AgriEngineering

MDPI

Article

Detection of Coffee Leaf Miner Using RGB Aerial Imagery and Machine Learning

Emerson Ferreira Vilela [1,*], Cileimar Aparecida da Silva [1], Jéssica Mayara Coffler Botti [1], Elem Fialho Martins [1], Charles Cardoso Santana [2], Diego Bedin Marin [1], Agnaldo Roberto de Jesus Freitas [3], Carolina Jaramillo-Giraldo [1], Iza Paula de Carvalho Lopes [1], Lucas de Paula Corrêdo [4], Daniel Marçal de Queiroz [5], Giuseppe Rossi [6], Gianluca Bambi [6], Leonardo Conti [6] and Madelaine Venzon [1,*]

1 Agriculture and Livestock Research Enterprise of Minas Gerais (EPAMIG-Sudeste), Viçosa 36570-000, MG, Brazil
2 Minas Gerais Agricultural Research Agency (EPAMIG)—Pitangui Institute of Agricultural Technology (ITAP), Pitangui 35650-000, MG, Brazil
3 Department of Soil, Federal University of Viçosa, Viçosa 36570-900, MG, Brazil
4 Department of Agronomy, Federal University of Viçosa, Viçosa 36570-900, MG, Brazil
5 Department of Agricultural Engineering, Federal University of Viçosa, Viçosa 36571-900, MG, Brazil
6 Department of Agriculture, Food, Environment and Forestry, University of Florence, 50121 Florence, Italy
* Correspondence: efvilela@yahoo.com.br (E.F.V.); madelaine@epamig.br (M.V.)

Abstract: The sustainability of coffee production is a concern for producers around the world. To be sustainable, it is necessary to achieve satisfactory levels of coffee productivity and quality. Pests and diseases cause reduced productivity and can affect the quality of coffee beans. To ensure sustainability, producers need to monitor pests that can lead to substantial crop losses, such as the coffee leaf miner, *Leucoptera coffeella* (Lepidoptera: Lyonetiidae), which belongs to the Lepidoptera order and the Lyonetiidae family. This research aimed to use machine learning techniques and vegetation indices to remotely identify infestations of the coffee leaf miner in coffee-growing regions. Field assessments of coffee leaf miner infestation were conducted in September 2023. Aerial images were taken using remotely piloted aircraft to determine 13 vegetative indices with RGB (red, green, blue) images. The vegetation indices were calculated using ArcGis 10.8 software. A comprehensive database encompassing details of coffee leaf miner infestation, vegetation indices, and crop data. The dataset was divided into training and testing subsets. A set of four machine learning algorithms was utilized: Random Forest (RF), Logistic Regression (LR), Support Vector Machine (SVM), and Stochastic Gradient Descent (SGD). Following hyperparameter tuning, the test subset was employed for model validation. Remarkably, both the SVM and SGD models demonstrated superior performance in estimating coffee leaf miner infestations, with kappa indices of 0.6 and 0.67, respectively. The combined use of vegetation indices and crop data increased the accuracy of coffee leaf miner detection. The RF model performed poorly, while the SVM and SGD models performed better. This situation highlights the challenges of tracking coffee leaf miner infestations in fields with varying ages of coffee plants, different cultivars, and other environmental variables.

Keywords: remote sensing; *Leucoptera coffeella*; artificial intelligence

Citation: Vilela, E.F.; Silva, C.A.d.; Botti, J.M.C.; Martins, E.F.; Santana, C.C.; Marin, D.B.; Freitas, A.R.d.J.; Jaramillo-Giraldo, C.; Lopes, I.P.d.C.; Corrêdo, L.d.P.; et al. Detection of Coffee Leaf Miner Using RGB Aerial Imagery and Machine Learning. *AgriEngineering* **2024**, *6*, 3174–3186. https://doi.org/10.3390/agriengineering6030181

Academic Editor: Murali Krishna Gumma

Received: 13 July 2024
Revised: 27 August 2024
Accepted: 2 September 2024
Published: 5 September 2024

1. Introduction

Coffee (*Coffea arabica*) is recognized as a global commodity, impacting the lives of millions from producers to consumers [1]. In 2023, Brazil was the top producer of Coffea arabica, processing 38.9 million bags [2]. However, production can be impacted by pest infestations, such as the coffee leaf miner (CLM), *Leucoptera coffeella*. This insect belongs to the Lepidoptera order and the Lyonetiidae family. It is a significant pest in unshaded, arid, and hot regions like the Cerrado, a Brazilian biome resembling savanna vegetation [3–8].

The CLM females lay eggs on the leaves of coffee plants and, after hatching in 3–21 days, the larvae feed on the palisade parenchyma cells [4,9,10]. The damage may cause defoliation up to 70% at high population levels, presenting a decrease in photosynthesis rate and reducing coffee yields by 80% [3,11].

For effective CLM control in coffee crops, it is extremely necessary to conduct decision-making at the right moment of intervention. For this reason, the coffee crops need to be constantly monitored, mainly during the critical infestation period. The CLM infestation leads to defoliation of the plant from top to bottom, and a more intense attack can lead to complete defoliation of the upper part of the coffee trees [3]. These features can help monitor the miner infestation through remote sensing.

Remote sensing research using drones has become increasingly vital for advancing coffee farming, especially in crop monitoring [12]. This approach facilitates the ongoing evaluation of extensive coffee plantations, providing an efficient and low-cost method for early detection of phytosanitary issues. Thus, some research has demonstrated the possibility of using machine learning techniques and aerial images to monitor important damage caused mainly by pests and diseases in coffee farming, such as rust [13,14], nematodes [15], and coffee leaf miner [16,17], and damage by frost in coffee plants [18]. The study by Dos Santos et al. [16] aimed to identify lesions on leaves infested with coffee leaf miners, cultivar Catuaí Vermelho IAC 99, using images from drones flying at a height of 3 m, in a coffee plant less than one year old. In another study [17], to identify plants infested with CLM, a drone flight was carried out at a height of 50 m, in a cultivar Catuaí Vermelho IAC 99, also less than 1 year old, and the Random Forest algorithm was used.

However, there are still research gaps that need to be investigated, as several factors influence the spectral response of the plant, which can interfere in the assessment of coffee leaf miner infestation through remote sensing on a larger scale. Therefore, studies should be carried out in coffee plantations at different phenological stages and ages, and with different cultivars and management effects, among other environmental effects, as these may interfere with large-scale monitoring, in addition to evaluating different machine learning models. Remote monitoring of the CLM infestation is a complex practice that must take into account various environmental and crop characteristics, and management factors. Therefore, more research must be carried out to better understand the detection of miner infestation through machine learning and remote sensing techniques.

Thus, with this research we aimed to detect the infestation of CLM in coffee crops of different ages, using different crop management practices, cultivars, machine learning techniques, and vegetation indices derived from aerial images captured by an RGB (red, green, blue) camera mounted on a remotely piloted aircraft.

2. Materials and Methods

To detect CLM infestation using a remotely piloted aircraft, infestation data were first collected from identified plants, and then a flight was performed. Vegetation indices were calculated from the orthomosaic generated. Machine learning algorithms were used from this database and compared using performance metrics (Figure 1).

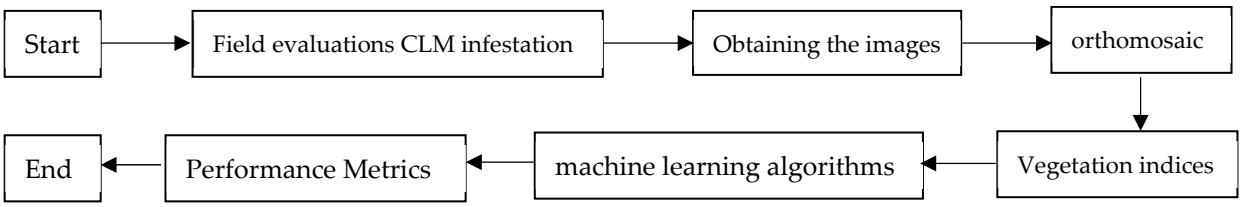

Figure 1. Flowchart of the methodology.

The study was carried out on eight farms situated in the municipalities of Coromandel, Presidente Olegário, Varjão de Minas, Carmo do Paranaíba, and Patrocínio in the State of

Minas Gerais, Brazil, within the Cerrado Biome (Table 1, Figure 2). In each study area, a plot with 0.28 ha of coffee (*Coffea arabica* L.) was selected.

Table 1. Coffee cultivars and characteristics of the study area.

Study Area	Latitude (South)	Longitude (West)	Cultivar	Planting (Year)	Elevation (Meters)
1—Presidente Olegário	18°33′49.82″	46°19′45.51″	IAC 125 (IBC 12)	11 December 2020	1090
2—Varjão de Minas	18°31′22.09″	46°3′46.36″	IPR 100	12 December 2021	0926
3—Varjão de Minas	18°31′18.35″	46°3′46.04″	IPR 100	12 December 2021	0926
4—Carmo do Paranaíba	18°57′57.99″	46°15′46.44″	IPR 100	12 December 2020	1100
5—Carmo do Paranaíba	19°0′6.52″	46°13′51.66″	Catucai 144	12 January 2021	1100
6—Patrocínio	18°59′10.86″	46°58′57.48″	Paraíso	12 December 2008	0987
7—Patrocínio	18°56′28.57″	47°17′18.59″	Paraíso	12 December 2017	1070
8—Coromandel	18°37′18.41″	46°49′52.43″	Paraíso 2	12 December 2020	1140

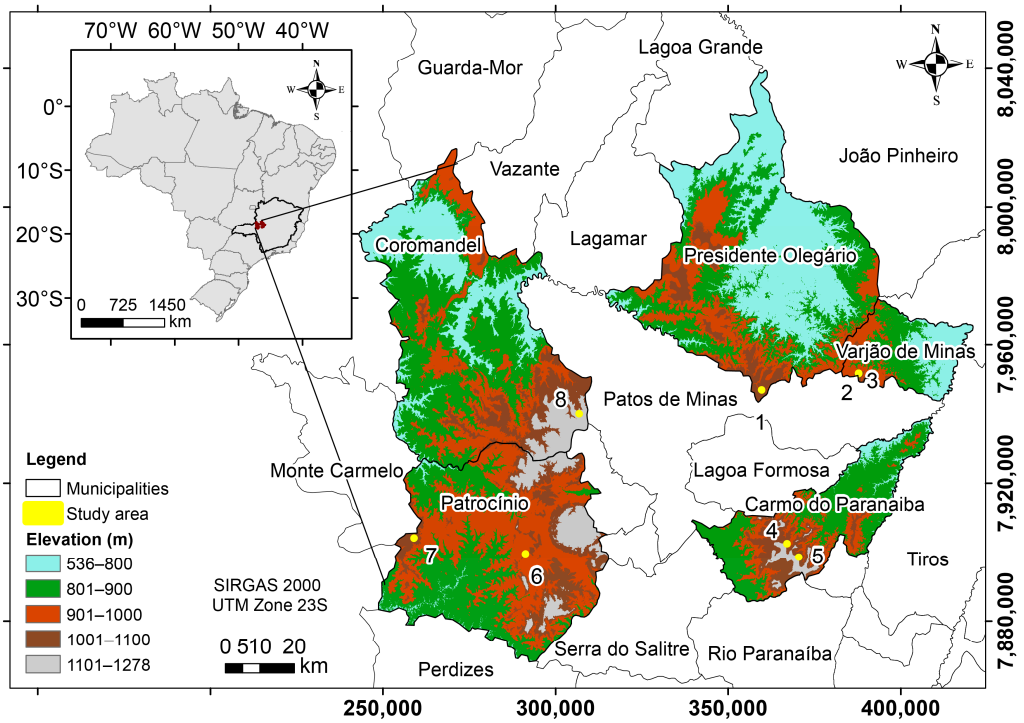

Figure 2. Location of the eight study areas.

2.1. Pest Monitoring

The highest coffee leaf miner infestations in the Cerrado region of Minas Gerais typically occur in September and October [3]. To address this, CLM infestation was assessed in September 2023. Leaves were collected randomly from 10 plants per plot, with one plant chosen from each row. All sampled plants were marked with control points. Samples were collected from the middle section of two branches on opposite sides of each plant, using leaves from the third or fourth pair counted from the tip (adapted from Souza et al., 1998) [9], and a total of four leaves were collected per plant. The coffee leaves were examined to determine active mines (infestation). The CLM infestation rate was calculated using the following formula:

$$\text{CLM Infestation rate (\%)} = (\text{Number of coffee leaf with active mines} \times 100)/n^* \quad (1)$$

* n = Total number of collected coffee leaves

The CLM infestation data were summarized into two classes: "infested", when there was presence of infestation in the study area, and "healthy", when there was no presence of infestation in the experimental field

2.2. Aerial Data Collection

In September, aerial surveys were conducted with the remotely piloted DJI Mavic 2 Zoom aircraft. An RGB sensor from the aircraft was employed to capture aerial images. This sensor can acquire images with a 12-megapixel resolution (4000 × 3000 pixels). The flight plan was established using the DroneDeploy application—5.43.0, with the following parameters: a flight altitude of 60 m and image overlaps of 80% longitudinally and 75% laterally, with the survey conducted at a speed of 5 m/s. The images were processed using Agisoft software, Version 1.5.1. Approximately 40 photos per farm were processed. All collected images were used to generate the othomosaics. Thus, utilizing the raster calculator tool to perform arithmetic operations on the red, green, and blue bands with the reflectance values obtained from the ArcGis 10.8 software, 13 spectral indices were used, according to Table 2. A 50 cm buffer, with 0.78 m^2, was generated on top of each plant, and the vegetation index (VI) values were extracted; the average value was then determined with R 4.3.3 software.

Table 2. Vegetation indices were calculated from RGB images.

Vegetation Index	Equation	Reference
Normalized Redness Intensity—NRI	$\frac{RED}{RED+GREEN+BLUE}$	[19]
Excess Green Index—EXG	$2 \times RED - (GREEN + BLUE)$	[20]
Green Red Ratio Vegetation Index—GRRI	$\frac{GREEN}{RED}$	[21]
Green Blue Ratio Index—GBRI	$\frac{GREEN}{BLUE}$	[22]
Red Blue Ratio Index—RBRI	$\frac{RED}{BLUE}$	[22]
Woebbecke Index—WI	$\frac{GREEN-BLUE}{RED-GREEN}$	[23]
Normalized Pigment Chlorophyll Ratio Index—NPCI	$\frac{BLUE-RED}{BLUE+RED}$	[24]
Normalized Green–Red Difference Index—NGRDI	$\frac{GREEN-RED}{GREEN+RED}$	[25]
Redness Index—RI	$\frac{RED^2}{GREE^3 * BLUE}$	[26]
Primary Colors Hue Index—HI	$\frac{2 \times RED-GREEN-BLUE}{GREEN-BLUE}$	[27]
Green Leaf Index—GLI	$\frac{2 \times GREEN-RED-BLUE}{2 \times GREEN+RED+BLUE}$	[28]
Spectra Slope Saturation Index—SI	$\frac{RED-BLUE}{RED+BLUE}$	[29]
Normalized Blueness Intensity—NBI	$\frac{BLUE}{RED+GREEN+BLUE}$	[19]

A detailed statistical analysis was performed on the reflectance values of the vegetation index data. To analyze the VI in infested and healthy coffee plants, we used the generalized linear mixed model (GLMM). When necessary, the BoxCox transformation was applied to each of the numerical variables. The infested and healthy states were defined as fixed effects, and the farm, cultivar, and planting year were random effects. Coffee cultivars present great variability in their morphology, plant height, leaf, and canopy shape, among others. Therefore, the cultivar was set as a random effect. The vegetation index data were analyzed using principal component analysis (PCA) and permutational multivariate analysis of variance (PERMANOVA), with 1000 permutations and Euclidean distance, to explore patterns among various treatment groups (infested and healthy). The data were normalized by subtracting the mean and dividing by the standard deviation, and eigenvalues and eigenvectors were computed from the covariance matrix. All statistical analyses were performed with R 4.3.3 software (R Core Team, 2024).

2.3. Model Development

A database containing infestation data and 13 vegetation indices for September 2023 was created. The algorithms employed included Random Forest (RF), Stochastic Gradient Descent (SGD), Support Vector Machine (SVM), and Logistic Regression (LR). A database was created with infestation data and 13 vegetation indices for September 2023. The algorithms used comprised Random Forest (RF), Stochastic Gradient Descent (SGD), Support Vector Machine (SVM), and Logistic Regression (LR). The implementation was conducted in Python (version 3.9), utilizing libraries like Numpy [30], Pandas [31], and Scikit learn [32].

The data were split into 80% for training and 20% for testing. Subsequently, variable selection and hyperparameter tuning were implemented for each machine learning model (Table 3). Accuracy was selected as the evaluation metric, utilizing 5-fold cross-validation. To determine the most relevant variables, recursive feature elimination with cross-validation was performed. Hyperparameter tuning was carried out using 5-fold cross-validation. Value determination and hyperparameter selection were guided by the Scikit-learn library [32]. Primary performance indicators commonly used in machine learning, such as the confusion matrix, precision, kappa, auc, recall, and F1 score, were utilized to assess the efficacy of the CLM monitoring models [17].

Table 3. Hyperparameters optimized for machine learning algorithms.

Algorithm	Hyperparameter	Value	Algorithm	Hyperparameter	Value
RF	Number of trees	123	SGD	Power t	3
	Criterion	gini		Loss	hinge
	Maximum depth	17		Penalty	l2
SVM	C	5	LR	C	2
	Kernel	rbf		Tol	0
	Degree	10		Max iter	152
	Gamma	scale		Intercept scaling	0

3. Results

3.1. Exploratory Data Analysis

Monitoring of Coffee Leaf Miner Infestation and Image Selection

The summary of the descriptive analysis data of the two classes, healthy and infested, based on the CLM infestation monitoring data recorded in the eight experimental plots in September of 2023, is presented in Table 4. The CLM infestation varied between 0 and 75%, with an average of 10%. Approximately 75% of the plants evaluated presented an infestation below 25% (Table 4). The descriptive analysis of the values of the vegetative indices are shown in Table 5.

Table 4. Descriptive analysis of coffee leaf miner infestation in September 2023.

Year	Minimum	Q1	Mean	Q3	Maximum	Sd
2023	0.000	0.000	10.58	25.00	75.00	19.54

Q1: Quartile 1; Q3: Quartile 3; Sd: Standard deviation.

Table 5. Summary of descriptive statistical values for VI in healthy and infested plants.

VI	State of the Plants	Mean	Mediana	SD	Min	Max
EXG	healthy	30.323	12.586	44.915	−14.752	140.502
	Infested	22.042	15.458	26.812	−2.347	130.963
GBRI	healthy	2.334	2.366	0.588	1.147	3.913
	Infested	2.215	2.150	0.468	1.308	3.226
GLI	healthy	0.205	0.225	0.063	0.034	0.274
	Infested	0.192	0.196	0.046	0.080	0.264

Table 5. *Cont.*

VI	State of the Plants	Mean	Mediana	SD	Min	Max
GRRI	healthy	1.279	1.301	0.134	1.014	1.572
	Infested	1.221	1.206	0.099	1.076	1.405
HI	healthy	3.537	2.538	4.062	−0.304	23.929
	Infested	3.304	2.908	2.008	−0.114	10.365
NBI	healthy	0.225	0.217	0.029	0.181	0.306
	Infested	0.223	0.221	0.023	0.190	0.290
NGRDI	healthy	0.115	0.127	0.050	0.006	0.201
	Infested	0.095	0.091	0.038	0.036	0.162
NPCI	healthy	−0.218	−0.226	0.065	−0.342	−0.059
	Infested	−0.231	−0.238	0.056	−0.336	−0.081
NRI	healthy	0.342	0.343	0.015	0.302	0.376
	Infested	0.351	0.350	0.013	0.324	0.372
RBRI	healthy	1.744	1.713	0.343	1.132	2.752
	Infested	1.757	1.728	0.289	1.188	2.510
RI	healthy	0.000	0.000	0.000	0.000	0.000
	Infested	0.000	0.000	0.000	0.000	0.000
SI	healthy	0.218	0.226	0.065	0.059	0.342
	Infested	0.231	0.238	0.056	0.081	0.336
WI	healthy	−2.614	−2.445	0.855	−5.437	−0.914
	Infested	−3.385	−3.127	1.221	−6.909	−1.880

The EXG index presented a greater standard deviation, that is, greater dispersion of data, which is also evident from the larger difference between the median and mean compared to the other indices. The NRI index presented greater uniformity of data, and lower standard deviation.

The GLMM was applied to compare the mean VI between the two plant conditions studied (Figure 3). The statistical results provide insight into the variations of VI in both healthy and infested plants.

For all indices, the *p*-value was greater than 0.05, suggesting no significant difference in the means between the infested and healthy plants (Figure 3).

The entire dataset of vegetation index and coffee leaf miner infestation was analyzed by principal component analysis (PCA) (Figure 4). The PCA diagram shows the relationship among the different vegetation indices and the two plant states (infested and healthy).

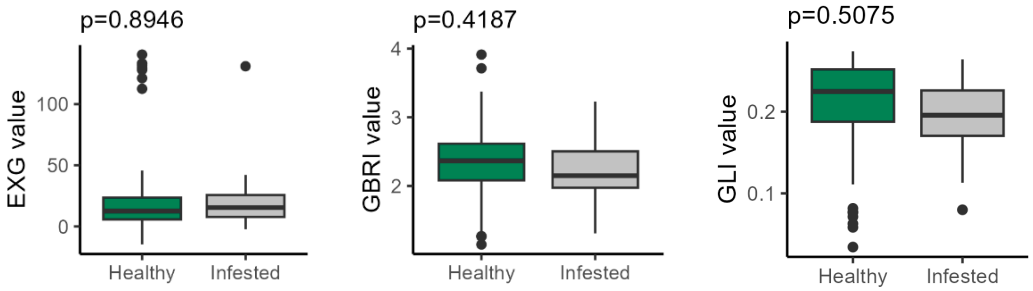

Figure 3. *Cont.*

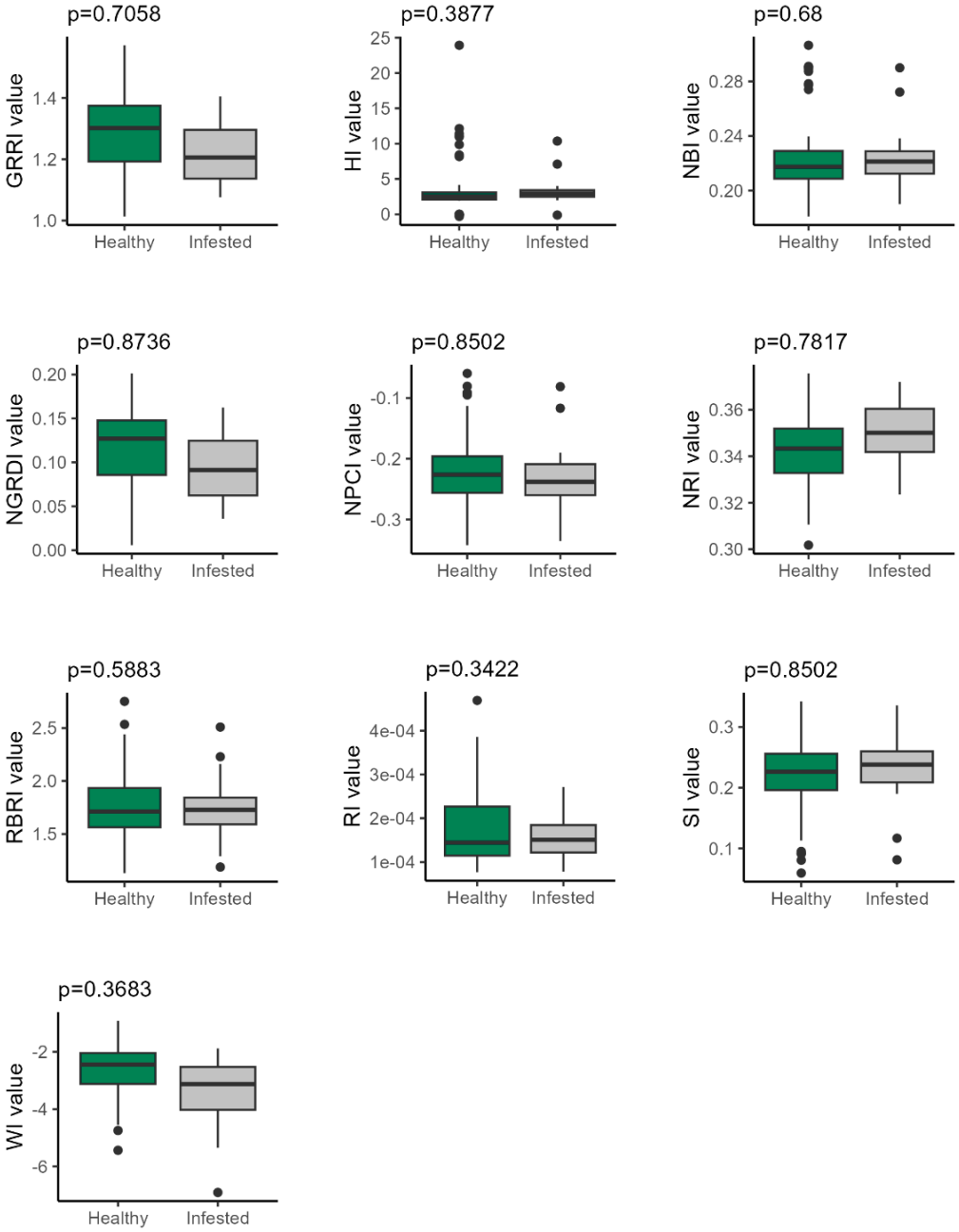

Figure 3. Box plot of VI values for healthy and infested conditions, showing the *p*-value from the generalized linear mixed model.

Thus, PC1 and PC2, the first and second components of PCA, explained 53.1% and 26.8% of the total variance, respectively. Sample homogeneity within each treatment group was confirmed (homogeneous dispersion) (PERMIDISP: F = 3.05, $p = 0.08$). No significant differences were found in the detection patterns of coffee leaf miner infestation using the vegetation indices (F = 1.96, $p = 0.12$, Figure 4).

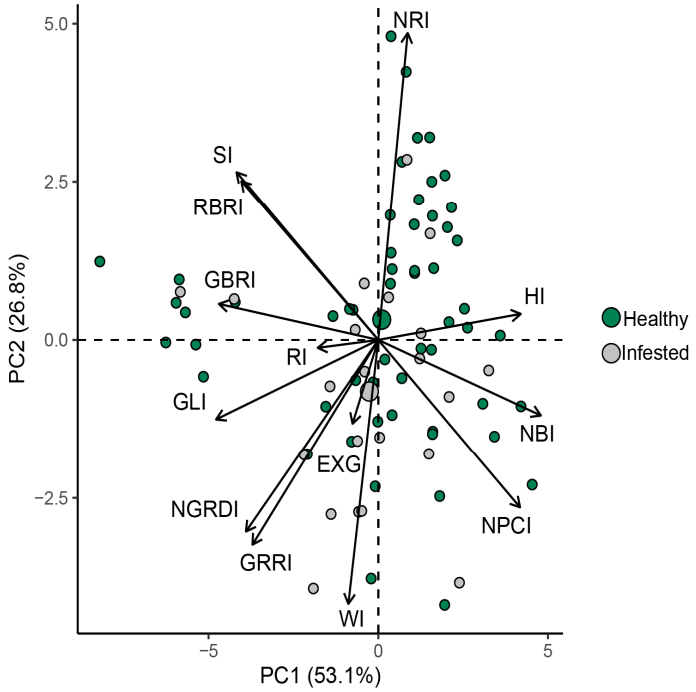

Figure 4. Principal component analysis (PCA). The percentages on the axes represent the proportion of variance explained by each principal component. Loadings of the principal components for resource contributions, with arrow directions illustrating the relative importance of resources in the first and second components.

3.2. Machine Learning

Apart from the vegetative indices, the age of the plants and the cultivar were the most significant variables for the machine learning models (Figure 5). Plant age was the non-vegetative index that presented significant importance, reaching 30% in the SVM models and 28% in the LR. The vegetative index HI presented the best performance in the models presented. The RF model selected only the NRI index.

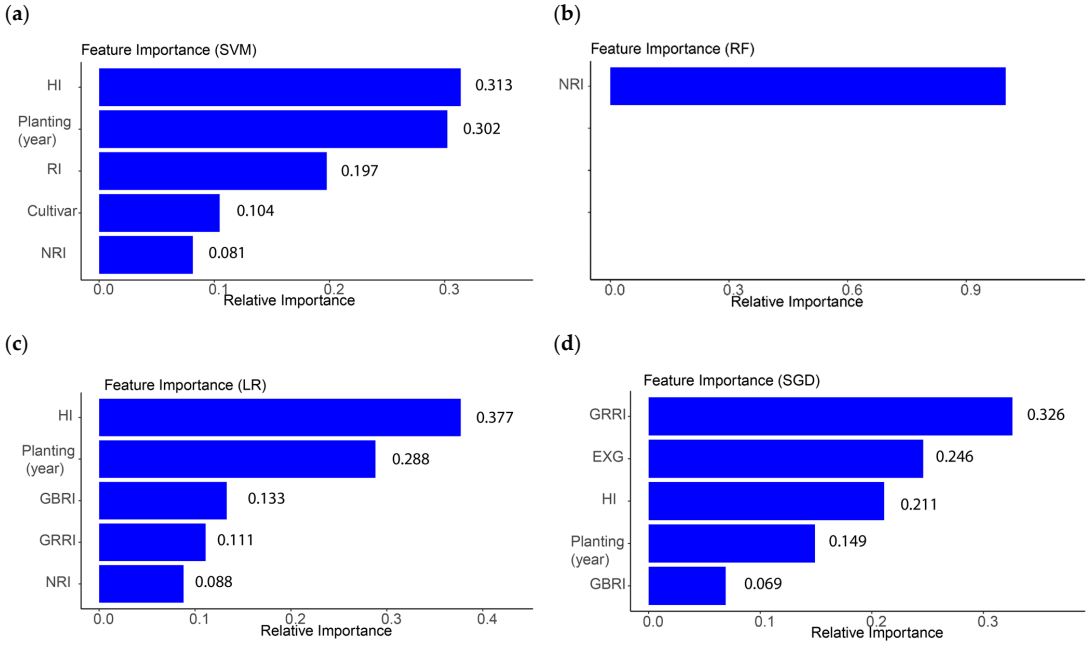

Figure 5. Key variables for predicting coffee leaf miner infestation. (**a**) Support Vector Machine (SVM). (**b**) Random Forest (RF). (**c**) Logistic Regression (LR). (**d**) Stochastic Gradient Descent (SGD).

Algorithm Performance

The RF algorithms were unable to identify plants with CLM infestation. All predictions from the RF algorithm indicated that the plants were not infested with CLM. The SGD algorithm showed the best performance in identifying the infestation, followed by the SVM algorithm, with 75% and 50% accuracy in identifying plants with infestation, respectively (Figure 6).

Figure 6. Confusion matrices for machine learning algorithms used to forecast coffee leaf miner infestation (Infested = coffee plants affected by leaf miner; Healthy = plants free from leaf miner). The algorithms employed included Random Forest (RF), Support Vector Machine (SVM), and Stochastic Gradient Descent (SGD). The confusion matrices illustrate the proportion of accurate and inaccurate predictions for each category.

To evaluate the performance metrics, the Random Forest model showed precision, recall, F1 e Kappa, and had a value of zero. On the other hand, the SGD and SVM algorithms presented satisfactory results, with Kappa of 0.67 and 0.6, and precision of 0.75 and 1, respectively (Table 6).

Table 6. Performance metrics for the classification algorithms.

Algorithm	Precision	Recall	Auc	f1_Score	Log_Loss	Kappa
SVM	1	0.5	0.75	0.66	8.63489	0.60
RF	0	0	0.5	0	8.63489	0
LR	1	0.37	0.68	0.54	11.5131	0.44
SGD	0.75	0.75	0.83	0.75	8.63489	0.67

4. Discussion

In the study areas, CLM infestation had an average of 10.58% and a standard deviation of 19.54%, however, it reached 75% in some plants (Table 4). The months of September/October are considered the months with the highest incidence of CLM infestation due to low relative humidity, low precipitation, and high temperatures. These weather conditions favor the development of the CLM population [3].

The identified variability highlights a highly heterogeneous distribution of infestation, indicating that local factors such as microclimate and management practices play important roles in determining infestation levels. The need for specific and localized management strategies became evident with the presence of plants with no infestation (0%), alongside highly infested plants (75%). Although the average infestation was relatively low, the standard deviation was high (Table 5), demonstrating the importance of continuous monitoring

to quickly identify critical areas and carry out targeted management. Additionally, the variability of the observed data helps to understand the complexity and the most influential factors in the infestation, which can be incorporated into machine learning models to improve their predictive capacity.

The descriptive analysis and PCA indicated some vegetation indices with the potential to be used to identify coffee leaf miner infestation, such as NRI, EXG, GBRI, GRRI, WI, and HI (Table 5). The data were similar to those found by dos Santos et al. (2024) [16], who used the vegetation indices GRRI and MPCI to separate healthy plants. However, when evaluating the indices using generalized linear models, the indices showed no difference between infested and healthy plants (Figure 3). When considering the variability in areas with different management, plants of different ages, and cultivars, the univariate analysis of the vegetation indices was not able to differentiate between infested and healthy plants. Due to the complexity of evaluating areas with coffee plantations with different characteristics, such as age, nutritional status, and cultivar, the use of vegetation indices derived from RGB images were not sensitive enough to detect CLM infestation in univariate analyses (Figure 3). When monitoring large-scale CLM infestations in coffee plants across different environments, vegetation indices that include the infrared band may offer greater sensitivity in univariate analyses. This is because the reflectance of the infrared band is influenced by alterations in the leaf parenchyma [33], which is the tissue targeted by coffee leaf miner larvae, leading to necrosis [3,8]. For CLM monitoring, the GRNDVI index showed the best performance in differentiating infested and healthy plants. Another situation that may have influenced the result is the flight height. The present study carried out the flight at a height of 60 m. Flights at lower heights may be more suitable as they generate more detailed images and, consequently, more sensitive vegetation indices to detect changes caused by CLM infestation in a univariate analysis. Detection of Cercospora disease symptoms in coffee plants was achieved through vegetation indices generated from RGB images taken by a remotely piloted aircraft flying at 30 m altitude [34].

The four algorithms based on machine learning employed to differentiate between coffee leaf miner-infested and healthy plants demonstrated varying levels of effectiveness. All models had high accuracy in identifying healthy plants, but most of them performed poorly when identifying infested plants. The Random Forest model was not able to identify CLM in infested plants through remote sensing. In an effort to assess the potential of RGB-based VI for monitoring nitrogen levels in a coffee-growing area using a remotely piloted aircraft at a fixed flight altitude of 50 m, the Random Forest algorithm performed poorly when assessing and forecasting nitrogen content in coffee plants, as indicated by a Kappa value of −0.02 [35]. Another study utilized the Random Forest algorithm along with vegetation indices obtained from a remotely piloted aircraft to estimate the nitrogen content in coffee plants [36]. The RF algorithm exhibited reduced effectiveness when applied to vegetation indices derived from RGB bands compared to those incorporating the infrared band [36]. The RF algorithm showed lower performance when using vegetation indices based on RGB bands, when compared to indices that used indices that had the infrared band [36]. The Random Forest model has shown good results in detecting CLM infestation [16]. However, this study employed vegetative indices that included green, red, near-infrared (NIR), and edge bands.

The selection of variables is important to make the algorithm simpler and more capable of generalization. The algorithms selected between five and one variable, according to the models. The planting (year) variable was chosen in almost all models, indicating that the age of the plant should be considered when evaluating the spectral response of the plant. The vegetation indices HI and NRI were the most important indices, as they were chosen by most models. The algorithms mostly presented Kappa values between 0.4 and 0.8, which can be considered moderate/substantial [37]. High precision indicates that the models presented better performance when predicting healthy plants.

Our results demonstrate that machine learning-based approaches have promising results for detecting patterns of changes due to CLM infestation that are not found in

univariate behavioral assessments when evaluating indices formed with RGB bands. However, more studies should be carried out using cameras that have images with bands other than RGB. Lower flight heights that present more details are also necessary for a better assessment of plant infestation with coffee leaf miner. Furthermore, other machine learning algorithms can be evaluated to detect coffee leaf miner infestation, such as convolutional neural networks. The inclusion of a larger amount of databases, in addition to the effect of other variables such as other diseases and pests that can influence monitoring, should be considered. Finally, the possibility of monitoring via plant stress signals, as well as thermal sensors, can represent an alternative for future evaluations. However, further research is needed to improve the detection of coffee leaf miner infestations using aerial images and machine learning techniques.

5. Conclusions

The combined use of vegetation indices and crop data increased the accuracy of coffee leaf miner detection. The machine learning models performed variably. The RF model performed poorly, while the SVM and SGD models performed best. This dynamic reflects the complexity of monitoring coffee leaf miner infestation in areas with coffee of different ages, cultivars, and other environmental factors. The present work contributes substantially to the advancement of knowledge about monitoring coffee leaf miners. However, there are limitations due to the complexity involved in the remote monitoring of coffee leaf miner infestation, as several factors influence the spectral response of the plant, such as stress caused by diseases and pests, water vigor, nutritional status, and pruning, among others. For future work, other machine learning algorithms and lower flight heights can be tested.

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

Funding: This research was funded by the "Fundação de Amparo à Pesquisa de Minas Gerais" (FAPEMIG), the "Conselho Nacional de Desenvolvimento Científico e Tecnológico" (CNPq), and the "Consórcio Brasileiro de Pesquisa e Desenvolvimento do Café" (CBP&D-Café).

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author/s.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study, in the collection, analyses, or interpretation of the data, in the writing of the manuscript, or in the decision to publish the results.

References

1. Pancsira, J. International Coffee Trade: A literature review. *J. Agric. Inform.* **2022**, *13*, 26–35. [CrossRef]
2. CONAB Companhia Nacional de Abastecimento. Historical Series—Arabica Coffee—Brazil. Available online: https://www.conab.gov.br/info-agro/safras/serie-historica-das-safras#café-2 (accessed on 1 July 2024).
3. Reis, P.R.; Souza, J.C.; Silva, R.A.; Santa-Cecília, L.V.C. Principais pragas do cafeeiro no Cerrado Mineiro: Reconhecimento e manejo. In *Cafeicultura do Cerrado*; Carvalho, G.R., Ferreira, A.D., Andrade, V.T., Botelho, C.E., Carvalho, J.P.F., Eds.; EPAMIG: Belo Horizonte, Brazil, 2021; pp. 321–346.
4. Pereira, E.J.G.; Picanço, M.C.; Bacci, L.; Bacci, L.; Della Lucia, T.M.C.; Silva, E.M.; Fernandes, F.L. Natural mortality factors of *Leucoptera coffeella* (Lepidoptera:Lyonetiidae) on Coffea arabica. *Biocontrol Sci.* **2007**, *17*, 441–455. [CrossRef]
5. Fernandes, F.L.; Mantovani, E.C.; Neto, H.B.; Nunes, V.V. Efeitos de variáveis ambientais, irrigação e vespas predadoras sobre *Leucoptera coffeella* (Guérin-Méneville) (Lepidoptera: Lyonetiidae) no cafeeiro. *Neotrop. Entomol.* **2009**, *38*, 410–417. [CrossRef] [PubMed]
6. Pantoja-Gomez, L.M.; Corrêa, A.S.; Oliveira, L.O.; Guedes, R.N.C. Common origin of Brazilian and Colombian populations of the Neotropical coffee leaf miner, *Leucoptera coffeella* (Lepidoptera: Lyonetiidae). *J. Econ. Entomol.* **2019**, *112*, 924–931. [CrossRef]

7. Leite, S.A.; Santos, M.P.; Resende-Silva, G.A.; Costa, D.R.; Moreira, A.A.; Lemos, O.L.; Guedes, R.N.C.; Castellani, M.A. Area-wide survey of chlorantraniliprole resistance and control failure likelihood of the Neotropical coffee leaf miner *Leucoptera coffeella* (Lepidoptera: Lyonetiidae). *J. Econ. Entomol.* **2020**, *113*, 1399–1410. [CrossRef]
8. Venzon, M. Agro-ecological Management of Coffee Pests in Brazil. *Front. Sustain. Food Syst.* **2021**, *5*, 721117. [CrossRef]
9. Souza, J.C.; Reis, P.R.; Rigitano, R.L.D.O. *Bicho-Mineiro do Cafeeiro: Biologia, Danos e Manejo Integrado*; Boletim Técnico, EPAMIG: Belo Horizonte, Brazil, 1998; p. 48.
10. Giraldo-Jaramillo, M.; Garcia-Gonzalez, J.; Rugno, J.B. Fertility life table of *Leucoptera coffeela* (Guérin-Mèneville) (Lepidoptera: Lyonetiidae) at seven temperatures in coffee. *Am. J. Entomol.* **2019**, *3*, 70–76. [CrossRef]
11. Carvalho, C.F.; Carvalho, S.M.; Souza, B. *Coffee*; Souza, B., Vázquez, L.L., Marucci, R.C., Eds.; Natural Enemies of Insect Pests in Neotropical Agroecosystems; Springer Nature: Basel, Switzerland, 2019; pp. 277–292.
12. Bento, N.L.; Ferraz, G.A.e.S.; Santana, L.S.; Silva, M.d.L.O.e. Coffee Growing with Remotely Piloted Aircraft System: Bibliometric Review. *AgriEngineering* **2023**, *5*, 2458–2477. [CrossRef]
13. Velásquez, D.; Sánchez, A.; Sarmiento, S.; Toro, M.; Maiza, M.; Sierra, B. A Method for Detecting Coffee Leaf Rust through Wireless Sensor Networks, Remote Sensing, and Deep Learning: Case Study of the Caturra Variety in Colombia. *Appl. Sci.* **2020**, *10*, 697. [CrossRef]
14. Soares, A.d.S.; Vieira, B.S.; Bezerra, T.A.; Martins, G.D.; Siquieroli, A.C.S. Early Detection of Coffee Leaf Rust Caused by *Hemileia vastatrix* Using Multispectral Images. *Agronomy* **2022**, *12*, 2911. [CrossRef]
15. Pereira, F.V.; Martins, G.D.; Vieira, B.S.; de Assis, G.A.; Orlando, V.S.W. Multispectral images for monitoring the physiological parameters of coffee plants under different treatments against nematodes. *Precis. Agric.* **2022**, *23*, 2312–2344. [CrossRef]
16. 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. [CrossRef]
17. 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]
18. Marin, D.B.; Schwerz, F.; Barata, R.A.P.; de Oliveira Faria, R.; Dias, J.E.L. Unmanned Aerial Vehicle to Evaluate Frost Damage in Coffee Plants. *Precis. Agric.* **2021**, *22*, 1845–1860. [CrossRef]
19. Liu, K.; Li, Y.; Han, T.; Yu, X.; Ye, H.; Hu, H.; Hu, Z. Evaluation of grain yield based on digital images of rice canopy. *Plant Methods* **2019**, *15*, 28. [CrossRef] [PubMed]
20. Torres-Sánchez, J.; Pena, J.M.; De Castro, A.I.; López-Granados, F. Multi-temporal mapping of the vegetation fraction in early season wheat fields using images from UAV. *Comput. Electron. Agric.* **2014**, *103*, 104–113. [CrossRef]
21. Lu, J.; Cheng, D.; Geng, C.; Zhang, Z.; Xiang, Y.; Hu, T. Combining plant height, canopy coverage and vegetation index from UAV-based RGB images to estimate leaf nitrogen concentration of summer maize. *Biosyst. Eng.* **2021**, *202*, 42–54. [CrossRef]
22. Sellaro, R.; Crepy, M.; Trupkin, S.A.; Karayekov, E.; Buchovsky, A.S.; Rossi, C.; Casal, J.J. Cryptochrome as a Sensor of the Blue/Green Ratio of Natural Radiation in Arabidopsis. *Plant Physiol.* **2010**, *154*, 401–409. [CrossRef]
23. Woebbecke, D.M.; Meyer, G.E.; Von Bargen, K.; Mortensen, D.A. Color indices for weed identification under various soil, residue, and lighting conditions. *Trans. ASAE* **1995**, *38*, 259–269. [CrossRef]
24. Peñuelas, J.; Gamon, J.A.; Fredeen, A.L.; Merino, J.; Field, C.B. Reflectance indices associated with physiological changes in nitrogen-and water-limited sunflower leaves. *Remote Sens. Environ.* **1994**, *48*, 135–146. [CrossRef]
25. Tucker, C.J. Red and Photographic Infrared Linear Combinations for Monitoring Vegetation. *Remote Sens. Environ.* **1979**, *8*, 127–150. [CrossRef]
26. Levin, N.; Ben-Dor, E.; Singer, A. A digital camera as a tool to measure colour indices and related properties of sandy soils in semi-arid environments. *Int. J. Remote Sens.* **2005**, *26*, 5475–5492. [CrossRef]
27. Mathieu, R.; Pouget, M.; Cervelle, B.; Escadafal, R. Relationships between Satellite-Based Radiometric Indices Simulated Using Laboratory Reflectance Data and Typic Soil Color of an Arid Environment. *Remote Sens. Environ.* **1998**, *66*, 17–28. [CrossRef]
28. Louhaichi, M.; Borman, M.M.; Johnson, D.E. Spatially Located Platform and Aerial Photography for Documentation of Grazing Impacts on Wheat. *Geocarto Int.* **2001**, *16*, 65–70. [CrossRef]
29. Escadafal, R.; Belghith, A.; Ben-Moussa, H. Indices spectraux pou la télédetection de la dégradation des milieux naturels en Tunisie aride. In *Symposium International sur les Mesures Physiques et Signatures en Télédétection, 6., 1994, Val d'Isère, France. Anaux... Toulouse*; Centre National d'Etude Spatiale: Paris, France, 1994; pp. 253–259.
30. Harris, C.R.; Millman, K.J.; Van der Walt, S.J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N.J.; et al. Array Programming with Numpy. *Nature* **2020**, *585*, 357–362. [CrossRef]
31. McKinney, W. Data structures for statistical computing in Python. In Proceedings of the 9th Python in Science Conference, Austin, TX, USA, 28 June–3 July 2010; pp. 56–61. [CrossRef]
32. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
33. Ponzoni, F.J.; Shimabukuro, Y.E.; Kuplich, T.M. *Sensoriamento Remoto Aplicado ao Estudo da Vegetação*, 2nd ed.; Parêntese: São José dos Campos, Brazil, 2012; 160p.

34. Santos, L.M.; Ferraz, G.A.F.; Santana, L.S.; Barbosa, B.D.S.; Xavier, L.A.G.; Andrade, M.T. *Índice de Vegetação (ExGR) Aplicado a Imagens rgb Obtidas por UAV para Detecção de Doença em Cafeeiros, Proceedings of the X Simpósio de Pesquisa dos Cafés do Brasil, Vitoria, Brazil, 8 November 2019*; Centro de Convenções de Vitoria: Vitoria, Brazil, 2019.
35. Mincato, R.L.; Parreiras, T.C.; Lense, G.H.E.; Moreira, R.S.; Santana, D.B. Using unmanned aerial vehicle and machine learning algorithm to monitor leaf nitrogen in coffee. *Coffee Sci.* **2020**, *15*, e151736.
36. Marin, D.B.; Ferraz, G.A.E.S.; Guimarães, P.H.S.; Schwerz, F.; Santana, L.S.; Barbosa, B.D.S.; Barata, R.A.P.; Faria, R.d.O.; Dias, J.E.L.; Conti, L.; et al. Remotely Piloted Aircraft and Random Forest in the Evaluation of the Spatial Variability of Foliar Nitrogen in Coffee Crop. *Remote Sens.* **2021**, *13*, 1471. [CrossRef]
37. McHugh, M.L. Interrater Reliability: The Kappa Statistic. *Biochem. Med.* **2012**, *22*, 276–282. [CrossRef]