

SILVIA: An eXplainable Framework to Map Bark Beetle Infestation in Sentinel-2 Images

Giuseppina Andresini , Annalisa Appice , and Donato Malerba , *Senior Member, IEEE*

Abstract—Recent long spells of high temperatures and drought-hit summers have fostered the conditions for an unprecedented outbreak of bark beetles in Europe. This phenomenon has ruined vast swathes of European conifer forests creating a need among forest managers to find effective methods to gather information about the mapping of bark beetle infestation hotspots. Sentinel-2 data have been recently established as an alternative to field surveys for certain inventory tasks. Hence, this study explores the achievements of machine learning to perform the inventory mapping of bark beetle infestation hotspots in Sentinel-2 images. In particular, we investigate the accuracy performance of a spectral classifier that is learned for the study task by leveraging spectral vegetation indices and performing self-training. We use a dataset of Sentinel-2 images acquired in nonoverlapping forest scenes from the North-east of France acquired in October 2018. The selected scenes host bark beetle infestation hotspots of different sizes, which originate from the mass reproduction of the bark beetle in the 2018 infestation. We perform a learning stage by accounting for the ground-truth bark beetle infestation masks of a subset of images in the study imagery dataset (training imagery set). The goal is to produce a prediction of the bark beetle infestation masks for the remaining images in the study imagery dataset (working imagery set). Moreover, we use an explainable artificial intelligence technique to study the relevance of spectral information and explain the effect of both self-training and spectral vegetation indices on the mapping decisions.

Index Terms—Bark beetle infestation mapping, explainable artificial intelligence (XAI), forest tree die-back, self-training, Sentinel-2 image processing, spectral classification, spectral vegetation indices.

I. INTRODUCTION

BARK beetles are small insects that reproduce beneath the bark of coniferous trees, depositing their eggs under the bark. The beetles commonly feed dying trees with weakened defense mechanisms, but they can also attack healthy trees when beetle populations expand. In particular, bark beetles swarm only when the temperature is sufficiently high, while recent trends toward extremely high temperatures and drought-hit summers

have led the beetles to reproduce at a faster rate. The current infestation of European spruce bark beetles began in 2012. However, the phenomenon started growing significantly in the middle of 2018. In fact, since 2018, Europe has experienced repeated large-scale hot, dry weather that has led bark beetles, which normally reproduce twice a year, to start reproducing three times a year. This has fostered an unprecedented outbreak of bark beetles that are ruining vast swathes of conifer forests. For example, in 2018, the volume of affected conifer trees in France was already much higher than during the previous beetle infestation in 2003–2007, from which the forests took years to recover.

Although higher temperatures and drought increase bark beetle infestations on a large scale, local damage is often patchy. So the mapping of the spatial spread of the bark beetle infestation is an important challenging problem in both forest management and ecological research. This mapping is traditionally performed in Europe by foresters during field surveys. However, fieldwork based on human assessment of individual trees is laborious and, in the case of large forest zones, it is time-consuming and expensive [1]. On the other hand, the increasing free availability of remote sensing images enables remote sensing image classification to be used as a key tool for (partially) automating the mapping of the forest health status by systematically reducing the amount of fieldwork.

Remotely sensed satellite data rely on spectral signatures from different regions in the electromagnetic spectrum. The recent literature has already linked unique spectral signatures to different functional and structural plant traits (e.g., pigments, leaf structure, plant water content, and nitrogen concentration) by showing that stress sources that affect a plant's biophysical and biochemical properties also affect spectral signatures [2]. Various studies have shown that spectral data of satellite images commonly disclose useful information for monitoring the forest health status comprising the presence of bark beetle infestation hotspots [1], [3], [4].

The recent study of [5] has discussed the importance of the early detection of the bark beetle infestation (i.e., before the bark swarm). In fact, early detection is indispensable for the timely sanitation of infested trees. The same study has also described the early detection symptoms of bark beetle infestation, which comprise the presence of entrance holes, resin flow from entrance holes and boring dust that occur when the beetles attack the tree, penetrate the bark, and excavate mating chambers and breeding galleries. All these symptoms are commonly monitored through terrestrial monitoring [5]. On the other hand, gradual

Manuscript received 25 May 2023; revised 3 August 2023; accepted 2 September 2023. Date of publication 6 September 2023; date of current version 14 November 2023. The work of G. Andresini was supported by the project FAIR - Future AI Research (PE00000013), Spoke 6 - Symbiotic AI, under the NRRP MUR program funded by the NextGenerationEU. The work of A. Appice and D. Malerba was supported by the project SWIFTT "Satellites for Wilderness Inspection and Forest Threat Tracking"¹ funded by the European Union under Grant Agreement 101082732. (Corresponding author: Giuseppina Andresini.)

The authors are with the Dipartimento di Informatica, Università degli Studi Aldo Moro di Bari and Consorzio Interuniversitario Nazionale per l'Informatica - CINI, via Orabona, 4-70125 Bari, Italy (e-mail: giuseppina.andresini@uniba.it; annalisa.appice@uniba.it; donato.malerba@uniba.it).

Digital Object Identifier 10.1109/JSTARS.2023.3312521

¹<https://swiftt.eu/>

needle discoloration and loss of green or discolored needles are considered late symptoms indicating an advanced stage of infestation [5]. Often the next bark beetle generation has already emerged at this point in time, so the recognition of these symptoms will likely not enable timely sanitation. Based on these considerations, Kautz et al. [5] highlight the need to combine terrestrial monitoring and remote sensing monitoring for the early stage detection. In any case, a delayed detection performed with remote sensing monitoring may still serve to perform an inventory task. For example, the French Ministry of Agriculture and Food commissioned the inventory map of the bark beetle infestation hotspots observed in October 2018, to assess the damage in spruce forests of the North-east of France following the 2018 bark beetle outbreak.² This kind of inventory results, although not distinguishing the stage of the infestation, may also support locating undiscovered previous breeding sites and prompting the application of field monitoring in the vicinity of the detected hotspots, to identify recent infestations in the neighbor areas [6], [7].

This study explores how machine learning methods can be used to map bark beetle infestation hotspots in Sentinel-2 images by postponing the problem of boosting early detection by integrating terrestrial monitoring and satellite data monitoring to future work. Therefore, it can be regarded as contributing to establishing opportunities and limits of using machine learning on Sentinel-2 data as an alternative to field surveys for bark beetle infestation inventory tasks. In particular, we explore the effect of spectral vegetation indices coupled with self-training on the accuracy of a spectral classifier trained and evaluated in Sentinel-2 images. Notably, various recent studies have already shown that spectral vegetation indices can aid in gaining accuracy in training a spectral classifier for bark beetle infestation mapping (e.g., [1], [3], [4], and [8]). However, to the best of our knowledge, no previous study has explored the achievements of self-training in this remote sensing problem.

Another contribution of this study is the use of an eXplainable Artificial Intelligence (XAI) technique to strengthen the mapping results to explore the effect of both spectral bands and spectral vegetation indices on decisions concerning forest patches with health trees and forest patches with dying trees, respectively. This allows us to explain how spectral input information contributes to recognizing zones of the opposite classes and how self-training changed the relevance of the input information in this problem. Explaining how spectral input information contributes to recognizing zones of the opposite classes and how self-training changed the relevance of the input information in this problem can help gain the trust of forest managers and remote sensing stakeholders in classifier decisions.

This article is organized as follows. The related work is presented in Section II. Section III gives problem formulation. Section IV describes the study area and data. Section V illustrates the spectral vegetation indices, while Section VI describes the methods used in this work. Section VII discusses the results obtained. Finally, Section VIII concludes this article.

²[Online]. Available: <https://agriculture.gouv.fr/cartographie-des-degats-de-scolytes-de-lepicea-dans-le-nord-est-de-la-france>

II. RELATED WORK

Several remote sensing studies describe methods for mapping forest stress related to bark beetle attacks in Sentinel-2 imagery data. In fact, [2] notes that several forest stress sources (comprising bark beetle infestation) that may affect biophysical and biochemical properties of trees also influence spectral data. For example, [2] shows that chlorophyll degradation and nitrogen deficiency lead to an increase in reflectance spectra in the visible region (in particular, the red and green bands). On the other hand, the NIR and SWIR bands commonly provide information to assess water content and nitrogen concentration in trees. In fact, they help to highlight changes caused by a reaction to vitality losses and cell structure alterations when chlorophyll and leaf water are reduced [2]. Finally, the Red Edge spectral bands are more sensitive to diseased and insect attacks [2]. This analysis has prompted a plethora of studies [1], [3], [4], [8], [9], [10], which explore the ability of various spectral vegetation indices, that mainly combine red, green, NIR, and SWIR bands, to enhance the accuracy of classifiers trained to map bark beetle stress in spectral data. In any case, to the best of our knowledge, previous studies neither leverage self-training, to gain accuracy, nor use XAI, to explain the effect of spectral input information on the classifier decisions.

Regarding the classification algorithm, with the recent boom of deep learning also in remote sensing, some studies have started to explore sophisticated deep neural architectures for mapping bark beetle infestation in satellite images. For example, Zhang et al. [8] experiment a UNet++ with attention mechanism modules for detecting forest damage induced by bark beetles in Sentinel-2 images. Liu et al. [11] test a combination of a deep neural network, a support vector machine (SVM), and a random forest, while Nguyen et al. [12] explore the performance of a convolution neural network in UAV images of forest patches damaged by bark beetle infestations. In any case, several recent studies on bark beetle mapping (e.g., [1], [3], [9], and [10]) still resort to machine learning algorithms and, particularly, to Random Forest [13] and SVM [14]. On the other hand, a few recent studies have started to experiment the performance of XGBoost [15] in a limited number of remote sensing problems such as landslide susceptibility mapping [16] and cloud mask generation [17]. Notably, Singh et al. [17] show that XGBoost can achieve better generalization ability than Random Forest, SVM, and convolutional neural network in cloud mask generation on Sentinel-2 data. In addition, recent studies, developed outside the remote sensing field, have repeatedly assessed the better performance and generalization power of XGBoost in compliance check problems [18], solar radiation mapping [19], and earthquake classification [20]. For these reasons, in this study, we use XGBoost for the spectral mapping of bark beetle infestation in satellite images.

Finally, there are three stages of bark beetle infestation.

- 1) Green attack—A period with no visible abnormal colors in the tree crowns during bark beetle colonization.
- 2) Red attack—A period when the crowns turn a yellow or reddish color with significantly decreased water content in the needles.

- 3) Gray attack—When coniferous trees gradually lose needles after dying.

As reported in a recent survey [21], green attacks can be reliably detected from the ground or verifying the presence early stage bark beetle symptoms in neighbors of red-attacked patches. In fact, most sensing methods, comprising our method, are developed to map red or gray attacks. However, a recent study [22] has concluded that the detection of green attacks in spectral data requires the detection of subtle reflectance changes in the imagery time series collected with very high temporal revisit frequency. Therefore, a few recent studies, [1], [3], [10], have started to explore the time series of spectral data, to perform the early detection of green attacks. The analysis and explanation of the achievements of self-training, coupled with spectral vegetation indices in the times series of Sentinel-2 images of forest areas, will be a future direction of this research.

III. PROBLEM FORMULATION

Let us consider a Sentinel-2 imagery dataset, that is a collection of optical images of nonoverlapping, rectangular, forest areas. Images were acquired in a specific period using the Sentinel-2 sensor in the visible, near infrared, and short wave infrared part of the spectrum. In particular, every Sentinel-2 image \mathbf{S} is an hypercube of size $n_S \times m_S \times d$, which represents a collection of spectral vectors measured on a d -dimensional selection of the Sentinel-2 spectrum over a grid of $n_S \times m_S$ pixels. Every pixel (i, j) of \mathbf{S} is a region of around a few square meters of the Earth's surface, which is a function of the satellite spatial resolution. $\mathbf{S}(i, j)$ is a 1-D real-valued spectrum section of hypercube \mathbf{S} indexed by spatial coordinates i and j within the satellite resolution. Let us assume that the status of the bark beetle infestation of pixels of \mathbf{S} is described by a binary mask \mathbf{Y} , that is a 2-D matrix with size $n_S \times m_S$. Every frame $\mathbf{Y}(i, j)$ is a binary value with two opposite labels: “healthy” (0) and “damaged” (1), where “healthy” denotes the absence of a bark beetle infestation hotspot in the pixel, while “damaged” denotes the presence of a bark beetle infestation hotspot in the pixel.

In this study, we have the binary mask of every study Sentinel-2 image. We partitioned images and masks associated with images in the provided Sentinel-2 imagery dataset into a training imagery set and a working imagery set. Specifically, every Sentinel-2 image and its mask of the starting imagery set is assigned to either the training imagery set or the working imagery set so that the intersection of the training imagery set and working imagery set is empty, while the union of the training imagery set and working set is the entire starting imagery set.

In the learning stage, we considered the Sentinel-2 images of the training imagery set with the binary masks associated with the images, and the Sentinel-2 images of the working imagery set (without the binary masks associated with). We resorted to a self-training strategy and spectral vegetation indices to learn a spectral classifier from these partially labeled spectral data. We used the spectral classifier to predict the binary masks of the Sentinel-2 images in the working imagery set. Finally, in the evaluation stage, we evaluated the accuracy of the spectral classifier by computing several metrics measuring the pixel-wise

differences between the binary labels of both the predicted masks and the ground-truth masks associated with the images of the working imagery set. We note that labels of ground-truth masks of images of working imagery set are neglected in the learning stage, while they are considered in the evaluation stage.

IV. STUDY AREA AND DATA

The study area is in the North-east of France and is predominated by coniferous forests. Due to the mass reproduction of the bark beetle in 2018 and 2019, at the end of April 2019, the French National Forestry Office estimated that 50% of the spruce trees in France were infested with bark beetles, while, under normal conditions, the figure for dead or diseased trees was 15%. In this area, no large windthrows occurred in the years before 2018, and therefore, the observed attacks at the regional scale were very likely caused by the hot summer droughts in 2018 [23]. In addition, the study area is fully covered by the map of the bark beetle infestation hotspots observed in October 2018 and created by Sertit (University of Strasbourg), an organization specializing in emergency mapping.³ This map was commissioned by the French Ministry of Agriculture and Food to assess the damage in spruce forests of the North-east of France following the 2018 bark beetle outbreak. This map does not provided information on the bark beetle infestation stages, but it delineates the boundaries of bark beetle hotspots that caused forest tree die-back in the North-east of France in October 2018.

The remote sensing company Wildsense⁴ rechecked and fixed the infestation hotspot polygons of this map also with the help of foresters. In particular, to avoid mixed reflectance from various causes in discoloration and defoliation of conifer, Wildsense manually selected 92 squared, imagery tiles covering spruce forestry areas fully under bark beetle attacks in October 2018. The selected imagery tiles have spatial sizes varying from 33×41 m to 225×364 m. The percentage of infested territory per tile varies from 0.003% to 0.30% of the tile surface. Selected imagery tiles of this study cover 1072 979 pixels in the North-east of France, at 10-m² resolution.

Based on these premises, Sentinel-2 satellite images, acquired in October 2018, were downloaded from the Sentinel hub for all the selected 92 imagery tiles spanned across the study area. The downloaded imagery set went through atmospheric and reflectance distribution function correction, cloud clearing, and band selection performed by Wildsense. The spectral bands considered are: Coastal aerosol, Blue, Green, Red, Red Edge 1, Red Edge 3, NIR, Water Vapor, SWIR 1, and SWIR 2. In Sentinel-2, the spectral bands Blue, Green, Red, and NIR are all at 10-m resolution, while the bands Coastal aerosol and Water Vapor bands are at 60-m resolution. The remaining bands are at 20-m resolution. So, during the imagery set preparation, spectral bands with different resolutions from 10 m were resampled to 10 m to achieve an equal spatial resolution across all spectral

³The map can be accessed via <https://macarte.ign.fr/carte/3bd52aa2b6422a3a58b5086576f91080/Foyers+de+scolytes+dans+les+peSSI%3%A8res+et+les+sapini%3%A8res+du+Nord-Est+de+la+France,+automne+2018-printemps+2019>

⁴[Online]. Available: <https://swifft.eu/consortium/partners>

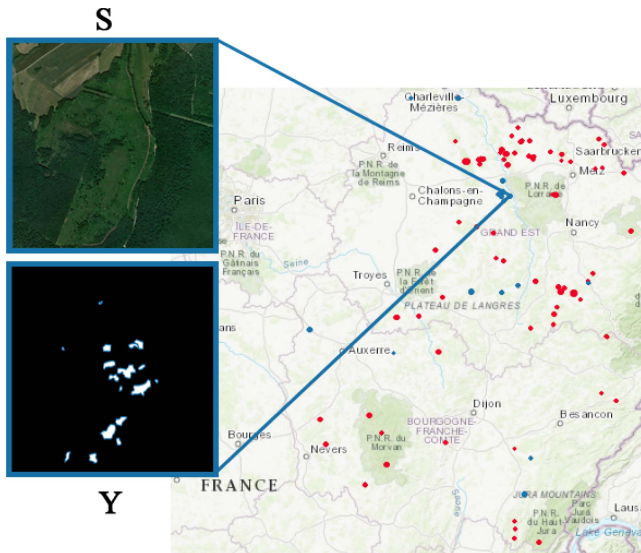


Fig. 1. Location of the study 92 images in the North-east of France. The red tiles fed the training imagery set, while the blue tiles fed the working set of this study.

TABLE I
DISTRIBUTION OF CLASSES “HEALTHY” AND “DAMAGED” IN THE TRAINING IMAGERY SET AND THE WORKING IMAGERY SET OF THE SET OF 92 SENTINEL-2 IMAGES CONSIDERED IN THIS STUDY

Imagery pixels	Healthy (%)	Damaged (%)	Total
Training imagery set	828 825 (97%)	27 237 (3%)	856 062
Working imagery set	213 932 (99%)	2 985 (1%)	216 917
Entire imagery set	104 275 (97%)	30 222 (3%)	1072 979

bands. This dataset was produced with the ground-truth map of the bark beetle infestation in October 2018 by the Wildsense company, to allow the research partners of the EU research project SWIFTT⁵ to partially fulfil project objectives referred to the development of powerful machine learning models to provide forest managers with affordable, simple, and effective remote sensing tools for forest health monitoring of various types of risks comprising tree die-back due to bark beetle infestation.

Within the prepared imagery set, we performed the learning stage by using 76 images (covering 856062 pixels at 10² m² resolution) with their labels (training imagery set) and 16 images (216917 pixels) without labels (working imagery set). A map of the selected imagery scenes, and their partitioning in training imagery set and working imagery set is shown in Fig. 1. In addition, Table I reports the distribution of the pixel classes (“healthy” and “damaged”) in both the training and the working imagery set of this study. These statistics show the high imbalanced condition of the classification problem addressed in this study. Fig. 2 shows the box plots of the spectral bands plotted independently of each other in the entire imagery set with respect to the two opposite ground-truth classes (“damaged” and “healthy”). The box plots show that a greater divergence between the opposite classes can be observed on Red Edge 3, NIR, and Water Vapor spectral bands. This conclusion is

⁵[Online]. Available: <https://swifft.eu/>

TABLE II
F METRIC RESULTS OF ONE-WAY ANOVA ANALYSIS PERFORMED ON THE SPECTRAL BANDS IN THE ENTIRE IMAGERY SET, THE TRAINING IMAGERY SET, AND THE WORKING IMAGERY SET

Spectral band	Entire set	Training set	Working set
Coastal aerosol	2531.33 (*)	3111.50 (*)	235.33 (*)
Blue	1889.21 (*)	2180.71 (*)	97.82 (*)
Green	7651.26 (*)	7223.38 (*)	786.18 (*)
Red	568.90 (*)	574.04 (*)	92.47 (*)
Red Edge 1	10343.63 (*)	8971.99 (*)	1458.48 (*)
Red Edge 3	46754.69 (*)	40189.50 (*)	5357.61 (*)
NIR	39529.89 (*)	34183.18 (*)	4506.64 (*)
Water vapor	40124.71 (*)	34920.57 (*)	4186.97 (*)
SWIR 1	5794.21 (*)	4886.50 (*)	101.41 (*)
SWIR 2	2394.93 (*)	2079.82 (*)	390.40 (*)

(*) denotes that p -value ≤ 0.001 so that the null hypothesis (i.e., the means of the compared groups of pixels labeled with the two opposite ground truth classes “healthy” and “damaged” are equals) is rejected.

also supported by the results of the one-way ANOVA analysis that finds out whether there exists a statistically significant difference between the mean values of the two groups of pixels labeled with the two opposite ground-truth classes. Results of the one-way ANOVA analysis reported in Table II show that the null hypothesis that the means of the compared groups are equals is always rejected with p -value ≤ 0.001 and the highest F-statistic (i.e., the ratio of the variation between sample means on the variation within the samples) is measured on Red Edge 3, NIR, and Water Vapor spectral bands. This conclusion can be equally drawn by performing the one-way ANOVA analysis on the entire imagery set, the training imagery set, and the working imagery set, respectively.

V. SPECTRAL VEGETATION INDEX ANALYSIS

We consider three spectral vegetation indices, namely NGDRI, NMDI, and MCARI. These indices, that were computed from the Sentinel-2 spectral bands, allowed us to capture forest greenness, chlorophyll, and water content, and are selected particularly based on the considerations illustrated in [2] on the sensitivity of Sentinel-2 bands to stress-induced variations in chlorophyll content (Visible), biomass (NIR), and water content (SWIR). In fact, according to [2], Green, Red, and NIR bands are mainly sensitive to chlorophyll degradation. SWIR 1 and SWIR 2 bands are mainly sensitive to vitality losses and cell structure alterations when chlorophyll and leaf water are reduced. Red-edge spectral bands are mainly sensitive to disease and insect attacks.

Therefore, the normalized difference green/red index (NGDRI) [24] was computed to capture the chlorophyll degradation, by combining the green and red bands in the visible spectrum

$$\text{NGDRI} = \frac{\text{Green} - \text{Red}}{\text{Green} + \text{Red}}. \quad (1)$$

The authors in [1], [8], and [9] explore the use of this index for bark beetle mapping in forestry regions in the Southeast of the Skeena region of Canada, the Vysočina region in the central part of the Czech Republic and the Remningstorp, Västra Götaland region in Sweden, respectively.

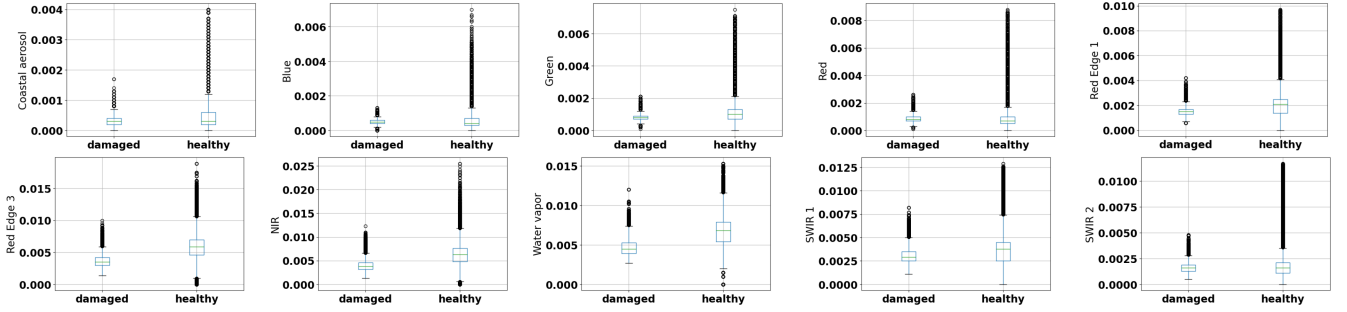


Fig. 2. Box plot distribution of spectral bands in the entire imagery set.

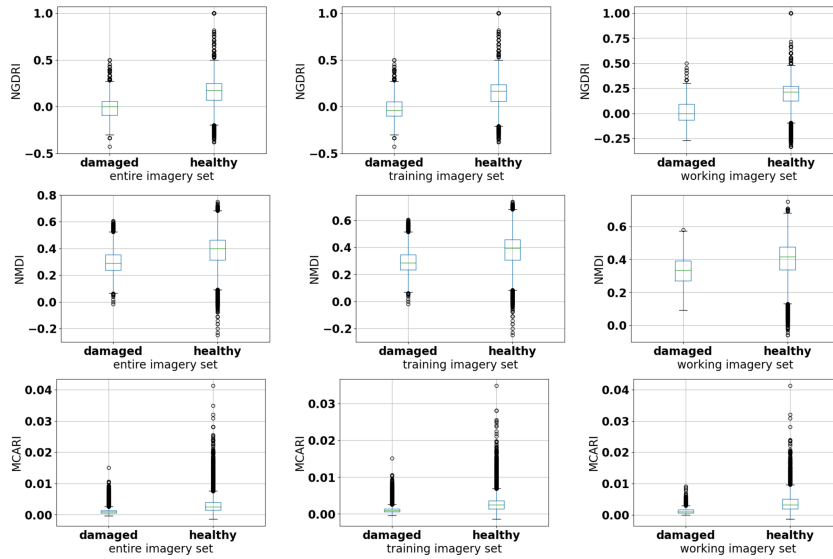


Fig. 3. Box plot distribution of NGDRI, NMDI, and MCARI in the entire imagery set (left side), the training imagery set portion (central side), and the working imagery set portion (right side).

The normalized multiband drought index (NMDI) [25] was adopted for detecting vegetation water by using NIR, SWIR 1, and SWIR 2 bands

$$\text{NMDI} = \frac{\text{NIR} - (\text{SWIR1} - \text{SWIR2})}{\text{NIR} + (\text{SWIR1} + \text{SWIR2})}. \quad (2)$$

Wang and Qu. [25] show that the lower the NMDI, the greater the severity of the vegetation drought, and consequently, the tree vitality loss. The index is also considered in [10] for bark beetle mapping in a forestry ecosystem in the North-eastern part of Italy and Southern Austria, and in [26], for forest fire detection in both southern Georgia, USA, and southern Greece.

The modified chlorophyll absorption in reflectance index (MCARI) [27] was adopted for estimating the leaf chlorophyll concentration from leaf and canopy reluctance. It combines spectral data enclosed in the Sentinel-2 bands Red, Green, and Red Edge 1 (RE1) bands

$$\text{MCARI} = ((\text{RE1} - \text{Red}) - 0.2 * (\text{RE1} - \text{Green})) * \frac{\text{RE1}}{\text{Red}}. \quad (3)$$

The index is also considered in [3] for mapping the bark beetle infestation in a forestry region in Northern Italy.

Fig. 3 shows the box plots of the three indices NGDRI, NMDI, and MCARI for the two opposite ground-truth classes “healthy” and “damaged.” The plots are computed on the entire imagery set, as well as on the training and the working imagery set of this study. The box plots show that a good divergence can be observed between the opposite classes in the three spectral vegetation indices. The divergence is equally visible in the entire imagery set, as well as in the training and the working portion of the imagery set considered for this study. Notably, the greatest divergence between the damaged and healthy pixels can be observed in the NGDRI box plots. The conclusions of this visual analysis are also supported by the statistical results of the one-way ANOVA analysis reported in Table III. Fig. 4 shows the heatmaps of the three indices in a sample Sentinel-2 image of the working imagery set considered in this study. These plots show that both NGDRI and NMDI roughly separate the forest patches infested by the bark beetle from the healthy ones surrounding them better than MCARI.

Finally, Fig. 5 shows the results of the bivariate correlation analysis performed by computing the Spearman’s rank correlation coefficient over the collection of spectral bands and spectral vegetation indices. The Spearman’s rank correlation coefficient

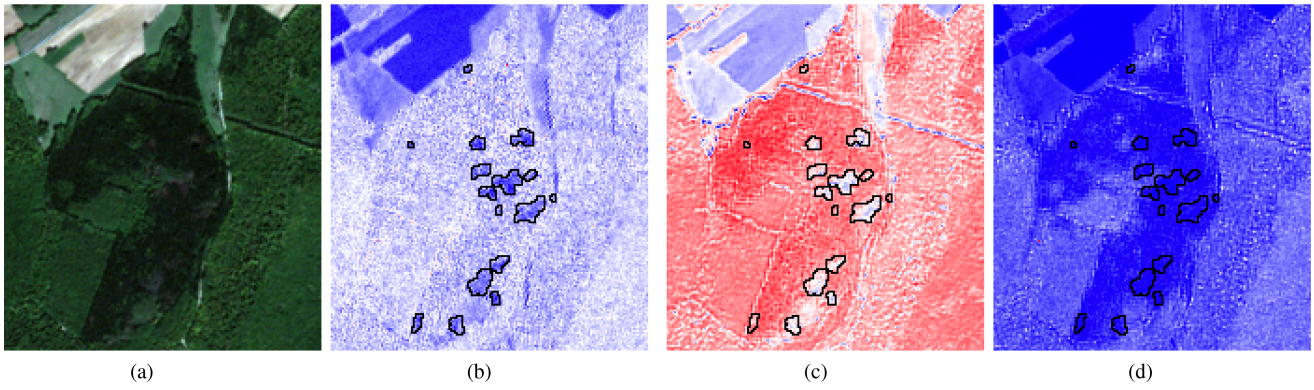


Fig. 4. (a) RGB, (b) NGDRI, (c) NMDI, and (d) MCARI of the sample image of the study working imagery set that is shown in the zoom of Fig. 1. The black-colored polygons delimit the ground-truth boundary of the bark beetle infestation patches. The heatmaps of NGDRI, NMDI, and MCARI are produced with the “bwr” color map. (a) RGB. (b) NGDRI. (c) NMDI. (d) MCARI.

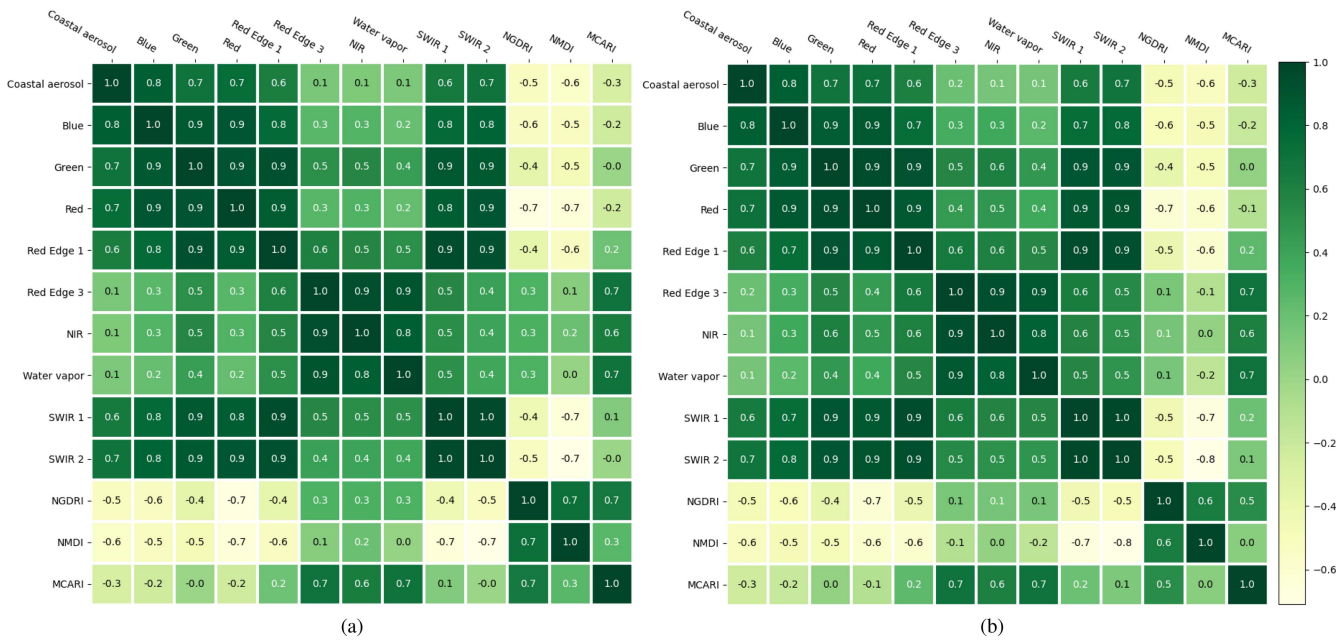


Fig. 5. Spearman’s rank correlation analysis performed on the spectral bands and spectral vegetation indices in both the (a) training imagery set and (b) working imagery set. (a) Training imagery set. (b) Working imagery set.

TABLE III
F METRIC RESULTS OF ONE-WAY ANOVA ANALYSIS PERFORMED ON THE SPECTRAL VEGETATION INDICES NGDRI, NMDI, AND MCARI ON THE ENTIRE IMAGERY SET, THE TRAINING IMAGERY SET, AND THE WORKING IMAGERY SET

Spectral band	Entire set	Training set	Working set
NGDRI	42896.52 (*)	37411.49 (*)	4068.45 (*)
NMDI	17249.09 (*)	15003.49 (*)	1227.25 (*)
MCARI	25864.72 (*)	22263.13 (*)	3243.99 (*)

(*) denotes that p -value ≤ 0.001 so that the null hypothesis (i.e., the means of the compared groups of pixels labeled with the two opposite ground truth classes “healthy” and “damaged” are equals) is rejected.

is a nonparametric measure of rank correlation that assesses how well the relationship between two compared variables can be described using a monotonic function. It varies between -1

and $+1$ with 0 implying no correlation, -1 implying an exact monotonic relationship with negative correlation and $+1$ implying an exact monotonic relationship with positive correlation. This correlation analysis shows that although the spectral bands are all positively correlated with each other, they can be grouped into two blocks. Block 1 contains: Coastal aerosol, Blue, Green, Red, Red Edge 1, SWIR 1, and SWIR 2, while Block 2 contains: Red Edge 3, NIR, and Water vapor. In particular, the intrablock positive correlation computed between each pair of spectral bands belonging to the same block is commonly greater than 0.6 , while the interblock correlation computed between each pair of spectral bands belonging to the two different blocks is commonly lower than 0.6 . In addition, we note that the interblock positive correlation computed between the Green and Red Edge 1 spectral bands of Block 1 and the spectral bands of Block 2 is greater than the interblock positive correlation measured

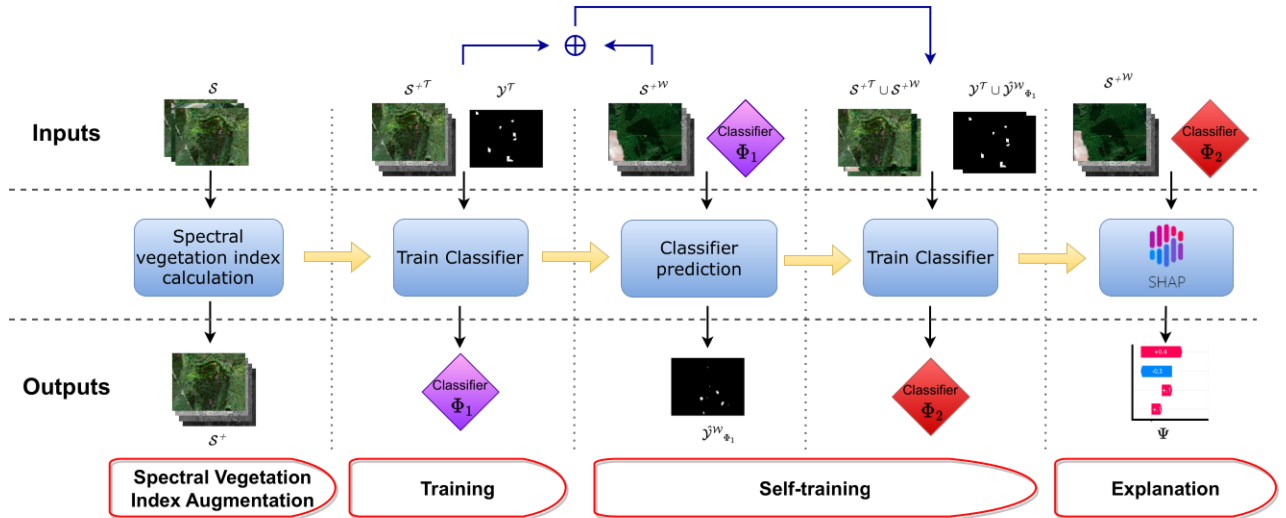


Fig. 6. Workflow of the framework SILVIA for mapping bark beetle infestation in unlabeled Sentinel-2 images of a forest located in the North-east of France by leveraging spectral vegetation indices, self-training with labeled images of the training imagery set and unlabeled images of the working imagery set, and XAI.

between the Coastal Aerosol, Blue, and Red spectral bands of Block 1 and the spectral bands of Block 2.

By extending this correlation analysis to the spectral vegetation indices, we note that each index is positively correlated with each other. In addition, both NGDRI and NMDI have a more negative correlation with spectral bands of Block 1, while they have a slight positive correlation, closer to 0, with spectral bands of Block 2. On the other hand, MCARI has a correlation closer to 0 with spectral bands of Block 1, while it has a more positive correlation with spectral bands of Block 2. Based on this analysis, we expect that MCARI may have a different effect than NGDRI and NMDI into the learning stage. To investigate this idea, we performed an ablation study of spectral vegetation indices, whose results are illustrated in Section VII-B, in order to explore the effect of several combinations of the three spectral vegetation indices on the performance of the spectral classifier trained in the learning stage.

VI. METHODS

Fig. 6 graphically shows the learning stage of the workflow of the self-training and spectral vegetation indices for spectral classification (SILVIA) framework experimented in this study. The learning framework resorts to a self-training strategy and leverages spectral vegetation indices to learn a spectral classifier from a set of partially labeled images (that comprise labeled training images and unlabeled working images). It uses the learned classifier to predict labels of working images. It also integrates an XAI component to explain classifier decisions. The evaluation framework computes several accuracy metrics by comparing pixel-wise ground-truth labels and predicted labels of masks of images in the working set.

A. Notation

Let \mathcal{S} denote the starting imagery set and \mathcal{Y} denote the starting mask set so that every binary mask $\mathbf{Y} \in \mathcal{S}$ is one-to-one

associated with a Sentinel-2 image $\mathbf{S} \in \mathcal{S}$. Let \mathcal{S}^+ the imagery set derived from \mathcal{S} by concatenating the spectral vector of every pixel $(i, j) \in \mathbf{S}$ of every image $\mathbf{S} \in \mathcal{S}$ with the vector of spectral vegetation indices computed from $\mathbf{S}(i, j)$ (see details in Section V). Let \mathbf{F} denote the vector of spectral bands and spectral vegetation indices recorded in every pixel section of images in \mathcal{S}^+ . Let $\mathcal{S}^{+\mathcal{T}}$ and $\mathcal{S}^{+\mathcal{W}}$ denote the training and the working imagery set so that $\mathcal{S}^{+\mathcal{T}} \cup \mathcal{S}^{+\mathcal{W}} = \mathcal{S}^+$ and $\mathcal{S}^{+\mathcal{T}} \cap \mathcal{S}^{+\mathcal{W}} = \emptyset$. Let $\mathcal{Y}^{\mathcal{T}}$ and $\mathcal{Y}^{\mathcal{W}}$ denote the training and working mask set so that $\mathcal{Y}^{\mathcal{T}}$ contains ground-truth masks of images comprised in $\mathcal{S}^{+\mathcal{T}}$ and $\mathcal{Y}^{\mathcal{W}}$ contains ground-truth masks of images comprised in $\mathcal{S}^{+\mathcal{W}}$.

B. Self-Training

In the learning stage, the spectral vegetation indices are computed and the self-training approach works on $\mathcal{S}^{+\mathcal{T}}$, $\mathcal{Y}^{\mathcal{T}}$, and $\mathcal{S}^{+\mathcal{W}}$ with a wrapper classification algorithm.

The self-training strategy, also known as decision-directed or self-taught machine learning, is a simple semisupervised learning approach [28] that accounts for supervisory information obtained from unlabeled data, by leveraging the underlying data structure to predict the unobserved or hidden part of the input. It enriches the labeled training data by training a new spectral classifier with labeled input in conjunction with pseudo-labeled samples for prediction. Cascante-Bonilla et al. [29] have recently shown that the use of pseudolabeling makes the classifier more resilient to out-of-distribution samples in the unlabeled set, while Zhang et al. [30] have proved that unlabeled data can improve training convergence.

The self-training step is performed with XGBoost [15] as a wrapper classification algorithm. This is a well-known tree-boosting algorithm that applies gradient boosting to couple weak tree classifiers and learn a strong classifier in an iterative manner. This is done by using residual errors computed for each tree

to decrease the error margin for the successor tree. In addition, XGBoost adopts a regularization term to prevent overfitting, and parallel computation to speed up the learning stage.

XGBoost is initially used to train a spectral classifier $\Phi_1 : \mathbf{F} \mapsto \{\text{healthy}, \text{damaged}\}$ by minimizing a regularized empirical loss over the collection of labeled pixels in the training imagery set $(\mathcal{S}^{+T}, \mathcal{Y}^T)$. Notice that the spectral input space \mathbf{F} of Φ_1 contains the concatenation of both the spectral bands and the spectral vegetation indices. The output space contains the “damaged” and “healthy” classes. Subsequently, Φ_1 is adopted to estimate the pseudolabels $\hat{\mathcal{Y}}_{\Phi_1}^W$ of the collection of pixels of \mathcal{S}^{+W} , i.e., $\hat{\mathcal{Y}}_{\Phi_1}^W = \Phi_1(\mathcal{S}^{+W})$. We note that $\hat{\mathcal{Y}}_{\Phi_1}^W$ is a prediction of \mathcal{Y}^W , while \mathcal{Y}^W is unavailable in the learning stage. Finally, the pseudolabeled pixels of \mathcal{S}^{+W} are leveraged to train a new classifier $\Phi_2 : \mathbf{F} \mapsto \{\text{healthy}, \text{damaged}\}$ by minimizing the regularized empirical loss of the wrapper classification algorithm over the entire imagery set $(\mathcal{S}^{+T} \cup \mathcal{S}^{+W}, \mathcal{Y}^T \cup \hat{\mathcal{Y}}_{\Phi_1}^W)$. Specifically, we use XGBoost to learn Φ_2 by taking advantage of supervision provided by ground-truth labels \mathcal{Y}^T for pixels of images collected in \mathcal{S}^{+T} and predicted pseudolabels $\hat{\mathcal{Y}}_{\Phi_1}^W$ for pixels of images collected in \mathcal{S}^{+W} .

XGBoost, similarly to most machine learning algorithms, expects training samples with a balanced class distribution. To handle the imbalanced condition of the study data, the two classifiers of the self-training step are trained by adopting a cost-sensitive approach [31]. This approach consists in learning a classifier by accounting for the costs assigned to confidence scores of different labels, i.e., the highest cost should be assigned to the minority class (“damaged”), while the lowest cost should be assigned to the majority class (“healthy”). For each classifier, the cost schema is automatically selected according to a fivefold stratified cross-validation of training samples by performing a grid search of the cost space.

Finally, Φ_2 is used to output final predictions $\hat{\mathcal{Y}}_{\Phi_2}^W = \Phi_2(\mathcal{S}^{+W})$, which represent the final predictions of infestation labels of masks associated with images in \mathcal{S}^{+W} .

C. Explainable Artificial Intelligence (XAI)

SHapley Additive exPlanations (SHAP) [32] is an explainer, to yield local and global explanations of inputs that contribute to decisions. SHAP has been recently used in [33] for interpretable land cover classification. In the proposed framework, SHAP is used to explain the effect of the input feature space \mathbf{F} , composed of both spectral bands and spectral vegetation indices, on classifier decisions. It can be used to explain the final decisions of the two classifiers Φ_1 and Φ_2 . These explanations allow us to identify which spectral information mainly conditions classifiers’ decisions, as well as to understand how classifier decisions may be changed through the self-training strategy.

SHAP is a local-level XAI technique that performs a theoretic game approach. It determines the contribution of each input feature to the prediction (relevance value) of a classifier as the average marginal contribution of a feature value for all possible predictions [34]. Let $\Phi : \mathbf{F} \mapsto \{\text{healthy}, \text{damaged}\}$ denote the classifier to be explained. In this study $\Phi = \Phi_1$ or $\Phi = \Phi_2$. For

each imagery spectral pixel section $\mathbf{p} \in \mathcal{S}^{+W}$ and for each feature $X \in \mathbf{F}$, SHAP measures the effect of X on the confidence scores estimated by Φ on \mathbf{p} for the opposite classes $y = \text{“healthy”}$ and $y = \text{“damaged,”}$ respectively. Let $\Phi(\mathbf{p})[y]$ denote the confidence score according to Φ sees \mathbf{p} assigned to class y with $y = \text{“healthy”}$ or $y = \text{“damaged.”}$ Notice that Φ , finally, decides to classify \mathbf{p} in the class for which the highest confidence score is computed. For each class y , SHAP measures the effect of X on $\Phi(\mathbf{p})[y]$, by comparing the confidence scores that are estimated by considering, for each subspace $\mathbf{X} \subseteq \mathbf{F}/\{X\}$, the pair of surrogate classifiers $\Phi_{\mathbf{X}} : \mathbf{X} \mapsto Y$ and $\Phi_{\mathbf{X} \cup \{X\}} : \mathbf{X} \cup \{X\} \mapsto Y$ trained with input feature spaces \mathbf{X} and $\mathbf{X} \cup \{X\}$, respectively. Specifically, for each subspace $\mathbf{X} \subseteq \mathbf{F}/\{X\}$, the confidence scores produced by $\Phi_{\mathbf{X}}$ and $\Phi_{\mathbf{X} \cup \{X\}}$ are compared according to the following formula:

$$\varphi_{\mathbf{X}, X, y}(\mathbf{p}) = \Phi_{\mathbf{X} \cup \{X\}}(\pi_{\mathbf{X} \cup \{X\}}(\mathbf{p})) [y] - \Phi_{\mathbf{X}}(\pi_{\mathbf{X}}(\mathbf{p})) [y] \quad (4)$$

where $\pi_{\mathbf{X}} : \mathbf{F} \mapsto \mathbf{X}$ and $\pi_{\mathbf{X} \cup \{X\}} : \mathbf{F} \mapsto \mathbf{X} \cup \{X\}$ denote the projection operators that return the spectral information enclosed in \mathbf{X} and $\mathbf{X} \cup \{X\}$, respectively.

Finally, Shapley values are computed, for each input dimension X , as a weighted average of all possible differences as follows:

$$\Psi_{X, y}(\mathbf{p}) = \sum_{\mathbf{X} \subseteq \mathbf{F}/\{X\}} \frac{|\mathbf{X}|! (|\mathbf{F}| - |\mathbf{X}| - 1)!}{|\mathbf{F}|!} (\varphi_{\mathbf{X}, X, y}(\mathbf{p})) \quad (5)$$

where $|\bullet|$ denotes the vector size. The higher the value of $\varphi_{X, y}(\mathbf{p})$, the greater the effect of X on the decision of assigning pixel \mathbf{p} to class y .

D. Accuracy Metrics

To explore the effectiveness of the SILVIA framework described, we performed an evaluation stage by measuring multiple accuracy metrics, i.e., Precision (P), Recall (R), and FScore (F) computed for the two opposite classes, false detection rate (FDR), missed alarm rate (MAR), overall accuracy (OA), average accuracy (AA), and GMean (G). Notice that AA and GMean metrics are commonly used in imbalanced classification problems. Let us consider: tp —the number of pixels of the imagery set in the “damaged” class that are correctly predicted as belonging to that class type; fp —the number of pixels not belonging to the “damaged” class that are wrongly predicted as belonging to the class “damaged”; tn —the number of pixels belonging to the “healthy” class that are correctly predicted as belonging to that class type; and fn —the number of pixels of the “damaged” class that are wrongly predicted as not belonging to that class type. The accuracy metrics were computed as follows.

- 1) P(h) and P(d) measure the precision (also called user accuracy) of the “healthy” and the “damaged” class, respectively, i.e., $P(h) = \frac{tn}{tn+fp}$ and $P(d) = \frac{tp}{tp+fn}$.
- 2) R(h) and R(d) measure the recall (also called producer accuracy) of the “healthy” and the “damaged” class, respectively, i.e., $R(h) = \frac{tn}{tn+fn}$ and $R(d) = \frac{tp}{tp+fp}$.
- 3) F(h) and F(d) measure the FScore, that is, the harmonic mean of precision and recall of the “healthy” and the

“damaged” class respectively, i.e., $F(h) = \frac{2P(h)R(h)}{P(h)+R(h)}$ and

$$F(d) = \frac{2P(d)R(d)}{P(d)+R(d)}.$$

- 4) FDR measures the probability that a pixel belonging to the “healthy” class is wrongly classified in the “damaged” class, i.e., $FDR = \frac{fp}{fp+tp}$.
- 5) MAR measures how many pixels belonging to the “damaged” class are wrongly predicted in the “healthy” class on the number of pixels predicted in the “damaged” class, i.e., $MAR = \frac{fn}{fn+tp}$.
- 6) OA measures the proportion of correctly classified pixels, i.e., $OA = \frac{tp+tn}{tp+tn+fp+fn}$.
- 7) AA measures the average of each accuracy value per class, i.e., $AA = \frac{1}{2}(\frac{tp}{tp+fn} + \frac{tn}{tn+fp})$.
- 8) GMean measures the geometric mean of *specificity* and *recall* by equally considering the errors in both classes, i.e., $G = \sqrt{\text{specificity} \times \text{sensitivity}}$, where *specificity* measures how many pixels are correctly predicted for the “healthy” class given all occurrences of that class, i.e., $\text{specificity} = \frac{tn}{tn+fp}$, while *sensitivity* measures how many pixels are correctly predicted for the “damaged” class given all occurrences of that class type, i.e., $\text{sensitivity} = \frac{tp}{tp+fn}$.

All the accuracy metrics were computed on predictions produced for pixels in the working imagery set of this study.

VII. RESULTS AND DISCUSSION

The SILVIA framework was implemented in Python 3.8. The evaluation study was conducted on a Linux machine with an Intel(R) Core(TM) i7-9700F CPU at 3.00 GHz and 32-GB RAM. All the experiments were executed on a single GeForce RTX 2080.

A. Spectral Vegetation Indices and Self-Training

We performed an ablation study, to explore the effect of both spectral vegetation indices and self-training on spectral classifiers trained with XGBoost in SILVIA. To this end, we analyzed the performance of spectral classifiers trained in the following configurations of SILVIA.

- 1) S, which discards both the spectral vegetation indices and the self-training strategy, and maps the bark beetle infestation by using the spectral classifier trained by the spectral data of the study training imagery set.
- 2) S + SELF, which discards the spectral vegetation indices, and maps the bark beetle infestation by using the spectral classifier trained by the labeled spectral data of the training imagery set and the unlabeled spectral data of the working imagery set through the self-training strategy.
- 3) S + I, which discards the self-training strategy, and maps the bark beetle infestation by using the spectral classifier trained by both the spectral data and the spectral vegetation index data of the training imagery set.
- 4) S + I + SELF, which is the default configuration of the SILVIA framework as described in Section VI.

Fig. 7 compares the time spent (in minutes) completing the computation process with configurations S, S + SELF, S + I,



Fig. 7. Computation time (in minutes) of configurations S, S + SELF, S + I, and S + I + SELF of SILVIA.

and S + I + SELF, while Table IV reports the accuracy metrics measured on the bark beetle maps produced for the working imagery set with S, S + SELF, S + I, and S + I + SELF, respectively. As expected, performing the self-training strategy is more time consuming than accounting for spectral vegetation indices in the input space of classifiers. Precision, Recall, and FScore results show that the use of spectral vegetation indices and the self-training strategy allow us to improve the precision of the “healthy” class $P(h)$ and the recall of the “damaged” class $R(d)$, while they reduce the precision of the “damaged” class $P(d)$ and the recall of the “healthy” class $R(h)$. These metrics show that a better ability to recognize infested patches is achieved at the cost of a higher number of healthy pixels being wrongly assigned to the damaged patch. This consideration is also supported by the analysis of both MAR and FDR. In fact, both performing self-training and leveraging spectral vegetation indices allow us to reduce the MAR by diminishing the number of pixels infested by the bark beetle, but missed by the classifier. In any case, a better ability to recognize infested patches is achieved at the cost of a higher number of healthy pixels being wrongly assigned to the damaged patch. In fact, the FDR increases with the use of both self-training and spectral vegetation indices. Notably, learning the spectral classifier through the self-training strategy (S + SELF) causes a higher number of false alarms than training the classifier by leveraging the spectral vegetation indices (S + I). On the other hand, we highlight that the decrease of MAR is highly desirable in this problem, also at the cost of an equal increase of FDR, considering that the “damaged” class is very imbalanced in the study problem, and from the point of view of forest management, a missed alarm is more expensive than a false alarm.

Regarding OA, AA, and GMean, we note that OA decreases negligibly, but OA is dominated by the majority “healthy” class. On the other hand, both AA and GMean improve significantly due to the use of self-training and spectral vegetation indices. We recall that AA gives equal weight to the accuracy achieved in both classes, so it is more appropriate than OA for measuring improvements in the study problem without being conditioned by the imbalanced phenomenon. Similarly, GMean is commonly considered a more appropriate measure of performance than OA in imbalanced classification problems, as it combines sensitivity and specificity with equal weights. Sensitivity might be more interesting than specificity in the imbalanced problems. Notably,

TABLE IV
ACCURACY METRICS OF S, S + SELF, S + I, AND S + I + SELF

Configuration	P(h)	R(h)	F(h)	P(d)	R(d)	F(d)	FDR	MAR	OA	AA	GMean
S	0.9950	0.9944	0.9947	0.6186	0.6445	0.6313	0.3814	0.3554	0.9896	0.8195	0.8006
S+SELF	0.9958	0.9921	0.9940	0.5577	0.7055	0.6229	0.4423	0.2945	0.9882	0.8489	0.8367
S+I	0.9956	0.9930	0.9943	0.5819	0.6901	0.6314	0.4181	0.3099	0.9889	0.8416	0.8279
S+I+SELF	0.9961	0.9912	0.9936	0.5364	0.7289	0.6180	0.4635	0.2710	0.9876	0.8601	0.8500

The best results are in bold.

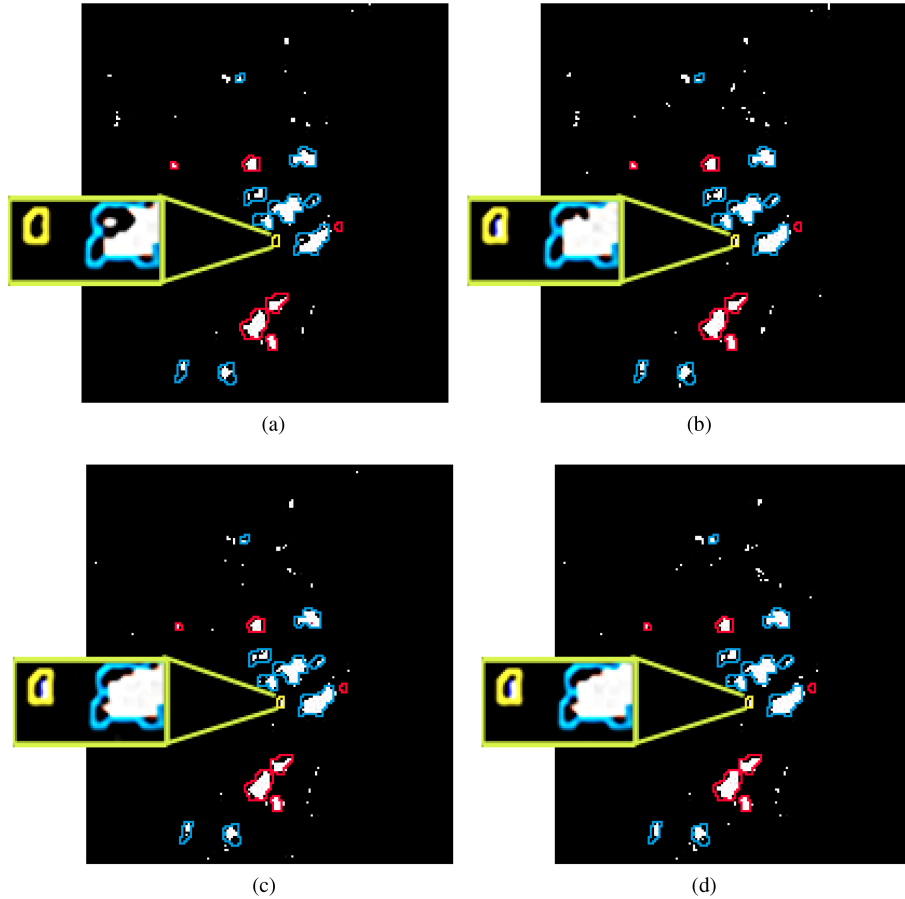


Fig. 8. Bark beetle maps of the sample image of the study working imagery set that is shown in the zoom of Fig. 1. The maps were produced with (a) S, (b) S + SELF, (c) S + I, and (d) S + I + SELF. The white pixels are the ones predicted in the “damaged” class. The red, blue, and yellow polygons delimit the boundary of the ground-truth hotspots of the bark beetle infestation in this sample image. No pixel in the yellow polygon was correctly detected by S as infested by the bark beetle. The number of infested pixels in the blue polygons correctly detected by S + SELF, S + I, and S + I + SELF is higher than the number discovered by S. (a) S. (b) S + SELF. (c) S + I. (d) S + I + SELF.

a greater improvement in both AA and GMean is achieved by coupling self-training with vegetation indices.

The conclusions drawn from the accuracy performance analysis can be visually supported by the exploration of the bark beetle maps that are produced with the spectral classifiers, trained in the configurations compared. Fig. 8 shows the bark beetle maps produced by configurations S, S + SELF, S + I, and S + I + SELF in the sample image of the study’s unlabeled imagery set already considered in Fig. 4. Notably, both the self-training strategy and the spectral vegetation indices allow us to discover a small forest patch infested by the bark beetle that is otherwise ignored [i.e., the yellow polygon in Fig. 8(a)]. On the other hand, both the self-training strategy and the spectral vegetation indices allow

us to fill in the surface of the forest patches infested by the bark beetle better [i.e., the blue polygons in Fig. 8(a)].

B. Spectral Vegetation Index Sensitivity

We performed a sensitivity study to explore the effect of the different combinations of spectral vegetation indices on the accuracy of the spectral classifiers trained accounting for the selected indices. For this purpose, we considered the spectral classifiers trained in the two configurations of SILVIA that account for the vegetation indices: (S + I) (that discards the self-training strategy) and S + I + SELF (that uses the self-training strategy).

TABLE V
ACCURACY METRICS OF S + I AND S + I + SELF BY VARYING THE COMBINATION OF SPECTRAL VEGETATION INDICES CONSIDERED

Configuration	Vegetation indices	P(h)	R(h)	F(h)	P(d)	R(d)	F(d)	FDR	MAR	OA	AA	GMean
S+I	-	0.9950	0.9944	0.9947	0.6186	0.6445	0.6313	0.3814	0.3554	0.9896	0.8195	0.8006
	NGDRI	0.9955	0.9931	0.9943	0.5801	0.6807	0.6264	0.4199	0.3193	0.9888	0.8369	0.8222
	NMDI	0.9944	0.9958	0.9951	0.6656	0.5993	0.6307	0.3344	0.4007	0.9903	0.7976	0.7725
	MCARI	0.9937	0.9969	0.9953	0.7096	0.5484	0.6187	0.2904	0.4516	0.9907	0.7726	0.7394
	NGDRI, NMDI	0.9956	0.9929	0.9942	0.5740	0.6831	0.6238	0.4260	0.3169	0.9887	0.8380	0.8236
	NGDRI, MCARI	0.9954	0.9931	0.9943	0.5784	0.6737	0.6224	0.4216	0.3263	0.9888	0.8334	0.8180
	NMDI, MCARI	0.9951	0.9943	0.9947	0.6142	0.6476	0.6305	0.3858	0.3524	0.9896	0.8209	0.8024
	NGDRI, NMDI, MCARI	0.9956	0.9930	0.9944	0.5819	0.6901	0.6314	0.4181	0.3099	0.9889	0.8416	0.8279
S+I+SELF	-	0.9958	0.9921	0.9940	0.5577	0.7055	0.6229	0.4423	0.2945	0.9882	0.8489	0.8367
	NGDRI	0.9959	0.9923	0.9941	0.5601	0.7045	0.6240	0.4399	0.2955	0.9883	0.8484	0.8361
	NMDI	0.9953	0.9941	0.9947	0.6122	0.6626	0.6364	0.3878	0.3374	0.9896	0.8284	0.8116
	MCARI	0.9951	0.9949	0.9950	0.6399	0.6506	0.6452	0.3601	0.3494	0.9902	0.8227	0.8045
	NGDRI, NMDI	0.9951	0.9943	0.9947	0.6157	0.6489	0.6319	0.3843	0.3511	0.9896	0.8216	0.8033
	NGDRI, MCARI	0.9960	0.9913	0.9936	0.5334	0.7159	0.6114	0.4666	0.2841	0.9875	0.8536	0.8424
	NMDI, MCARI	0.9954	0.9940	0.9947	0.6070	0.6690	0.6365	0.3930	0.3310	0.9895	0.8315	0.8155
	NGDRI, NMDI, MCARI	0.9961	0.9912	0.9936	0.5364	0.7289	0.6180	0.4635	0.2710	0.9876	0.8601	0.8500

“-” denotes classifiers trained by neglecting the spectral vegetation indices. The best results are in bold.

TABLE VI
ACCURACY METRICS OF XGBOOST, RANDOM FOREST, AND SVM

Classifier	P(h)	R(h)	F(h)	P(d)	R(d)	F(d)	FDR	MAR	OA	AA	GMean
XGBoost	0.9961	0.9912	0.9936	0.5364	0.7289	0.6180	0.4635	0.2710	0.9876	0.8601	0.8500
Random Forest	0.9946	0.9947	0.9947	0.6220	0.6164	0.6192	0.3779	0.3836	0.9896	0.8056	0.7831
SVM	0.9954	0.9936	0.9945	0.5989	0.6753	0.6348	0.4011	0.3246	0.9893	0.8345	0.8192

The best results are in bold.

The results reported in Table V show that MCARI allows us to better delineate the healthy forest patches causing an increase in the amount of missed alarms, while NGDRI and NMDI allow us to better delineate infested forest patches causing an increase in the number of false alarms. In fact, the configuration of both S + I and S + I + SELF with MCARI achieves the highest R(h) and P(d), as well as the lowest FDR. The configuration with NGDRI and NMDI achieves the highest P(h) and R(d), as well as the lowest MAR. NGDRI is more effective than NMDI at recognizing infested forest patches. Finally, we note that, although MCARI elaborated alone helps in recognizing the healthy forest patches better than the infested forest patches, the spectral classifier trained leveraging NGDRI, NMDI, and MCARI achieves the highest P(h), R(d), AA, and GMean of this study. This result is welcome in a scenario characterized by a strong imbalance of the minority class “damaged.” This behavior of spectral vegetation indices is equally observed regardless the use of the self-training strategy.

C. Wrapper Classification Algorithm

We analyze the performance of the wrapper classification algorithm by comparing the performance of SILVIA with XGBoost to the performance of SILVIA with Random Forest and SVM. These classification algorithms are selected as they have been recently used in [1], [3], [9], and [10] to address various bark beetle mapping problems in different European forestry regions. As for XGBoost, we adopt a cost-based approach to handle the imbalanced condition of the study data in both Random Forest and SVM. Similarly to XGBoost, the cost schema of each classifier is selected with a grid-search performed on the fivefold cross-validation of samples processed to train the classifier with either Random Forest or SVM.

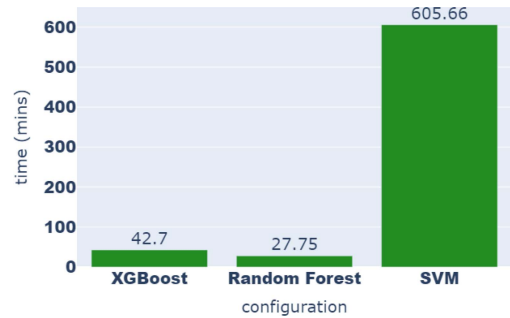


Fig. 9. Computation time (in minutes) of XGBoost, Random Forest, and SVM.

Fig. 9 compares the time that the SILVIA framework spent (in minutes) completing the computation process with XGBoost, Random Forest, and SVM, respectively. Table VI reports the accuracy metrics measured on the bark beetle infestation maps yielded by the three wrapper classification algorithms. These results show that Random Forest achieves the lowest FDR, which results in a small gain in both R(h) and P(d), as well as OA. On the other hand, XGBoost achieves the lowest MAR that results in a significant gain in both AA and GMean. Notably, the SVM achieves a good tradeoff between FDR and MAR, as well as between F(h) and F(d), in this study, and it is the runner-up in terms of OA, AA, and GMean. However, wrapping the SVM into SILVIA is much more time consuming than wrapping XGBoost or Random Forest.

D. XAI Exploration

We perform the analysis of the explanations that the XAI module of SILVIA produced. For this purpose, we analyze separately

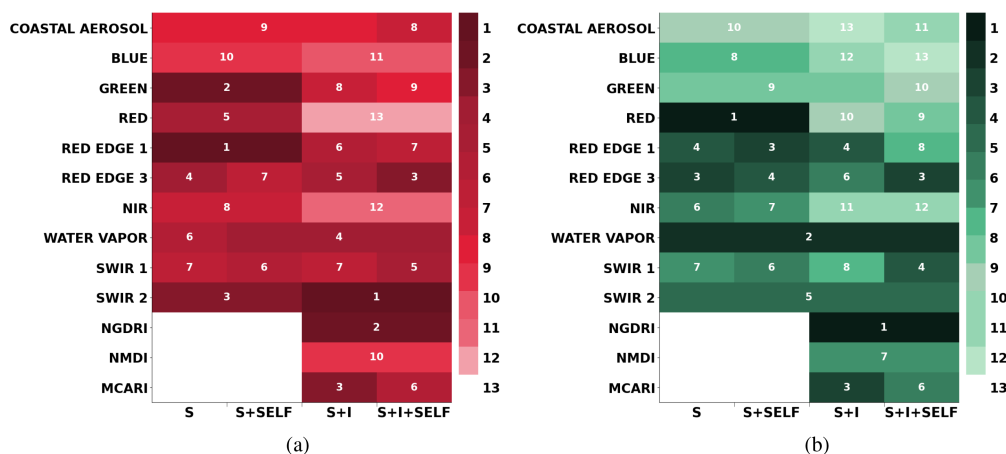


Fig. 10. Average Shapley vectors of the spectral input space of the classifiers learned with configurations S, S + SELF, S + I, and S + I + SELF, respectively. (a) Damaged. (b) Healthy.

the average Shapley vector of the input space of each spectral classifier on pixels belonging to the two opposite classes.

Fig. 10(a) and (b) compares the maps of the average Shapley vectors generated for the “healthy” and “damaged” classes in the configurations S, S + SELF, S + I, and S + I + SELF, respectively. These Shapley maps show that the Coastal Aerosol spectral band has, in general, little effect on all the spectral classifiers trained in this problem independently of the class to be recognized. This outcome is expected since Coast Aerosol commonly conveys useful information for mapping a coastal habitat [35], but no coastal habitat was covered by this study area. On the other hand, the Shapley maps show that the effect of the spectral input on the decisions yielded by the same spectral classifier may change according to the classes. For example, SWIR 2, which is one of the top-relevant spectral bands of this study, has greater effect on decisions concerning pixels infested by the bark beetle than decisions concerning pixels not infested by the bark beetle, regardless of the computation of the spectral vegetation indices or the use of the self-training strategy. On the other hand, SWIR 2 has the greatest effect on decisions concerning forest patches infested by the bark beetle when this band is processed with spectral vegetation indices, while Red Edge 1 has the greatest effect on decisions concerning the same patches when spectral vegetation indices are neglected for training the classifier. Notably, this outcome is in line with the analysis reported in [2], which highlights the importance of SWIR and Red Edge spectral bands for recognizing a forest patch infested by insects, and particularly, by bark beetles. In addition, our XAI study shows that SWIR 2 also appears among the five most relevant spectral bands considered by each classifier to recognize healthy forest patches. On the other hand, the Red Edge spectral bands appear among the four most relevant spectral features for recognizing healthy forest patches with classifiers trained by leveraging the original Sentinel-2 data and in the eighth most relevant features to recognize healthy forest patches with classifiers trained leveraging the spectral vegetation indices. This confirms the relevance of this band when additional information disclosed by the spectral vegetation indices is kept away in this study. Focusing attention

on the healthy forest patches Water Vapor is consistently the second most important feature, regardless of the computation of the spectral vegetation indices or the use of self-training. This outcome shows that Water Vapor can be a relevant indicator of the presence of healthy forest patches.

Further considerations concern how the decisions of classifiers change according to the computation of the spectral vegetation indices and the use of the self-training strategy. Using the self-training strategy to train the classifier introduces only small changes in how the spectral input space conditions the classifier decisions. Therefore, this XAI study clarifies that, even though the self-training strategy is able to estimate the parameters of a spectral classifier in such a way as to delimit the infested patches better, it does not change significantly the relevance of the dimensions of the input feature space on the classifier decisions. On the contrary, the computation of the spectral vegetation indices has a greater effect on how the various dimensions of the input feature space contribute to the classifier decisions. In fact, both NGDRI and MCARI are among the six most important dimensions of the input space of a classifier, regardless of the use of the self-training strategy, and the specific class considered. Particularly, the consideration of NGDRI, which is a combination of Red and Green, causes a decrease in the relevance of the Red and Green bands in decisions on forest patches belonging to both classes. The final comments concern the NIR band. The XAI analysis in this study reveals that the NIR band results are less appropriate for bark beetle detection, even though the evidence in [2] highlighted significant differences in NIR values between healthy and infested trees. However, our conclusion on NIR band is in line with recent achievements that have reported in [1] the poor performance of NIR for early bark beetle detection.

Let us focus on the spectral classifier trained in the S + I + SELF configuration of SILVIA. Fig. 11 plots the Shapley value, measured for each input dimension (Y-axis), with respect to the input dimension value (X-axis), for both pixels of the healthy forest patches (green pixels) and pixels of the forest patches infested by the bark beetle (red pixels). For example, this plot shows that Shapley values of NGDRI follow an opposite

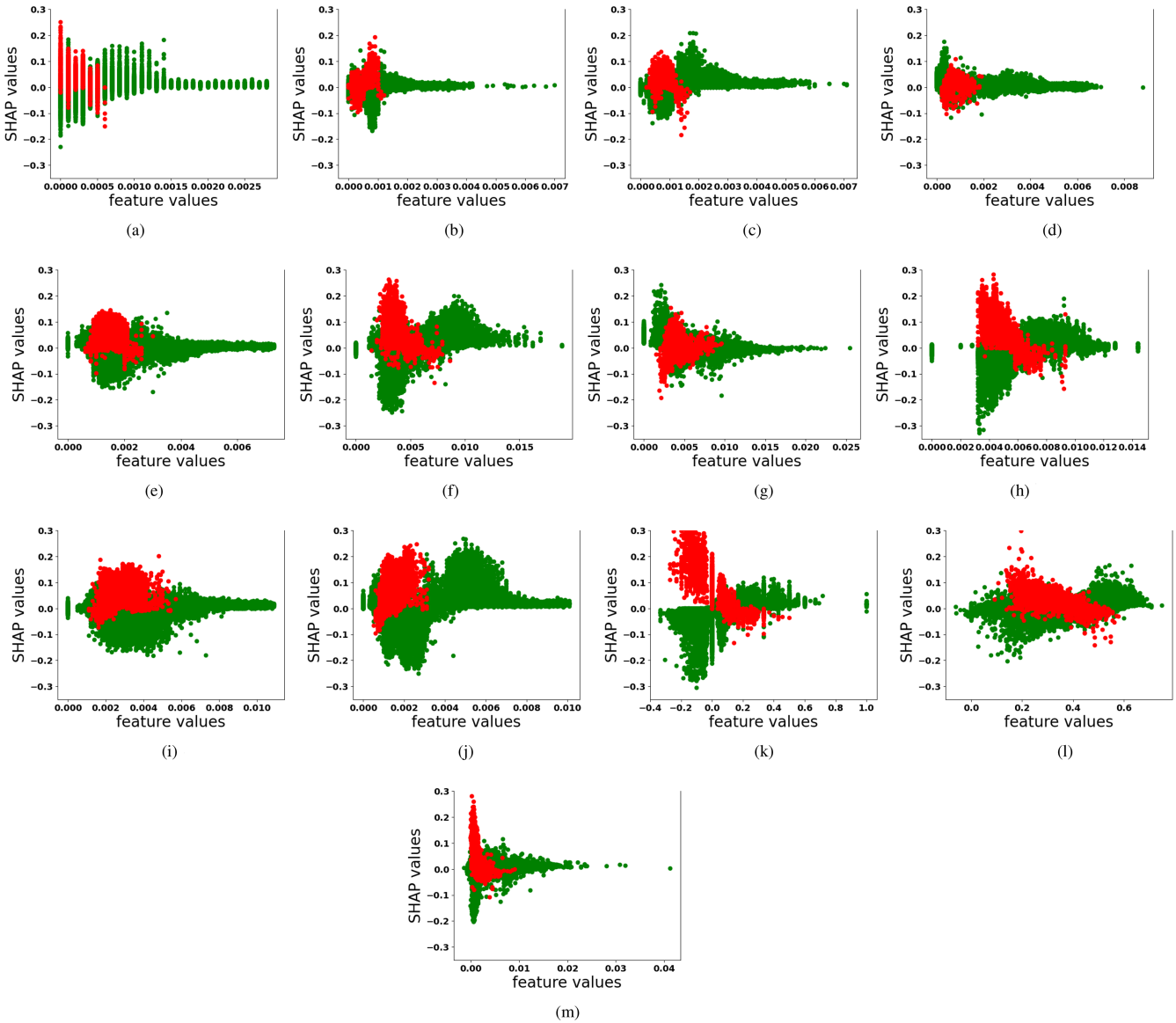


Fig. 11. Plots of local Shapley values (Y-axis) measured for the spectral classifier trained in the default configuration $S + I + SELF$ of SILVIA for each input dimension of the spectral classifier (spectral band or vegetation index on the X-axis) for both pixels belonging to the healthy forest patches (green pixels) and pixels belonging to forest patches infested by the bark beetle (red pixels). (a) Coastal Aerosol. (b) Blue. (c) Green. (d) Red. (e) Red Edge 1. (f) Red Edge 3. (g) NIR. (h) Water Vapor. (i) SWIR 1. (j) SWIR 2. (k) NGDRI. (l) NMDI. (m) MCARI.

trend on pixels belonging to opposite classes. Specifically, the lower the value of NGDRI, the greater the effect of NGDRI on decisions concerning pixels belonging to forest patches infested by the bark beetle. On the other hand, the higher the value of Water Vapor or NGDRI, the greater the relevance of NGDRI for decisions on pixels belonging to healthy forest patches. A similar trend is observed in Shapley values of Water Vapor. However, Shapley values of NIR show that similar values of NIR may result in an opposite effect of this dimension on decisions concerning pixels in the same class. We observe that Shapley values of NIR are distributed approximately symmetrically around the X-axis for both pixels belonging to forest patches infested by the bark beetles and healthy forest patches. This supports the conclusions drawn previously on the poor performance of NIR for early bark beetle detection.

Finally, Fig. 12(a) and (b) compares the maps of the average Shapley vectors generated for the “healthy” and “damaged” classes in the configuration $S + I + SELF$ of SILVIA with XGBoost, Random Forest, and SVM, as wrapper classification algorithms, respectively. We note that NGDRI is among the two most important dimensions of the input space of both classes, regardless of the classification algorithm. On the other hand, Coastal Aerosol, Blue, and Green are the less relevant dimensions of the input spectral space for both classes, regardless of the classification algorithm. Instead, the NIR band that is among the least appropriate dimension for the bark beetle detection with the spectral classifier trained with XGBoost gains importance with Random Forest and SVM. In particular, the NIR band is the most important input dimension of the SVM for deciding on pixels in the healthy forest patches, while it is the

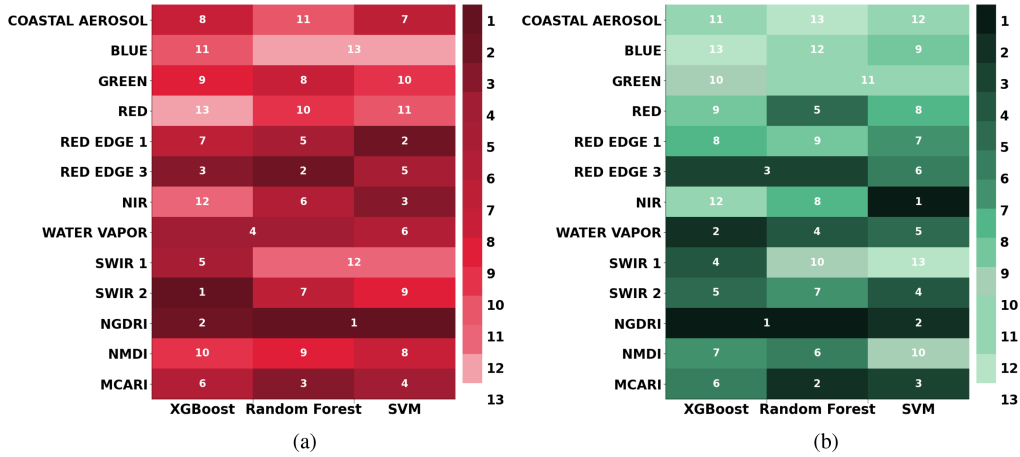


Fig. 12. Average shapley vectors of the spectral input space of XGBoost, Random Forest, and SVM. (a) Damaged. (b) Healthy.

TABLE VII
ACCURACY METRICS OF SPECTRAL CLASSIFIERS TRAINED IN THE CONFIGURATION S + I + SELF OF SILVIA BY SELECTING THE TOP- k FEATURES (SPECTRAL BANDS AND SPECTRAL VEGETATION INDICES) SORTED ACCORDING TO THE RESULTS OF THE AVERAGE SHAPLEY VALUE ANALYSIS

k	P(h)	R(h)	F(h)	P(d)	R(d)	F(d)	FDR	MAR	OA	AA	GMean
13	0.9961	0.9912	0.9936	0.5364	0.7289	0.6180	0.4635	0.2710	0.9876	0.8601	0.8500
12	0.9956	0.9933	0.9944	0.5852	0.6821	0.6300	0.4148	0.3179	0.9890	0.8377	0.8231
11	0.9959	0.9920	0.9940	0.5542	0.7092	0.6222	0.4458	0.2908	0.9881	0.8506	0.8388
10	0.9956	0.9936	0.9946	0.5986	0.6831	0.6381	0.4014	0.3169	0.9893	0.8383	0.8238
9	0.9956	0.9935	0.9945	0.5942	0.6827	0.6354	0.4058	0.3173	0.9892	0.8381	0.8236
8	0.9959	0.9924	0.9941	0.5648	0.7069	0.6279	0.4352	0.2931	0.9885	0.8496	0.8376
7	0.9957	0.9923	0.9940	0.5545	0.6908	0.6152	0.4455	0.3092	0.9881	0.8415	0.8279
6	0.9954	0.9930	0.9942	0.5725	0.6744	0.6193	0.4275	0.3256	0.9886	0.8337	0.8183
5	0.9952	0.9926	0.9939	0.5554	0.6586	0.6026	0.4446	0.3414	0.9880	0.8256	0.8086

The best results are in bold.

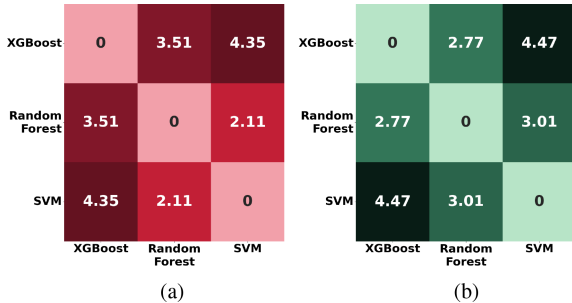


Fig. 13. Euclidean distance computed between the Shapley rank vectors of the inputs space of decisions produced in SILVIA with the wrapper classification algorithms: XGBoost, Random Forest, and SVM, for the opposite labels: "healthy" and "damaged." (a) Damaged. (b) Healthy.

third most important input dimension of the SVM for deciding on pixels in the forest patches infested by the bark beetle.

To complete this analysis, Fig. 13 shows the maps of the Euclidean distances computed between the Shapley rank vectors computed on the dimensions of the input space of decisions produced with XGBoost, Random Forest, and SVM for the opposite labels: "healthy" and "damaged." These maps show that XGBoost and Random Forest are the most similar in terms of the relevance that the different dimensions of the input space have on decisions concerning pixels in the forest patches infested

by the bark beetle. On the other hand, Random Forest and SVM are the most similar in terms of the relevance that the different dimensions of the input space have on decisions concerning pixels in the healthy forest patches.

E. XAI-Based Feature Sensitivity

In this section, we explore to what extent it would be meaningful to train the spectral classifier on a reduced set of spectral bands and indices, which correspond to the best performing ones according to the XAI analysis. For this purpose, we sorted features according to their average Shapley values and trained a spectral classifier with the top- k features. We performed this study considering the spectral classifier trained with XGBoost in the configuration S + I + SELF of SILVIA.

The results reported in Table VII show that the highest P(h), R(d), AA, and GMean, and the lowest MAR are achieved when all spectral bands and spectral vegetation indices are used. On the other hand, the highest OA and the lowest FDR are achieved when the top-ten features are used for training the spectral classifier. In particular, this spectral classifier was trained through self-training by discarding Blue, Red, and NIR according to the XAI ranking. Notably, the spectral classifier trained with the top-ten features is also the classifier that achieves the highest FScore in the opposite classes F(h) and (F(d)).

TABLE VIII
ACCURACY PERFORMANCE OF THE CONFIGURATION S + I + SELF OF SILVIA, U – Net, AND FCN – 8

Configuration	P(h)	R(h)	F(h)	P(d)	R(d)	F(d)	FDR	MAR	OA	AA	GMean
SILVIA	0.9962	0.9912	0.9940	0.5365	0.7290	0.6181	0.4635	0.2710	0.9876	0.8601	0.8500
U-Net	0.9960	0.9889	0.9924	0.4701	0.7089	0.5653	0.5298	0.2911	0.9850	0.8488	0.8372
FCN-8	0.9956	0.9825	0.9890	0.3530	0.6831	0.4654	0.6470	0.3169	0.9784	0.8328	0.8192

The best results are in bold.

F. Deep Learning Competitors

Finally, considering the recent boom of deep learning in remote sensing, we compare the accuracy of the methodology considered in this study with two state-of-the-art deep neural networks for semantic segmentation, i.e., U-Net [36] and FCN-8 [37]. Both U-Net and FCN-8 are convolutional neural networks. The U-Net was developed for biomedical image segmentation and has been recently used in several remote sensing problems (e.g., [38] and [39]). The FCN-8 has also recently been adopted various remote sensing tasks (e.g., [40] and [41]). In this study, we used standard implementations of both U-Net⁶ and FCN-8.⁷ In both cases, we adopted the Tversky loss to address the issue of data imbalance [42].

Table VIII reports the accuracy metrics measured on the bark beetle infestation maps, produced with the S + I – SELF configuration of the SILVIA framework, the U-Net, and the FCN-8. The results show that, even though SILVIA performs a pixel-wise approach that neglects the spatial arrangement of pixels, it can outperform both U-Net and FCN-8 along all measured metrics.

VIII. CONCLUSION

In this study, we explore the effect of performing the self-training strategy and leveraging spectral vegetation indices on a spectral classifier, trained for mapping bark beetle infestation in Sentinel-2 images, acquired in October 2018 in a forestry area in the North-east of France. The spectral classifier is trained using the XGBoost algorithm that allows us to find a good tradeoff between accuracy and efficiency requirements. A cost-based approach is adopted in the classifier training to handle the imbalanced condition. Finally, we use a local XAI technique—SHAP—to explain what the effect of the input spectral data is on the classifier decisions, as well as to understand how the use of self-training and spectral vegetation indices change the classifier decisions.

As a future development, we plan to continue the research in this direction, by processing the time series of Sentinel-2 images and using XAI to understand how the temporal spectral information may have an effect on the ability of a spectral classifier to recognize forest patches infested at different stages of bark beetle attack. In addition, we plan to explore possible transfer learning techniques [43] to update the spectral classifier trained in a specific geographic area or in a specific period to different geographic areas or periods. We also intend to explore the trasferability of a spectral classifier trained for mapping

forest tree die-back hotspots caused by bark beetle infestation, to perform the inventory of tree die-back hotspots caused by different families of fungal forest pathogens. As further future work, we plan to go spectral-spatial classifiers into detail (e.g., [44]), as well as explore the achievements of the proposed methodology in mapping various risks (e.g., windthrow forest damage and forest fires risks) across various regions of Europe. Finally, we plan to extend this study by focusing on the more challenging task of recognizing the different stages of bark beetle infestation to prompt the early detection of new infestations before the beetles swarm. In fact, cutting down affected trees during the early stages may help mitigating the diffusion of infestation. However, as recently discussed in [5], optical systems, comprising Sentinel-2 based systems, facilitate detection of infested trees beyond the visible range (e.g., subtle decreases in chlorophyll and water content of the needles), while terrestrial monitoring may be more advantageous for detecting early crown discoloration caused by bark beetle infestation. Hence, a future direction of this study is that of extending the proposed methods to handle both optical and terrestrial monitoring data and exploring the performances of these methods in the early detection task.

ACKNOWLEDGMENT

The authors would like to thank Wildsense for preparing the data used in this study, L. Rudd for her help in reading the manuscript and anonymous reviewers for helpful suggestions they provided to improve this article.

Authors' Contribution: Giuseppina Andresini: Conceptualization, methodology, software, data curation, investigation, validation, visualization, writing—original draft, writing—review and editing; Annalisa Appice: Conceptualization, methodology, software, investigation, validation, supervision, writing—original draft, writing—review and editing; and Donato Malerba: Conceptualization, writing—review and editing.

REFERENCES

- [1] V. Bárta, P. Lukeš, and L. Homolová, "Early detection of bark beetle infestation in Norway spruce forests of central Europe using Sentinel-2," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 100, 2021, Art. no. 102335.
- [2] H. Abdullah, A. K. Skidmore, R. Darvishzadeh, and M. Heurich, "Sentinel-2 accurately maps green-attack stage of European spruce bark beetle (*Ips typographus*, L.) compared with Landsat-8," *Remote Sens. Ecol. Conservation*, vol. 5, no. 1, pp. 87–106, 2019.
- [3] M. Dalponte, Y. T. Solano-Correa, L. Frizzera, and D. Gianelle, "Mapping a European spruce bark beetle outbreak using Sentinel-2 remote sensing data," *Remote Sens.*, vol. 14, no. 13, 2022, Art. no. 3135.
- [4] A. Fernandez-Carrillo, Z. Patočka, L. Dobrovolný, A. Franco-Nieto, and B. Revilla-Romero, "Monitoring bark beetle forest damage in central Europe. A remote sensing approach validated with field data," *Remote Sens.*, vol. 12, no. 21, 2020, Art. no. 3634.

⁶[Online]. Available: <https://github.com/karolzak/keras-unet/tree/master>

⁷[Online]. Available: <https://github.com/naineshhulke/keras-fcn-segmentation-model>

- [5] M. Kautz, F. Peter, L. Harms, S. Kammen, and H. Delb, "Patterns, drivers and detectability of infestation symptoms following attacks by the European spruce bark beetle," *J. Pest Sci.*, vol. 96, pp. 403–414, 2022.
- [6] R. Duraciová et al., "A bark beetle infestation predictive model based on satellite data in the frame of decision support system TANABBO," *iForest - Biogeosci. Forestry*, no. 3, pp. 215–223, 2020.
- [7] M. Kautz, K. Dworschak, A. Gruppe, and R. Schopf, "Quantifying spatio-temporal dispersion of bark beetle infestations in epidemic and non-epidemic conditions," *Forest Ecol. Manage.*, vol. 262, no. 4, pp. 598–608, 2011.
- [8] J. Zhang, S. Cong, G. Zhang, Y. Ma, Y. Zhang, and J. Huang, "Detecting pest-infested forest damage through multispectral satellite imagery and improved UNet++," *Sensors*, vol. 22, no. 19, 2022, Art. no. 7440.
- [9] L. Huo, H. J. Persson, and E. Lindberg, "Early detection of forest stress from European spruce bark beetle attack, and a new vegetation index: Normalized distance red & SWIR (NDRS)," *Remote Sens. Environ.*, vol. 255, 2021, Art. no. 112240.
- [10] A. Candotti, M. De Giglio, M. Dubbini, and E. Tomelleri, "A sentinel-2 based multi-temporal monitoring framework for wind and bark beetle detection and damage mapping," *Remote Sens.*, vol. 14, no. 23, 2022, Art. no. 6105.
- [11] X. Liu, J. Frey, M. Denter, K. Zielewska-Büttner, N. Still, and B. Koch, "Mapping standing dead trees in temperate montane forests using a pixel- and object-based image fusion method and stereo WorldView-3 imagery," *Ecological Indicators*, vol. 133, 2021, Art. no. 108438.
- [12] H. T. Nguyen, M. L. Lopez Caceres, K. Moritake, S. Kentsch, H. Shu, and Y. Diez, "Individual sick fir tree (*Abies mariesii*) identification in insect infested forests by means of UAV images and deep learning," *Remote Sens.*, vol. 13, no. 2, 2021, Art. no. 260.
- [13] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
- [14] T. Joachims, "Text categorization with support vector machines: Learning with many relevant features," in *Proc. 10th Eur. Conf. Mach. Learn.*, 1998, pp. 137–142.
- [15] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2016, pp. 785–794.
- [16] R. Can, S. Kocaman, and C. Gokceoglu, "A comprehensive assessment of XGBoost algorithm for landslide susceptibility mapping in the upper basin of Ataturk Dam, Turkey," *Appl. Sci.*, vol. 11, no. 11, 2021, Art. no. 4993.
- [17] R. Singh, M. Biswas, and M. Pal, "Cloud detection using Sentinel 2 imageries: A comparison of XGBoost, RF, SVM, and CNN algorithms," *Geocarto Int.*, pp. 1–32, 2022.
- [18] F. Lore, P. Basile, A. Appice, M. de Gemmis, D. Malerba, and G. Semeraro, "An AI framework to support decisions on GDPR compliance," *J. Intell. Inf. Syst.*, pp. 1–28, 2023.
- [19] J. Fan et al., "Comparison of support vector machine and extreme gradient boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates: A case study in China," *Energy Convers. Manage.*, vol. 164, pp. 102–111, 2018.
- [20] T. Wang, Y. Bian, Y. Zhang, and X. Hou, "Classification of earthquakes, explosions and mining-induced earthquakes based on XGBoost algorithm," *Comput. Geosci.*, vol. 170, 2023, Art. no. 105242.
- [21] S. M. Marvasti-Zadeh, D. Goodsman, N. Ray, and N. Erbilgin, "Early detection of bark beetle attack using remote sensing and machine learning: A review," 2022, *arXiv:2210.03829*.
- [22] European Commission et al., "Monitoring of forests through remote sensing—Final report," Publications Office of the European Union, Luxembourg, 2020.
- [23] D. Nardi, H. Jactel, E. Pagot, J.-C. Samalens, and L. Marini, "Drought and stand susceptibility to attacks by the European spruce bark beetle: A remote sensing approach," *Agricultural Forest Entomol.*, vol. 25, no. 1, pp. 119–129, 2023.
- [24] C. J. Tucker, "Red and photographic infrared linear combinations for monitoring vegetation," *Remote Sens. Environ.*, vol. 8, no. 2, pp. 127–150, 1979.
- [25] L. Wang and J. J. Qu, "NMDI: A normalized multi-band drought index for monitoring soil and vegetation moisture with satellite remote sensing," *Geophysical Res. Lett.*, vol. 34, no. 20, pp. 1–5, 2007.
- [26] L. Wang, J. J. Qu, and X. Hao, "Forest fire detection using the normalized multi-band drought index (NMDI) with satellite measurements," *Agricultural Forest Meteorol.*, vol. 148, no. 11, pp. 1767–1776, 2008.
- [27] C. Daughtry, C. Walthall, M. Kim, E. d. Colstoun, and J. McMurtry, "Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance," *Remote Sens. Environ.*, vol. 74, no. 2, pp. 229–239, 2000.
- [28] M.-R. Amiri and N. Usunier, *Learning With Partially Labeled and Interdependent Data*. Berlin, Germany: Springer, 2015.
- [29] P. Cascante-Bonilla, F. Tan, Y. Qi, and V. Ordóñez, "Curriculum labeling: Revisiting pseudo-labeling for semi-supervised learning," in *Proc. 35th AAAI Conf. Artif. Intell.*, 2021, pp. 6912–6920.
- [30] S. Zhang, M. Wang, S. Liu, P. Chen, and J. Xiong, "How unlabeled data improve generalization in self-training? A one-hidden-layer theoretical analysis," in *Proc. 10th Int. Conf. Learn. Representations*, 2022, pp. 1–36.
- [31] N. Thai-Nghe, Z. Gantner, and L. Schmidt-Thieme, "Cost-sensitive learning methods for imbalanced data," in *Proc. Int. Joint Conf. Neural Netw.*, 2010, pp. 1–8.
- [32] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proc. Annu. Conf. Neural Inf. Process. Syst.*, 2017.
- [33] A. Temenos, N. Temenos, M. Kaselimi, A. Doulamis, and N. Doulamis, "Interpretable deep learning framework for land use and land cover classification in remote sensing using SHAP," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, 2023.
- [34] L. S. Shapley, "A value for n-person games, contributions to theory games," in *Proc. 24th Int. Conf. Discov. Sci.*, vol. 2, pp. 307–317, 1953.
- [35] D. Poursanidis, D. Traganos, P. Reinartz, and N. Chrysoulakis, "On the use of Sentinel-2 for coastal habitat mapping and satellite-derived bathymetry estimation using downscaled coastal aerosol band," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 80, pp. 58–70, 2019.
- [36] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Med. Image Comput. Comput.-Assist. Intervention*, 2015, pp. 234–241.
- [37] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3431–3440.
- [38] S. Ge, H. Gu, W. Su, J. Praks, and O. Antropov, "Improved semisupervised UNet deep learning model for forest height mapping with satellite SAR and optical data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 5776–5787, Mar. 2022.
- [39] K. Heidler, L. Mou, C. Baumhoer, A. Dietz, and X. X. Zhu, "HED-UNet: Combined segmentation and edge detection for monitoring the Antarctic coastline," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–14, 2022.
- [40] K. Bittner, S. Cui, and P. Reinartz, "Building extraction from remote sensing data using fully convolutional networks," *Int. Arch. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. XLII-1/W1, pp. 481–486, 2017.
- [41] N. Zang, Y. Cao, Y. Wang, B. Huang, L. Zhang, and P. T. Mathiopoulos, "Land-use mapping for high-spatial resolution remote sensing image via deep learning: A review," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 5372–5391, 2021.
- [42] S. S. M. Salehi, D. Erdogmus, and A. Gholipour, "Tversky loss function for image segmentation using 3D fully convolutional deep networks," in *Proc. Int. Workshop Mach. Learn. Med. Imag.*, 2017, pp. 379–387.
- [43] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.
- [44] A. Appice and D. Malerba, "Segmentation-aided classification of hyperspectral data using spatial dependency of spectral bands," *ISPRS J. Photogrammetry Remote Sens.*, vol. 147, pp. 215–231, 2019.



Giuseppina Andresini received the Ph.D. degree in computer science from the University of Bari Aldo Moro, Bari, Italy, in 2022.

She is currently an Assistant Professor (non-tenure) with the Department of Computer Science, University of Bari Aldo Moro. She has authored and co-authored several papers in international journals and international conferences. Her current research interests mainly include deep learning, explainable artificial intelligence, and adversarial learning with applications in cybersecurity and remote sensing.



Annalisa Appice received the Ph.D. degree in computer science from the University of Bari Aldo Moro, Bari, Italy, in 2005.

She is an Associate Professor with the Department of Computer Science, University of Bari Aldo Moro. She has authored and co-authored more than 180 papers in international journals and conferences. Her current research interests include data mining with spatio-temporal data, data streams, and event logs with applications to remote sensing, process mining, and cybersecurity.

Dr. Appice has been the Program co-Chair of European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD) 2015, International Symposium on Methodologies for Intelligent System (ISMIS) 2017, and Discovery Science (DS) 2020. She was Journal Track co-Chair of ECML-PKDD 2021. She has participated in the organization of several workshops in machine learning, process mining, and cybersecurity. She is a member of the editorial board of Machine Learning Journal (MACH), Data Mining and Knowledge Discovery Journal (DAMI), Engineering Applications of Artificial Intelligence Journal (EAAI) and Journal of Intelligent Information Systems (JIIS). She is responsible for the local research unit of the University of Bari Aldo Moro in the EU project SWIFTT.



Donato Malerba (Senior Member, IEEE) received the M.Sc. degree in computer science from the University of Bari, Italy, in 1987.

He is a Full Professor with the Department of Computer Science, University of Bari Aldo Moro, Italy. He has authored and co-authored more than 330 papers in international journals and conferences. His research interests include machine learning, data mining, Big Data analytic, and their applications.

Prof. Malerba has been responsible for the local research unit of several European and national projects, and was the recipient of an IBM Faculty Award in 2004. He was Program (co-)Chair of International Conference on Industrial, Engineering & Other Applications of Applied Intelligent Systems (IEA-AIE) 2005, International Symposium on Methodologies for Intelligent System (ISMIS) 2006, Symposium on Advanced Database System (SEBD) 2007, and European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD) 2011, and the General Chair of Algorithmic Learning Theory (ALT) and Discovery Science (DS) 2016 and International Joint Conference on Learning & Reasoning (IJCLR) 2023. He is scientific responsible of Spoke 6 - Symbiotic AI of the Future AI Research (FAIR) project. He is in the Editorial Board of several international journals.

Open Access funding provided by 'Universit? degli Studi di Bari "Aldo Moro"' within the CRUI CARE Agreement