



Technical Note

# Use of Images Obtained by Remotely Piloted Aircraft and Random Forest for the Detection of Leaf Miner (*Leucoptera coffeella*) in Newly Planted Coffee Trees

Luana Mendes dos Santos <sup>1,\*</sup>, Gabriel Araújo e Silva Ferraz <sup>1</sup>, Nicole Lopes Bento <sup>1</sup>, Diego Bedin Marin <sup>2</sup>, Giuseppe Rossi <sup>3</sup>, Gianluca Bambi <sup>3</sup> and Leonardo Conti <sup>3</sup>

<sup>1</sup> Department of Agricultural Engineering, Federal University of Lavras (UFLA), Lavras 37203-202, MG, Brazil; gabriel.ferraz@ufla.br (G.A.e.S.F.); nicole.bento1@estudante.ufla.br (N.L.B.)

<sup>2</sup> Agriculture Research Company of Minas Gerais (EPAMIG), Viçosa 36571-000, MG, Brazil; diego.marin@ufv.br

<sup>3</sup> Department of Agriculture, Food, Environment and Forestry, University of Florence, 50145 Florence, Italy; giuseppe.rossi@unifi.it (G.R.); gianluca.bambi@unifi.it (G.B.); leonardo.conti@unifi.it (L.C.)

\* Correspondence: luana.goncalves1@ufla.br

**Abstract:** Brazil is the largest producer and exporter of coffee beans in the world. Given this relevance, it is important to monitor the crop to prevent attacks by pests. This study aimed to detect leaf miner (*Leucoptera coffeella*) infestation in a newly planted crop based on vegetation indices (VI) derived from aerial images obtained by a multispectral camera embedded in a remotely piloted aircraft (RPA) using random forest (RF). The study was conducted on the Cafua farm in the municipality of Lavras in southern Minas Gerais. The images were collected using a multispectral camera attached to a remotely piloted aircraft (RPA). Collections were carried out on 30 July 2019 (infested crop) and 16 December 2019 (post chemical control). The RF package in R software was used to classify the infested and healthy plants. The *t* test revealed significant differences in band means between healthy and infested plants, favouring higher means in healthy plants. VI also exhibited significant differences, with EXR being higher in infested plants and GNDVI, GOSAVI, GRRI, MPRI, NDI, NDRE, NDVI and SAVI showing higher averages in healthy plants, indicating distinct spectral responses and light absorption patterns between the two states of the plant. Due to the spectral differences between the classes, it was possible to classify the infested and healthy plants, and the RF algorithm performed very well.

**Keywords:** unmanned aircraft systems; vegetation index; *Coffea arabica*; multispectral; artificial intelligence



**Citation:** dos Santos, L.M.; Ferraz, G.A.e.S.; Bento, N.L.; Marin, D.B.; Rossi, G.; Bambi, G.; Conti, L. Use of Images Obtained by Remotely Piloted Aircraft and Random Forest for the Detection of Leaf Miner (*Leucoptera coffeella*) in Newly Planted Coffee Trees. *Remote Sens.* **2024**, *16*, 728. <https://doi.org/10.3390/rs16040728>

Academic Editor: Xanthoula Eirini Pantazi

Received: 1 December 2023

Revised: 3 February 2024

Accepted: 13 February 2024

Published: 19 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

Brazil is widely recognized as the largest producer of Arabica coffee (*Coffea arabica*) worldwide. The country ranks first in the global production of arabica beans, accounting for 46.4% to the total production. The estimated total production of grain for the 2023/2024 crop worldwide is 96.33 million bags (60 kg) [1]. In addition, the state of Minas Gerais has a significant share of national production, contributing 27.8 million bags of processed grain, representing 50.8% of the country's total production [2]. The significant contribution by this state is due to its ideal climatic and geographical conditions for the cultivation of this coffee variety, including a combination of elevation, temperature, rainfall and suitable soils [3].

However, coffee cultivation still faces challenges, including predation by the leaf miner (*Leucoptera coffeella*), a key crop pest that affects coffee production and quality [4,5] in hotter and drier seasons particularly. According to Franklin et al. [6], the leaf miner moth is a small silvery-white moth with a diurnal mating habit. The development of the eggs varies from 5 to 21 days according to temperature. When the eggs hatch, the larvae penetrate the epidermis of the leaf to feed and to form “mines” which have a light brown colour and a rounded shape, eventually reaching a darker colour due to the accumulation of

excretions [6]. Attacks by this pest on coffee plants result in reduced productivity, causing losses of hundreds of millions of dollars every year according to Rosado et al. [7].

To address this challenge, emerging technologies such as remotely piloted aircraft (RPA) have been used to identify symptoms of pest attacks, e.g., in preliminary studies conducted by Santos et al. [8]. These studies aimed to mitigate challenges and improve the production of this crop in Brazil. RPA with multispectral sensors can fly over coffee plantations and capture high-resolution (in cm) images at different wavelengths, as highlighted in recent studies by Bento et al. [9], Santana et al. [10] and Santos et al. [11] using such technologies for various applications in coffee farming. Such images can be processed to generate vegetation indices (VI), can provide information on the phytosanitary conditions of crops [12], be used to identify frost damage [13] and be used to identify diseases such as rust in coffee trees [14].

With the analysis of VI, it is possible to provide early detection of areas affected by leaf miner because the leaves damaged by the pest have spectral characteristics that differ from those of healthy leaves, which allows the identification and mapping of infested areas. Studies conducted by Santos et al. [8] found that the mean NDVI values for healthy plants were higher than those for plants infested with the leaf miner. Such information can lead farmers to make more precise and targeted decisions, contributing to the application of specific pesticides, reducing indiscriminate pesticide use, minimizing environmental impacts and reducing costs associated with coffee crops, thus making coffee cultivation more sustainable.

To monitor the leaf miner currently, two sampling plans can be performed manually: common sampling and sequential sampling. In the study by Souza et al. [15], common sampling was performed: 10 samples were collected consisting of the fourth pair of leaves from the upper third of the branches, with one pair of leaves collected from each quadrant of the coffee plant. In this common sampling scheme, the number of leaves with mines is examined, obtaining the percentage of mined leaves. The threshold for level control (LC) is 40% of leaves mined in the winter and 20% of leaves mined in the summer. According to Berzoti and Aquino [16], sequential sampling has an advantage over common sampling, in which a fixed number of samples is collected, in addition to having a gain of considerable time throughout the cycle of a crop. The path for sampling the leaves in the affected area should form a spiral.

Both manual identification methods require professionals on site to detect the leaf miner pest. When identifying an outbreak, they write down the information for analysis, calculate the LC and then the conventional treatment process begins with the application of pesticides throughout the area. To minimize and optimize the identification of leaf miners in coffee plantations, remote detection methods are recommended.

A study of the application of Sentinel-2 satellite images to identify coffee leaf miner infestation levels in coffee plantations was conducted by Vilela et al. [17]. The study considered bands of the Sentinel-2 satellite with a spatial resolution of 10 m and a revisit period of approximately five days. Another study by Marin et al. [18] aimed to identify and map coffee crop stress caused by biotic and abiotic variables, including the leaf miner, using VI derived from multispectral Landsat-5 Thematic Mapper (TM) images with spatial resolution of 30 m. Despite having obtained good results, the methodologies proposed by both studies have some limitations such as the low spatial and temporal resolution when compared to the images that RPA can offer. RPAs have high temporal resolution compared to satellites, and therefore obtain images more frequently. Finally, with RPA it is possible to obtain constant data as long as there are favourable weather conditions (i.e., no rain).

Relying upon traditional manual sampling and monitoring of satellite images, the identification of extensive areas of infestation by the leaf miner in a timely manner can become difficult and require processing of large amounts of samples. Thus, the use of RPAs coupled with image processing and VI analysis allows the continuous monitoring of coffee plantations, enabling early detection of phytosanitary problems, favouring more

efficient planning of actions and increased effectiveness in decision making related to crop management.

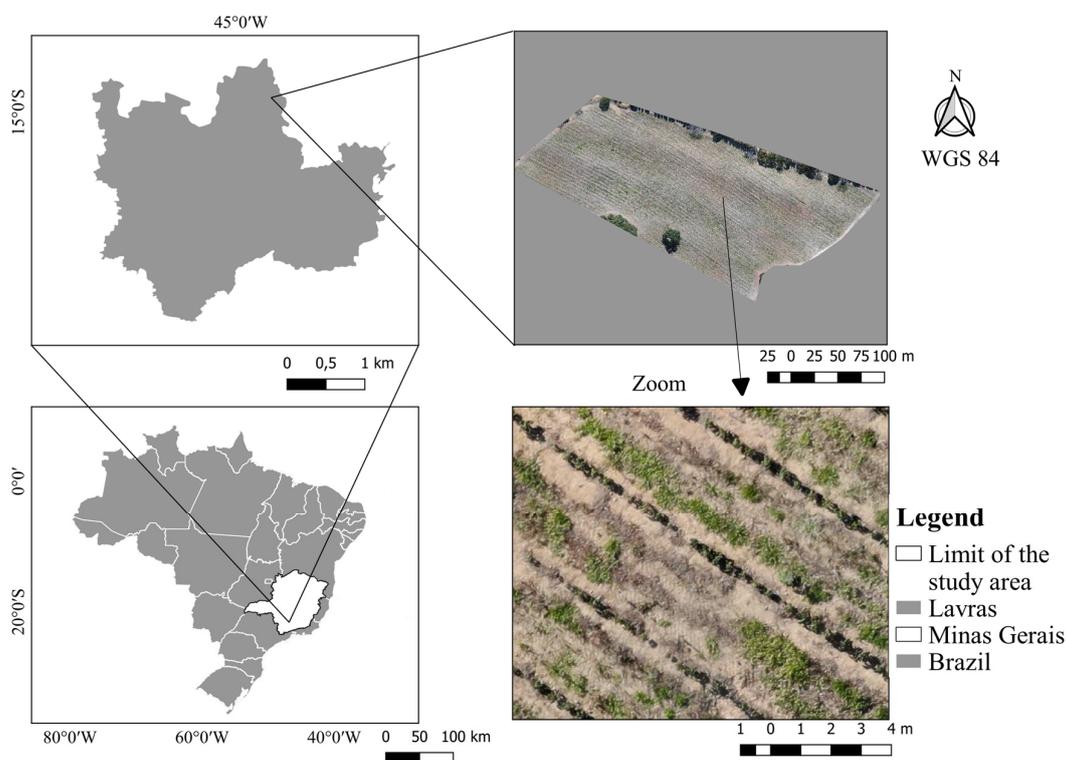
In addition, the advent of agricultural solutions based on artificial intelligence (AI), artificial neural networks, machine learning, deep learning, remote sensing, digital image processing and robotics [19] has improved production in this sector. Neural networks in particular have a strong learning ability and are highly accurate in the identification of leaf pests and diseases [20].

Thus, this study aimed to detect leaf miner infestation in a newly planted crop using VI analysis and the application of the random forest (RF) algorithm in aerial images obtained by a multispectral camera embedded on a RPA.

## 2. Materials and Methods

### 2.1. Study Area

The study was conducted on the Cafua commercial farm, located in the municipality of Lavras in southern Minas Gerais in an area of 0.2 ha of coffee (*Coffea arabica* L.) of the cultivar Catuaí Vermelho IAC 99, planted in November 2018, with a spacing of 3.5 m between rows and 0.5 m between plants. The geographical coordinates are 21°09'50''S and 44°59'35''W, with an average elevation of 934 m (Figure 1).



**Figure 1.** Location of the study area.

The climate in the region, according to the Koppen classification, is Cwa (subtropical climate, characterized by a dry season in the winter and a rainy season in the summer) with an average temperature of 20 °C and average annual rainfall of 1153 mm [21].

### 2.2. Aerial Data Collection

Visual sampling was performed in the field to classify the coffee crop as infested or healthy. The first aerial data collection was performed on 30 July 2019 in an area of crop classified as infested. After this collection, the area was subjected to chemical control through the application of 0.69 kg.ha<sup>-1</sup> Actara 250 WG® (Syngenta, São Paulo, Brazil). A second visual sampling was performed after chemical control. The crop was classified as

healthy, and a second collection of aerial data was performed on 16 December 2019. Two aerial data collections were then carried out on different dates to remotely analyze and identify changes in the health status of the plants.

A 3DR solo model (3D Robotics, Berkeley, CA, USA) RPA was used. The aircraft was remotely operated at a height of 50 m from the ground and integrated with a global positioning satellite system (GNSS) (Figure 2a). The images were captured by a Parrot Sequoia™ multispectral camera (MicaSense, Seattle, WA, USA) onboard the RPA. The camera has four spectral sensors: green (G) (530–570 nm), red (R) (640–680 nm), red edge (EDGE) (730–740 nm) and near infrared (NIR) (770–810 nm), with a resolution of 1.2 megapixels (1280 × 960). It is self-calibrated through a solar sensor integrated into the camera and has a RGB sensor with a resolution of 16 MP (Figure 2b).



**Figure 2.** Equipment used for acquisition of images: (a) RPA used for field survey; (b) camera (image and irradiance sensors).

### 2.3. Image Processing

The processing of the images obtained with the RPA was performed with Pix4Dmapper Pro version 4.8.0 software, and orthoimages with reflectance values in the 4 bands, i.e., green, red, red edge and near infrared, were generated as described by Santos et al. [22].

The postprocessing of the images was performed using Quantum GIS (QGIS) software version 3.10. Arithmetic operations were performed for the R, NIR, EDGE and G bands using the raster calculator tool, and the following VI were calculated: GNDVI, GO-SAVI, NDVI, GRRRI, MPRI, NDI, NDRE, SAVI and EXR (Table 1). Such VI were selected according to the relevance in the literature for studies of diseases in coffee trees and VI used for monitoring coffee leaf miners. A total of 10,987 plants were identified and a 15 cm buffer was generated on top of each plant to obtain spectral information.

**Table 1.** VI applied to images obtained using the bands of the multispectral camera.

Vegetation Indices	Equation	Source
Excessive red (EXR)	$EXR = 1.44 \times R - G$	[23]
Green normalized difference vegetation index (GNDVI)	$GNDVI = \frac{NIR - G}{NIR + G}$	[24]
Green optimal soil adjusted vegetation index (GOSAVI)	$GOSAVI = (1 + 0.16) \frac{NIR - G}{NIR + G + 0.16}$	[25]
Green–red ratio index (GRRRI)	$GRRRI = \frac{G}{R}$	[26]
Modified photochemical reflectance index (MPRI)	$MPRI = \frac{G - R}{G + R}$	[27]
Normalized different index (NDI)	$NDI = \frac{G - R}{G + R + 0.01}$	[28]
Normalized difference red edge (NDRE)	$NDRE = \frac{NIR - EDGE}{NIR + EDGE}$	[29]
Normalized difference vegetation index (NDVI)	$NDVI = \frac{NIR - R}{NIR + R}$	[30]
Soil-adjusted difference vegetation index (SAVI)	$SAVI = \frac{(1+L)(NIR - R)}{NIR + R + L}$	[31]

L = 0.5 (adjustment factor for the SAVI index).

### 2.4. Classification Method

A descriptive statistical analysis of the reflectance values of the bands and the VI values was carried out. The Shapiro–Wilk normality test ( $W$ ;  $p \leq 0.05$ ), homogeneity of variances test and  $t$  test for independent samples ( $p \leq 0.05$ ) was applied in order to verify significant

differences between the bands and VI obtained in the infested leaves and healthy leaves. Analysis was performed using R software version 4.3.2 (R Core Team, Vienna, Austria).

The training and validation stages of the classification algorithm consisted of exploratory analysis of the input data, classification by the RF algorithm and validation and verification of the performance of the classifier by performance metrics.

In the classification process, two classes of interest were established: coffee plants with leaf miner infestation and coffee plants without leaf miner infestation (after chemical control), each with 10,897 samples. The data used for the classification consisted of the individual spectral bands and the previously mentioned vegetation indices.

R version 4.3.2 software and the “randomForest” package were used for the RF classification process [32]. Initially, a random division was created between training and validation samples, with proportions of 70% and 30%, respectively. For the RF classification, the number of decision trees (Ntree) was set to 100 and the number of variables to be tested (Mtry) corresponded to the square root of the number of input variables [33]. Gini index analysis was also applied to describe the importance of each input variable in the algorithm to the classification process.

Validation was performed by comparing the validation sample to the reference data. The evaluation of the performance and quality of the classification was performed using information from a confusion matrix and metrics of global accuracy, sensitivity and specificity and the F1 score [34] in addition to the external metric, i.e., the area under the ROC curve (AUC) based on a confusion matrix.

The application of algorithms combined with classification and spectral characterization tools allows for the segmentation of pixels according to their reflectance values [35,36]. Consequently, these techniques can contribute to the identification of pests and foliar diseases. The RF algorithm uses the ensemble paradigm, in which multiple models are combined to classify the input. Therefore, a combination of several decision trees is used which are trained with the input data, obtained via a simple random sample and the attributes to be considered are randomly chosen [37]. The positive characteristics for its use include the fact that this algorithm is affected by outliers and noisy data, it is non-parametric, it supports data with different statistical distributions, it has a high capacity for processing data on a large scale and from different sources and has greater accuracy in terms of classification when compared to other classifier algorithms [38,39].

### 3. Results

#### 3.1. Exploratory Data Analysis

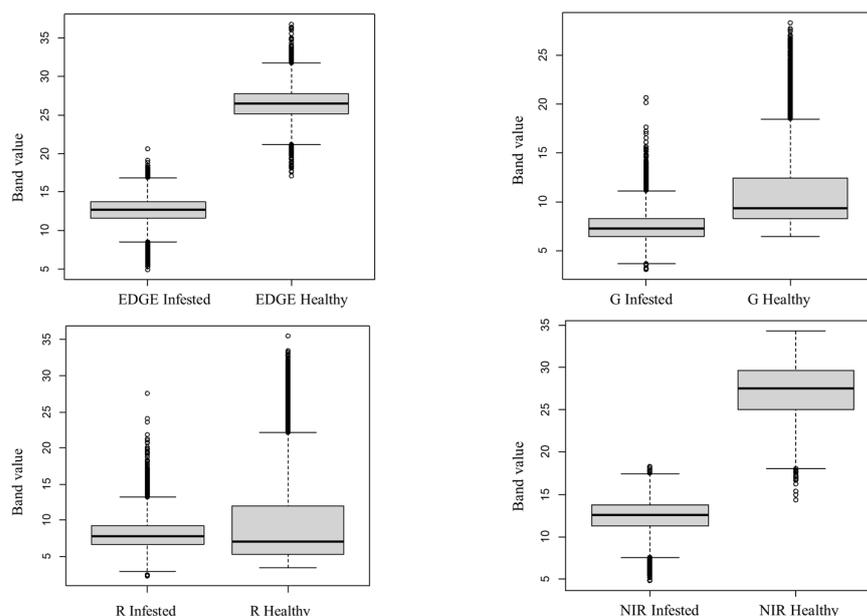
Table 2 shows the results of the exploratory analysis of the values of the bands used to calculate the VI. With the statistical results, it was possible to analyze the spectral characteristics of the plants relative to the state of the plants (healthy or infested with leaf miners).

**Table 2.** Summary of descriptive statistical values for the bands used to detect healthy and infested plants.

Band	State of the Plants	Average	SD	Mean	Min	Max
EDGE	Infested	12.644	1.709	12.707	4.902	20.627
	Healthy	26.498	2.110	26.506	17.112	36.794
G	Infested	7.563	1.537	7.318	3.135	20.634
	Healthy	11.083	4.005	9.376	6.480	28.293
NIR	Infested	12.509	1.879	12.598	4.869	18.348
	Healthy	17.209	3.080	27.507	14.377	34.277
R	Infested	8.149	2.225	7.778	2.315	27.524
	Healthy	9.711	6.287	7.033	3.454	35.516

SD—standard deviation, Min—minimum VI value, Max—maximum VI value.

The  $t$  test was used to compare the averages of the EDGE, G, NIR and R bands in the two states of plants studied. The four bands did not present normality and did not present homogeneity of variance. Given this result, a specific  $t$  test was used for independent samples that present non-homogeneous variances. The  $p$ -value obtained for the four bands was less than 0.05, indicating a statistically significant difference in the band means (EDGE, G, R and NIR) in the two states of the plants. The results obtained suggest that the average values of the EDGE, G, NIR and R bands is higher in healthy plants than in infested plants, which indicates a more intense spectral response in these bands in healthy plants and a greater absorption of light in these bands by healthy plants (Table 2 and Figure 3).



**Figure 3.** Box plots for band values in the states of infested and healthy plants with statistically significant differences by the  $t$  test ( $p \leq 0.05$ ).

Table 3 shows the exploratory analysis of the values of the VI obtained in the study. The statistical results provide an overview of the characteristics of the different VI in healthy plants and plants infested with leaf miner. These results indicate differences in the means, standard deviations, medians and ranges of values between the two conditions.

**Table 3.** Summary of descriptive statistical values for VI in healthy and infested plants.

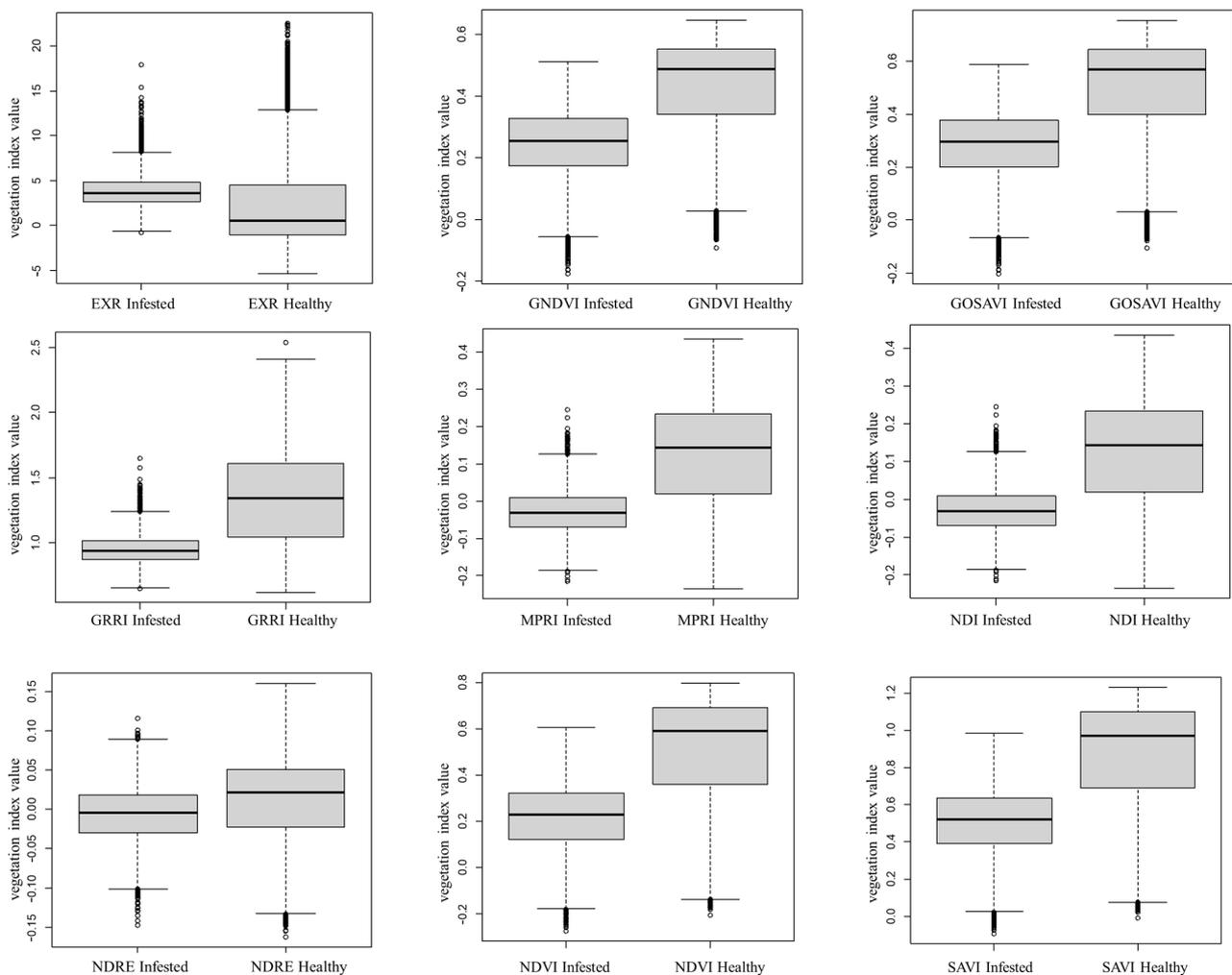
Band	State of the Plants	Average	SD	Mean	Min	Max
EXR	Infested	3.845	1.797	3.591	−0.771	17.899
	Healthy	2.512	5.037	0.537	−5.342	22.528
GNDVI	Infested	0.245	0.111	0.257	−0.175	0.512
	Healthy	0.429	0.162	0.488	−0.090	0.647
GOSAVI	Infested	0.283	0.128	0.295	−0.202	0.589
	Healthy	0.499	0.188	0.569	−0.105	0.753
GRR1	Infested	0.951	0.111	0.939	0.646	1.650
	Healthy	1.336	0.343	1.342	0.621	2.537
MPRI	Infested	−0.028	0.057	−0.031	−0.215	0.245
	Healthy	0.1223	0.1131	0.143	−0.234	0.4335
NDI	Infested	−0.028	0.057	−0.031	−0.214	0.244
	Healthy	0.123	0.131	0.143	−0.234	0.434
NDRE	Infested	−0.006	0.034	−0.005	−0.147	0.116
	Healthy	0.011	0.054	0.022	−0.162	0.160

Table 3. Cont.

Band	State of the Plants	Average	SD	Mean	Min	Max
NDVI	Infested	0.217	0.145	0.228	−0.275	0.608
	Healthy	0.501	0.242	0.589	−0.206	0.798
SAVI	Infested	0.508	0.177	0.522	−0.092	0.984
	Healthy	0.864	0.297	0.973	−0.007	1.231

SD—standard deviation, Min—minimum VI value, Max—maximum VI value.

The *t* test was used to compare the average VI in the two states of the plants studied. The indices did not present normality and did not present homogeneity of variance. Given this result, a specific *t* test was used for independent samples that present non-homogeneous variances. The *p*-value obtained for the indices was less than 0.05, indicating a statistically significant difference in their averages in the two states of the plants. Accounting for the comparison of means observed in Table 3 and Figure 4, the mean value of EXR was higher for infested plants than for healthy plants. For the VI GNDVI, GOSAVI, GRRI, MPRI, NDI, NDRE, NDVI and SAVI, the means were higher for healthy plants than for infested plants.



**Figure 4.** Box plot for VI values for infested and healthy states of the plants with statistically significant differences by the *t* test ( $p \leq 0.05$ ).

### 3.2. Overall Classifier Accuracy

The general accuracy, sensitivity, specificity, F1 score and area under the ROC curve for the classifications performed by the RF algorithm are shown in Table 4. In general, the

values of the parameters analyzed demonstrate a satisfactory performance of the classifier in the differentiation of classes. Favourable results were obtained with the algorithm.

**Table 4.** Performance parameters for the RF classification algorithm.

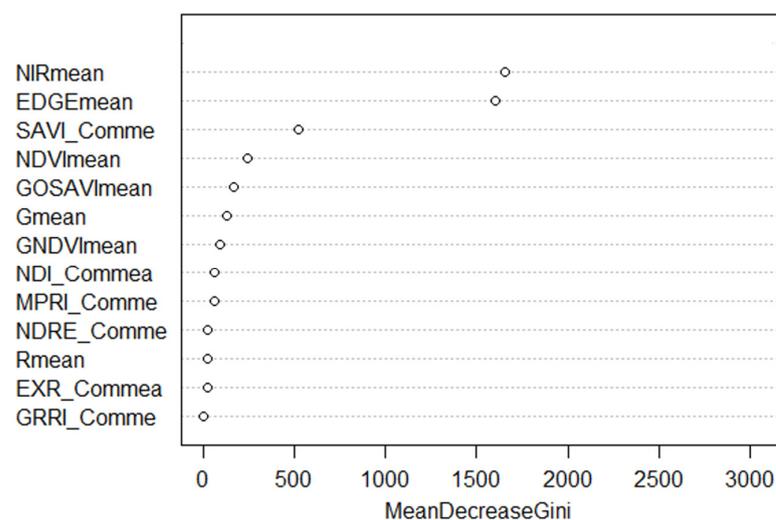
Algorithm	Overall Accuracy	Sensitivity	Specificity	F1	AUC
RF	0.9997	0.9997	0.997	0.9996	0.9994

Confusion matrix analyses were performed to validate the results of the classification algorithm. The previously described satisfactory values of the performance parameters illustrate that there were few errors in the classification of study classes, as evidenced by the confusion matrices created for data from the classification algorithm (Table 5).

**Table 5.** Confusion matrix of plants with and without leaf miner infestation classified using RF algorithm.

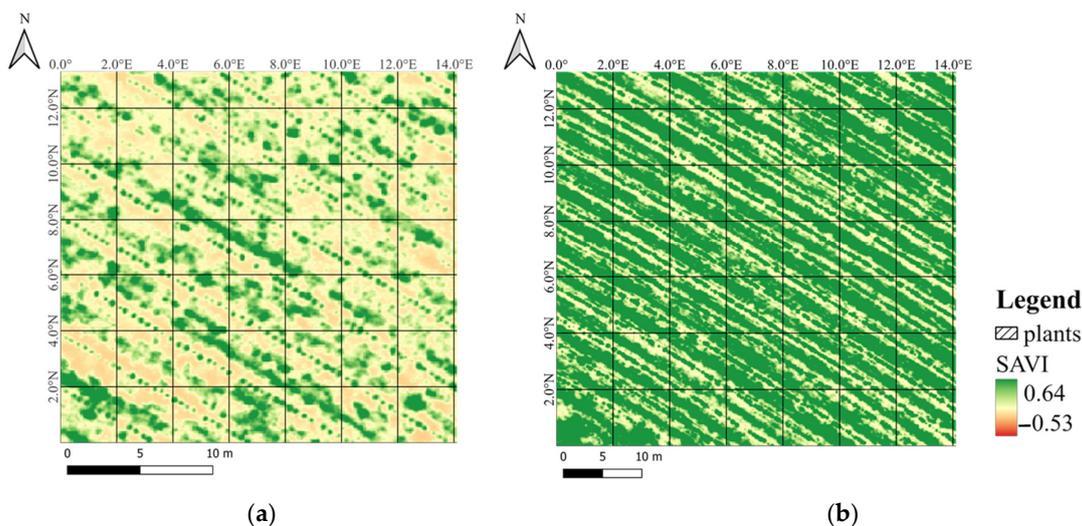
Classes	Reference			Total
		With	Without	
Prediction	With	3268	1	3269
	Without	1	3268	3269
	Total	3269	3269	6538

Using the RF classifier and the Gini index, it was possible to describe the importance of each variable for the classification process as shown in Figure 5. It is noted that for this study, the NIR spectral band and the EDGE band were more important in the classification, whereas the R band was the least important. The most important vegetation index in the classification was SAVI, whereas the GRRR vegetation index was the least important in the proposed classification.



**Figure 5.** Variables of importance according to the Gini index.

According to the classification, SAVI was the most important vegetation index. Thus, the spatial distribution of VI in the study area can be observed in Figure 6. As observed in Table 3, SAVI had an average value of 0.508 in infested plants, shown by the yellowish and red colour in Figure 6, and an average value of 0.864 in healthy plants, shown by the green colour in Figure 6.



**Figure 6.** Spatial distribution of the vegetation index SAVI in the study area. (a) Area with leaf miner infestation; (b) area without leaf miner infestation.

#### 4. Discussion

In this study, it was observed that healthy plants had a higher mean reflectance of the EDGE, G, NIR and R bands than infested plants. This result is consistent with the studies conducted by Vilela et al. [17] in which the authors observed that the spectral response in the NIR region of a coffee crop with a higher percentage of leaf miner infestation was lower than that of a crop with a lower percentage of infestation. That is, the greater the reflectance, especially in the NIR region, the more this spectral behaviour is related to vegetative vigour. When infestation occurs, the leaf miner larvae feed on the mesophyll of the coffee leaves, creating mines. These mines cause leaf necrosis, contributing to a lower photosynthetic rate [4] and, therefore, lower vigour and lower spectral response in the NIR range. Thus, the infestations of pests that affect the internal structures of the leaves can be evaluated using the NIR band, confirming the results of this study and some previous studies [17,40] and highlighting the importance of spectral bands in RF classifications, as demonstrated by the Gini index (Figure 5).

The EDGE spectral band also showed greater importance in the classification process, as shown in Figure 5. This band is highly sensitive to the concentration of chlorophyll, allowing high absorption of the incident energy that penetrates the surface layer while reducing the high level of reflectance observed in the innermost part of the leaf structure [24].

In addition, the descriptive analysis of vegetation index values can help to identify patterns and differences between healthy and infested plants, which can be useful for monitoring and managing coffee plantations affected by the leaf miner. As the results showed a higher average for healthy plants than for infested plants for the VI GNDVI, GOSAVI, GRRI, MPRI, NDI, NDRE, NDVI and SAVI, it is suggested that the values of the indices tend to be higher in healthy plants, indicating greater vegetative vigour. Similar results were found in the studies conducted by Santos et al. [8], in which the authors performed studies of VI in different sites infested by leaf miner. In the studies, the authors found higher values of VI for healthy plants than for infested plants on average. However, the opposite results were found for the vegetation index EXR when comparing the average values of infested and healthy plants. The spectral sensitivity of EXR demonstrates the difference between red and green bands. In addition, it is noteworthy that the index is intended to segment soil leaf regions [41]. This behaviour is attributed to the higher percentage of exposed soil in areas with infested plants (plant age: eight months) than in areas with healthy plants (plant age: one year and one month).

Regarding the VI that were found to have the greatest importance in the classification process, SAVI stands out. This index makes use of the constant L, known as the adjustment

factor of the SAVI index, the value of which varies from 0.25 to 1 depending on the soil coverage. According to Huete [31], an L value of 0.25 is recommended for areas with dense vegetation, an L value of 0.5 is recommended for areas of intermediate density and an L value equal to 1 indicates areas with low vegetation density. Importantly, when the SAVI value is equal to 0, its results are similar to those of NDVI, which was found in the second-best classification. This vegetation index considers the soil influences visible in the images during analysis and was suitable because the plants analyzed were young and the interrow space had no vegetation cover, minimizing the interference of the exposed soil in this analysis. Thus, the fact that this index eliminates the effects of the soil gives us a more realistic view of the status in the study area [42].

The results obtained by the RF algorithm indicate that the performance of this approach was satisfactory for the differentiation of study classes, since high values were observed for the performance metrics evaluated such as the overall accuracy, sensitivity, specificity, F1 score and area under the ROC curve (Table 4). The precision values described in this study indicate the frequency with which the classifier correctly estimates the classes. Values close to 100% were observed, highlighting the excellent fit of the model to the proposed classes based on the satisfactory values of the performance metrics, thus showing few errors between the thematic classes identified in the confusion matrix (Table 5). Therefore, based on the results presented, RF is considered a suitable classification algorithm for this study.

It is also noteworthy that the NIR band had the highest importance weight in the classifier according to the Gini index (Figure 5). However, the NIR band alone was not sufficient to differentiate between the study classes, and using other spectral bands resulted in better classification results. In general, satisfactory and accurate classifications of plants with and without the presence of leaf miner were obtained by using images with a high spatial resolution combined with spectral bands and VI that were able to capture the responses of the plants to stress conditions imposed by pest infestation. Notably, in a study by Bento et al. [9], the authors used classification algorithms to differentiate weeds from coffee plants and obtained satisfactory parameter values for the RF algorithm based on the analyzed performance metrics.

The application of a classification algorithm such as RF, in the identification of the leaf miner in Brazilian coffee cultivation, has the potential to improve the early detection and management of this pest. However, there are challenges to be overcome, including issues related to data quality, generalizability of models and interpretability. With continuous advances in technology and research, these approaches may play a key role in the future of precision coffee farming in Brazil. It is recommended that future studies consider using data acquired by cameras with high spectral resolutions to improve the quality of the training images. In addition, the application of supervised learning and subsequent transfer learning techniques may allow training of supervised models and eventually pretrained models on large image datasets, favouring adjustments for the detection of specific diseases to save time and resources. Another recommendation is for future studies to apply other classifiers and algorithms to obtain better fits. The integration of these models into real-time monitoring systems can support assertive and proactive decision making. Learning models with good interpretability are suggested to help farmers better understand the decisions made by the algorithms.

Finally, we recommend experiments that follow various stages of plant development. Due to the fact that the leaf miner develops quickly, its damage can seriously compromise productivity. Although there is already a coffee variety resistant to the pest [43], there are still no significant commercial plantations. Knowing how to identify the leaf miner in coffee makes it easier to apply the correct insect control techniques and, consequently, ensures that the crop does not suffer drastic losses. However, producers still do not use remote detection techniques or precision agriculture techniques for early identification. What happens in the field is the initial identification and independent of the LC, and the application of pesticides occurs. This approach entails high costs and unfavourable environmental implications.

Therefore, the relevance of studies like this one is justified to provide subsidies to producers on the topic in question.

Furthermore, as recommendations for future work, hyperspectral remote sensing of solar-induced chlorophyll fluorescence (SIF) studies stand out for monitoring vegetation health. Such studies can be significant and replicable for detecting pests in coffee farming, as the caterpillars begin to eat the leaves from the upper part, thus forming a kind of mine for shelter. These lesions have a brown color and cause necrosis in the leaves resulting in defoliation, which contributes to a reduction in the plants' photosynthetic rate and, consequently, a drop in productivity. SIF, according to the available literature [44,45], can be an effective tool for detecting photosynthesis and identifying variations in photosynthesis caused by the miner pest.

## 5. Conclusions

In the present research, leaf miner infestation detection has been achieved in newly planted coffee seedlings through the effective combination of VIs obtained by a multispectral camera embedded in an RPA and aerial images.

The *t* test revealed significant differences in EDGE, G, NIR and R bands between healthy and infested plants. The highest band averages were in healthy plants. The VIs also showed significant differences through the *t* test between healthy and infested plants, with higher mean EXR in infested plants and higher means for GNDVI, GOSAVI, GRRI, MPRI, NDI, NDRE, NDVI and SAVI in healthy plants.

Due to spectral differences between the classes, it was possible to use the RF classification algorithm. The algorithm demonstrated a satisfactory performance in classifying healthy and infested plants. Parameters such as accuracy, sensitivity, specificity, F1 score and area under the ROC curve indicated favourable results. The RF classifier and the Gini index identified the NIR and EDGE bands as the most important for classification, while the R band was the least important. Among the vegetation indices, SAVI was the most important and GRRI the least important.

Such results contribute to the identification of patterns and differences between healthy and infested plants, marking an unprecedented advancement in coffee farming. This achievement not only improves the effectiveness and precision of agricultural practices but also offers an efficient means of monitoring for producers, thus driving the advancement of coffee farming through the adoption of advanced technologies. Furthermore, the lack of widespread research on the use of advanced technologies in coffee farming highlights the distance of these resources from producers. Thus, the integration of these technologies stands out for its efficiency in identifying pests, providing an innovative approach for optimizing the management of coffee crops.

**Author Contributions:** Conceptualization, L.M.d.S.; methodology, L.M.d.S., N.L.B. and D.B.M.; software, L.M.d.S., N.L.B. and D.B.M.; validation, L.M.d.S., N.L.B. and D.B.M.; formal analysis L.M.d.S., N.L.B. and D.B.M.; investigation, L.M.d.S., N.L.B. and D.B.M.; resources, G.A.e.S.F.; data curation, L.M.d.S., N.L.B. and D.B.M.; writing—original draft preparation, L.M.d.S. and N.L.B.; writing—review and editing, L.M.d.S., N.L.B., D.B.M., G.A.e.S.F., G.R., G.B. and L.C.; visualization, G.R., G.B. and L.C.; supervision, G.A.e.S.F.; project administration, G.A.e.S.F.; funding acquisition, G.R., G.B. and L.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Embrapa Café—Coffee Research Consortium (project 10.18.20.041.00.00), the National Council for Scientific and Technological Development (CNPq) (project 30595 3/2020-6), and the Minas Gerais Research Support Foundation (FAPEMIG) (projects PPE-00118-22 and BPD-00040-22).

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author.

**Acknowledgments:** We would like to thank the Federal University of Lavras (UFLA) and the postgraduate program in Agricultural Engineering (PPGEA) for supporting this study, Embrapa Café—Coffee Research Consortium for funding the research (project 10.18.20.041.00.00), to the National

Council for Scientific and Technological Development (CNPq) (project 305953/2020-6), to the Minas Gerais Research Support Foundation (FAPEMIG) (projects PPE-00118-22 and BPD-00040-22), and to the Coordination of Improvement of Higher Education Personnel (CAPES) for financial support, and University of Firenze (UniFI) for the support and development of the research.

**Conflicts of Interest:** Author Diego Bedin Marin was employed by the company Agriculture Research Company of Minas Gerais (EPAMIG). The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

1. United States Department of Agriculture (USDA). Available online: <http://usda.mannlib.cornell.edu/MannUsda/viewDocumentInfo.do?documentID=1801> (accessed on 20 June 2023).
2. Companhia Nacional de Abastecimento (CONAB). Available online: <https://www.conab.gov.br/info-agro/safras/cafe/boletim-da-safra-de-cafe> (accessed on 20 June 2023).
3. Mesquita, C.M.; Melo, E.M.; Rezende, J.E.; Carvalho, J.S.; Fabri, M.A., Jr.; Moraes, N.C.; Araujo, W.G. *Manual do Café: Implantação de Cafezais Coffea arábica L.*; EMATER-MG: Belo Horizonte, Brazil, 2016; p. 50.
4. Dantas, J.; Motta, I.O.; Vidal, L.A.; Nascimento, E.F.M.B.; Bílio, J.; Pupé, J.M.; Veiga, A.; Carvalho, C.; Lopes, R.B.; Rocha, T.L. A comprehensive review of the coffee leaf miner *Leucoptera coffeella* (Lepidoptera: Lyonetiidae)—A major pest for the coffee crop in Brazil and others Neotropical countries. *Insects* **2021**, *12*, 1130. [[CrossRef](#)]
5. Matiello, J.B.; Santinato, R.; Almeida, S.R.; Garcia, A.W.R. *Cultura de Café No Brasil: Manual de Recomendações*, 2nd ed.; Futurama Editora: São Paulo, Brasil, 2020; p. 585.
6. Franklin, A.M.; Martins, F.K.; Costa, E.S.; Nunes, J.F. Community of parasitoids associated with the 'bicho-mineiro' (*Leucoptera coffeella* Guérin-Mêneville) in coffee leaves (*Coffea arabica* L.) grown in the Southwest of Minas Gerais. *Ciência et PRAXIS* **2017**, *10*, 25–30.
7. Rosado, M.C.; de Araújo, G.J.; Pallini, A.; Venzon, M. A consorciação de culturas de cobertura aumenta o controle biológico em lavouras de café. *Biol. Controle* **2021**, *160*, 104675.
8. Santos, L.M.D.; Ferraz, G.A.E.S.; Marin, D.B.; Carvalho, M.A.D.F.; Dias, J.E.L.; Alecrim, A.D.O.; Silva, M.D.L.O.E. Vegetation Indices Applied to Suborbital Multispectral Images of Healthy Coffee and Coffee Infested with Coffee Leaf Miner. *AgriEngineering* **2022**, *4*, 311–319. [[CrossRef](#)]
9. Bento, N.L.; Ferraz, G.A.E.S.; Amorim, J.D.S.; Santana, L.S.; Barata, R.A.P.; Soares, D.V.; Ferraz, P.F.P. Weed Detection and Mapping of a Coffee Farm by a Remotely Piloted Aircraft System. *Agronomy* **2023**, *13*, 830. [[CrossRef](#)]
10. Santana, L.S.; Ferraz, G.A.E.S.; Santos, G.H.R.D.; Bento, N.L.; Faria, R.D.O. Identification and Counting of Coffee Trees Based on Convolutional Neural Network Applied to RGB Images Obtained by RPA. *Sustainability* **2023**, *15*, 820. [[CrossRef](#)]
11. Santos, L.M.D.; Ferraz, G.A.E.S.; Carvalho, M.A.D.F.; Teodoro, S.A.; Campos, A.A.V.; Menicucci Neto, P. Use of RPA Images in the Mapping of the Chlorophyll Index of Coffee Plants. *Sustainability* **2022**, *14*, 13118. [[CrossRef](#)]
12. Martins, G.D.; Galo, M.D.L.B.T.; Vieira, B.S. Detecting and mapping root-knot nematode infection in coffee crop using remote sensing measurements. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 5395–5403. [[CrossRef](#)]
13. Marin, D.B.; Ferraz, G.A.E.S.; Schwerz, F.; Barata, R.A.P.; Faria, R.O.; Dias, J.E.L. Unmanned aerial vehicle to evaluate frost damage in coffee plants. *Precis. Agric.* **2021**, *22*, 1845–1860. [[CrossRef](#)]
14. Marin, D.B.; Santana, L.S.; Barbosa, B.D.S.; Barata, R.A.P.; Osco, L.P.; Ramos, A.P.M.; Guimarães, P.H.S. Detecting coffee leaf rust with UAV-based vegetation indices and decision tree machine learning models. *Comput. Electron. Agric.* **2021**, *190*, 106476. [[CrossRef](#)]
15. Souza, T.P.; Castellani, M.A.; Lemos, R.N.S.D.; Maluf, R.P.; Moreira, A.A.; Silva, B.S.; Ribeiro, E.B. Seasonal occurrence, predation and parasitism of *leucoptera coffeella* (Guérin-Mêneville, 1842) (Lepidoptera: Lyonetiidae) in coffee plants associated with grevillea trees. *Coffee Sci.* **2014**, *9*, 34–50.
16. Berzoti, E.; Aquino, L.H. Sequential sampling plan to evaluate the infestation of the coffee leaf miner (Lepidoptera: Lyonetiidae) in southern Minas Gerais state, Brazil. *Pesqui. Agropecuária Bras.* **1994**, *29*, 695–705.
17. Vilela, E.F.; Ferreira, W.P.M.; Castro, G.D.M.D.; Faria, A.L.R.D.; Leite, D.H.; Lima, I.A.; Venzon, M. New Spectral Index and Machine Learning Models for Detecting Coffee Leaf Miner Infestation Using Sentinel-2 Multispectral Imagery. *Agriculture* **2023**, *13*, 388. [[CrossRef](#)]
18. Marin, D.B.; Alves, M.D.C.; Pozza, E.A.; Gandia, R.M.; Cortez, M.L.J.; Mattioli, M.C. Multispectral remote sensing in the identification and mapping of biotic and abiotic coffee tree variables. *Rev. Ceres* **2019**, *66*, 142–153. [[CrossRef](#)]
19. Srivastava, A.; Jain, S.; Maity, R.; Desai, V.R. Demystifying artificial intelligence amidst sustainable agricultural water management. *Curr. Dir. Water Scarcity Res.* **2022**, *7*, 17–35.
20. Wang, Z.; Ren, Z.; Li, X. Identification of Coffee Leaf Pests and Diseases based on Transfer Learning and Knowledge Distillation. *Front. Comput. Intell. Syst.* **2023**, *5*, 15–18. [[CrossRef](#)]
21. Alvares, C.A.; Stape, J.L.; Sentelhas, P.C.; Gonçalves, J.L.M.; Sparovek, G. Köppen's climate classification map for Brazil. *Meteorologische* **2013**, *22*, 711–728. [[CrossRef](#)] [[PubMed](#)]

22. Santos, S.A.D.; Ferraz, G.A.E.S.; Figueiredo, V.C.; Volpato, M.M.L.; Machado, M.L.; Silva, V.A. Evaluation of the Water Conditions in Coffee Plantations Using RPA. *Agriengineering* **2023**, *5*, 65–84. [[CrossRef](#)]
23. Meyer, G.E.; Hindman, T.W.; Laksmi, K. Machine Vision Detection Parameters for Plant Species Identification. In *Precision Agriculture and Biological Quality*; SPIE: Boston, MA, USA, 1999; Volume 3543, pp. 327–335.
24. Gitelson, A.A.; Merzlyak, M.N. Signature analysis of leaf reflectance spectra: Algorithm development for remote sensing of chlorophyll. *J. Plant Physiol.* **1996**, *148*, 494–500. [[CrossRef](#)]
25. Rondeaux, G.; Steven, M.; Baret, F. Optimization of soiladjusted vegetation indices. *Remote Sens. Environ.* **1996**, *55*, 95–107. [[CrossRef](#)]
26. Gamon, J.A.; Surfus, J.S. Assessing leaf pigment content and activity with a reflectometer. *New Phytol.* **2013**, *143*, 105–117. [[CrossRef](#)]
27. Yang, Z.; Willis, P.; Mueller, R. Impact of band-ratio enhanced AWIFS image to crop classification accuracy. In Proceedings of the 17th William Pecora Memorial Remote Sensing Symposium, Denver, CO, USA, 18–20 November 2008; American Society for Photogrammetry & Remote Sensing: Bethesda, MD, USA, 2008; pp. 1–11.
28. Mao, W.; Wang, Y.; Wang, Y. Real-time detection of between-row weeds using machine vision. In Proceedings of the 2003 ASAE Annual Meeting; American Society of Agricultural and Biological Engineers, Las Vegas, NV, USA, 27–30 July 2003.
29. Gitelson, A.; Merzlyak, M.N. Quantitative estimation of chlorophyll-a using reflectance spectra: Experiments with autumn chestnut and maple leaves. *J. Photochem. Photobiol. B Biol.* **2004**, *22*, 247–252. [[CrossRef](#)]
30. Rouse, J.W.; Haas, R.H.; Schell, J.A.; Deering, D.W. Monitoring vegetation systems in the great plains with ERTS. In Proceedings of the Earth Resources Technology Satellite (ERTS) Symposium, NASA SP-351, Washington, DC, USA, 10–14 December 1974; pp. 309–317.
31. Huete, A.R. A soil-adjusted vegetation index (SAVI). *Remote Sens. Environ.* **1988**, *25*, 95–309. [[CrossRef](#)]
32. Breiman, L. *Manual Sobre Configuração, Uso e Compreensão de Florestas Aleatórias V3.1*; Universidade de Berkeley: Berkeley, CA, USA, 2001; p. 33.
33. Gislason, P.O.; Benediktsson, J.A.; Sveinsson, J.R. Random Forest classification of multisource remote sensing and geographic data. In Proceedings of the IGARSS 2004. 2004 IEEE International Geoscience and Remote Sensing Symposium, Anchorage, AK, USA, 20–24 September 2004; IEEE: Piscataway, NJ, USA, 2004; Volume 2, pp. 1049–1052.
34. Chinchor, N.; Sundheim, B.M. MUC-5 evaluation metrics. In Proceedings of the Fifth Message Understanding Conference (MUC-5), Baltimore, MD, USA, 25–27 August 1993.
35. Pérez-Ortiz, M.; Peña, J.M.; Gutiérrez, P.A.; Torres-Sánchez, J.; Hervás-Martínez, C.; López-Granados, F. A semi-supervised system for weed mapping in sunflower crops using unmanned aerial vehicles and a crop row detection method. *Appl. Soft Comput. J.* **2015**, *37*, 533–544. [[CrossRef](#)]
36. Whyte, A.; Ferentinos, K.P.; Petropoulos, G.P. A new synergistic approach for monitoring wetlands using Sentinels-1 and 2 data with object-based machine learning algorithms. *Environ. Model. Softw.* **2018**, *104*, 40–54. [[CrossRef](#)]
37. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
38. Mahdianpari, M.; Salehi, B.; Mohammadimanesh, F.; Motagh, M. Classificação aleatória de zonas úmidas florestais usando banda L ALOS-2, banda C RADARSAT-2 e imagens TerraSAR-X. *ISPRS J. Fotogram. Sens. Remoto* **2017**, *130*, 13–31. [[CrossRef](#)]
39. De Almeida Furtado, L.F.; Silva, T.S.F.; de Moraes Novo, E.M.L. Avaliação SAR de banda C de dupla estação e polarimétrica total para mapeamento de vegetação nas áreas úmidas de várzea amazônica. *Sensoriamento Remoto Meio Ambiente* **2016**, *174*, 212–222.
40. Almeida, K.E.C.; Souza, T.T.C.; Costa, A.L.; Marcelo, A.S.D.S.; Laia, L. Near-infrared spectroscopy for the evaluation and prediction of injuries associated with the attack of *Thaumastocoris peregrinus* (Carpintero & Dellapé) in *Eucalyptus camaldulensis* (Dehnh). *Sci. For.* **2018**, *46*, 9–16.
41. Guijarro, M.; Pajares, G.; Riomoros, I.; Herrera, P.J.; Burgos-Artizzu, X.P.; Ribeiro, A. Automatic segmentation of relevant textures in agricultural images. *Comput. Electron. Agric.* **2011**, *75*, 75–83. [[CrossRef](#)]
42. Gameiro, S.; Teixeira, C.P.B.; Silva Neto, T.A.; Lopes, M.F.L.; Duarte, C.R.; Souto, M.V.S.; Zimback, C.R.L. Evaluation of vegetation by means of vegetation index (NDVI, SAVI, and IAF) in the lower jaguaribe hydrographic sub-basin, CE. *Terrae* **2016**, *13*, 15–22.
43. Andreazi, E.; Sera, G.H.; Faria, R.T.D.; Sera, T.; Shigueoka, L.H.; Carvalho, F.G.; Chamlet, D. Performance of f1 hybrids of arabica coffee with simultaneous resistance to leaf rust, bacterial blight and leaf miner. *Coffee Sci.* **2015**, *10*, 375–382.
44. Zhang, X.; Zhang, Z.; Zhang, Y.; Zhang, Q.; Liu, X.; Chen, J.; Wu, L. Influences of fractional vegetation cover on the spatial variability of canopy SIF from unmanned aerial vehicle observations. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *107*, 102712. [[CrossRef](#)]
45. Wu, L.; Wang, L.; Shi, C.; Yin, D. Detecting mangrove photosynthesis with solar-induced chlorophyll fluorescence. *Int. J. Remote Sens.* **2022**, *43*, 1037–1053. [[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.