

Article

A Novel Technique Based on Machine Learning for Detecting and Segmenting Trees in Very High Resolution Digital Images from Unmanned Aerial Vehicles

Loukas Kouvaras and George P. Petropoulos * 

Department of Geography, Harokopio University of Athens, El. Venizelou St., 70, 17671 Athens, Greece; kouvarasloukas1@gmail.com

* Correspondence: gpetropoulos@hua.gr

Abstract: The present study proposes a technique for automated tree crown detection and segmentation in digital images derived from unmanned aerial vehicles (UAVs) using a machine learning (ML) algorithm named Detectron2. The technique, which was developed in the python programming language, receives as input images with object boundary information. After training on sets of data, it is able to set its own object boundaries. In the present study, the algorithm was trained for tree crown detection and segmentation. The test bed consisted of UAV imagery of an agricultural field of tangerine trees in the city of Palermo in Sicily, Italy. The algorithm's output was the accurate boundary of each tree. The output from the developed algorithm was compared against the results of tree boundary segmentation generated by the Support Vector Machine (SVM) supervised classifier, which has proven to be a very promising object segmentation method. The results from the two methods were compared with the most accurate yet time-consuming method, direct digitalization. For accuracy assessment purposes, the detected area efficiency, skipped area rate, and false area rate were estimated for both methods. The results showed that the Detectron2 algorithm is more efficient in segmenting the relevant data when compared to the SVM model in two out of the three indices. Specifically, the Detectron2 algorithm exhibited a 0.959% and 0.041% fidelity rate on the common detected and skipped area rate, respectively, when compared with the digitalization method. The SVM exhibited 0.902% and 0.097%, respectively. On the other hand, the SVM classification generated better false detected area results, with 0.035% accuracy, compared to the Detectron2 algorithm's 0.056%. Having an accurate estimation of the tree boundaries from the Detectron2 algorithm, the tree health assessment was evaluated last. For this to happen, three different vegetation indices were produced (NDVI, GLI and VARI). All those indices showed tree health as average. All in all, the results demonstrated the ability of the technique to detect and segment trees from UAV imagery.

Keywords: UAVs; machine learning; Detectron2; tree detection; trees health



Citation: Kouvaras, L.; Petropoulos, G.P. A Novel Technique Based on Machine Learning for Detecting and Segmenting Trees in Very High Resolution Digital Images from Unmanned Aerial Vehicles. *Drones* **2024**, *8*, 43. <https://doi.org/10.3390/drones8020043>

Academic Editor: Diego González-Aguilera

Received: 2 January 2024

Revised: 27 January 2024

Accepted: 28 January 2024

Published: 1 February 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The study of tree characteristics is one of the most important applications in various ecological sciences, such as forestry and agriculture [1,2]. Assessing the health of trees is perhaps the most significant and widely used application in studies that focus on assessing tree properties. The health of trees can effectively be assessed using remote sensing data from satellite sensors. These sensors are essentially satellite cameras that utilize the spectrum from ultraviolet to the far-infrared. Therefore, the estimation of their health can be accomplished by calculating vegetation indices from remote sensing data collected in the electromagnetic spectrum of visible and near-infrared [3]. In recent years, studies evaluating the health of trees and vegetation using such data have been increasing, highlighting the contribution of remote sensing systems. Recent increased developments in unmanned aerial systems have further enhanced these applications by providing the capability to use very high-resolution spatial data [4].

In the recent past, the detection of individual trees was accompanied by significant difficulty, as it depended on the homogeneity of the data, and accuracy tended to decrease as the variety of natural elements increased [1]. This problem was possible to easily overcome by using hyperspectral data in combination with LiDAR (Light Detection and Ranging) data. However, challenges in tree detection persisted due to the influence of atmospheric conditions. The use of unmanned aerial vehicles (UAVs) successfully addressed this obstacle. UAVs have the capability to adjust their flight altitude, eliminating limitations posed by atmospheric conditions and cloud cover. Simultaneously, these aircraft are equipped with exceptionally high spatial resolution and can provide data quickly and reliably. Unlike satellite systems limited to specific orbits around the Earth, UAVs can swiftly capture and acquire aerial images, thus providing information much faster. Furthermore, these systems are easily adaptable to capturing images with the best possible clarity while considering noise limitations [5].

The isolation of tree information has traditionally relied on classifiers that categorized pixels based on their spectral characteristics. However, in recent years, these classifiers have been replaced by ML techniques, such as convolutional neural networks, which aim to identify objects in images [6,7]. Convolutional neural networks were originally designed for object recognition in ground-based images, but some researchers have adapted these algorithms for object recognition from aerial images using remote sensing data. Nevertheless, most researchers have focused on pattern recognition through computer vision for the detection of objects such as buildings or infrastructure with distinctive shapes such as airports, vehicles, ships, and aircraft [8,9]. For instance, Xuan et al. [10] used convolutional networks for ship recognition from satellite systems. One of the most popular applications is the detection and segmentation of trees in agricultural fields [11–13]. For this application, the use of remote sensing data from UAV platforms is recommended. The aerial images provided by such systems are particularly suitable for these applications due to their high spatial resolution. These applications are of significant utility in precision agriculture.

Many researchers have used convolutional neural networks for tree detection from RGB aerial photographs, aiming to assess the performance of those networks in applications involving a wide variety of tree species and developmental stages. Most of them have utilized the Mask R-CNN algorithm [7,14–16]. Additionally, many researchers and practitioners have applied these techniques to precision agriculture applications, such as assessing tree health [16] and estimating biomass [15]. The Mask R-CNN algorithm was preferred over others for object detection in remote sensing products due to its additional capability of object segmentation by creating masks, thereby providing precise object boundaries. Consequently, in this study, tree detection and segmentation were performed using the Detectron2 algorithm based on the Mask R-CNN algorithm. Detectron2 offers the ability to create masks for each detectable object and delivers better performance in terms of speed and accuracy [17].

By having precise boundaries for each tree, it is possible to estimate their health. This can be achieved by calculating specific vegetation indices using remote sensing data. These indices are computed by combining spectral channels in the visible and near-infrared regions. Vegetation appears to be better represented in the green and red channels of the visible spectrum and in the near-infrared channel, as these channels capture its characteristics and morphology effectively [18,19]. Therefore, vegetation indices can be utilized for assessing vegetation health, detecting parasitic diseases, estimating crop productivity, and for other biophysical and biochemical variables [20]. Calculating these indices from data collected by unmanned aerial vehicles (UAVs) provides extremely high accuracy in assessing vegetation health, eliminating the need for field-based research. The use of UAV systems for forestry and/or crop health using machine learning techniques for tree detection has been previously conducted by various researchers. For instance, Safonova et al. [15] and Sandric et al. [16] utilized images acquired from UAV systems to study the health of cultivated trees.

Based on the above, in the present study, an automated method is proposed for assessing vegetation health, providing its accurate boundaries. Accordingly, the Detectron2 algorithm is configured to read UAV imagery, recognize the trees present in it, and segment their boundaries by creating a mask of boundaries. The accuracy of the herein proposed algorithm is compared against the Support Vector Machine (SVM) method, which is a widely used and accurate technique for isolating information in remote sensing data. A further accuracy assessment step involves the comparisons of the outputs produced from the two methods against those obtained from the direct digitization method using photointerpretation. To the authors' knowledge, this comparison is conducted for the first time, and this constitutes one of the unique aspects of contribution of the present research study.

2. Study Area

The study area is a cultivated citrus orchard located in Palermo, Sicily, Italy, at coordinates $38^{\circ}4'53.4''$ N, $13^{\circ}25'8.2''$ E (Figure 1). Each tree occupies an area of approximately 5×5 m, with a planting density of 400 trees per hectare of land. The climate in this region is typically Mediterranean semi-arid, characteristic of the central Mediterranean. The area is situated at an elevation of 30 to 35 m above sea level, with a slope ranging from 1% to 4%. The study area is divided into two sections separated by a dirt road, with each section covering an area of about 4000 square meters. These two sections differ in terms of irrigation, with the southern section receiving significantly less irrigation compared to the northern section [4]. The selection was based on the fact that we had already multispectral UAV imagery available for the site which resulted from an activity performed within the EU-funded HARMONIOUS research project (<https://www.costharmonious.eu/>, accessed on 15 January 2024).

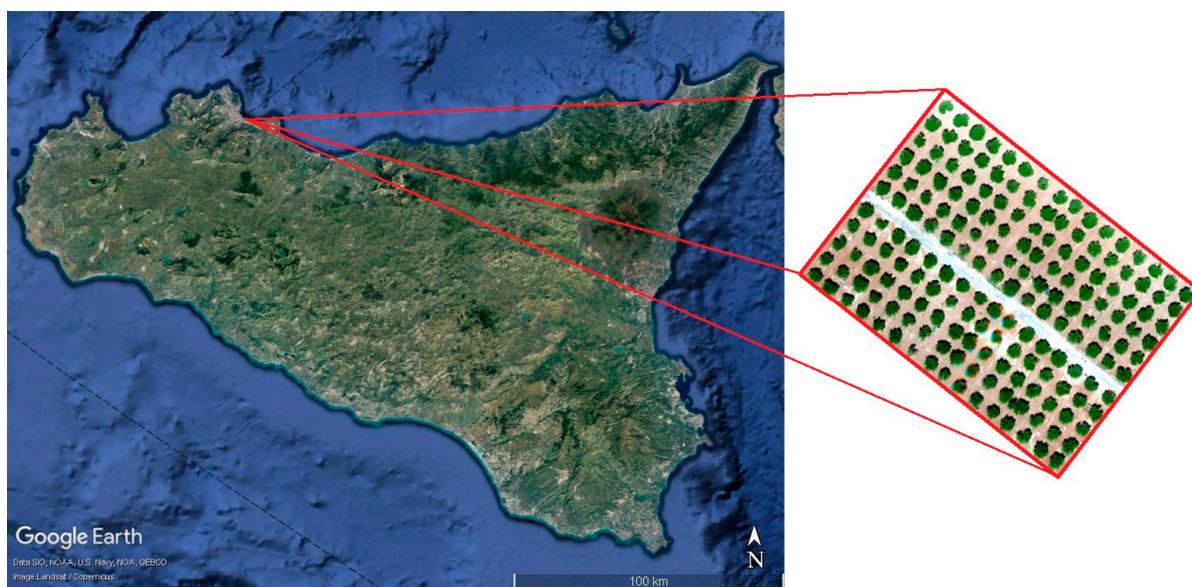


Figure 1. Study area, Palermo, Sicily. Cultivated area with citrus which divided into two sections, with the northern part receiving more extensive irrigation compared to the southern part.

3. Data and Pre-Processing

3.1. Data Acquisition

The data utilized for this study consisted of imagery captured using an unmanned aerial vehicle (UAV) in July 2019. Precise coordinates of the area were calculated to facilitate the image capture. To achieve this, nine black and white ground control points, along with nine aluminum targets, were strategically placed, creating a grid across the entire cultivation area. The coordinates were obtained using an NRTK (Network Real-Time

Kinematic) system with a Topcon Hiper V receiver, simultaneously utilizing both GPS and Glonass systems. Multispectral images were acquired using a NT8 contras octocopter carrying a RikolaDT-17 Fabry-Perot camera (manufactured by Rikola Ltd., Oulu, Finland). The multispectral camera has a 36.5° Field of View. It was set up to acquire images in nine spectral bands with a 10 nm bandwidth. The central wavelengths were 460.43, 480.01, 545.28, 640.45, 660.21, 700.40, 725.09, 749.51, and 795.53 nm. At a flight altitude of 50 m above ground (a.g.l), the average Ground Sampling Distance (GSD) was 3 cm [21,22]. Eight of these channels cover portions of the visible spectrum, while one is in the near-infrared range. For more detailed information regarding data acquisition and preprocessing, one can refer to the publication by Petropoulos et al. in 2020 [4].

3.2. Pre-Processing

The pre-processing of the UAV imagery was carried out as part of the implementation of a previous scientific research study by [4]. To orthorectify the multispectral and thermal images, a standard photogrammetric/SfM approach was applied via Pix4D mapper (by Pix4D Inc., Denver 4643 S. Ulster Street, Suite 230, Denver, CO 80237, USA). Thus, based on the GPS and Glonass systems, ground control points were geometrically corrected. The average position dilution of precision (PDOP) and the geometric dilution of precision (GDOP) were 1.8 and 2.0, respectively. The control targets were positioned with an average planimetric and altimetric accuracy of ± 2 cm, which can be considered within acceptable geometrical configuration limits to orthorectify the UAV images, considering that these latter are characterized by a spatial resolution of 4 cm once orthorectified. The spectral channels of the visible and near-infrared were calibrated according to the ground reflectance using the empirical line method, as it allows for simultaneous correction of atmospheric effects. For a more comprehensive description of data pre-processing, the reader is pointed to [4].

4. Methodology

4.1. Tree Crown Detection Using a Machine Learning Model

During the first part of the analysis (Figure 2), the objective is to detect the boundaries of trees through the training of a Detectron2 algorithm written in the Python programming language. This algorithm represents a parameterized (tailored to the needs of the current study) version of the Detectron2 algorithm. By taking images in which objects have been defined, essentially setting boundaries on their image elements, the algorithm undergoes training and subsequently gains the ability to autonomously set the boundaries of these objects in new images. The training for tree detection was conducted using an already trained model capable of recognizing various living and non-living objects such as humans, cars, books, dogs, airplanes, etc. It should be noted that the class of trees is absent from this specific model. Even if this class were present, it would not be able to recognize the trees in aerial photographs, as the model is not trained to observe objects from a vertical perspective above the ground. The algorithm's code was developed herein in the Google Colab environment due to the availability of free GPU resources it provides, with the code running on Google's servers.

It is essential for the training process to define the boundaries of trees so that the algorithm can correctly identify whether the object presented to it is a tree or not. To facilitate the algorithm in terms of the computational power required, the aerial photograph was divided into equal-sized regions (300 × 300 pixels). This division was performed using the "Raster Divider" plugin in the QGIS environment. Tree boundary delineation was accomplished through digitization using the "label studio" program. This choice was made because it has the capability to export data in COCO (Common Objects in Context) format [23]. This format was necessary, as it is the most widely used format for defining large-scale data for object detection and segmentation applications in computer vision utilizing neural networks. The more trees presented to the algorithm, the more accurate it becomes in detecting them. Therefore, two-thirds of the trees in the area were utilized

for training. The code sequence result described is the storage of images representing the boundaries. The next necessary step is to extract this information. The images were first georeferenced to restore their coordinates (which were lost during their introduction to the algorithm). Afterward, they were merged into a new mosaic. Subsequently, only the information presenting the trees was isolated. The georeferencing of the images and the conversion into a unified mosaic were carried out in the QGIS environment, while tree extraction was performed in SNAP software. Finally, this information was transformed from raster format to vector format.

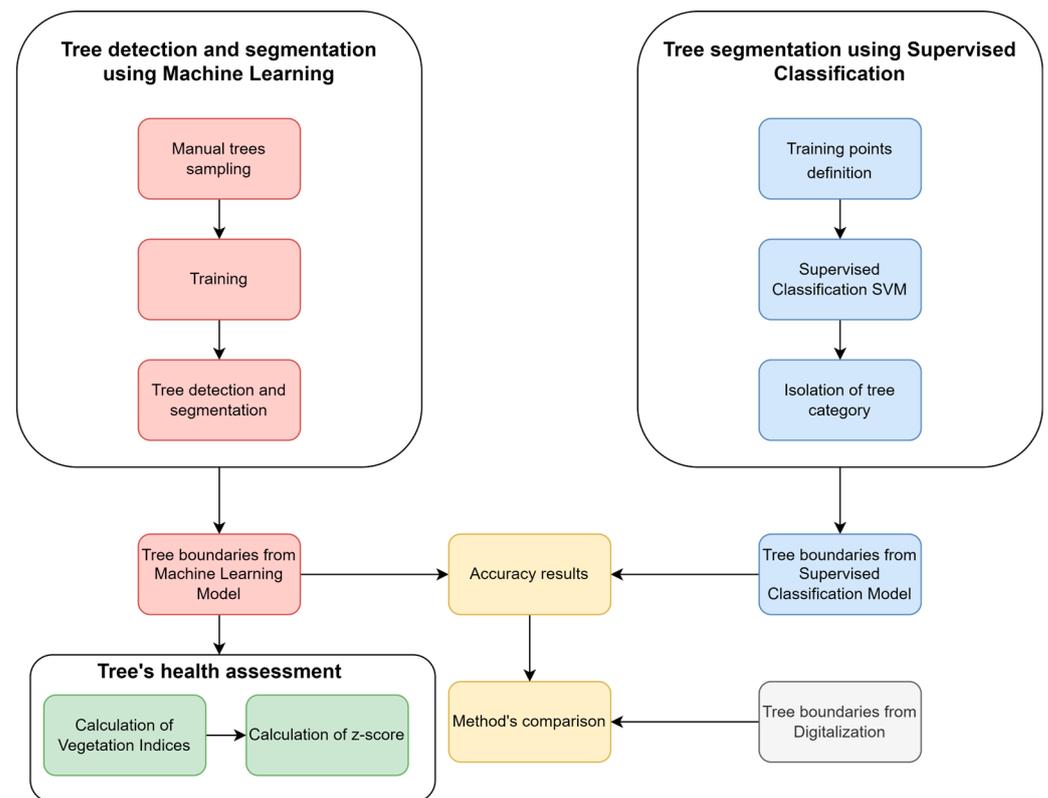


Figure 2. Flowchart of the methodology.

4.2. Tree Crown Detection Using Supervised Classification

It is crucial to examine the accuracy of the algorithm, i.e., its ability to correctly delineate the trees it detects, compared to other methods for isolating this information. Therefore, in this study it was decided to compare the results against those of the supervised classification method. Such a method relies on the reflective characteristics of image pixels (pixel-based) in an image [24]. Specifically, the chosen supervised classification method is SVM, which is based on machine learning. The algorithm essentially uses kernel functions to map non-linear decision boundaries from the primary data to linear decisions in higher dimensions [25]. It was selected, among other methods, because it is associated with high precision results and handles classes with similar spectral characteristics and data with high noise (Noisy Data) more effectively [26].

For the implementation, spectral characteristics of each pixel in the aerial photograph were utilized across all available channels, including eight channels in the visible spectrum and one channel in the near-infrared spectrum. Each pixel was classified into one of the predefined categories (classes) based on the entities (natural and man-made) observed in the aerial photograph. These categories were as follows:

- Trees.
- Tree shadow.
- Grass.

- Bare Soil.
- Road.

In the class of trees, it is expected that all the trees in the aerial photograph will be classified. This is the most crucial class because it is the one against which the algorithm will be compared. Due to the sun's position, most trees cast shadows that are ignored by the Detectron2 algorithm. Therefore, it is important to create a corresponding class (tree shadow) to avoid any confusion with the class of trees. Additionally, three other classes were created to represent the soil, the grass observed in various areas, and the road passing through the cultivated area. For each class, a sufficient number of training points (ROIs) are defined (approximately 6000 pixels) that are considered representative. Thus, the values displayed by these points in each spectral channel of the image are considered as thresholds for classifying all other pixels. It is essential to examine the values of the training points to ensure they exhibit satisfactory separability to avoid confusion between the values of different classes. During this process, the spectral similarities of the training points are compared to determine how well they match. The lower the similarity between two spectral signatures, the higher the separability they exhibit. The supervised classification process results in a thematic map where each pixel of the aerial photograph has been classified into one of the five classes. From those classes, only the information related to the class of trees that are detected is isolated. Thus, this information, as was the case with the algorithm, is transformed from a raster into a polygonal format.

4.3. Calculation of Accuracy for Each Method and Method Comparison

4.3.1. Accuracy of the Machine Learning Algorithm

It is important to assess the accuracy of each method individually and to compare the accuracy of the results among them. The accuracy of the Detectron2 algorithm was evaluated in terms of its correct recognition of objects, based on three statistical indicators that were calculated [7,16].

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Negative} \times 100 \quad (1)$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Positive} \times 100 \quad (2)$$

$$F1\ score = \frac{2 * Precision * Recall}{Precision + Recall} \times 100 \quad (3)$$

where Precision is calculated as the number of correctly identified trees (True Positive) divided by the sum of correctly identified trees (True Positive) and objects that were incorrectly identified as trees (False Negative). Recall is calculated as the number of correctly identified trees (True Positive) divided by the sum of correctly identified trees (True Positive) and the number of objects that were identified as trees in areas with zero tree presence (False Positive). The F1 score represents the overall accuracy (OA) of the model and is calculated as twice the product of the two aforementioned indicators divided by their sum.

4.3.2. Supervised Classification Accuracy

Regarding supervised classification accuracy, it is important to calculate its OA, i.e., how well the pixels of the aerial image have been classified correctly into the various assigned classes [24,27,28]. To achieve this, certain validation samples (approximately 2000 pixels) are defined for each class. The pixels of the validation samples are evaluated for their correct classification into each of the classification classes, thus giving the OA. A significant factor in the accuracy presented by the classification is the Kappa coefficient. This index indicates to what extent the accuracy of the classification is due to random agreements and how statistically significant it is [29]. It is beneficial to examine the producer's accuracy (PA) and user accuracy (UA). The PA shows how well the training points of the classification

were classified into each class. It is calculated as the percentage of correctly classified pixels compared to the total number of control pixels for each class. UA indicates the probability that the classified pixels actually belong to the class into which they were classified. It is calculated as the percentage of correctly classified pixels compared to the total number of classified pixels for each class.

4.3.3. Comparison of the Two Methods

It is important to compare the Detectron2 algorithm with one of the most common methods for vegetation boundary delineation in aerial imagery, the supervised classification method, in order to assess its performance. The results of this comparison demonstrate the performance of the model, highlighting its significance. The two models are compared in terms of their accuracy with the digitization method. While the results of digitization, although more accurate as they are carefully executed by the researcher, lack automation and are thus considered a very time-consuming method. Therefore, three accuracy indexes are calculated [30,31].

$$\text{Detected area efficiency} = \frac{\text{Detected area}}{\text{Detected area} + \text{Skipped area}} \quad (4)$$

$$\text{Skipped area rate} = \frac{\text{Skipped area}}{\text{Detected area} + \text{Skipped area}} \quad (5)$$

$$\text{False area rate} = \frac{\text{False area}}{\text{Detected area} + \text{False area}} \quad (6)$$

From the above equations, the detected area is the common area between the trees generated by the algorithm and the trees resulting from the digitization process. The false area (or commission error) is the area present in the algorithm's trees but absent from the digitization trees. Finally, the skipped area (or omission error) is the area in the digitization trees that exists but is missing in the algorithm's trees. The above equations, in addition to the algorithm's trees, are also calculated for the trees that result from the classification process. This enables the comparison of the two methods.

4.4. Assessment of Vegetation Health in Detected Trees

The next step involved the calculation of the health of the trees identified by the Detectron2 algorithm in the second section. This step essentially constitutes an application whereby the detection and segmentation of trees within a cultivation could find a useful application. Such an application could provide quick updates to the grower regarding the health status of their crop, enabling them to take necessary actions. To calculate the health, three different vegetation health indices are defined: NDVI, VARI, and GLI. The implementation steps are described in more detail next.

4.4.1. Normalized Difference Vegetation Index (NDVI)

The index presented by Rouse et al. (1974) [32] was introduced to indicate a vegetation index that distinguishes vegetation from soil brightness using Landsat MSS satellite data. This index minimizes the impact of topography. Due to this property, it is considered the most widely used vegetation index. The index is expressed as the difference between the near-infrared and red bands of visible light, divided by their sum. It yields values between -1 (indicating the absence of vegetation) and 1 (indicating the presence of vegetation), with 0 representing the threshold for the presence of vegetation [32].

$$NDVI = \frac{NIR - Red}{NIR + Red} \quad (7)$$

4.4.2. Visible Atmospherically Resistant Index (VARI)

Introduced by Kaufman & Tanre (1992) [33], this index was designed to represent vegetation among the channels of the visible spectrum. The index is less sensitive to atmospheric conditions due to the presence of the blue band in the equation, which reduces the influence of water vapor in the atmospheric layers. Because of this characteristic, the index is often used for detecting vegetation health [33].

$$VARI = \frac{Green - Red}{Green + Red + Blue} \quad (8)$$

4.4.3. Green Leaf Index (GLI)

Presented by Lorenzen & Madsen (1986) [34], this index was designed to depict vegetation using the three spectral channels of visible light, taking values from -1 to 1 . Negative values indicate areas with the absence of vegetation, while positive values indicate areas with the presence of vegetation [34].

$$GLI = \frac{(Green - Red) + (Green - Blue)}{2 \times Green + Red + Blue} \quad (9)$$

4.4.4. Standard Score

For the calculation of vegetation health, the Standard Score, or z-score, of all three vegetation indices is computed. The z-score is calculated as the difference between the values of each tree's vegetation index and the mean value of these indices for all trees, divided by the standard deviation of these indices for all trees.

$$Standard\ score = \frac{Value - Mean}{Standard\ deviation} \quad (10)$$

5. Results

5.1. Results of Machine Learning Algorithm

The sequence of the algorithm's code for the detection of trees through machine learning, as described in the first part of the methodology, resulted in the detection and segmentation of the cultivation trees (see Figure 3). The algorithm successfully recognized all the trees (a total of 175) within the land area, applying case segmentation (see Figure 4). All the trees were successfully categorized under the 'Tree' class without any other classes that existed in the training model being presented. All the images that were inputted at the beginning of the code were processed in such a way that the boundaries of the trees were presented through the algorithm's prediction. However, the output obtained was not the result of the case segmentation. For the sake of easy isolation of tree information, each tree was given a red hue (see Figure 4).

5.2. Results of Supervised Classification

The sequence of the SVM supervised classification process resulted in the classifying of each pixel of the aerial image into one of the predefined classes. To achieve this, specific training points were assigned for each class separately. These points served as reference points for classification, based on their spectral characteristics, for all the remaining pixels. It is important to note the presence of good separability between the training points to avoid confusion between the assigned classes. The separability between points of different classes exhibited values ranging from 1.92 to 2.

Only the class representing the trees in the area was isolated from the classification classes (see Figure 5).



Figure 3. Detected trees using the Detectron2 algorithm.

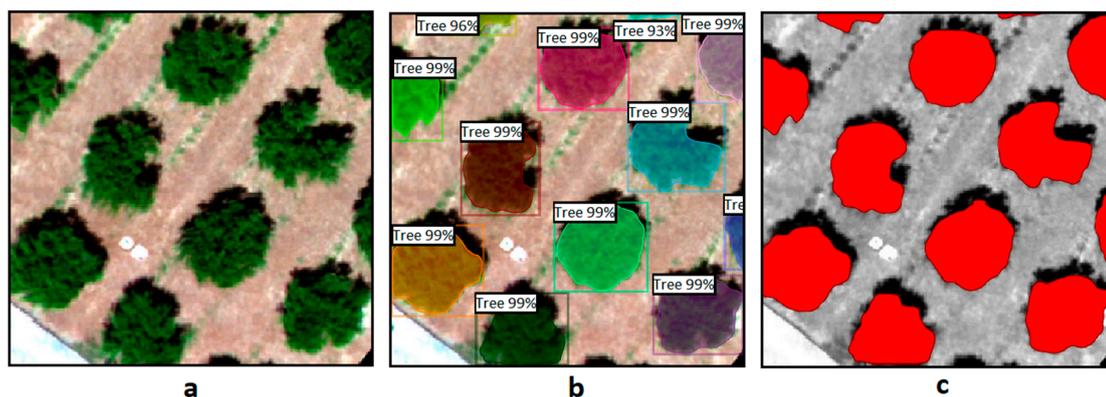


Figure 4. (a) The image as input to the algorithm, (b) case segmentation, and (c) the image as output from the algorithm.

5.3. Accuracy Results and Method's Comparison

The detection accuracy of the Detectron2 model in correctly identifying objects exhibited an F1 score of 100%. This is attributed to the fact that the algorithm correctly detected all the trees without detecting other objects as trees or detecting any trees in areas where they were not present. Regarding classification accuracy, when comparing the training points with the validation points, it is demonstrated that the OA reaches 97.0332%. This means that 97% of the classified pixels were correctly classified based on the validation points. The Kappa coefficient index shows a value of 0.959, indicating a high level of agreement in the results. Such a strong agreement relationship suggests that the overall classification accuracy is not due to random agreements during validation, and, therefore, the OA is statistically significant. According to Table 1, the majority of classes exhibited high PA, indicating that the training points of the classification were generally correctly classified into each defined class. However, the grass class presented the lowest percentage (86.1%). As for UA, it also showed high percentages in most classes, indicating a high likelihood that the classified pixels indeed belong to the class they were classified into. However, here again, the class of grass had a lower percentage (65.8%).



Figure 5. Detected Trees from SVM Supervised Classification.

Table 1. Results from the classification accuracy assessment: PA & UA.

Classes	PA (%)	UA (%)
Tree	92.2	98.73
Tree shadow	100	100
Grass	86.1	65.8
Bare Soil	100	98.93
Road	99.18	99.55

The comparison between the two boundary detection methods is made by comparing these methods with the digitization method, which is considered the most accurate method of creating boundaries. Whichever method shows results closest to the digitization process, which represents the most accurate boundary creation method, will be considered the most accurate. As presented in Table 2, the Detectron2 method exhibits a higher detection area efficiency, with a percentage of 0.959% compared to the 0.902% displayed by classification. This method also shows a lower skipped area rate compared to classification, with percentages of 0.041% and 0.097%, respectively. In the case of the false area rate index, machine learning seems to have a higher value at 0.056%, while classification performs better with a percentage of 0.035%. The conclusion drawn from these three indices is that the Detectron2 method yields more accurate results as it exhibits a larger common (detected) area with the digitization method while simultaneously showing a smaller skipped area. However, it should be noted that it also exhibited a higher false detected area compared to the supervised classification method. This result is attributed to the fact that the algorithm appeared to be weaker in detecting abrupt changes in the tree boundaries caused by their branches, resulting in the trees being more rounded in shape (see Figure 6).

Table 2. Accuracy indices of the two methods.

Method	Detected Area Efficiency (%)	Skipped Area Rate (%)	False Area Rate (%)
Detectron2	0.959	0.041	0.056
SVM	0.902	0.097	0.035

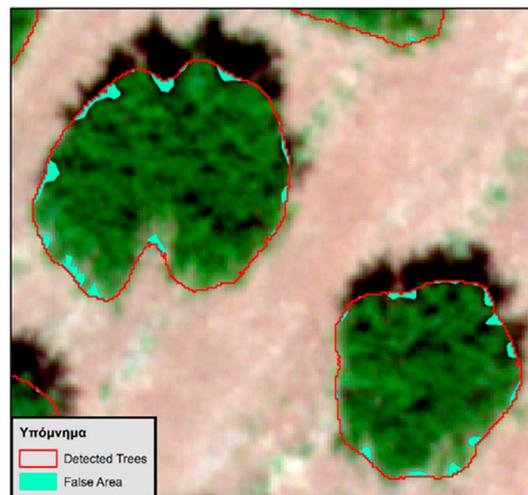


Figure 6. Illustrative example of the detected false area index by the algorithm.

5.4. Assessment of Vegetation Health

In order to calculate the vegetation health of the individual trees, three indices were applied, representing this parameter. These indices are the NDVI, GLI, and VARI. For the assessment of health, a standard score was calculated for each of these indices. Values of this index greater than $+1.96$ indicate very healthy vegetation, while values of -1.96 indicate low vegetation health values for a 95% confidence level. The results of all three vegetation indices showed standard score values within the set thresholds. This means that the trees were characterized by moderate health. Only one tree exhibited good vegetation health in the GLI index; however, this information may be erroneous, as this particular tree was located at the boundaries of the aerial photograph, where strong ground reflectance was observed, and the values reported for the tree may be inaccurate. In general, the indices calculated exclusively from the visible spectrum channels (GLI & VARI) yielded similar tree health results. However, these results seem to be contradictory to those of the NDVI index (see Figure 7).

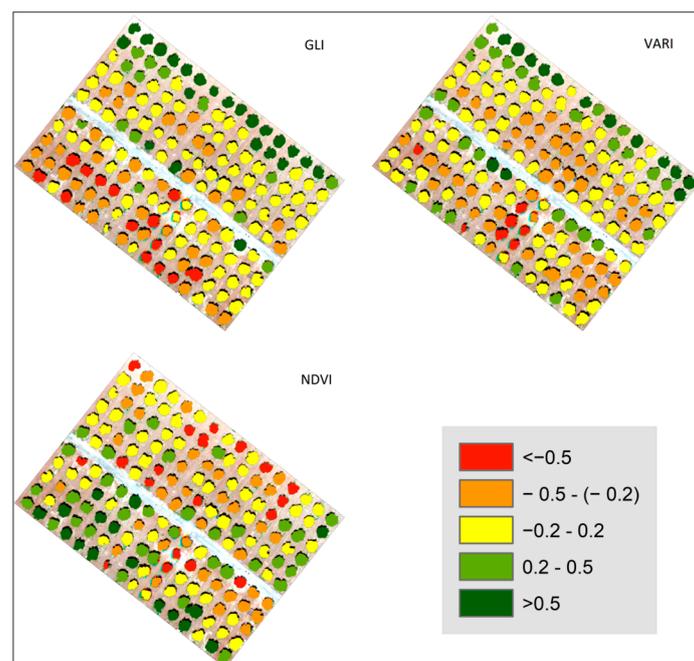


Figure 7. Assessment of vegetation health for the three indices.

6. Discussion

Over the past decade, the use of unmanned systems in precision agriculture has experienced significant growth for monitoring crops and real-time disease prevention [35,36]. Utilizing artificial intelligence has brought unprecedented computational time savings, adding an automation element. This study successfully managed to detect and segment the boundaries of trees presented in an UAV aerial image through an automated process, by fine-tuning a ML object detection algorithm. This algorithm demonstrated exceptionally high accuracy in tree recognition, achieving a 100% F1 score. It correctly identified all trees without classifying other objects as trees or falsely detecting trees in areas with no presence of trees. In comparison, other similar studies where soil and vegetation colors were very similar also reported very high accuracy [7,14–16]. This highlights the high precision exhibited by convolutional neural networks in such applications.

The supervised classification method of SVM demonstrated practical high accuracy in classifying pixels based on their spectral characteristics. The separability of the training points yielded excellent results, with the majority of them approaching absolute separability (a value equal to 2). The OA reached a very high percentage, and the kappa coefficient showed a very satisfactory result. Both the PA and UA indices exhibited high accuracy in the context of efficient classification. Their results indicated a generally high likelihood of correctly classifying the training points into various classes. However, the classes of trees and grass showed lower values in these indices. It is highly likely that the pixels pertaining to this class may belong to both the tree and grass classes and vice versa. These two classes are closely related as they both belong to the vegetation of the area. Therefore, the spectral signatures of their pixels have the highest similarity compared to any other pair of classes, potentially leading to confusion between these two classes.

The accuracy of the algorithm's model, compared to the supervised classification SVM model, was found to be superior in segmenting the boundaries of trees. When comparing the results of these two models with the digitization process, the algorithm's model exhibited higher accuracy in the common boundary detection index and the missed boundary detection index. However, the classification model showed greater accuracy in the false boundary detection index. This can be attributed to the fact that the algorithm appeared to be less capable of detecting abrupt changes in the boundaries of trees, making them appear more rounded.

The estimation of vegetation was conducted using three different vegetation indices. Two of these indices were calculated exclusively from the visible spectrum channels. Although the VARI index is primarily used for detecting the vigor of trees and the development of their branches, and the GLI index is used for detecting chlorophyll in their foliage, the two indices did not show significant differences. Despite the small difference between them, the GLI index is considered more reliable due to its ability to capture the interaction of the green spectral channel with the tree foliage [16]. On the other hand, the NDVI vegetation index, which includes the near-infrared channel in its calculation, appeared to produce contrasting results compared to the other two indices. The major differences were observed mainly in extreme values, with trees that exhibited relatively high vegetation health in the two visible indices showing relatively low vegetation health in the NDVI, and vice versa. However, the NDVI index is considered the most reliable among the three indices because it utilizes the near-infrared spectrum, which provides better information about vegetation characteristics and morphology than other parts of the spectrum [18,19]. Regarding tree health, all three indices indicated their health as moderate, as none of the indices showed extreme z-score values at a 95% confidence level.

Another particularly important factor influencing the results of this study is the flight conditions of the UAV, such as flight altitude and acquisition angles. The flight altitude of the UAV is proportional to the spatial resolution of the UAV's imagery. Segmentation methods, in this case Detectron2, are influenced by the spatial resolution of the imagery. As the altitude of the UAV's imagery increases, the detection accuracy of the algorithm will be deteriorated. In addition, increased altitude can lead to geometric distortions and

reduced image resolution, further complicating object identification. Apart from flight altitude, the imagery acquisition angle also impacts the results and the performance of the algorithm. UAV images captured from non-vertical angles can introduce perspective distortion, making it difficult to accurately determine object dimensions and orientations. For example, in a recent study by [37] are discussed the key challenges in UAV imagery object detection, including small object detection, object detection under complex backgrounds, object rotation, and scale change. As the detection and segmentation effect of deep learning (DL) algorithms from UAV images is affected by the different heights and shooting angles of the UAV images acquisition, it would be interesting to perform an evaluation of the algorithms herein at different altitudes and shooting angles to determine its sensitivity to changes in object size and image quality.

All in all, from the results obtained herein, it is evident that the Detectron2 algorithm, upon which the analysis algorithm relied, is capable of object detection and segmentation in aerial images beyond ground-based images. The detection of trees achieved perfect accuracy (F1 score = 100%), and the segmentation of their boundaries also yielded satisfactory results. However, it is essential to note that the algorithm may exhibit reduced accuracy when dealing with tree species that differ significantly in shape and color from citrus trees. This limitation of the model can be effectively resolved through further training on other types of trees. Although the training process requires a significant amount of time for both data preparation and computational time, the more the algorithm is trained, the more accurate it becomes. Therefore, while training the algorithm demands time and a vast amount of data, after extensive training on various datasets, the algorithm will be capable of detecting and segmenting objects with the same accuracy as a human researcher. This process clearly demonstrates the significant contribution of artificial intelligence to the automation of processes through machine learning in the context of UAV data exploitation for environmental studies.

7. Conclusions

The present study proposed a new algorithm to detect and segment the boundaries of trees in an aerial image of a cultivated area with mandarin trees using an automated approach, using a ML algorithm developed in the Python programming language. The outcome of the algorithm was utilized in assessing tree health. The innovation of the methodology is highly significant, as it demonstrates that, by employing artificial intelligence techniques, it is possible to create a tool for automated object recognition and boundary delineation in an image, thereby saving valuable time for researchers. Moreover, perhaps for the first time, there is a presentation of the comparison of the accuracy of an object segmentation algorithm with the SVM method.

Following the methodology presented herein, the detection and segmentation of trees in the cultivated area became feasible. The use of the tool makes it possible to save valuable time by offering automation for a similar research study, where object detection and segmentation would otherwise be performed manually by the researcher through the digitization process. The algorithm successfully detected all the trees presented in the study, assigning the correct classification category to each tree without falsely categorizing them into any other existing categories of the training model. The detection accuracy was exceptionally high, achieving an F1 score of 100%. The comparison of the accuracy results shows that the Detectron2 algorithm is more efficient in segmenting the relevant data when compared to the supervised classification model in the indices of common detected and skipped area rate.

In summary, this study credibly demonstrated the significant contribution of artificial intelligence in precision agriculture, utilizing high-resolution data for the timely monitoring of crop conditions. The combination of remote sensing data with ML algorithms provides cost-effective and rapid solutions, eliminating the need for field research while introducing automation. Technological advancements have led to unprecedented growth in agriculture development and modernization over the last decade. Artificial intelligence, as exemplified

in this article, is poised to further accelerate this progress by offering solutions that save both time and money. The development of artificial intelligence has already demonstrated its value and is expected to gain higher recognition in the future as it finds applications in an increasingly diverse range of fields. The latter remains to be seen.

Author Contributions: Conceptualization, L.K. and G.P.P.; methodology, L.K. and G.P.P.; software, L.K.; validation, L.K.; formal analysis, L.K.; data curation, L.K.; writing—original draft preparation, L.K.; writing—review and editing, L.K. and G.P.P.; visualization, L.K.; supervision, G.P.P. funding acquisition, G.P.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The UAV data available in the present study were acquired during the implementation of the HARMONIOUS project (<https://www.costharmonious.eu>, accessed on 15 January 2024), which is an EU-funded Cost Action, and the data is available upon request.

Acknowledgments: The authors would like to thank the anonymous reviewers for their valuable feedback, which resulted in improving the initially submitted manuscript. Also, the authors would like to thank the HARMONIOUS project COST Action Salvatore Manfreda, as well as Giuseppe Ciruolo and Antonino Maltese for their local support in the UAV data acquisition used in the present study.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Anand, A.; Pandey, P.C.; Petropoulos, G.P.; Pavlides, A.; Srivastava, P.K.; Sharma, J.K.; Malhi, R.K.M. Use of hyperion for mangrove forest carbon stock assessment in Bhitarkanika Forest Reserve: A contribution towards blue carbon initiative. *Remote Sens.* **2020**, *12*, 597. [CrossRef]
2. Fragou, S.; Kalogeropoulos, K.; Stathopoulos, N.; Louka, P.; Srivastava, P.K.; Karpouzas, S.; Kalivas, D.; Petropoulos, G.P. Quantifying land cover changes in a Mediterranean environment using landsat TM and support vector machines. *Forests* **2020**, *11*, 750. [CrossRef]
3. Srivastava, P.K.; Petropoulos, G.P.; Prasad, R.; Triantakostas, D. Random forests with bagging and genetic algorithms coupled with least trimmed squares regression for soil moisture deficit using SMOS satellite soil moisture. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 507. [CrossRef]
4. Petropoulos, G.P.; Maltese, A.; Carlson, T.N.; Provenzano, G.; Pavlides, A.; Ciruolo, G.; Hristopoulos, D.; Capodici, F.; Chalkias, C.; Dardanelli, G.; et al. Exploring the use of Unmanned Aerial Vehicles (UAVs) with the simplified ‘triangle’ technique for soil water content and evaporative fraction retrievals in a Mediterranean setting. *Int. J. Remote Sens.* **2020**, *42*, 1623–1642. [CrossRef]
5. Achille, C.; Adami, A.; Chiarini, S.; Cremonesi, S.; Fassi, F.; Fregonese, L.; Taffurelli, L. UAV-based photogrammetry and integrated technologies for architectural applications—Methodological strategies for the after-quake survey of vertical structures in Mantua (Italy). *Sensors* **2015**, *15*, 15520–15539. [CrossRef] [PubMed]
6. Plesoiianu, A.I.; Stupariu, M.S.; Sandric, I.; Patru-Stupariu, I.; Dragu, L. Individual tree-crown detection and species classification in very high-resolution remote sensing imagery using a deep learning ensemble model. *Remote Sens.* **2020**, *12*, 2426. [CrossRef]
7. Hao, Z.; Lin, L.; Post, C.J.; Mikhailova, E.A.; Li, M.; Chen, Y.; Yu, K.; Liu, J. Automated tree-crown and height detection in a young forest plantation using mask region-based convolutional neural network (Mask R-CNN). *ISPRS J. Photogramm. Remote Sens.* **2021**, *178*, 112–123. [CrossRef]
8. Deng, Z.; Sun, H.; Zhou, S.; Zhao, J.; Lei, L.; Zou, H. Multi-scale object detection in remote sensing imagery with convolutional neural networks. *ISPRS J. Photogramm. Remote Sens.* **2018**, *145*, 3–22. [CrossRef]
9. Ding, P.; Zhang, Y.; Deng, W.; Jia, P.; Kuijper, A. A light and faster regional convolutional neural network for object detection in optical remote sensing images. *ISPRS J. Photogramm. Remote Sens.* **2018**, *141*, 208–218. [CrossRef]
10. Xuan, N.; Duan, M.; Ding, H.; Hu, B.; Wong, E.K. Attention Mask R-CNN for Ship Detection and Segmentation from Remote Sensing Images. *IEEE Access* **2020**, *8*, 9325–9334. [CrossRef]
11. Li, W.; Fu, H.; Yu, L.; Cracknell, A. Deep learning based oil palm tree detection and counting for high-resolution remote sensing images. *Remote Sens.* **2017**, *9*, 22. [CrossRef]
12. Weinstein, B.G.; Marconi, S.; Bohlman, S.; Zare, A.; White, E. Individual tree-crown detection in RGB Imagery using semi-supervised deep learning neural networks. *Remote Sens.* **2019**, *11*, 1309. [CrossRef]
13. Neupane, B.; Horanont, T.; Hung, N.D. Deep learning based banana plant detection and counting using high-resolution red-green-blue (RGB) images collected from unmanned aerial vehicle (UAV). *PLoS ONE* **2019**, *14*, e0223906. [CrossRef] [PubMed]
14. Nuri Erkin, O.; Gordana, K.; Firat, E.; Dilek, K.M.; Ugur, A. Tree extraction from multi-scale UAV images using Mask R-CNN with FPN. *Remote Sens. Lett.* **2020**, *11*, 847–856. [CrossRef]

15. Safonova, A.; Guirado, E.; Maglinets, Y.; Alcaraz-Segura, D.; Tabik, S. Olive Tree Biovolume from UAV Multi-Resolution Image Segmentation with Mask R-CNN. *Sensors* **2021**, *21*, 1617. [CrossRef] [PubMed]
16. Sandric, I.; Irimia, R.; Petropoulos, G.; Anand, A.; Srivastava, P.; Plesoianu, A.; Faraslis, I.; Stateras, D.; Kalivas, D. Tree's detection & health's assessment from ultrahigh resolution UAV imagery and deep learning. *Geocarto Int.* **2022**, *37*, 10459–10479. [CrossRef]
17. FAIR. Detectron2. 2019. Available online: <https://github.com/facebookresearch/detectron2> (accessed on 15 December 2022).
18. Thenkabail, P.S.; Smith, R.B.; De Pauw, E. Hyperspectral vegetation indices and their relationships with agricultural crop characteristics. *Remote Sens. Environ.* **2000**, *71*, 158–182. [CrossRef]
19. Gitelson, A.A.; Peng, Y.; Huemmrich, K.F. Relationship between fraction of radiation absorbed by photosynthesizing maize and soybean canopies and NDVI from remotely sensed data taken at close range and from MODIS 250 m resolution data. *Remote Sens. Environ.* **2014**, *147*, 108–120. [CrossRef]
20. Iost Filho, F.H.; Heldens, W.B.; Kong, Z.; de Lange, E.S. Drones: Innovative technology for use in precision pest management. *J. Econ. Entomol.* **2020**, *113*, 1–25. [CrossRef]
21. Ciralo, G.; Tauro, F. Chapter 11. Tools and datasets for UAS applications. In *Unmanned Aerial Systems for Monitoring Soil, Vegetation, and Riverine Environments*; Elsevier: Amsterdam, The Netherlands, 2023. [CrossRef]
22. Ciralo, G.; Capodici, F.; Maltese, A.; Ippolito, M.; Provenzano, G.; Manfreda, S. UAS dataset for Crop Water Stress Index computation and Triangle Method applications (rev 1). *Zenodo* **2022**. [CrossRef]
23. Lin, T.L.; Maire, M.; Belongie, S.; Bourdev, L.; Girshick, R.; Hays, J.; Perona, P.; Ramanan, D.; Zitnick, C.L.; Dollar, P. Microsoft COCO: Common Objects in Context. *arXiv* **2014**, arXiv:1405.0312. [CrossRef]
24. Petropoulos, G.; Kalivas, D.; Georgopoulou, I.; Srivastava, P. Urban vegetation cover extraction from hyperspectral imagery and geographic information system spatial analysis techniques: Case of Athens, Greece. *J. Appl. Remote Sens.* **2015**, *9*, 096088. [CrossRef]
25. Huang, C.; Davis, L.S.; Townshend, J.R.G. An assessment of support vector machines for land cover classification. *Int. J. Remote Sens.* **2002**, *23*, 725–749. [CrossRef]
26. Chuvieco, E. *Fundamentals of Satellite Remote Sensing: An Environmental Approach*, 2nd ed.; CRC Press: Boca Raton, FL, USA, 2016. [CrossRef]
27. Tehrany, M.S.; Pradhan, B.; Jebuv, M.N. A comparative assessment between object and pixel-based classification approaches for land use/land cover mapping using SPOT 5 imagery. *Geocarto Int.* **2013**, *29*, 351–369. [CrossRef]
28. Rujoiu-Mare, M.-R.; Olariu, B.; Mihai, B.-A.; Nistor, C.; Săvulescu, I. Land cover classification in Romanian Carpathians and Subcarpathians using multi-date Sentinel-2 remote sensing imagery. *Eur. J. Remote Sens.* **2017**, *50*, 496–508. [CrossRef]
29. Foody, G.M. Status of land cover classification accuracy assessment. *Remote Sens. Environ.* **2002**, *80*, 185. [CrossRef]
30. Kontoes, C.; Poilve, H.; Florsch, G.; Keramitsoglou, I.; Paralikidis, S. A comparative analysis of a fixed thresholding vs. a classification tree approach for operational burn scar detection and mapping. *Int. J. Appl. Earth Obs. Geoinf.* **2009**, *11*, 299–316. [CrossRef]
31. Petropoulos, G.; Kontoes, C.; Keramitsoglou, I. Land cover mapping with emphasis to burnt area delineation using co-orbital ALI and Landsat TM imagery. *Int. J. Appl. Earth Obs. Geoinf.* **2011**, *18*, 344–355. [CrossRef]
32. Rouse, J.W., Jr.; Haas, R.H.; Deering, D.W.; Schell, J.A.; Harlan, J.C. *Monitoring the Vernal Advancement and Retrogradation (Green Wave Effect) of Natural Vegetation*; NASA/GSFC Type III Final Report; NASA: Greenbelt, MD, USA, 1974; 371p.
33. Kaufman, Y.J.; Tanre, D. Atmospherically resistant vegetation index (ARVI) for EOS-MODIS. *IEEE Trans. Geosci. Remote Sensing* **1992**, *30*, 261–270. [CrossRef]
34. Lorenzen, B.; Madsen, J. Feeding by geese on the Filso Farmland, Denmark, and the effect of grazing on yield structure of Spring Barley. *Ecography* **1986**, *9*, 305–311. [CrossRef]
35. Zhang, C.; Valente, J.; Kooistra, L.; Guo, L.; Wang, W. Orchard management with small unmanned aerial vehicles: A survey of sensing and analysis approaches. *Precis. Agric* **2021**, *22*, 2007–2052. [CrossRef]
36. Manfreda, S.; Ben Dor, E. (Eds.) *Unmanned Aerial Systems for Monitoring Soil, Vegetation, and Riverine Environments*; Earth Observation Series; Elsevier: Amsterdam, The Netherlands, 2023; ISBN 9780323852838.
37. Tang, G.; Ni, J.; Zhao, Y.; Gu, Y.; Cao, W. A Survey of Object Detection for UAVs Based on Deep Learning. *Remote Sens.* **2024**, *16*, 149. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.