

# Mapping Coconut Plantations in the Brazilian Semi-Arid Region Using Machine Learning Techniques and Sentinel Imagery

Eduardo Vicente do Prado

Professor, Vale do Acaraú State University, Acaraú, Ceará, Brazil  
Corresponding Author

Date of Submission: 10-08-2025 Date of Acceptance: 20-08-2025

**ABSTRACT:** Coconut farming in the Brazilian semi-arid region, especially in the state of Ceará, plays a significant economic role but faces challenges such as the limited use of modern agricultural technologies. The use of remote sensing techniques, through both multispectral satellite imagery and radar data, offers an opportunity to efficiently, quickly, and accurately monitor crop conditions, promoting more effective and sustainable management. This study aimed to map coconut plantations in the municipality of Paraipaba, in the state of Ceará, using machine learning techniques. Multispectral images from Sentinel-2 satellites were used, covering the period from January 1, 2024, to March 8, 2025, and radar images from Sentinel-1 satellites, from January 1, 2023, to May 30, 2024. The study employed Support Vector Machine (SVM) and Random Forest (RF) classifiers, with RF demonstrating superior accuracy in both multispectral and radar images. The RF classifier achieved an accuracy of 97.8% and a kappa coefficient of 94.2% with multispectral images, outperforming SVM, which yielded an accuracy of 86.3% and a kappa of 71.9%. For radar imagery, RF achieved a global accuracy of 98.6% and a kappa of 93.3%. This study highlights the effectiveness of machine learning techniques, particularly RF, for the accurate and efficient mapping of coconut plantations using Sentinel-1 and Sentinel-2 satellite imagery.

**KEYWORDS:** Coconut plantations, Machine learning, Remote sensing, Sentinel satellites, Semi-arid regions.

## I. INTRODUCTION

The coconut palm (*Cocos nucifera* L.) is a widely distributed crop throughout the tropics,

cultivated in approximately 86 countries, playing a prominent role in food security, income generation, and international trade especially due to the growing demand for fresh coconut water [1][2]. Brazil ranks fifth in global coconut production, with around 2.4 million tons per year [3], with the Northeast Region serving as the country's main production hub. In the state of Ceará, the municipality of Paraipaba stands out as the largest producer, with an estimated output of 115.5 million fruits, underscoring the economic and social importance of coconut cultivation in the semi-arid region.

Despite its productive relevance, coconut farming in Brazil's Northeast still lacks robust technological strategies for the systematic mapping and monitoring of cultivated areas. The scarcity of studies focused on the use of advanced geospatial technologies for crop management represents a critical gap, particularly in regions affected by climate variability and structural limitations. In this context, remote sensing using radar and multispectral imagery has emerged as a viable, non-destructive, and scalable alternative for extracting biophysical vegetation parameters, such as the Normalized Difference Vegetation Index (NDVI), which is widely applied in precision agriculture [4][5].

Among the main advantages of remote sensing are its broad spatial coverage, high temporal frequency of data acquisition, and independence from prior information about soil conditions and production systems [6][7]. However, the high computational cost associated with processing these images remains a challenge, especially in environments with limited technological infrastructure. Cloud computing, through platforms such as Google Earth Engine (GEE), has helped to overcome this barrier by integrating global geospatial data repositories with analytical tools optimized for large-scale parallel processing [8][9].

GEE provides direct access to imagery from optical and radar sensors, including Landsat, Sentinel-1, Sentinel-2, and MODIS, eliminating the need for specialized hardware and facilitating the implementation of complex analytical workflows. The Sentinel-1 mission uses C-band synthetic aperture radar (SAR), offering images with varying spatial resolutions depending on the acquisition mode, reaching up to 5 meters. The Sentinel-2 mission, with a spatial resolution of 10 meters and a revisit time of up to five days, has become a strategic tool for agricultural monitoring in tropical regions, enabling the detection of phenological variations and accurate land cover mapping [10][11].

In addition, the incorporation of machine learning (ML) algorithms, such as Random Forest (RF) and Support Vector Machines (SVM), has revolutionized supervised classification approaches. These models, being non-parametric and independent of assumptions regarding data distribution, are particularly effective in agricultural contexts characterized by spatial and spectral heterogeneity [12][13][14]. Recent literature highlights the success of combining multispectral data and ML algorithms in small-scale farming, showing promising results in developing countries [15][16].

Alongside optical imagery, synthetic aperture radar (SAR) data, such as those provided by Sentinel-1, have enhanced the analytical potential of agricultural mo-

onitoring systems due to their cloud penetration capability and independence from illumination conditions [17].

Although susceptible to topographic distortions, SAR data complements spectral information in discriminating surface features, especially in tropical regions [18].

Moreover, indices such as the Normalized Difference Built-up Index (NDBI) have proven effective in identifying impervious surfaces and built-up areas, contributing to the refinement of classifications in urban-rural transition zones [19].

In light of this scenario, the present study aimed to map coconut cultivation areas in the municipality of Paraipaba, Ceará, using radar imagery from the Sentinel-1 mission and optical imagery from Sentinel-2, processed on the Google Earth Engine platform. The classification of coconut cultivation areas employed the Random Forest and Support Vector Machine algorithms to assess the accuracy and applicability of these approaches within the context of coconut farming in the Brazilian semi-arid region.

## II. METHODOLOGY

### Study area

The municipality of Paraipaba, located in the state of Ceará, Brazil, is situated at coordinates 3° 26' 20" S and 39° 08' 52" W, at an altitude of 39 meters and with a total area of 289,231 m<sup>2</sup> (Figure 1). According to the Köppen classification, the local climate is of type Aw (tropical savanna climate), characterized by a dry season. The region has an average annual rainfall of 1,131 mm and maximum and minimum temperatures of 31.2 °C and 21.1 °C, respectively [20].



Figure 1. Boundaries of the municipality of Paraipaba

### Remote sensing data

Radar images from the Sentinel-1 satellite, made available by the European Space Agency (ESA) through the Google Earth Engine (GEE) platform, were used in this study. Sentinel-1 consists of satellites equipped with C-band (~5.4 GHz) Synthetic Aperture Radar (SAR), offering global coverage and spatial resolution of up to 10 meters.

For this study, images acquired in Interferometric Wide Swath (IW) mode were selected, using VV and VH polarizations, captured in descending orbit. The images were filtered based on the absence of severe atmospheric interference and spaced at a maximum temporal interval of 24 days.

The images were pre-processed in GEE through the following steps:

- **Precise orbit correction** (similar to recalibrating the satellite's GPS path to ensure accurate positioning);
- **Radiometric calibration to sigma nought ( $\sigma^0$ )** (standardizing the radar's brightness values so different images speak the same "language");
- **Speckle filtering using a 3x3 Lee filter** (like smoothing out "static noise" on a TV screen to reveal the real picture);
- **Geometric correction and orthorectification using the SRTM elevation model** (correcting for terrain distortions, similar to flattening a crumpled map);
- **Logarithmic conversion to decibels (dB)** (transforming the data scale to better compare values, like adjusting sound levels to a decibel scale for clarity).

#### Construction of the training and validation dataset

Georeferenced samples of areas with and without coconut cultivation were collected based on high-resolution imagery available in Google Earth Engine (GEE). These samples were divided into two datasets:

- **Training set (70% of the samples):** used to train the algorithms;
- **Validation set (30%):** used to assess classification accuracy.

The sampling followed a stratified random model, ensuring proportional representation between classes and minimizing spatial bias.

#### Supervised classification

The classification of coconut cultivation areas was performed using two widely recognized machine learning algorithms:

- **Random Forest (RF):** an algorithm based on an ensemble of decision trees built from random subsets of the dataset. The number of trees and the maximum depth were fine-tuned through cross-validation.
- **Support Vector Machine (SVM):** an algorithm based on maximizing the

margin between hyperplanes. The Radial Basis Function (RBF) kernel was used, with optimization of the parameters performed via grid search.

Both models were implemented on the Google Earth Engine (GEE) platform, using backscatter values ( $\sigma^0$ , VV, and VH) and the ratio between polarizations (VV/VH) to improve classifier performance.

#### Accuracy assessment

Model accuracy was evaluated using confusion matrices and statistical metrics, including:

- Overall accuracy;
- Kappa coefficient (K);
- User's Accuracy and Producer's Accuracy for each class;
- Omission and commission errors.

Additionally, classification maps were generated to visualize the classified areas and identify boundaries between classes.

#### Satellite imagery

Multispectral images of the study area, acquired by the Sentinel-2 satellites, were used for the period from January 1, 2024, to March 8, 2025, to identify areas under coconut cultivation. Sentinel-2 images have a spatial resolution of 10 meters, a temporal resolution of 5 days, and a radiometric resolution of 12 bits per pixel, meaning that the intensity of each pixel ranges from 0 to 4096 gray levels, ensuring a high level of detail in the generated information.

To ensure better data processing quality, multispectral images captured on days with low cloud cover were selected. The cloud filter available on the GEE platform was applied using the command `filter(ee.Filter.lt('CLOUDY_PIXEL_PERCENTAGE', 2))`. This command selects images with less than 2% cloud cover over the total image area.

#### Biophysical parameters

The algorithms for processing biophysical parameters were implemented on the Google Earth Engine (GEE) platform. GEE was chosen due to its capability to efficiently access and process large

volumes of remote sensing data, enabling fast and reproducible analyses.

The processing was conducted in the following steps:

- **Data acquisition:** multispectral images from the Sentinel-2 satellites, available in the GEE catalog, were used.
- **Pre-processing:** this included atmospheric correction and cloud filtering to ensure image quality.
- **Calculation of vegetation indices (VIs):** indices such as NDVI, NDBI, and others relevant to monitoring the coconut orchard were computed.

### Vegetation indices (VIs)

The multispectral images from the Sentinel-2 satellites provide the spectral bands necessary for calculating various vegetation indices (VIs).

The processing and calculation of VIs were performed on the Google Earth Engine (GEE) platform, which is recognized for its efficiency in analyzing large volumes of data and enabling process automation. After calculation, thematic VI maps were generated to visualize and identify spatial patterns and trends within the study area.

There are categories of VIs related to structural, biochemical, and physiological aspects (associated with stress), as well as those specific to pest and disease detection. The VIs utilized in this study are listed below (Table 1).

### Statistical analysis

Thematic maps were generated to illustrate the spatial distribution of the vegetation indices (VIs), promoting clear visual interpretation of the data and facilitating their use in the classifiers.

Confusion matrices and the Kappa statistic were employed to assess the accuracy of the estimated coconut cultivation area. Classification accuracy was determined by comparing pixel allocations in the classified image with corresponding reference data. A confusion matrix was constructed to compile pixels of agreement and disagreement, with rows and columns representing all classes, and matrix elements indicating pixel counts in the test dataset. Accuracy metrics, such as overall

accuracy and the Kappa coefficient, were derived from the error matrix.

Overall accuracy (OA) was calculated using Equation 1.

$$OA = \frac{P_o}{P_n} \times 100 \quad \text{Equation 1}$$

Where  $P_n$  is the total number of pixels and  $P_o$  is the total number of correctly classified pixels.

Considering  $N$  as the total number of pixels analyzed;  $N_{ii}$  as the number of pixels correctly classified for class  $i$ ;  $N_{i.}$  as the total number of pixels classified as class  $i$  (sum of row  $i$  in the confusion matrix); and  $N_{.i}$  as the total number of pixels that truly belong to class  $i$  (sum of column  $i$  in the confusion matrix).

Thus, the proportion of observed agreement  $P_o$  (Equation 2) and the proportion of expected agreement by chance  $P_e$  (Equation 3) are calculated, respectively, by:

$$P_o = \frac{\sum N_{ii}}{N} \quad \text{Equation 2}$$

$$P_e = \frac{\sum (N_{i.} \cdot N_{.i})}{N^2} \quad \text{Equation 3}$$

To evaluate the effectiveness of the classifiers, the Kappa statistic is a robust method that allows a comparative analysis of the results from maps obtained through remote sensing, within certain limits [23]. The Kappa coefficient (K) is calculated using Equation 4.

$$K = \frac{P_o - P_e}{1 - P_e} \quad \text{Equation 4}$$

## III. RESULTS

The mapping results indicated the spatial distribution of coconut cultivation areas in the municipality of Paraipaba, Ceará state. Generally, coconut palms are cultivated throughout almost the entire municipality, with a higher concentration in the central region. The spatial distribution of coconut plantations were relatively dispersed across the central, northern, and southern areas (Figures 2 and 3).

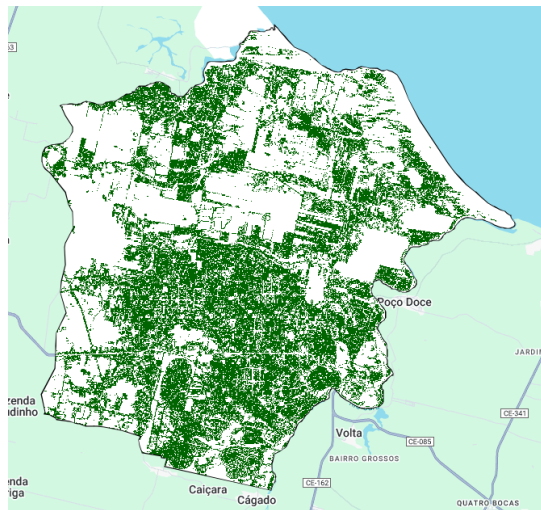


Figure 2. Classification result using the SVM classifier.

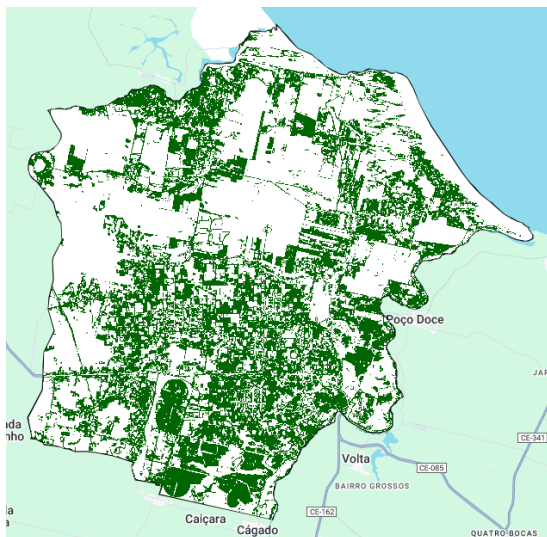


Figure 3. Classification result using the RF classifier.

In the tested scenarios, the vegetation indices NDVI, mNDVI, and NDBI were used. The NDVI and NDBI vegetation indices yielded the best classification results (Figure 3), both for optical and radar images (Figure 4). The best classifier was Random Forest (RF). The RF classification using optical data achieved an Overall Accuracy and Kappa of 97.8% and 94.2%, respectively (Figure 3), outperforming the SVM classification, which reached an Overall Accuracy of 86.3% and a Kappa of 71.9%. The RF classification of radar images obtained an Overall Accuracy and Kappa of 98.6% and 93.3%, respectively (Figure 4).

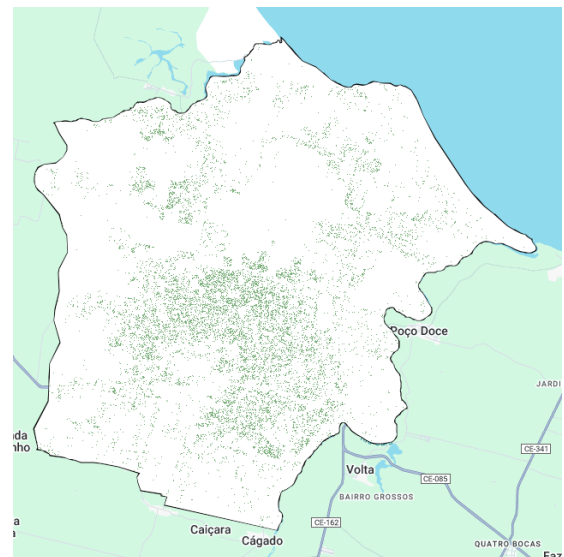


Figure 4. Classification result using the RF classifier with radar images.

Table 1. Vegetation indices used.

| Index                                                    | Equation                                                          | Characteristics                                                                 | Reference |
|----------------------------------------------------------|-------------------------------------------------------------------|---------------------------------------------------------------------------------|-----------|
| Normalized Difference Vegetation Index (NDVI).           | $\frac{\rho_{IR} - \rho_R}{\rho_{IR} + \rho_R}$                   | Structural index related to biomass production.                                 | [21]      |
| Modified Normalized Difference Vegetation Index (mNDVI). | $\frac{\rho_{NIR} - \rho_{RedEdge}}{\rho_{NIR} + \rho_{RedEdge}}$ |                                                                                 | [22]      |
| Normalized Difference Built-up Index (NDBI).             | $\frac{(\rho_{SWIR2} - \rho_{NIR})}{(\rho_{SWIR2} + \rho_{NIR})}$ | Index adjusted to extract information about impervious surfaces in urban areas. | [19]      |

IR: infrared; R: red;  $\rho$ : reflectance at specific wavelength; SWIR2: surface reflectance of the second shortwave infrared band; NIR: surface reflectance of the near-infrared band.

Source: author.

#### IV. DISCUSSION

Based on the metrics of Accuracy and the Kappa coefficient, the poorest result was obtained with the SVM classifier using multispectral images from the Sentinel-2 satellites. The SVM classifier showed significant difficulty in distinguishing pasture areas from coconut cultivation areas, achieving a Kappa of 71.9%. The RF classifier was better able to differentiate between pasture and coconut cultivation areas. The texture characteristic was better captured by the RF classifier when using multispectral images.

Using radar images, the RF classifier also achieved better classification metrics for coconut cultivation areas, with a Kappa of 93.3%. It is noteworthy that the RF classifier with radar images from the Sentinel-1

satellite classified coconut areas into smaller classification units (Figure 4); however, the coconut cultivation regions corresponded to the same areas identified with multispectral optical imagery.

The NDVI and NDBI vegetation indices provided the strongest support for image classification, both for optical and radar images. The NDVI index, among other effects, highlights regions in the images with photosynthetically active plants and distinguishes them from senescent vegetation, which helps the classifier to segment coconut cultivation areas from other land uses. The NDBI index highlights regions with impervious surfaces, such as urban areas, thus facilitating the classifier in separating impervious areas from permeable areas cultivated with coconut palms.

#### V. CONCLUSION

The results obtained in this study demonstrate the effectiveness of applying remote sensing techniques and machine learning algorithms to map coconut cultivation areas in the municipality of Paraipaba, Ceará. Among the evaluated methods, the Random Forest (RF) classifier showed the best performance, particularly when applied to radar images from Sentinel-1, achieving an Overall Accuracy of 98.6% and a Kappa coefficient of 93.3%.

The superiority of RF over SVM was evident in distinguishing between pasture areas and coconut plantations, especially due to the algorithm's ability to handle the spectral and textural variability of the images. The vegetation indices NDVI and NDBI significantly contributed to classification accuracy by improving class separability, highlighting key features such as photosynthetically active vegetation and impervious surfaces.

The integration of optical and radar imagery, combined with robust classifiers, proves to be a promising strategy for monitoring perennial crops in tropical semi-arid regions. The findings of this study can support agricultural planning, land management, and the development of public policies

aimed at strengthening the coconut production chain in semi-arid regions.

#### REFERENCES

- [1]. Martins, C. R.; Jesus Júnior, L. A., 2011. "Evolução da produção de coco no Brasil e o comércio internacional: panorama 2010". Embrapa Tabuleiros Costeiros-Documents. Disponível em: <https://www.infoteca.cnptia.embrapa.br/Bits>. Acesso em: 10 dez. 2024.
- [2]. Muniz, V. I. M. S.; Braga, P. E. T.; Alves, J. E., 2020. "Frequência de coletores florais do coqueiro (*Cocos nucifera* L.) no litoral do município de Acaraú (Ceará Brasil)". Revista Brasileira de Meio Ambiente, v.8, n.2, p. 1-10.
- [3]. Brainer, M. S. C. P., 2021. "Coco: produção e mercado". Cadernos setorial ETENE, Banco do Nordeste, ano 6, n.206, 13p.
- [4]. Jensen, J. R., 2000. "Remote sensing of the environment: an earth resource perspective". New Jersey: Prentice Hall, 544p.
- [5]. Marin, D. B.; Alves, M. C.; Pozza, E. A.; Gandia, R. M.; Cortez, M. L. J.; Mattioli, M. C., 2019. "Sensoriamento remoto multiespectral na identificação e mapeamento das variáveis bióticas e

- abióticas do cafeeiro”. Revista Ceres, Viçosa, v.66, n.2, p.142-153.
- [6]. Bastiaanssen, W. G. M., 2000. “Sebal - Based sensible and latent heat fluxes in the irrigated Gediz Basin, Turkey”. Journal of Hydrology, Amsterdam, v.229, n.12, p.87-100, 2000.
- [7]. Bastiaanssen, W. G. M.; Noordman, E. J. M.; Pelgrum, H.; Davids, G.; Thoreson, B. P.; Allen, R. G., 2005. “SEBAL model with remotely sensed data to improve water-resources management under actual field conditions”. Journal of Irrigation and Drainage Engineering, New York, v.131, n.1, p.85-93.
- [8]. Gorelick N., Hancher M., Dixon M., 2017. “Google Earth Engine: Planetary-scale geospatial analysis for everyone”. Remote Sensing of Environment, v.202, p.18-27.
- [9]. Koskinen, J., Leinonen, U., Vollrath, A., 2019. “Participatory mapping of forest plantations with Open Foris and Google Earth Engine”. ISPRS Journal of Photogrammetry and Remote Sensing, v.148, p.63-74.
- [10]. Zhao, Z.; Islam, F.; Liaqat, L. A.; Tariq, A.; Nawaz, M.; Islam, I. U.; Bibi, T.; Rehman, N. U.; Ahmad, W.; Aslam, R. W.; Raza, D.; Hatamleh, W. A., 2024. “Comparison of three machine learning algorithms using Google Earth Engine for land use land cover classification”. Rangeland Ecology & Management, v.92, p.129-137.
- [11]. Avci, C. et al., 2023. “Comparison between random forest and support vector machine algorithms for LULC classification”. International Journal of Engineering and Geosciences, v. 8, n. 1, p. 1-10.
- [12]. Tan, Y.C.; Duarte, L.; Teodoro, A.C., 2024. “Comparative Study of Random Forest and Support Vector Machine for Land Cover Classification and Post-Wildfire Change Detection”. Land, v.13, n.1878, p.1-24.
- [13]. Ghimire, B.; Rogan, J.; Galiano, V.R.; Panday, P.; Neeti, N., 2012. “An evaluation of bagging, boosting, and random forests for land-cover classification in Cape Cod, Massachusetts, USA”. Giscience Remote Sens. 49, 623-643.
- [14]. Wulder, M.A.; Coops, N.C.; Roy, D.P.; White, J.C.; Hermosilla, T., 2018. “Land cover 2.0”. Int. J. Remote Sens. 39, 4254-4284.
- [15]. Zhang, W. et al., 2019. “From woody cover to woody canopies: How Sentinel-1 and Sentinel-2 data advance the mapping of woody plants in savannas”. Remote Sensing of Environment, v. 234, p. 111465.
- [16]. Mutanga, O.; Dube, T.; Galal, O., 2017. “Remote sensing of crop health for food security in Africa: Potentials and constraints”. Remote Sensing Applications: Society and Environment, Amsterdam, v. 8, p. 231-239.
- [17]. Lopes, M. et al., 2020. “Combining optical and radar satellite image time series to map natural vegetation: savannas as an example”. Remote Sensing in Ecology and Conservation, v. 6, n. 3, p. 316-326.
- [18]. Alencar, B. P. B., 2013. “Análise de áreas de sombras em imagens ALOS/PALSAR”. 74f. Dissertação (Mestre em Ciências Geodésicas e Tecnologias da Geoinformação) - Universidade Federal de Pernambuco.
- [19]. Kabede, T. A.; Hailu, B. T.; Suryabagavan, K. V., 2022. “Evaluation of spectral built-up indices for impervious surface extraction using Sentinel-2A MSI imageries: A case of Addis Ababa city, Ethiopia”. Environmental Challenges, p. 1-14.
- [20]. EMBRAPA., 2021. “Caracterização e classificação de solos do campo experimental do Curu, Paraipaba, CE”. Documento 196. Fortaleza: Embrapa Agroindústria Tropical, 18p.
- [21]. Rouse, J. W. J.; Haas, H. R.; Schell, A. J.; Deering, W. D., 1973. “Monitoring vegetation systems in the Great Plains with ERTS”. In: NASA special publication 1, p. 309-317.
- [22]. Sims, D. A.; Gamon, J. A., 2002. “Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages”. Remote Sensing of Environment, v. 81, n. 2-3, p. 337-354.
- [23]. Landis, J. R.; Koch, G. G., 1977. “The measurement of observer agreement for categorical data”. Biometrics, v.33, n.1, p.159-174.