Int J Appl Earth Obs Geoinformation 85 (2020) 101983

Contents lists available at ScienceDirect

Int J Appl Earth Obs Geoinformation

journal homepage: www.elsevier.com/locate/jag

Detection of coffee berry necrosis by digital image processing of landsat 8 oli satellite imagery

Jonathan da Rocha Miranda[a,*], Marcelo de Carvalho Alves[b], Edson Ampélio Pozza[c], Helon Santos Neto[c]

[a] Agricultural Engineering Department, Federal University of Lavras, University Campus, P.O.Box 3037, 37200-000, Lavras, Minas Gerais, Brazil
[b] Department of Agricultural Engineering at the Federal University of Lavras, University Campus, P.O.Box 3037, 37200-000, Lavras, Minas Gerais, Brazil
[c] Plant Pathology Department, Federal University of Lavras, University Campus, P.O.Box 3037, 37200-000, Lavras, Minas Gerais, Brazil

ARTICLE INFO

Keywords:
Data mining
Spectral behavior
Accuracy
Colletotrichum ssp.
Atmospheric correction

ABSTRACT

Coffee berry necrosis is a fungal disease that, at a high level, significantly affects coffee productivity. With the advent of surface mapping satellites, it was possible to obtain information about the spectral signature of the crop on a time scale pertinent to the monitoring and detection of plant phenological changes. The objective of this paper was to define the best machine learning algorithm that is able to classify the incidence CBN as a function of Landsat 8 OLI images in different atmospheric correction methods. Landsat 8 OLI images were acquired at the dates closest to sampling anthracnose field data at three times corresponding to grain filling period and were submitted to atmospheric corrections by DOS, ATCOR, and 6SV methods. The images classified by the algorithms of machine learning, Random Forest, Multilayer Perceptron and Naive Bayes were tested 30 times in random sampling. Given the overall accuracy of each test, the algorithms were evaluated using the Friedman and Nemenyi tests to identify the statistical difference in the treatments. The obtained results indicated that the overall accuracy and the balanced accuracy index were on an average around 0.55 and 0.45, respectively, for the Naive Bayes and Multilayer Perceptron algorithms in the ATCOR atmospheric correction. According to the Friedman and Nemenyi tests, both algorithms were defined as the best classifiers. These results demonstrate that Landsat 8 OLI images were able to identify an incidence of the coffee berry necrosis by means of machine learning techniques, a fact that cannot be observed by the Pearson correlation.

1. Introduction

Coffee farming has always played a prominent role in Brazilian commodities. In the 2017/2018 harvesting, Brazilian coffee production accounted for about 32.4% of the world market for *in natura* coffee (IOC, 2018). The technologies employed from planting to commercialization are being increasingly demanded mainly with the advent of precision agriculture, which can provide in gaining productivity.

Within the mechanisms adopted in crop management, knowledge of tools that monitor pests and diseases is essential. Regarding coffee diseases, the Coffee Berry Necrosis – CBN – is one of the main coffee diseases, since it has a direct action on coffee productivity. CBN, in most cases, is related to fungi of the genus *Colletotrichum*, which has reported a reduction of up to 80% in productivity (Griffiths et al., 1971; Varzea et al., 2002).

Although the fungus acts on fruits, the presence of *Colletotrichum* spp. in coffee branches cause changes in the normal stem and leaf structure due to physiological disorders that are associated with disease onset. Among them, the most common are high pending fruit loads, nutritional deficiency, physical and chemical impediments in the soil (Paradela Filho, 2001). Sera et al. (2005), observed a negative correlation between the increased incidence of *Colletotrichum* ssp. and the vegetative vigor, which was evaluated in the visual perception of the plant, observing the leaf tone and branch dryness.

The incidence of the disease can change the density of the canopy and the leaf area, factors that can be identified by spectral signature mainly in the infrared region (Franke and Menz, 2007). The combination of different wavelengths may be able to detect diseases by multispectral sensors, given that the disease signals may influence peculiarly the spectral signature of the target (Mahlein et al., 2013). Therefore, multispectral satellite imagery can aid the detection and control of the pathogen due to its ability to infer in physiological aspects of plants

* Corresponding author.
E-mail addresses: jhonerocha@estudante.ufla.br (J.d.R. Miranda), marcelo.alves@ufla.br (M.d.C. Alves), eapozza@ufla.br (E.A. Pozza), helonsantosneto@gmail.com (H. Santos Neto).

https://doi.org/10.1016/j.jag.2019.101983
Received 28 June 2019; Received in revised form 21 September 2019; Accepted 1 October 2019
Available online 09 November 2019

(Lopresti et al., 2015).

The use of orbital images for detection, quantification and classification of coffee diseases has been used and improved over time (Chemura et al., 2018a, 2017; Price et al., 1993; Tucker et al., 2013). Accurate and reliable detection of diseases is facilitated by highly sophisticated and innovative methods of data analysis that lead to new insights derived from sensor data for complex plant-pathogen systems (Mahlein, 2016).

Machine learning algorithms make no assumptions about frequency distribution and are becoming increasingly popular to classify remote sensing data, which rarely have normal distributions (Belgiu and Drăgu, 2016). It is estimated that the techniques of machine learning can find a classifier capable of identifying the incidence of coffee berry necrosis based on Landsat 8 OLI images, once the relation of the spectral signature of the coffee canopy under the effect of coffee berry necrosis incidence is known.

This paper aimed to define which methods of atmospheric correction combined with machine learning techniques can approximate the process of evaluating the disease in the field data.

2. Material and methods

The analyzed area is located in the southern region of Minas Gerais, in a coffee crop, in the municipality of Carmo do Rio Claro, centered at coordinates of latitude 21°00′28″ South of Ecuador and longitude 46°01′30″ West of Greenwich (Fig. 1). The planting of coffee (Coffea arabica L.) cultivar Acaiá 474/19 was arranged with a spacing of 3.6 m between rows and 0.70 m between plants in a total area of 11 ha. The crop was irrigated through drip irrigation with management based on the water demand, measured through properly installed tensiometer batteries.

2.1. Berry necrosis assessment in the field

The sample mesh was composed of georeferenced points in a spacing of 40 to 40 m measured in the field with a GPS TRIMBLE 4600 LS ® and Total Station Leica TC600 ® (Fig. 1). In each georeferenced point five plants were assessed, and in each plant two branches in the third middle of the plant canopy were randomly defined and marked using a wire in order to assess the same branch in all evaluated periods. The percentage of reproductive nodes with diseased berries in relation to reproductive nodes of the branch was evaluated (Santos Neto, 2017).

During disease assessment, necrotic berry samples were collected and sent to the diagnostic and control laboratory of plant diseases of the Federal University of Lavras (UFLA), where it was possible to observe the presence of the fungus *Colletotrichum* ssp. The isolates were morphologically characterized as belonging to *Colletotrichum gloeosporioides* species complex and had its pathogenicity proved on coffee berries.

The assessment of the berries disease was carried out in three periods (December, 15, 2013; January, 18 and February, 26, 2014), corresponding to the period of grain development according to the scale of evaluation of phenological stages proposed by Pezzopane et al. (2003). All assessments were considered over the months in a single sample, which allowed to separate the intensity of coffee berry necrosis in 4 bands of classes defined by the quartile (0–25, 25–50, 50–75, 75–100% of the total number of samples), which corresponded to 32 samples per established class.

2.2. Constitution of the sample mesh to remote sensing analysis

The sample mesh was composed of georeferenced points in a spacing of 40 by 40 m collected by means of a GPS TRIMBLE 4600 LS ® and Total Station Leica TC600 ®. In these points, it is considered a buffer of 7.2 m of radius corresponding to the spacing between two lines of planting.

At these points, its framing was not considered in the pixels of the Landsat 8 OLI image. In this sense, a criterion of selection of the points was put into practice, in which it guarantees that it is representative of a single pixel. The criterion of point selection was established in the condition that the polygon of the buffer contained in a single pixel of the Landsat 8 OLI image, discarding all points in an intersection between two pixels. The resampling was carried out by moving the pointer to the closest position to the center of the pixel in which it contained to reorganize a mesh structure coincident with the images

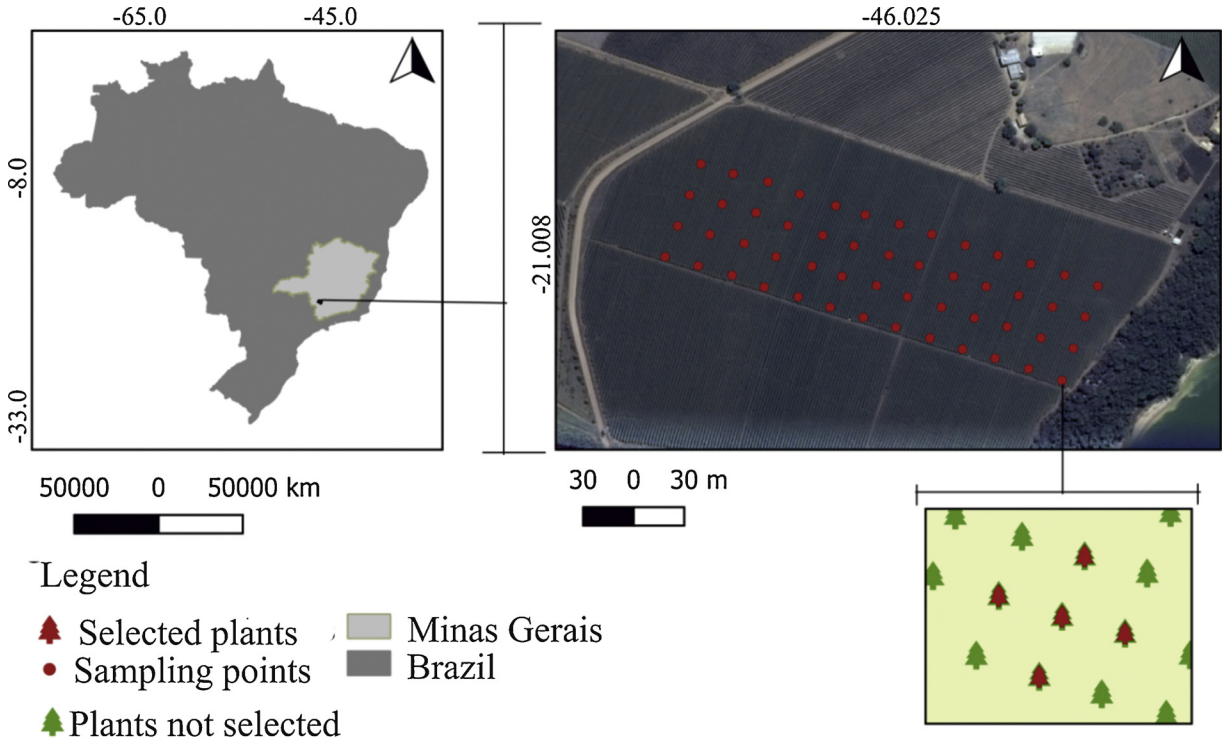

Fig. 1. The spatial location of the analyzed area highlighted the distribution of the sample set, emphasizing the five coffee plants selected in each sample.

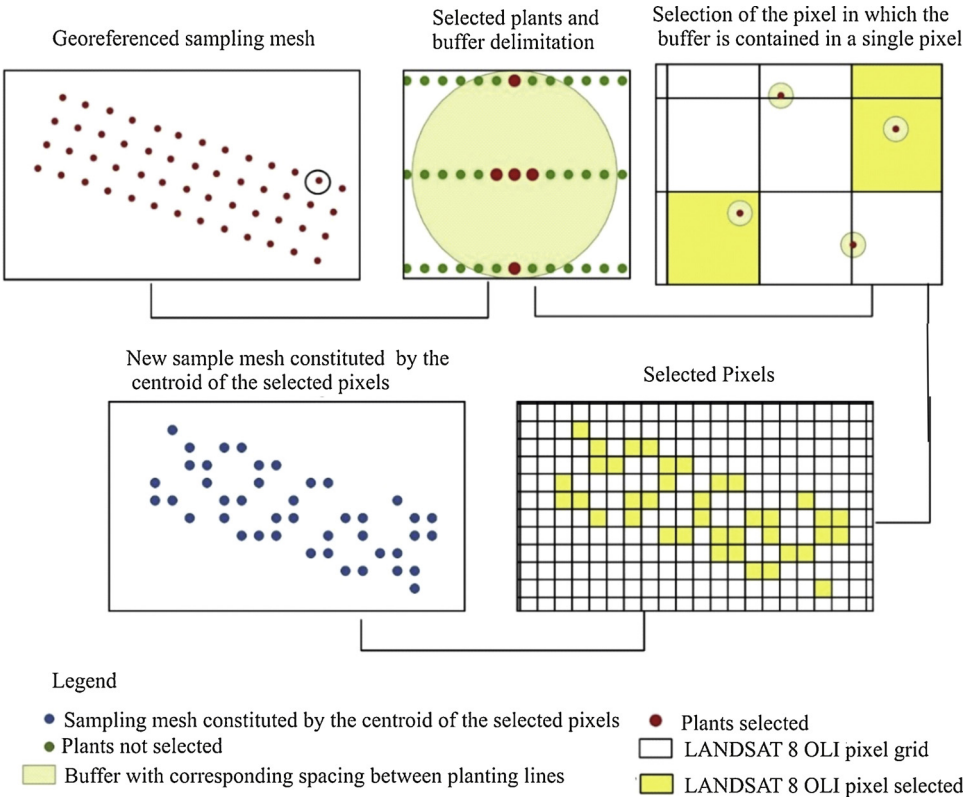

Fig. 2. Scheme of the sample selection process coincident with Landsat 8 OLI images.

(Fig. 2).

2.3. Atmospheric corrections

Images from the Landsat 8 OLI satellite at orbit 219 points 75 which are freely available through the Earth Explore online interface. These images were obtained as close as possible to the date of the field sampling to minimize the phenological transformations that occur over time, which may prevent the direct correlation between the intensity of the necrotic berries and its spectral characteristics. The selected images were collected from December, 6, 2013, January, 23 and February, 24, 2014 (Table 1 and Table 2).

Three atmospheric correction models were used to compare atmospheric correction methods more suitable to characterize the features in the coffee canopy. It has been used the Dark Object Subtraction (DOS), Atmospheric and Topographic Correction for Satellite Imagery (ATCOR) and Second Simulation at the Satellite Signal in the Solar Vector Spectrum (6SV).

The atmospheric correction by the DOS method (Chavez, 1988) performed considering the histogram of the image of the region of smaller wavelength, in which for Landsat 8 OLI images refers to the band of blue of a wavelength of 0.43 μm. The input information were the raw images and their metadata. Based on this information, the atmospheric interference in each spectral band was estimated followed by calculations for the transformation of the digital number into radiance values and then for reflectance values. All procedures of the equations can be consulted in the study of Chavez (1988).

Table 1
Images Landsat 8 OLI used in the process and metadata information.

Date	12/06/2013	01/23/2014	02/24/2014
Land cloud cover (%)	17.38	23.39	38.89
Sun Azimuth angle	76.08	94.12	100.16

Table 2
Description of Landsat 8 OLI used products.

Bands	Wavelength (μm)	Description
B2	0.452-0.512	Blue
B3	0.533-0.590	Green
B4	0.636-0.673	Red
B5	0.851-0.879	Near Infrared
B6	1.566-1.651	Short-wave infrared 1
B7	2.107-2.294	Short-wave infrared 2

For ATCOR atmospheric correction, the algorithm proposed by Richter (1996) consisted in the entry of a fog-free image, cloud shadow and full pixel mask. However, in the absence of this information, the azimuth and zenith angle information were used as the basis, the calibration coefficients contained in the metadata Image. Consequently, the top atmosphere reflectance (TOA) and the cloud mask were defined.

To mitigate the effects of cloudiness that masked the actual reflectance, the values of the highest and lowest brightness pixels and the maximum magnitude (in pixels) present in each cloud have been adjusted. The altered adjustment values followed the criterion in which the current mask overlies the maximum on the cloud cover.

The adjustment of the scene lighting conditions was established based on the digital elevation model (DEM) imagery from the Shuttle Radar Topographic Mission (SRTM). This feature allows the radar, transmittance and irradiance values to be obtained in conjunction with the TOA image. This process was performed iteratively to recover the surface reflectance value for each pixel.

Considering the effects of aerosols on the atmosphere, the rural model was selected to represent the aerosol conditions that are not influenced by urban or industrial sources. It is a product of the reactions between atmospheric gases and the effects of the dust particles (Richter and Schläpfer, 2011, 2003).

For atmospheric correction 6SV (Vermote et al., 2016), the product already processed by NASA has been used. Atmospheric data for the

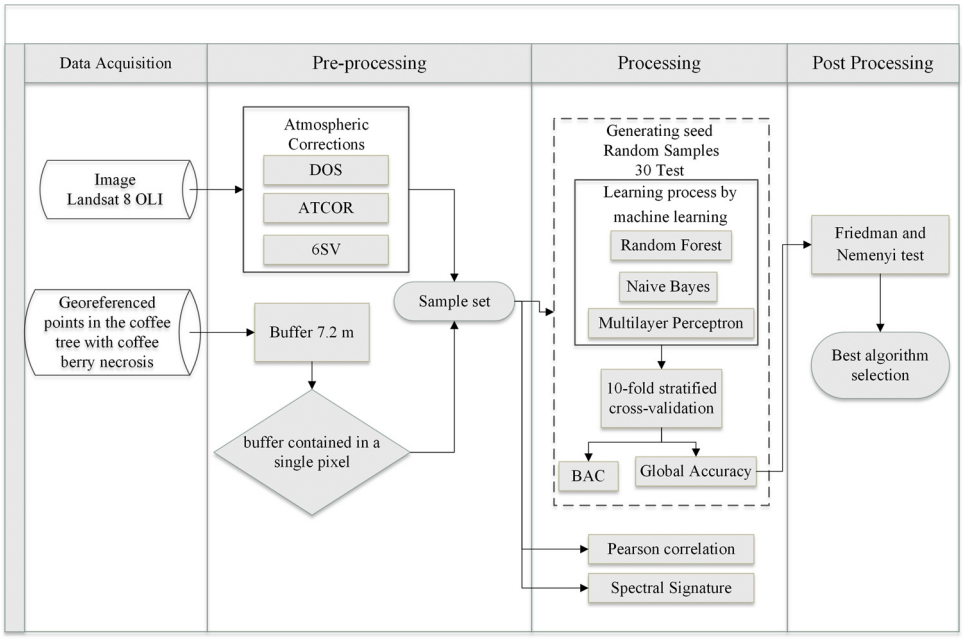

Fig. 3. Scheme of the methodology used to choose the best combination of data for the machine learning process in the classification of the incidence of coffee berry necrosis.

latest 6SV models use the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor as the source of information for applying the atmospheric correction to Landsat 8 OLI images.

With these reflectance values corrected by atmospheric effects, it was possible to perform radiometric transformations in order to enhance the vegetation information. The normalized difference vegetation index (NDVI) proposed by Rouse (1973), the improved vegetation index (EVI) proposed by Huete et al. (1997) and the normalized difference water index (NDWI) proposed by Gao (1996) have been used. (Equations 1, 2 and 3).

$$NDVI = \frac{\rho_{nir} - \rho_{red}}{\rho_{nir} + \rho_{red}} \quad (1)$$

$$EVI = G^{*}\frac{\rho_{nir} - \rho_{red}}{L+\rho_{nir} + C_1{*}\rho_{red} - C_2{*}\rho_{blue}} \quad (2)$$

$$NDWI = \frac{\rho_{nir} - \rho_{swir1}}{\rho_{nir} + \rho_{swir1}} \quad (3)$$

Where: ρ_{nir} is the band reflectance of Near-Infrared; ρ_{red} is the reflectance band of red; ρ_{swir1} is the band reflectance of short-wave infrared 1; ρ_{blue} is the band reflectance of blue; L is the soil adjustment factor adopted for value 1; C_1 is the coefficient for the aerosol effect adopted in value 6; C_2 also refers to aerosol, however the adopted value was 7.5; G refers to gain factor of 2.5.

2.4. Machine learning process

We have used the Python-based learning algorithm Naive Bayes (John and Langley, 1995), Random Forest (Breiman, 2001) and the Multilayer Perceptron (Hinton, 1990).

For the Naive Bayes algorithm, the classification that has been performed started by estimating the probabilities of each class, followed by the calculation of the respective mean, so the algorithm constructed the covariance matrices forming the discriminant function for each type according to Bhargavi and Jyothi (2009).

In the classification by Random Forest, we used a set of 500 decision trees that were formed by values of the set of data sampled by bootstrap. According to Lawrence et al. (2006), estimation errors tend to stabilize before this reached tree number. The entropy criterion was used in tree hierarchization to reduce the randomness of the classifier.

However, the classifier by the Multilayer Perceptron configured so that the found errors were less than 0.000001 in the classification, or that reached a maximum iteration of 1000, the optimization of the weights adjusted the errors based on the Adam stochastic function gradient (Kingma and Ba, 2014).

The algorithms were validated by the 10-fold stratified cross-validation method, which consisted of the iteration number of the algorithms in which each round is a new set of training data, and test was changed and organized with the same amount of repetitions of each sample class. The results of the classification were defined by the final mean of alliterations, as recommended by the authors (Hall et al., 2009).

2.5. Evaluation of machine learning models

We performed 30 tests of the algorithms, modifying the sample set by randomizing the generating seed from 1 to 30. In each analysis, the global accuracy, user accuracy, producer accuracy and Balanced Accuracy (BAC) have been calculated. The global accuracy was defined by the number of the correctness of the error matrix by the total of evaluated samples. The accuracy of the user is associated with the commission error, in which the committed error is attributed to a pixel that does not belong to the true class. The producer's accuracy is associated with the omission error, which occurs when we fail to map a pixel in the true class and the BAC is the average specificity and sensitivity.

The best classifier defined by the method of Friedman and Nemenyi was the one that performs a non-parametric analysis of variance for a single factor of variation and makes comparisons between independent samples by ordering the data by increasing values; and then the original values are replaced by the order number in a set of ordered series.

The global accuracy values of each test ranked in increasing order among the classifiers in different atmospheric corrections. If there are no statistically significant differences between the two classifiers, they will be connected in the diagram by a straight line (Rodríguez et al., 2010). All discussed processing is expressed in the flowchart below (Fig. 3).

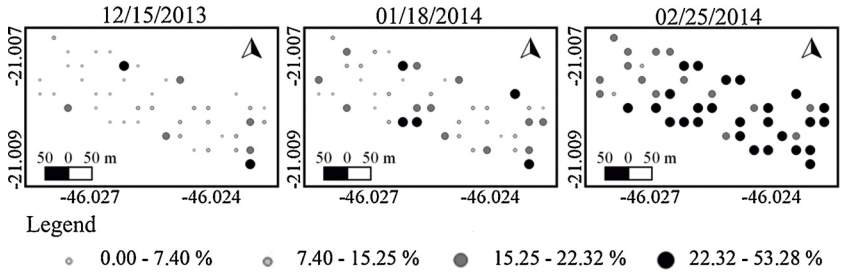

Fig. 4. Quantile spatial distribution of the incidence of coffee berry necrosis field data throughout December (2013), January and February (2014).

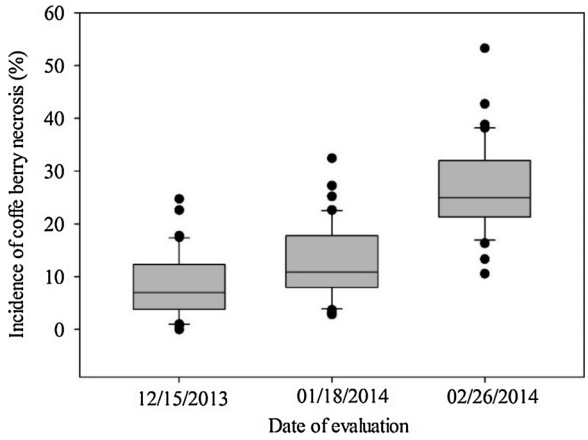

Fig. 5. Box plot of field data evaluations of the coffee berry necrosis incidence during December (2013), January, and February (2014).

3. Results and discussion

On-site monitoring showed a peak infestation in February 2014 (Fig. 4 and 5). During this period, berries were in the phenological phase of expansion, in which there was a greater proliferation of the attack of the pathogen on the reproductive nodes. We believe that the occurrence of diseases in the coffee crop could be influenced by factors related to the pathogen's virulence, as well as host resistance, climatic conditions and crop management (Maia et al., 2013).

There were isolated points in February in which the incidence was lower than that found in January. This was a normal diseased because berries tend to fall. The infestation of the *Colletotrichum gloeosporioides* in an advanced stage presents necrotic centers that when reaches the leaves and berries of the coffee tree tend to cause the early fall beside the dry of the branches. However, in these situations, the same value of incidence of the previous evaluation was considered, since the berry fall did not represent if there was a decrease in the disease (Fernandes and Vieira Junior, 2015; Paradela Filho, 2001).

In the evaluation of the atmospheric correction models, there was

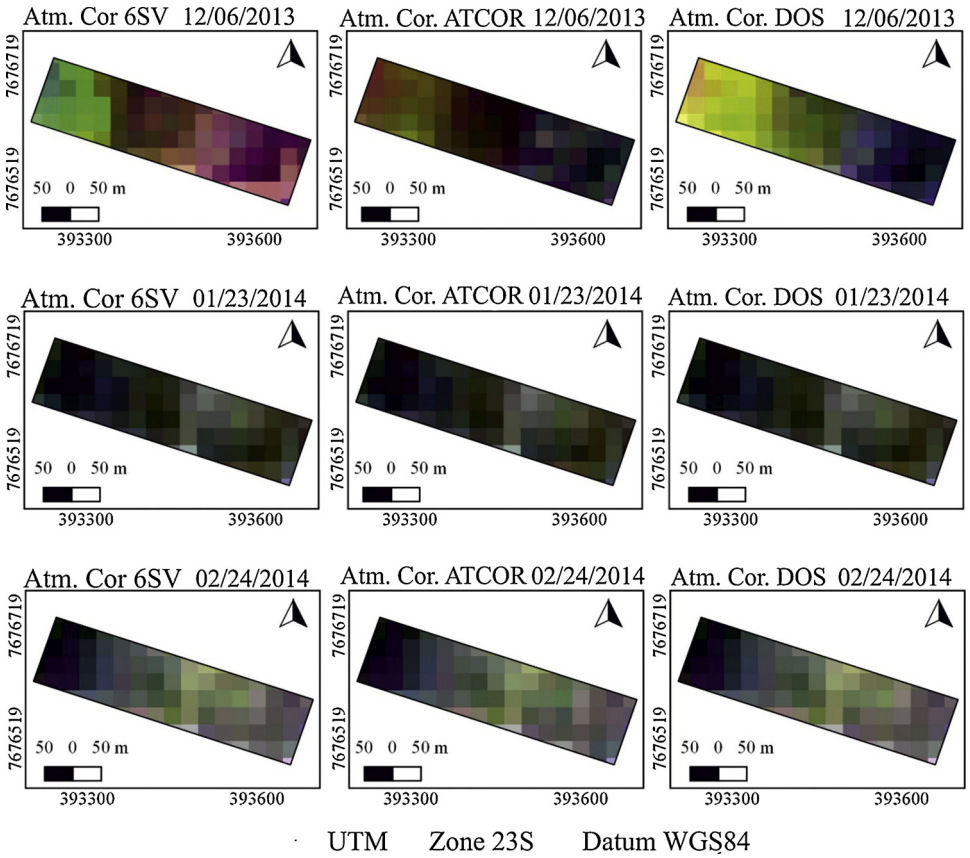

Fig. 6. RGB-321 Landsat 8 OLI color composition for the 6SV, ATCOR and DOS atmospheric correction methods at the dates closest to the field data assessments of coffee berry necrosis incidence.

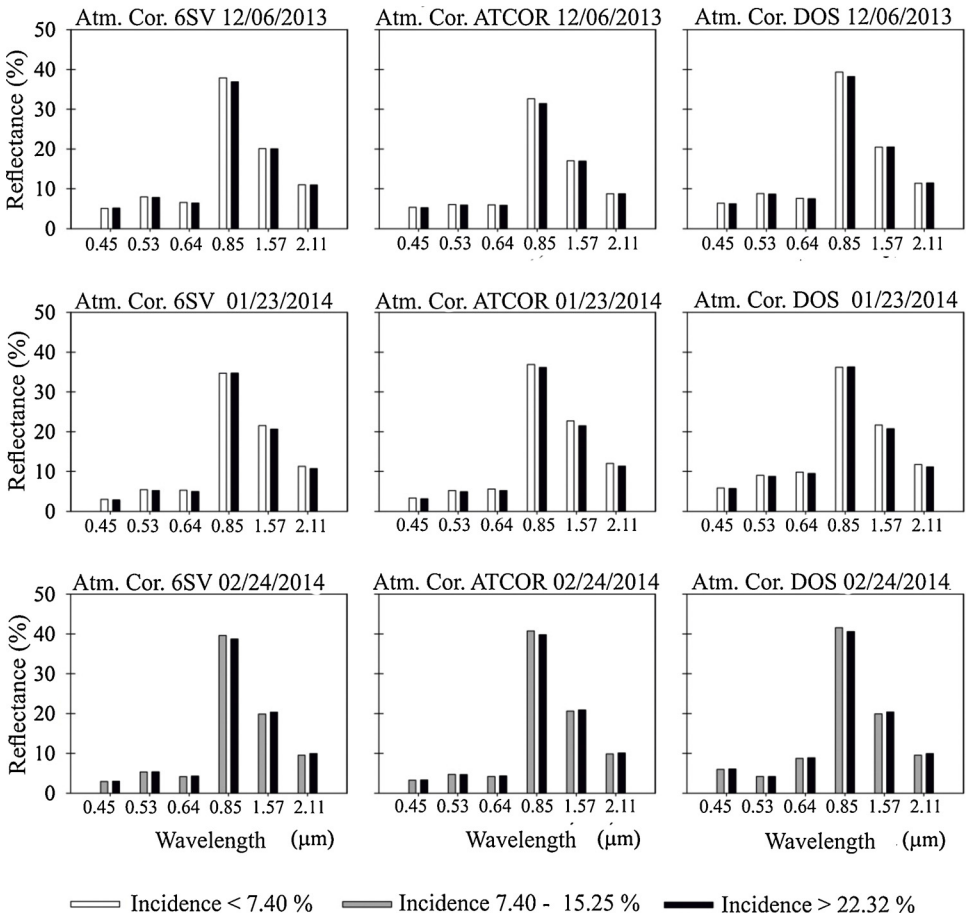

Fig. 7. Spectral signature of the upper and lower classes of the coffee berry necrosis incidence during December, January and February for the 6SV, ATCOR, and DOS atmospheric correction methods.

Table 3
Vegetation index values for the lowest and highest coffee berry necrosis incidence classes in the analyzed period in images of different atmospheric corrections.

Atmospheric correction	Index vegetation	12/06/2013 Incidence (%)		01/23/2014 Incidence (%)		02/24/2014 Incidence (%)	
		< 7.40	> 22.32	< 15.40	> 22.32	< 7.40	> 22.32
6SV	NDVI	0.716	0.685	0.741	0.738	0.816	0.796
	NDWI	0.325	0.278	0.249	0.235	0.342	0.302
	EVI	0.581	0.532	0.515	0.510	0.618	0.603
ATCOT	NDVI	0.742	0.700	0.579	0.575	0.864	0.842
	NDWI	0.332	0.281	0.267	0.253	0.361	0.321
	EVI	0.600	0.531	0.439	0.434	0.819	0.795
DOS	NDVI	0.701	0.662	0.737	0.734	0.819	0.800
	NDWI	0.331	0.273	0.246	0.232	0.338	0.299
	EVI	0.535	0.474	0.517	0.512	0.626	0.612

not a discrepant difference in the visualizations in true color compositions (RGB-432), with an exception of December (Fig. 6). According to the quality of the pixel that is available in the images corrected by 6SV, specifically in this month, there was a higher concentration of aerosol and the presence of clouds. These factors mainly affect the bands in the spectrum in the region of the visible that corresponds to the bands of 0.43, 0.56 and 0.69 μm. The effects of the aerosol concentration on bands centered at shorter wavelengths (0.43 μm) make the surface reflectance generally small, with a robust aerosol signal. In this case, there will be a greater Rayleigh-type dispersion and gas absorption of electromagnetic energy (Vermote et al., 2016). The other configurations of colored compositions presented similar tonalities and visual appearance among themselves. It was the color-accurate color composition that performed the best contrast between the atmospheric corrections.

The reflectance in the visible region was lower in February than in previous months, and this aspect can be observed independently regarding the atmospheric correction method and coffee berry necrosis incidence (Fig. 7). Because the crop is relatively young, approximately three-years old, the plant is in full development, being the month of February with greater leafing, which consequently absorbed more energy in the visible region.

The atmospheric correction in the DOS method presented higher values of reflectance compared to the other methods, possibly because the algorithm does not consider meteorological factors in the calculation. The DOS, the atmospheric effects scaled by the distribution of the histogram of the assigned image in the bands of shorter wavelengths and formulates a linear equation of atmospheric correction for each

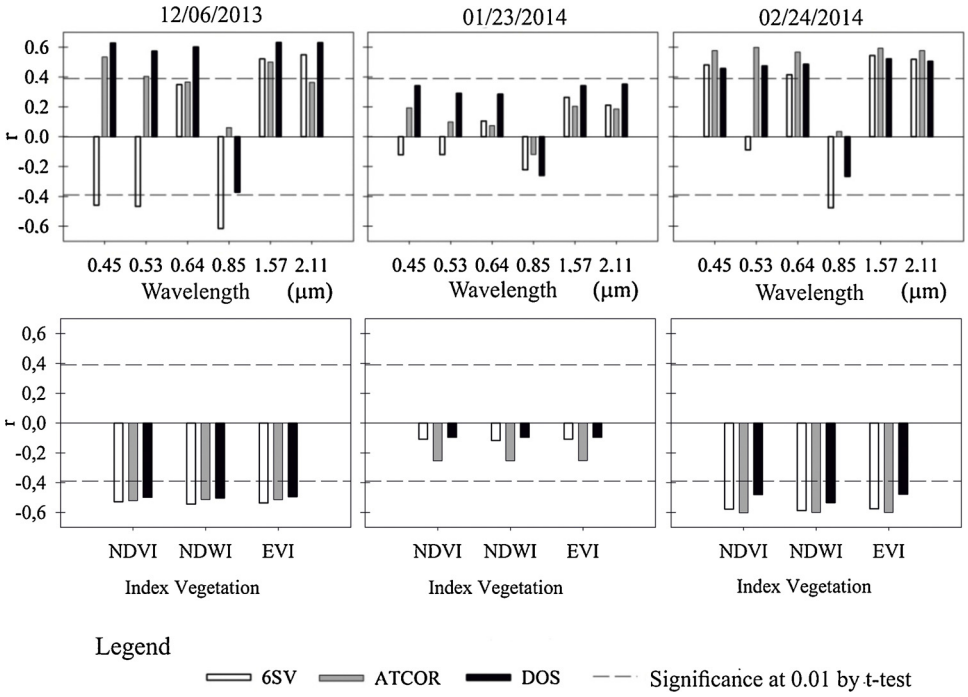

Fig. 8. Pearson correlation coefficient (r) for the reflectance of the wavelengths and index vegetation of the Landsat 8 OLI images between the incidence of coffee berry necrosis ($\alpha \leq 0.01$).

Table 4
Global Accuracy and the Balanced Accuracy (BAC) mean index of the 30 random seed evaluations of the Multilayer Perceptron, Random Forest and Naive Bayes algorithms in detecting the incidence of coffee berry necrosis.

	Atm. Cor. 6SV		Atm. Cor. ATCOR		Atm. Cor. DOS	
All Bands						
Machine Learning Algorithm	Global Accuracy	BAC	Global Accuracy	BAC	Global Accuracy	BAC
Multilayer Perceptron	0.516	0.523	0.577	0.589	0.509	0.518
Random Forest	0.498	0.504	0.487	0.494	0.499	0.503
Naive Bayes	0.549	0.539	0.585	0.577	0.463	0.459
All Bands and Index vegetation						
Multilayer Perceptron	0.520	0.521	0.579	0.580	0.505	0.505
Random Forest	0.504	0.507	0.493	0.493	0.519	0.521
Naive Bayes	0.534	0.526	0.545	0.540	0.475	0.463

band (King, 2003). This method was not sufficient to mitigate the influence of the atmosphere on the image.

According to Martinelli et al. (2015), pathogens cause a reduction in chlorophyll content in leaves due to necrotic lesions, tending to change in spectral signature, causing a change in the value of index vegetation. It has been observed that small differences between changes in reflectance values (Fig. 7) and vegetation indices between incidence less than 7.40% and greater than 22.32% did not exceed 10%, but even subtle there was a signal that fruit necrosis decreases in the value of vegetation indices (Table 3).

The relationship between coffee berry necrosis and reflectance had higher correlations in the medium infrared region and also in the index vegetation, specifically for December and February (Fig. 8). In this region, the internal scattering of the electromagnetic radiation occurred as a consequence of its interaction with the leaf mesophyll, external factors such as the disease can be altering the water and air relation in the mesophyll, resulting in a lower reflectance (Franke and Menz, 2007).

Table 5
Producer and user accuracy for the coffee berry necrosis incidence classes obtained by the Multilayer Perceptron, Random Forest and Naive Bayes classifier algorithms based on the Landsat 8 OLI image reflectance in the 6SV, ATCOR and DOS atmospheric correction methods.

Machine Learning Algorithm	Incidence Class (%)	Atm. Cor. 6SV		Atm. Cor. ATCOR		Atm. Cor. DOS	
		Producer Accuracy	User Accuracy	Producer Accuracy	User Accuracy	Producer Accuracy	User Accuracy
Multilayer Perceptron	00.00 - 07.40	0.606	0.606	0.576	0.704	0.576	0.633
	07.40 - 15.25	0.452	0.333	0.806	0.439	0.452	0.318
	15.25 - 22.32	0.250	0.615	0.250	0.727	0.125	0.571
	22.32 - 53.28	0.727	0.585	0.758	0.735	0.758	0.521
Random Forest	00.00 - 07.40	0.545	0.486	0.576	0.514	0.576	0.500
	07.40 - 15.25	0.355	0.324	0.290	0.281	0.323	0.323
	15.25 - 22.32	0.281	0.333	0.344	0.393	0.375	0.414
	22.32 - 53.28	0.636	0.677	0.636	0.656	0.636	0.677
Naive Bayes	00.00 - 07.40	0.576	0.760	0.576	0.655	0.576	0.633
	07.40 - 15.25	0.613	0.396	0.742	0.404	0.355	0.324
	15.25 - 22.32	0.219	0.636	0.250	0.727	0.219	1.000
	22.32 - 53.28	0.758	0.556	0.758	0.781	0.818	0.466

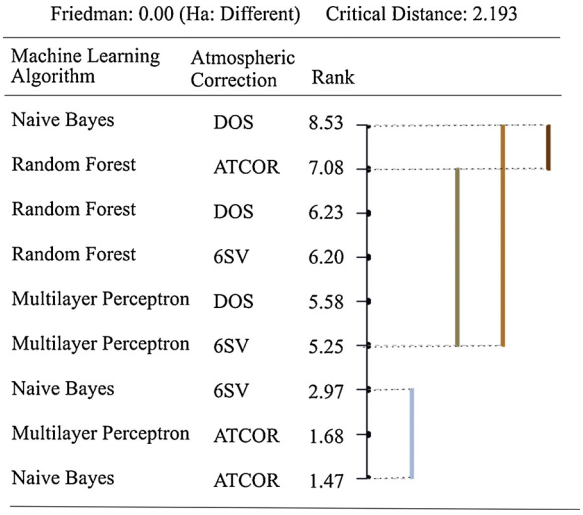

Fig. 9. Friedman and Nemenyi test for the Multilayer Perceptron, Random Forest and Naive Bayes classifier algorithms to identify the coffee berry necrosis incidence classes based on the reflectance of Landsat 8 OLI images in the ATCOR, DOS, and 6SV atmospheric corrections.

Similar results were found by Boechat et al. (2014), who evaluated the spectral signature of the bean white mold using a spectroradiometer field data. The authors observed that in the near-infrared region, the leaves infected by the fungi presented lower reflectance than in healthy leaves, caused by the destruction of plant tissues during colonization of the leaves, as correlations between NDVI and a white mold severity, which were not statistically understood between loading and grain maturation.

Other studies also report the existence of the relationship between reflectance and pathogen infection in plants. In a study by Zhao et al. (2014), the authors evaluated the effect of the severity of yellow rust (*Puccinia striiformis*) on wheat by hyperspectral reflectance of a spectrophotometer. The correlations were highest in the visible region and reached the maximum correlation at about 0.85 with the disease. In the near and medium infrared region, there was a decrease in the mean correlation to 0.45. Prabhakar et al. (2011), in a study on the cotton stress attacked by green spittlebugs (*Hemiptera cicadellidae*), obtained a high negative correlation in the near-infrared region with the pest, reaching a value of -0.77, which corroborated the average R^2 determination index of 0.68 in the NDVI ratio and the leafhopper. In the infrared medium and in the visible region, the correlation was high and positive, with the maximum in the red wavelength reaching approximately 0.79. These studies revealed the potential of using spectral signatures as indicative of physiological disturbances in crops, from satellite imagery, such as Landsat 8 OLI.

Concerning January, the correlations were not significant in the test. In this case, the hypothesis is that the reflectance relations and the incidence of the coffee berry necrosis were not linear. In some instances, not even hyperspectral satellites operating in narrow and specific bands obtained linear relationships, probably due to the spectral mixing components (Chemura et al., 2018b).

From the use of the machine learning algorithms, it was possible to identify the level of the coffee berry necrosis incidence as a function of the reflectance of the Landsat 8 OLI bands for every month in the analysis, even in different atmospheric corrections, surpassing the answers obtained by the Pearson correlation. Specifically for biological analysis, modern machine learning techniques were capable of describing patterns that exceed the estimates determined by conventional statistical methods, such as regression and linear correlation (Ma et al., 2014).

In the results concerning the identification of the class intensity of the coffee berry necrosis as a function of the reflectance, the ATCOR atmospheric correction presented the higher values of accuracy and BAC index was compared to the 6SV and DOS ones (Table 4). The inclusion of the SRTM images in the corrections by the ATCOR allowed a reference of illumination originated by the terrain effect, which guaranteed standardization between the images of different evaluated dates. According to Richter and Schläpfer (2011), this mode of atmospheric correction is especially important in cases of multi-temporal, multi-sensor or multi-condition images and must be standardized so that they can be compared.

Vermote et al. (2016) specified that the 6SV correction was limited to uniform and flat targets. Tan et al. (2013) indicated that in cases of rough terrain, the topographic effects may introduce interference in the reflectance promoted by the shading of the image that in this method are not counted in the 6SV atmospheric correction. The application of the DOS method in soft or smooth undulating areas may lead to the underestimation of the atmospheric effects on the images due to the low presence of shading (Chavez, 1988).

The global accuracy and BAC using all bands were similar with respect to analysis of all bands with the addition of vegetation indices. All indices have in common the use of the near-infrared band that obtained significant correlation only with the application of 6SV

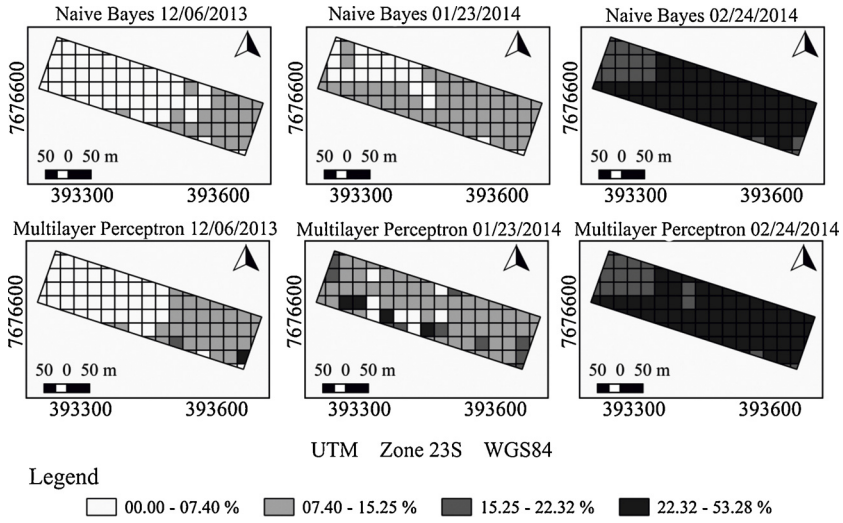

Fig. 10. Incidence classification of coffee berry necrosis from the Naive Bayes and Multilayer perceptron algorithm in Landsat 8 OLI images used in the ATCOR atmospheric correction during December, January and February.

atmospheric correction, so its relationship with fruit necrosis was not enough to increase the accuracy of machine learning models. By the principle of parsimony, which refers to choosing a smaller set of information to elucidate the problem according to Powers and Turk (2012), it has been decided to use only the spectral bands for further analysis.

Considering that the disease level control has an incidence of 5%, being tolerable up to 12%, the primary attention in the accuracy was given to 0–7.4% incidence class. The Multilayer Perceptron algorithm had the highest efficiency classification in images corrected by the 6SV method, and the accuracy was around 0.6. Regarding the user accuracy, the most significant was 0.76, determined by the Naive Bayes in the atmospheric correction 6SV. This adjustment can be advantageous, once the sites that have not been affected yet by the pathogen were identified in the images, which makes targeting the regions possible, in which they need an application of fungicides.

In the lowest incidence class interval between 7.40 and 15%, the highest producer accuracy and user accuracy (of 0.8 and 0.43) was found by the Multilayer Perceptron algorithm in the ATCOR atmospheric correction, respectively (Table 5). These results may aid field data monitoring, in which the indication of the intensity of the disease can be evaluated punctually, thus providing a rapid identification and a previous mapping of the places where a control measure should be taken. Decisions on the timely management of diseases in coffee are particularly important because they are closely linked to the yield losses (Martinelli et al., 2015).

The Friedman test indicated that there was a significant difference (p-value < 0.01) between the classifications, given that the Naive Bayes in the ATCOR correction suggested by the Nemenyi test is the best classifier (Fig. 9). Although simple, Naive Bayes can often overcome some sophisticated classification methods (Farid et al., 2012). This algorithm based on the common assumption that all the characteristics are independent from one another, there is a tendency to be less tolerable to the changes in values in the spectral bands, being able to identify patterns of behavior even with few samples (Xu, 2018). This algorithm is known to be more sensitive to changes in the training set because it has a fixed structure and a small number of parameters (Rodríguez et al., 2013).

The Multilayer Perceptron in the ATCOR correction also stood out among the best classifiers and statically equal to the Naive Bayes, as already pointed out the averages of the accuracy and BAC (Table 1) and the spatial distribution of the classes (Fig. 10). Researches report on the efficiencies of classification techniques with the use of neural networks to produce results according to the presence/absence of the disease in crops and possibly severity levels. West et al. (2004), employing spectral images aboard a spectrograph mounted at the level of the spray bar, obtained increased performance of the classifier in 99% accuracy to differentiate healthy wheat from disease wheat using the algorithm of neural network perceptron multi-layered. Abdulridha et al. (2016) selected the appropriate wavelengths to correctly classify healthy trees of stressed trees with an accuracy of 98% through the neural network. Li et al. (2009) had an accuracy of 95% of the classification by neural networks for unhealthy rice stressed by rice diseases and healthy rice pests based on spectral leaf behavior.

However, the performance of the classifiers depends on the arrangement of the data sets, being that each algorithm can have a better performance in each case. In this paper, Naive Bayes presented the best results, but this does not mean that it is the best classifier. As an example, Ma et al. (2015) identified the activities of relationships of biological structures in medicines by making use of Random Forest, obtaining the best performance in most of the data sets when compared to the classifier of deep learning by neural networks. In the study carried out by Russo et al. (2018), to predict compounds for endocrine-disrupting abilities, such as estrogen receptor binding, Random Forest once again had the best performance when compared to Naive Bayes and Multilayer Perceptron. Other researches evidence the potential of the Multilayer Perceptron. Were et al. (2015), mapping soil organic carbon variations, the Neural Networks algorithm obtained a superior performance of up to 36% when compared to Random Forest. This research evidenced the need for tests to choose the best learning algorithm of the machine adjusted to the data set.

4. Conclusion

The global accuracy and BAC were generally less than 0.60 for the trained data set. With a more robust database of samples from other coffee crops, it is believed that this result may be more accurate. The results are indicative that Landsat 8 OLI images may provide pertinent information for decision-making in agricultural planning, as well as for the timely application of pesticides. Machine learning tools were more efficient than Pearson's correlation to detect the incidence of coffee necrosis. The best classifier performance was Naive Bayes and Multilayer Perceptron in atmospheric images corrected by the ATCOR method.

Acknowledgements

The authors wish to thank (i) the Agricultural Engineering Department the Federal University of Lavras (UFLA) for providing office space and infrastructure to achieve this article, (ii) the Foundation for Supporting Research of the State of Minas Gerais (FAPEMIG).

References

Abdulridha, J., Ehsani, R., de Castro, A., 2016. Detection and differentiation between laurel wilt disease, phytophthora disease, and salinity damage using a hyperspectral sensing technique. Agriculture 6, 56. https://doi.org/10.3390/agriculture6040056.

Belgiu, M., Drăgu, L., 2016. Random forest in remote sensing: a review of applications and future directions. ISPRS J. Photogramm. Remote Sens. 114, 24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011.

Bhargavi, P., Jyothi, S., 2009. Applying naive bayes data mining technique for classification of agricultural land soils. IJCSNS Int. J. Comput. Sci. Netw. Secur. 9, 117–122.

Boechat, L.T., de Carvalho Pinto, F., de, A., de Paula, T.J., Queiroz, D.M., Teixeira, H., 2014. Detection of white mold in dry beans using spectral characteristics. Rev. Ceres 61, 907–915. https://doi.org/10.1590/0034-737X201461060004.

Breiman, L., 2001. Random forest. Mach. Learn. 45, 1–33. https://doi.org/10.1023/A:1010933404324.

Chavez, P.S., 1988. An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data. Remote Sens. Environ. 24, 459–479. https://doi.org/10.1016/0034-4257(88)90019-3.

Chemura, A., Mutanga, O., Dube, T., 2017. Separability of coffee leaf rust infection levels with machine learning methods at Sentinel-2 MSI spectral resolutions. Precis. Agric. 18, 859–881. https://doi.org/10.1007/s11119-016-9495-0.

Chemura, A., Mutanga, O., Odindi, J., Kutywayo, D., 2018a. Mapping spatial variability of foliar nitrogen in coffee (Coffea arabica L.) plantations with multispectral Sentinel-2 MSI data. ISPRS J. Photogramm. Remote Sens. 138, 1–11. https://doi.org/10.1016/j.isprsjprs.2018.02.004.

Chemura, A., Mutanga, O., Sibanda, M., Chidoko, P., 2018b. Machine learning prediction of coffee rust severity on leaves using spectroradiometer data. Trop. Plant Pathol. 43, 117–127. https://doi.org/10.1007/s40858-017-0187-8.

Farid, D.M., Zhang, L., Rahman, C.M., Hossain, M.A.A., Strachan, R., Mofizur, C., Hossain, M.A.A., Strachan, R., 2012. Expert systems with applications hybrid decision tree and naïve Bayes classifiers for multi-class classification tasks. Expert Syst. Appl. 41, 1937–1946. https://doi.org/10.1016/j.eswa.2013.08.089.

Fernandes, C., de, F., Vieira Junior, J.R., 2015. Coffee diseases, in: Embrapa Rondônia-Chapter of Scientific Book (ALICE). In: MARCOLAN, A.L., ESPINDULA, M.C. (Eds.), Coffee in the Amazônia, Brasília, DF ….

Franke, J., Menz, G., 2007. Multi-temporal wheat disease detection by multi-spectral remote sensing. Precis. Agric. 8, 161–172. https://doi.org/10.1007/s11119-007-9036-y.

Gao, B.-C., 1996. NDWI—a normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sens. Environ. 58, 257–266. https://doi.org/10.1016/S0034-4257(96)00067-3.

Griffiths, E., Gibbs, J.N., Waller, J.M., 1971. Control of coffee berry disease. Ann. Appl. Biol. 67, 45–74. https://doi.org/10.1111/j.1744-7348.1971.tb02907.x.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H., 2009. The WEKA data mining software: an update. SIGKDD Explor. Newsl. 11, 10–18. https://doi.org/10.1145/1656274.1656278.

Hinton, G.E., 1990. Connectionist learning procedures. Mach. Learn. 555–610. https://doi.org/10.1016/B978-0-08-051055-2.50029-8.

Huete, A.R., Liu, H.Q., Batchily, K.V., Van Leeuwen, W., 1997. A comparison of vegetation indices over a global set of TM images for EOS-MODIS. Remote Sens. Environ. 59, 440–451. https://doi.org/10.1016/S0034-4257(96)00112-5.

John, G.H., Langley, P., 1995. Estimating Continuous Distributions in Bayesian Classifiers George, in: Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence., UAI'95. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp. 338–345. https://doi.org/10.1109/TGRS.2004.834800.

King, R.B., 2003. Remote sensing geology. The Photogrammetric Record. Springerhttps://doi.org/10.1046/j.0031-868x.2003.024_04.x.

Kingma, D.P., Ba, J., 2014. Adam: A Method for Stochastic Optimization.

Lawrence, R.L., Wood, S.D., Sheley, R.L., 2006. Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (randomForest). Remote Sens. Environ. 100, 356–362. https://doi.org/10.1016/j.rse.2005.10.014.

Li, B., Liu, Z., Huang, J., Zhang, L., Zhou, W., Shi, J., 2009. Hyperspectral identification of rice diseases and pests based on principal component analysis and probabilistic neural network. Nongye Gongcheng Xuebao/Trans. Chin. Soc. Agric. Eng. 25, 143–147. https://doi.org/10.3969/j.issn.1002-6819.2009.09.026.

Lopresti, M.F., Di Bella, C.M., Degioanni, A.J., 2015. Relationship between MODIS-NDVI data and wheat yield: a case study in Northern Buenos Aires province, Argentina. Inf. Process. Agric. 2, 73–84. https://doi.org/10.1016/J.INPA.2015.06.001.

Ma, C., Zhang, H.H., Wang, X., 2014. Machine learning for Big Data analytics in plants. Trends Plant Sci. 19, 798–808. https://doi.org/10.1016/j.tplants.2014.08.004.

Ma, J., Sheridan, R.P., Liaw, A., Dahl, G.E., Svetnik, V., 2015. Deep neural nets as a method for quantitative structure-activity relationships. J. Chem. Inf. Model. 55, 263–274. https://doi.org/10.1021/ci500747n.

Mahlein, A.-K., Rumpf, T., Welke, P., Dehne, H.-W., Plümer, L., Steiner, U., Oerke, E.-C., 2013. Development of spectral indices for detecting and identifying plant diseases. Remote Sens. Environ. 128, 21–30. https://doi.org/10.1016/J.RSE.2012.09.019.

Mahlein, A.K., 2016. Present and future trends in plant disease detection. Plant Dis. 100, 1–11. https://doi.org/10.1007/s13398-014-0173-7.2.

Maia, F.G.M., Armesto, C., Ogoshi, C., Vieira, J.F., Maia, J.B., de Abreu, M.S., 2013. Behavior of isolated of Colletotrichum Gloeosporioides inoculated micropropagated in seedlings of coffee. Biosci. J. 29.

Martinelli, F., Scalenghe, R., Davino, S., Panno, S., Scuderi, G., Ruisi, P., Villa, P., Stroppiana, D., Boschetti, M., Goulart, L.R., Davis, C.E., Dandekar, A.M., 2015. Advanced methods of plant disease detection. A review. Agron. Sustain. Dev. 35, 1–25. https://doi.org/10.1007/s13593-014-0246-1.

Paradela Filho, O., et al., 2001. The Colletotrichum Coffee Complex. Campinas Inst. Agronômico, Bol. Técnico IAC.

Pezzopane, J.R.M., Pedro Júnior, M.J., Thomaziello, R.A., Camargo, M.B.Pde, 2003. Coffee phenological stages evaluation scale. Bragantia 62, 499–505. https://doi.org/10.1590/S0006-87052003000300015.

Powers, D.M.W., Turk, C.C.R., 2012. Machine Learning of Natural Language. Springer Science & Business Media.

Prabhakar, M., Prasad, Y.G., Thirupathi, M., Sreedevi, G., Dharajothi, B., Venkateswarlu, B., 2011. Use of ground based hyperspectral remote sensing for detection of stress in cotton caused by leafhopper (Hemiptera: Cicadellidae). Comput. Electron. Agric. 79, 189–198. https://doi.org/10.1016/j.compag.2011.09.012.

Price, T.V., Gross, R., Ho Wey, J., Osborne, C.F., 1993. A comparison of visual and digital image-processing methods in quantifying the severity of coffee leaf rust (Hemileia Vastatrix). Aust. J. Exp. Agric. 33, 97–101. https://doi.org/10.1071/EA9930097.

Richter, R., 1996. Atmospheric correction of satellite data with haze removal including a haze/clear transition region. Comput. Geosci. 22, 675–681. https://doi.org/10.1016/0098-3004(96)00010-6.

Richter, R., Schläpfer, D., 2011. Atmospheric/Topographic Correction for Satellite Imagery. In : DLR Report DLR-IB 565-02/11. DLR Rep. DLR-IB 565, 202.

Richter, R., Schläpfer, D., 2003. Atmospheric/topographic correction for satellite imagery: ATCOR-2/3 user guide, version 9.1.1. February 2017. ReSe Appl. Schläpfer 3, 270.

Rodríguez, J.D., Pérez, A., Lozano, J.A., 2013. A general framework for the statistical analysis of the sources of variance for classification error estimators. Pattern Recognit. 46, 855–864. https://doi.org/10.1016/j.patcog.2012.09.007.

Rodríguez, J.D., Pérez, A., Lozano, J.A., 2010. Sensitivity analysis of k-Fold cross validation in prediction error estimation. IEEE Trans. Pattern Anal. Mach. Intell. 32, 569–575. https://doi.org/10.1109/TPAMI.2009.187.

Rouse, J.W. 1974, 1973. Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation.

Russo, D.P., Zorn, K.M., Clark, A.M., Zhu, H., Ekins, S., 2018. Comparing multiple machine learning algorithms and metrics for estrogen receptor binding prediction. Mol. Pharm. 15, 4361–4370. https://doi.org/10.1021/acs.molpharmaceut.8b00546.

Santos Neto, H., 2017. Temporal Space Analysis of Patosystem Relations with the Magnesium, Calcium and Potassium Nutrients. Universidade Federal de Lavras.

Sera, G.H., Altéia, M.Z., Sera, T., Petek, M.R., Ito, D.S., 2005. Correlation among the Colletotrichum spp. incidence with some coffee agronomic traits. Bragantia 64, 435–440. https://doi.org/10.1590/S0006-87052005000300013.

Tan, B., Masek, J.G., Wolfe, R., Gao, F., Huang, C., Vermote, E.F., Sexton, J.O., Ederer, G., 2013. Improved forest change detection with terrain illumination corrected Landsat images. Remote Sens. Environ. 136, 469–483. https://doi.org/10.1016/J.RSE.2013.05.013.

Tucker, C.J., Grant, D.M., Dykstra, J.D., 2013. NASA's global orthorectified landsat data set. Photogramm. Eng. Remote Sens. 70, 313–322. https://doi.org/10.14358/pers.70.3.313.

Varzea, V.M.P., Rodrigues, C.J., Lewis, B.G., 2002. Distinguishing characteristics and vegetative compatibility of Colletotrichum kahawe in comparison with other related species from coffee. Plant Pathol. 51, 202–207. https://doi.org/10.1046/j.1365-3059.2002.00622.x.

Vermote, E., Justice, C., Claverie, M., Franch, B., 2016. Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. Remote Sens. Environ. 185, 46–56. https://doi.org/10.1016/j.rse.2016.04.008.

Were, K., Bui, D.T., Dick, Ø.B., Singh, B.R., 2015. A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape. Ecol. Indic. 52, 394–403. https://doi.org/10.1016/j.ecolind.2014.12.028.

West, J., Wahlen, S., McCartney, A., Ramon, H., Bravo, C., Moshou, D., Bravo, C., West, J., Wahlen, S., McCartney, A., Ramon, H., 2004. Automatic detection of 'yellow rust' in wheat using reflectance measurements and neural networks. Comput. Electron. Agric. 44, 173–188. https://doi.org/10.1016/j.compag.2004.04.003.

Xu, S., 2018. Bayesian Naïve Bayes classifiers to text classification. J. Inf. Sci. 44, 48–59. https://doi.org/10.1177/0165551516677946.

Zhao, J., Huang, L., Huang, W., Zhang, D., Yuan, L., Zhang, J., Liang, D., 2014. Hyperspectral measurements of severity of stripe rust on individual wheat leaves. Eur. J. Plant Pathol. 139, 401–411. https://doi.org/10.1007/s10658-014-0397-6.