



ENGINEERING SCIENCES

Classifiers based on artificial intelligence in the prediction of recently planted coffee cultivars using a Remotely Piloted Aircraft System

NICOLE L. BENTO, GABRIEL ARAÚJO E.S. FERRAZ, RAFAEL ALEXANDRE P. BARATA, DANIEL V. SOARES, SABRINA A. TEODORO & PEDRO HENRIQUE DE O. ESTIMA

Abstract: The classification and prediction methods through artificial intelligence algorithms are applied in different sectors to assist and promote intelligent decision-making. In this sense, due to the great importance in the cultivation, consumption and export of coffee in Brazil and the technological application of the Remotely Piloted Aircraft System (RPAS) this study aimed to compare and select models based on different data classification techniques by different classification algorithms for the prediction of different coffee cultivars (*Coffea arabica* L.) recently planted. The attributes evaluated were height, crown diameter, total chlorophyll content, chlorophyll A and chlorophyll B, Foliar Area Index (LAI) and vegetation indexes NDVI, NDRE, MCARI1, GVI, and CI in six months. The data were prepared programming language Python using algorithms of Decision Trees, Random Forest, Support Vector Machine and Neural Networks. It was evaluated through cross-validation in all methods, the distribution by FreeViz, the hit rate, sensitivity, specificity, F1 score, and area under the ROC curve and percentage and predictive performance difference. All algorithms showed good hits and predictions for coffee cultivars (0.768% Decision Tree, 0.836% Random Forest, 0.886 Support Vector Machine and 0.899 Neural Networks) and the Neural Networks algorithm produced more accurate predictions than other tested algorithm models, with a higher percentage of hits for the classes considered.

Key words: *Coffea arabica* L., neural networks, precision farming, remote sensing.

INTRODUCTION

The coffee crop is of great importance to Brazil, accounting for about 47% of world production, and expected exports for the 2020/2021 crop show an average total of 47 million bags (of 60 kg) (Conab 2021). Among the species cultivated, *Coffea arabica* L. refers to a coffee of greater appreciation by consumers and therefore has significant economic value (Embrapa 2020).

Due to the importance of this commodity, the use of tools from new scientific fields such as digital agriculture and artificial intelligence have been applied for research and prediction

of behavior, development, and damage in coffee culture (Marin et al. 2021, Maciel et al. 2020, Marujo et al. 2017, Alves et al. 2016). Machine learning, big data and data mining allied to technologies based on remote sensing currently evidenced by the application of the Remotely Piloted Aircraft System (RPAS) seek to optimize the activities to be developed in the field, providing gain in time, profitability to producers (Liakos et al. 2018).

The recognition and classification of standards is a natural activity of human beings, however, due to the complexity of the insertion of many variables we seek the applicability

of statistical methodologies that enable the technological automation of this practice, using computational algorithms built in the form of pre-defined rules (Borsato et al. 2011). The use of data-intensive approaches, driven by the high performance of predictive models, highlights the ability to generate information to provide support and understanding of the environment, which condition more accurate results and intelligent decision making (Fernandes & Chiavegatto Filho 2019). Besides obtaining data, they must also be stored, processed, analyzed, and interpreted (Sarri et al. 2017).

Prediction models have been applied in several areas of knowledge, including agriculture (Mincato et al. 2020, Liakos et al. 2018, Chlingaryan et al. 2018). In coffee culture, several applications of the use of artificial intelligence are found in the literature, of which the quality classification of coffee beans (Oliveira et al. 2019), prediction of the degree of roasting of coffee (Leme et al. 2019), detection of diseases such as rust using neural networks (Da Silva et al. 2017) and decision tree (Meira et al. 2009), geographical identification of coffee samples (Borsato et al. 2011), mapping of coffee areas (Souza et al. 2019, Marujo et al. 2017, Souza et al. 2016) among other applications. Regarding crop forecasting studies, computational models resulting from the performance of artificial intelligence techniques and data science have great contributions and advantages of being used, as they allow the interconnected study of management, climate and soil variables presenting themselves as useful tools for the study of dynamic and complex environments (Van Keulen & Asseng 2019).

In this sense, understanding and identifying the area planted with coffee cultivars is of fundamental importance, especially in the initial period of post-planting to the first year of crop formation, since such information helps in

understanding the development of the cultivar, as well as its specific needs related to its fixation and cultivation, variant among cultivars (Mesquita et al. 2016). Classifying coffee cultivars in the field is extremely important, especially on farms with large areas planted with coffee trees, since it is a perennial crop that remains under development and productivity in the field for a long time, there is a need to know how to certain which cultivar is implemented in each field, and the use of sensors coupled to remotely piloted aircraft combined with the help of artificial intelligence methodologies refer to an optimized and efficient application for this purpose.

Usually, classification and mapping studies of coffee areas consist of differentiating areas with different land-use activities (Hunt et al. 2020, Souza et al. 2019, Kelley et al. 2018, Chemura & Mutanga 2017, Kawakubo & Pérez Machado 2016) and not individualization of coffee cultivars. Works with this focus are not so widespread in the literature because the prediction techniques, application of artificial intelligence models and machine learning are recent in the agricultural sector, both due to computational issues and the interest of the agricultural sector, that is, its inclusion in the field is still initial, however, despite being at the beginning of their implementation, these technologies already demonstrate a vast potential for use in the field, especially with high-resolution images with multispectral cameras, showing the innovation and applicability of studies with this focus.

In this context, the objective was to identify the best performance classification algorithm for predicting different newly planted coffee cultivars, based on field collection data and Vegetation Indices (VIs) from aerial images of the Remotely Piloted Aircraft System (RPAS) using artificial intelligence resources.

MATERIALS AND METHODS

Study area

The study area encompasses three sub-areas of recently planted coffee crops at the beginning of the 5-month studies of *Coffea arabica* L. cultivars Catucaí (2SL), Catuaí (IAC 62) and Bourbon (IAC J10) according to the National Registry of Cultivars - RNC, of the Brazilian Ministry of Agriculture, Livestock and Supply (Mapa 2018). The area is located in the municipality of Santo Antônio do Amparo, western region of Minas, Minas Gerais, between the meridians 506000 and 508000 m W and parallel 7680000 and 7690000 m S, in the UTM projection zone 23 S and geodesic reference Sirgas 2000 (Figure 1).

The area has an average altitude of 1022 m, is inserted in the Atlantic Forest Biome with its soil classification as Distrophical Red Yellow Latosol (Embrapa 2018) and according to Köppen's classification, the climate refers to the Humid Subtropical (Cwb) and average temperatures between 18°C and 22°C (Alvares et al. 2013).

Each sub-area of the study was standardized in 0.60 ha, with 15 planting rows and 200 plants per study row, totaling 3000 plants per area. The coffee plants have 3.8 m spacing between rows, 0.5 m between plants, and the presence of brachiaria (*Brachiaria decumbens*) in the between rows. In each area, there was a systematic distribution of 5 sample points, composed of 4 coffee plants, two in the central street, and two in each side street, according to the methodology proposed by Ferraz et al. (2017), totaling 20 plants sampled per study area.

Field data

In each studied plant were collected data of height, crown diameter, total chlorophyll content, chlorophyll A and chlorophyll B, and Leaf Area Index (LAI) measured in the field, in addition to Vegetation Indices (VIs) obtained

by the acquisition of aerial images through the Remotely Piloted Aircraft System (RPAS).

The measurements of plant height and crown diameter were performed with the aid of a conventional ruler. The total chlorophyll contents, A and B were obtained by portable chlorophyll meter atLEAF Chl meter (atLEAF 2019) using an average of the reading of three plant leaves and calculation of chlorophyll contents according to the equation proposed by Padilla et al. (2018) (Equation 1, 2, 3). The Leaf Area Index (LAI) was obtained according to the equation proposed by Favarin et al. (2002) (Equation 4).

$$\text{Total chlorophyll content} = \text{chlorophyll A} + \text{chlorophyll B} \quad (1)$$

$$\text{Chlorophyll A} = -5,774 + (0,43 * \text{atLEAF}) + (0,0045) * (\text{atLEAF}^2) \quad (2)$$

$$\text{Chlorophyll B} = 0,04 * (\text{atLEAF}^1,57) \quad (3)$$

$$\text{LAI} = 0,0134 + 0,7276 * D^2 * h \quad (4)$$

where Chl t - Total chlorophyll content, Chl A - Chlorophyll A, Chl B - Chlorophyll B ($\mu\text{g}/\text{cm}^2$); atLEAF- measurement obtained in the chlorophyll meter atLEAF (IRC); LAI - Leaf Area Index (adimensional); D - plant crown diameter (m); h - height of plants (m).

The VIs data were obtained from images taken by Matrice 100 Remotely-Piloted Aircraft (Dji 2015) with on-board sensor Parrot Sequoia multispectral camera (Micasense 2016) with reflectance values in the spectral bands of green (550 to 590 nm), red (660 to 700 nm), red (735 to 745 nm), near-infrared (760 to 820 nm) and RGB (380 to 720 nm), being the calculations performed with the average values of the bands of the spectral bands and with radiometric calibration of the sensor before and after flights, with the help of a calibration plate. The flight plan was realized in the Precision Flight software (Precision Hawk 2010) with fixed parameters of 50 meters for flight height, 8 m/s flight speed, 80%

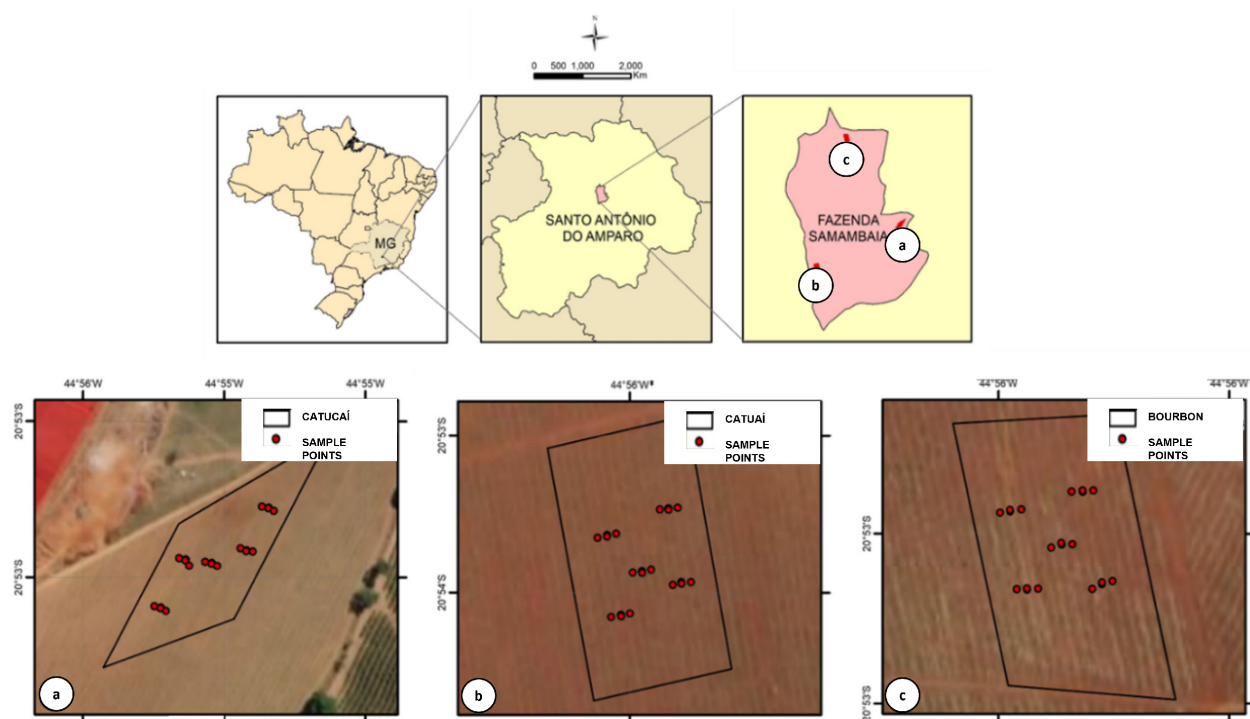


Figure 1. Location map of the studied sub-areas a) Catucaí (2SL), b) Catucaí (IAC 62) and c) Bourbon (IAC J10).

X 80% overlap level, and transversal direction of the flight to the planting row, with sampled ends and plants, demarcated with control points (targets).

The processing of the aerial images occurred in the PIX4D Mapper software (Pix4d SA, 2019) as described in Figure 2 with all items configured in high resolution and the VIs calculated according to equations and references described in Table I which were obtained through the ArcGIS 10.4 software (Esri 2018).

Database

The attributes available on record were height, crown diameter, total chlorophyll content, chlorophyll A and chlorophyll B, Leaf Area Index (LAI), NDVI, NDRE, MCARI1, GVI, and CI of coffee plants in the six months (May, July, September, November 2019 and January, March 2020) totaling 3960 records.

Data processing

The data were prepared with routines developed in Python programming language in the Orange Canvas 3.25.0 software (Demsar et al. 2013). The Orange Canvas is an open-source software. It is based on data mining and machine learning components. Its structure is ordered inflow building blocks for system visual programming named Widgets that are grouped according to different functions and possible direct coding in Python language (Demsar et al. 2013).

This study considered the classification utilizing algorithms based on prediction models such as Decision Tree- DT, Random Forest- RF, Support Vector Machines- SVM, and Neural Networks- NW. The processing is described in Figure 3.

Decision Trees were built using the C4.5 classification algorithm proposed by Quinlan (1993). This algorithm selects the data attribute

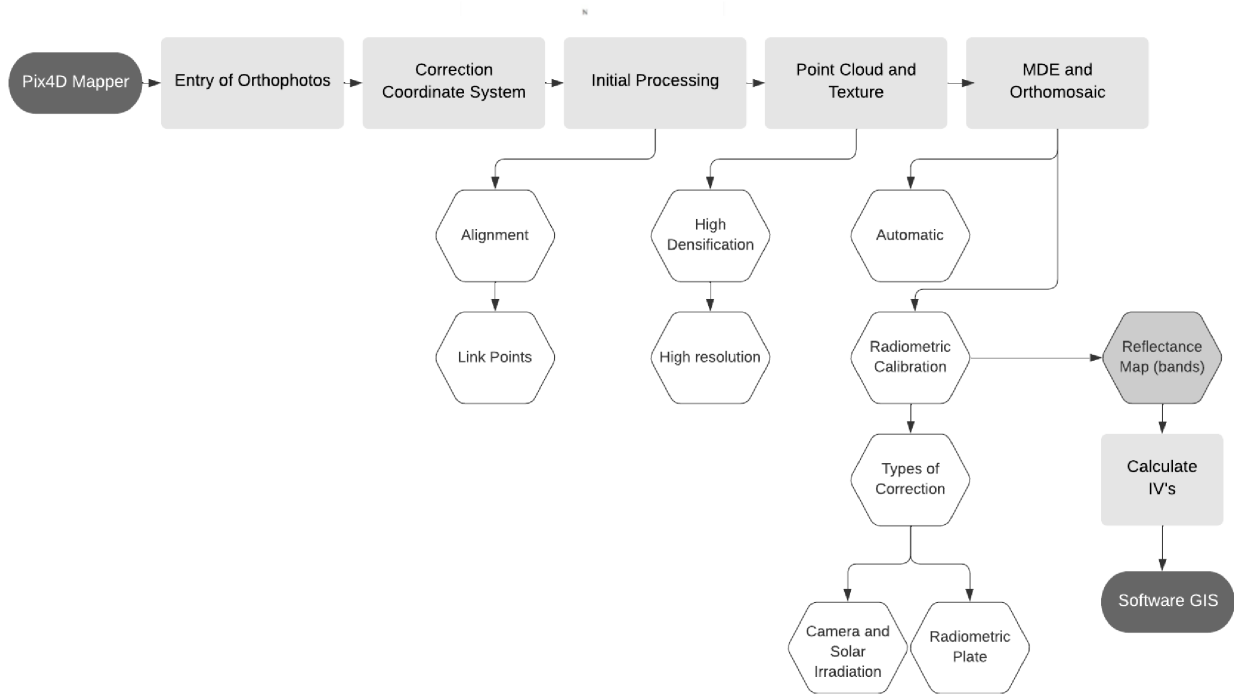


Figure 2. PIX4D processing flowchart.

by the utility of the attribute for the classification according to the reference of greatest gain of normalized information (GI), by the difference in entropy, partitioning the set of samples into subsets, and the sample with greater GI is chosen in the decision and then repeats the previous step in the smaller partitions to structure the whole Decision Tree (Witten & Frank 2011). Binary tree induction, the minimum number of instances on leaves of 2, no division of subsets smaller than 5, maximum depth of the tree of 10, and stop when the majority reaches 95% have been defined.

The Random Forest was built through the Random Forest classification algorithm (RF) proposed by Breiman (2001). This algorithm is based on the ensemble learning method and builds a set of Decision Trees. Each tree is developed using a self-initialization sample forming arbitrary subsets of attributes (bootstrapping) (Han et al. 2011). The final model generated for the rating case is based on the

majority vote of the individual trees generated (Breiman 2001). The number of trees in the forest was set at 10, the number of attributes to be arbitrarily drawn for consideration at each node at 5, and the number of attributes equal to the square root of the number of attributes in the data. The pruning pre-mowing was defined in not less than 10.

The Support Vector Machine was built through the LIBSVM package with C-SVC and nu-SVC classifier algorithm proposed by Chang & Lin (2011). This algorithm separates the attribute space with a hyperplane as a decision surface and maximizes the margin between the instances of different classes or class values performing the classification (Hastie et al. 2009). The minimization of the error function was defined as v-SVM (for cases of application in classification and regression), with cost in 1.00 and model parameter in 0.50. The Kernel function used was the RBF (radial base function) being the allowed deviation from the expected

Table I. Vegetation indices used, followed by their acronyms, equations, and references.

Vegetation Index	Acronyms	Equations	References
Normalized Difference Vegetation Index	NDVI	$(R_{NIR} - R_R) / (R_{NIR} + R_R)$	Rouse et al. (1973)
Index of the Standardized Difference - Red Edge	NDRE	$(R_{NIR} - R_{REG}) / (R_{NIR} + R_{REG})$	Buschmann & Nagel (1993)
First Modification to the Chlorophyll Absorption Ratio	MCARI1	$1,2[2,5(R_{NIR} - R_R) - 1,3(R_{NIR} - R_G)]$	Haboudane et al. (2004)
Red Edge Chlorophyll Index	CI	$(R_{NIR} / R_{REG}) - 1$	Gitelson et al. (2003)
Index of Chlorophyll Content in the Canopy	GCI	$(R_{NIR} / R_G) - 1$	Gitelson et al. (2005)

R_{NIR} , reflectance values obtained by the sensor in the near-infrared range; R_{REG} , reflectance in the range between red and infrared; R_R , reflectance in the red range; R_G , reflectance in the green range.

value in numerical tolerance of 0.0010 and Iteration Limit of 100.

The Neural Network was built through the Multilayer Perceptron classification algorithm (MLP) proposed by Rumelhart et al. (1986). This algorithm uses in its structure one or more hidden layers with an undetermined number of neurons being the training effected by backpropagation (Da Silva et al. 2010). It was defined 100 neurons per hidden layer of the net, being the activation function of the hidden layer ReLu that refers to the function of the rectified linear unit, stochastic optimizer based on a gradient as a solver for weight optimization, a parameter of penalty L2 (regularization term) and the maximum amount of iterations fixed at 200.

Performance metrics

The models of the algorithms were evaluated using the 10-part cross-validation method (10

fold cross-validation) with four metrics derived from the confusion matrix: (i) hit rate (accuracy); (ii) sensitivity; (iii) specificity; (iv) F1 score; and another metric complementary to the confusion matrix: (v) area of the ROC curve (AUC) used in the pre-processing and learning stages with training data set for the initial months (May, July, September, November 2019 and January 2020). In the prediction and evaluation stages, the percentage and difference of correctness between the three classes of study with test data sets of the proposed models for the last month (March 2020) of the study were verified.

RESULTS

The FreeViz method can be used to graph the multidimensional results. It selects, based on gradient descent modeling, the graphical optimization representation for compaction and separation between instances of the same class,

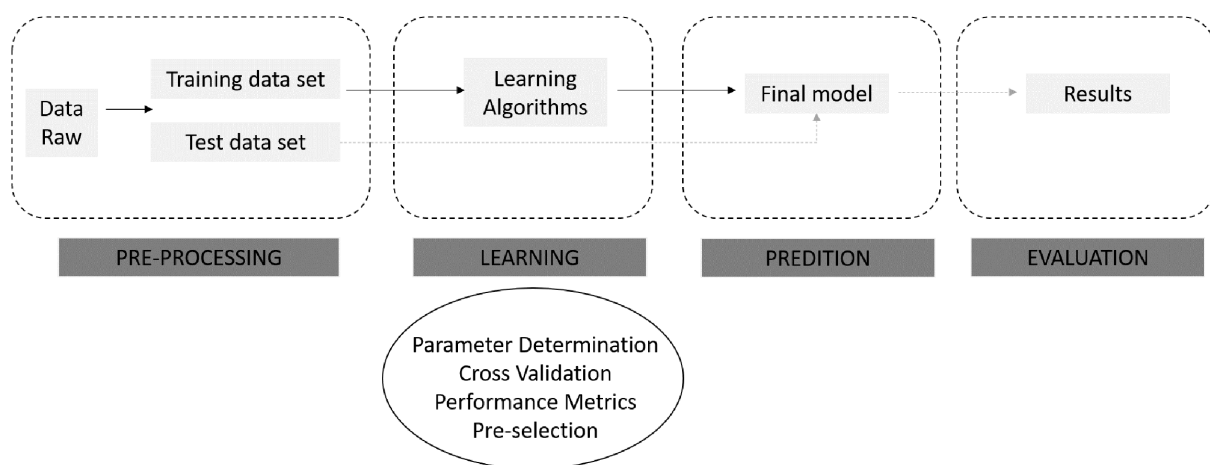


Figure 3. Orange Canvas 3.25.0 processing flowchart.

evaluated through mean scores (Rousseeuw 1987) as described in Figure 4.

According to the representation of FreeViz (Figure 4), the cultivars Catucaí and Catuaí present a greater similarity between themselves that differ slightly from the Bourbon cultivar for most analyzed variables. It is observed that a large part of the study variables emerged as a large cluster being the differentiation and individualization among coffee cultivars more efficient for the variables of height and crown diameter for the Bourbon coffee cultivar, for the variables of the Vegetation Indices MCARI1 and CI for the Catucaí coffee cultivar and the variables of Chlorophyll Content A and B for the Catuaí coffee cultivar.

The performance of the algorithms was evaluated from metrics derived from the confusion matrix, which registers in its rows and columns the errors and hits of the prediction for recently planted coffee cultivars as described in Figure 5.

According to the analysis of the confusion matrices (Figure 5) the cultivar Bourbon was always more confused with the cultivar Catuaí for all the algorithms tested. The cultivar Catucaí

was always more confused with the cultivar Catuaí for all the algorithms tested and the cultivar Catuaí was more confused with the cultivar Bourbon for the algorithms Decision Tree and Neural Network and more confused with the cultivar Catucaí for the algorithms Random Forest and Support Vector Machine.

Table II describes the performance metrics, applied to the training database set, using cross-validation in all methods, relative to hit rate (accuracy), sensitivity, specificity, F1 score, and area under the ROC curve (AUC) for the study classifying algorithms.

Table II shows that the algorithms were able to predict the classes of coffee cultivars with metric values from moderate to high, with excellent performance justified by the values of the study metrics, highlighting the Neural Networks algorithm with higher values for the study performance metrics, with a hit rate (accuracy) about 13% higher when compared to the algorithm with a lower hit rate (Decision Tree).

Figure 6 shows the ROC curve and its respective area (AUC) for the classification algorithms for each coffee cultivar under study.

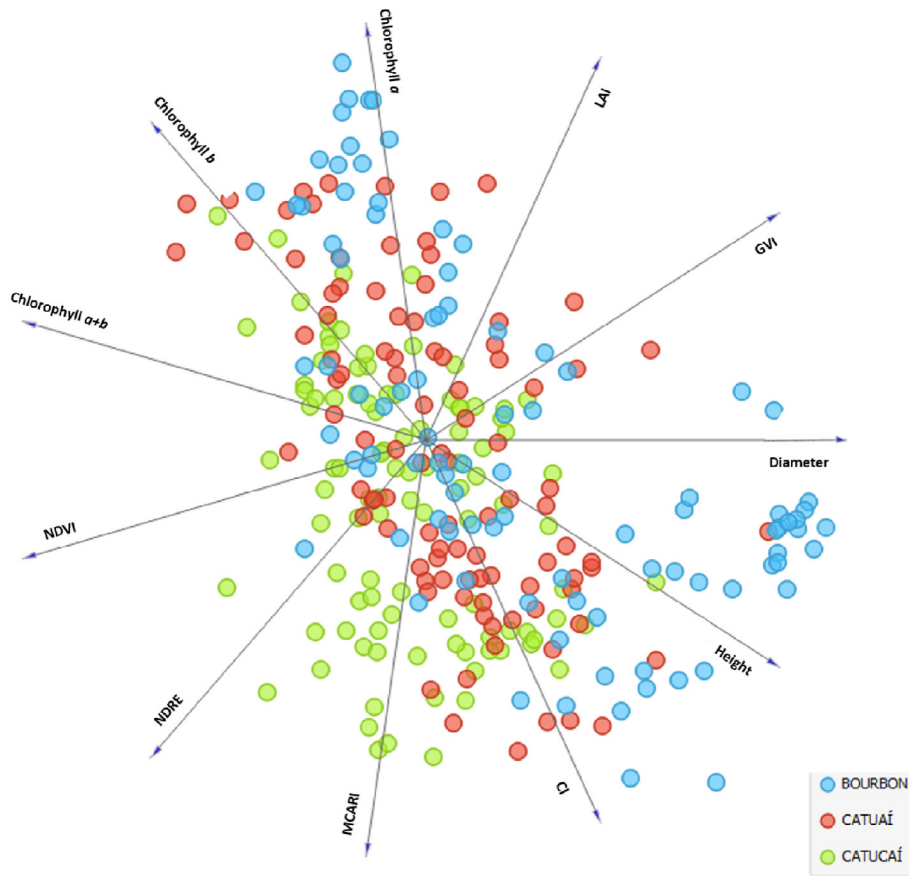


Figure 4. FreeViz representation of the dispersion of the simulated data for the study variables.

Table II. Performance metrics for each studied classifier algorithm.

Algorithms	Accuracy	Sensitivity	Specificity	F1	AUC
Decision Tree	0.768	0.769	0.768	0.793	0.857
Random Forest	0.836	0.837	0.836	0.836	0.938
Support Vector Machine	0.886	0.899	0.886	0.887	0.976
Neural Network	0.899	0.900	0.899	0.899	0.986

As demonstrated by the AUC values described in Table II, the performance evaluation of the models presented by the behavior described by the ROC curve and its respective area (AUC) considering, therefore, the individual study for each cultivar (Figure 6) highlights the Neural Network algorithm (green curve) as best and Decision Trees (purple curve) as worst suited for the prediction of coffee cultivars recently

planted from the study, being observed AUC ROC higher than 85% for all algorithms.

Table III shows the values in percent of the classes regarding the hits and differences of the hits for each algorithm according to the prediction and evaluation steps of the models with the test data set.

In Table III, the Decision Tree algorithm presented a higher error percentage,

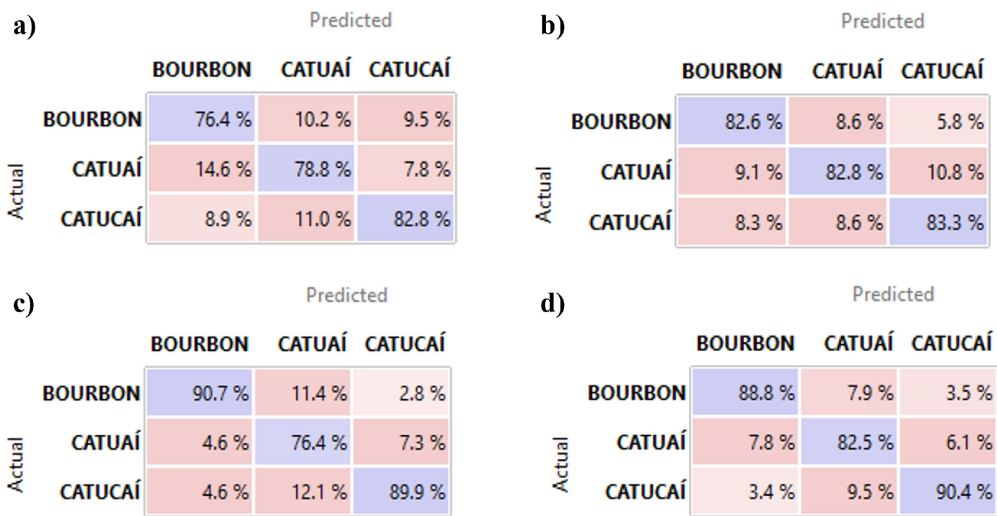


Figure 5. Confusion Matrices a) Decision Tree; b) Random Forest; c) Support Vector Machine; and d) Neural Network.

Table III. Percentage and difference of predictive performance hits for each classification algorithm for the actual percentage for the three study cultivars.

Algorithm/Cultivars	Catuaí	Catuaí	Bourbon
Real Class			
% Hit	33.898	33.898	32.203
Decision Tree			
% Hit	35.593	28.814	35.593
Difference of the Hit	-1.694	5.0843	-3.389
Random Forest			
% Hit	33.898	28.814	37.288
Difference of the Hit	0.0003	5.0843	-5.084
Support Vector Machine			
% Hit	33.898	32.203	33.898
Difference of the Hit	0.0003	1.695	-1.694
Neural Network			
% Hit	32.203	35.593	32.203
Difference of the Hit	1.695	-1.694	0.0003

underestimating the class’s prediction referring to the Catuaí cultivar and overestimating the classes referring to the Catucaí and Bourbon cultivars. The Random Forest and Support Vector Machine algorithms underestimated the class referring to the Catuaí cultivar and overestimated the class referring to the Bourbon cultivar, and almost totally hit the class referring to the Catucaí cultivar. The Neural Network algorithm

underestimated the class referring to Cultivar Catucaí, overestimated the class referring to Cultivar Catuaí, and almost totally hit the class referring to Cultivar Bourbon.

DISCUSSION

In this study, the classes must be recognized by the classification system, although they are only three classes, it is observed that the values of the study variables for each class are quite similar and/or overlapping, so they need to be effectively individualized for segmentation and classification by the proposed algorithm models.

FreeViz refers to an algorithm that optimizes a linear projection and displays the projected data in a scatter plot (Demsar et al. 2007). The procedure results in informative projections that are subject to simple interpretation with differentiation and separation of instance classes as proposed in this study for the differentiation of coffee cultivars.

According to Figure 4, cultivars Catucaí and Catuaí were close and differed from cultivar Bourbon for the study variables. As evidenced by Carvalho (2007) this fact is justified since the cultivars Catucaí and Catuaí present low to medium size and medium crown diameter,

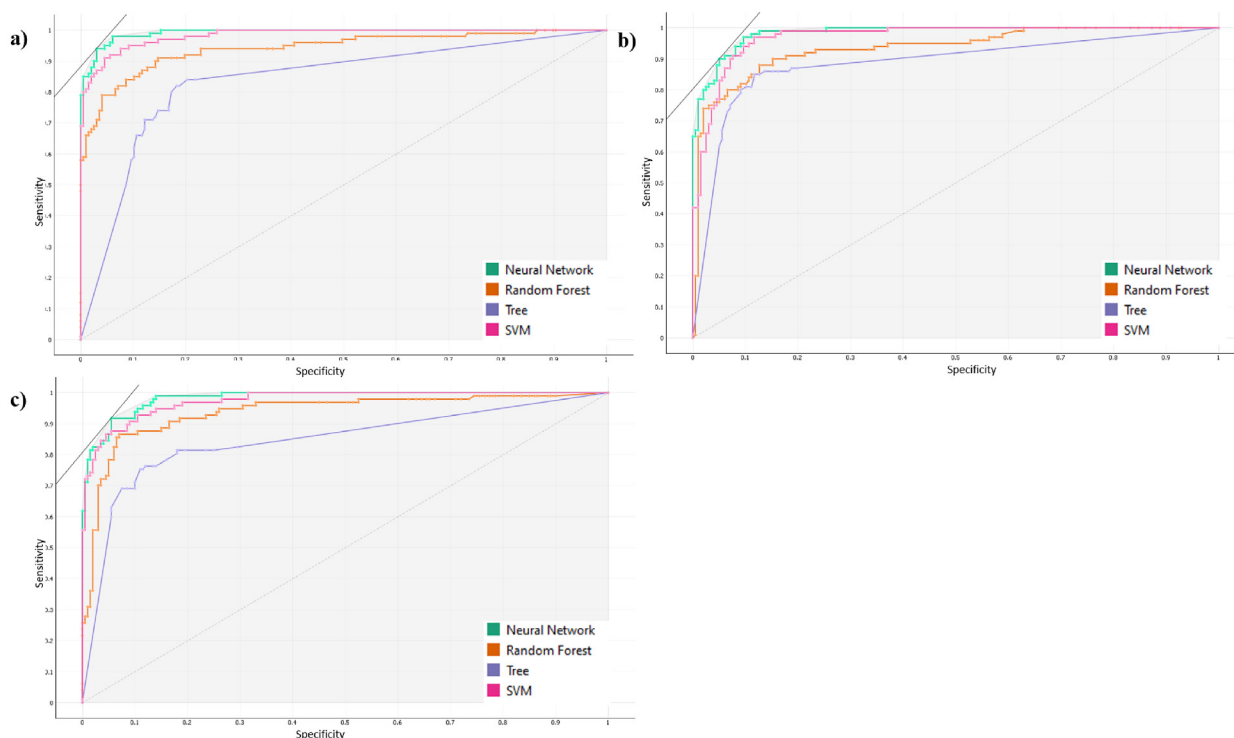


Figure 6. ROC and AUC curve a) Catucaí; b) Catuaí; and c) Bourbon.

and the Bourbon cultivar presents the high size and large crown diameter. Also, the Catucaí cultivar results from a cross between the Catuaí cultivar and the Icatu cultivar, which justifies great proximity between such coffee cultivars with a higher degree of overlap between them regarding the study variables. This fact of the proximity of morphological characteristics, mainly related to the aerial part of height and canopy diameter, of the cultivars Catucaí and Catuaí was also seen in studies by Ávila et al. (2020) and Veiga et al. (2020) and corroborating the results verified in this study.

According to the results described in the confusion matrices (Figure 5), it can be observed that the prediction hit percentages presented a lower rate of 76.4% for the Decision Tree algorithm to cultivate Bourbon and the same value for the Support Vector Machine algorithm to cultivate Catuaí and a higher rate of 90.7%

also in the Support Vector Machine algorithm to cultivate Bourbon. However, the more the prediction model responds correctly in comparison the more assertive the algorithm is, and in this case, the prediction rates found presented considerable values, which reinforces the good prediction of classification for the algorithms tested in this study.

Analyzing the confusion matrix, it can be observed that for all study algorithms for the classification of coffee cultivars the prediction percentage for the correct cultivar was considerably higher when compared to the error classes, demonstrating good representation of the proposed algorithms with great confidence with the reference data. Results for classification among coffee cultivars are not found in the literature, but in studies of classification of land use areas, which include coffee areas, the percentages of prediction correctness by the

confusion matrix were superior to the studies of Silveira et al. (2016) and Andrade et al. (2013).

It is observed in Table II that the algorithms were able to predict the classes of coffee cultivars with metric values from moderate to high, thus resulting in optimum performance justified by the values of the study metrics. For the training database on the prediction of the three recently planted coffee cultivars (Table II), the model trained by the Neural Network algorithm obtained better results, with higher values found for all performance metrics tested, being accuracy of 0.899, sensitivity of 0.900, specificity of 0.899, F1 of 0.899 and AUC of 0.986. The other algorithms also presented good values to the performance metrics, with values close to those obtained by the Neural Network models, being observed the largest metrics for the Vector Support Machine, Random Forest and Decision Tree respectively and the possible use of such algorithms, since they presented moderate to high values for the performance metrics considered in this study.

The accuracy or general error rate represents the agreement between the predicted and observed study classes and determines classifiers' specific performance according to the classes of the response of interest (James et al. 2013). It should be noted that accuracy values between 70 and 100% represent satisfactory results from moderate to high for classification (Kuhn & Johnson 2013), as seen in this study.

However, the accuracy must be combined with other metrics that affect the appropriate choice of classifier (James et al. 2013, Kuhn & Johnson 2013). We used the sensitivity metrics that refer to the proportion of true positives among the instances classified as positive and specificity the proportion of false negatives among the instances classified as negative. The metric score F1 takes into account, in turn, the harmonic mean between the metrics of

sensitivity and specificity, thus excluding the possible determination of metrics that include true and false values (positive and negative), and the higher the percentage values, the better the prediction made by the classifier (Geron 2019), as evidenced in all the values of such metrics in this study.

Works using different algorithms are applied for the classification and mapping of coffee areas. Souza et al. (2016) found good applicability and adequate performance metrics for the Support Vector Machine algorithm; Li et al. (2014) also found promising results using classifier algorithms for land use mapping in a region of China; Hussain et al. (2014) demonstrated that the Decision Tree algorithm could show disadvantages, justifying lower performance metrics, as the tree can contain many branches, which makes interpretation of the classification difficult. The high values for performance metrics demonstrated in this study in the classification of coffee cultivars is mainly justified using high-resolution images and spectral vegetation indices that highlight essential characteristics of the vegetation, since spectral information is essential to obtain higher accuracy.

The Receiver Operating Characteristic (ROC) curve is used to evaluate the accuracy performance based on the shortest distance between sensitivity and specificity resulting from the cutoff points, and the ideal ROC curve tends to the upper left corner in the graphical representation (Gonzaga 2011). The area under the ROC curve (AUC) quantifies the discriminatory power of a model and thus orders the risks from the highest to the lowest to the study models. However, for the ROC and AUC curve, the performance is represented in the range of normalized limits between 0 and 1 (James et al. 2013, Kuhn & Johnson 2013).

As shown in Figure 6 with the representation of the ROC curve, the performance of the classifier indicates the Neural Network algorithm (green curve) as the best and Decision Trees (purple curve) as the worst fit for prediction of newly transplanted coffee cultivars under study. However, it is noteworthy that both values are acceptable evidencing discrimination ability for the prediction and classification proposed.

Although the models have been adjusted with results presented by the relevant performance metrics indicating good adjustment performance, they can generate inaccurate predictions. So, it is necessary to apply the evaluation step of the models, which consists of evaluating the behavior of new data to verify the correct prediction for the models, data that were not used to develop the model and definition of parameters.

The Neural Network and Support Vector Machine algorithms presented similar prediction performance regarding the percentage and difference of hits, changing, however, the total hit classes, evidencing what is observed in Table II regarding the performance metrics for each study classifier algorithm, with great proximity of the values for the two algorithms in question. However, it is important to point out that the Neural Network algorithm presents more coherence than expected and observed in the field, since the classes referring to the cultivars Catucaí and Catuaí have been more confused among themselves. This fact was previously highlighted by the proximity of the characteristics between such cultivars, regarding height, canopy diameter and Leaf Area Index, differently from that observed for the class referring to Bourbon cultivar, which in the Support Vector Machine algorithm was more confused with the class referring to Catuaí cultivar, a fact not commonly observed by the behavior of coffee cultivars.

The validation of pattern recognition by artificial intelligence algorithms is essential, as it guarantees the application of external data to the algorithm's testing and training stage, showing the applicability and obtaining adequate results for intelligent decision-making.

Studies using prediction and recognition algorithms are applicable in various fields of coffee production, in which the classification of coffee beans samples is cited (Oyama et al. 2013), the occurrence of rust (Souza et al. 2016), identification and classification of foliar disease varieties (Sasirekha & Swetha 2015), the incidence of pests and diseases (Aparecido et al. 2020) and coffee quality (Suarez-Peña et al. 2020). On the other hand, the current application for estimating yield and consequent prediction of the coffee crop through machine learning models are use images to recognize patterns of biomass estimation and correlation with productivity (Nascimento 2019) and algorithms for detection of fruits and classify it according to its maturation (Kazama 2019), therefore dealing with great applications for forecasting studies of coffee crops.

Thus, it was possible to verify that all the algorithms used in this study to predict coffee cultivars in the first year of fixation in the post-planted field were satisfactory. However, the Neural Network algorithm presented better values the performance metrics of the study in the prediction and evaluation phases of the models, and its use is indicated for the objective proposed in this study. With the greater diffusion of applications in the field, digital agriculture combined with machine learning techniques and prediction algorithms can be essential for the optimization of activities in the field. They can therefore be applied with various objectives of agricultural studies, providing more assertive and appropriate decision-making, based on the recognition of patterns of the analyzed data,

thus encouraging further studies on this topic in the agricultural field, expanding the diffusion of works, which includes the coffee culture.

CONCLUSIONS

- Catucaí, Catuaí, and Bourbon coffee cultivars were satisfactorily predicted in the first year after planting based on the evaluated performance metrics. The algorithm model based on Neural Networks produced more accurate predictions than other algorithm models tested, with a higher percentage of hits for the classes considered.

- Prediction of agricultural cultivars are of fundamental importance for agricultural producers, especially in cultivars with perennial characteristics, since they remain for a long time fixed in the field, it is necessary to correctly identify for intelligent decision making and application of efficient management variant between cultivars, mainly in the study of coffee, due to the economic importance of the commodity.

Acknowledgments

We would like to thank the Consórcio Brasileiro de Pesquisa do Café - Embrapa for the financial support of this study. We would also like to thank the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), the Universidade Federal de Lavras (UFLA) and the Samambaia Farm for all the support to this research.

REFERENCES

ALVARES CA, STAPE JL, SENTELHAS PC, DE MORAES GONÇALVES JL & SPAROVEK G. 2013. Köppen's climate classification map for Brazil. *Meteorol Z* 22(6): 711-728.

ALVES HMR, VIEIRA TGC, VOLPATO MML, LACERDA MPC & BORÉM FM. 2016. Geotechnologies for the characterization of specialty coffee environments of Mantiqueira de Minas in Brazil. In: THE INTERNATIONAL SOCIETY FOR PHOTOGRAMMETRY AND REMOTE SENSING, v. 23.

ANDRADE LN, VIEIRA TGC, LACERDA WS, VOLPATO MML & DAVIS JR CA. 2013. Application of artificial neural networks in the classification of coffee areas in Machado-MG. *Coffee Sci* 8(1): 78-90. ISSN 1984-3909.

APARECIDO LE DE O, DE SOUZA ROLIM G, DE MORAES JR DSC, COSTA CTS & DE SOUZA PS. 2020. Machine learning algorithms for forecasting the incidence of *Coffea arabica* pests and diseases. *Int J Meteorol* 64(4): 671-688.

atLEAF. 2019. atLEAF chlorophyll meter. Available in: <https://www.agriculturesolutions.com/atleaf-digital-chlorophyll-meter>. Access in: May. 2020.

ÁVILA EAS, SOUSA CM, PEREIRA W, ALMEIDA VG, SARTI JK & SILVA DP. 2020. Growth and Productivity of Irrigated Coffee Trees (*Coffea arabica*) in Ceres-Goiás. *J Agric Sci* 12(2).

BORSATO D, PINA MVR, SPACINO KR, DOS SANTOS SCHOLZ MB & ANDROCIO FILHO A. 2011. Application of artificial neural networks in the geographical identification of coffee samples. *Eur Food Res Technol* 233(3): 533.

BREIMAN L. 2001. Random forests. *Machine Learning* 45(1): 5-32.

BUSCHMANN C & NAGEL E. 1993. In vivo spectroscopy and internal optics of leaves as basis for remote sensing of vegetation. *Int J Remote Sens* 14(4): 711-722.

CARVALHO CHS DE. 2007. *Cultivares De Café*. / (Ed) Brasília: EMBRAPA, 247 p.

CHANG CC & LIN CJ. 2011. LIBSVM: A library for support vector machines. *ACM transactions on intelligent systems and technology (TIST)* 2(3): 1-27.

CHEMURA A & MUTANGA O. 2017. Developing detailed age-specific thematic maps for coffee (*Coffea arabica* L.) in heterogeneous agricultural landscapes using random forests applied on Landsat 8 multispectral sensor. *Geocarto Int* 32(7): 759-776.

CHLINGARYAN A, SUKKARIEH S & WHELAN B. 2018. Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. *Comput Electron Agric* 151: 61-69.

CONAB - COMPANHIA NACIONAL DE ABASTECIMENTO. 2021. Acompanhando a colheita brasileira: café - v. 6, n. 3 (2020) - Brasília: Conab, 2020. ISSN 2318-7913. Available in: <https://www.conab.gov.br/>. Access in: Oct. 2020.

DA SILVA IN, SPATTI DH & FLAUZINO RA. 2010. Artificial Neural Networks for engineering and applied sciences practical course. São Paulo: Artliber.

DA SILVA IN, SPATTI DH, FLAUZINO RA, LIBONI LHB & DOS REIS ALVES SF. 2017. Artificial neural network architectures and

- training processes. In *Artificial neural networks* Springer, Cham 1: 21-28.
- DEMSAR J ET AL. 2013. Orange: data mining toolbox in Python. *J Mach Learn Res* 14(1): 2349-2353.
- DEMSAR J, LEBAN G & ZUPAN B. 2007. FreeViz - An intelligent multivariate visualization approach to explorative analysis of biomedical data. *J Biomed Inform* 40(6): 661-671.
- DJI. 2015. UAV MATRICE 100. Available in: <https://www.dji.com/br/matrice100/>. Access in: Apr. 2020.
- EMBRAPA - EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA. 2018. Sistema brasileiro de classificação de solos / Humberto Gonçalves dos Santos ... [et al.]. - 5. ed., rev. e ampl. Brasília, DF.
- EMBRAPA - EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA. 2020. Monitoramento da safra brasileira de café. 2020. Terceiro inquérito, Brasília 6: 1-54.
- ESRI. 2018. ArcGIS: SOFTWARE. Available in: <http://www.esri.com/software/arcgis/index>. Access in: May. 2020.
- FAVARIN JL, DOURADO NETO D, GARCÍA Y GARCÍA A, VILLA NOVA NA & FAVARIN MDGGV. 2002. Equações para a estimativa do índice de área foliar do cafeeiro. *Pesqui Agropecu Bras* 37(6): 769-773.
- FERNANDES FT & CHIAVEGATTO FILHO ADP. 2019. Perspectives on the use of data mining and machine learning in occupational health and safety. *Braz J Occup Ther*, 44.
- FERRAZ GA, SILVA FMD, OLIVEIRA MSD, CUSTÓDIO AAP & FERRAZ PFP. 2017. Spatial variability of plant attributes in a coffee plantation. *Agron Sci* 48(1): 81-91.
- GÉRON A. 2019. Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems. O'Reilly Media, 210 p.
- GITELSON AA, VIÑA A, ARKEBAUER TJ, RUNDQUIST DC, KEYDAN G & LEAVITT B. 2003. Remote estimation of leaf area index and green leaf biomass in maize canopies. *Geophys Res Lett* 30(5).
- GITELSON AA, VINA A, CIGANDA V, RUNDQUIST DC & ARKEBAUER TJ. 2005. Remote estimation of canopy chlorophyll content in crops. *Geophys Res Lett* 32(8).
- GONZAGA A. 2011. Methods for evaluating classifiers, 111 p.
- HABOUDANE D, MILLER JR, PATTEY E, ZARCO-TEJADA PJ & STRACHAN IB. 2004. Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. *Remote Sens Environ* 90(3): 337-352.
- HAN J, PEI J & KAMBER M. 2011. Data mining: concepts and techniques. Elsevier.
- HASTIE T, TIBSHIRANI R & FRIEDMAN J. 2009. The elements of statistical learning: data mining, inference, and prediction. Springer Science & Business Media.
- HUNT DA ET AL. 2020. Review of Remote Sensing Methods to Map Coffee Production Systems. *Remote Sens* 12(12): 2041 p.
- HUSSAIN A, BHALLA P & PALRIA S. 2014. Remote sensing based analysis of the role of land use/land cover on surface temperature and temporal changes in temperature a case study of Ajmer district, Rajasthan- The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XL-8, 2014, ISPRS Technical Commission VIII Symposium, 9-12 December 2014, Hyderabad, India.
- JAMES G, WITTEN D, HASTIE T & TIBSHIRANI R. 2013. An introduction to statistical learning. New York: Springer, 112 p.
- KAWAKUBO FS & PÉREZ MACHADO RP. 2016. Mapping coffee crops in southeastern Brazil using spectral mixture analysis and data mining classification. *Int J Remote Sens* 37(14): 3414-3436.
- KAZAMA EH. 2019. Prescription harvesting for coffee, is it possible? (Doctoral Thesis Faculty of Agrarian and Veterinary Sciences) - Universidade Estadual Paulista (Unesp), Jaboticabal- São Paulo.
- KELLEY LC, PITCHER L & BACON C. 2018. Using Google Earth engine to map complex shade-grown coffee landscapes in Northern Nicaragua. *Remote Sens* 10(6): 952.
- KUHN M & JOHNSON K. 2013. Applied predictive modeling. New York: Springer, Vol. 26.
- LEME DS, DA SILVA SA, BARBOSA BHG, BORÉM FM & PEREIRA RGFA. 2019. Recognition of coffee roasting degree using a computer vision system. *Comput Electron Agric* 156: 312-317.
- LIAKOS KG, BUSATO P, MOSHOU D, PEARSON S & BOCHTIS D. 2018. Machine learning in agriculture: A review. *Sensors* 18(8): 2674.
- MACIEL DA ET AL. 2020. Leaf water potential of coffee estimated by landsat-8 images. *PLoS ONE* 15(3): e0230013.
- MAPA - MINISTÉRIO DA AGRICULTURA. 2018. Pecuária e Abastecimento. Dados sobre a agricultura cafeeira. Available in: <https://www.gov.br/agricultura/pt-br/> Access in: Oct. 2020.

- MARIN DB ET AL. 2021. Remotely Piloted Aircraft and Random Forest in the Evaluation of the Spatial Variability of Foliar Nitrogen in Coffee Crop. *Remote Sens* 13(8): 1471.
- MARUJO R DE FB, MOREIRA MA, VOLPATO MML & ALVES HMR. 2017. Coffee crop detection by automatic classification using spectral and textural attributes and illumination factor. *Coffee Sci* ISSN 1984-3909, [S. L.] 12(2): 164-175.
- MEIRA CAA, RODRIGUES LHA & MORAES SAD. 2009. Modelos de alerta para o controle da ferrugem do cafeeiro em áreas de cultivo com grande carga de frutos. *Pesqui Agropecu Bras* 44(3): 233-242.
- MESQUITA CD ET AL. 2016. Manual do café: implantação de cafezais *Coffea arabica* L. Belo Horizonte: EMATER-MG, v. 50.
- MICASENSE PARROT SA. 2016. Available in: <https://www.parrot.com/global/> Access in: Oct. 2020.
- MINCATO RL, PARREIRAS TC, LENSE GHE, MOREIRA RS & SANTANA DB. 2020. Using unmanned aerial vehicle and machine learning algorithm to monitor leaf nitrogen in coffee. *Coffee Sci* ISSN 1984-3909, [S. L.] 15: e151736.
- NASCIMENTO ALD. 2019. Estimativa de produtividade de café usando métodos de aprendizado de máquina. (Agricultural Engineering Doctoral Thesis) - Universidade Federal de Viçosa (UFV), Viçosa- Minas Gerais.
- OLIVEIRA AJ, ASSIS GA, GUIZILINI V, FARIA ER & SOUZA JR. 2019. Segmenting and Detecting Nematode in Coffee Crops Using Aerial Images. In *International Conference on Computer Vision Systems*. Springer, Cham, p. 274-283.
- OYAMA PIC, JORGE LAC & RODRIGUES ELL. 2013. Methodology to Classify Coffee Beans Samples through Shape, Colour and Texture Descriptors, IX Workshop de Visão Computacional, Rio de Janeiro, RJ.
- PADILLA FM, DE SOUZA R, PEÑA-FLEITAS MT, GALLARDO M, GIMENEZ C & THOMPSON RB. 2018. Different responses of various chlorophyll meters to increasing nitrogen supply in sweet pepper. *Front Plant Sci* 9: 1752.
- PIX4D - PIX4D MAPPER. 2019. Available in: <https://www.pix4d.com/product/pix4dmapper-photogrammetry-software>. Access in: Oct. 2020.
- PRECISION HAWK. 2010. Precision Flight. Available in: <https://www.precisionhawk.com/>. Access in: Oct. 2020.
- QUINLAN JR. 1993. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers Inc., San Mateo, CA, USA.
- ROUSE JW, HAAS RH, SCHELL JA & DEERING DW. 1973. Monitoring vegetation systems in the Great Plains with ERTS In *Third Earth Resources Technology Satellite-1*. December. Goddard Space Flight Center: NASA, p. 309-317.
- ROUSSEEUW PJ. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math* 20: 53-65.
- RUMELHART DE, HINTON GE & WILLIAMS RJ. 1986. Learning representations by back-propagating errors. *Nature* 323(6088): 533-536.
- SARRI D, MARTELLONI L & VIERI M. 2017. Development of a prototype of telemetry system for monitoring the spraying operation in vineyards. *Comput Electron Agric* 142: 248-259.
- SASIREKHA N & SWETHA N. 2015. An Identification of Variety of Leaf Diseases Using Various Data Mining Techniques, *Int J Adv Res Comput Commun Eng* 4(10).
- SILVEIRA LSD, VALENTE DSM, PINTO FDAC & SANTOS FL. 2016. Case studies of classification of areas cultivated with coffee using texture descriptors. *Coffee Sci* 11(4): 15.
- SOUZA CG, ARANTES TB, CARVALHO LMTD & AGUIAR P. 2019. Variáveis multitemporais para o mapeamento de áreas de cultivo de café. *Pesqui Agropecu Bras*. v. 54.
- SOUZA CG, CARVALHO L, AGUIAR P & ARANTES TB. 2016. Machine learning algorithms and remote sensing variables for coffee crop mapping. *Geodetic Sci Bul* 22: 751-773.
- SUAREZ-PEÑA JA, LOBATON-GARCÍA HF, RODRÍGUEZ-MOLANO JI & RODRIGUEZ-VAZQUEZ WC. 2020. Machine Learning for Cup Coffee Quality Prediction from Green and Roasted Coffee Beans Features. In *Workshop on Engineering Applications*. Springer, Cham, p. 48-59.
- VAN KEULEN H & ASSENG S. 2019. Simulation models as tools for crop management. *Crop Sci*, p. 433-452.
- VEIGA A, GUERRA A, BARTHOLO G, ROCHA O, RODRIGUES G & CARVALHO MDF. 2020. Desempenho agrônômico de genótipos de café arábica resistentes à ferrugem no Cerrado Central. *Embrapa Cerrados-Boletim Pesquisa e Desenvolvimento (INFOTECA-E)*.
- WITTEN IH & FRANK E. 2011. Data mining: practical machine learning tools and techniques with Java implementations. *Acm Sigmod Record* 31(1): 76-77.

How to cite

BENTO NL, FERRAZ GAES, BARATA RAP, SOARES DV, TEODORO SA & ESTIMA PHO. 2023. Classifiers based on artificial intelligence in the prediction of recently planted coffee cultivars using a Remotely Piloted Aircraft System. *An Acad Bras Cienc* 95: e20210534. DOI 10.1590/0001-3765202320210524.

*Manuscript received on April 9, 2021;
accepted for publication on May 19, 2022*

NICOLE L. BENTO

<https://orcid.org/0000-0002-0432-7655>

GABRIEL ARAÚJO E.S. FERRAZ

<https://orcid.org/0000-0001-6403-2210>

RAFAEL ALEXANDRE P. BARATA

<https://orcid.org/0000-0002-6575-2364>

DANIEL V. SOARES

<https://orcid.org/0000-0001-6566-6960>

SABRINA A. TEODORO

<https://orcid.org/0000-0002-3249-9147>

PEDRO HENRIQUE DE O. ESTIMA

<https://orcid.org/0000-0002-4704-3257>

Universidade Federal de Lavras/UFLA, Departamento de Engenharia, Avenida Aqueanta Sol, Caixa Postal 3037, 37200-000 Lavras, MG, Brazil

Correspondence to: **Nicole Lopes Bento**

E-mail: nicolelbento@gmail.com

Author contributions

Conceived and designed the experiments: NLB, GASF and DVS. Performed the experiments: NLB, RPB, SAT, PHDOE. Analyzed the data: NLB, RPB and GASF. Contributed reagents/materials/analysis tools: NLB, GASF and DVS. Wrote the paper: NLB, GASF and RPB. All authors read and approved the final manuscript.

