



HAL
open science

DIAMANTE: A data-centric semantic segmentation approach to map tree dieback induced by bark beetle infestations via satellite images

Giuseppina Andresini, Annalisa Appice, Dino Ienco, Vito Recchia

► **To cite this version:**

Giuseppina Andresini, Annalisa Appice, Dino Ienco, Vito Recchia. DIAMANTE: A data-centric semantic segmentation approach to map tree dieback induced by bark beetle infestations via satellite images. *Journal of Intelligent Information Systems*, 2024, 10.1007/s10844-024-00877-6 . hal-04687503

HAL Id: hal-04687503

<https://hal.science/hal-04687503v1>

Submitted on 4 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **DIAMANTE: A data-centric semantic segmentation**
2 **approach to map tree dieback induced by bark**
3 **beetle infestations via satellite images**

4 **Giuseppina Andresini · Annalisa**
5 **Appice · Dino Ienco · Vito Recchia**

6
7 Received: date / Accepted: date

8 **Abstract** Forest tree dieback inventory has a crucial role in improving forest
9 management strategies. This inventory is traditionally performed by forests
10 through laborious and time-consuming human assessment of individual trees.
11 On the other hand, the large amount of Earth satellite data that is publicly
12 available with the Copernicus program and can be processed through advanced
13 deep learning techniques has recently been established as an alternative to field
14 surveys for forest tree dieback tasks. However, to realize its full potential, deep
15 learning requires a deep understanding of satellite data since the data collec-
16 tion and preparation steps are essential as the model development step. In this
17 study, we explore the performance of a data-centric semantic segmentation ap-
18 proach to detect forest tree dieback events due to bark beetle infestation in
19 satellite images. The proposed approach prepares a multisensor data set col-
20 lected using both the SAR Sentinel-1 sensor and the optical Sentinel-2 sensor
21 and uses this dataset to train a multisensor semantic segmentation model. The
22 evaluation shows the effectiveness of the proposed approach in a real inventory
23 case study that regards non-overlapping forest scenes from the Northeast of
24 France acquired in October 2018. The selected scenes host bark beetle infes-
25 tation hotspots of different sizes, which originate from the mass reproduction
26 of the bark beetle in the 2018 infestation.

27 **Keywords** Data-centric remote sensing, Sentinel-1 data, Sentinel-2 data,
28 Data fusion, Data imbalance, Semantic segmentation, Bark beetle detection

A. Andresini, A. Appice, V. Recchia
Department of Informatics, Università degli Studi di Bari Aldo Moro, via Orabona, 4 - 70125
Bari - Italy
Consorzio Interuniversitario Nazionale per l'Informatica - CINI, Bari, Italy
E-mail: giuseppina.andresini@uniba.it,
annalisa.appice@uniba.it, vito.recchia@uniba.it

D. Ienco
INRAE, UMR TETIS, University of Montpellier, Montpellier, France
INRIA, University of Montpellier, Montpellier, France
E-mail: dino.ienco@inrae.fr

1 Introduction

Forests and woodlands cover roughly one-third of Earth's surface and play a critical role in providing many ecosystem services, including carbon sequestration, water flow regulation, timber production, soil protection, and biodiversity conservation. However, the accelerating pace of climate change and its impact on species distribution and biome composition are leading to an increase in various types of disturbances, whether biotic, abiotic, or a combination of both, which are now affecting this vital natural resource and resulting in forest loss. Consequently, the decline in key forest ecosystem services is becoming more and more apparent. Among all the disturbances, insect infestations and disease outbreaks (e.g., bark beetle infestations) can induce massive tree dieback and, subsequently, significantly disrupt ecosystem dynamics [18]. This is why forest surveillance is crucial to monitor, quantify and possibly prevent outbreak diseases and enable foresters to perform informed decision-making for effective environmental management. Nevertheless, common strategies used to evaluate the health of forested regions primarily rely on laborious and time-consuming field surveys [8]. Consequently, they are restricted in their ability to cover extensive geographical areas, thereby preventing large-scale analysis across vast territories. To this end, the substantial amount of remote sensing information collected today via modern Earth observation missions constitutes an unprecedented opportunity to scale up forest dieback assessment and surveillance over large areas. As an exemplar, the European Space Agency's Sentinel missions [7] provide a set of quasi-synchronous synthetic aperture radar (SAR) and optical data, systematically acquired worldwide, at high spatial (order of 10m) and temporal (an acquisition up to every five/six days) resolution. This information can be of paramount interest to support large-scale forest dieback assessment and surveillance systems.

While the research community is investigating the benefit related to exploiting multisensor remote sensing information via recent deep learning approaches [23, 29], there is still the necessity to design effective and well-tailored approaches to get the most out of multisensor remote sensing information [22]. This is the case for the large-scale assessment of tree dieback events induced by insect infestations and disease outbreaks where, to the best of our literature survey, existing works (e.g., [4, 8, 9, 15, 17, 45]) mainly focus on optical data analysis, while no works exist that achieve improvements by leveraging multisensor remote sensing data (e.g., SAR and optical data). In particular, the literature studies to monitor bark beetle infestation in optical data pay high attention to both the data engineering step, through the synthesis of spectral vegetation indices, and the model development step, through the test of various machine learning and deep learning algorithms. On the other hand, similar to research communities where data play a major central role (e.g., computer vision, machine learning, information retrieval), also researchers coming from the remote sensing field are investing efforts towards more systematic and effective exploitation of available data sources. To this end, research actions in this direction have been proposed under the umbrella of data-centric Artificial

Intelligence (AI) [43]. Under this movement, the attention of researchers and practitioners is gradually shifting from advancing model design (model-centric AI) to enhancing the quality, quantity and diversity of the data (data-centric AI). Moreover, when remote sensing data are considered, the data-centric AI perspective is even more important since it can steer the community towards developing a methodology to provide further improvements related to the use of highly heterogeneous information to ameliorate the generalization ability with impact on real-world relevant problems and applications [38]. Nevertheless, the two perspectives (model-centric and data-centric AI) play a complementary role in the larger remote sensing deployment cycle, since standard approaches still struggle to manage and exploit valuable data coming from different and heterogeneous sources as, for instance, in the case of leveraging multisensor complementary information.

With the objective to find a trade-off between data-centric and model-centric achievements in remote sensing and map bark beetle-induced tree dieback events in remote sensing data adopting a semantic segmentation approach (e.g., categorization of pixels into a class), in this paper, we propose DIAMANTE (Data-centric semAntic segMentation to mAp iNfestations in saTellite imagEs): a data-centric semantic segmentation approach to train a U-Net like model from a labelled remote sensing dataset prepared using both SAR Sentinel-1 (S1) and multi-spectral optical Sentinel-2 (S2) remote sensing data sources. In particular, for the model development, we compare the achievements of several multisensor data fusion schema that are performed via early, middle or late stages fusion in an underlining U-Net architecture [37]. The U-Net is considered thanks to its wide versatility and increasing popularity, as well as due to the fact that it has been recently used to map bark beetle-induced tree dieback in Sentinel-2 images [4,5,45]. In addition, in this study, we consider that model recycling is one of the achievements to be evaluated in developing a data-centric AI approach. Hence, we start a preliminary investigation of how the multisensor fusion approaches considered in this study may allow us to train a semantic segmentation model for bark beetle detection, which still achieves good performance in a future data setting. The following are the main contributions of this work:

- The definition of a remote sensing data collection and curation pipeline to prepare multisensor, Sentinel-1 and Sentinel-2 images of forest areas for which the ground truth map of the bark beetle infestation is available at a specific time. The defined pipeline pays particular attention to the quality of the Sentinel-1 and Sentinel-2 data prepared for the model development.
- The adoption and comparison of several multisensor data fusion schemes to combine Sentinel-1 and Sentinel-2 data via early, middle or late stages fusion considering the underlying U-Net architecture.
- The extensive assessment of our proposal using a ground truth map of tree dieback induced by bark beetle infestations in the Northeast of France in October 2018. The evaluation examines the performance of models trained and tested using images acquired over non-overlapping scenes in the same

119 period, as well as the temporal forecasting and transferability of the model
120 [to an upcoming data setting](#).

121 The rest of the manuscript is organized as follows. Related literature is re-
122 viewed in Section 2. The proposed methodology is described in Section 4 while
123 the study site and the associated multisensor remote sensing dataset are in-
124 troduced in Section 3. Section 5 reports the experimental evaluation and it
125 discusses the related findings. Section 6 concludes.

126 **2 Related work**

127 This related work overview is organised into two main fronts. Firstly, we delve
128 into recent remote sensing studies that incorporate machine learning and deep
129 learning to map bark beetle infestation in Sentinel-1 (S1) and Sentinel-2 (S2)
130 images. On the other front, we address the recent achievements of the data-
131 centric artificial intelligence paradigm in remote sensing applications.

132 **2.1 Bark beetle detection in remote sensing**

133 Remote sensing studies to map forest stress related to bark beetle attacks
134 have mainly focused on the analysis of Sentinel-2 data [16]. These studies are
135 mainly inspired by the analysis conducted in [1] to explore the effect of several
136 forest disturbances sources (comprising bark beetle infestation) on S2 data.
137 This study shows that the bark beetle infestation, which may affect the bio-
138 physical and biochemical properties of trees, is commonly visible via Sentinel-2
139 multi-spectral imagery. In particular, the chlorophyll degradation and nitro-
140 gen deficiency lead to an increase in reflectance spectrum in the visible re-
141 gion (particularly, red and green bands). Changes caused by the reduction
142 of chlorophyll and leaf water have also an effect on Near Infrared (NIR) and
143 Water vapor bands, while diseased and insect attacks affect red-edge bands.
144 This analysis has boosted a plethora of studies [4,5,8,9,15,17,25,26,45] that
145 explore the ability of various spectral vegetation indices to enhance the ac-
146 curacy of decision models trained on Sentinel-2 data. Notice that explored
147 spectral vegetation indices mainly combine red, green, NIR and SWIR (short
148 wave infrared) bands.

149 Regarding the classification algorithms used to map bark beetle infesta-
150 tions in Sentinel-2 images, the most recent studies have mainly used machine
151 learning algorithms such as Random Forest [4,5,8,9,25], Support Vector Ma-
152 chine [4,9,15] and XGBoost [4,5]. Instead, [4,5,45] explore the performance of
153 deep learning algorithms under semantic segmentation settings such as U-Net
154 [4,5,45] and FCN-8 [4]. To handle the data imbalance situation, [4,5,15] use a
155 cost-based learning strategy in combination with Random Forest and Support
156 Vector Machine, while [4,5] use the Tversky loss in combination with U-Net
157 and FCN-8. Finally, some studies consider Sentinel-2 time series data to train
158 either Random Forest [5,8,17] or U-Net models [5].

159 On the other hand, only recently, few remote sensing studies have started
160 exploring the potential of Sentinel-1 data to detect bark beetle infestations.
161 Sentinel-1 data are traditionally used in deforestation [detection](#) on [21]. How-
162 ever, [22] has recently hypothesized that the joint exploitation of Sentinel-1 and
163 Sentinel-2 satellite information can disclose useful information to detect bark
164 beetle infestation hotspots. In particular, this study finds significant differences
165 between Sentinel-1 values measured in infested and healthy sites, respectively.
166 Similar conclusions are drawn in [2]. However, [2,22] perform a statistical
167 analysis of Sentinel-1 data distribution without exploring how the use of the
168 Sentinel-1 information can contribute to learning accurate decision models to
169 characterise bark beetle infestations. In general, based on the literature survey,
170 [22] highlights that significant research effort is still needed to explore the full
171 potential of multisensor data in insect-induced forest disturbance mapping. In
172 this direction, [25] shows that the joint analysis of Sentinel-1 and Sentinel-2
173 data marginally contributes to improving the performance of Random For-
174 est models. This conclusion has been recently confirmed also by [28] where
175 poor performances have been achieved for bark beetle infestation mapping
176 exploiting only Sentinel-1 radar data and negligible amelioration by the joint
177 exploitation of multisensor (Sentinel-1 and Sentinel-2) data considering both
178 Bayesian and Random Forest classification models. [Notably, in \[28\], the multi-
179 sensor data are stacked in a single feature vector that is used as input space
180 for training a classification model. This corresponds to an early fusion schema
181 that concatenates pixel-wise the feature vectors which are acquired with the
182 Sentinel-1 and Sentinel-2 sensors before starting the training stage.](#)

183 On the other hand, some recent studies have started to investigate how to
184 combine multisensor remote sensing data (e.g., Sentinel-1 and Sentinel-2 data)
185 for the underlying task of land use land cover mapping under a semantic seg-
186 mentation setting [39]. [The authors of \[29\] have surveyed recent deep learning
187 architectures developed to handle multisensor data comprising Sentinel-1 and
188 Sentinel-2 data. However, this survey mainly considers problems of change
189 detection and biomass estimation without any attention to bark beetle detec-
190 tion problems. In addition, this study points out that the majority of deep
191 neural architectures trained with multisensor satellite data adopt an early fu-
192 sion mechanism to concatenate pixel-wise data acquired with the Sentinel-1
193 and Sentinel-2 satellites. The output of the concatenation step is subsequently
194 used as input space for the deep neural model development. In particular, the
195 authors of both \[31\] and \[41\] learn a U-Net model for land cover classifica-
196 tion and flood detection via an early fusion of the Sentinel-1 and Sentinel-2
197 data. The authors of \[3\] introduce the Principal Component Analysis \(PCA\)
198 to combine stacked Sentinel-1 and Sentinel-2 imagery before training a U-Net
199 model for the downstream task of tropical mountain deforestation delineation.
200 \[On the other hand, a few studies have recently started the investigation of late
201 fusion mechanisms to combine Sentinel-1 and Sentinel-2 data through a deep
202 learning architecture. For example, the authors of \\[24\\] describe a two-branch
203 architecture that separately extracts features from data acquired with the two
204 distinct satellites and perform the late convolutional fusion before the final de-\]\(#\)](#)

205 cision. A similar late fusion schema is also investigated in [19] for a problem of
206 urban change detection. This study describes an architecture composed of two
207 separate, identical U-Net architectures that process Sentinel-1 and Sentinel-2
208 image pairs in parallel, and lately fuses extracted features from both sensors
209 at the final decision stage. A middle fusion mechanism is introduced in [6] to
210 perform the fusion of Infrared-Red-Green (IRRG) images and Digital Surface
211 Model (DSM) data extracted from the Lidar point cloud through a SegNet
212 model. Middle fusion is performed at the encoder layers with a simple sum-
213 mation. Imagery data fusion schemes are also discussed in the survey paper
214 [46].

215 In any case, to the best of our knowledge, no previous studies have been
216 proposed yet to explore the opportunity of combining Sentinel-1 and Sentinel-
217 2 data via modern deep learning architecture (i.e., U-Net) for the downstream
218 bark beetle detection task. In addition, this is the first study that frames
219 the investigation of different multisensor fusion schemes (i.e., early fusion,
220 middle fusion and late fusion) in a U-Net development step performed under
221 the umbrella of data-centric AI. On the other hand, neither previous studies
222 have experimented with a fusion mechanism that operates at the encoder
223 level of semantic segmentation models trained on Sentinel-1 and Sentinel-2
224 data, nor these studies have started the investigation of achievements of data
225 fusion schemes for model development done under the possible lens of model
226 recycling.

227 2.2 Data-centric artificial intelligence in remote sensing

228 Data plays a fundamental role in several remote sensing problems, comprising
229 satellite imagery-based forest health monitoring. As a consequence, the emerg-
230 ing data-centric artificial intelligence paradigm [44] has recently started receiv-
231 ing attention in remote sensing where the big satellite image collections (e.g.,
232 the Earth Sentinel-1 and Sentinel-2 image collections acquired via the Coper-
233 nicus programme) are freely available. [38] describes the main principles of the
234 data-centric artificial intelligence paradigm in geospatial data applications by
235 highlighting that data acquisition and curation should receive as much atten-
236 tion as data engineering and model development and evaluation. This study
237 describes one of the first data-centric remote sensing pipelines experimented
238 for land cover classification in satellite imagery. [35] describes a data-centric
239 approach that uses deep feature extraction to prepare a Sentinel-2 dataset to
240 improve the performance of insect species distribution models. [12] describes a
241 data-centric approach that combines semantic segmentation and Geographical
242 Information Systems (GIS) to obtain instance-level predictions of wind plants
243 by using free orbital satellite images. Specifically, this study achieves an im-
244 provement of the model performance by including the wind plant shadows
245 to increase the mapped area and facilitate target detection. [11] investigates
246 the application of iterative sparse annotations for semantic segmentation in
247 remote-sensing imagery, focusing on minimizing the labor-intensive and costly

248 data labeling process. Finally, [40] describes a data-centric approach for RGB
249 imagery dataset creation that reduces annotation ambiguity for RGB images
250 by combining semi-supervised classification and clustering. To the best of our
251 knowledge, no previous studies have explicitly defined a data-centric semantic
252 segmentation approach that pays specific attention to the data curation step,
253 in addition to the model development step, to support bark beetle infestation
254 mapping considering multisensor remote sensing data provided by Sentinel-1
255 and Sentinel-2 satellites.

256 3 Study area and data preparation

257 This section describes the pipeline realised to prepare the datasets used to
258 train and test the semantic segmentation models. We used Microsoft Planetary
259 Computer¹ that provides the API to access petabytes of environmental
260 monitoring data comprising Sentinel-1 and Sentinel-2 images from 2016 to the
261 present. Datasets are accessed via Azure Blob Storage. The study site denoted
262 as *Northeast France*, situated in the northeastern region of France, is
263 predominantly covered by coniferous forests. In 2018 and 2019, a significant
264 proliferation of bark beetles occurred, leading to an estimation by the French
265 National Forestry Office in late April 2019 that approximately 50% of spruce
266 trees in France were infested, contrasting with the typical rate of 15% for dead
267 or diseased trees under normal circumstances. Notably, preceding 2018, there
268 were no instances of substantial windthrows in this area, suggesting that the
269 observed regional-scale attacks were likely spurred by the hot summer droughts
270 experienced in 2018. Satellite data covering the *Northeast France* study site
271 consists of a Synthetic Aperture Radar (SAR) image acquired via the Sentinel-
272 1 sensor and an optical multi-spectral image acquired via the Sentinel-2 sensor.

273 3.1 Sentinel-1 and Sentinel-2 data collection

274 The Sentinel-1 satellite constellation collects [polarization](#) data via a C-band
275 synthetic-aperture radar instrument. [The C-band denotes a nominal frequency](#)
276 [range from 8 to 4 GHz \(3.75 to 7.5 cm wavelength\) within the microwave](#)
277 [\(radar\) portion of the electromagnetic spectrum. Imaging radars equipped](#)
278 [with C-band are generally not hindered by atmospheric effects. They are capa-](#)
279 [ble of imaging in all-weather \(even through tropical clouds and rain showers\),](#)
280 [day or night.](#) The constellation is composed of two satellites (Sentinel-1A and
281 Sentinel-1B), and it offers a 6-day exact repeat cycle. This means that, over
282 the same geographical area, one SAR can be accessed every 6 days. Due to
283 the nature of the radar signal, the raw information needs calibration correc-
284 tion related to the terrain topography. For this reason, we adopt the level-1
285 Radiometrically Terrain Corrected (RTC) product available via the Microsoft

¹ <https://planetarycomputer.microsoft.com/>

Table 1: Sentinel-1 band description

Band	Spatial Resolution	Wavelength	Band Name
VV	10 m	C Band / 3.75 – 7.5 cm	Vertical-Vertical
VH	10 m	C Band / 3.75 – 7.5 cm	Vertical-Horizontal

286 Planetary platform². This product provides SAR images at 10m of spatial res-
 287 olution. Here we consider the two polarizations VV (Vertical-Vertical) and VH
 288 (Vertical-Horizontal). In particular, VV is a mode of radar polarisation where
 289 the microwaves of the electric field are oriented in the vertical plane for both
 290 signal transmission and reception by means of a radar antenna. VH is a mode
 291 of radar polarisation where the microwaves of the electric field are oriented in
 292 the vertical plane for signal transmission and where the horizontally polarised
 293 electric field of the back-scattered energy is received by the radar antenna.
 294 The list of Sentinel-1 bands considered in this study is reported in Table 1.

295 The Sentinel-2 satellite constellation retrieves multi-spectral radiometric
 296 data (13 bands) in the visible, near infrared, and short wave infrared parts
 297 of the spectrum through two satellites (Sentinel-2A and Sentinel-2B). The
 298 Sentinel-2 constellation permits covering the majority of the Earth’s surface
 299 with a repeat cycle of 5 days. The optical imagery is acquired at high spatial
 300 resolution (between 10m and 60 m) over land and coastal water areas. The mis-
 301 sion supports a broad range of services and applications such as agricultural
 302 monitoring, emergency management or land cover classification. Similarly to
 303 the SAR signal, also the optical signal collected by the Sentinel-2 sensors re-
 304 quires corrections. To this end, we adopt the level 2A product available via
 305 the Microsoft Planetary platform³ that provides atmospherically corrected
 306 surface reflectances. Here we consider all the multi-spectral bands at a spatial
 307 resolution of 10m. While bands B2, B3, B4 and B8 are originally at a spatial
 308 resolution of 10m, for all the other bands we downscale them at 10m of spa-
 309 tial resolution via the nearest-neighbor resampling based interpolation [34].
 310 This technique selects the value of the pixel that is nearby the surrounding
 311 coordinates of the intended interpolation point. Finally, we ignore the B10
 312 (SWIR – Cirrus) band that is reserved for atmospheric corrections. The final
 313 list of Sentinel-2 bands considered in this study is reported in Table 2.
 314 In particular, for each Sentinel-2 band, we report the spatial resolution, the
 315 central wavelength, and the band name. The central wavelength refers to the
 316 midpoint wavelength at the centre of the spectral band range (barycenter)
 317 that the satellite sensor captures. For example, for the B1 band that captures
 318 wavelengths from 433 to 453 nanometers (nm), the central wavelength is 443
 319 nm.

² <https://planetarycomputer.microsoft.com/dataset/sentinel-1-rtc>

³ <https://planetarycomputer.microsoft.com/dataset/sentinel-2-12a>

Table 2: Sentinel-2 band description.

Band	Spatial Resolution	Central Wavelength	Band Name
B1	60 m	443 nm	Coastal aerosol
B2	10 m	490 nm	Blue
B3	10 m	560 nm	Green
B4	10 m	665 nm	Red
B5	20 m	705 nm	Red edge 1
B6	20 m	740 nm	Red edge 2
B7	20 m	783 nm	Red edge 3
B8	10 m	842 nm	Near Infrared Narrow
B8A	20 m	865 nm	Narrow Near Infrared Narrow
B9	60 m	940 nm	Water vapor
B11	20 m	1610 nm	SWIR 1
B12	20 m	2190 nm	SWIR

320 3.2 Multisensor data alignment

321 Let us consider a collection of scenes in *Northeast France* for which we know
322 the coordinates of each scene geometry and the timestamp in which scenes
323 were observed using both Sentinel-1 and Sentinel-2 sensors. For each scene, we
324 perform two geospatial queries to select a Sentinel-1 and a Sentinel-2 image
325 acquired in a given time interval. The two queries are performed over the
326 Sentinel-1 and Sentinel-2 collections, respectively, using the coordinates of the
327 selected scenes and the selected time interval as query filters. The queried
328 Sentinel-1 and Sentinel-2 images are recorded in the World Geodetic System
329 1984 ensemble using metric units. As each query may return a resultset of
330 images, we adopt a pipeline to select a representative image from each resultset.

331 In particular, images are downloaded from Planetary using the STAC API.⁴
332 For each scene in the study area, we first retrieve the Sentinel-2 image of the
333 scene in a given month by formulating a STAC API query with the parameters
334 "catalogue", "bbox" and "datetime" set as follows: the value "sentinel-2-l2a" is
335 used as "catalogue", the "list of the coordinates of the four vertices of the
336 rectangular box of the scene" is used as value for "bbox", and the "date inter-
337 val from the first day to the last day of a given month" is used as value for
338 "datetime". As the Sentinel-2 satellite may record images of the Earth every
339 five days, the resultset of such query may contain several Sentinel-2 images
340 recorded in the sentinel-2-l2a catalogue, covered by the given bbox, and ac-
341 quired by the satellite within the selected datetime interval. The motivation
342 for querying the sentinel-2-l2a catalogue with a time interval (one month in
343 this study) is that cloud cover, shadows and defective pixels are among the
344 main issues that may affect the Sentinel-2 imagery. The assumption for the
345 success of a model development step performed with Sentinel-2 images is that
346 images have to be as much as possible cloud and defective pixels-free. For this
347 reason, we query Sentinel-2 imagery on a time interval (of one month in this
348 study), to improve the possibility of choosing low-affected Sentinel-2 images
349 in terms of clouds and defective pixels. Hence, we select the Sentinel-2 image

⁴ <https://planetarycomputer.microsoft.com/docs/quickstarts/reading-stac/>

of the resultset that achieves the lowest value of "cloud index". If several images achieve the minimum value of the cloud index in the resultset, then we select the most recent Sentinel-2 image of this selection. The cloud index is computed based on the output of the Scene Classification Level (SCL) algorithm [30]. This information is also recorded as a band in the sentinel-2-l2a catalogue. Specifically, the SCL algorithm uses the reflectance properties of imagery bands to establish the presence or absence of clouds or defective pixels in an image. In this way, it identifies clouds, snow and cloud shadows thus, generating a classification map, which consists of three different cloud classes (including cirrus), together with six additional classes covering shadows, cloud shadows, vegetation, not vegetated, water and snow land covers. For a candidate Sentinel-2 image, the index of cloud is computed as the percentage of imagery pixels that the SCL algorithm recognises as noise, defective, dark, cloud, cloud shadow or thin cirrus.

Given the Sentinel-2 image retrieved for a given scene in the given month, then we formulate the STAC API query to retrieve the Sentinel-1 image that is co-located in space and time with this Sentinel-2 image. The new query is performed by setting the "bbox" parameter as in the query performed to obtain the Sentinel-2 image while setting "catalogue" equal to "sentinel-1-rtc" and "datetime" equal to the "interval from three days before the date of the Sentinel-2 image and three days after the date of the Sentinel-2 image". The time interval of this query depends on the fact that we would extract a Sentinel-1 image that should be roughly co-located in time with the Sentinel-2 image. On the other hand, Sentinel-1 images are collected every three days with any weather by using a technology not affected by clouds or weather. In addition, we note that noise has been already removed from the Sentinel-1 images that are recorded in the "sentinel-1-rtc" catalogue of Planetary thanks to the application of the Radiometrically Terrain Corrected (RTC) process. This process has been performed before recording the images in the "sentinel-1-rtc" catalogue by using the Ground Range Detected (GRD) Level-1 products produced by the European Space Agency with the RTC processing performed by Catalyst⁵. Hence, we limit to search the Sentinel-1 images potentially collected before and after the Sentinel-2 image and select the Sentinel-1 image that is the closest in time to the respective Sentinel-2 image.

3.3 Ground truth data, datasets and statistics

We use the ground truth map of the bark beetle infestation hotspots that caused tree dieback in the Northeast of France in October 2018.⁶ This map was commissioned by the French Ministry of Agriculture and Food to Sertit

⁵ <https://catalyst.earth/>

⁶ <https://macarte.ign.fr/carte/3bd52aa2b6422a3a58b5086576f91080/Foyers+de+scolytes+dans+les+pessi%C3%A8res+et+les+sapini%C3%A8res+du+Nord-Est+de+la+France,+automne+2018-printemps+2019>

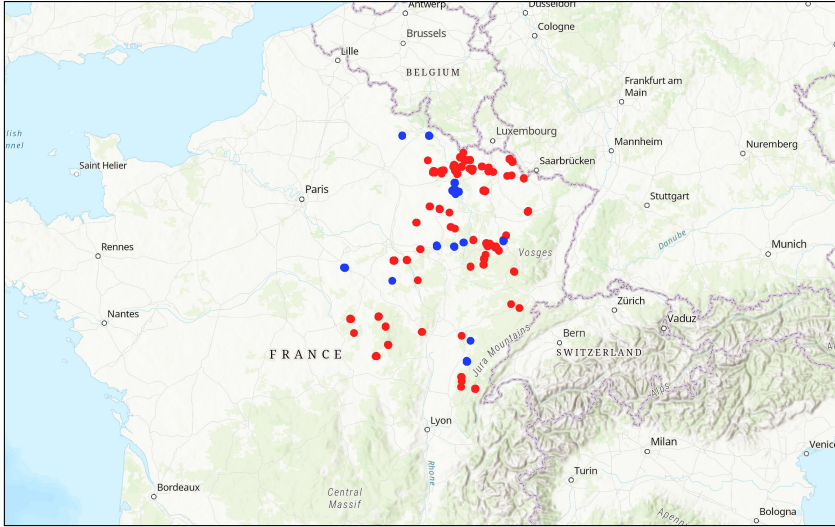


Fig. 1: Location of the centroids of the study 87 scenes in the Northeast of France area. The red circles correspond to scenes considered for training semantic segmentation models, while the blue circles correspond to scenes considered for evaluating semantic segmentation models.

388 (University of Strasbourg), to assess the damage in spruce forests of the North-
 389 east of France following the 2018 bark beetle outbreak. The remote sensing
 390 company WildSense assessed and fixed the infestation hotspot polygons of this
 391 map. In particular, to avoid mixed reflectance from various causes in discoloration
 392 and defoliation of conifer, WildSense manually selected 87 squared,
 393 imagery tiles, covering spruce forestry areas fully under bark beetle attacks
 394 in October 2018. The scenes of the final collection cover 1004020 pixels at 10
 395 square meters resolution. The size of the scenes varies from 27×16 to 296×319
 396 pixels at 10 square meters resolution, while the percentage of infested territory
 397 per scene varies from 0.35% to 34.4% of the scene surface. The total
 398 percentage of damaged territory of the entire scene collection is 2.92%. For
 399 the experimental evaluation of this research work, we consider 71 scenes (covering
 400 772844 pixels at 10 squared meters resolution) as training scene set and
 401 16 scenes (covering 231176 pixels) as testing scene set. A map of the study
 402 scene location and their partitioning in the training set and testing set is depicted
 403 in Figure 1. In addition, WildSense identified an extra scene covering
 404 spruce forestry areas fully under bark beetle attacks, according to a ground
 405 truth map acquired in March 2020. The geographic location of this scene is
 406 shown in Figure 2. This scene is a tile with size 205×135 covering 27675 pixels
 407 with 10 squared meters as spatial resolution with a percentage of infested
 408 territory equal to 3.55%.

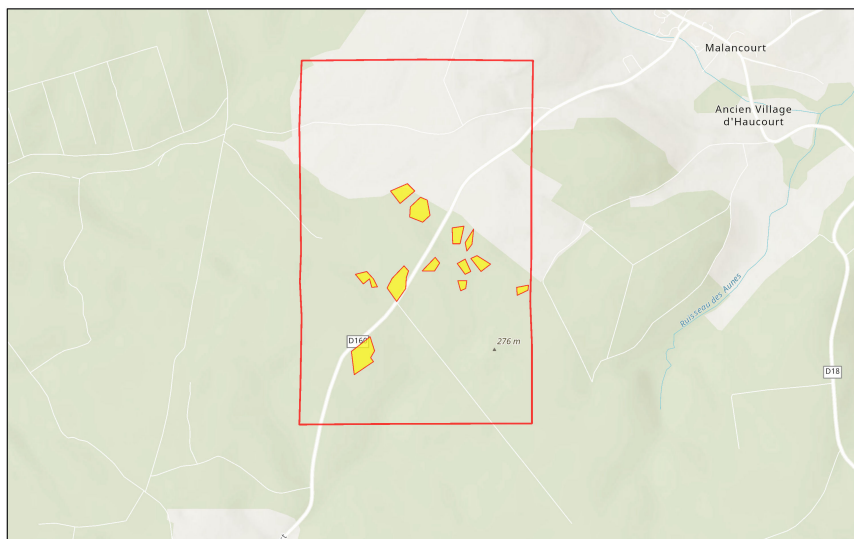


Fig. 2: Location of the scene for which the ground truth mask of the bark beetle infestation was acquired in March 2020. The yellow patches map the forest areas with tree dieback caused by the bark beetle.

409 In this study, we prepare four multisensor, satellite datasets populated
 410 with both the Sentinel-1 and Sentinel-2 images acquired for each scene in
 411 the study area in the Northeast of France. Hence, each dataset is populated
 412 with 87 Sentinel-1 images and 87 Sentinel-2 images roughly co-located in time
 413 within the same month. Specifically, the four multisensor satellite datasets were
 414 obtained by considering Sentinel-1 and Sentinel-2 images acquired monthly for
 415 the 87 study scenes in July 2018, August 2018, September 2018 and October
 416 2018, respectively. We partition each imagery dataset into a training set and
 417 a testing set by using the same split ratio for each month. In particular, as
 418 mentioned above, we select 71 multisensor images as the training set and 16
 419 multisensor images as the testing set for each of these four datasets. Notably,
 420 the multisensor images assigned to the four training sets were acquired for the
 421 same 71 training scenes although in different months. Similarly, the multisensor
 422 images assigned to the four testing sets were acquired for the same 16 testing
 423 scenes although in different months.

424 The dataset collected in October 2018 – the time at which the ground truth
 425 map of the bark beetle-induced tree dieback of the study scenes was produced –
 426 is elaborated to analyse the ability to map bark beetle-induced tree dieback in
 427 October, while datasets collected for the same scenes from July to September
 428 2018 are elaborated to analyse the ability to predict as earlier as possible signs
 429 related to the bark beetle infestation (before trees start dying). Notice that
 430 the analysis of satellite imagery data collected in October 2018 follows some
 431 communications with foresters reported by [8], according to the beginning of

432 the autumn, i.e., October in this study, may be considered the most suitable
433 period for proactive measures, i.e., for looking for areas of infested trees and
434 removing them from the forest before next spring. On the other hand, the anal-
435 ysis of satellite imagery data collected in [July](#), August and September 2018
436 is done to explore the performance of the proposed approach in predicting
437 where bark beetle infestation disturbance events are likely to cause future tree
438 dieback. This evaluation is done according to the considerations reported in
439 [27] that the early detection symptoms of bark beetle infestation, which com-
440 prise the presence of entrance holes, resin flow from entrance holes and boring
441 dust that occur when the beetles attack the tree, penetrate the bark, and ex-
442 cavate mating chambers and breeding galleries that can be observed through
443 terrestrial fieldwork inventory. So, counting on manually produced labels in
444 the summer months may help the training of semantic segmentation models
445 for automated early detection in scenes uncovered by the forestry fieldwork.

446 Figure 3 shows the box plots of Sentinel-1 and Sentinel-2 data collected in
447 the datasets prepared for this study. All bands are plotted independently of
448 each other for the two opposite ground truth classes (“damaged” and “healthy”).
449 The box plots show that the range of both Sentinel-1 and Sentinel-2 data
450 changes over time. Sentinel-2 data, particularly B5, B6, B7, B8, B8A and B9,
451 show a greater divergence between the opposite classes than Sentinel-1 data,
452 over all the datasets. So, this visual data exploration confirms the general idea
453 that Sentinel-2 contains the most important information to recognize bark
454 beetle infestation hotspots, while Sentinel-1 data can be considered ancillary
455 data that may be used to support analysis of Sentinel-2 data, to gain accuracy
456 in the bark beetle infestation inventory.

457 In addition, Figure 4 shows the results of the bivariate correlation analysis
458 performed by computing the Spearman’s rank correlation coefficient between
459 Sentinel-1 and Sentinel 2 bands in images acquired between [July](#) and October
460 2018. Spearman’s rank correlation coefficient is a non-parametric measure of
461 rank correlation that assesses how well the relationship between two compared
462 variables can be described using a monotonic function. It varies between -1 and
463 +1 with 0 implying no correlation, -1 implying an exact monotonic relationship
464 with negative correlation and +1 implying an exact monotonic relationship
465 with positive correlation. This correlation analysis shows that the Sentinel-1
466 bands VV and VH are negatively correlated to the Sentinel-2 bands B1, B2,
467 B3, B4, B5, B11 and B12, while they are positively correlated to Sentinel-2
468 bands B7, B8, B8A and B9. [The Sentinel-2 band B6 passes from showing a low
469 negative correlation with the Sentinel-1 bands VV and VH in July to showing a
470 low positive correlation with the same Sentinel-1 bands in August, September
471 and October. In general, the intensity of the correlation between the Sentinel-2
472 bands B6, B7, B8, B8A and B9 and the Sentinel-1 bands VV and VH increases
473 from July to August, September and October.](#) In any case, the correlation is
474 close to zero independently of the sign, especially on the bands B6, B7, B8,
475 B8A and B9, which are the Sentinel-2 bands that better separate the opposite
476 classes in the box plot analysis of the same data. Hence, this visual inspection
477 of the collected data confirms a limited correlation between Sentinel-1 and

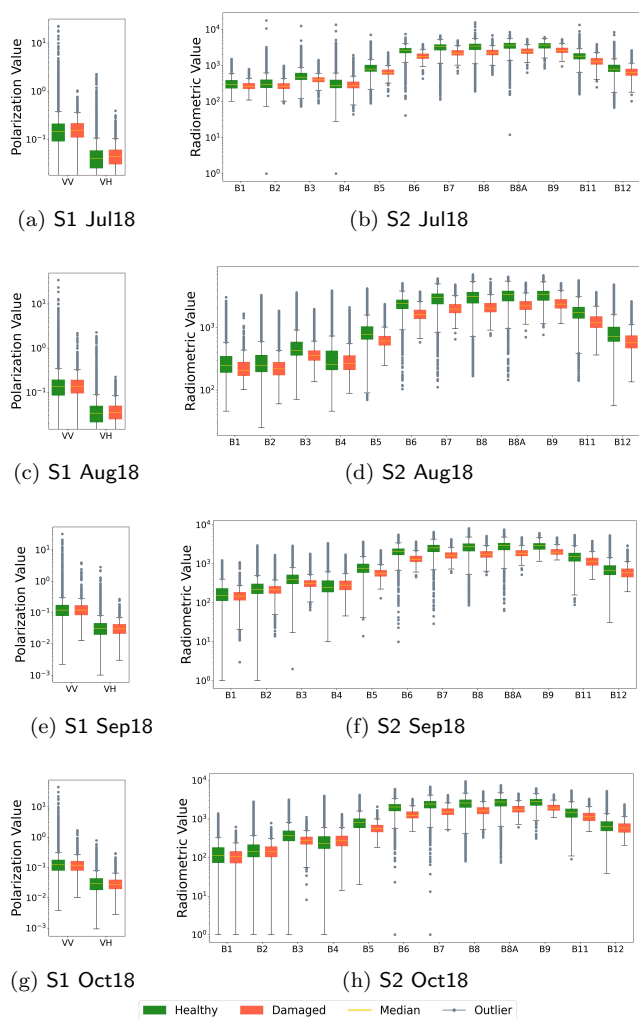
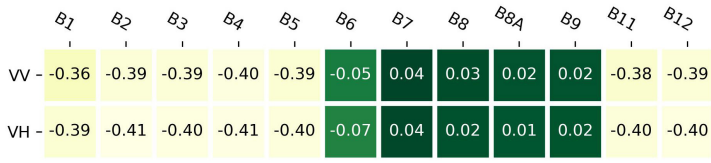


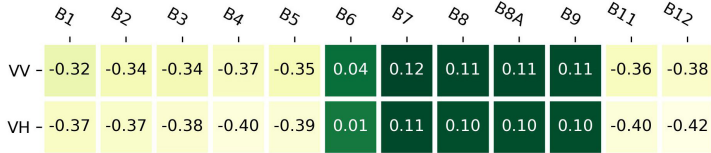
Fig. 3: Box plot distribution of the polarization values measured for the Sentinel-1 bands and the radiometric values measured for the Sentinel-2 bands recorded in the datasets of Sentinel-1 and Sentinel-2 images acquired in the study site in July, August, September and October 2018. Bands are plotted independently with respect to the two opposite classes in the logarithmic scale.

478 Sentinel-2 data, which is one of the prerequisites for taking advantage of a
 479 multisensor approach in model development.

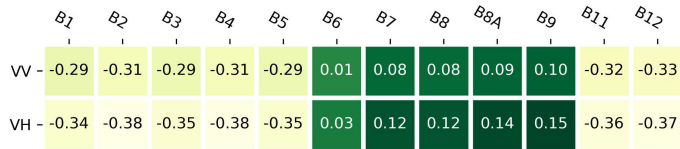
480 Figure 5 shows the box plot of the cloud index of the Sentinel-2 images
 481 selected for this study. This plot shows the high quality of Sentinel-2 images
 482 selected in each month. In fact, we are unable to select images with a cloud
 483 index lower than 5% only in one image in August 2018 and two images in



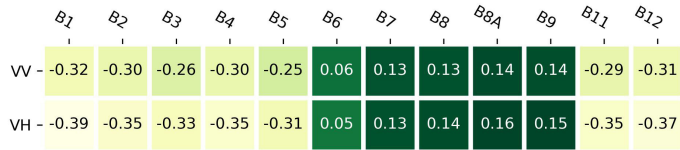
(a) Jul18



(b) Aug18



(c) Sep18



(d) Oct18

Fig. 4: Spearman's rank correlation coefficient computed between Sentinel-1 and Sentinel-2 bands in the images acquired in the study site in July, August, September and October 2018

484 October 2018. We also note that differences between the box-plot quartiles are
 485 slightly higher in October 2018 than in the period July-September 2018. This
 486 depends on the expected increase in the frequency of cloudiness as autumn
 487 advances.

488 Finally, we collect and prepare the pair of Sentinel-1 and Sentinel-2 images
 489 of the scene for which the ground truth map was acquired in March 2020. This
 490 pair of images is used in the evaluation stage only, to explore the transferabil-
 491 ity of the semantic segmentation model learned in October 2018 to subsequent
 492 periods. The Sentinel-2 image acquired for this scene in March 2020 and se-
 493 lected in this study has a low noise and cloud index equal to 0.16%. Finally,

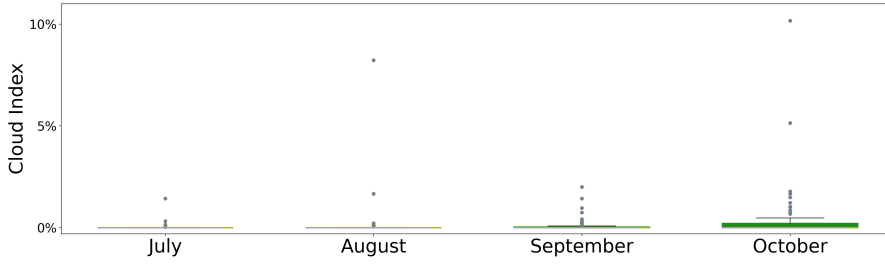


Fig. 5: Box plot of cloud index of Sentinel-2 images acquired in the study site in July, August, September and October 2018.

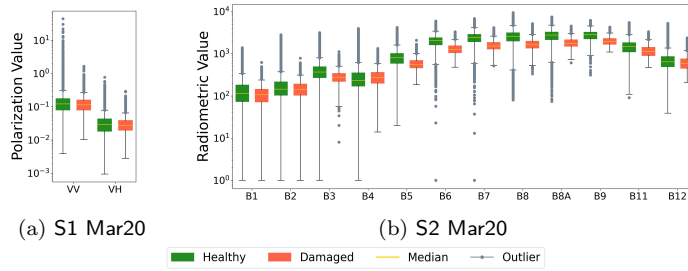


Fig. 6: Box plot distribution of the polarization values measured for the Sentinel-1 bands and the radiometric values measured for the Sentinel-2 bands recorded in the Sentinel-1 image and the Sentinel-2 image acquired in March 2020 for the scene seen in Figure 2. Bands are plotted independently to the two opposite classes in the logarithmic scale.

494 Figure 6 shows the box plots of both Sentinel-1 and Sentinel-2 data collected
 495 in March 2020 for this scene. We note that the outliers of Sentinel-1 data are
 496 spread across a lower heat range than that observed in the images collected in
 497 the summer and autumn months of 2018. On the other hand, B5, B6, B7, B8,
 498 B8A and B9 of Sentinel-2 data still show a remarkable divergence between the
 499 opposite classes as in the images collected in the summer and autumn months
 500 in 2018.

501 4 Semantic segmentation model development

502 The model development step is performed by leveraging the aligned Sentinel-
 503 1 and Sentinel-2 images of scenes for which the ground truth mask of bark
 504 beetle infestation is available. Let us consider $\mathcal{D} = \{(\mathbf{X}_{S1}, \mathbf{X}_{S2}, \mathbf{Y}) \mid \mathbf{X}_{S1} \in$
 505 $\mathbb{R}^{H \times W \times 2}, \mathbf{X}_{S2} \in \mathbb{R}^{H \times W \times 12}, \mathbf{Y} \in \mathbb{R}^{H \times W \times 1}\}$ a collection of labelled Sentinel-
 506 1 and Sentinel-2 images of forest scenes, where every ground truth mask \mathbf{Y}
 507 is associated with the images \mathbf{X}_{S1} and \mathbf{X}_{S2} , acquired from Sentinel-1 and

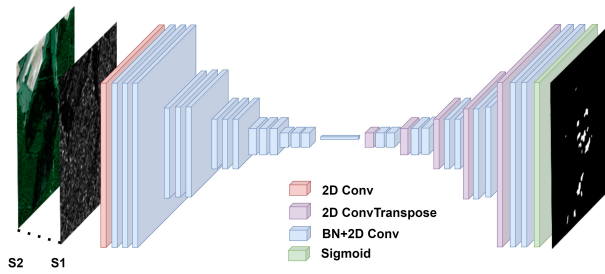


Fig. 7: Early fusion. Abbreviations: 2D Conv = 2D Convolutional layer; BN=Batch Normalization; S1=Sentinel-1; S2=Sentinel-2

508 Sentinel-2 satellites, respectively. For each scene, H and W denote the spatial
 509 extent of the monitored scene in terms of scene height and scene width, respec-
 510 tively. The model development step trains a semantic segmentation network
 511 from \mathcal{D} through a U-Net-like architecture that is also in charge of learning the
 512 data fusion.

513 The U-Net architecture is composed of an encoder part and a decoder part.
 514 The encoder extracts features. It consists of multiple blocks, where each block
 515 is composed of a Batch Normalization layer and a 2D Convolutional layer
 516 followed by Max-Pooling for downsampling. At each downsampling step, the
 517 height and width of the tensor halves, while the number of channels remains
 518 unchanged. The decoder part upsamples the encoded feature maps to the
 519 original input shape. It consists of one transposed Convolutional layer for
 520 upsampling, followed by multiple blocks, each of which each block consists of a
 521 Batch Normalization layer and a 2D Convolutional layer. Skipping connections
 522 between the decoder part and the encoder part are used to propagate the
 523 spatial information from the earlier layers to the deeper layers to alleviate
 524 the vanishing gradients problem [42]. The final classification of each imagery
 525 pixel is obtained by using the Sigmoid activation function. The U-Net used in
 526 this study is trained via the Tversky loss, which is commonly used to handle
 527 imbalanced data [20].

528 The data fusion mechanism is implemented through three different strate-
 529 gies, namely, Early fusion, Middle fusion and Late fusion, which are defined
 530 according to the general classification of multimodal data fusion methods re-
 531 ported in the survey of [46]. The Early fusion strategy is the first mechanism
 532 adopted in literature for the multimodal data fusion in the deep neural scenario
 533 [13]. It is implemented via a simple concatenation, performed at an early stage,
 534 of features from different modalities (i.e., sensors in this study). The concate-
 535 nation produces a single input space for the model development. In our study,
 536 the Early fusion strategy, shown in Figure 7, concatenates each pair of images
 537 \mathbf{X}_{S1} and \mathbf{X}_{S2} obtaining a single hypercube with dimension $H \times W \times 14$. A
 538 traditional U-Net architecture is trained on the newly stacked hypercubes.

539 The Middle fusion strategy combines features learned with the separate
 540 branches of a multi-input deep neural network that takes data acquired from

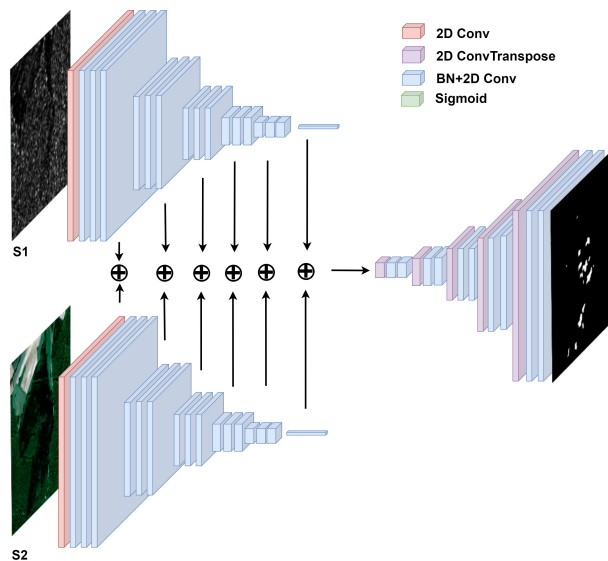


Fig. 8: Middle fusion. Abbreviations: 2D Conv = 2D Convolutional layer; BN=Batch Normalization; S1=Sentinel-1; S2=Sentinel-2

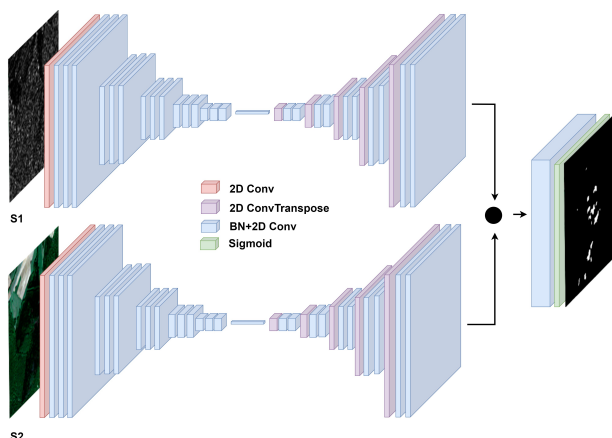


Fig. 9: Late fusion. Abbreviations: 2D Conv = 2D Convolutional layer; BN=Batch Normalization; S1=Sentinel-1; S2=Sentinel-2

541 different modalities as separate inputs. The fusion is performed at an inter-
 542 mediate layer of the deep neural network. The output of this combination
 543 performed at the fusion layer is processed across the subsequent layers of the
 544 network until the decision layer. In our study, the Middle fusion strategy, de-
 545 picted in Figure 8 uses an architecture with two encoder branches, each taking
 546 \mathbf{X}_{S1} and \mathbf{X}_{S2} as input, respectively. The output of these branches is fed into a

single decoder. The two encoder branches are mapped into a common feature space via a fusion operation and the fusion output is used for the skipping connections. Two fusion operators, named SUM and CONC, are considered in this work for the middle fusion. The SUM operator performs an element-wise summation between the outputs of two parallel blocks in the encoder parts. The CONC operator produces a single hypercube by stacking the outputs of two parallel blocks in the encoder parts. Subsequently, it employs a 2D Convolutional layer to halve the channel size of the output hypercube. This is done to align with the number of channels of the corresponding decoder block for skipping connections. Both the concatenation [13,47] and the element-wise summation [33,36] are two common fusion operators used in the literature to fuse multimodal features enclosed in RGB images and Depth images by using CNN-based algorithms. We select these two operators for the Middle fusion strategy performed in this study since they implement two different mechanisms in terms of information retention. In particular, the concatenation operator (CONC) allows us to keep all the information from both Sentinel-1 and Sentinel-2 data, where each feature is entirely preserved. On the other hand, the summation operator (SUM) provides a more compact representation than the concatenation. In fact, it fuses the features originated from the two sensors into a single vector having the same size of the combined vectors. This operator can be particularly useful when the features are aligned and represent the same spatial locations or attributes.

The Late fusion strategy processes separately input data provided by each modality through distinct deep neural models, and their outputs are combined at the later stage, usually at the classification stage. In our study, the Late fusion strategy, illustrated in Figure 9, uses an architecture with two identical, parallel encoder and decoder paths that take as input \mathbf{X}_{S1} and \mathbf{X}_{S2} , respectively. The outputs returned by the two decoders are stacked into a single hypercube and the Sigmoid activation function is employed in the final layer. Final considerations concern the expected behaviour of the three data fusion schemes. According to the discussion reported in [46], the Early fusion strategy is expected to better leverage cross-modal information interaction as early as possible in the learning stage. On the other hand, the Late fusion strategy is considered flexible, but it may lack sufficient cross-modal correlation. Finally, the Middle fusion strategy is expected to find a trade-off between Early fusion and Late fusion, with possible advantages in terms of final performances.

5 Empirical evaluation and discussion

5.1 Implementation details

We implemented DIAMANTE in Python 3.0. The source code is available online.⁷ In this study, we consider a U-Net architecture optimized for satellite

⁷ <https://github.com/gsndr/DIAMANTE>

587 images implemented using the Keras 2.15 and TensorFlow as back-end ⁸. Both
588 encoder and decoder components of the different variants of U-Net architec-
589 tures tested in this study are composed of five main blocks. In the encoder
590 part, each block consists of 3 blocks containing a Batch Normalization layer
591 and a 2D Convolutional layer, followed by a 2×2 Max-Pooling operation or
592 downsampling. The stride of the Max-Pooling operation was set equal to 2.
593 In the decoder part, each main block consists of a transposed Convolutional
594 layer (for upsampling) followed by 3 blocks containing a Batch Normaliza-
595 tion layer and a 2D Convolutional layer. The kernel size of each Convolutional
596 layer was set equal to 3×3 . In all hidden layers the Rectified Linear Unit func-
597 tion (ReLU) was used as the activation function, while the Sigmoid activation
598 function was used in the final semantic segmentation layer. The SUM operator
599 was implemented using the Add layer available in TensorFlow.⁹ The training
600 of the U-Net architectures was performed using imagery tiles of size 32×32
601 extracted from the imagery scenes by using tiler library.¹⁰ Both Sentinel-1 and
602 Sentinel-2 data were scaled between 0 and 1 using the Min-Max scaler (as it is
603 implemented in the Scikit-learn 0.22.2 library) In addition, we considered a tile
604 augmentation strategy to improve the performance of the U-Net architecture
605 by using the Albumentations library ¹¹. Specifically, we quintupled the number
606 of training imagery tiles by creating new tiles applying traditional computer
607 vision augmentation operators (i.e., Horizontal Flip, Vertical Flip, Random
608 Rotate, Transpose and Grid Distortion). We used the tree-structured Parzen
609 estimator algorithm to optimize hyper-parameters of U-Net architectures (i.e.,
610 mini-batch size in $\{2^2, 2^3, 2^4, 2^5, 2^6\}$, learning rate between 0.0001 and 0.01
611 and image augmentation in $\{\text{True}, \text{False}\}$), by using 20% of the training set
612 as the validation set. In particular, the hyper-parameter configuration that
613 achieves the highest F1 score on the minority class ("damaged") in the valida-
614 tion set was automatically selected as the best semantic segmentation model.
615 We performed the gradient-based optimisation using the Adam update rule.
616 Finally, each U-Net model was trained with a maximum number of epochs
617 equal to 150, using an early stopping approach to retain the best semantic
618 segmentation model.

619 5.2 Metrics

620 To evaluate the accuracy of the semantic segmentation masks, we measured the
621 following metrics: F1 score (F1) computed for the two opposite classes, Macro
622 F1 score (Macro F1) averaged on the two opposite classes and Intersection-over-
623 Union (IoU). Specifically, the F1 score measures the harmonic mean of Precision
624 and Recall. The Precision = $\frac{TP}{TP+FP}$ is the fraction of pixels correctly classified
625 in a specific class (TP) among pixels of the considered class (TP + FP). The

⁸ <https://github.com/karolzak/keras-unet/tree/master>

⁹ https://www.tensorflow.org/api_docs/python/tf/keras/layers/Add

¹⁰ <https://github.com/the-lay/tiler>

¹¹ <https://albumentations.ai/>

626 $\text{Recall} = \frac{TP}{TP+FN}$ is the fraction of pixels correctly classified in a specific class
 627 (TP) among pixels classified in the considered class ($TP + FN$). In this study,
 628 we computed the F1 score for the two opposite classes of both case studies:
 629 “healthy” (F1(h)) and “damaged” (F1(d)). Macro F1 measures the average of
 630 each F1 score value per class, that is, $\text{Macro F1} = \frac{F1(h)+F1(d)}{2}$. The IoU score
 631 is the ratio of the intersected area to the combined area of prediction and
 632 ground truth, that is, $\text{IoU} = \frac{TP}{TP+FP+FN}$. This is commonly used to evaluate
 633 the accuracy of models trained in both semantic segmentation and object
 634 detection problems. All metrics are reported in percentages and computed
 635 on the images collected for the testing scenes. For each metric, the higher
 636 the value, the better the performance of the semantic segmentation masks
 637 predicted.

638 5.3 Results

639 The illustration of results is organised as follows. Section 5.3.1 presents the
 640 results achieved by processing the multisensor imagery dataset collected in the
 641 study area in October 2018. This analysis is done to evaluate the performance
 642 of the data fusion strategies at the same time the ground truth masks of the
 643 study scenes were created. Section 5.3.2 presents a temporal study where we
 644 explore the performance of the models trained and evaluated considering satel-
 645 lite images acquired in July, August and September 2018. This analysis is done
 646 to explore the ability of the considered data fusion strategies to learn a model
 647 capable to perform early detection of tree dieback phenomena. Finally, Sec-
 648 tion 5.3.3 illustrates the results achieved by considering multisensor semantic
 649 segmentation models trained from satellite images acquired in October 2018
 650 to predict the mask of tree dieback caused by a bark beetle infestation in a
 651 new scene located in the Northeast of France, but monitored in March 2020.
 652 This analysis explores the transferability over time of a semantic segmentation
 653 model.

654 5.3.1 Performance Analysis

655 In this Section, we analyse the performance of the semantic segmentation
 656 masks produced for the testing scenes of the *Northeast France* study by us-
 657 ing the multisensor semantic segmentation models trained via the three data
 658 fusion schemes illustrated in Section 4. As baselines, we consider the single-
 659 sensor semantic segmentation models trained with a traditional U-Net by pro-
 660 cessing either the Sentinel-1 images (S1 U-Net) or the Sentinel-2 images (S2
 661 U-Net) alone. With regard to the Middle fusion strategy, we report the results
 662 achieved with the two fusion operators: SUM and CONC. This evaluation was
 663 conducted by processing the dataset of images acquired in October 2018 for
 664 both the training and evaluation stages. The accuracy metrics measured on
 665 the semantic segmentation masks produced for the images of the testing scenes
 666 are reported in Table 3. As we expected, the output of the stand-alone use

Table 3: Accuracy performance of semantic segmentation produced with S1 U-Net, S2 U-Net, Early fusion U-Net, Middle fusion (SUM) U-Net, Middle fusion (CONC) U-Net and Late fusion U-Net in the imagery collections acquired in October 2018. The best results are in bold

Configuration	F1(h)	F1(d)	IoU	Macro F1
S1 U-Net	61.90	5.47	2.81	33.68
S2 U-Net	99.14	64.65	47.76	81.89
Early fusion U-Net	99.42	69.09	52.78	84.25
Middle fusion (SUM) U-Net	99.35	70.93	54.92	85.13
Middle fusion (CONC) U-Net	99.40	69.90	53.73	84.65
Late fusion U-Net	99.29	67.66	51.25	83.47

667 of Sentinel-1 images is unsatisfactory for this inventory task. In fact, the con-
 668 figuration S1 U-Net achieves the lowest performance in all accuracy metrics.
 669 Better performance can be achieved by processing Sentinel-2 images in place of
 670 Sentinel-1 images. However, this evaluation study shows that the data fusion
 671 of Sentinel-1 and Sentinel-2 images can help us to improve the performance of
 672 the semantic segmentation model regardless of the type of data fusion strategy
 673 employed. In fact, the Early fusion U-Net, Late fusion U-Net and Middle fusion
 674 U-Net all achieve better performance than S2 U-Net that considers Sentinel-
 675 2 images only. More in detail, the best configuration in terms of F1(d), IoU
 676 and Macro F1 is achieved with the Middle fusion schemes having Middle fusion
 677 (CONC) U-Net as runner-up of Middle fusion (SUM) U-Net. These conclusions
 678 are consistent with the observations on the expected behaviour of the data
 679 fusion schemes reported in Section 4. Figures 10b-10g show the semantic seg-
 680 mentation masks of a sample testing scene predicted by the compared models,
 681 while Figure 10a shows the RGB image of this sample scene. The masks high-
 682 light how the use of a data fusion strategy helps us to reduce the number of
 683 false alarms in this case. Specifically, the bark beetle infestation masks pre-
 684 dicted using the multisensor U-net trained with both Early fusion and Middle
 685 fusion schemes show only one false infested patch, while the U-Net trained
 686 from Sentinel-1 data shows large extensions of false infested areas and the
 687 U-Net trained from Sentinel-2 data shows two false infested patches. Notably,
 688 the multisensor U-Net trained with Late fusion strategy removes one of the
 689 false patches discovered by S2 U-Net, but, at the same time, it alerts a new
 690 false patch that is undetected in the other masks. We note that the Late fusion
 691 strategy is the worst-performing fusion strategy of this experiment. This result
 692 suggests that although the Late fusion strategy may allow us to correct some
 693 false patches detected processing Sentinel-2 data only, it may also produce
 694 some artefacts at the decision level, which may cause false alarms unseen in
 695 the remaining configurations. Finally, the masks of this example show that the
 696 use of SUM operator performs better than the CONC operator in delineating
 697 the large damaged patch located on the left side of the scene.

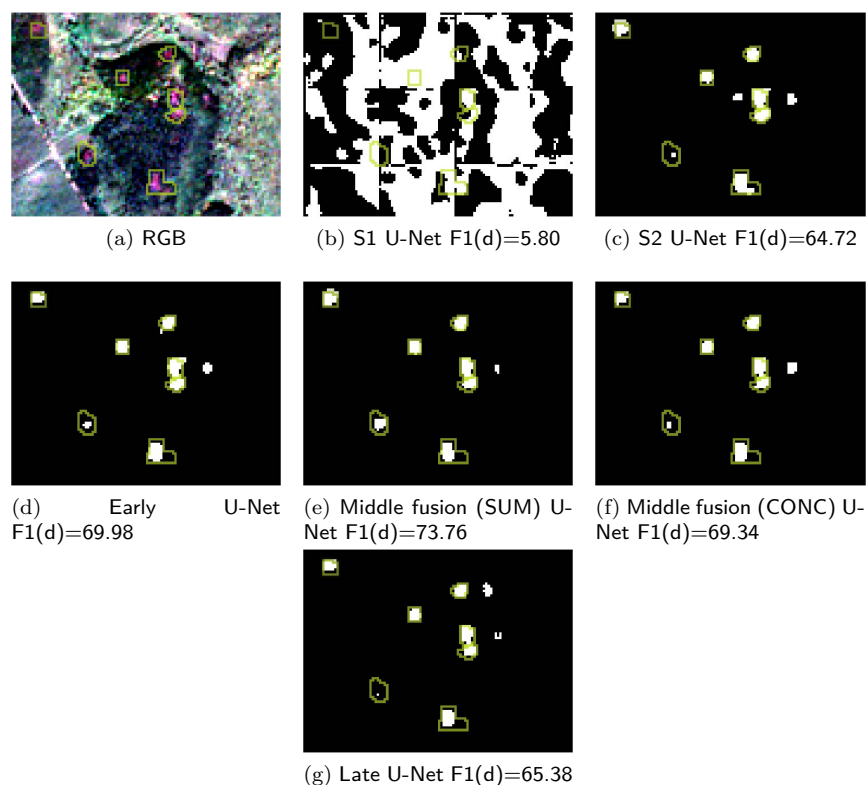


Fig. 10: RGB of the Sentinel-2 image acquired in October 2018 for a testing scene of the study area in the Northeast of France (10a). Inventory masks of tree dieback areas caused by bark beetle hotspots in this scene as they are predicted by S1 U-Net (10b), S2 U-Net (10c), Early fusion U-Net (10d), (10g), Middle fusion U-Net with operators SUM (10e) and CONC (10f) and Late fusion U-Net trained on the imagery set acquired in October 2018 for the training scenes of the study area.

698 5.3.2 Temporal Analysis

699 To complete this investigation, we illustrate the results of a temporal study
 700 conducted to explore the accuracy performance of the semantic segmentation
 701 maps produced when the Sentinel-1 and Sentinel-2 images were acquired in
 702 [the middle of summer \(i.e., July 2018\)](#) and the late summer (i.e., August 2018
 703 and September 2018), while the ground truth map of the tree dieback was
 704 observed in early autumn (October 2018). This analysis is done to explore
 705 the performance of the presented data fusion strategies in the early detection
 706 of areas where bark beetle infestation disturbance events are likely to cause
 707 (near-)future tree dieback. [The temporal snapshots of this experiment were](#)
 708 [selected according to the recent achievements of the analysis on the spectral](#)

Table 4: Accuracy performance of the semantic segmentation models produced with S1 U-Net, S2 U-Net, Early fusion U-Net, Middle fusion (SUM) U-Net, Middle fusion (CONC) U-Net and Late fusion U-Net in the [multisensor](#) images acquired with both Sentinel-1 and Sentinel-2 satellites from July 2018 to October 2018. For each configuration, the best results are in bold, while the runner-up configuration is underlined.

Imagery set	Configuration	F1(h)	F1(d)	IoU	Macro F1
Jul18	S1 U-Net	73.53	4.03	2.06	38.68
	S2 U-Net	97.23	30.18	17.78	63.71
	Early fusion U-Net	98.26	31.49	18.69	64.87
	Middle fusion (SUM) U-Net	98.39	36.46	22.29	67.43
	Middle fusion (CONC) U-Net	98.13	<u>35.27</u>	<u>21.41</u>	<u>66.70</u>
	Late fusion U-Net	<u>98.36</u>	34.20	20.63	66.28
Aug18	S1 U-Net	40.87	4.05	2.07	22.46
	S2 U-Net	95.58	26.46	15.25	61.01
	Early fusion U-Net	98.22	32.49	19.40	65.35
	Middle fusion (SUM) U-Net	<u>98.92</u>	37.41	23.00	68.16
	Middle fusion (CONC) U-Net	99.35	36.37	<u>22.22</u>	<u>67.29</u>
	Late fusion U-Net	98.02	36.15	22.06	67.08
Sep18	S1 U-Net	72.66	5.26	2.70	38.96
	S2 U-Net	99.32	58.92	41.77	79.13
	Early fusion U-Net	99.14	63.42	46.44	81.28
	Middle fusion (SUM) U-Net	99.35	<u>64.28</u>	<u>47.36</u>	<u>81.81</u>
	Middle fusion (CONC) U-Net	<u>99.29</u>	63.16	46.15	81.22
	Late fusion U-Net	99.28	64.99	48.14	82.14
Oct18	S1 U-Net	61.90	5.47	2.81	33.68
	S2 U-Net	99.14	64.65	47.76	81.89
	Early fusion U-Net	99.42	69.09	52.78	84.25
	Middle fusion (SUM) U-Net	99.35	70.93	54.92	85.13
	Middle fusion (CONC) U-Net	99.40	<u>69.90</u>	<u>53.73</u>	<u>84.65</u>
	Late fusion U-Net	99.29	67.66	51.25	83.47

709 separability between the healthy and bark beetle attacked trees illustrated
710 in [14]. In particular, this study shows that bark beetle attacks commonly
711 occur in the summer, while the spectral separability between the two opposite
712 classes ("Healthy" and "Damaged") increases moving from July to October. In
713 addition, it highlights that a time span of approximately one month commonly
714 occurs between the attack of the beetles to a tree and the development of the
715 first symptoms (green-attack) in the tree. Hence, based on the conclusions
716 drawn in this study, the green attack detection stage can reasonably arise in
717 the summer period spanned from July to August. Based on these premises,
718 the accuracy metrics measured on the semantic segmentation maps produced
719 for the testing scenes of this study in each month between July and October
720 2018 are reported in Table 4.

721 These results show that the data fusion of Sentinel-1 and Sentinel-2 con-
722 tinues to help us to gain accuracy also when the multisensor semantic seg-
723 mentation model is trained to forecast tree dieback areas caused by the bark
724 beetle infestation. Notably, Middle fusion (SUM) U-Net achieves the highest
725 F1(d), IoU and Macro F1 in segmentation maps produced in [experiments per-](#)
726 [formed in July 2018, August 2018 and October 2018.](#) The only exception is

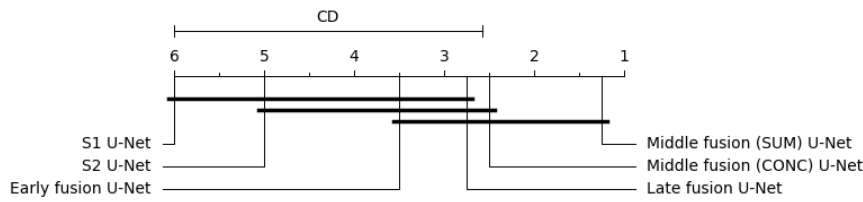


Fig. 11: Comparison of the configurations: Macro F1 measured for S1 U-Net, S2 U-Net, Early fusion U-Net, Middle fusion (SUM) U-Net, Middle fusion (CONC) U-Net and Late fusion U-Net, performed with the Friedman-Nemenyi test run on Macro F1 measured in the temporal analysis performed from July 2018 to October 2018 (computed $pvalue = 0.013$)

727 observed in the segmentation maps produced for the evaluation in September
 728 2018. However, also in the experiment conducted in September 2018, the
 729 Middle fusion (SUM) U-Net still achieves good performance by ranking as the
 730 runner-up of the Late fusion U-Net. To draw conclusive conclusions on the bet-
 731 ter data fusion strategy, we perform the Friedman-Nemenyi test to compare
 732 the Macro F1 measured for S1 U-Net, S2 U-Net, Early fusion U-Net, Middle
 733 fusion (SUM) U-Net, Middle fusion (CONC) U-Net and Late fusion U-Net on the
 734 multiple segmentation maps produced for the testing data of the multisensor
 735 datasets of this temporal analysis. This non-parametric test ranks the model
 736 configurations compared for each dataset separately, so the best-performing
 737 model is given a rank of 1, the second-best rank of 2 and so on. The results of
 738 the Friedman-Nemenyi test reported in Figure 11 shows that the test groups
 739 the configurations adopting a multisensor data fusion strategy as statistically
 740 different from the configurations that consider either Sentinel-1 data only (S1
 741 U-Net) or Sentinel-2 data only (S2 U-Net). In addition, the Middle fusion (SUM)
 742 U-Net achieves the highest rank by having the Middle fusion (CONC) U-Net as
 743 runner-up. Notably, these results of the comparative test support the conclu-
 744 sions already drawn in 5.3.1 and 5.3.3 on the superior performance of a Middle
 745 fusion strategy to combine Sentinel-1 and Sentinel-2 data for bark beetle in-
 746 festation detection.

747 5.3.3 Transferability Analysis

748 In this Section, we examine the accuracy of the semantic segmentation models
 749 learned in October 2018 when used to detect the tree dieback events caused
 750 by bark beetle infestations in March 2020. The accuracy metrics measured in
 751 this experiment are reported in Table 5. These results show that also in this
 752 evaluation scenario, the data fusion of Sentinel-1 and Sentinel-2 may help us
 753 to improve the performance of a semantic segmentation model even when it
 754 was trained on past images and used for mapping the bark beetle infestation
 755 in future images. The only exception is observed for the Late fusion strategy

Table 5: Accuracy performance of semantic segmentation produced with S1 U-Net, S2 U-Net, Early fusion U-Net, Middle fusion (SUM) U-Net, Middle fusion (CONC) U-Net and Late fusion U-Net in the pair of Sentinel-1 and Sentinel-2 images acquired in March 2020. The best results are in bold.

Configuration	F1(h)	F1(d)	IoU	Macro F1
S1 U-Net	82.04	13.55	7.27	37.22
S2 U-Net	96.20	42.18	26.73	69.20
Early fusion U-Net	98.07	50.43	33.72	74.25
Middle fusion (SUM) U-Net	98.03	49.45	32.84	73.74
Middle fusion (CONC) U-Net	97.65	55.49	38.40	76.57
Late fusion U-Net	96.00	41.08	25.85	68.54

756 that achieves lower performance than S2 U-Net. In general, the highest F1(d),
757 IoU and Macro F1 are achieved with the Middle fusion (CONC) U-Net schema
758 having Middle fusion (SUM) U-Net as runner-up. This confirms the conclu-
759 sions on the better performance of the Middle fusion strategy already drawn in
760 Section 5.3.1. Finally, Figures 12b-12g show the semantic segmentation masks
761 predicted for the scene under evaluation. The RGB image of the scene in
762 March 2020 is shown in Figure 12a. The extracts show that the data fusion
763 schemes, except for Late fusion, allow us to reduce the extension of the false
764 alarm areas detected. In both Early fusion and Middle fusion (SUM) schemes,
765 the higher precision is achieved at the cost of a lower recall. Both data fusion
766 configurations allow us to map correctly a percentage of the infested area that
767 is lower than the one mapped processing Sentinel-2 data only. Instead, the
768 use of the Middle fusion (CONC) strategy allows us to achieve the best trade-
769 off between precision and recall in detecting the tree dieback areas caused by
770 the bark beetle infestation. In general, these maps confirm the idea that also
771 when the semantic segmentation model is trained on historical data, the main
772 contribution to the correct detection of the bark beetle infestation is given
773 by Sentinel-2 data, while Sentinel-1 data can aid in reducing false alarms and
774 better delimiting infested areas.

775 5.4 Considerations and Findings

776 The experimental assessment highlights the general advantages of using mul-
777 tisensor data over a single data source in various scenarios of bark beetle
778 detection, including early disease detection and out-of-year temporal transfer.
779 While Sentinel-1 alone is not suitable for the considered downstream mapping
780 task, using Sentinel-2 alone yields satisfactory results. However, the combined
781 use of these two publicly available and freely accessible remote sensing data
782 sources provides the best overall results.

783 More specifically, the joint use of Sentinel-1 and Sentinel-2 data signifi-
784 cantly reduces false alarms and improves the delineation of infested areas in
785 the resulting binary maps. Regarding the early detection of bark beetle attacks
786 (Section 5.3.2), signs of the attack can be detected with reasonable accuracy

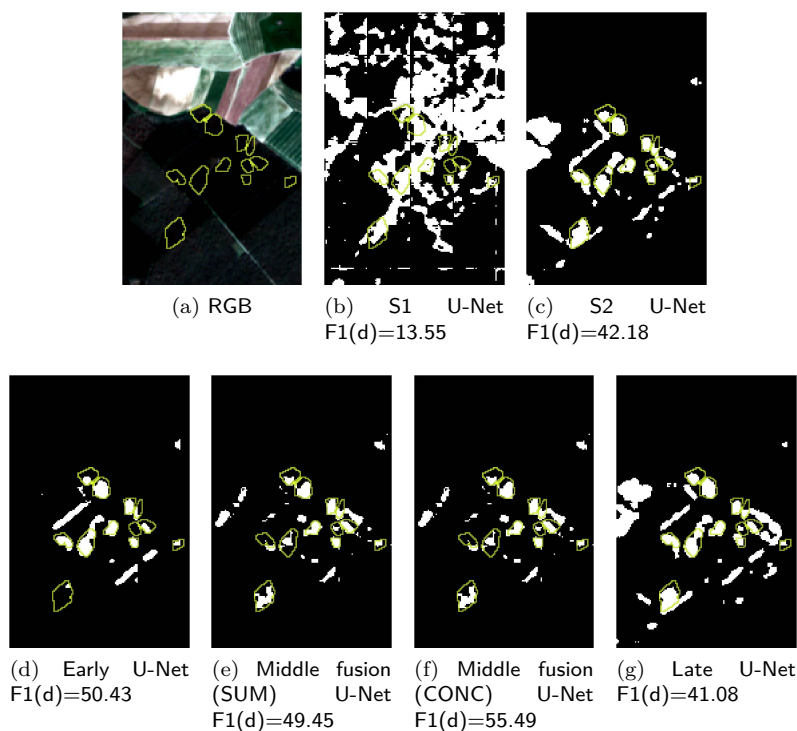


Fig. 12: RGB of the Sentinel-2 image acquired in March 2020 (12a). Inventory masks of tree dieback areas caused by bark beetle hotspots in this scene as they are predicted by S1 U-Net (12b), S2 U-Net (12c), Early fusion U-Net (12d), Middle fusion U-Net with operators SUM (12e) and CONC (12f) and Late fusion U-Net (12g) trained on the imagery set acquired in October 2018 for the training scenes of the study area

787 one month before the acquisition of ground truth data (September 2018). How-
 788 ever, the disease's early stages (before July 2018) are weakly detectable via
 789 satellite imagery.

790 An additional challenge is represented by the out-of-year transfer of the
 791 model trained on 2018 data to 2020 data. Recent studies in the domain of
 792 remote sensing analysis have highlighted that spatial and temporal distribution
 793 shifts can hinder the direct deployment of a model trained on a particular
 794 area or time period to a different area or time period [10,32]. The results
 795 obtained in Section 5.3.3 confirm this point, indicating that there is still room
 796 for research activities in the way historical data can be leveraged in order to
 797 improve current mapping results.

798 Finally, the comparison of the different approaches indicates that all fusion
 799 strategies are statistically significant compared to single source analysis, with
 800 the Middle fusion (SUM) U-Net model exhibiting the best average performance.

801 This finding underscores once more the importance of combining multisensor
802 satellite data for mapping tree dieback induced by bark beetle infestation.

803 6 Conclusion

804 In this study, we investigate the effectiveness of a data-centric semantic seg-
805 mentation approach to map forest tree dieback areas caused by bark beetle
806 hotspots. First, we define a data-centric pipeline to collect and prepare im-
807 ages acquired from both the SAR Sentinel-1 sensor and the optical Sentinel-
808 2 sensor. Then, we explore the accuracy performance of several data fusion
809 strategies, namely Early fusion, Middle fusion and Late fusion adopted for the
810 development of a U-Net-like model combining both Sentinel-1 and Sentinel-2
811 images acquired in the Northeast of France. Finally, we investigate the per-
812 formance of the proposed strategies in multisensor imagery data acquired in
813 Northeast of France with the map of bark beetle infestation available in Octo-
814 ber 2018. We conducted the evaluation with imagery data prepared according
815 to the data curation pipeline presented in this study. The experimental results
816 show that multisensor data can actually help us to improve the ability of the
817 U-Net model to detect tree dieback areas caused bark beetle infestations. The
818 evaluation also explores the transferability of the output of the model devel-
819 opment step, as well as the performance of the proposed approach in early
820 detection of infestations that will cause tree dieback.

821 As future work, we plan to continue the investigation of multisensor data
822 fusion strategies in combination with ecological and weather data, as well tem-
823 poral data trend information. In addition, we plan to extend the investigation
824 of the transferability of the semantic segmentation model, trained with the
825 described multisensor data fusion techniques to unseen data settings. In par-
826 ticular, we intend to start a systematic exploration of some transfer learning
827 approaches to obtain the transferability of a "general" semantic segmentation
828 model trained for a specific disturbance agent to different disturbance agents.
829 For example, we intend to investigate the transferability of a semantic seg-
830 mentation model trained for mapping forest tree die-back hotspots caused
831 by bark beetle infestation to perform the inventory of tree die-back hotspots
832 caused by different families of fungal forest pathogens. In addition, we hope
833 to be able to acquire large-scale data within the experimental phase of the EU
834 project SWIFTT to be able of investigating, on large scale, the transferability
835 of a semantic segmentation model trained in a geographic area to a different
836 geographic area, in addition to a future time.

837 Acknowledgment

838 Annalisa Appice acknowledges support from the SWIFTT project, funded by
839 the European Union under Grant Agreement 101082732. Dino Ienco acknowl-
840 edges support from the Eco2Adapt project, funded by the European Union

841 under Grant Agreement 101059498. Giuseppina Andresini and Vito Recchia
842 are supported by the project FAIR - Future AI Research (PE00000013), Spoke
843 6 - Symbiotic AI, under the NRRP MUR program funded by the NextGener-
844 ationEU. The authors wish to thank the remote sensing company WildSense
845 for preparing the ground truth masks of the evaluation study.

846 **Declarations**

847 6.1 Conflict of Interests

848 The authors declare that they have no conflict of interest.

849 6.2 Competing Interests

850 No competing interests.

851 6.3 Ethics Approval

852 We declare that this submission follows the policies as outlined in the Guide
853 for Authors. The current research involves no Human Participants and/or
854 Animals.

855 6.4 Data, Material, and/or Code Availability

856 The source code is available at <https://github.com/gsndr/DIAMANTE>

857 6.5 Authors' contributions

858 **Giuseppina Andresini**: Conceptualization, Methodology, Software, Validation,
859 Investigation, Supervision, Writing - original draft, Writing - review &
860 editing **Annalisa Appice**: Conceptualization, Methodology, Validation, Visu-
861 alization, Investigation, Writing - original draft, Writing - review & editing, Su-
862 pervision, Project administration. **Dino Ienco**: Conceptualization, Method-
863 ology, Investigation, Writing - original draft, Writing - review & editing. **Vito**
864 **Recchia**: Conceptualization, Methodology, Data curation, Software, Valida-
865 tion, Visualization, Investigation, Writing - review & editing

866 **References**

- 867 1. Abdullah, H., Skidmore, A.K., Darvishzadeh, R., Heurich, M.: Sentinel-2 accurately
868 maps green-attack stage of European spruce bark beetle (*ips typographus*, l.) compared
869 with landsat-8. *Remote Sensing in Ecology and Conservation* **5**(1), 87–106 (2019)

- 870 2. Alshayef, M.S., Musthafa, M.: Identification of bark beetle infestation in part of bo-
871 hemian forest using sentinel-1 time series insar. In: 2021 IEEE International India
872 Geoscience and Remote Sensing Symposium (InGARSS), pp. 5–9 (2021)
- 873 3. Altarez, R.D.D., Apan, A., Maraseni, T.: Deep learning u-net classification of sentinel-1
874 and 2 fusions effectively demarcates tropical montane forest’s deforestation. *Remote*
875 *Sensing Applications: Society and Environment* **29**, 1–21 (2023)
- 876 4. Andresini, G., Appice, A., Malerba, D.: SILVIA: An explainable framework to map bark
877 beetle infestation in Sentinel-2 images. *IEEE Journal of Selected Topics in Applied*
878 *Earth Observations and Remote Sensing* **16**, 10050–10066 (2023)
- 879 5. Andresini, G., Appice, A., Malerba, D.: Leveraging sentinel-2 time series for bark beetle-
880 induced forest dieback inventory. In: The 39th ACM/SIGAPP Symposium on Applied
881 Computing, SAC 2024, pp. 875–882. CAM (2024)
- 882 6. Audebert, N., Le Saux, B., Lefèvre, S.: Beyond rgb: Very high resolution urban remote
883 sensing with multimodal deep networks. *ISPRS Journal of Photogrammetry and Remote*
884 *Sensing* **140**, 20–32 (2018)
- 885 7. Berger, M., Moreno, J., Johannessen, J.A., Levelt, P.F., Hanssen, R.F.: Esa’s sentinel
886 missions in support of earth system science. *Remote Sensing of Environment* **120**, 84 –
887 90 (2012)
- 888 8. Bárta, V., Lukeš, P., Homolová, L.: Early detection of bark beetle infestation in norway
889 spruce forests of central europe using Sentinel-2. *International Journal of Applied Earth*
890 *Observation and Geoinformation* **100**, 102335 (2021)
- 891 9. Candotti, A., De Giglio, M., Dubbini, M., Tomelleri, E.: A Sentinel-2 based multi-
892 temporal monitoring framework for wind and bark beetle detection and damage map-
893 ping. *Remote Sensing* **14**(23) (2022)
- 894 10. Capliez, E., Ienco, D., Gaetano, R., Baghdadi, N.N., Hadj-Salah, A.: Temporal-domain
895 adaptation for satellite image time-series land-cover mapping with adversarial learning
896 and spatially aware self-training. *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.*
897 **16**, 3645–3675 (2023)
- 898 11. Ferreira de Carvalho, O.L., Oline de Albuquerque, A., Luiz, A.S., Henrique
899 Guimarães Ferreira, P., Mou, L., e Silva, D.G., Abílio de Carvalho Junior, O.: A data-
900 centric approach for rapid dataset generation using iterative learning and sparse anno-
901 tations. In: IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing
902 Symposium, pp. 5650–5653 (2023)
- 903 12. de Carvalho, O.L.F., de Carvalho Junior, O.A., de Albuquerque, A.O., Orlandi, A.G.,
904 Hirata, I., Borges, D.L., Gomes, R.A.T., Guimarães, R.F.: A data-centric approach for
905 wind plant instance-level segmentation using semantic segmentation and gis. *Remote*
906 *Sensing* **15**(5), 1–23 (2023)
- 907 13. Couprie, C., Farabet, C., Najman, L., Lecun, Y.: Indoor Semantic Segmentation using
908 depth information. In: First International Conference on Learning Representations
909 (ICLR 2013), pp. 1–8 (2013)
- 910 14. Dalponte, M., Cetto, R., Marinelli, D., Andreatta, D., Salvadori, C., Pirotti, F., Frizzera,
911 L., Gianelle, D.: Spectral separability of bark beetle infestation stages: A single-tree
912 time-series analysis using planet imagery. *Ecological Indicators* **153**, 110349 (2023)
- 913 15. Dalponte, M., Solano-Correa, Y.T., Frizzera, L., Gianelle, D.: Mapping a European
914 spruce bark beetle outbreak using Sentinel-2 remote sensing data. *Remote Sensing*
915 **14**(13) (2022)
- 916 16. Estrada, J.S., Fuentes, A., Reszka, P., Auat Cheein, F.: Machine learning assisted remote
917 forestry health assessment: a comprehensive state of the art review. *Frontiers in Plant*
918 *Science* **14** (2023)
- 919 17. Fernandez-Carrillo, A., Patočka, Z., Dobrovolný, L., Franco-Nieto, A., Revilla-Romero,
920 B.: Monitoring bark beetle forest damage in central europe. a remote sensing approach
921 validated with field data. *Remote Sensing* **12**(21) (2020)
- 922 18. Gomez, D.F., Riggins, J.J., Cognato, A.I.: *Forest Entomology and Pathology: Volume*
923 *1: Entomology – Bark Beetles*, pp. 299–337. Springer International Publishing, Cham
924 (2023)
- 925 19. Hafner, S., Nascetti, A., Azizpour, H., Ban, Y.: Sentinel-1 and sentinel-2 data fusion
926 for urban change detection using a dual stream u-net. *IEEE Geoscience and Remote*
927 *Sensing Letters* **19**, 1–5 (2022)

- 928 20. Hinton, G.E., Vinyals, O., Dean, J.: Distilling the knowledge in a neural network. CoRR,
929 abs/1503.02531 (2015)
- 930 21. Hoekman, D., Kooij, B., Quiñones, M., Vellekoop, S., Carolita, I., Budhiman, S., Arief,
931 R., Roswintiarti, O.: Wide-area near-real-time monitoring of tropical forest degradation
932 and deforestation using sentinel-1. *Remote Sensing* **12**(19), 1–32 (2020)
- 933 22. Hollaus, M., Vreugdenhil, M.: Radar satellite imagery for detecting bark beetle out-
934 breaks in forests. *Current Forestry Reports* **5**, 240–250 (2019)
- 935 23. Hong, D., Gao, L., Yokoya, N., Yao, J., Chanussot, J., Du, Q., Zhang, B.: More diverse
936 means better: Multimodal deep learning meets remote-sensing imagery classification.
937 *IEEE Transactions on Geoscience and Remote Sensing* pp. 1–15 (2020)
- 938 24. Hu, J., Mou, L., Schmitt, A., Zhu, X.X.: Fusionet: A two-stream convolutional neural
939 network for urban scene classification using polsar and hyperspectral data. In: 2017
940 Joint Urban Remote Sensing Event (JURSE), pp. 1–4 (2017)
- 941 25. Huo, L., Persson, H.J., Lindberg, E.: Early detection of forest stress from European
942 spruce bark beetle attack, and a new vegetation index: Normalized distance red & swir
943 (ndrs). *Remote Sensing of Environment* **255**, 112240 (2021)
- 944 26. Jamali, S., Olsson, P.O., Ghorbanian, A., Müller, M.: Examining the potential for early
945 detection of spruce bark beetle attacks using multi-temporal sentinel-2 and harvester
946 data. *ISPRS Journal of Photogrammetry and Remote Sensing* **205**, 352–366 (2023)
- 947 27. Kautz, M., Peter, F., Harms, L., Kammen, S., Delb, H.: Patterns, drivers and detectabil-
948 ity of infestation symptoms following attacks by the European spruce bark beetle. *Jour-
949 nal of Pest Science* **96**, 403–414 (2022)
- 950 28. König, S., Thonfeld, F., Forster, M., Dubovyk, O., Heurich, M.: Assessing combinations
951 of landsat, Sentinel-2 and Sentinel-1 time series for detecting bark beetle infestations.
952 *GIScience & Remote Sensing* **60**(1), 2226515 (2023)
- 953 29. Li, J., Hong, D., Gao, L., Yao, J., Zheng, K., Zhang, B., Chanussot, J.: Deep learning
954 in multimodal remote sensing data fusion: A comprehensive review. *Int. J. Appl. Earth
955 Obs. Geoinformation* **112**, 102926 (2022)
- 956 30. Louis, J., Debaecker, V., Pflug, B., Main-Knorn, M., Bieniarz, J., Mueller-Wilm, U.,
957 Cadau, E., Gascon, F.: Sentinel-2 sen2cor: L2a processor for users. In: *Proceedings of
958 the Living Planet Symposium 2016*, pp. 1–8. Spacebooks Onlin (2016)
- 959 31. Muszynski, M., Hölzer, T., Weiss, J., Fraccaro, P., Zortea, M., Brunschweiler, T.: Flood
960 event detection from sentinel 1 and sentinel 2 data: Does land use matter for performance
961 of u-net based flood segmenters? In: *2022 IEEE International Conference on Big Data
962 (Big Data)*, pp. 4860–4867 (2022)
- 963 32. Nyborg, J., Pelletier, C., Lefèvre, S., Assent, I.: Timematch: Unsupervised cross-region
964 adaptation by temporal shift estimation. *ISPRS Journal of Photogrammetry and Re-
965 mote Sensing* **188**, 301–313 (2022)
- 966 33. Park, S.J., Hong, K.S., Lee, S.: Rdfnet: Rgb-d multi-level residual feature fusion for
967 indoor semantic segmentation. In: *The IEEE International Conference on Computer
968 Vision (ICCV)* (2017)
- 969 34. Patil, M.: Interpolation techniques in image resampling. *International Journal of Engi-
970 neering and Technology* **7**, 567–570 (2018)
- 971 35. Phillips, J., Zhang, C., Williams, B., Jarvis, S.: Data-driven sentinel-2 based deep feature
972 extraction to improve insect species distribution models. In: *EGU General Assembly
973 2022*, pp. 1–1 (2022)
- 974 36. Qian, X., Lin, X., Yu, L., Zhou, W.: Fasfnnet: feature adaptive selection and fusion
975 lightweight network for rgb-d indoor scene parsing. *Opt. Express* **31**(5), 8029–8041
976 (2023). DOI 10.1364/OE.480252
- 977 37. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical
978 image segmentation. In: *Medical Image Computing and Computer-Assisted Interven-
979 tion, MICCAI 2015*, pp. 234–241. Springer (2015)
- 980 38. Roscher, R., Rußwurm, M., Gevaert, C., Kampffmeyer, M., dos Santos, J.A.,
981 Vakalopoulou, M., Hänsch, R., Hansen, S., Nogueira, K., Prexl, J., Tuia, D.: Data-
982 centric machine learning for geospatial remote sensing data. CoRR abs/2312.05327
983 (2023)
- 984 39. Sainte Fare Garnot, V., Landrieu, L., Chahata, N.: Multi-modal temporal attention
985 models for crop mapping from satellite time series. *ISPRS Journal of Photogrammetry
986 and Remote Sensing* **187**, 294–305 (2022)

- 987 40. Schmarje, L., Santarossa, M., Schröder, S.M., Zelenka, C., Kiko, R., Stracke, J., Volk-
988 mann, N., Koch, R.: A data-centric approach for improving ambiguous labels with com-
989 bined semi-supervised classification and clustering. In: S. Avidan, G. Brostow, M. Cissé,
990 G.M. Farinella, T. Hassner (eds.) *Computer Vision – ECCV 2022*, pp. 363–380. Springer
991 Nature Switzerland, Cham (2022)
- 992 41. Solórzano, J.V., Mas, J.F., Gao, Y., Gallardo-Cruz, J.A.: Land use land cover classifi-
993 cation with u-net: Advantages of combining sentinel-1 and sentinel-2 imagery. *Remote*
994 *Sensing* **13**(18), 1–23 (2021)
- 995 42. Wu, J., Zhang, Y., Wang, K., Tang, X.: Skip connection u-net for white matter hyper-
996 intensities segmentation from mri. *IEEE Access* **7**, 155194–155202 (2019)
- 997 43. Zha, D., Bhat, Z.P., Lai, K., Yang, F., Jiang, Z., Zhong, S., Hu, X.: Data-centric artificial
998 intelligence: A survey. *CoRR* **abs/2303.10158** (2023)
- 999 44. Zha, D., Lai, K., Yang, F., Zou, N., Gao, H., Hu, X.: Data-centric AI: techniques and
1000 future perspectives. In: A.K. Singh, Y. Sun, L. Akoglu, D. Gunopulos, X. Yan, R. Ku-
1001 mar, F. Ozcan, J. Ye (eds.) *Proceedings of the 29th ACM SIGKDD Conference on*
1002 *Knowledge Discovery and Data Mining, KDD 2023*, pp. 5839–5840. ACM (2023)
- 1003 45. Zhang, J., Cong, S., Zhang, G., Ma, Y., Zhang, Y., Huang, J.: Detecting pest-infested
1004 forest damage through multispectral satellite imagery and improved UNet++. *Sensors*
1005 **22**(19) (2022)
- 1006 46. Zhang, Y., Sidibé, D., Morel, O., Mériaudeau, F.: Deep multimodal fusion for semantic
1007 image segmentation: A survey. *Image and Vision Computing* **105**, 1–17 (2021)
- 1008 47. Zhou, W., Yue, Y., Fang, M., Qian, X., Yang, R., Yu, L.: Bcinet: Bilateral cross-modal
1009 interaction network for indoor scene understanding in rgb-d images. *Information Fusion*
1010 **94**, 32–42 (2023)