eISSN: 1984-3909
Coffee Science, e182170, 2023

Doi: https://doi.org/10.25186/.v18i.2170

New vegetation index for monitoring coffee rust using sentinel-2 multispectral imagery

Gabriel Dumbá Monteiro de Castro[1], Emerson Ferreira Vilela[2], Ana Luísa Ribeiro de Faria[1], Rogério Antonio Silva[3], Williams Pinto Marques Ferreira[4]

[1]Universidade Federal de Viçosa/UFV, Departamento de Engenharia Agrícola, Viçosa, MG, Brasil
[2]Empresa de Pesquisa Agropecuária de Minas Gerais/EPAMIG Sudeste, Viçosa, MG, Brasil
[3]Empresa de Pesquisa Agropecuária de Minas Gerais/EPAMIG Sudeste, Lavras, MG, Brasil
[4]Empresa Brasileira de Pesquisa Agropecuária/EMBRAPA, Centro Nacional de Pesquisa do Café. Viçosa, MG, Brasil
Contact authors: gdumbamonteirodecastro@gmail.com; efvilela@yahoo.com.br; ana.faria3@ufv.br; rogeriosilva@epamig.br; salva_terra@yahoo.com.br
Received in August 30, 2023 and approved in December 11, 2023

ABSTRACT

Coffee Rust (*Hemileia vastatrix)* is considered the primary coffee disease in the world. The pathogenic fungus can find favorable environmental conditions in different countries, constantly threatening coffee producers. The previous detection of the incidence of coffee rust in a region is crucial because it provides an overview of the disease's progress aiding in coffee plantations management. The objective of this work was the development of a vegetation index for remote monitoring of coffee rust infestation. Using satellite images from the MSI/Sentinel-2 collection, the Machine Learning classifier algorithm - Random Forest, and the cloud processing platform - Google Earth Engine, a supervised classification was performed to determine the most sensitives bands in coffee rust detection, namely B4 (Red), B7 (Red Edge 3) and B8A (Red Edge 4). Thus, the Triangular Vegetation Index method was used to create a new vegetative index for remote detection of coffee rust infestation on a regional scale, named Coffee Rust Detection Index (CRDI). A linear regression model was created to estimate rust infestation based on the performance of the new index. The model presented a coefficient of determination (R^2) of 63.21%, and a root mean square error (RMSE) of 0.103. In addition, a comparison analysis of the new index with eight other vegetative indices commonly used in the literature was carried out. The CRDI obtained the best performance in coffee rust detection among the others. This study shows that the new index CRDI could serve as a crucial tool for monitoring coffee rust infestation on a regional scale.

Key words: *Hemileia vastatrix*; disease monitoring; triangular vegetation index method; google earth engine.

1 INTRODUCTION

Planting coffee (*Coffea arabica* L.) has significant economic, social and cultural relevance worldwide. The global coffee production in 2022 was approximately 174.3 million 60-kg bags (United States Department of Agriculture - USDA, 2023). The success of cultivation often depends on natural factors such as the weather and the infestation of pests and diseases. This way, proper crop management ensures good productivity, generates greater economic returns for producers, and causes a less environmental impact on the ecosystem.

Among the diseases that affect coffee, coffee leaf rust represents a significant threat to producers worldwide, especially in countries with warm and damp climates, where the pathogenic fungus finds favorable environmental conditions to proliferate (Pozza; Carvalho; Chalfoun, 2010). The disease can cause losses of up to 50% if no control measures are employed (Kushalappa; Eskes, 1989; Zambolim, 2016). The situation worsens in coffee-growing regions where there is still widespread use of susceptible cultivars. The continued utilization of this cultivar type primarily stems from the challenge of replacing coffee plants, considering the crop's prolonged growth cycle.

The evolution of coffee rust in crops is associated with a combination of three factors: the environment (climatic conditions), the host (coffee plant), and the pathogen, with interaction between them (Moraes, 1983). Climatic factors favorable to the disease are temperature in the range of 20°C to 25°C and humidity at levels suitable for spore germination (Pereira; Camargo; Camargo, 2008). Currently, the control of coffee rust is carried out in a total area with systemic protective fungicides on the leaves or soil application. In large coffee producing countries, such as Brazil, Colombia and Vietnam, agrochemical spray applications are calendar based, starting in November/December and continuing through April (Empresa Brasileira de Pesquisa Agropecuária - EMBRAPA, 1999). It is important to emphasize that the spatial and temporal distribution of the incidence of coffee rust is heterogeneous in the plantations, suggesting that the current strategy of controlling rust in the total area can be replaced by the management located in the foci of incidence of the disease (Alves et al., 2009).

The adoption of localized management techniques is a common practice in precision agriculture. With the advancement of geotechnologies and remote sensing, pest and disease management has become increasingly efficient and accurate, directing management only to the necessary places (Queiroz et al., 2022). This occurs because variations in reflectance of plants in specific regions of the spectrum can provide important information about senescence problems and plant stresses (Jensen, 2009). For this purpose, vegetation study methods are adopted for important measure characteristics of the crop, such as vegetative vigor, productive potential, and infestation of pests and diseases. These analyzes are performed using vegetation indices (VIs).

Succinctly, VIs are mathematical operations involving two or more bands, allowing spatial and temporal intercomparisons of photosynthetic activity of the canopy structure of the vegetation (Huete et al., 2002). The principle of using VI is that the reflected energies of specific bands in the electromagnetic spectrum are directly related to the photosynthetic activity of the canopy, as well as the assumption that the use of two or more spectral bands can substantially minimize the primary sources of noise that affect the response of vegetation (Ferreira; Ferreira; Ferreira, 2008).

The development of new specific vegetative indices for disease monitoring can fill gaps left by traditional generic indices, allowing for a more targeted and adaptable assessment of the specific characteristics of diseases that affect vegetation. For this detection, it is necessary to capture the reflectance of specific sensitive bands of the spectrum. The Multi-Spectral Instrument (MSI) sensor embedded in the Sentinel-2 satellite obtains spectral reflectance information from the Earth's surface by recording it in 13 different electromagnetic spectrum bands (European Space Agency - ESA, 2019). It is important to emphasize that five of these 13 bands are located in the near-infrared (NIR) spectral region. More specifically, three of these five are located in the zone of rapid reflectance growth when evaluating the plant canopy, known as Red-Edge. Chemura, Mutanga and Dube (2017) discriminated the severity of coffee rust on coffee leaves under greenhouse conditions based on the reflectance of Sentinel-2 satellite sensor bands. According to the authors, the bands located in the spectral position of the Red Edge can help detect diseases and evaluate the status of the coffee crop.

By determining the most sensitive bands for detecting a pest or disease, VIs can be developed specifically for this function. Recently, many methods have been used to select the most sensitive traits for pest and disease detection in crops (Domingues; Brandão; Pereira, 2022). Among the methods used, the Random Forest (RF) algorithm has performed this function well. This method makes it possible to rank variables based on their importance when performing a supervised classification, providing valuable information for building models and designing vegetation indices. Using the RF algorithm, Fletcher and Reddy (2016) demonstrated that shortwave-infrared bands were the most critical variables in discriminating the pigweeds (*Amaranthus viridis*) on soybean. Chemura, Mutanga and Dube (2017) used the importance ranking of variables to select the four most essential bands in the discrimination of coffee leaf rust on a canopy scale. The RF algorithm has shown satisfactory results in the scientific literature, helping to create models for detecting pests and diseases in crops.

The study and processing of bands in satellite images have been advancing over the years. Analyzing and creating VIs has been facilitated using Geographic Information Systems (GIS) software, like the popular ones QGIS and ArcGIS, and online platforms such as iSpatial and Google Earth Engine. In the development of geoprocessing works, it is essential to use one of these programs, as they provide the tools for analyzing satellite images and obtaining parameters of interest, thus enabling the achievement of the proposed objectives.

Among the software and platforms available, the current highlight is the recent platform launched in 2010, developed by Google, the Google Earth Engine (GEE). The GEE is a Google-hosted cloud-based computing platform that provides direct access to satellite imagery and geospatial datasets, including the entire Landsat catalog from EROS (USGS/NASA), MODIS, and Sentinel-2. In addition, it is possible to obtain updated climate data such as precipitation, altitude, and surface temperature from any part of the world. The great advantage of GEE is to provide an Application Programming Interface (API) enabling the development of algorithms in JavaScript or Python programming language, performing all the processing in the cloud. This fact allows analysis on a planetary scale, for any time scale, with operational capacity and speeds higher than conventional software. In addition, the GEE platform is currently free and open for developing non-profit research, making it an accessible tool for elaborating works in several areas.

Starting from the hypothesis that it is possible to detect changes in the spectral behavior of plants due to the presence of coffee rust through orbital remote sensing this work's objective was to develop a new VI for remote detection of coffee rust incidence on a regional scale using the GEE platform and the RF algorithm.

2 MATERIAL AND METHODS

2.1 Study Sites and data

The study was carried out in four experimental fields of the Agricultural Research Company of Minas Gerais (Epamig), located in the municipalities of Machado (EFMA), São Sebastião do Paraíso (EFSP), Patrocínio (EFPC) and Três Pontas (EFTP) in the state of Minas Gerais (MG). Data were collected over four consecutive years (2019, 2020, 2021 and 2022), generating a database used throughout the study. A rural property near the coffee region of Ribeirão de São Domingos-MG was selected to test the new index. In Figure 1, we can see the location of the municipalities considered in the study.

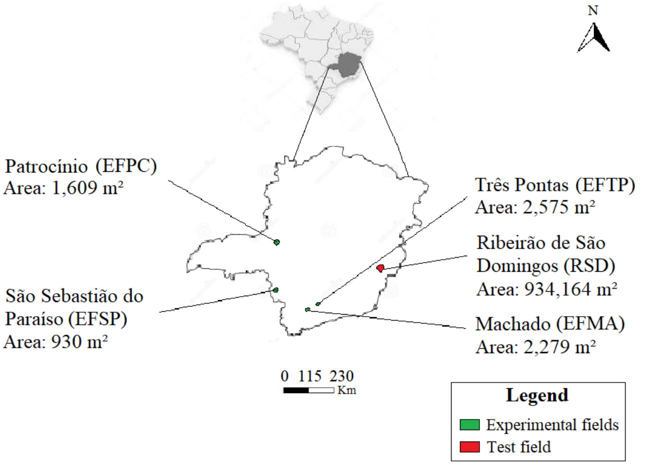

Figure 1: Location and area of the experimental fields and the local selected for test in the state of Minas Gerais. EFTP: experimental field of Três Pontas, EFSP: experimental field of São Sebastião do Paraíso, EFMA: experimental field of Machado, EFPC: experimental field of Patrocínio, and RSD: Test field of Ribeirão de São Domingos.

The database used in the study contains information about four consecutive years (2019, 2020, 2021 and 2022), with three categories of data, climatological data of the regions: average temperature and precipitation; phenological data of the plants: number of leaves, number of internodes and average foliage; and data on the infestation of pests and diseases: percentage of coffee rust (*Hemileia vastatrix*), leaf miner (*Leucoptera coffeella*) and cercosporiosis (*Mycosphaerella coffeicola*). The Data were recorded month by month over the three years.

The edaphoclimatic characteristics of experimental fields from 2019 to 2022 and the information on the cultivar planted in each stand are shown in Table 1. The rainfall data were obtained from the Climate Hazards Group InfraRed Precipitation with Stations – CHIRPS database (Funk et al., 2015) and temperature from the Latest Climate Reanalysis Produced by ECMWF / Copernicus Climate Change Service - ERA 5 database.

The observation and analysis of all the information available in the database were fundamental for elaborating the work, as it made it possible to direct the study to a single variable of interest, which in this case was the percentage of rust infestation in the stands. For this purpose, data from months with high precipitation (>150 mm) and a record of high infestation (>30%) of leaf miner and Cercosporiosis were excluded from the analysis.

Based on the database, the study sought a more suitable month for analysis. A month that provided data with different degrees of infestation and the presence of coffee rust was the most relevant factor among all other available data. The importance of choosing a month with a record of different degrees of infestation was to analyze the different spectral behavior of the stands. After searching the entire database, the most representative month chosen for the study was August 2021. The data recorded for the month can be seen in Table 2.

2.2 Obtaining and processing the satellite images

After defining the month and the data to be analyzed in the study, the Google Earth Engine (GEE) platform was used to process and obtain satellite images of the regions of interest. The use of GEE provided agility in all stages of the work due to cloud processing and the code creation tool. Most of the work was developed by elaborating an algorithm to perform the necessary analyses. All the steps described below were carried out by developing programming codes within the GEE platform in JavaScript.

Analyzing all the collections of satellite images available on the GEE platform was defined to use the Sentinel - 2A collection in the study due to its excellent spectral resolution of 13 bands and temporal resolution of five days, parameters considered adequate for the proposal of the work. Table 3 shows the number of bands recorded by this collection, their denominations, the central wavelength value, the width of the spectral bands, and their respective spatial resolutions.

For the elaboration of the study, a plot was selected in each experimental field with the following areas: EFTP (2,575 m²); EFSP (930 m²); EFMA (2,279 m²), and EFPC (1,609 m²). Considering that the spatial resolution of the Sentinel 2 A satellite is 10 meters for bands B2, B3, B4, and B8; and 20 meters for bands B5, B6, and B7, the plot areas were enough to meet the Sentinel 2A resolution.

Table 1: Edaphoclimatic characteristics and cultivar type of experimental fields

Characteristics	EFMA	EFSP	EFPC	EFTP
Cultivar	Catuaí 99	Catuaí 99	Rubi	Mundo Novo
Altitude (m)	970	880	997	916
Precipitation† (mm)	816	857	845	720
Temperature† (°C)	21.5	22	21.7	21.4
Geographic Coordinates	45⁰ 28' 59.452"W 21⁰ 20' 37.014"S	47⁰ 7' 20.341"W 20⁰ 54' 42.023"S	46⁰59'21.700"W 18⁰ 59'28.284"S	45⁰ 28' 59.452"W 21⁰ 20' 37.014"S

W: West, S: South.
†Annual average.

Table 2: Infestation data, monthly average temperature, and precipitation from experimental fields for August 2021

Characteristics	EFMA	EFSP	EFPC	EFTP
Coffee rust infestation (%)	100	46	24	0
Monthly average temperature (°C)	20.0	21.6	21.4	19.8
Precipitation (mm)	6.2	2.0	1.8	16.0

From the geographic coordinates of the experimental fields, satellite images, also called scenes, were searched for each region for the selected month of August 2021. Scenes from 08/17/2021 were selected for EFPC and EFSP, and scenes from 08/19/2021 for EFMA and EFTP. The scenes were chosen following criteria of less cloud coverage and being as close as possible to the data recorded in the field.

Subsequently, image processing was performed by applying a cloud filter through a cloud mask function within the GEE. After processing, the scenes of each experimental field were joined into a single image to obtain only one object of study for the development of further analyses. All the following steps were performed on this single image already processed.

2.3 Selection of sensitive bands for coffee rust identification

One of the study's main objectives was to identify the most influential bands for coffee rust detection for the development of the index. In the scientific literature, different methods have already been used for this purpose in different cultures. The most used methodologies are based on a combination of indexes, linear regression models, or Machine Learning algorithms, as can be seen in the works of Zheng, Huang, and Cui (2018), Liu et al. (2020), and Marin et al. (2021). This study identified the most influential bands for coffee rust detection based on the Machine Learning classifier algorithm – Random Forest, a method widely used for supervised classification and regression problems. According to Marin et al. (2021), RF is one of the most used and relevant tree-based algorithms since it can return a more complex model and solve non-linear tasks.

The Random forest (RF) is an ensemble of learning algorithms proposed by Breiman (Fletcher; Reddy, 2016). It consists of a set of independent, unpruned decision trees. According to Genuer, Poggi and Tuleau-Malot (2015), RF assumes an initial training set with "N" instances and each instance with "M" attributes. During the forest construction process, the algorithm performs in two aspects: (1) It samples a new training set with replacement at each iteration, and the new training set is the same size as the original set; (2) Rather than choosing the best split among all attributes, "m" attributes are randomly chosen from "M" at each node, and then these "m" attributes are used to split the node according to the principle of the decision tree algorithm, where "m" << "M", and it is held constant during the forest construction process.

Once the algorithm does not use all samples for model training at one time, it is possible to use the remaining samples (out-of-bag data) to evaluate the out-of-bag error (OOB error). Moreover, the principle of the feature importance ranking is to compare the difference in OOB error of each feature before and after adding noise to determine the importance of each feature. Thus, the importance of each feature is directly proportional to the calculated difference (Genuer; Poggi; Tuleau-Malot, 2015).

Table 3: Spectral bands and resolutions of the sentinel-2A sensor MSI.

Spectral Band		Central Wavelength (nm)	Bandwidth (nm)	Spatial Resolution (m)
B1	Coastal aerosol	443	20	60
B2	Blue†	490	65	10
B3	Green†	560	35	10
B4	Red†	665	30	10
B5	Red-edge 1 (Re1) †	705	15	20
B6	Red-edge 2 (Re2) †	740	15	20
B7	Red-edge 3 (Re3) †	783	20	20
B8	Near-infrared (NIR) †	842	115	10
B8A	Red-edge 4 (Re4) †	865	20	20
B9	Water vapor	945	20	60
B10	Shortwave infrared/cirrus	1375	30	60
B11	Shortwave infrared 1 (SWIR1) †	1610	90	20
B12	Shortwave infrared 2 (SWIR2) †	2190	180	20

† Bands used by the Random Forest algorithm.
Source: (ESA, 2019).

The supervised classification performed in the study was made by indicating the infested and healthy samples based on the database. The experimental field of the municipality of Machado (EFMA), which recorded 100% infestation in August 2021, was defined as the infested sample, and the experimental field of the municipality of Três Pontas (EFTP), which recorded 0 % infestation as of August 2021, as the healthy sample.

Based on this information, the algorithm performs a reflectance analysis of bands of each pixel at samples indicated as infested and healthy, using this information to classify other regions. The reflectance of the main bands captured by the Sentinel-2A satellite collection was defined as the base parameter of the classification, namely: B2 (Blue), B3 (Green), B4 (Red), B5 (Red Edge1), B6 (Red Edge 2), B7 (Red Edge 3), B8 (NIR), B8a (Red Edge 4), B11 (SWIR 1) and B12 (SWIR2). The bands: B1 (Coastal aerosol), B9 (Water vapor), and B10 (Shortinfrared/cirrus) were not considered, according to Chemura, Mutanga and Dube (2017), due to their low significance in studies of this nature.

The methodology developed in this study focused solely on employing the Random Forest model to identify and rank the most important variables used to classify infested and healthy plots. For this, only the model training phase was perfomed. Therefore, as it is not common to evaluate only the training stage, no conventional results of machine learning performance metrics will be presented.

After running the algorithm, a ranking of the most relevant bands used for classification was generated. As explained, this importance is measured by comparing the out bag error. In this way, each band's influence on the algorithm appointment was quantified.

2.4 Development of the new index

With the information of three relevant bands to discriminate a particular feature and the central wavelengths of each one, it was possible to create a new index using the Triangular Vegetation Index (TVI) method described by Broge and Leblanc (2001). The TVI method is based on calculating the area of a triangle on the multispectral graph with vertices in the three selected bands. The area of this triangle is calculated by the determinant of a matrix composed of the wavelengths of the bands and their respective central wavelengths, as can be seen in equations 1 and 2.

$$\begin{bmatrix} B_1 & \cdots & C_1 & \cdots & 1 \\ \vdots & \ddots & \vdots & & \vdots \\ B_2 & \cdots & C_2 & \cdots & 1 \\ \vdots & & \vdots & \ddots & \vdots \\ B_3 & \cdots & C_3 & \cdots & 1 \end{bmatrix} \tag{1}$$

$$Area_{Tri} = \frac{(C_3 - C_1)\times(B_2 - B_1) - (C_2 - C_1)\times(B_3 - B_1)}{2} \tag{2}$$

where: B1, B2, and B3 are the recorded reflectance values of the selected bands; and C1, C2, and C3 are the central wavelengths of each band, respectively. The central wavelength values are fixed values obtained from the official Sentinel 2-A mission website (Table 3).

2.5 Models comparison with other vegetative indices

According to the literature, several vegetation indices (VIs) have been used to identify crop diseases. Eight indices commonly used for vegetation studies were selected in the study for comparison with the new index created. The aim was to analyze its ability to detect coffee rust in stands. The VIs chosen for comparison were NDVI, NDVIre1, GNDVI, NREDI1, NREDI2, NREDI3, EVI, and SR, their definitions, formulas, and references can be seen in Table 4.

It was then developed for each index, including for the new index created, linear regression equations using as parameters the average value of the index in the stands and the rust infestation registered in the field. Thus, the value of rust infestation estimated for each VIs was obtained, making it possible to compare with the real infestation value measured in the field.

The regression models created for each VI were compared using the performance criteria: coefficient of determination (R^2) and root mean error square (RMSE). These parameters were calculated on the statistical software "R". It was used for these analyses satellite Images from August 2019, 2020, 2021 and 2022, and the infestation data recorded in each experimental field available in the study (Table 5).

Data that recorded infestation greater than or equal to 90% were removed from this analysis to avoid the possible effects of saturation of the vegetative indices, which is very common in this kind of study. This phenomenon occurs when the vegetative index reaches a maximum invariant value, even with differences in vegetation characteristics (Ponzoni et al., 2012). In addition, there was no record in the database for August 2019 in EFPC, data that could not be used for analysis.

2.6 Monitoring coffee rust on a regional scale

Finally, an appropriate value of infestation percentage was sought in the scientific literature in which it would be great to issue a phytosanitary alert to the producer. According to the manual of recommendations for coffee cultivation in Brazil, infection percentages above 40% already indicate significant losses in production due to defoliation caused by coffee rust (Matiello et al., 2010). Thus, the value of infestation percentage equal to or greater than 40% was defined in the work as an alert parameter for monitoring coffee rust on a regional scale.

Table 4. Indices commonly used for vegetation studies

Names	Formula	Reference
Normalized Difference Vegetation Index (NDVI)	$\frac{(NIR-RED)}{(NIR+RED)}$	Rouse et al. (1974)
Normalized difference vegetation index red-edge1 (NDVIre1)	$\frac{NIR-RE1}{NIR+RE1}$	Rouse et al. (1974)
Green normalized difference vegetation index (GNDVI)	$\frac{NIR-GREEN}{NIR+GREEN}$	Gitelson and Merzlyak (1998)
Normalized red-edge1 index (NREDI1)	$\frac{RE2-RE1}{RE2+RE1}$	Gitelson and Merzlyak (1994)
Normalized red-edge2 index (NREDI2)	$\frac{RE3-RE1}{RE3+RE1}$	Gitelson and Merzlyak (1994)
Normalized red-edge3 index (NREDI3)	$\frac{RE3-RE2}{RE3+RE2}$	Gitelson and Merzlyak (1994)
Enhanced Vegetation Index (EVI)	$\frac{2,5*(NIR-RED)}{(NIR+6*RED-0,5*BLUE+1)}$	Justice et al. (1998)
Simple ratio (SR)	$\frac{NIR}{RED}$	Jordan (1969)

Table 5: Coffee rust infestation data from the experimental fields in August 2019, 2020, 2021 and 2022

Coffee rust infestation (%)	EFMA	EFSP	EFPC	EFTP
August 2019	91†	12	x†	29
August 2020	95†	10	50	4
August 2021	100†	24	46	0
August 2022	49	14	11	12

† Data excluded from the analysis.

Using the equation obtained in the linear regression model that estimates the rust infestation based on the new index, a value representing the critical level of 40% infestation was determined. In this way, the remote detection of coffee rust proposed in work was done by elaborating a map indicating the infested areas (>40%). To determine if the pixel is considered infested or healthy, the new index value calculated for that pixel is compared with the adopted critical value. The pixel is considered infested when the calculated value is equal to or greater than the critical value. If lower, the pixel is considered healthy.

It was proposed in the study the monitoring using the new index of coffee rust in a different region from the experimental fields studied. The intent of testing in a different location is to validate the application of the index to any location in the state. Thus, a rural property in the coffee region to the east of the state was chosen, located near Ribeirão de São Domingos - MG, geographical coordinates: 42º 20' 46.9854" O; 20º 27' 29.4804" W (Figure 1). Since the main database used for the development of the new index was related to the month of August, it was decided to carry out this regional monitoring analysis also for the month of August, but for the year 2022.

The following steps were carried out in the GEE to perform the monitoring: obtaining all satellite images from the Sentinel – 2A collection of the study sites, available in August 2022; processing the images through a cloud filter function; selection of the most representative satellite image of the month using the median function, processing that sorts the collection of images obtained and selects the median value of each pixel in the image, selecting then the best pixel among all available images; application of the new index in the processed image for slicing regions based on the adopted critical infestation value. Finally, these processes resulted in a map of the rural property with the indication of possible healthy regions and regions with a degree of infestation equal to or greater than 40%.

Regrettably, due to practical constraints, it was not possible to carry out a numerical validation of the results, relying solely on visual assessments. However, it is important to remember that the inclusion of this topic in the work was not with the intention of validation but rather providing insights into the application of the developed index. The intention is that future studies will be carried out with this aim.

3 RESULTS

3.1 The Spectral characteristic of the fields

After processing the satellite images of the experimental fields on the 17th and 19th of August 2021, the spectral characteristics of each field were analyzed by elaborating multispectral graphs. The result of the analysis can be seen in Figure 2.

It is possible to observe similar reflectance values from all fields in bands B1 to B5. The multispectral graphs show considerable differences from bands B6 to B9. The bands B11 and B12 did not show any noticeable pattern. As expected, it is possible to observe that the significant differences in reflectances are between the least infested sample (EFTP) and the most infested sample (EFMA), as highlighted in Figure 2.

3.2 Supervised Classification using Random Forest

The RF algorithm was used to perform the supervised classification of satellite images from the experimental fields. Figure 3 displays the points sampled in each plot used for model training. There were a total of 12 points in the infested sample and 14 in the healthy sample. The background image in the figure does not depict the satellite image used for analysis; instead, a higher-resolution base map was chosen to enhance visualization.

Thus, a model that estimated the possible healthy areas and areas with rust infestation was trained. The classification was based on the reflectance of the ten selected bands from the Sentinel 2A satellite collection (Table 3), resulting in the following ranking of bands based on their importance for coffee rust detection (Figure 4).

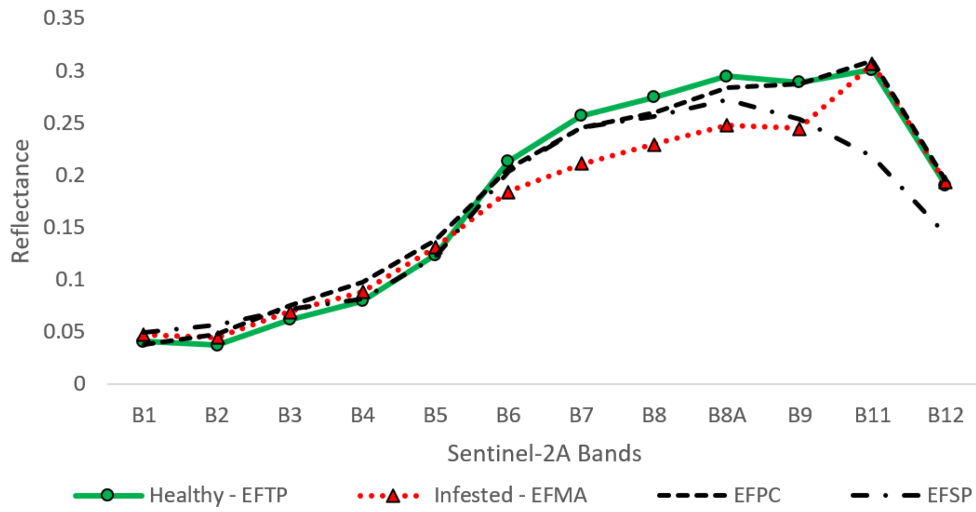

Figure 2: Multispectral reflectance graph of the experimental fields located in the municipalities of Três Pontas (EFTP), Machado (EFMA), São Sebastião do Paraíso (EFSP), and Patrocínio (EFPC) for August 2021.

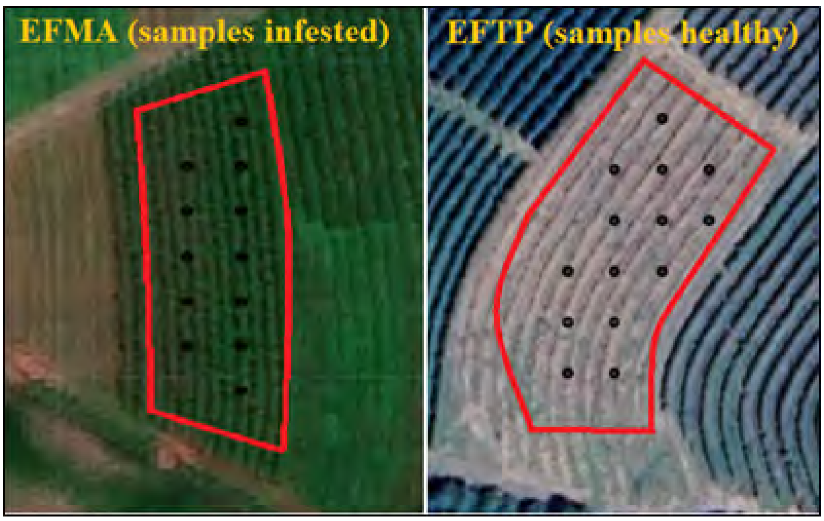

Figure 3: Location of sampling points in the EFMA and EFTP plots that were used for model training.

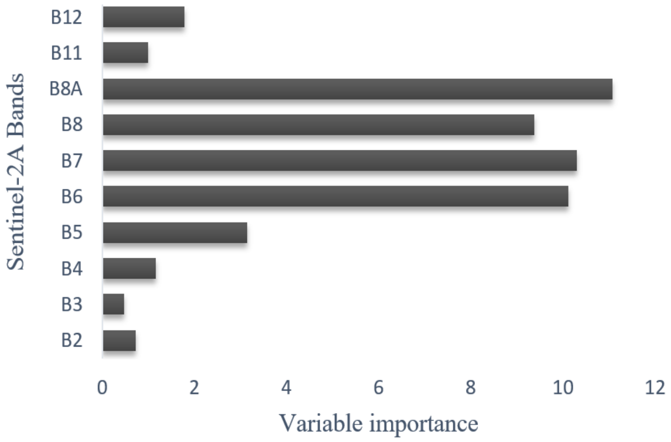

Figure 4. Ranking of Sentinel-2A bands based on their importance for coffee rust detection through RF algorithm.

It is possible to observe the great relevance of the bands in the Red-Edge spectrum (B5, B6, B7, and B8a) on coffee rust detection. Surprisingly, the B8 band (NIR), traditionally used for vegetation studies, came in fourth place in the ranking of importance. The B4 band (Red) had the highest sensitivity for coffee rust detection at the visible spectrum.

3.3 The Conception of the new index

To obtain a more representative triangle area, a modification of the Triangular Vegetation Index method was proposed. Instead of selecting the three most important bands identified in the ranking, only the two most important were selected: B8A (Red Edge 4) and B7 (Red Edge 3); the third selected band was the most important band in the visible spectrum: B4 (Red). It is understood in the present study that the proposed modification will improve the performance of the new index because it increases the calculated difference obtained between a healthy and a diseased plant, allowing better discrimination of the coffee rust infestation.

So, the triangles' vertices were established in the selected bands B4, B7, and B8A, in the multispectral graph (Figure 5), as indicated by the TVI method.

Thus, the creation of a new index named "Coffee Rust Detection Index" (CRDI) was proposed, in which its value represents the area of the triangle of vertices in the bands B4, B7, and B8A. This area is calculated from the determinant of the matrix according to the TVI method (Equations 1 and 2). In such a manner, the new index is determined as shown in Equation 3:

$$CRDI = \left| \frac{(665-865)\times(B7-B8A)-(783-865)\times(B4-B8A)}{2} \right| \quad (3)$$

where: B4, B7, and B8A are the reflectance values of bands of a particular pixel in the image; 665, 783, and 865 are the center wavelengths of each band, respectively.

3.4 Comparison with other vegetative indices

Using satellite images and data from August 2019, 2020, 2021 and 2022 (Table 5), a model was created using the infestation percentage values estimated by the CRDI index and the actual infestation values measured in the field recorded in the study database. This analysis resulted in a coefficient of determination (R^2) of 0.6321 and a root-mean-square error (RMSE) of 0.103, as shown in Figure 6.

The same analysis was done with the others eight VIs presented in Table 4, thus making it possible to compare the performance in detecting rust infestation of the new CRDI index with the other relevant VIs in the literature. The result of this analysis can be seen in Table 6.

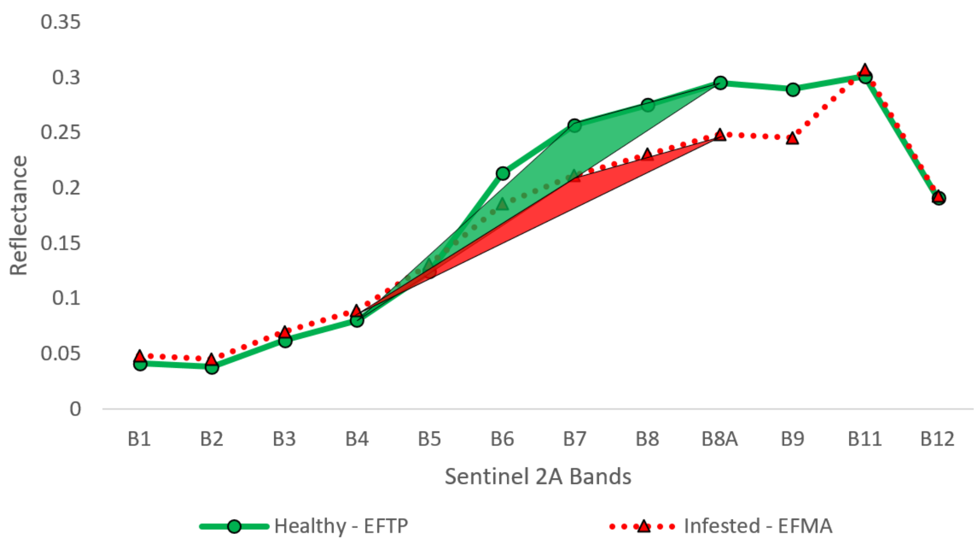

Figure 5: Triangles in the multispectral graphs of the EFTP and EFMA experimental fields with vertices in the three most sensitive bands in coffee rust detection (B4, B7, and B8A).

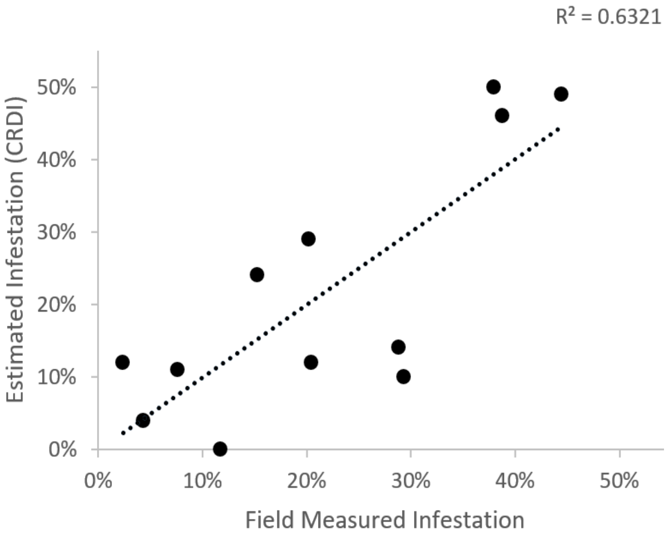

Figure 6: Comparison between the model created to estimate coffee rust infestation based on the CRDI and data recorded in the field.

Table 6: R^2 and RMSE of models that estimate coffee rust infestation, based on vegetative indices, for August from 2019 to 2021.

Vegetative Index	R^2	RMSE
CRDI	0.632	0.103
NDVI	0.373	0.139
NDVIre1	0.395	0.136
GNDVI	0.472	0.128
NREDI1	0.448	0.130
NREDI2	0.406	0.135
NREDI3	0.063	0.170
EVI	0.447	0.131
SR	0.325	0.144

As we can observe in Table 6, the new CRDI index model had the best performance among the others since it had the highest R^2 value and the lowest RMSE value, followed by the GNDVI and later the NREDI1. The worst performance was presented by the model of the NREDI3 index. Considering that the same methods and data were used for the analysis, we can assume that the new CRDI index is a good option for detecting coffee rust on a regional scale.

3.5 Using the new index for detecting coffee rust on a regional scale

According to the equation obtained in the linear regression model (Figure 6), the numerical value of the CRDI that represents 40% of coffee rust infestation was 3.942. Thus, to exemplify the application of the CRDI, an analysis for the detection of coffee rust infestation was proposed in a coffee plantation near the city of Ribeirão de São Domingos - MG.

In Figure 7, it is possible to observe the location of the plantation and the representative satellite image obtained for August 2022. In Figure 8, we can observe the application of the CRDI index for the study site. Finally, in Figure 9, we can see the final objective of the analysis, the map of coffee rust detection for values greater than 40% of an infestation.

The producer's visual assessment indicated that the generated map successfully detected coffee rust outbreaks in the plots. However, it's recognized that a numerical assessment is essential to validate this observation.

4 DISCUSSION

The results obtained in the study highlight the usefulness and potential of the Sentinel 2 – MSI sensor characteristics for discriminating disease infections in commercial crops such as coffee, a previously challenging task using the multispectral broadband sensors, for example, from the Landsat series.

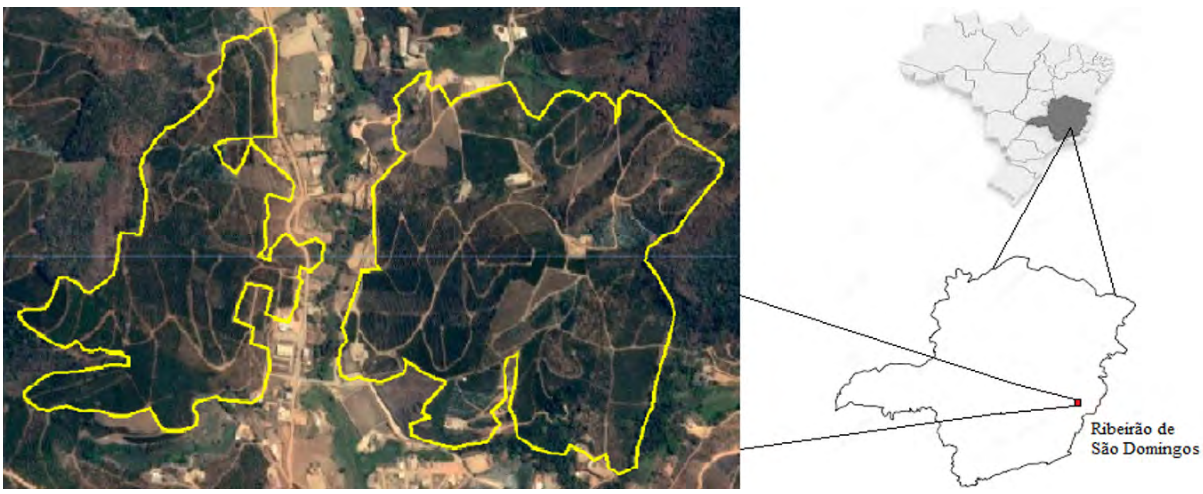

Figure 7: Satellite images from the Sentinel-2 collection for August 2022 of the coffee plantation located near the city of Ribeirão de São Domingos – MG.

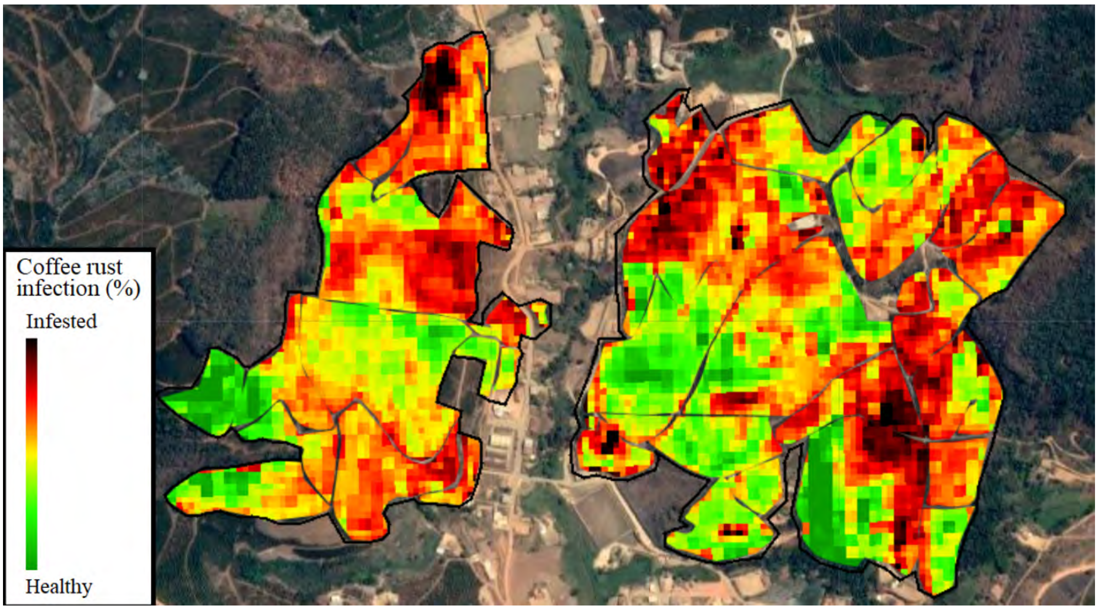

Figure 8: Utilization of the CRDI index in the satellite image of the Sentinel 2A collection for a coffee plantation near the city of Ribeirão de São Domingos - MG.

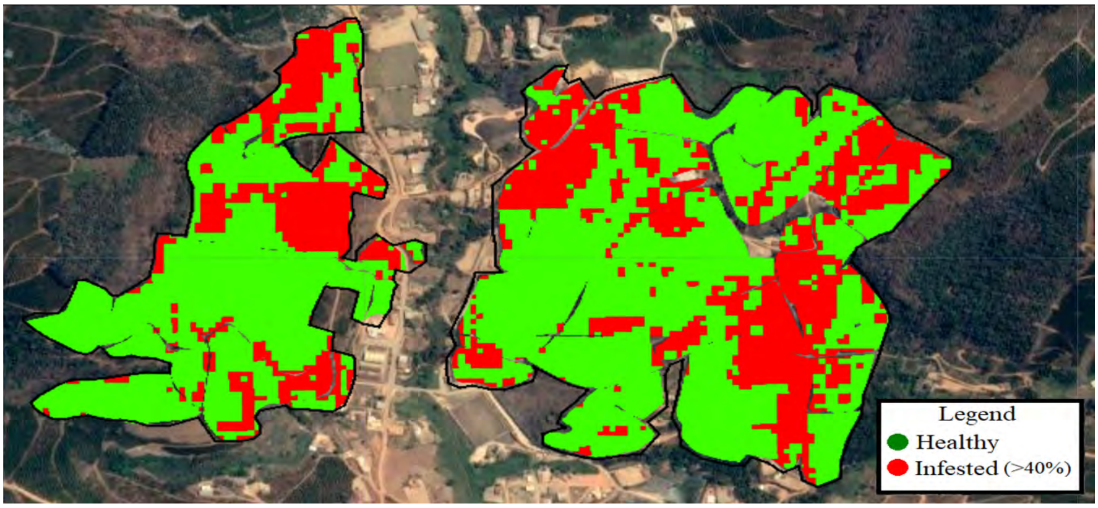

Figure 9: Mapping of the detection of critical levels of coffee rust infestation (>40%) based on the CRDI index, in August 2022, for a coffee plantation near the city of Ribeirão de São Domingos - MG.

That reflectance differences point to a reasonable possibility of discriminating the infestation through satellite images using vegetative indices, considering the capacity of VIs to highlight reflectance characteristics of specific bands in the spectrum, allowing comparisons between patterns in the multispectral graph (Rumpf et al., 2010). Thus, by identifying sensitive regions for a particular trait (infested or healthy), it was possible to develop a specific vegetative index to discriminate this aspect accurately.

The presence of stress factors alters the properties of plants, which in turn influences the radiation emitted across the spectrum. Alteration in the spectrum is produced mainly by changes in leaf water content (Mottram et al., 1983; Pinter et al., 1979), and this can also be detected in the early stage of coffee rust (Chaerle et al., 1999; Costa; Grant; Chaves, 2013; Omasa, 1990). The different reflectance patterns observed in the samples can be adequately justified by the physiological symptoms that coffee rust causes in plants.

The multispectral analysis carried out in the study confirms the general effect of vegetation under different types of biotic stresses of increases in reflectance in the red edge region. The authors Chemura, Mutanga and Dube (2017) and Mahlein et al. (2013), when comparing the reflectances of diseased plants with healthy plants, also observed more significant differences in the region of the red edge in the electromagnetic spectrum in their studies.

The result of the selection of the most sensitive bands for coffee rust detection (B4, B7, B8A) using the RF algorithm follows the study carried out by Chemura, Mutanga and Dube (2017), which sought to discriminate the levels of rust infection in coffee leaves using Machine Learning methods at Sentinel-2 MSI spectral resolutions. Two bands selected in the current work (B7 and B4) were identified in common as the most sensitive bands for detecting coffee rust, indicating a similar reflectance pattern between the healthy and infested plants studied. The divergent result of selecting the third most sensitive band in detecting rust infestation can be justified by the scale adopted in each study. Chemura, Mutanga and Dube (2017) performed the reflectance analyses in greenhouses with a controlled environment, while in the present work, the reflectance analyses were performed on coffee plantations in the field.

Regarding the TVI method, as seen in Figure 5, different area values were obtained using the selected bands B4, B7, and B8A as vertices of the triangle. This fact reinforces the ability to distinguish between a healthy and an infested plant of the new index, as its value represents precisely the area of the triangle in the multispectral graph. As seen in Figure 5, a healthier region will have a larger triangle area and, consequently, a higher CRDI index value. The opposite reasoning will also be valid; a more infested region will have a smaller triangle area and, consequently, a lower CRDI index value. This direct relationship is due to the reflectance pattern of the bands that compose the new index and the reflectance that plant infested by rust present.

The comparison of the CRDI index with the other eight vegetative indices showed good performance of the model for detecting coffee rust infestation as seen in Table 6. The CRDI model obtained the best performance in the statistical criteria R^2 and RMSE, reaching values of 0.631 and 0.103, respectively, followed by the GNDVI and NREDI1. The consistent performance also achieved by these two other indices can be justified by the bands involved in their calculations (NIR, Re1, and Re2), also identified in the study as relevant in the discrimination of coffee rust infestation.

The use of the new index to detect critical levels of coffee rust infestation can be a valuable tool for controlling the disease in large coffee plantations. The generated map provides essential information on the disease progression overview in the region, which can be used to issue phytosanitary alerts. In addition, the monitoring through the CRDI index can also be carried out month by month over the years, allowing the development of studies to characterize the incidence of coffee rust in any region of the world.

Overall, the incidence of any coffee pest or disease considerably affects the spectral reflectance of the crop. Its crucial to emphasize that the analysis conducted in this research attempted to focus on coffee rust, by excluding data with high infestation records (>50%) of leaf miner and cercosporiosis. However, this filtering approach also brings practical limitations to the work, considering that this situation frequently occurs in the field.

5 CONCLUSION

Based on the conducted study, three bands of the Sentinel 2A sensor (Red, Red-edge 3, and Red-edge 4) have been identified as efficient for monitoring coffee rust infestation. The findings suggest that the newly developed vegetative index (CRDI) could serve as a crucial tool for monitoring this disease in coffee plants. Moreover, further studies using the CRDI must be carried out to validate the index and support global coffee rust control efforts. Additionally, the methodology employed in this research has the potential to be extended to other types of crops and diseases, offering improved monitoring strategies for this issue that impacts multiple crops.

6 AUTHORS' CONTRIBUTION

GDMC wrote the manuscript and performed the experiment, EFV supervised the experiment and co-work the manuscript, RAS and ALRF review and approved the final version of the work, WPMF conducted all statistical analyses.

7 REFERENCES

ALVES, M. C. et al. Modeling spatial variability and pattern of rust and brown eye spot in coffee agroecosystem. **Journal of pest science**, 82(2):137-148, 2009.

BROGE, N. H.; LEBLANC, E. Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. **Remote sensing of environment**, 76(2):156-172, 2001.

CHAERLE, L. et al. Presymptomatic visualization of plant–virus interactions by thermography. **Nature biotechnology**, 17(8):813-816, 1999.

CHEMURA, A.; MUTANGA, O.; DUBE, T. Separability of coffee leaf rust infection levels with machine learning methods at Sentinel-2 MSI spectral resolutions. **Precision Agriculture**, 18(5):859-881, 2017.

COSTA, J. M.; GRANT, O. M.; CHAVES, M. M. Thermography to explore plant: Environment interactions. **Journal of Experimental Botany**, 64(13):3937-3949, 2013.

DOMINGUES, T.; BRANDÃO, T.; FERREIRA, J. C. Machine learning for detection and prediction of crop diseases and pests: A comprehensive survey. **Agriculture**, 12(9):1350, 2022.

EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA - EMBRAPA. **Controle da ferrugem o cafeeiro (*Hemileia vastatrix*) no estado de Rondônia**. 1999. Available in: <https://ainfo.cnptia.embrapa.br/digital/bitstream/item/76700/1/FOL-7730-0001.pdf>. Access in: 18 December 2023.

EUROPEAN SPACE AGENCY - ESA. **Sentinel-2 user handbook**. 2019. Available in: <https://sentinel.esa.int/documents/247904/685211/Sentinel-2_User_Handbook>. Access in: 18 December 2023.

FERREIRA, L. G.; FERREIRA, N. C.; FERREIRA, M. E. Sensoriamento remoto da vegetação: Evolução e estado da arte. **Acta Scientiarum. Biological Sciences**, 30(4):379-390, 2008.

FLETCHER, R. S.; REDDY, K. N. Random forest and leaf multispectral reflectance data to differentiate three soybean varieties from two pigweeds. **Computers and Electronics in Agriculture**, 128:199-206, 2016.

FUNK, C. et al. The climate hazards infrared precipitation with stations: A new environmental record for monitoring extremes. **Scientific data**, 2(1):1-21, 2015.

GITELSON, A.; MERZLYAK, M. N. Quantitative estimation of chlorophyll-a using reflectance spectra: Experiments with autumn chestnut and maple leaves. **Journal of Photochemistry and Photobiology B: Biology**, 22(3):247-252, 1994.

GENUER, R.; POGGI, J.-M.; TULEAU-MALOT, C. VSURF: An R package for variable selection using random forests. **The R Journal**, 7(2):19-33, 2015.

GITELSON, A.; MERZLYAK, M. N. Spectral reflectance changes associated with autumn senescence of *Aesculus hippocastanum L.* and *Acer platanoides L.* leaves. Spectral features and relation to chlorophyll estimation. **Journal of plant physiology**, 143(3):286-292, 1994.

GITELSON, A. A.; MERZLYAK, M. N. Remote sensing of chlorophyll concentration in higher plant leaves. **Advances in Space Research**, 22(5):689-692, 1998.

HUETE, A. et al. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. **Remote sensing of environment**, 83(1-2):195-213, 2002.

JORDAN, C. F. Derivation of leaf- area index from quality of light on the forest floor. **Ecology**, 50(4):663-666, 1969.

JUSTICE, C. O. et al. The moderate resolution imaging spectroradiometer (MODIS): Land remote sensing for global change research. **IEEE transactions on geoscience and remote sensing**, 36(4):1228-1249, 1998.

JENSEN, J. R. **Sensoriamento remoto do ambiente**: Uma perspectiva em recursos terrestres. Tradução José Carlos Neves Epiphanio et al., São José dos Campos/SP, 2009. 40p.

KUSHALAPPA A. C.; ESKES A. B. **Coffee rust**: Epidemiology, resistance, and management. CRC Press, Boca Raton. p. 171-292, 1989. 54p.

LIU, L. et al. A disease index for efficiently detecting wheat fusarium head blight using sentinel-2 multispectral imagery. **IEEE Access**, 8:52181-52191, 2020.

MAHLEIN, A.-K. et al. Development of spectral indices for detecting and identifying plant diseases. **Remote Sensing of Environment**, 128:21-30, 2013.

MARIN, D. B. et al. Detecting coffee leaf rust with UAV-based vegetation indices and decision tree machine learning models. **Computers and Electronics in Agriculture**, 190:106476, 2021.

MATIELLO, J. B. et al. **Cultura de café no Brasil**: Manual de recomendações. Rio de Janeiro: MAPA/Procafé, 2010.

MORAES, S. A. **A ferrugem do cafeeiro**: Importância, condições predisponentes, evolução e situação no Brasil. Campinas: Instituto Agronômico, 1983. 50p.

MOTTRAM, R.; DE JAGER, J. M.; DUCKWORTH, J. R. Evaluation of a water stress index for maize using an infra-red thermometer. **Gewasproduksie Crop production**, 1983.

OMASA, K. Image instrumentation methods of plant analysis. **Physical methods in plant sciences**, 1990. 243p.

PEREIRA, A. R.; CAMARGO, A. P.; CAMARGO, M. P. **Agrometeorologia de cafezais no Brasil**. Campinas: Instituto Agronômico, 2008. 127p.

PONZONI, F. J.; SHIMABUKURO, Y. E.; KUPLICH, T. M **Sensoriamento remoto no estudo da vegetação**. 2. ed. São Paulo: 176p. 2012.

POZZA, E. A.; CARVALHO, V. L.; CHALFOUN, S. M. Sintomas de injúrias causadas por doenças em cafeeiro. *In*: GUIMARÃES, R. J., MENDES, A. N.; BALIZA, D. P. **Semiologia do cafeeiro:** Sintomas de desordens nutricionais, fitossanitárias e fisiológicas. Lavras: Editora UFLA, p. 68-106, 2010.

PINTER JR, P. J. et al. Remote detection of biological stresses in plants with infrared thermometry. **Science**, 205(4406):585-587, 1979.

QUEIROZ, D. M. et al. **Agricultura digital**. 2 ed. São Paulo: Oficina de Textos, 2022. 224p.

ROUSE, J. W. et al. Monitoring vegetation systems in the great plains with ERTS proceeding. **Third Earth Reserves Technology Satellite Symposium**, Greenbelt: NASA SP-351. 1974.

RUMPF, T. et al. Early detection and classification of plant diseases with support vector machines based on hyperspectral reflectance. **Computers and electronics in agriculture**, 74(1):91-99, 2010.

UNITED STATES DEPARTMENT OF AGRICULTURE - USDA. **Coffee**: World markets and trade. United States Department of Agriculture: Washington, DC, USA, 2023. Available in: https://apps.fas.usda.gov/psdonline/circulars/coffee.pdf. Access in: 21 December 2023.

ZAMBOLIM, L. Current status and management of coffee leaf rust in Brazil. **Tropical Plant Pathology**, 41(1):1-8, 2016.

ZHENG, Q.; HUANG, W.; CUI, X. New spectral index for detecting wheat yellow rust using sentinel-2 multispectral imagery. **Sensors**, 18(3):868, 2018.