ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

REPORT

THE SPREAD OF THE BARK BEETLE IN EUROPEAN FORESTS

EPFL COURSEWORK

(IMAGE PROCESSING FOR EARTH OBSERVATION)

Mathurin Kiss (262585)
Philine Witzig (331038)

8th January 2021

Introduction

The bark beetle is an insect from the subfamily scolytinae. It lives and reproduces under the bark of trees which gives it the respective name. It particularly infests coniferous trees, especially spruce, and therefore this species is widely spread in the forests of central and northern Europe as well as in the forests of North America. Even though the bark beetle is a relatively small insect (2-5mm), it has a huge impact on the state of the forests which it lives in. In order to survive, the bark beetle feeds itself from the sap of the tree, which kills the tree from the inside.

Due to climate change, we experience more and more droughts in central and northern Europe. This also resulted in a great increase of the bark beetle population. Droughts expose trees to stress and limit them in the production of sap. Unhealthy and dying trees are not only a problem by themselves, they also have negative effects on the healthy ones. When a tree is defoliated it cannot grow anymore, but it still pumps the water from the soil and loses it by evapotranspiration. By doing so, it reduces the already little amount of water available for the healthy trees. This can become a dangerous cycle.

This development becomes even worse since unhealthy trees produce a stress hormone, which is recognized by the bark beetle. Unhealthy trees do not produce enough sap to defend themselves from insects, thus making it a perfect prey for the bark beetle.

Recently, there has been a succession of very warm summers followed by soft winters (2003-2006 and 2017-2020), so that the defoliating species were able to proliferate. That has led to a dehydration of the trees throughout Europe. From that point, a big amount of stress signals were sent, which allowed an unprecedented growth in population of the insects of the scolytinae subfamily. As expected this eventually led to a big amount of killed trees.

The measure taken from different governments to avoid a total spread of the beetle was to identify the "hotspots" where the concentration of the insects was the highest and to make clear cuts around them. However, sanitary cuttings come with some drawbacks. The higher is the infestation, the bigger has to be the cut, and the more damages are done to the ecosystems present in the area. Moreover, the costs related to such operations are relatively high, which represents another problem from a political point of view. For those reasons it is important to be able to identify those "hotspots" as early as possible before the concentration is too much spread.

The aim of this project is to use data from aerial imagery to identify those "hotspots". Based on that, we want to be able to monitor the evolution of the spread over time. In particular, we propose a general processing pipeline for aerial imagery which can be used to automatically identify infested areas in forests. Based on that, we develop a detector which can be used for early bark beetle infestation detection. Moreover, the processing pipeline allows for estimating the areas for sanitary cutting.

Background

In a study made by the Swiss Federal Institute for Forest, Snow and Landscape Research (WSL), researchers analyzed the typical life cycle of the scolytes [1]. The parent generation hibernates under the bark of the trees during the winter and when the temperature is about 16,5°C (roughly around April in the "plateau" of Switzerland) they fly to infest another tree to nest the first generation of the ongoing year. If the present nutritional resources are sufficient they can take another fly to nest the sister generation, which usually happens in June. The first generation will stay inside the tree, feeding itself from its sap and takes its fly around July to go to nest what will be the parent generation of the next year. The period where the concentration of beetles is the highest and therefore the damages done to the forest are the most important is during the month of August.

Thus, the development of the bark beetle depends on the temperature of the air. Every step will be delayed if the air is colder, or accelerated if the air is warmer. It is worth mentioning that the temperature of the air is highly related to the altitude.

From this analysis, we firstly concluded that the aerial imagery should be taken from September. Since the hibernation starts at the end of the month, the biggest part of the damages are already done for the year and they will be visible in the images. Furthermore, choosing this month has the advantage - compared to going for a date later in the year - that the damaged trees can be selected because of the beetle and not because of the natural cycle of trees loosing leaves during fall. Secondly, we decided to investigate on a region with a low altitude, because it has a better chance to show significant results.

Selection of a Region of Interest

For this project, aerial imagery was used. Images were collected from the satellite Sentinel-2 L1C [2] which can be accessed through Sentinel Hub [3]. Since we want to monitor the spread of the bark beetle as accurate as possible, a resolution of 10m was chosen. Ideally, an even finer resolution would have benefited our project. However such imagery, e.g. provided through Planet [4], comes with high costs and often didn't suffice our data acquisition criteria. We collected a region of interest (ROI) for testing our processing pipeline based on the following conditions: First, data had to be available regularly over a period of time of 2 years.Secondly, the blue, green, red and NIR bands could be individually selected. Thirdly, the bark beetle had to be present in the ROI.

Since the project is conducted as part of the IPEO class at EPFL in Lausanne, we first decided that the ROI would be a forest region near the Lac de Joux in the canton of Vaud. However, this area introduced a high amount of cloud coverage in the data available. Since we require data from different years collected at roughly the same point of the year, this location did not suffice to be analyzed properly. Therefore, it was impossible to monitor the development of the bark beetle in this area. Eventually another ROI, i.e. the forest Königsforst next to Cologne in Germany, was chosen. Since the climate in Germany and Switzerland is comparable, the assumptions made before still hold. The ROI was selected for the following reasons: First, we had evidence that the bark beetle was present in this area [5]. Secondly, cloud-free data was available for around the same point of time in the year for 2018 and 2020 (i.e. 06.10.2018 and 21.09.2020) which we required for change detection.

Procedure

We now propose the image processing pipeline for analyzing the spread of the bark beetle in European forests. The pipeline is fully implemented using Python. One major goal is to design this pipeline as general as possible, such that it can be easily transferred to new satellite imagery from different ROIs. The pipeline aims at realizing the following analysis tasks: First, it should identify infected "hotspots". Secondly, the pipeline should propose area for sanitary cutting based on these hotspots. Thirdly, it should be able to perform change detection over time.

To achieve these goals, we designed the image processing pipeline as suggested in Figure 1. We will go through the individual processing parts step by step now.

Data Acquisition First of all, satellite images had to be pulled through the API of Sentinel Hub [3], which we achieved by making use of the Python package `sentinelhub` [6]. For this, we set the timestamps to the dates mentioned above and store the three bands Blue, Green, Red and NIR in form of a `.npy` file. Furthermore, the longitude/latitude coordinates of the bounding box of the ROI had to be passed. For the described ROI in Germany, this corresponded to $[7.16842, 50.94826, 7.10489, 50.91959]$. The code

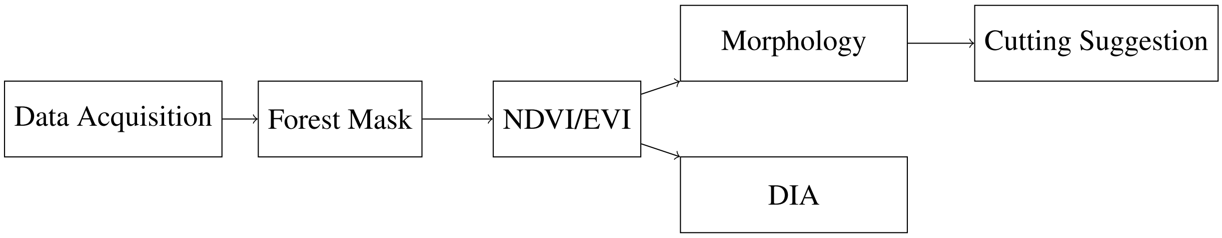


FIGURE 1
Image processing pipeline for analyzing the spread of the bark beetle.

required for data acquisition can be found in data.py. Configurations such as the credentials for the pull request, coordinates, dates, orbital periods etc. are located in config.py.

FOREST MASK We then suggest computing a forest mask on the first input image in time, e.g. in our particular example for 2018. With this step, we would like to be able to extract the forest area in the image only since the bounding boxes often contain non-forest areas, which later computations might be sensitive to, e.g. agriculture. For example, we don't want to take into account the change in the vegetation index of fields since they are exposed to rapid changes due to extensive land use throughout the entire year.

Originally, we aimed at solving this problem through an unsupervised approach, i.e. by applying k-means to the RGB image with two centers (corresponding to forest vs. non-forest). Our hope was that the RGB values between the forest region and the non-forest areas would differ in a way that k-means would pull the clusters to representative centers. The cluster centers were chosen randomly initially to have general approach that would work for any input. However, since k-means is highly sensitive to the initialization of cluster centers and the input images mostly consist of forest pixels, this lead to a segmentation of areas within the forest.

Thus, we changed our strategy to a supervised approach. We trained a support vector machine (SVM) for land cover/land use classification. The SVM was trained on the EuroSat dataset [7] made available on Kaggle [8]. The dataset consists in satellite imagery of size 64×64 taken from Sentinel2 at a resolution of 10m and thus suits our project. It contains ten land cover classes, such that each 64×64 patch is assigned to one class label. Since we are only interested in differentiating between forest and non-forest areas, we kept all samples of the forest class and selected a random subset of the remaining samples being instances of any of the other classes. It is worth mentioning that the random subset with the non-forest instances was of the same size as the forest class in order to have a balanced dataset.

We performed a 80:20 train-test split on the modified dataset. After training, the SVM achieved a precision of 0.98, a recall of 0.91 and an F1-score of 0.94 on the non-forest class. Furthermore, it achieved a precision of 0.91, a recall of 0.98 and an F1-score of 0.94 on the forest class. The code for training the SVM and reproducing the above results can be found in train_svm.py.

Applying the above forest mask to a new image will divide the input image into 64×64 patches and will assign one of the two labels to each patch. Figure 2 shows an example for such an application of he forest mask, where non-forest patches are indicated by the black squares.

NDVI/EVI The major goal of this project is to identify hotspots in forests which are infected by the bark beetle and to monitor the spread of the insect over time. To achieve this, vegetation indices are quite helpful as they can be used to represent the "healthiness" of vegetation. In particular, a spectral transformation of multiple bands is performed to enhance vegetation characteristics.

The normalized difference index (NDVI) is an example of a vegetation index commonly used to assess the

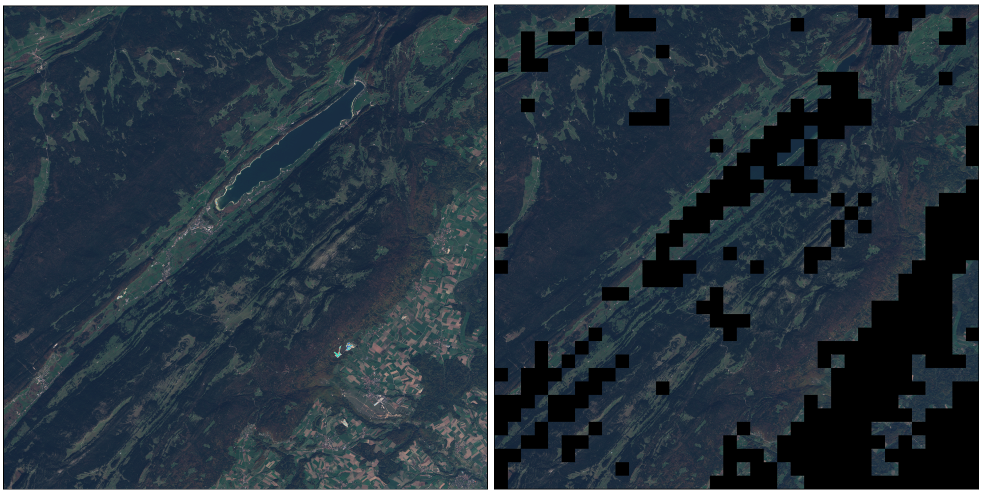


FIGURE 2
Application of the trained forest mask (right) to a test image (left) where black squares indicate the non-forest patches.

amount of live green vegetation. It is computed as stated in formula (1), where NIR and Red correspond to the bands we collect during the data acquisition phase.

$$NDVI = \frac{NIR - Red}{NIR + Red} \in [-1, 1] \tag{1}$$

The NDVI ranges between -1 and 1, where negative values correspond to clouds, water or snow due to a rather low reflectance in both spectral bands. The larger the NDVI value, the more dense and healthy the vegetation. For easier understanding, we discretize the continuous value to four bins corresponding to no vegetation ($[-1, 0]$), very unhealthy plants ($(0, 0.33]$), moderately healthy vegetation ($(0.33, 0.66]$) and very dense, healthy vegetation ($(0.66, 1]$) [9]. While the NDVI is easy to compute, it comes with disadvantage of being sensitive to atmospheric effects (e.g. amount of aerosols in the atmosphere).

Due to this limitation of the NDVI, we were interested in looking at another vegetation index which reduces the amount of atmospheric effects. We opted for the enhanced vegetation index (EVI) which claims to be more sensitive to dense vegetation [10]. It can be computed through formula (2). L, C_1, C_2 and G are coefficients set to $L = 1, C_1 = 6, C_2 = 7.5$, and $G = 2.5$.

$$EVI = G \cdot \frac{NIR - Red}{NIR + C_1 * Red - C_2 * Blue + L} \tag{2}$$

The EVI is not bounded as the NDVI. Thus, we discretize it to 20 values as suggested in [11].

Furthermore, We do not only compute the two indices stand-alone but are also interested in finding out whether on index is more suitable for change detection. This means that we would like to use the index, which is capable of detecting a change in the healthiness of the vegetation earlier. The code for this and every following processing step is located in processing.py.

SANITARY CUTTING Based on the vegetation index, the processing pipeline can be used to make predictions for forest sanitary cuttings. This refers to the process of identifying areas which are affected the most and proposing an area which should be cut. To achieve this, we opted for a simple approach, i.e. applying a cascade of morphological operators to a binary image (dead/unhealthy vegetation corresponding to 1, the rest to 0).

First, opening is applied to remove small responses for the dead/unhealthy vegetation. This is based on the assumption that small unhealthy areas are not as urgent as the larger infected parts, i.e. the "hotspots". Mathematically, opening is achieved through

$$S \circ I = (S \ominus I) \oplus I, \tag{3}$$

where I is the binary image and S the structuring element. $\oplus$ corresponds to dilation, $\ominus$ to erosion. This is followed by applying closing. This is done to fuse close responses for dead vegetation obtained above together. Closing can also close potential holes in detected unhealthy vegetation areas. Mathematically, closing is achieved through

$$S \bullet I = (S \oplus I) \ominus I, \tag{4}$$

where all parts have the same meaning as in equation (3). Last but not least, we can apply dilation in order to make the dead/unhealthy vegetation area larger. This can give the forester hints for the area for sanitary cutting. This is based on the assumption that immediately around a dead/unhealthy region, the trees are highly stressed and potentially infected as well. The size of the structuring element should correspond to the radius for sanitary cutting around the explicit hotspot done for bark beetles. The output of this last step of the pipeline branch will be areas that the forester should cut.

CHANGE DETECTION On the other hand, the pipeline can be used for monitoring the spread of the bark beetle over time. If we take this branch in the pipeline, we would like to compare two input images taken at two different timesteps. It is important to mention that the two input images will be processed differently. Let I_1 be the first image in time (from 2018 in our case) and I_2 the second image in time (from 2020 in our case). Then we only compute the forest mask on I_1 and apply this mask to I_2 since I_1 is our ground truth image for change detection. We then compute the vegetation index on the extracted forest areas for, both, I_1 and I_2. After discretizing the vegetation index values, we perform an image different analysis by simply subtracting the two images. Since we are interested in the area where the healthiness of the vegetation declined, we filter the difference image for positive values:

$$I_{diff} = I_1 - I_2 \tag{5}$$

$$I_{IDA} = \begin{cases} 1 & , I_{diff} > 0 \\ 0 & , I_{diff} \leq 0. \end{cases} \tag{6}$$

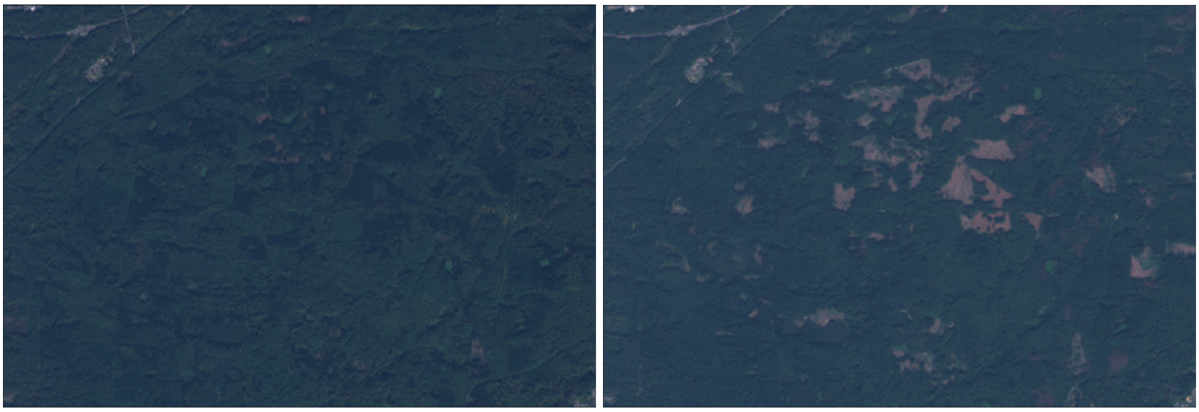
As a result, we obtain the binary image I_{IDA} where white pixels indicate areas where the forest's healthiness became worse.

RESULTS

We then applied this image processing pipeline to the ROI described above to perform the analysis tasks described in the Procedure section. We now report the results for each processing pipeline step.

For the data acquisition phase, we collected the four bands Red, Green Blue and NIR for two comparable timesteps, the first corresponding to 06.10.2018 and the second to 21.09.2020. We then applied the trained SVM to the input image which comes first in time, i.e. 2018 in this case, and apply the non-forest regions to the second input image. Figure 3 shows the enhanced RGB result (each channel multiplied by 3.5). We can observe an increased amount of brownish patches which could correspond to blades in the 2020 image compared to the 2018 image. Since the SVM classified every 64×64 patch in this stimulus to be forest, we do not have to reject areas for further computations.

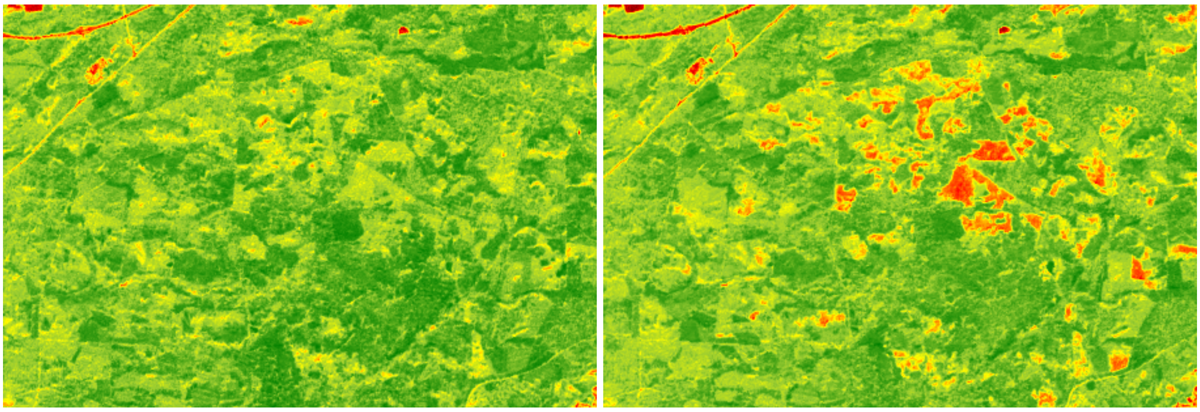
The processing pipeline then continues with computing the vegetation indices. Figure 4 visualizes the results after computing the NDVI with red corresponding to dead or very unhealthy vegetation and dark



.

(A) 2018 (B) 2020

FIGURE 3
RGB images after the first two processing steps.



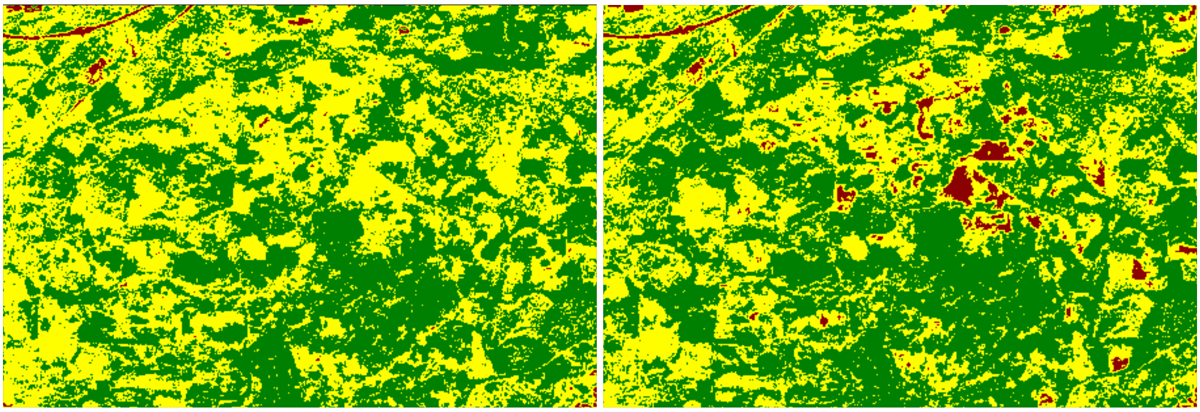
(A) 2018 (B) 2020

FIGURE 4
NDVI images (dark red = no vegetation, dark green = dense vegetation).

green to very dense and healthy vegetation. In Figure 5, we can see the discretized NDVI with the color encoding as above. In those images, we can observe that the amount of red areas increases between 2018 and 2020. In contrast, the amount of yellow area decreased and green patches became more dense. Figure 6 visualizes the output after computing the EVI, with dark green areas corresponding to very healthy vegetation and light green to white corresponding to very unhealthy or dead vegetation. We can also observe that the number of white areas increased, while in contrast the dark green areas became more dense.

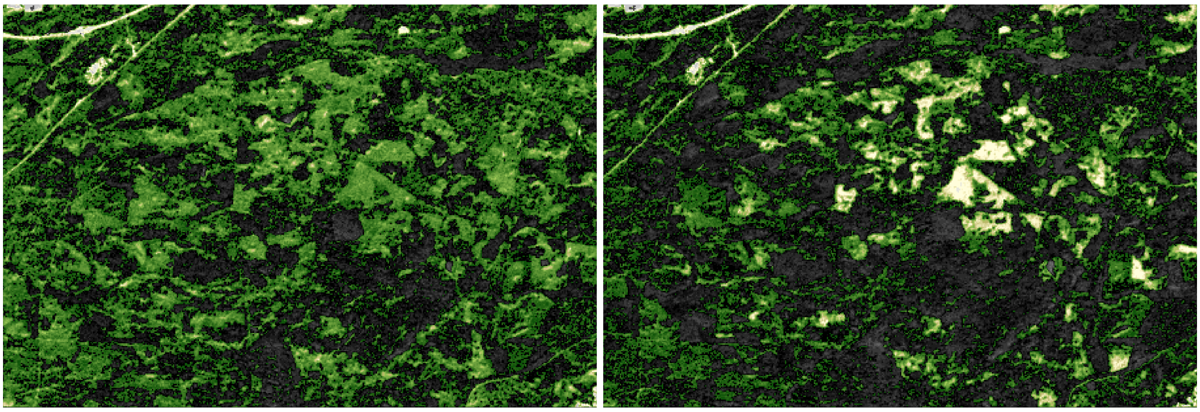
When following the top branch of the processing pipeline, we are interested in computing the cutting suggestion based on the first image in time. This was achieved through a cascade of morphology operators on the very unhealthy vegetation areas in the discretized vegetation index outputs. Figure 7 reports the outputs after each step of the morphology cascade applied to the unhealthy vegetation areas of the NDVI output. Figure 8 does the same, this time on the EVI output, respectively. The last image (D) contains the final cutting suggestion indicated through white pixels for both images. Comparing Figure 7 and Figure 8, we can observe that the EVI suggests more areas to be cut.

When following the bottom branch in the processing pipeline, we are interested in monitoring the change over time. Figure 9 reports the output of the image difference analysis for, both, discretized NDVI and



(A) 2018 (B) 2020

FIGURE 5
Discretized NDVI images (dark red = no vegetation, dark green = dense vegetation).



(A) 2018 (B) 2020

FIGURE 6
Discretized EVI images (white= no/dead vegetation, dark green = dense vegetation).

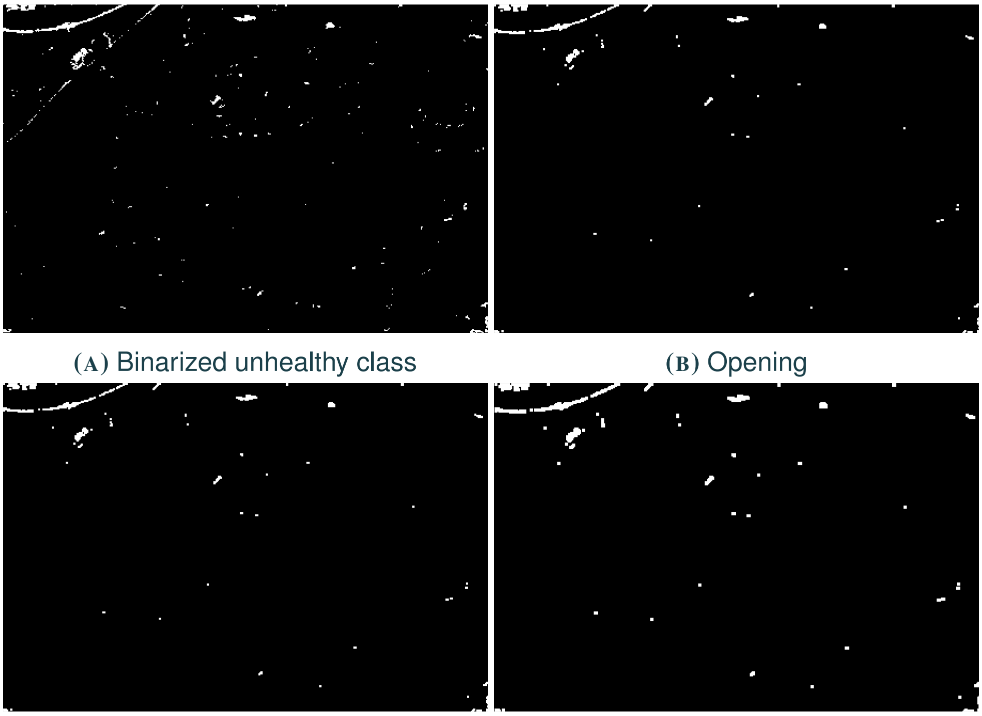
(A) Binarized unhealthy class

(B) Opening

(C) Closing

(D) Dilation

FIGURE 7
Morphological operation cascade on discrtized NDVI.



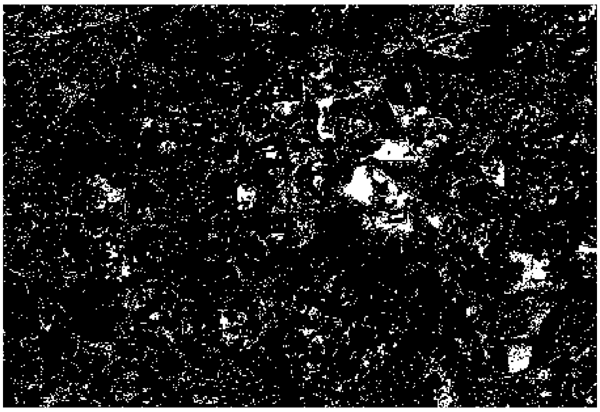
(A) Binarized unhealthy class

(B) Opening

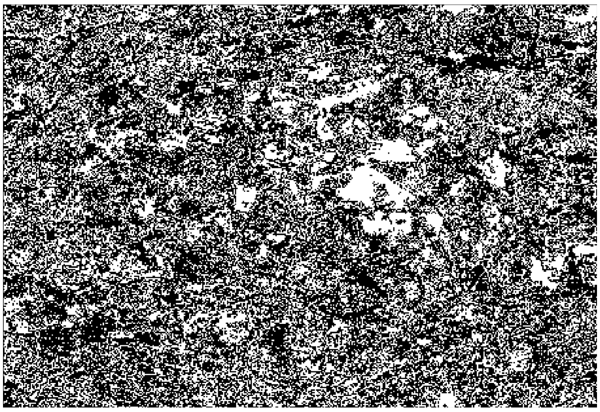
(C) Closing

(D) Dilation

FIGURE 8
Morphological operation cascade on discretized EVI.



(A) Discretized NDVI



(B) Discretized EVI

FIGURE 9
IDA on different vegetation indices.

discretized EVI. White pixels indicate a decline in the healthiness while black pixels indicate no change or an improvement. Comparing the IDA between the discrtized NDVI and the discretized EVI, the EVI reports more areas to have a decline in the healthiness of the vegetation.

Discussion

Let's now discuss the reported results. Starting with the data acquisition, it is very easy to spot the differences between the two years from Figure 3. Where there used to be a homogeneous forest in 2018, there appeared brown spots mostly in the center of the image in 2020. Considering how clean the border of those spots is and looking at their color it is reasonable to assume that there is no forest available anymore and that it was cut intentionally.

We now evaluate the quality of the forest mask. Every square that has been "blacked" in Figure 2 is indeed not containing any forest area. On the other hand, some squares have been kept despite the fact that they contain only a small fraction of forest inside of them. Thus, the main advantage of this mask is that it effectively removes large areas that are not forest, such as agriculture or water, while keeping the forest area integrally. Its main weakness is that it still keeps a reasonable amount of grass area, which can have an impact later when computing the vegetation indices. This issue could be solved by having a dataset of a higher resolution during training, i.e. patches smaller than 64×64 pixels. However, we did not find such a dataset for Sentinel2 L1C at 10m resolution.

In terms of the vegetation indices, the NDVI (cf. Figure 4) seems to give the same information as the RGB images at first sight. However on the 2018 image, it is already obvious that the forest that looked homogeneous is actually in different states of health. The yellow areas that are very present in the middle of the image are indeed a sign of stress. In the 2020 image, the condition even worsens as indicated by the red spots. This represents, as assumed from the RGB image, dead vegetation - or no vegetation at all. Something that is particularly interesting in this case is that, where there is a red spot in 2020, there used to be a yellow spot in 2018. With the discretized NDVI values of Figure 5, the conclusions made for the continuous NDVI values become more clear. Firstly, is it obvious how much yellow - unhealthy vegetation - is already present in 2018. It represents almost half of the image. Furthermore, it is also easier to notice that the red spots appeared on the large yellow areas in the middle of the image. Something new that also appeared is the increase of green areas, i.e. healthy vegetation. It suggests that the fact that vegetation was removed somewhere has somehow benefited to vegetation in different areas. Figure 6 represents the forest using the discretized EVI index. On this images it is possible to observe the same result that was observed using the discretized NDVI images, i.e. that the healthy vegetation is more dense in 2020 than it was in 2018. Thus, some areas seem to have recovered. These areas might contain tree types which are not too vulnerable to the bark beetle, i.e. trees other than spruce trees.

We now discuss the results of the morphology cascade for computing a cutting suggestion. We compare the suggestion after dilation made on the 2018 image taken from either discretized NDVI (Figure 5) or EVI (Figure 6). The white pixels indicate which parts should be cut. It appears that the EVI suggests that more areas should be considered than what the NDVI suggests. It is important to mention that the discretized EVI images considers 20 different values whereas the discretized NDVI only considers 4. Therefore, the EVI is more sensitive and will take more areas into consideration when categorizing them as unhealthy. Note that these discretization levels were taken from literature. Thus, one can say that the EVI is more conservative when it comes to computing a cutting suggestion compared to the NDVI, which is more optimistic. Last but not least, both indices have trouble identifying areas which are actually not vegetation, i.e. which might correspond to big paths (cf. bow in top left corner), and suggest to cut it. Having a finer forest mask might have solved this issue since we would not have processed this area then.

Looking at the change detection in Figure 9, this observation between the EVI and the NDVI is also supported since the EVI detects a lot more areas where the forest's health decreased (white pixels indicate decrease in health).

The result from this analysis is that the EVI should be the index used when determining a cutting suggestions and when monitoring change. Although we want the smallest part of the forest to be cut, we also want to take efficient measure against the beetle. Furthermore, the computed selection for cutting can be considered as reasonable considering the fraction of the forest it represents.

Finally, one major limitation of our proposed pipeline is that the vegetation index only cannot prove the presence of the bark beetle. It only represents the health status of forest and thus, we need extra references for its presence to use the vegetation indices for monitoring the beetles' spread. On the other hand, the main advantages of this method are its generalizability and transferability to any ROI.

Conclusion

We proposed a image processing pipeline which can be used to perform various analysis task related to the spread of the bark beetle. The pipeline was kept as general as possible to be able to apply it to any ROI. We tested the pipeline on a ROI in Germany for which we had evidence on the presence of the bark beetle.

The analysis indicated that in the ROI, the health of the forest between 2018 and 2020 declined. We can assume that (at least partially) the decrease in health is due to the the presence of the bark beetle. However, other factors such as droughts or storms could also contribute to this change. The pipeline outputs an explicit cutting suggestion, which is more conservative when we use the EVI and more optimistic in case we use the NDVI. In case resources are available, we would recommend following the suggestions based on the EVI to be sure we are in control of the beetle. Also in the change detection, the EVI identifies a stronger change. However, comparing results based on the two different vegetation indices should be taken with care since their discretized versions are based on a different number of values.

In the future, we could think of expanding this project by implementing a spread predictor. This means, we would train an auto-encoder on image pairs from two consecutive years. The auto-encoder would learn the change of the forest's health situation and could predict the change on a new query image. This could further help to make sanitary cuttings more precise compared to our trivial approach with a morphology cascade. However, since we were not able to find a suitable dataset for this, this would involve the creation of our own dataset which was not in the scope of this project.

BIBLIOGRAPHY

[1] WSL Ltd. *Bostryche*. 2020. URL: http://www.bostryche.ch/buchdrucker/biologie/index_FR (visited on 9th May 2020).

[2] *Sentinel Online: Level 1C*. ESA. URL: https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi/product-types/level-1c (visited on 6th Jan. 2021).

[3] Ltd. Sinergise Laboratory for geographical information systems. *Sentinel Hub*. 2020. URL: http://www.sentinel-hub.com (visited on 29th Dec. 2020).

[4] *Global Monitoring, Daily Insight*. Planet. 2020. URL: https://www.planet.com/products/monitoring/ (visited on 6th Jan. 2021).

[5] *Die Mär vom bösen Borkenkäfer*. Bündnis Heideterrasse e. V. 2019. URL: https://xn--knigsforst-ecb.net/die-maer-vom-boesen-borkenkaefer,608,1.html (visited on 11th Nov. 2020).

[6] Matej Aleksandrov. *sentinelhub-python*. 2020. URL: https://github.com/sentinel-hub/sentinelhub-py/ (visited on 29th Dec. 2020).

[7] Patrick Helber et al. 'Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification'. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 12.7 (2019), pp. 2217–2226.

[8] Gotam Dahiya. *EuroSat Dataset*. 2020. URL: https://www.kaggle.com/apollo2506/eurosat-dataset (visited on 31st Dec. 2020).

[9] *NDVI FAQ: ALL YOU NEED TO KNOW ABOUT INDEX*. Earth Observing System. 2019. URL: https://eos.com/blog/ndvi-faq-all-you-need-to-know-about-ndvi/ (visited on 2nd Jan. 2021).

[10] USGS. *Landsat EVI*. URL: https://www.usgs.gov/core-science-systems/nli/landsat/landsat-enhanced-vegetation-index?qt-science_support_page_related_con=0#qt-science_support_page_related_con (visited on 2nd Jan. 2021).

[11] *EVI (Enhanced Vegetation Index)*. SentinelHub. URL: https://custom-scripts.sentinel-hub.com/sentinel-2/evi/# (visited on 4th Jan. 2021).