eISSN: 1984-3909
Coffee Science, e212401, 2026

Doi: https://doi.org/10.25186/.v21i.2401

Estimating Arabica Coffee Productivity Using Planetscope Images and Artificial Neural Networks

Rozymario Bittencourt Fagundes[1*], Luiz Patric Kayser[1], Ana Carolina Paim Benedetti[2], Lúcio de Paula Amaral[1], Édson Luiz Bolfe[3], Odair Lacerda Lemos[4]

[1]Programa de Mestrado Profissional em Agricultura de Precisão no Colégio Politécnico da Universidade Federal de Santa Maria, Santa Maria, RS, Brasil
[2]Colégio Politécnico da Universidade Federal de Santa Maria, Santa Maria, RS, Brasil
[3]Empresa Brasileira de Pesquisa Agropecuária/Embrapa Agricultura Digital e Instituto de Geociências da Universidade Estadual de Campinas/Unicamp, Campinas, SP, Brasil
[4]Programa de Pós-Graduação em Agronomia da Universidade Estadual do Sudoeste da Bahia, Vitória da Conquista, BA, Brasil
*Corresponding author: rozymariofagundes@gmail.com
Editor: Samuel Pereira de Carvalho
Received in May 16, 2025 and approved December 4, 2025

ABSTRACT

Research on orbital remote sensing and artificial neural networks (ANN) for estimating coffee productivity has advanced in Brazil, but applications remain scarce in states such as Bahia. This study aimed to develop and validate a model to estimate Arabica coffee (Coffea arabica L.) yield using PlanetScope images and ANN, based on two commercial areas with contrasting production systems (rainfed and irrigated). Georeferenced productivity samples (2 per hectare) were collected during the 2024 harvest, and PlanetScope images were analyzed from August 2023 to May 2024. In R software, vegetation indices related to vigor, nutritional status, and water stress were calculated, and their minimum, mean, and maximum values were extracted. Most variables followed a normal distribution (Shapiro–Wilk test, $p < 0.05$), enabling Pearson's correlation analysis. The best correlation was observed in Santa Vera, where the CCCI index of 08/20/2023 showed a moderate positive correlation with yield ($r = 0.66$; $p = 0.001$). After ANN training in SPSS software, the model achieved very high predictive performance in both areas, with R^2 and adjusted R^2 above 0.99, low errors (MAE = 0.031–0.072; RMSE = 0.042–0.087; relative RMSE = 0.029–0.043), and almost no bias (BIAS = 0.001; relative BIAS = 0). Among the vegetation indices, CCCI was the most relevant, followed by NDRE, NDVI, NDWI, and EVI. Overall, the findings demonstrate that combining PlanetScope imagery with ANN provides a highly accurate approach for estimating Arabica coffee productivity under different production systems in Bahia.

Key words: Precision Agriculture; Coffee Farming; Artificial Intelligence; *Coffea arabica* L.

1 INTRODUCTION

The estimation of coffee productivity at the farm level is fundamental information for agricultural management, as it contributes to the proper planning of the harvest, the strategic allocation of inputs, labor, and financial resources (Bolaños, Corrales, & Campo, 2023). In the Brazilian context, the world's largest producer and exporter of coffee, the modernization of estimation techniques becomes imperative in the face of the growing challenges imposed by climate change and the need for increased productive efficiency (Dias et al., 2024).

However, in the current panorama, the use of traditional methods for estimating productivity predominates, both on a macro and microeconomic scale, without the use of technologies such as satellite images or artificial intelligence algorithms. On the macro scale, official estimates are carried out by two government agents, CONAB (National Supply Company) and IBGE (Brazilian Institute of Geography and Statistics); and within the farms, by the producers themselves, through visual observations during the flowering period, a method that, although widely used, presents significant limitations as it depends on the subjective experience of the observer and does not allow for the identification of the spatial variability of production within the crops (Pereira, 2019). This is what happens, for example, in the production areas of the state of Bahia, the 4th largest coffee producer in Brazil, which stands out for its production of specialty coffees. This technical gap becomes particularly critical considering that productivity forecasting is an essential element for the adaptation of production systems to the new climate reality (Silva & Pinto, 2024; Fagundes & Bolfe, 2022; Santana et al., 2022).

In this context, Precision Agriculture, and its correlate for coffee production, Precision Coffee Growing, emerges as a technological paradigm capable of overcoming these limitations, by employing promising geotechnology tools, such as orbital Remote Sensing, from which it is possible to obtain spectral indices capable of characterizing biophysical parameters of the coffee crop. From these data, it is possible to derive spectral indices capable of characterizing biophysical parameters of coffee crops, such as NDVI (Normalized Difference Vegetation Index) and EVI (Enhanced Vegetation Index), which have been explored for assessing water conditions, although with varying levels of effectiveness (Santos et al., 2023); the NDRE (Normalized Difference Red Edge Index) as an indicator of leaf nitrogen (Zanzarini et al., 2013; Formaggio & Sanches, 2017); the CCCI (Canopy Chlorophyll Content Index) for estimating canopy chlorophyll (Gitelson et al., 2005; Santos et al., 2022); and the GNDVI (Green Normalized Difference Vegetation Index) for correlation with leaf nitrogen (Chemura et al., 2018). Additionally, indices such as SAVI

(Soil-Adjusted Vegetation Index) and MSAVI2 (Modified Soil-Adjusted Vegetation Index 2) have been used to reduce the influence of soil reflectance (Ramirez & Júnior, 2010), while the NDWI (Normalized Difference Water Index) shows effectiveness in detecting water stress (Santos et al., 2023).

The most significant advance in this area has been the integration between remote sensing techniques and artificial intelligence. Martello et al. (2022) successfully predicted Arabica coffee productivity one year in advance using data from the PlanetScope sensor combined with multiple linear regression and Random Forest models. In a similar approach, machine learning algorithms applied to MSI/Sentinel-2 sensor data have demonstrated potential for estimating productivity up to three months in advance. In the international context, Viet and Thuy (2023) developed a predictive model for Canephora coffee in Vietnam, combining MODIS images and multiple linear regression with 6 to 8 months anticipation.

Studies with other crops reinforce the potential of this approach. In the cotton crop, Oliveira et al. (2020) developed an automated system based on digital images that allowed yield prediction with an error of 17.86%, demonstrating the feasibility of computationally efficient solutions. In sweet potato, Tedesco et al. (2021) established that vegetation indices can predict crop yield before harvest, with smaller errors during the growth stage (2.63 t ha^{-1} in summer with NDVI and 3.06 t ha^{-1} in winter with GNDVI). In peanut, Souza et al. (2022) demonstrated that data from satellite and UAV (Unmanned Aerial Vehicles) images, combined with artificial neural networks, can predict the maturity index with high precision ($R^2 = 0.88$ with NDRE). Specifically for coffee, Thao et al. (2022) developed multiple linear regression models with satellite-derived biophysical variables that explained 64-69% of the productivity variability in Vietnam.

In the context of coffee production and the use of Precision Coffee Growing geotechnologies in the state of Bahia, local research has advanced in the application of orbital remote sensing for different purposes. Dias et al. (2019) using LANDSAT 7 images and the SEBAL algorithm in western Bahia, determined evapotranspiration and the crop coefficient for coffee, demonstrating the model's efficiency in estimating these biophysical variables, which are fundamental for crop water management. Despite these initiatives, there remains a lack of models that integrate orbital remote sensing and artificial intelligence techniques for estimating crop productivity, a limitation that can be attributed to the initial stage of consolidation of Precision Coffee Growing in the state (Silva & Alves, 2013).

This gap is particularly worrying considering that Bahia traditionally represents an important state for Brazilian coffee farming, both in the production of Arabica (Coffea arabica L.) and Canephora (Coffea canephora) coffees, with distinct edaphoclimatic characteristics that demand specific technological solutions. But the absence of predictive models adapted to Bahian conditions significantly limits the ability of producers to carry out truly precise and anticipatory management, as well as to analyze their productivity from the point of view of spatial variability.

Therefore, this work aims to develop a predictive model for the productivity of Arabica coffee (*Coffea arabica* L.) in Bahia, using data from spectral indices derived from high spatial resolution orbital sensors, combined with Artificial Neural Networks of the Multilayer Perceptron type. This work seeks to provide producers with a robust and precise tool for managing the spatial variability of production at the farm level, overcoming the limitations of traditional methods and promoting more efficient and sustainable decision-making.

2 MATERIAL AND METHODS

2.1 Orbital remote sensing images

To carry out this research, high spatial and temporal resolution images provided by constellations of small, standardized satellites, called nanosatellites or CubeSats, were used. These images are also referenced in the coffee SR, including for predicting productivity. Currently, such images are provided via a subscription service, but they can be obtained free of charge through academic research programs (Ruiz et al., 2022).

One of the companies that provide images of this type is Planet Labs, which has a constellation of 130 PlanetScope satellites in orbit with a daily temporal and spatial resolution of 3 meters, from different instruments.

The images from the PSB.SD instrument, whose characteristics are shown in Table 1, are offered with reflectance correction at the base of the surface and have a radiometric resolution of 12 bits and 8 spectral bands, according to the PlanetScope Product Specifications document (2022), available at https://www.planet.com/. Planet allows researchers free access to images for a period of one year, through the Education and Research Program (Planet, 2025).

2.2 Spectral indices

With images from PlanetScope (PSB.SD), the indices NDVI (Normalized Difference Vegetation Index), EVI (Enhanced Vegetation Index), SAVI (Soil-Adjusted Vegetation Index), MSAVI2 (Modified Soil-Adjusted Vegetation Index 2), NDRE (Normalized Difference Red Edge Index), GNDVI (Green Normalized Difference Vegetation Index), NDWI (Normalized Difference Water Index) and CCCI were generated. (Canopy Chlorophyll Content Index), whose equations are described in Table 2.

Table 1: PlanetScope sensor characteristics (PSB.SD).

Band Number	Spectral Bands	Wavelength (nm)	Spatial Resolution (m)
B01	Coastal Blue	431-452	3
B02	Blue	465-515	3
B03	Green I	513-549	3
B04	Green	547-583	3
B05	Yellow	600-620	3
B06	Red	650-680	3
B07	Red Edge	697-713	3
B08	NIR	845-885	3

Credit: Planet (2025).

2.3 Regional contextualization of the study areas

The sample areas correspond to plots of the farms Santa Vera, located in Bonito (BA), in the Chapada Diamantina region; and Requinte, located in Encruzilhada (BA), in the Planalto da Conquista region (SEI, s.d), as shown in Figure 1.

The Chapada Diamantina region, in the central area of the state known as Serra Geral, is characterized by its high plateaus and the production of specialty coffees, recognized for their full-bodied, sweet drink, with citric acidity and notes of nuts and chocolate. In 2024, the region received the Geographical Indication (GI) seal for Denomination of Origin by the National Institute of Industrial Property (INPI), due to the high levels of organic acids, chlorogenic acids and lipids in the beans. Coffee is grown in 24 municipalities with altitudes above 1,100 meters, in a mild climate with temperatures between 15 ºC and 21 ºC. Traditionally producing Arabica coffee since the 1970s, the Planalto da Conquista, located in the southwest of Bahia, comprises ten municipalities with altitudes above 700 meters and temperatures between 15 ºC and 25 ºC. This region has consolidated itself as the main producing area of Arabica coffee in Bahia, standing out in the production of both commodity and specialty coffees, with nationally awarded properties and the recent creation of the Specialty Coffee Route (Alves, 2023; Agência Sebrae, 2024).

2.4 Characterization of the study areas

2.4.1 Santa Vera Farm

Santa Vera Farm is located in the municipality of Bonito, and the climate is classified by the Köppen method as tropical highland, with summer rains and winter drought, an average annual temperature of 20.1º C, Caatinga biome and predominance of Neosol. The sample area (Figure 2) has geographic coordinates of 12º 00’ 26.15’’ South and 41º 15’ 48.96’’ West and an altitude of 993 meters.

The sample area used in this research is 10 ha, with a density of 4,649 Arabica coffee plants per hectare, of the Rubi 1192 variety, in a rainfed system. The crop was planted in 1999 and renewed with new planting in 2016, with an average productivity of 34.6 bags/ha in 2023. For the 2023/2024 harvest, two flowerings occurred, on October 5 and November 11 (with the highest volume). The fruit was set from January to May and ripened from May to August 2024. Fertilization management followed the agronomic recommendation of average application in kg

Table 2: Spectral indices and their equations with spectral bands.

Name	Spectral indice	Equation
Normalized Difference Vegetation Index	*NDVI*	$\frac{(NIR - Red)}{(NIR + Red)} (Kogan, 2012)$
Green Normalized Difference Vegetation Index	*GNDVI*	$\frac{(NIR - Greeen)}{(NIR + Greeen)} (Gitelson\ et\ al., 1996)$
Enhanced Vegetation Index	*EVI*	$2.5x \frac{(NIR - Red)}{(NIR + 6.Red - 7.5\ x\ Blue + 1)} (Huete\ et\ al., 2002)$
Soil-Adjusted Vegetation Index	*SAVI*	$\frac{(1 + L)(NIR - Red)}{(NIR + Red + L)} (Huete, 1988)$
Modified Soil-Adjusted Vegetation Index 2	*MSAVI2*	$\frac{[2 * NIR + 1 - \sqrt{(2 * NIR + 1)^2} - 8\ (NIR - Red)]}{2} (Qi\ et\ al., 1994)$
Normalized Difference Red Edge Index	*NDRE*	$\frac{(NIR - Red\ Edge)}{(NIR + Red\ Edge)} (Boiarskii\ \&\ Hasegawa, 2019)$
Canopy Chlorophyll Content Index	*CCCI*	$\frac{(NDRE)}{(NDVI)} (Fitzgerald\ et\ al., 2006)$
Normalized Difference Water Index	*NDWI*	$\frac{(Green - NIR)}{(Green + NIR)} (Gao, 1996)$

per hectare (555 kg of nitrogen applied in two stages, 97 kg of phosphorus, and 196 kg of potassium), as well as against pests and diseases. There were no significant complications during the period analyzed. In Figure 2, in addition to the sampling area and points where coffee productivity samples were collected, it is possible to see the sampling units in the planting line and at the intersection of the pixels in the PlanetScope image.

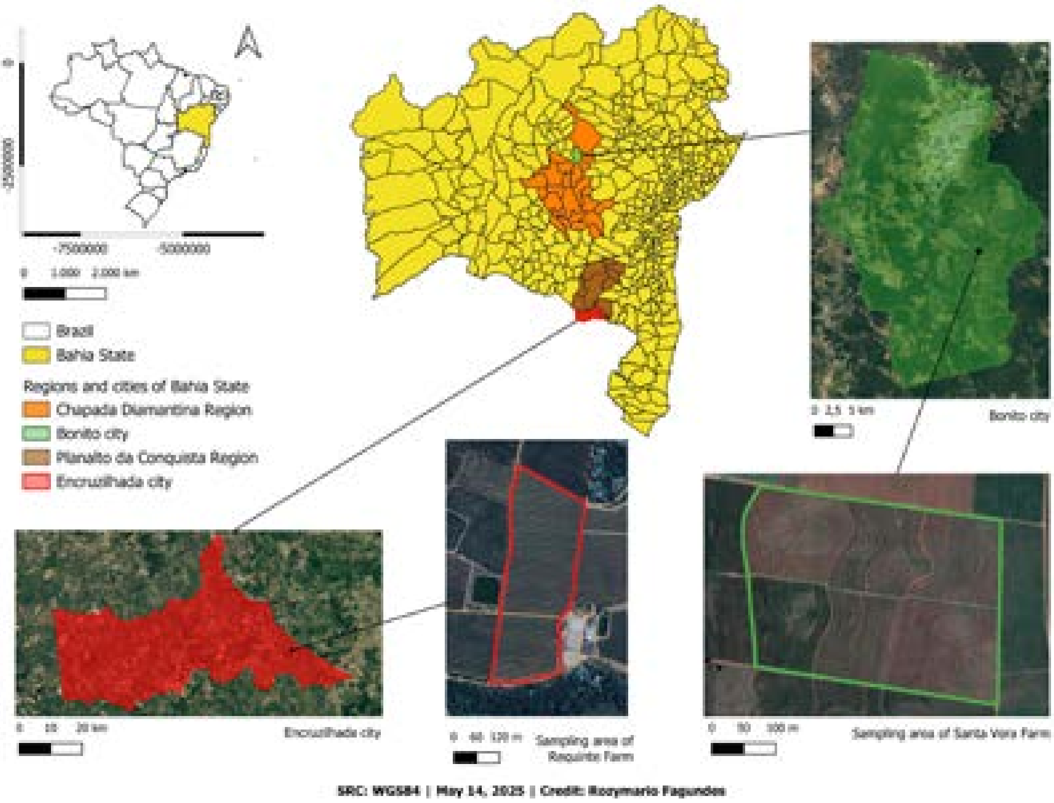

Figure 1: State of Bahia in Brazil; regions of Chapada Diamantina and Planalto da Conquista; the municipalities of Bonito and Encruzilhada; and the sample areas of the Santa Vera and Requinte farms.

Ground image credit: Google Earth.

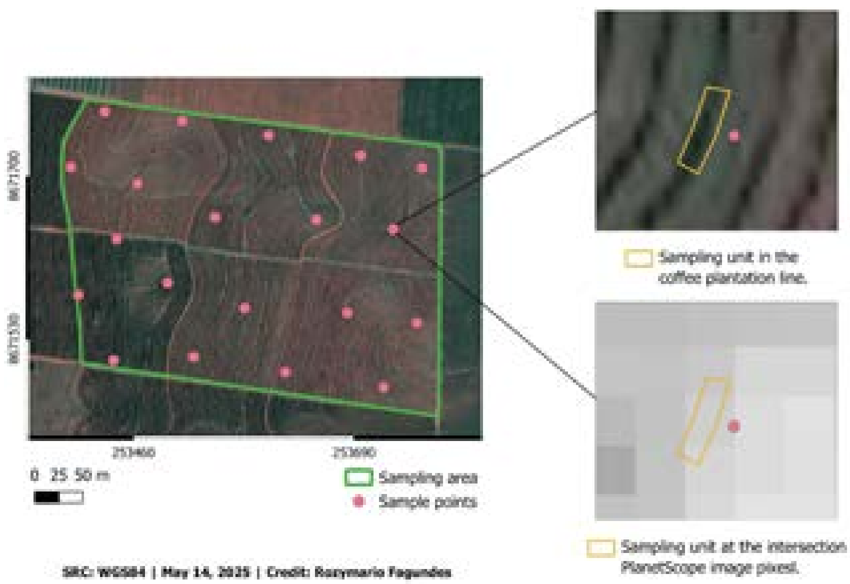

Figure 2: Santa Vera Farm, municipality of Bonito, Bahia, Brazil.

Ground image credit: Google Earth.

2.4.2 Requinte Farm

The Requinte Farm is located in Encruzilhada, where the climate of the region is classified by the Köppen method as a sub-humid tropical climate, Atlantic Forest biome and predominance of Latosols, with an average annual temperature of 22.1º C. The geographic coordinates are 15º 30' 29.33'' South and 40º 40' 51.91'' West, and the altitude is 901 meters. The sample area has 8.6 ha, with 3,125 Arabica coffee plants per hectare of the Catuai IAC 144 variety, grown in a drip irrigation system. The crop was established in 2008, with a spacing of 4 m x 0.80 m. The average productivity of the sample area in 2023 was 57 sc/ha. In the 2023/2024 harvest, three flowerings occurred, on September 12, November 10 (with the highest volume) and December 18, 2023. The fruit graining developed between January and May and ripening from May to August 2024. Fertilization management followed the agronomic recommendation of average application in kg per hectare (345 kg of nitrogen applied in two stages, 50 kg of phosphorus and 137 kg of potassium), as well as against pests and diseases, without recording significant complications that could reduce production.

Figure 3 shows details of the sampling area and the collection points of the coffee productivity samples, in addition to the sampling units in the planting line and at the intersection of the PlanetScope image pixels.

2.5 Auravant Digital Farming System

Auravant is a Digital Agriculture platform that allows for the georeferenced recording of field activities, allowing for the recording of various data, such as manual or mechanized harvesting of agricultural crops. More than one person can access the system and enter data, making it collaborative. Data can be downloaded in point files in .xlsx or .shp format for more advanced analysis. In the .xlsx file, each sample point and its corresponding data come with the code column and associated geographic coordinates. The .shp file also includes the code column of the points associated with the samples. The platform uses the WGS84 Coordinate Reference System (CRS). Auravant allows free use limited to 100 ha for one year, through the Student program, for academic research. Access is available at https://www.auravant.com/pt/home-pt/ (Auravant, 2024).

2.6 Computational tools

The computational tools used in this research were QGIS, Google Earth, R, and SPSS. Quantum GIS (QGIS) is a free and open-source Geographic Information System (GIS), created by Gary Sherman in 2002 and subsequently made available by a global community of developers. It allows the creation, editing, visualization, and analysis of geospatial data in vector and raster formats, as well as other output formats such as figures and tables, in addition to file export. Version 3.34.7 was used

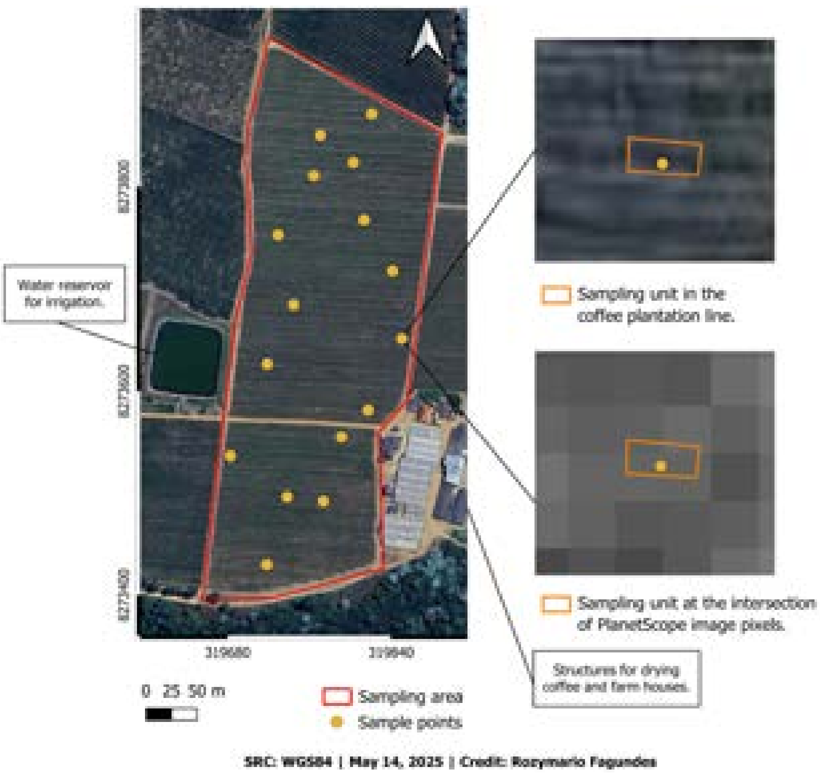

Figure 3: Requinte Farm, municipality of Encruzilhada, Bahia, Brazil.
Ground image credit: Google Earth.

for this research. QGIS can be purchased at https://qgis.org/. Google Earth is a free program provided by Google that allows for the acquisition of three-dimensional spatial data of the Earth. This research used the Pro version, which can be purchased at https://google-earth-pro.softonic.com.br/. R is an open-source programming language, owned by Posit, that enables statistical and geospatial analysis, with the ability to manipulate large amounts of data. Version 4.4.0 was used in this research. It can be downloaded for free at https://posit.co/download/rstudio-desktop/. The Statistical Package for the Social Sciences (SPSS) is IBM's proprietary advanced statistics software that requires a paid license. This computational tool enables advanced analysis of large volumes of data using machine learning and artificial intelligence models. IBM SPSS Statistics version 22 was used in this research. SPSS can be purchased at https://www.ibm.com/br-pt.

2.7 Methodology

The methodology for structuring data for modeling and analysis by ANN was divided into two parts: obtaining georeferenced samples of coffee productivity and vegetation index values (minimum, average and maximum) at the intersection of sampling units in the planting line and creating Excel spreadsheets with productivity models, as described below.

2.7.1 Coffee productivity samples

The sampling of Arabica coffee productivity followed the methodology used by Santinato et al. (2016) for Precision Agriculture in Coffee. Thus, a sampling grid of 2 points per hectare was created, with five coffee plants harvested at each sampling point. At Santa Vera Farm, manual harvesting took place on July 11, 2024, with 20 sampling points. At Requinte Farm, manual harvesting took place on June 12, 2024, with 16 sampling points generated for the harvest of mature coffee. On both farms, georeferenced harvest data was recorded in the Auravant system by the farm owners. The design and sample size were determined based on established protocols for Precision Coffee Farming, which prioritize operational and logistical feasibility under real field conditions while maintaining statistical robustness for calibrating machine learning models. While larger sample sizes are generally ideal, seminal studies demonstrate that advanced models, such as Artificial Neural Networks (ANNs), are capable of learning complex and generalizable patterns even from limited datasets, provided they are properly designed and validated (with cross-validation). Their effectiveness depends more on data quality and representativeness than purely on quantity (Tuia et al., 2022). The ability of ANNs, specifically the Multilayer Perceptron (MLP) model, to model complex nonlinear relationships between spectral variables and crop yield makes them particularly suitable for this context, where the number of predictor variables (spectral indices) is significant relative to the number of samples (Chlingaryan, Sukkarieh, & Whelan, 2018).

An important difference to consider is the way the data were recorded: in "liters" for the Santa Vera farm and in "cans" for the Requinte farm. These harvest recording methods are traditional in Brazilian coffee farming and vary from region to region. To unify the data and allow for more accurate analyses and direct comparisons, these values were converted to kilograms (kg) of mature processed coffee. Productivity data for the Santa Vera and Requinte farms are recorded in Tables 3 and 4, respectively, in their original values ("liters" or "cans") and their equivalent in kg. It is important to emphasize that Arabica coffee has a biennial cycle; in one year it produces more (a positive biennial cycle) and in the other (a negative biennial cycle). The year 2024 was a positive biennial cycle (Companhia Nacional de Abastecimento - CONAB, 2025).

2.7.2 Selection of orbital remote sensing images

Images from the PlanetScope constellation on the Planet Labs platform, available from August 1, 2023 to May 31, 2024, were selected from the two sample areas (Santa Vera and Requinte farms), covering the three phases of the Arabica coffee reproductive cycle: flowering, grain formation and fruit ripening. Since this is orbital remote sensing, the images were selected according to the quality criteria for analysis regarding the presence of clouds and shadows. Thus, the number of images selected was as follows::

- Santa Vera Farm: 36 images;
- Requinte Farm: 54 images.

To ensure that all images were in the same geographic coordinate projection, which in this case is WGS84, a polygon was created in QGIS (*.shp* file) for each sample area, thus serving as a reference for possible corrections. Directories were created only with images in *.tiff* extension files.

2.7.3 Obtaining spectral indices and creating sampling units

After the images of each sample area were acquired, directories were created for the respective orbital sensors. The following procedure was carried out entirely in the R software, where the .tiff files of the orbital images had their geographic coordinates and projections verified to obtain the spectral indices. The raster and tools packages were used. The images of the spectral indices were saved in a new directory. To reduce the influence of the soil and other vegetation of the coffee plantations on the spectral response of the coffee plant and on the subsequent analysis of the data, sampling units were created in the planting lines, corresponding to five coffee plants, using as reference the point where the harvests were carried out. The polygon of the sampling units was created in

Table 3: Productivity data by sampling point of the Santa Vera farm, municipality of Bonito, Bahia, Brazil – 2023/2024 harvest.

Points	Auravant Code	Latitude	Longitude	Liters	Mature kg
1	7X60-0PM-ZAA0	12°0’23.643”S	41°15’41.437”W	16	1.44
2	7X60-0PM-276S	12°0’25.706”S	41°15’42.453”W	20	1.8
3	7X60-0PM-R0Q4	12°0’28.892”S	41°15’41.659”W	26	2.34
4	7X60-0PM-JQ9S	12°0’31.060”S	41°15’42.827”W	20	1.8
5	7X60-0PM-C0JA	12°0’23.207”S	41°15’43.556”W	30	2.7
6	7X60-0PM-09V5	12°0’25.371”S	41°15’45.089”W	22	1.98
7	7X60-0PM-HJ2H	12°0’28.540”S	41°15’44.068”W	39	3.51
8	7X60-0PM-78T7	12°0’30.527”S	41°15’46.212”W	44	3.96
9	7X60-0PM-5Y47	12°0’22.493”S	41°15’46.711”W	36	3.24
10	7X60-0PM-TJFT	12°0’25.255”S	41°15’48.582”W	20	1.8
11	7X60-0PM-816R	12°0’28.346”S	41°15’47.602”W	59	5.31
12	7X60-0PM-7K0J	12°0’29.982”S	41°15’49.369”W	20	1.8
13	7X60-0PM-TLH3	12°0’21.985”S	41°15’49.688”W	40	3.6
14	7X60-0PM-QVCR	12°0’24.095”S	41°15’51.262”W	28	2.52
15	7X60-0PM-8X4F	12°0’27.463”S	41°15’50.236”W	52	4.68
16	7X60-0PM-2WH8	12°0’30.078”S	41°15’52.126”W	35	3.15
17	7X60-0PM-RGRL	12°0’21.649”S	41°15’52.358”W	45	4.05
18	7X60-0PM-8MS6	12°0’23.514”S	41°15’53.555”W	3	0.27
19	7X60-0PM-TVV6	12°0’25.965”S	41°15’51.986”W	49	4.41
20	7X60-0PM-JVXH	12°0’27.857”S	41°15’53.326”W	50	4.5

Table 4: Arabica coffee productivity data by sampling point at Requinte farm, municipality of Encruzilhada, Bahia, Brazil – 2023/2024 harvest.

Points	Auravant Code	Latitude	Longitude	Cans	Mature kg
1	2Z9V-0LH-11JK	15°36’38.001”S	40°40’53.695”W	3	5.4
2	2Z9V-0LH-F3F1	15°36’35.844”S	40°40’53.004”W	3,5	6.3
3	2Z9V-0LH-7KJH	15°36’35.996”S	40°40’51.813”W	4	7.2
4	2Z9V-0LH-N0S0	15°36’34.526”S	40°40’54.860”W	3,1	5.58
5	2Z9V-0LH-TM3G	15°36’33.954”S	40°40’51.205”W	2,3	4.14
6	2Z9V-0LH-F7GW	15°36’33.106”S	40°40’50.320”W	3,5	6.3
7	2Z9V-0LH-XS78	15°36’31.616”S	40°40’53.624”W	4,5	8.1
8	2Z9V-0LH-BTTN	15°36’30.837”S	40°40’49.185”W	4,4	7.92
9	2Z9V-0LH-TBM0	15°36’29.727”S	40°40’52.741”W	3,5	6.3
10	2Z9V-0LH-GZAG	15°36’28.681”S	40°40’49.470”W	4	7.2
11	2Z9V-0LH-1CCT	15°36’27.490”S	40°40’53.233”W	3	5.4
12	2Z9V-0LH-7YS8	15°36’27.042”S	40°40’50.404”W	2,8	5.04
13	2Z9V-0LH-Q01W	15°36’25.620”S	40°40’52.045”W	3,5	6.3
14	2Z9V-0LH-K6F1	15°36’25.209”S	40°40’50.737”W	3,2	5.76
15	2Z9V-0LH-KR5Z	15°36’23.669”S	40°40’50.124”W	4	7.2
16	2Z9V-0LH-AMGV	15°36’24.330”S	40°40’51.817”W	3	5.4

Google Earth, with manual vectorization and generation of a *.kmz* file, which was later used in R to generate a *.gpkg* file with information from the *.shp* file of points. In R, the sampling units were created using the *sf* and *tidyverse* packages. Next, using the R packages *sf, raster, dplyr* and *openxlsx*, the minimum, average and maximum values for the pixels at the intersection of each sample unit were extracted and an .xlsx spreadsheet was created with the values of the spectral indices for each sample of the coffee harvest at the sampling points.

2.7.4 Creating spreadsheets for data analysis

Na In the spreadsheet with data on the productivity of mature Arabica coffee, a column was created for each sampling point, similar to the one in the spreadsheet with the minimum, average and maximum values of the spectral indices. In this way, it was also possible in R, using the *readxl, openxlsx, dplyr* and *tidyr* packages, to generate a code that combined the two spreadsheets (productivity and vegetation indices) into a single one, with the data necessary for the analysis of the results of this research. The spreadsheets were as follows:

• Santa Vera Farm Spreadsheet:
Total rows: 20;
Total columns: 865.
• Requinte Farm Spreadsheet:
Total rows: 16;
Total columns: 1,297.

2.7.5 Correlation analysis between productivity and vegetation indices

Among the types of correlations that can be used, there are Pearson's and Spearman's. Pearson's correlation (r) assesses whether two variables (x and y) have a linear relationship, as well as the intensity and direction of this relationship, assuming that the data follow a normal distribution. This method can be used, for example, to verify the relationship between coffee productivity and vegetation indices (Silvestre & Bezerra, 2015). Pearson's correlation is calculated according to Equation 1:

$$r = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2} \cdot \sqrt{\sum(Y_i - \bar{Y})^2}} \quad (1)$$

where X_i and Y_i are the values of the X and Y variables, respectively; and $\bar{X}$ and $\bar{Y}$ are the means of the X and Y variables.

The Spearman correlation p (rho) measures the intensity and direction of a monotonic relationship between two variables, which may or may not be linear. This method is useful when the data do not follow a normal distribution, which may be the case when predicting coffee productivity based on vegetation indices (Silvestre & Bezerra, 2015). The Spearman correlation is calculated according to Equation 2:

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2 - 1)} \quad (2)$$

where d_i is the difference between the ranks of the variables X and Y for each observation *i*. And n is the total number of observations.

Both correlations range from -1 to 1. A value equal to or close to 0 indicates a weak or non-existent relationship. Values close to -1 indicate a negative correlation, while values close to 1 indicate a positive correlation. In both Pearson and Spearman correlations, the p-value is used to assess the statistical significance of the correlation, generally adopting a significance level of 5% (0.05). Values of $p \leq 0.05$ indicate that the correlation is statistically significant, while values greater than 0.05 suggest that the correlation may have occurred by chance (Sousa, 2019).

To determine which type of correlation best fits the productivity data and vegetation indices, the Shapiro-Wilk normality test was performed in the R software, using the *dplyr* package, considering a significance level of 5%. The test considers that the data are normal when the p-value is > 0.05, which indicates that Pearson's correlation is the best option for evaluating the data; and if it is ≤ 0.05, normality is rejected, and Spearman's correlation should be chosen. After the test, an Excel spreadsheet was created in which the results were classified as "Yes" for normality of the variables and "No" for non-normality. Based on this response, the best method for evaluating the correlation was chosen, with execution in R, using the *readxl* and *dplyr* packages (Dalchiavon and Carvalho, 2012; Silvestre & Bezerra, 2015).

2.7.6 Productivity Estimation Assessment Statistics

The productivity forecasting models were evaluated in Excel, based on the results generated by the ANN in SPSS between the real and predicted values. For this evaluation, the Mean Absolute Error (MAE) (Equation 3) and the Root Mean Square Error (RMSE) absolute (Equation 4) and relative (%) (Equation 5) were used, which are used to evaluate the quality of regression or forecasting models, by measuring the average error between the real and predicted values. The difference between the two metrics is that while the MAE measures the average errors, the RMSE penalizes large errors more. In both, values close to 0 indicate a smaller number of errors in the evaluated model (Willmott & Matsuura, 2005). The MAE and the absolute and relative RMSE are calculated as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} | y_i - \hat{y}_i | \quad (3)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (4)$$

$$RMSE\ (100\%) = \frac{RMSE}{\frac{1}{n}\sum_{i=1}^{n} y_i} \quad (5)$$

where y_i is the actual productivity; $\hat{y}_i$ is the productivity estimated by the ANN; and n is the amount of sample area analyzed, in this case = 1.

The actual and predicted values generated by the ANN were also evaluated by the Coefficient of Determination (R^2) and the Adjusted Coefficient of Determination (R^2_{adj}), given that they are of great importance in evaluating the quality of the model's adjustment. It varies from 0 to 1, with a value equal to or close to zero indicating that the model evaluated does not explain the variance of the data, or that the values of the dependent variance are independent of the model. The R2 is calculated according to Equation 6:

$$R^2 = 1 - \frac{\sum res}{\sum tot} \quad (6)$$

where: *Σres* (sum of squared residuals) is the sum of the squared errors between the observed values and the values predicted by the model; and *Σtot* (sum of total squares) the total variation of observed values in relation to the mean of real values. In SPSS software, R^2 is the statistic used to assess whether the model used in the ANN explains the variability of the data (Anderson, 1984). Another analysis can be done using adjusted R^2 (adjusted coefficient of determination), especially when dealing with variables with different coefficients, as is the case with real productivity data and values predicted by the ANN. Adjusted R^2 is calculated according to Equation 7:

$$R^2_{adj} = 1 - \left(\frac{(1-R^2)(n-1)}{n-p-1}\right) \quad (7)$$

Where: R^2 = Coefficient of Determination; n = total number of observations; p = number of independent variables of the model (spectral indices). In a complementary manner, Bias and Relative Bias (%) were also used to assess the difference between the real value and that predicted by the model, indicating whether or not there is a systematic error. For values > 0, Bias indicates that the model overestimates the real values; < 0, the model underestimates the real values; and values close to 0 indicate a low tendency or equilibrium of the model (Ioannidis et al., 2017). Bias (Equation 8) and Relative Bias (%) (Equation 9) are calculated as follows:

$$Bias = \frac{1}{n}\sum_{i=1} n(Yi - \hat{y}i) \quad (8)$$

$$Bias_{relative} = \frac{Bias}{\bar{y}} x100 \quad (9)$$

where ȳ (Equation 10) is the average of the observed:

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \quad (10)$$

In SPSS, each productivity model was evaluated individually, considering the most relevant spectral indices for each sensor and sample area. The neural network used had a hidden layer of 50 neurons, a hyperbolic tangent activation function, and an identity function at the output. Initially, all indices were tested, and the 20 most important ones were selected for a second round of modeling, optimizing the input variables. SPSS automatically adjusted the division between training and testing according to the model's needs. The best results were defined based on the R^2 of the cross-validation. The spreadsheets by area and model were then analyzed in R, with metrics such as MAE, RMSE, relative RMSE, adjusted R^2, BIAS, and relative BIAS.

3 RESULTS

3.1 Shapiro-Wilk normality test results

The Shapiro-Wilk normality test, at a significance level of 5%, applied to all productivity variables and spectral indices of the two sample areas of the Santa Vera and Requinte Farms, indicated that most variables presented "Yes" as a response, as can be seen in Table 5, thus indicating Pearson's correlation as the most appropriate statistical model to evaluate the correlation between them.

3.2 Pearson correlation results

In this section we will present the 5 best correlation results between productivity and spectral indices of each sample area.

3.2.1 Santa Vera Farm

The productivity and spectral index data from the PlanetScope sensor, from the Santa Vera Farm sampling area, had a moderate positive correlation, with *r* ranging from 0.65 to 0.66, and the *p-value* practically stable at 0.001. The minimum value of the CCCI vegetation index, on August 20, 2023, had the best result, with *r* = 0.66 and *p-value* = 0.001, and the other indices with very close values, as shown in Table 6.

Table 5: Shapiro-Wilk normality test. "Yes" counts the number of normal variables and "No" counts the number of non-normal variables.

Santa Vera Farm	Yes	No	Requinte Farm	Yes	No
Results	765	103	Results	1051	249

Table 6: Pearson correlation (*r*) and *p-value* of productivity in "liters" X spectral indices of the PlanetScope sensor of the sampling area of the Santa Vera Farm, Bonito, Bahia.

Spectral Indice	r	p-valor
min_20230820_CCCI	0.667792534	0.001293131
min_20230914_NDRE	0.664033117	0.001408699
max_20230820_GNDVI	0.663873941	0.001413778
min_20230820_NDRE	0.65674272	0.001657647
media_20230820_NDRE	0.655768174	0.001693561

3.2.2 Requinte Farm

Pearson's correlation (*r*) between productivity in "cans" at Fazenda Requinte and spectral indices from the PlanetScope sensor showed a moderate positive correlation (*r* between 0.55 and 0.60), with *p-value* ranging from 0.024 to 0.012. The index that obtained the best result was the minimum GNDVI value on September 09, 2023, with *r* = 0.60 and *p-value* = 0.012. The average GNDVI value on the same date obtained similar results: *r* = 0.58 and *p-value* = 0.018. The average CCCI value on May 17, 2024 had *r* of 0.57 and *p-value* of 0.018, while the maximum values of the NDVI and SAVI indices, both on February 01, 2024, had similar results, as can be seen in Table 7.

Table 7: Pearson correlation (*r*) and *p-value* of productivity in "cans" X spectral indices of the PlanetScope sensor of the sampling area of the Requinte farm, Encruzilhada, Bahia.

Spectral Indice	*r*	*p-valor*
min_20230928_GNDVI	0.605365388	0.012958032
media_20230928_GNDVI	0.581899543	0.018043767
media_20240517_CCCI	0.578537724	0.018883782
max_20240201_NDVI	0.558489305	0.024540373
max_20240201_SAVI	0.558486961	0.024541103

3.3 Results of cross-validation in ANN to estimate productivity

3.3.1 Santa Vera Farm

The results of the cross-validation between the actual productivity and that predicted by the ANN for the sample area of Fazenda Santa Vera, in the municipality of Bonito, Bahia, showed good performance of the models, with R^2 and adjusted R^2 of 0.995 and low average errors of MAE, RMSE, and relative RMSE. The BIAS was also very close to zero and the relative BIAS at 0.001. Table 8 shows the data on the actual productivity and that predicted by the ANN.

The "Mature kg" model had results of $R^2 = 0.995$, adjusted $R^2 = 0.995$, MAE = 0.072, RMSE = 0.087, relative RMSE = 0.029, BIAS = 0.001, relative BIAS = 0.001, as shown in Figure 4.

Regarding the spectral indices that contributed most to the results of the cross-validation between the actual productivity and that predicted by the ANN of the sample area of Fazenda Santa Vera, the most prominent was the CCCI of 10/20/2023 (Figure 5).

The CCCI of December 03, 2023 and the EVI (August 20 and September 28, 2023, and March 31, 2024) also had a great contribution to the results.

3.3.2 Requinte Farm:

The performance of the cross-validation between the real productivity and that predicted by the ANN, for the PlanetScope sensor, of the sampling area of the Requinte farm, presented good results, with $R^2 = 0.999$, adjusted $R^2 = 0.992$, MAE = 0.093, RMSE = 0.042, relative RMSE = 0.043, BIAS = 0.001 and relative BIAS = 0. The Table 10 shows the values of the real productivity and that predicted by the ANN for "Maduro kg".

The model's performance analysis showed that it was fine-tuned, close to perfection. The Figure 6 the results of the statistics evaluating the model's performance.

The spectral indices that contributed most to the results of the cross-validation performance between the actual productivity and that predicted by the ANN, of the PlanetScope sensor, sampling area of Fazenda Requinte, were the CCCI (15/11/2023 and 09/10/2023) and NDRE (18/10/2023) indices, as shown in Figure 7.

The NDVI (11/24/2023) and NDWI (05/20/2024) also influenced the results, with the difference being that the latter index, important for monitoring crop water stress, is from the end of May, when the coffee is in the final stage of maturation, with the harvest having been carried out on June 12, 2023. Thus, it can be observed that the best results were with indexes that indicate the vigor of the vegetation from images of the plant crown and deeper in the canopy, and also the water content in the leaves, which is of great importance for capturing the spatial variability of the crop and making more assertive management decisions.

4 DISCUSSIONS

4.1 Discussion on Pearson correlation results

The best Pearson correlation results were observed in the sampling area of Santa Vera farm: the CCCI index (minimum value of 08/20/2023) presented a moderate and significant positive correlation with productivity (r = 0.66; p-value = 0.001). In the Requinte farm, the best performance was obtained with the minimum value of GNDVI of 09/28/2023 (r = 0.60; p-value = 0.012), also configuring a moderate positive correlation. These results indicate that CCCI and GNDVI were the most relevant vegetation indices in relation to coffee productivity, reinforcing their potential use in predictive analyses.

Table 8: Real value and value predicted by ANN for the "Mature kg" model, PlanetScope sensor, sampling area of Fazenda Santa Vera, Bonito, Bahia.

Points	Latitude	Longitude	Liters	Mature kg	Predicted value RNA
1	12°0'23.643"S	41°15'41.437"W	16	1.44	1.51
2	12°0'25.706"S	41°15'42.453"W	20	1.8	1.94
3	12°0'28.892"S	41°15'41.659"W	26	2.34	2.27
4	12°0'31.060"S	41°15'42.827"W	20	1.8	1.71
5	12°0'23.207"S	41°15'43.556"W	30	2.7	2.71
6	12°0'25.371"S	41°15'45.089"W	22	1.98	2.05
7	12°0'28.540"S	41°15'44.068"W	39	3.51	3.6
8	12°0'30.527"S	41°15'46.212"W	44	3.96	4.01
9	12°0'22.493"S	41°15'46.711"W	36	3.24	3.24
10	12°0'25.255"S	41°15'48.582"W	20	1.8	1.84
11	12°0'28.346"S	41°15'47.602"W	59	5.31	5.17
12	12°0'29.982"S	41°15'49.369"W	20	1.8	1.92
13	12°0'21.985"S	41°15'49.688"W	40	3.6	3.57
14	12°0'24.095"S	41°15'51.262"W	28	2.52	2.51
15	12°0'27.463"S	41°15'50.236"W	52	4.68	4.53
16	12°0'30.078"S	41°15'52.126"W	35	3.15	3.14
17	12°0'21.649"S	41°15'52.358"W	45	4.05	4.14
18	12°0'23.514"S	41°15'53.555"W	3	0.27	0.11
19	12°0'25.965"S	41°15'51.986"W	49	4.41	4.36
20	12°0'27.857"S	41°15'53.326"W	50	4.5	4.56

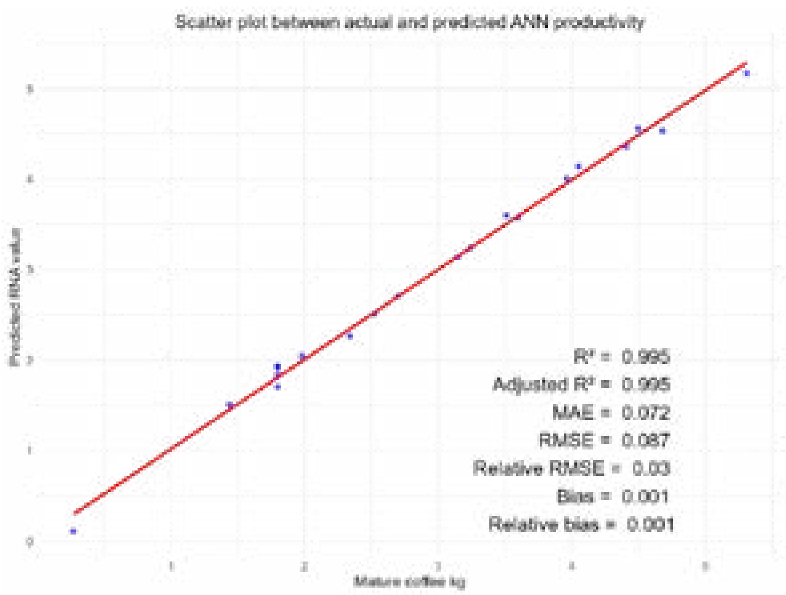

Figure 4: Statistical evaluation of the cross-validation performance between the real productivity "Maduro kg" and that predicted by the ANN for the PlanetScope sensor, sample area of Fazenda Santa Vera, Bonito, Bahia.

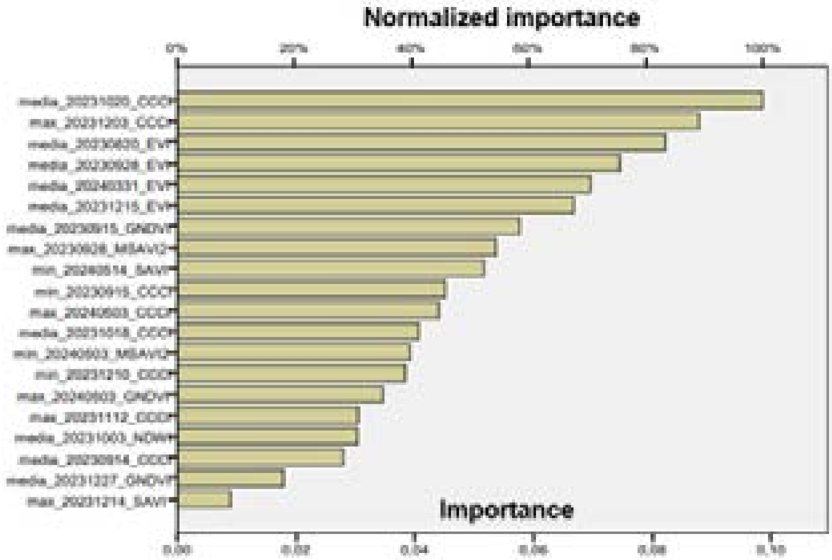

Figure 5: Importance of spectral indices, PlanetScope sensor, sampling area of the Santa Vera farm, Bonito, Bahia.

Table 10: Real value and value predicted by ANN for the “Maduro kg” model, PlanetScope sensor, sampling area of the Requinte farm, Encruzilhada, Bahia.

Points	Latitude	Longitude	Cans	Mature kg	Predicted value RNA
1	15°36’38.001”S	40°40’53.695”W	3	5,4	5.39
2	15°36’35.844”S	40°40’53.004”W	3,5	6,3	6.35
3	15°36’35.996”S	40°40’51.813”W	4	7,2	7.2
4	15°36’34.526”S	40°40’54.860”W	3,1	5,58	5.54
5	15°36’33.954”S	40°40’51.205”W	2,3	4,14	4.18
6	15°36’33.106”S	40°40’50.320”W	3,5	6,3	6.34
7	15°36’31.616”S	40°40’53.624”W	4,5	8,1	8.13
8	15°36’30.837”S	40°40’49.185”W	4,4	7,92	7.92
9	15°36’29.727”S	40°40’52.741”W	3,5	6,3	6.26
10	15°36’28.681”S	40°40’49.470”W	4	7,2	7.09
11	15°36’27.490”S	40°40’53.233”W	3	5,4	5.43
12	15°36’27.042”S	40°40’50.404”W	2,8	5,04	5.11
13	15°36’25.620”S	40°40’52.045”W	3,5	6,3	6.29
14	15°36’25.209”S	40°40’50.737”W	3,2	5,76	5.76
15	15°36’23.669”S	40°40’50.124”W	4	7,2	7.17
16	15°36’24.330”S	40°40’51.817”W	3	5,4	5.39

The relevance of CCCI for coffee growing is particularly notable, since this index is derived from the combination between NDRE and NDVI and is directly associated with the estimation of the chlorophyll content in the canopy, reflecting the photosynthetic capacity and nutritional status of the plant (Putra & Soni, 2018). The highlight of GNDVI in the Requinte farm, in turn, confirms its usefulness to evaluate vegetation density and vigor and to map the spatial variability of the productive potential of the coffee plant (Abreu Júnior et al., 2022).

The good performance of indices that use the Red Edge spectral band, such as CCCI and NDRE, was already expected, since previous studies in coffee growing reported similar results. Luna et al. (2020), for example, observed that NDRE presented a higher coefficient of variation in relation to NDVI and GNDVI, suggesting better sensitivity to detect spatial variability in cultivation. In this same work, NDVI and GNDVI had better performance in the Spearman correlation among themselves than NDRE, but all indices presented high correlations, above 0.9.

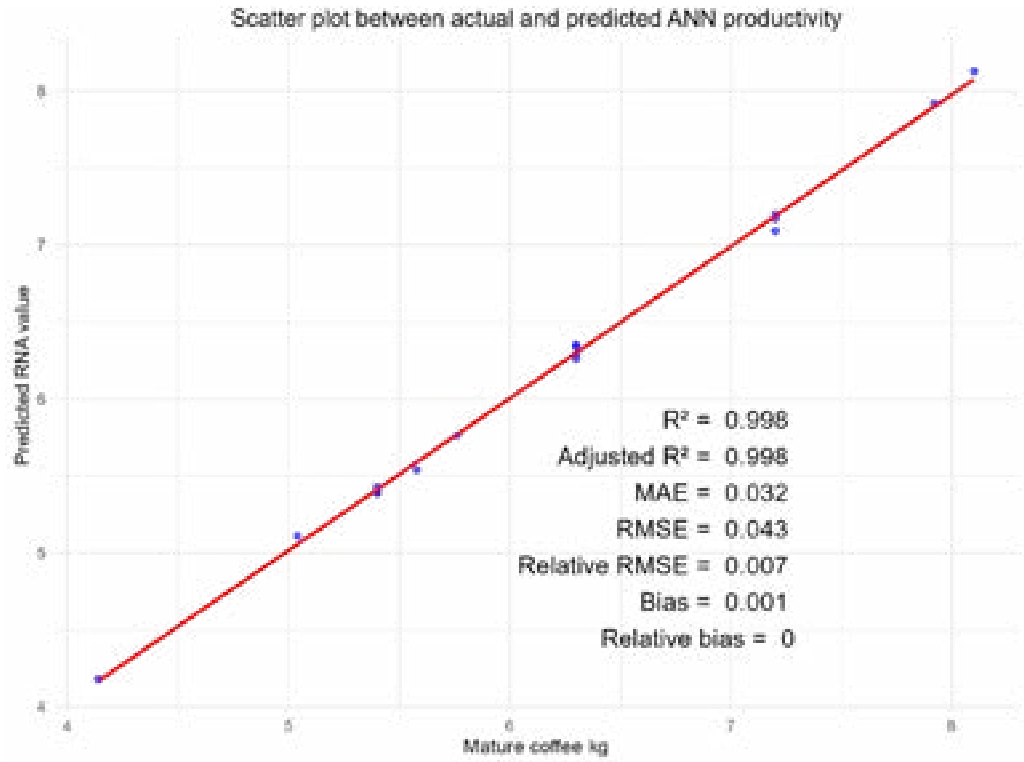

Figure 6: Statistical evaluation of the performance of the cross-validation between the real productivity, “Maduro kg” model, and that predicted by the ANN for the PlanetScope sensor, sampling area of the Requinte farm, Encruzilhada, Bahia.

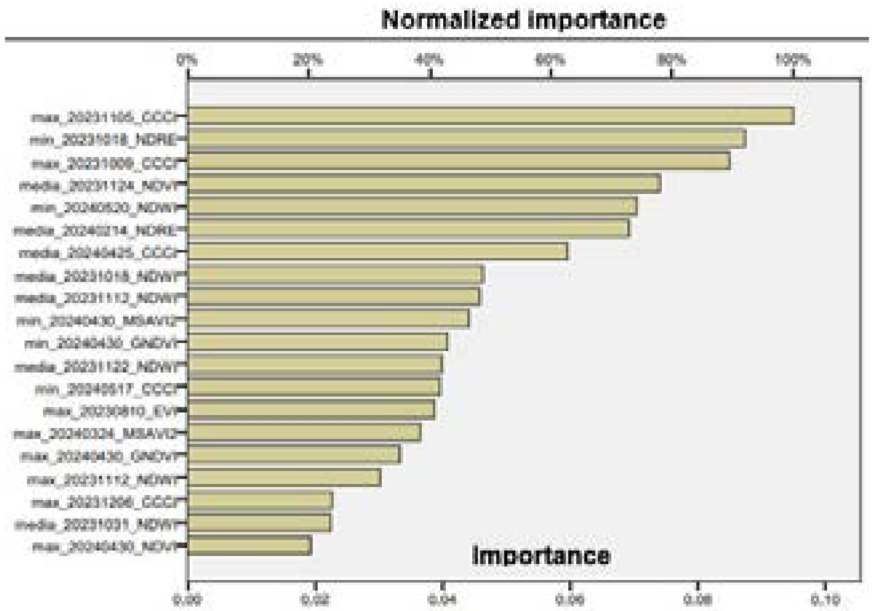

Figure 7: Importance of spectral indices, PlanetScope sensor, sampling area of the Requinte farm, Encruzilhada, Bahia.

In different contexts, NDRE has also shown to be consistent. Rodrigues et al. (2021) identified that this index was more sensitive than NDVI to detect variations after hailstorm events in different Arabica coffee varieties. Complementarily, Martins et al. (2021) proposed the Coffee Ripening Index (CRI) to monitor coffee ripening and demonstrated that it outperformed NDRE in discriminating ripening stages. Fabri et al. (2023), when evaluating vegetative vigor and biophysical characteristics (height and canopy diameters), verified that NDVI presented better correlation with these variables, followed by NDRE and GLI.

Considering that the performance of spectral indices may vary depending on the crop’s phenological stage, it should be noted that, at the Santa Vera farm, flowering for the 2023/2024 harvest occurred in two stages (10/05/2023 and 11/11/2023), on dates different from the CCCI acquisition. This condition, characterized by the simultaneous presence of fully expanded leaves and others in initial vegetative growth, may have influenced the behavior of the index.

4.2 Discussions on the results of cross-validation between actual productivity and that predicted by the ANN

The cross-validation between the actual productivity of the sampling areas of the Santa Vera and Requinte farms and the productivity predicted by the ANN, using spectral index data from the PlanetScope sensor, demonstrated high performance of the generated model. In both areas, the coefficients of determination were close to 1 and the prediction errors very reduced, confirming the accuracy and reliability of the method. This level of adjustment is unusual in crop yield prediction studies and highlights the potential of ANNs to capture nonlinear relationships between spectral variables and productivity.

At Santa Vera farm, the average sample productivity was 45.60 bags/ha, while the ANN predicted 47.5 bags/ha, with an absolute error of only 1.90 bags/ha. At Requinte farm, the difference was even smaller: 64.8 bags/ha observed against 64.81 bags/ha predicted by the ANN. Subsequently, however, the total average productivity of the sampling areas, as reported by the producers, was 38.91 bags/ha at Santa Vera farm and 56 bags/ha at Requinte, resulting in divergences of around 8.59 bags/ha and 8.81 bags/ha, respectively. This difference can be attributed to the contrast between sampling data, which privilege points representative of productive potential, and the general field averages, which include planting failures, less vigorous areas, and operational losses.

An important aspect is that the spectral indices were collected over ten months, covering all phenological phases of the coffee cycle, which allowed the model to capture information from different stages of plant development. This factor probably contributed to the slightly higher values predicted by the ANN, since more vigorous plants tend to dominate the spectral signals throughout the time series. Even so, the proximity between the values predicted and the field samples reinforces the robustness of the model.

The results obtained are in line with other research that applied remote sensing and machine learning techniques to predict coffee productivity. Zanella et al. (2024), for example, combined spectral and foliar nutritional data in regression models, showing that the use of orbital images anticipates productivity estimates months in advance, although with lower accuracy than that achieved in this research. Dhika et al. (2024), using ANN to predict productivity in Indonesia, also confirmed the feasibility of the technique, even with a dataset restricted to historical production series. Koaudio et al. (2021), by employing the Extreme Learning Machine (ELM) algorithm for canephora coffee in Vietnam, highlighted that approaches based on artificial intelligence can outperform traditional models, such as Multiple Linear Regression and Random Forest, in integrating spectral, climatic, and soil fertility variables.

In addition, there are concrete studies that applied ANNs or machine learning techniques to predict coffee productivity or similar crops, demonstrating the potential of these approaches. For example, Abreu Júnior et al. (2022) estimated coffee productivity using multispectral images (Sentinel-2) and several machine learning algorithms, such as Random Forest, SVM and neural networks, finding that the NN was the one that obtained the best performance, with R^2 of 0.82, RMSE of approximately 23% and MAPE of 20% (using 85% of the data for training and 15% for validation).

Another interesting study is the prediction of arabica crops using time series and ANNs, carried out in Thailand by Kittichotsatsawat et al. (2023), in which the ANN model obtained $R^2 = 0.9299$ and RMSE = 0.0642 when predicting annual productivity, outperforming the traditional ARIMA model ($R^2 = 0.7041$). Still in the context of coffee, Barbosa et al. (2021) used unmanned aerial vehicles (UAVs) with RGB images and deep learning algorithms applied to the estimation of plant height and canopy diameter (derived features) for yield prediction. They tested several techniques, including evolutionary algorithms such as NEAT, and obtained better results when they used only the most important variables (LAI and canopy diameter) and key months for prediction (such as December and April). This shows the importance of selecting and monitoring in the appropriate months, which in the case of this research proved to be those close to coffee flowering, corroborating other studies.

These works show that, at different scales and conditions, ANNs and machine learning methods have been effective in estimating coffee productivity, and even in phenological phases or internal spatial variation (as in the case of UAVs). This reinforces the relevance of your ANN model in the Bahian context and suggests that it is aligned with the most advanced trends in the agricultural prediction literature.

4.3 Importance of vegetation indices

In the normalized importance of vegetation indices, the CCCI of 10/20/2023, from Fazenda Santa Vera, and that of 11/05/2023, from Requinte, stood out, with greater contributions to the productivity models in the PlanetScope sensor.

It can be observed that for “Mature kg” in the PlanetScope sensor, the NDRE (10/18/2023), NDVI (11/24/2023) and NDWI (05/20/2024) indices were of relevant importance for Fazenda Requinte, and for Santa Vera also the EVI index (08/20/2023, 09/28/2023 and 03/31/2024).

The results were similar to those obtained by Martello et al. (2022) to estimate Arabica coffee productivity in Minas Gerais using images from the PlanetScope sensor, using NDVI and GNDVI. The study, carried out over three harvests (2019-2021) in an area of 10.24 ha, used productivity data obtained by a harvester equipped with a 3 m spatial resolution productivity monitor. RF and MLR models were adjusted to time series of spectral bands and vegetation indices, and both reproduced the spatial variability of production one year before harvest. In the results, the model obtained by RF with spectral bands obtained better performance ($R^2 = 0.93$ and smaller errors).

The dormant phase, during the dry months of July and August, was identified as the best period for image acquisition, allowing management decisions based on spatial and temporal variability, aligned with precision agriculture. Bahia, generally at this time, is in the harvest phase in most areas. However, changes have been increasingly observed in the periods related to the coffee cycle, with farms harvesting on different dates, depending on the management adopted.

Images from the PlanetScope, MSI/Sentinel-2 and Landsat-8 satellites have been used to map coffee landscapes, as in the study by Mosomtai et al. (2020) on the canephora species in Kenya. Spectral indices, such as GNDVI, texture variables and spectral bands were used as inputs to generate four RF classification models. The model based on the MSI/Sentinel-2 spectral bands presented the highest accuracy, with emphasis on the shortwave near infrared and the green band as the most relevant variables.

5 CONCLUSIONS

This study demonstrated that spectral indices from PlanetScope images effectively predict Arabica coffee productivity, with ANN models showing high performance ($R^2 > 0.99$ and low errors). CCCI, followed by NDRE, NDVI, EVI and NDWI, highlights the importance of the Red Edge band in explaining productivity variability. The results validate ANNs under Bahia's edaphoclimatic conditions. As a limitation, the analysis was restricted to one crop year and two sampling areas, suggesting future validations with multiple crop years.

6 AUTHOR CONTRIBUTIONS

Conceptual idea: Fagundes, R.; Kayser, L.P.; Methodological design: Fagundes, R.; Kayser, L.P.; Amaral, L.P.; Benedetti, A.C.; Bolfe, E.L.; Lemos, O.L.; Data collection: Fagundes, R., Data analysis and interpretation: Fagundes, R.; Kayser, L.P.; Amaral, L.P.; Benedetti, A.C.; Bolfe, E.L.; Lemos, O.L., and Writing and editing: Fagundes, R.; Kayser, L.P.; Amaral, L.P.; Benedetti, A.C.

7 DATA AVAILABILITY STATEMENT

Available upon Request to Authors.

8 REFERENCES

Abreu Júnior, C. A. M. et al. (2022). Estimating coffee plant yield based on multispectral images and machine learning models. *Agronomy*, *12*(12):3195.

Agência Sebrae. (2024). *Café da Chapada Diamantina (BA) é a mais nova Indicação Geográfica brasileira*. Available in: https://agenciasebrae.com.br/cultura-empreendedora/cafe-da-chapada-diamantina-ba-e-a-mais-nova-indicacao-geografica-brasileira/.

Alves, S. A. (2023). A inserção da cafeicultura no planalto da conquista - Bahia: transformações sociais e econômicas da região. *Geofronter*, *9*(1).

Anderson, T. W. (1984). Estimating linear statistical relationships. *Annals of Statistics*, *12*(1):1-45.

Auravant. (2024). *Auravant for Education*. Available in: <https://www.auravant.com/auravant-for-education/.>

Barbosa, B. D. S. et al. (2021). UAV-based coffee yield prediction utilizing feature selection and deep learning. *Smart Agricultural Technology*, *1*:100010.

Bolaños, J., Corrales, J. C., & Campo, L.V. (2023). Feasibility of early yield prediction per coffee tree based on multispectral aerial imagery: Case of arabica coffee crops in Cauca-Colombia. *Remote Sensing*, *15*(1):282.

BoiarskiI, B., & Hasegawa, H. (2019). Comparison of NDVI and NDRE indices to detect differences in vegetation and chlorophyll content. *Journal of Mechanics of Continua and Mathematical Sciences*, *14*:20-29.

Chemura, A. et al. (2018). Mapping spatial variability of foliar nitrogen in coffee (*Coffea arabica* L.) plantations with multispectral Sentinel-2 MSI data. *ISPRS Journal of Photogrammetry and Remote Sensing*, *138*:1-11.

Chlingaryan, A., Sukkarieh, S., & Whelan, B. (2018). Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. *Computers and Electronics in Agriculture*, *151*:61-69.

Companhia Nacional de Abastecimento - CONAB. (2025). *Acompanhamento da safra brasileira de café, v. 11 – Safra 2024, n.4- Quarto levantamento*. Brasília, p. 1-53. Available in: https://www.gov.br/conab/pt-br/atuacao/informacoes-agropecuarias/safras/safra-de-cafe/4o-levantamento-de-cafe-safra-2024/boletim-cafe-janeiro-2025.

Dias, Thalles Loiola et al. Evapotranspiração e coeficiente de cultura do cafeeiro irrigado a partir de imagens de sensores orbitais. IRRIGA, [S. l.], v. 1, n. 1, p. 31–37, 2019. DOI: 10.15809/irriga.2019v1n1p31-37.

Dias, C. G. et al. (2024). Climate risks and vulnerabilities of the Arabica coffee in Brazil under current and future climates considering new CMIP6 models. *Science of The Total Environment*, *907*:167753.

Dalchiavon, F. C., & Carvalho, M. de P. (2012). Correlação linear e espacial dos componentes de produção e produtividade da soja. *Semina: Ciências Agrárias*, *33*(2):541-552.

Dhika, H. R. et al. (2024). Predicting coffee productivity using artificial neural networks. *BIS Information Technology and Computer Science*, *1*:V124007.

Fagundes, R. B., & Bolfe, Édson L. (2022). Diagnosis about the perspectives of precision applications of coffee growing technologies in municipalities of Bahia, Brazil. *Coffee Science*, *17*:e172034.

Fabri, T. M. P. et al. (2023). Comportamento de índices de vegetação em cafeeiros utilizando sensor multi-espectral em aeronave remotamente pilotada. *Revista Ifes Ciência*, *9*(1):1-15.

Formaggio, A. R., & Sanches, I. D. A. (2017). *Sensoriamento remoto em agricultura*. São Paulo: Oficina de Textos, 285p.

Gao, Bo-Cai. (1996). NDWI-A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, *58*(3):257-266.

Gitelson, A. A. et al. (1996). Detection of red-edge position and chlorophyll content by reflectance measurements near 700 nm. *Journal of Plant Physiology*, *148*:501-508.

Gitelson, A. A. et al. (2005), Remote estimation of canopy chlorophyll content in crops. *Geophysical Research Letters*, *32*(8):L08403.

Ramirez, G. M., & Zullo Júnior, J. (2010). Topografia, fotogrametria e sensoriamento remoto. *Engenharia Agrícola*, *30*(3):468-479.

Huete, A. R. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, *25*(3):295-309.

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

Ioannidis, T. D. et al. (2017). The power of bias in economics research. *The Economic Journal*, *127*(605):F236-F265.

Fitzgerald, G. F. et al. (2006). Spectral and thermal sensing for nitrogen and water status in rainfed and irrigated wheat environments. *Precision Agriculture*, *7*(4):233-248.

Luna, D. R. et al. (2020). Análisis de los índices de vegetación NDVI, GNDVI y NDRE para la caracterización del cultivo de café (Coffea arábica). *Ingeniaría y desarrollo*, 38(2):2145-9371.

Martins, R. N. et al. (2021). A novel vegetation index for coffee ripeness monitoring using aerial imagery. *Remote Sensing*, *13*(2):263.

Martello, M. et al. (2022). Coffee-yield estimation using high-resolution time-series satellite images and machine learning. *AgriEngineering*, *4*(4):888-902.

Mosomtai, G. et al. (2021). Functional land cover scale for three insect pests with contrasting dispersal strategies in a fragmented coffee-based landscape in Central Kenya. 2020. *Agriculture, Ecosystems & Environment*, *319*:107558.

Koaudio, L. et al. (2021). Performance of a process-based model for predicting robusta coffee yield at the regional scale in Vietnam. *Ecological Modelling*, *443*:0304-3800.

Kittichotsatsawat, Y. et al. (2023). Forecasting arabica coffee yields by auto-regressive integrated moving average and machine learning approaches. *AIMS Agriculture and Food*, *8*(4):1052-1070.

Oliveira, D. T. et al. (2020). Convolutional neural networks in predicting cotton yield from images of commercial fields. *Computers and Electronics in Agriculture*, *171*:105307.

Pereira, A. (2019). Análise temporal do índice NDVI utilizando o Google Earth Engine: Estudo de caso na cafeicultura. In C. C. Santos. *Book: O Semiárido Brasileiro e suas especificidades*. Ponta Grossa, PR: Atena Editora, [recurso eletrônico].

Planet Labs. (2025). *PlanetScope product specifications*. Available in: <https://assets.planet.com/docs/Planet_PSScene_Imagery_Product_Spec_letter_screen.pdf.>.

Putra, B. T. W., & Soni, P. (2018). Enhanced broadband greenness in assessing Chlorophyll a and b, Carotenoid, and Nitrogen in Robusta coffee plantations using a digital camera. *Precision Agriculture*, *19*:238-256.

Qi, J. et al. (1994). A modified soil adjusted vegetation index. *Remote Sensing of Environment*, *48*(2):119-126.

Rodrigues, S. A. et al. (2021). Efeito da chuva de granizo em variedades do café arábica por meio de índices de vegetação. *Agrarian*, *14*(54):433-441.

Kogan, F. N. (2012). Historical Perspectives on AVHRR NDVI and Vegetation Drought Monitoring. In B. D. Wardlow., M. C. Anderson., & J. P. Verdin, (Ed.). *Remote sensing of drought: innovative monitoring approaches*. Boca Raton: CRC Press/Taylor & Francis.

Ruiz, I. et al. (2022). On the combined use of phenological metrics derived from different PlanetScope vegetation indices for classifying savannas in Brazil. *Remote Sensing Applications: Society and Environment*, *26*:100764.

Santana, L. S. et al. (2022). Precision coffee growing: a review. *Coffee Science*, *17*:e172007.

Santos, S. A. d. et al. (2023). Evaluation of the water conditions in coffee plantations using RPA. *AgriEngineering*, *5*(1):65-84.

Santos, L. M. d. et al. (2022). RPA images in the mapping of the chlorophyll index of coffee plants. *Sustainability*, *14*:13118.

Santinato, F. et al. (2016). Análise espacial da produtividade do cafeeiro arábica utilizando diferentes densidades amostrais. *Revista Brasileira de Engenharia Agrícola e Ambiental*, *20*(12):1075-1080.

Superintendência de Estudos Econômicos e Sociais da Bahia- SEI. (s.d). *Perfil dos Territórios de Identidade da Bahia*. Available in: <https://sei.ba.gov.br/index.php?option=com_content&view=category&id=1518&Itemid=284&lang=pt>.

Silva J. B. C. da., & Pinto, P. A. L. de A. (2024). Impactos socioeconômicos das mudanças climáticas na produção do café: uma revisão sistemática da literatura. 2024. *Revista Brasileira de Climatologia*, *35*(20):155-178.

Silva, F. M., & Alves, M. C. *Cafeicultura de Precisão*. Lavras: Editora UFLA, 2013.

Silvestre, M. R., & Bezerra, M. I. S. (2015). Modelos de decomposição temporal e de regressão harmônica: uma comparação para a série mensal da temperatura mínima e máxima média de Presidente Prudente (SP). *Revista Brasileira de Meteorologia*, *30*(4):457-466.

Sousa, A. (2019). *Coeficiente de correlação de Pearson e coeficiente de correlação de Spearman. O que medem e em que situações devem ser utilizados?*. «Correio dos Açores: Matemática», 19p.

Souza, J. B. C. et al. (2022). Integrating satellite and UAV data to predict peanut maturity upon artificial neural networks. *Agronomy*, *12*(7):1512.

Tedesco, D. et al. (2021). Use of remote sensing to characterize the phenological development and to predict sweet potato yield in two growing seasons. *European Journal of Agronomy*, *129*:126337.

Thao, N. T. T. T. et al. (2022). Early prediction of coffee yield in the central highlands of vietnam using a statistical approach and satellite remote sensing vegetation biophysical variables. *Remote Sensing*, *14*(13):2975.

Tuia, D. et al. (2022). Perspectives in machine learning for wildlife conservation. *Nature Communications*, *13*:792.

Viet, L. V., &Thuy, T. T. T. (2023). Improving the quality of coffee yield forecasting in Dak Lak Province, Vietnam, through the utilization of remote sensing data. *Environmental Research Communications*, *5*:095011.

Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, *30*(1):79-82.

Zanella, M. A. et al. (2024). Coffee yield prediction using high-resolution satellite imagery and crop nutritional status in Southeast Brazil. *Remote Sensing Applications: Society and Environment*, *33*:101092.

Zanzarini, F. V. et al. (2013). Correlação espacial do índice de vegetação (NDVI) de imagem Landsat/ETM+ com atributos do solo. *Revista Brasileira de Engenharia Agrícola e Ambiental*, *17*(6):608-614.