Estimating coffee crop parameters through multispectral imaging and machine learning algorithms

Fernando Vasconcelos Pereira[1], Vinicius Silva Werneck Orlando[1], George Deroco Martins[2], Bruno Sérgio Vieira[2], Eduardo Soares Nascimento[1], Aline Barrocá Marra[1], Maria de Lourdes Bueno Trindade Galo[1]

[1] São Paulo State University (UNESP), Presidente Prudente, São Paulo, Brazil - (fernando.v.pereira, vinicius.werneck, e.nascimento, aline.barroca, trindade.galo)@unesp.br
[2] Federal University of Uberlândia (UFU), Uberlândia, Minas Gerais, Brazil – deroco@ufu.br, brunovieira@iciag.ufu.br

Keywords: Multispectral Images, Machine Learning, Coffee Crop, Productivity Indicators, Plant Height, Canopy Diameter.

Abstract:

Brazil plays a crucial role in the global economy due to its significant contribution to the agricultural sector, particularly in coffee production, where it stands out as the largest producer and exporter of processed coffee. Various disturbances can influence coffee plants, causing abnormalities that can hinder their successful growth. Parameters such as plant height and canopy diameter play an essential role in assessing the health and productivity of the plants, reflecting their growth, development, and ability to capture sunlight. Additionally, height is also related to the balanced distribution of nutrients and water, providing valuable information about overall performance and the capacity for healthy production. In this regard, the application of methodologies involving remote sensing and machine learning algorithms has shown promising results in the rapid and safe acquisition of information about agricultural systems. This study evaluates different machine learning algorithms, using radiometric values from multispectral images obtained by remote sensing platforms as input datasets for estimating plant height and canopy diameter in coffee cultivation. The best performance was observed for architectures that showed lower RMSE and RMSE% values. For the plant height parameter (m), the RGB sensor exhibited the best performance using the Random Tree algorithm, with an RMSE (0.27) and RMSE% (8.80). For the canopy diameter (m), the sensor showed the best performance using the Random Forest algorithm, with an RMSE (0.15) and RMSE% (8.16).

1. Introduction

Brazil plays a crucial role in the global economy due to its significant contribution to the agricultural sector, especially in coffee production, where it currently stands out as the largest producer and exporter of processed coffee.

The Brazilian state of Minas Gerais (MG) produces over 50% of the country's entire coffee crop. The quality of coffee produced in Minas Gerais is globally recognized, making it the main export commodity of the agricultural sector in the state. Additionally, coffee production in the state is extremely important for generating employment in national agriculture and as a national source of income (EMATER, 2018).

Various disturbances can influence coffee plants, causing abnormalities that can hinder their successful growth. Through regular monitoring of coffee plant development, Martinez et al. (2007) demonstrated that productivity and crop longevity are directly correlated with certain plant parameters, such as canopy diameter and plant height.

Parameters such as plant height and canopy diameter play an essential role in assessing plant health and productivity. These parameters reflect plant growth, development, and the ability to capture sunlight. Moreover, height is also related to the balanced distribution of nutrients and water, providing valuable information about overall performance and healthy production capacity.

However, it is crucial to emphasize that the sustainability of Brazilian coffee farming requires the adoption of responsible agricultural practices and precise methodologies. In this context, remote sensing has proven to be efficient in detecting changes in coffee plantations to obtain reliable and timely agricultural statistics for remote management of large areas in a non-destructive manner.

In addition, advanced machine learning techniques are being used to create models that relate agricultural productivity to various factors that impact crop growth (Bocca and Rodrigues, 2016). These techniques explore different variables to develop these models.

Many studies have been conducted to evaluate parameters related to agricultural crops, both for monitoring and estimation, using data extracted from multispectral images through machine learning algorithms. These studies employ remote sensing to obtain detailed information on crop development, nutritional conditions, and the detection of diseases and pests, among other aspects (Pereira et al., 2022; Ndikumana et al., 2018; Wang et al., 2016; Arantes et al., 2021; Oliveira, 2022).

The application of methodologies involving remote sensing has shown promising results in facilitating the rapid and secure acquisition of information about agricultural systems, as obtaining up-to-date and reliable data can be challenging given the spatio-temporal dynamics of these systems (Barros, 2021). In this sense, the correlation between information acquired through remote sensing platforms and parameters related to productivity offers benefits not only for large-scale predictions but also for understanding relevant aspects related to spatial variability of productivity in agricultural systems (Bernardes, 2013).

Given the context described, the main aim of this study was to evaluate the potential of different machine learning algorithms using radiometric values extracted from multispectral images obtained through remote sensing platforms as input datasets for estimating coffee plant height and canopy diameter

2. Materials and Methods

2.1 Study Area

The experiment was conducted in an area located in the municipality of Monte Carmelo-MG, in the Mesoregion of Triângulo Mineiro and Alto Paranaíba, with an approximate area of 15,113 m² and average altitude of 826 meters (Figure 1). The climatic conditions of the study area are classified as Aw, according to the Köppen classification, indicating a hot and humid summer and a cold and dry winter.

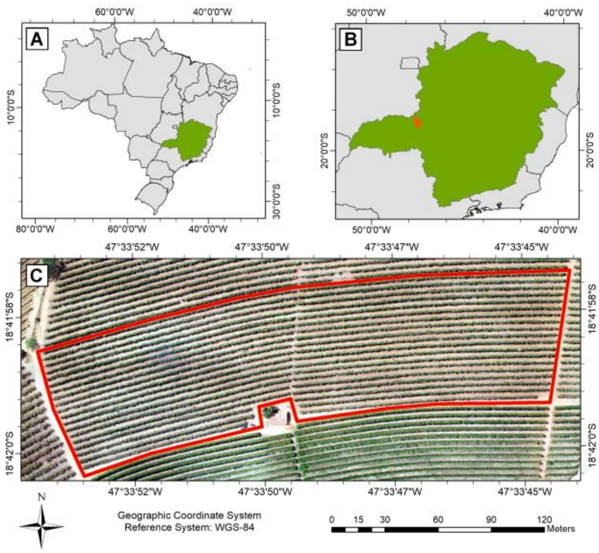

Figure 1. Location of the study area. (A) Minas Gerais State (MG) in Brazil. (B) Monte Carmelo-MG, located in the Triângulo Mineiro e Alto Paranaíba mesoregion. (C) Experimental area, highlighted in red.

2.2 Acquisition of agronomic parameters

The experiment was conducted in a commercial plot with an area of 15,113 m2, cultivating *Coffea arabica* L. cv. Yellow Bourbon, since 2013. The spacing between rows in the area was 3.8 m, and the spacing between plants within rows was 0.7 m.

A randomized complete block design (RCBD) was adopted for the experiment, consisting of 5 blocks with 12 plots each. Each block was separated by two rows of coffee plants, and each experimental plot consisted of 32 coffee plants. Among these, the three central plants of each plot were evaluated, resulting in a total of 165 plants assessed. Both the experimental area and the plots and evaluated plants were georeferenced using coordinates obtained from a dual-frequency GNSS receiver model Hiper V (L1/L2) through the Real-Time Kinematic (RTK) positioning method.

The plots were delimited using striped tape, clearly marking the beginning and end of each plot. For the three evaluated plants, small flags were fixed in the ground in front of the central plant to facilitate their location. Plant height (m) was measured from the base of the plant to the apical bud of the orthotropic branch, using a centimeter-graded aluminum fitting target and a 5-meter stadia rod. Canopy diameter (m) was measured from aerial multispectral images after geometric correction. The collected sample data were measured and evaluated on the same plants in all 9 assessments. After data collection, the values were tabulated, and the analyzed values correspond to the arithmetic mean of the three plants evaluated per plot.

2.3 Acquisition of multispectral images

During sample collection, an aerial survey was carried out using a Remotely Piloted Aircraft (RPA) model Phantom 4, equipped with two coupled cameras. The first camera (Complementary Metal Oxide Semiconductor sensor) integrated into the aircraft operated in the visible range of the spectrum, capturing images in Red (650 nm), Green (550 nm) and Blue (480 nm) colors (RGB), with a resolution of 20 megapixels and GSD (Ground Sample Distance) 1 cm per pixel (cm/px). The second camera, the Mapir Survey3W, contained the Red (660 nm), Green (550 nm) and Near InfraRed (NIR, 850 nm) spectral bands (RGN), with a resolution of 12 megapixels and a GSD of 5,519 cm/px. In the DroneDeploy© software, the flight was planned with an altitude set at 100 meters, speed of 10 m/s, longitudinal and lateral overlap of 70%, and flight time of approximately 6 minutes (Figure 2).

After the flight, the orthomosaics were generated and georeferenced based on identifiable points in the field, using the Pix4D software, educational version. Finally, the orthomosaics were radiometrically normalized using the ENVI 5.0 software.

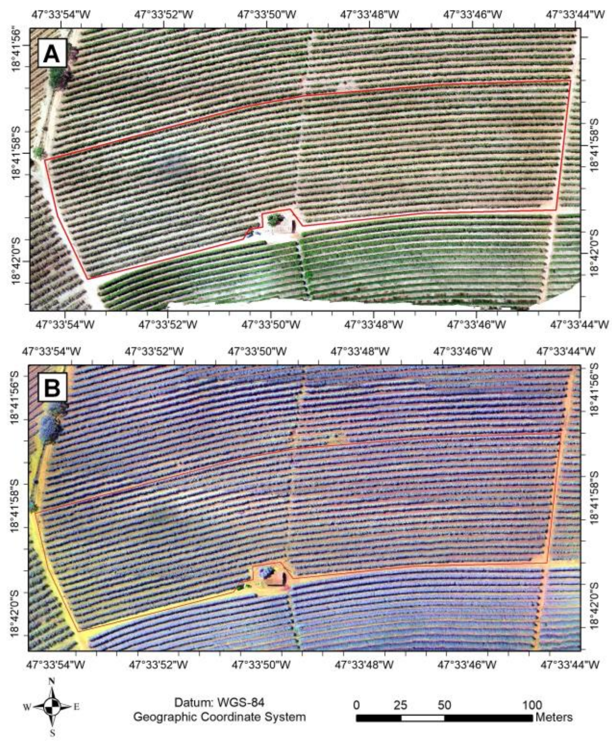

Figure 2. Color compositions of the study area generated from images captured by the sensors (A) RGB (true color) and (B) RGNIR (false color).

2.4 Construction of estimation models

After data acquisition, table-format files were created containing the values of plant height, canopy diameter, and radiometric values extracted from the multispectral images. Shapefiles were created with vectorized polygons around each

https://doi.org/10.5194/isprs-annals-X-3-2024-317-2024

plant evaluated in the field to extract average radiometric values using the Region of Interest (ROI) tool in ENVI 5.0 software. The radiometric values for each band were automatically extracted and calculated by the software itself, considering the mean values of the pixels within the vectorized polygons of each plot.

This dataset was used as input data for the prediction models. A total of ten supervised classification algorithms available in the Weka 3.9.4 software library were trained: Linear Regression, Multilayer Perceptron Neural Network, Simple Linear Regression, SMOreg, M5Rules, M5P, Random Forest, and Random Tree.

After several tests, the default adjustment parameters provided by the Weka 3.9.4 software were used for all algorithms. The training method employed was the Supplied Test Set, which utilizes an approach where a separate dataset, known as the test set, is provided by the user to evaluate the performance of the trained model. This test set generally contains data that was not used during the model training. A total of 55 samples were acquired, which were randomly divided into training sets, containing 80% of the data (44 samples), and a test set, containing 20% of the data (11 samples), using a supervised learning method. This choice was justified by the fact that there were no significant improvements in the prediction model's accuracy when experimenting with different adjustment values. To assess the relationship between the parameters and the radiometric values extracted from the images, two architectural structures were analyzed for constructing the prediction model: one using only spectral bands (RGNIR) obtained from the MAPIR camera, and the other containing only RGB bands obtained from the integrated sensor on the RPA.

2.5 Machine learning algorithms

The Random Forest algorithm, proposed by Breiman (2001), is a flexible technique for modeling high-dimensional data. It constructs multiple regression trees and calculates the average of their predictions. This algorithm uses kernels and nearest neighbor methods to weigh the predictions based on the weighted average of nearby observations. Unlike other methods, it uses data to determine which nearby observations receive more weight (Wager and Athey, 2018). For modeling using the Random Forest algorithm, the following network settings, proposed by the software, were applied: batch size of (100), number of iterations (100), and bag size percent (100).

The Random Tree algorithm, used for performing classifications and regressions, combines the features of decision trees with randomness by constructing a forest of Random Trees, where each tree is built from a random subset of the training data and uses a random selection of features to make splits at the nodes. The final classification is obtained through voting or averaging the outputs of all the trees (Breiman, 2001). For modeling using the Random Tree algorithm, the following network settings, proposed by the software, were applied: number of attributes and maximum depth of trees (0), number of threads (1), and number of instances (1).

The main difference between these algorithms is that RandomTree constructs trees by selecting a test based on a specified number of random features at each node, without applying pruning techniques, while RandomForest constructs Random Forests through the bagging ensemble of Random Trees. As it has multiple decision trees, it is less susceptible to overfitting than Random Tree (Witten and Frank, 2002).

Regression analysis is a statistical technique used to investigate the relationship between variables. These algorithms assess the relationship between a dependent variable Y and independent variables X, represented by a mathematical model (Rodrigues, 2012). For modeling using the Linear Regression algorithm, the following network settings, proposed by the software, were applied: batch size (100), and selection method set (M5).

The M5Rules algorithm constructs tree-based models, where the leaves of traditional regression trees contain individual values, while in M5Rules, the leaves can accommodate multivariate linear models. This unique feature allows M5Rules to effectively handle tasks characterized by high dimensionality, involving a significant number of attributes (Quinlan, 2006). For modeling using the M5Rules algorithm, the following network settings, proposed by the software, were applied: number of rules (2), and number of instances per rule also (2).

M5P uses decision trees to partition the data and linear regression models to estimate the values of the output variable at each leaf node, resulting in a flexible and interpretable model for regression problems (Quinlan, 2006). For modeling using the M5P algorithm, the following network settings, proposed by the software, were applied: number of iterations (0) and number of instances (4).

Multilayer Perceptron neural networks are structured with an input layer, an output layer, and one or more hidden layers consisting of sectoral units or neurons. The input signal travels through these layers, one layer at a time (Haykin, 2001). Neural networks of this type use an error backpropagation algorithm, also known as the backpropagation algorithm, which uses a rule-based approach to adjust the weights in the network (Haykin, 2001). For modeling using the Multilayer Perceptron algorithm, the following network settings, proposed by the software, were applied: learning rate (0.3), momentum (0.2), and training epochs (500).

Unlike other learning models, the Support Vector Machine (SVM) is based on the minimization of structural errors. This algorithm aims to create, from a dataset, a hyperplane that equally separates the closest data points of each class to achieve a maximum margin on each side of the hyperplane. Only the training data points from each class that fall on the boundary of these margins are considered for each hyperplane. These data points are called support vectors (Suárez, 2013). For modeling using the SMOreg algorithm, the following network settings, proposed by the software itself, were applied: batch size (100) and Polykernel as the kernel function.

2.6 Evaluation metrics

To compare the performance of the algorithms, the metrics used were RMSE (Root Mean Square Error) and RMSE% (relative root mean squared error) between canopy diameter and plant height for coffee plants, based on the average data from the RGN and RGB sensors.

RMSE is defined by Equation 1, and RMSE% is defined by Equation 2:

https://doi.org/10.5194/isprs-annals-X-3-2024-317-2024 | © Author(s) 2024. CC BY 4.0 License.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}} \quad (1)$$

$$RMSE\ \% = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}} X \left(\frac{100Xn}{\sum_{i=1}^{n}\bar{x}}\right) \quad (2)$$

where, x_i the estimated value
$\bar{x}$ the measured value
n the number of samples

From the trained models, the ones that showed the best and worst performance were selected based on the average values of the Root Mean Square Error and relative root mean squared error, considering the difference between the observed and estimated variables for plant height and canopy diameter parameters evaluated in the experimental plot.

3. Results and Discussion

For the estimation of plant height, the RGB sensor showed the best performance using the Random Tree algorithm, with an RMSE (0.11) and RMSE% (3.45). The worst performance was observed for the RGN sensor using the multilayer perceptron algorithm, with an RMSE (0.27) and RMSE% (8.80) (Table 1).

For the estimation of canopy diameter, the RGN sensor exhibited the best performance using the Random Forest algorithm, with an RMSE (0.15) and RMSE% (8.16). The worst performance was observed for the RGB sensor using the multilayer perceptron algorithm, with an RMSE (0,27) and RMSE% (14.52) (Table 2).

Algorithms	Metric	Sensors	
		RGN	RGB
Linear Regression	RMSE	0.19	0.14
	RMSE (%)	6.27	4.52
Multilayer Perceptron	RMSE	**0.27**	0.20
	RMSE (%)	**8.80**	6.45
Simple Linear Regression	RMSE	0.13	0.15
	RMSE (%)	6.27	4.85
SMOreg	RMSE	0.21	0.15
	RMSE (%)	6.92	4.79
M5Rules	RMSE	0.19	0.14
	RMSE (%)	6.12	4.52
M5P	RMSE	0.19	0.14
	RMSE (%)	6.19	4.52
Random Forest	RMSE	0.19	0.11
	RMSE (%)	6.25	3.65
Random Tree	RMSE	0.23	**0.11**
	RMSE (%)	7.54	**3.45**

Table 1. RMSE (Root Mean Square Error) and RMSE% (relative root mean squared error) of the algorithms used to estimate the height of coffee plants, for the RGN and RGB sensors.

Algorithms	Metric	Sensors	
		RGN	RGB
Linear Regression	RMSE	0.18	0,18
	RMSE (%)	9.82	9.52
Multilayer Perceptron	RMSE	0.18	**0.27**
	RMSE (%)	9.43	**14.52**
Simple Linear Regression	RMSE	0.18	0.19
	RMSE (%)	9.45	9.92
SMOreg	RMSE	0.18	0.19
	RMSE (%)	9.34	10.28
M5Rules	RMSE	0.17	.18
	RMSE (%)	9.19	9.52
M5P	RMSE	0.17	0.18
	RMSE (%)	9.19	9.52
Random Forest	RMSE	**0.15**	0.19
	RMSE (%)	**8.16**	10.19
Random Tree	RMSE	0.22	0.21
	RMSE (%)	11.59	11.14

Table 2. RMSE (Root Mean Square Error) and RMSE% (relative root mean squared error) of the algorithms used to estimate the canopy diameter of coffee plants, for the RGN and RGB sensors.

The Random Forest and Random Tree algorithms showed the best performances for estimating plant height and canopy diameter, respectively. The use of these algorithms in estimating biophysical parameters in coffee farming using remote sensing data offers significant benefits. Moreover, when applied to complex data such as remote sensing data, machine learning algorithms have shown varied and often positive performance. The complexity of these data, which includes high dimensionality and spatial variability, requires algorithms that can effectively handle such challenges.

The Random Forest algorithm is capable of handling the complexity of remote sensing data and capturing interactions between variables, allowing for precise estimations. According to Biau and Scornet (2016), this algorithm stands out for its applicability in various prediction problems, requiring few tuning parameters and its ability to handle large real-world systems. Its main characteristics include simplicity, the ability to work with small samples, and the capacity to deal with high-dimensional feature spaces.

Random Tree is particularly suitable for handling remote sensing data due to its ability to handle multidimensional and complex features. When constructing a Random Tree, a random subset of the training data and features is selected, which helps handle the high dimensionality of geospatial data. Additionally, the randomness introduced in tree construction aids in controlling overfitting and increasing the model's generalization ability.

The superior performance of the Random Tree and Random Forest algorithms compared to the Multilayer Perceptron can be attributed to the intrinsic characteristics of each algorithm and the nature of the data used. Decision trees, found in the Random

Tree and Random Forest algorithms, are simple and interpretable algorithms that capture nonlinear interactions between input variables. On the other hand, the Multilayer Perceptron, being a more complex neural network, requires careful tuning of hyperparameters and may be more susceptible to issues of overfitting and underfitting. Comparative studies have shown that tree-based algorithms tend to outperform neural networks in terms of accuracy and interpretability in certain problems (Breiman, 2001; Hastie et al., 2009).

Additionally, the number of samples can also influence algorithm performance. In limited datasets, tree-based algorithms are less prone to overfitting, while neural networks may be more sensitive to scaling and data variability issues (Hastie et al., 2009). Therefore, it is important to consider both the characteristics of the algorithms and the size and quality of the dataset when selecting the most appropriate machine learning algorithm.

In this study, an analysis is conducted on the use of data obtained from low-cost remote sensing sensors embedded in RPAs, along with machine learning algorithms, to estimate biophysical parameters related to coffee productivity. Numerous studies have been conducted to estimate these biophysical parameters, which are directly linked to plant productivity, using both linear and nonlinear regression models (Xavier et al., 2006; Ramirez and Junior, 2010), as well as different machine learning algorithms (Zheng et al., 2022, Pereira et al., 2022). However, this research adopts an approach utilizing the radiometric reflectance values extracted from remote sensing images to estimate plant height and canopy diameter parameters in coffee plants. These parameters are essential for evaluating productivity in coffee farming.

Sensors embedded in RPAs capture wavelengths in the visible (RGB) and near-infrared (NIR) regions. The sensors used for image acquisition vary in cost, depending on the number of bands they can capture. RGB sensors are the least expensive (Xie and Yang, 2020).

Overall, the estimation of these parameters is often based on the extraction of vegetation indices from the data. However, each spectral band has important characteristics that can be beneficial for predicting these variables. The relationship between these vegetation indices and plant variables occurs due to the high absorption capacity of healthy plants in the visible region, especially in the blue and red bands, while the green band exhibits a peak of reflectance (Yang et al., 2017). In the near-infrared region, high reflectance occurs due to the influence of internal leaf structures (Ollinger, 2011).

For these parameters, no similar research was found for estimation. However, some studies have been conducted using the same machine learning algorithms and data acquired from remote sensing platforms to estimate parameters related to productivity in various crops, validating the potential use of this technology.

Pereira et al. (2022) used machine learning algorithms to estimate biophysical parameters of coffee plants using vegetation indices generated from images acquired by RPAs. For the plant height and canopy diameter parameters, the Support Vector Machine algorithm showed the best performance with RMSE (0.1302) and RMSE% (7.7374), and RMSE (0.1128) and RMSE% (3.6929) compared to the other algorithms. Osco et al. (2020) attempted to estimate nitrogen content and plant height in corn plants using spectral bands and vegetation indices generated from multispectral images acquired by RPAs and machine learning. The Random Forest algorithm showed the best performance, with an RMSE of 1.9, and the use of vegetation indices as the input dataset significantly contributed to the model's performance compared to individual spectral bands.

4. Conclusion

The use of multispectral imagery combined with artificial intelligence-based algorithms has shown satisfactory results in estimating biophysical parameters in coffee farming. This approach provides an alternative way to obtain information about coffee crops, contributing to improvements in productivity.

It is important to note that the results obtained are specific to the context of the study, considering the relatively small sample size evaluated. Therefore, further studies with larger samples are needed for more robust validation. Additionally, it is important to consider that results may vary depending on the characteristics of the study area, the types of sensors used, and the selected algorithms.

Furthermore, it is relevant to highlight that remote sensing is a non-destructive technology, which means that information about plants and their characteristics can be obtained without causing direct harm to the studied environment. This approach offers several advantages, such as obtaining accurate data, continuous monitoring of crop development at different stages, and the ability to access remote or hard-to-reach areas.

In conclusion, the results of this study demonstrate the effectiveness of machine learning algorithms combined with remote sensing in estimating biophysical parameters, providing valuable information to enhance agricultural management and support strategic decisions in plant production systems. This approach represents a significant advancement in the field of crop monitoring and contributes to the development of more precise and sustainable agricultural practices.

Acknowledgements

This study was funded in part by the Coordination for the Improvement of Higher Education Personnel (CAPES) – Brazil (Grants: nº 88887.835305/2023-00, 88887.817758/2023-00, 88887.817769/2023-00, 88887.817766/2023-00), and Grant: 2021/06029-7 São Paulo Research Foundation (FAPESP).

References

Abd-Elrahman, A., Dalid, C., Whitaker, V., Zheng, C., 2022: Prediction of Strawberry Dry Biomass from UAV Multispectral Imagery Using Multiple Machine Learning Methods. *Remote Sens.*, 14, 4511. doi: 10.3390/rs14184511.

Barros, A. C. de., 2021: Estimating agricultural production through remote sensing techniques. PhD Thesis in Agronomy - Faculty of Agricultural Sciences, Botucatu.

Bernardes, T., 2023: Modeling spectral and agrometeorological data for coffee productivity estimation. PhD Thesis in Remote

Sensing - National Institute for Space Research, São José dos Campos.

Biau, G., Scornet, E. 2016: A random forest guided tour. *TEST*, vol., 197–227. 2016 doi: https://doi.org/10.1007/s11749-016-0481-7

Bocca, F. F., Rodrigues, L. H. A., 2016: The effect of tuning, feature engineering, and feature selection in data mining applied to rainfed sugarcane yield modelling. *Computers and Electronics in Agriculture*, vol. 128, Pages 67-76 doi: 10.1016/j.compag.2016.08.015.

Breiman, Leo., 2001: Random Forests. Machine Learning, vol., no., pp. 5-32.

Carneiro, F.M. et al., 2019: Comparison between vegetation indices for detecting spatial and temporal variabilities in soybean crop using canopy sensors. *Precision Agriculture*, p. 1-29. doi: 10.1007/s11119-019-09704-3.

EMATER - MG (Rural Technical Assistance and Extension Company of the State of Minas Gerais), 2018: Getting to Know Coffee Farming in Minas Gerais. emater.mg.gov.br/portal.do (June 20, 2023).

F. G Hall, D. E Strebe, J. E Nickeson, S. J Goetz., 1991: Radiometric rectification: Toward a common radiometric response among multidate, multisensor images. *Remote Sensing of Environment*, vol.35, pp.11- 27. doi: 10.1016/0034-4257(91)90062-B.

Hastie, T., Tibshirani, R., and Friedman, J., 2009: The elements of statistical learning: data mining, inference, and prediction. Springer Science & Business Media, Latest edition.

Hallak, R., Pereira Filho, A. J., 2011: Methodology for performance analysis of convective system simulations in the metropolitan region of São Paulo using the ARPS model: sensitivity to variations in advection schemes and data assimilation. *Revista Brasileira de Meteorologia*, vol. 26, no. 4. doi: 10.1590/S0102-77862011000400009.

Haykin, S., 2001. Neural Networks: Principles and Practice. 2nd ed. Porto Alegre: Bookman.

Martinez, H.E.P., Augusto, H.S., Cruz, C.D., Pedrosa, A.W., Sampaio, N.F. 2007: Vegetative growth of coffee cultivars (Coffea arabica L.) and its correlation with production in high-density plantings. *Acta Scientiarum Agronomy*, vol. 29, pp. 481–489. doi: 10.4025/actasciagron.v29i4.404

Ndikumana, E., Minh, D.H.T., Nguyen, H.T.D., Baghdadi, D.C., Hossard, L., Moussawi, I., 2018: Estimation of Rice Height and Biomass Using Multitemporal SAR Sentinel-1 for Camargue, Southern France. *Remote Sensing*, vol. 10(11), 1703. doi:10.3390/rs10111703.

Ollinger, S.V. 2011: Sources of variability in canopy reflectance and the convergent properties of plants. *New Phytologist*, vol.189, 375-394. doi: 10.1111/j.1469-8137.2010.03536.x.

Oliveira, R.P. de., 2022: Prediction of Biometric Parameters of Sugarcane using Unmanned Aerial Vehicle Images. Dissertation (Master's thesis) - São Paulo State University (Unesp), School of Agricultural and Veterinary Sciences, Jaboticabal.

Osco, L.P., Junior, J.M., Ramos, A.P.M., Furuya, D.E.G., Santana, D.C., Teodoro, L.P.R., Gonçalves, W.N., Baio, F.H.R., Pistori, H., Junior, C.A.d.S., et al., 2020: Leaf Nitrogen Concentration and Plant Height Prediction for Maize Using UAV-Based Multispectral Imagery and Machine Learning Techniques. *Remote Sens.*, 12, 3237. doi: 10.3390/rs12193237.

Pereira, F.V., Martins, G.D., Vieira, B.S., et al., 2022: Multispectral images for monitoring the physiological parameters of coffee plants under different treatments against nematodes. *Precision Agric,* 23, 2312–2344. doi: 10.1007/s11119-022-09922-2.

Quinlan, J. R., 1992: Learning with continuous classes. In Proceedings of the 5th Australian Joint Conference on Artificial Intelligence (pp. 343-348). Springer-Verlag. Disponível em: https://sci2s.ugr.es/keel/pdf/algorithm/congreso/1992-Quinlan-AI.pdf.

Ramirez, G. M., Júnior, J. Z., 2010: Estimation of Biophysical Parameters of Coffee Plantations Based on High-Resolution Satellite Images. *Eng. Agríc.*, 30(3). doi:10.1590/S0100-69162010000300011.

Rodrigues, S.C.A., 2012: Linear Regression Model and Its Applications. *Revista Brasileira de Biometria*, 30(2), 123-138

Rouse, J.W., Haas, R.H., Schell, J.A., Deering, D.W., 1973: Monitoring vegetation systems in the great plains with ERTS. *NASA*, vol., p. 309-317.

Sharma, P., Leigh, L., Chang, J., Maimaitijiang, M., Caffé, M., 2022: Above-Ground Biomass Estimation in Oats Using UAV Remote Sensing and Machine Learning. *Sensors*, 22, 601. doi: 10.3390/s22020601.

Stefan Wager and Susan Athey., 2018: Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. *Journal of the American Statistical Association* vol., 1228-1242. doi: 10.1080/01621459.2017.1319839.

Suárez-Vázquez, A., and Quevedo, J., 2015: Analyzing superstars’ power using support vector machines. *Empirical Economics*, 49(4), 1521-1542. doi: 10.1007/s00181-014-0889-6.

Wang, L., Zhou, X., Zhu, X., Dong, Z., Guo, W., 2016: Estimation of biomass in wheat using random forest regression algorithm and remote sensing data. The Crop Journal. vol. 4, Issue 3, Pages 212-219. doi: 10.1016/j.cj.2016.01.008

Xavier, A.C., et al., 2006: Hyperspectral field reflectance measurements to estimate wheat grain yield and plant height. *Sci. agric.*, vol. 63(2). doi: 10.1590/S0103-90162006000200004.

Xie, C., and Yang, C., 2020: A review on plant high-throughput phenotyping traits using UAV-based sensors. *Computers and Electronics in Agriculture*, vol.178, 105731. doi: 10.1016/j.compag.2020.105731.

Yang, G., Liu, J., Zhao, C., et al., 2017: Unmanned aerial vehicle remote sensing for field-based crop phenotyping:

Current status and perspectives. *Frontiers in Plant Science*, 8, 1111. doi: 10.3389/fpls.2017.01111.

Witten, I. H., and Frank, E., 2002: Data mining: practical machine learning tools and techniques with Java implementations. *Acm Sigmod Record*, vol.31(1), 76-77. doi: 10.1145/507338.507355.