Sydney, Australia, November 06, 2019 using MeDiB and SVM, patches of 25 × 25 pixels. Classes: low density urban (brown), high density urban (grey), forest (green), agriculture (yellow), water bodies (blue).

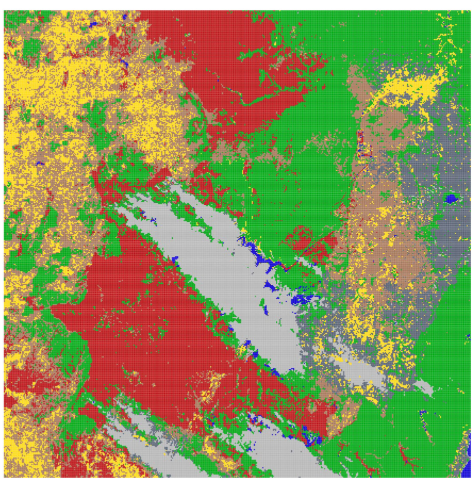


Fig. 16. Classified Sentinel-2 scene from West of Sydney, Australia, December 31, 2019 using MeDiB and SVM, patches of 25 × 25 pixels. Classes: low density urban (brown), high density urban (grey), forest (green), agriculture (yellow), water (blue), smoke (light grey), burned area (red).

Fig. 17. Print-screen from MyFireWatch web platform [62], covering approximately the same area as the analyzed Sentinel-2 scenes, showing fires from 2019 (grey), and fires from 2020 (black).

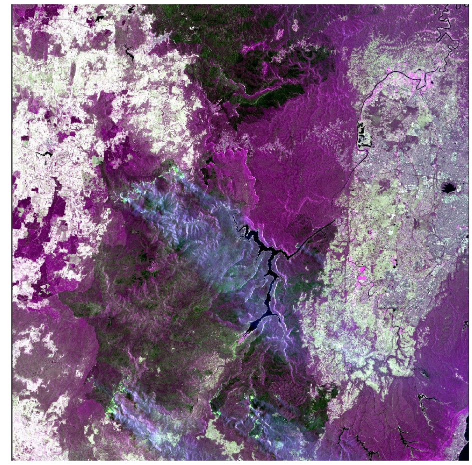


Fig. 18. FA representation for the Sentinel-2 scene from West of Sydney, Australia, acquired on December 31, 2019. The smoke covering the central and bottom area of the scene is greatly faded, enabling visual observation of the Earth surface, under the smoke. SNAP uses a randomly selected pseudo color scheme representation.

Due to its ability to handle high dimensionality data and perform well with limited training samples, SVMs gained widespread use in image analysis. Various implementations were also introduced for EO, as reviewed in [49]. Another "shallow learner" is random forest, an ensemble classifier that became popular in the EO community due to the accuracy of its classifications [50].

Over the past years, the popularity of DL algorithms has massively risen in the entire data science community. Although in the literature, there are hundreds of DL papers, only a few studies are on EO data and even fewer on multispectral medium resolution satellite images, including Sentinel-2 data.

Two comprehensive surveys of state-of-the-art EO DL research are presented in [51] and [52]. The main DL architectures are AE, deep belief networks, recurrent neural network (RNN). Most of the studies addressing the land use classification based on DL algorithms are using stable benchmark datasets, including RSSCN7 [53], UC-Merced [54], or WHU-RS [55]. Compared with the benchmark datasets used in multimedia scene classification, the volumes of available RS datasets are limited. To overcome this inconvenience, two new Sentinel-2 benchmark archives were created. EuroSAT consists of 10 classes, covers 13 different spectral bands, and includes a total of 27 000 labeled and geo-referenced images [56]. The performance of SVM, CNN, ResNet-50, and GoogleNet architecture was tested on this dataset. Training a shallow CNN architecture on the BigEarthNet (multilabel 590 326 Sentinel-2 image patches) provides much higher accuracy compared to a state-of-the-art CNN model pretrained on the ImageNet [57].

There are very few studies on DL architecture applied to real satellite images and most of them addressing high-resolution images [58]. A deep patch-based CNN system tailored for medium-resolution RS image classification was proposed in [59]. The achieved accuracies were 61.86% for SVM, 63.01% for pixel-based CNN, and 85.60% for the proposed deep patch-based CNN. The authors of [60] investigated the behavior of LSTM convolutional blocks integrated into fully convolutional deep architectures for urban change detection from multitemporal Sentinel-2 data. For the experiment, they used a single NVIDIA GeForce GTX TITAN with 12 Gb of GPU memory achieving an overall accuracy of more than 95%, and the training time for each model was approximately 70 min. Random forests (RFs), RNN, and temporal CNN (TempCNN) were compared for the classification of Sentinel-2 time series in [61]. For an accurate evaluation, they made all the experiments on the same machine with 12 Central Processing Units (CPUs) and 256 GB of RAM and on an NVIDIA Tesla V100 GPU. They obtained a similar result for RFs and TempCNN, with TempCNNs obtaining the highest overall accuracy values. The required time on CPU to classify about 120M pixels was 20 h for RNN, 2 h15 for TempCNN, and 17 min for RF.

DL architectures have proven capable to outperform humans and human-coded features [51]. However, various DL systems have many parameters and require a significant amount of training data to learn data representation for EO imagery, turning into a cornerstone for the field. Community shares a large amount of data but only a small number of labeled training samples. Other challenges of DL are to define the appropriate network topology and subsequently optimizing its hyperparameters. In many situations, DL systems are difficult to put at work without specific knowledge, coding skills, and extensive computational resources, being out of the reach of common users. Despite all the progress presented in the literature, scenarios depending on real-time response such as damage assessment and crisis monitoring are out of the scope of DL systems at this point, being more appropriate for traditional approaches.

Nevertheless, with the rise of thematic exploitation platforms, high-performance data processing infrastructure and cloud-based services will be within reach for the entire EO community, with full data access. This perspective is addressing the shortcomings of DL approaches for wide use. DAS-Tool is prepared to integrate new algorithms due to its modular architecture, aiming for an evolution closely connected to SNAP's development.

V. Conclusion

This article introduces a data analysis tool (DAS-Tool) that aims at enhancing the exploitation of Sentinel-2 data through fast image understanding and analysis. Based on a data mining concept for knowledge discovery and semantic annotation, DAS-Tool is integrated into SNAP, a standard, open-source operational platform dedicated to Sentinel data exploitation. As such, the proposed tool is enhanced with intuitive interfaces to encourage wide use even outside the EO scientific community. Centered on the characteristics of Sentinel-2 data, the concept behind DAS-Tool aims at increasing the accuracy of traditional algorithms by combining processes that are fit to the image content. The user gains flexibility in data exploration with multiple solutions to extract features (spectral, texture, and physical parameters) and model similarities (through automatic and supervised learning procedures), in multilevel processing (locally—at pixel level, contextually—at patch level). The methodology reduces the semantic gap by revealing to the user the kind of patterns that are statistically similar through exploratory visual analysis. This will increase the relevance of the training samples and the accuracy given a specific application. Experiments show high scores for the validation measures when comparing results with ground truth and demonstrate reproducibility at various scales (different patch sizes) in different applications (land cover classification and burned area detection in forest fires).

The proposed concept is aligning with the new study pointing that a great part of the so called breackthrouh algorithms are just one small step ahead their predecessor. With a careful parameter setup, the traditional machine learning approaches have the potential to reach similar performances for data exploitation. In fact, intelligent, DL it seems to be reaching a limit in its core progress in some fields [63].

Acknowledgment

This work will be made available through ESA's webpage as open source.

References

[1] PwC, Copernicus Market Report, 2019.
[2] C. Vaduva, F. A. Georgescu, and M. Datcu, "Understanding heterogeneous EO datasets. A framework for semantic representations," *IEEE Access*, vol. 6, pp. 11184–11202, 2018.
[3] J. Ball, D. Anderson, and C. Seng Chan, "A comprehensive survey of deep learning in remote sensing: Theories, tools and challenges for the community," *J. Appl. Remote Sens.*, vol. 11, no. 4, pp. 1–64, 2017. [Online]. Available: https://arxiv.org/abs/1709.00308
[4] Orfeo ToolBox for remote sensing images, [Online]. Available: https://www.orfeo-toolbox.org. Accessed: Aug. 2018.
[5] Google Earth Engine, [Online]. Available: https://earthengine.google.com. Accessed: Aug. 2018.
[6] [Online]. Available: https://step.esa.int/main/toolboxes/snap/. Accessed: Feb. 17, 2020.
[7] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Comput Elect. Eng.*, vol. 40, no. 1, pp. 16–28, 2014.
[8] A. Le Bris, N. Chehata, X. Briottet, and N. Paparoditis, "Spectral band selection for urban material classification using hyperspectral libraries," *ISPRS Ann. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. 3, 2016, Art. no. 33.
[9] J. Verrelst, J. PabloRivera, A. Gitelson, J. Delegido, J. Moreno, and G. Camps-Valls, "Spectral band selection for vegetation properties retrieval using Gaussian processes regression," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 52, pp. 554–567, 2016.
[10] D. Bratasanu, I. Nedelcu, and M. Datcu, "Interactive spectral band discovery for exploratory visual analysis of satellite images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 1, pp. 207–224, Feb. 2012.
[11] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *J. Educ. Psychol.*, vol. 24, no. 6, pp. 417–441, 1933.
[12] S. T. Roweis, "EM algorithms for PCA and SPCA," *Adv. Neural Inf. Process. Syst.*, 1998.
[13] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[14] R. A. Fisher, "The statistical utilization of multiple measurements," *Ann. Human Genetics*, vol. 8, no. 4, pp. 376–386, 1938.
[15] R. A. Reyment and K. G. Jvreskog, *Applied Factor Analysis in the Natural Sciences*. Cambridge, U.K.: Cambridge Univ. Press, 1996.
[16] J. E. Camargo and F. A. González, "Multimodal visualization based on latent topic analysis," in *Proc. IEEE Int. Conf. Multimedia Expo. Workshops*, 2014, pp. 1–6.
[17] L. van der Maaten and G. Hinton, "Visualizing data using T-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.
[18] P. Henderson, "Sammon mapping," *Pattern Recognit. Lett.*, pp. 1307–1316, 1997.
[19] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Sci.*, vol. 313, no. 5786, pp. 504–507, 2006.
[20] A. Griparis, D. Faur, and M. Datcu, "Feature space dimensionality reduction for the optimization of visualization methods," in *Proc. IEEE Int. Geosci Remote Sens. Symp.*, 2015, pp. 1120–1123.
[21] B. S. Manjunath, J. -. Ohm, V. V. Vasudevan, and A. Yamada, "Color and texture descriptors," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 6, pp. 703–715, Jun. 2001.

[22] F. P. G.Marchisio and C. Padwick, “On the relative predictive value of the new spectral bands in the WorldView-2 sensor,” in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2010, pp. 2723–2726.
[23] F.-A. Georgescu, D. Răducanu, and M. Datcu, “New MPEG-7 scalable color descriptor based on polar coordinates for multispectral earth observation image analysis,” *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 7, pp. 987–991, Jul. 2017.
[24] Y. Rui, T. S. Huang, and S.-F. Chang, “Image retrieval: Current techniques, promising directions, and open issues,” *J. Vis. Commun. Image Representation*, vol. 10, no. 1, pp. 39–62, 1999.
[25] F. Long, H. Zhang, and D. D. Feng, “Fundamentals of content-based image retrieval,” in *Multimedia Information Retrieval and Management*. Berlin, Germany: Springer, 2003, pp. 1–26.
[26] C.-R. Shyu, M. Klaric, G. J. Scott, A. S Barb, C. H. Davis, and K. Palaniappan, “GeoIRIS: Geospatial information retrieval and indexing system—Content mining, semantics modeling, and complex queries,” *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 4, pp. 839–852, Apr. 2007.
[27] S. Aksoy, “Spatial techniques for image classification,” in *Image Processing for Remote Sensing*. Boca Raton, FL, USA: CRC Press, 2007, pp. 241–284.
[28] V. P. Shah, N. H. Younan, S. S. Durbha, and R. L. King, “A systematic approach to wavelet-decomposition-level selection for image information mining from geospatial data archives,” *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 4, pp. 875–878, May 2007.
[29] M. Datcu *et al.*, “Information mining in remote sensing image archives: System concepts,” *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 12, pp. 2923–2936, Jan. 2003.
[30] K. W. Tobin *et al.*, “Automated feature generation in large-scale geospatial libraries for content-based indexing,” *Photogrammetric Eng. Remote Sens.*, vol. 72, no. 5, pp. 531–540, May 2006.
[31] M. Pietikäinen, T. Ojala, and Z. Xu, “Rotation-invariant texture classification using feature distributions,” *Pattern Recognit.*, vol. 33, no. 1, pp. 43–52, 2000.
[32] M. Molinier, J. Laaksonen, and T. Hame, “Detecting man-made structures and changes in satellite imagery with a content-based information retrieval system built on self-organizing maps,” *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 4, pp. 861–874, Apr. 2007.
[33] F. A. Georgescu, C. Vaduva, D. Raducanu, and M. Datcu, “Feature extraction for the patch-based classification of multispectral earth observation images,” *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 6, pp. 865–869, Jun. 2016.
[34] 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.
[35] F.A. Georgescu, M. Datcu, and D. Raducanu,” Patch-based multispectral image classification assessment for Sentinel-2 data analysis,” in *Proc. BiDS’16 Int. Conf.*, Tenerife, Spain, 2016.
[36] M. S. Nixon and A. S. Aguado, in *Feature Extraction & Image Processing for Computer Vision*. NY, USA: Academic, 2012.
[37] T. Cover and P. Hart, “Nearest neighbor pattern classification,” *IEEE Trans. Inf. Theory*, vol. 13, no. 1, pp. 21–27, Jan. 1967.
[38] M. Pal and M. Paul. Mather, “Decision tree-based classification of remotely sensed data,” in *Proc. Proc. 22nd Asian Conf. Remote Sens.*, 2001, vol. 5.
[39] C. J. C. Burges, “A tutorial on support vector machines for pattern recognition,” *Data Mining Knowl. Discov.*, vol. 2, no. 2, pp. 121–167, 1998.
[40] U. Maulik and D. Chakraborty, “Remote sensing image classification: A survey of support-vector-machine-based advanced techniques,” *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 1, pp. 33–52, Mar. 2017.
[41] X. Zhu, “Semi-supervised learning literature survey,” Univ. Wisconsin, Tech. Rep. pp. 1–39, 2005.
[42] C. Vaduva, I. Gavat, and M. Datcu, “Latent Dirichlet allocation for spatial analysis of satellite images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2770–2786, May 2013.
[43] 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.
[44] S. P. Lloyd, “Least squares quantization in PCM,” *IEEE Trans. Inf. Theory*, vol. IT-28, no. 2, pp. 129–137, Mar. 1982.
[45] *Manual of Remote Sensing*, 2nd ed. Bethesda, MD, USA: American Society for Photogrammetry and Remote Sensing, vol. 1.
[46] [Online]. Available: https://worldwind.arc.nasa.gov/. Accessed: Feb. 17, 2020.
[47] Copernicus Land Monitoring Service, [Online]. Available: https://land.copernicus.eu/. Accessed: Mar. 15, 2020.
[48] GEOFABRIK, [Online]. Available: http://www.geofabrik.de/. Accessed: Mar. 15, 2020.
[49] G. Mountrakis, J. Im, and C. Ogole, “Support vector machines in remote sensing: A review,” *ISPRS J. Photogrammetry and Remote Sens.*, vol. 66, no. 3, pp. 247–259, 2011.
[50] M. Belgiu and L. Drăguţ, “Random forest in remote sensing: A review of applications and future directions,” *ISPRS J. Photogrammetry and Remote Sens.*, vol. 114, pp. 24–31, 2016.
[51] J. E. Ball, D. T. Anderson, and C. S. Chan, “Comprehensive survey of deep learning in remote sensing: Theories, tools, and challenges for the community,” *J. Appl. Remote Sens.*, vol. 11, no. 4, 2017, Art. no. 042609.
[52] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, and B. A. Johnson, “Deep learning in remote sensing applications: A meta-analysis and review,” *ISPRS J. Photogrammetry Remote Sens.*, vol. 152, pp. 166–177, 2019.
[53] X. Yang, W. Liu, D. Tao, J. Cheng, and S. Li, “Multiview canonical correlation analysis networks for remote sensing image recognition,” *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1855–1859, Oct. 2017.
[54] G. J. Scott, M. R. England, W. A. Starms, R. A. Marcum, and C. H. Davis, “Training deep convolutional neural networks for land–cover classification of high-resolution imagery,” *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 4, pp. 549–553, Apr. 2017.
[55] W. Han, R. Feng, L. Wang, and Y. Cheng, “A semi-supervised generative framework with deep learning features for high-resolution remote sensing image scene classification,” *ISPRS J. Photogrammetry Remote Sens.*, vol. 145, pp. 23–43, 2018.
[56] P. Helber, B. Bischke, A. Dengel, and D. Borth, “Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 7, pp. 2217–2226, Jul. 2019.
[57] G. Sumbul, M. Charfuelan, B. Demir, and V. Markl, “BigEarthNet: A large-scale benchmark archive for remote sensing image understanding,” in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2019, pp. 5901–5904.
[58] T. Postadjian, A. L. Bris, H. Sahbi, and C. Mallet, “Investigating the potential of deep neural networks for large-scale classification of very high resolution satellite images,” *ISPRS Ann. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. IV, pp. 183–190, May 2017.
[59] A. Sharma, X. Liu, X. Yang, and D. Shic, “A patch-based convolutional neural network for remote sensing image classification,” *Neural Netw.*, vol. 95, pp. 19–28, 2017.
[60] M. Papadomanolaki, S. Verma, M. Vakalopoulou, S. Gupta, and K. Karantzalos, “Detecting urban changes with recurrent neural networks from multitemporal Sentinel-2 data,” in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2019, pp. 214–217.
[61] C. Pelletier, G. I. Webb, and F. Petitjean, “Deep learning for the classification of Sentinel-2 image time series,” in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2019, pp. 461–464.
[62] [Online]. Available: https://myfirewatch.landgate.wa.gov.au/map.html. Accessed: Feb. 17, 2020.
[63] B. M. Hutson, “Core progress in AI has stalled in some fields,” *Comput. Sci.*, vol. 368, 2020, Art. no. 927.

Alexandru Cosmin Grivei received the B.S. degree in electronics and telecommunications from the Military Technical Academy, Bucharest, Romania, in 2012 and the M.S. degree from Politehnica University of Bucharest, Bucharest, Romania, in 2014. He is currently working toward the Ph.D. degree in electronics and telecomunications with Politehnica University of Bucharest, Bucharest, Romania.

He joined CEOSpaceTech as a Research Assistant, in 2014. Since 2014, he has been an Associate Teacher with the Military Technical Academy “Ferdinand I” and covers subjects such as operating systems, computer networks and protocols, integrated systems, information technologies, and web programming technologies. His areas of interest include geospatial content-based information retrieval, data mining, computer vision and pattern recognition in multispectral Earth Observation images, and big data.

Mr. Grivei was a recipient of an Erasmus grant for a mobility to ERISCS Research Laboratories from Aix-Marseille University Politech, in Marseille, France, in 2012. Between 2015 and 2019, he was part of the TD COST Action TD 1403 - Big Data Era in Sky and Earth Observation. He took part in several national and international research projects. Since 2016, he has been the Publication Chair for The International Conference on Communications (COMM).

Iulia Coca Neagoe received the Engineering degree from the Faculty of Military Electronic Systems and Computer Science, Military Technical Academy, Bucharest, Romania, in 2015 and the M.S. degree in computer science engineering from the University of Bucharest, Bucharest, Romania, in 2017. She is currently working toward the Ph.D. degree with the Department of Applied Electronics and Information Engineering.

She joined the CEOSpaceTech Center from Politehnica University of Bucharest (PUB) as a Research Assistant, in 2016. Since then, she took part in several research projects. Since 2015, she has been working as an IT engineer within the Ministry of National Defense and covers subjects as integrated systems or information technologies and web programming technologies. In 2018, she participated on a three months internship at German Aerospace Center and also on BigSkyEarth Training School, which took place at the Vicomtech Research Center, in San Sebastian, Spain. Her areas of interest include geospatial content-based information retrieval, sensor deep learning, and analytics in multispectral earth observation images.

Florin Andrei Georgescu received the B.S. degree in civil engineering (topogeodesy and automatization of topogeodesy systems) from the Military Technical Academy, Bucharest, Romania, in 2011, the M.S. degree in electrical engineering (advanced techniques for image processing), in 2013, and the Ph.D. degree in civil engineering from the Military Technical Academy and the University Politehnica of Bucharest, Bucharest, Romania, under a study agreement, in 2016.

In 2013, he joined the Research Centre for Spatial Information, UPB, as a Research Assistant, where he conducted research on these topics within the frameworks of several national and international projects. In 2015, he held a five-month internship at the German Aerospace Center (DLR), Oberpfaffenhofen, Germany, as a Visiting Ph.D. Student with the Remote Sensing Technology Institute. His areas of research include geospatial content-based information retrieval, data mining, computer vision, and pattern recognition.

Andreea Griparis received the M.S. and Ph.D. degrees in electronics and telecommunications from the University Politehnica of Bucharest, Bucharest, Romania. After she received the Ph.D. degree in the field of visual data mining for image datasets, in 2018.

She was a Lecturer with the Department of Applied Electronics and Information Engineering, where she teaches subjects related to image processing such as Computer Vision or Medical Imaging. Since 2014, she has been a member of CEOSpaceTech, where she participated in various research projects including eVADE. She has large expertise in information theory, content-based image retrieval, visualization of high dimensional data sets, machine learning techniques, and visual data mining, focused on Earth Observation image datasets. Her new interest areas are deep neural networks and visualization techniques applied to satellite image time series.

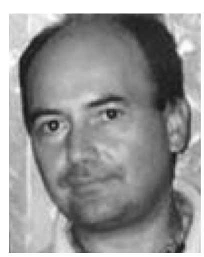
Corina Vaduva received the B.S. and Ph.D. degrees in electronics and telecommunications from the University Politehnica of Bucharest (UPB), Bucharest, Romania, in 2007 and 2011, respectively.

In 2010, she performed a six-month internship with the German Aerospace Center (DLR), Oberpfaffenhofen, Germany, as a Visiting Ph.D. Student with the Remote Sensing Technology Institute (IMF). In 2012, she returned at DLR as a Visiting Scientist for two months. She has been a Research Engineer with the Research Centre for Spatial Information, UPB, since 2007. Her research interests include earth observation image processing and data mining based on information theory, information retrieval, data fusion, change detection, temporal analysis in image time series, and semantic annotation for very high-resolution image understanding and interpretation.

Dr. Vaduva was a recipient of a competition Prize with the Digital Globe 8 bands Research Challenge, in 2010. She was the Chair and program committee member at the ESA Conference on Big Data from Space BiDS, in 2016, 2017, and 2019.

Zoltan Bartalis received the M.Sc. degree in space engineering from Luleå University of Technology, Luleå, Sweden and the Ph.D. degree in satellite remote sensing from Vienna University of Technology, Vienna, Austria.

He works as an Earth Observation (EO) Exploitation Engineer with the EO Directorate of the European Space Agency (ESA ESRIN, Italy). His main focus is on growing the prospects of satellite EO data and applications being used in a coherent and sustainable way in the day-to-day activities of various user groups and organizations. He is also actively supporting geospatial information service providers in ESA member states in developing their commercial activities. Thematically, his present interest is in EO applications for urban development, energy and natural resources, especially in the context of international development. His science and research background is in microwave remote sensing globally over land, in particular for the hydrological cycle.



Mihai Datcu (Fellow, IEEE) received the M.S. and Ph.D. degrees in electronics and telecommunications from the University Politechnica of Bucharest (UPB), Bucharest, Romania, in 1978 and 1986, respectively.

Since 1981, he has been a Professor in electronics and telecommunications with UPB. Since 1993, he has been a Scientist with the German Aerospace Center (DLR), Munich, Germany. From 1991 to 1992, he was a Visiting Professor with the Department of Mathematics, University of Oviedo, Oviedo, Spain. From 1992 to 2002, he was a Longer Invited Professor with the Swiss Federal Institute of Technology, Zürich, Switzerland. In 1994, he was a Guest Scientist with the Swiss Center for Scientific Computing, Manno, Switzerland. From 2000 to 2002, he was with Universitouis Pasteur, Strasbourg, France, and International Space University, Strasbourg. In 2003, he was a Visiting Professor with the University of Siegen, Siegen, Germany. He is currently a Senior Scientist and the Image Analysis Research Group Leader with the Remote Sensing Technology Institute (IMF), DLR, a Coordinator of the CNESDLR-ENST Competence Centre on Information Extraction and Image Understanding for Earth Observation, and a Professor with the Paris Institute of Technology/GET Telecom Paris. His research interests include Bayesian inference, information and complexity theory, stochastic processes, model-based scene understanding and image information mining for applications in information retrieval and understanding of high-resolution SAR, and optical observations.

Dr. Datcu is a member of the European Image Information Mining Coordination Group. In 1999, he was the recipient of the title Habilitation á diriger des recherches from Universitouis Pasteur.