

**SÃO PAULO STATE UNIVERSITY  
JABOTICABAL CAMPUS**

**COFFEE CLASSIFICATION ACCORDING TO ITS  
DETACHMENT FORCE: A DECISION TREE-BASED  
APPROACH**

**Mariana Dias Meneses**

Agricultural Engineer

**2023**

**SÃO PAULO STATE UNIVERSITY  
JABOTICABAL CAMPUS**

**COFFEE CLASSIFICATION ACCORDING TO ITS  
DETACHMENT FORCE: A DECISION TREE-BASED  
APPROACH**

**Mariana Dias Meneses**

**Advisor: Prof. Dr. Rouverson Pereira da Silva**

**Co-advisors: Prof. Dr. Welington Gonzaga do Vale**

**Prof. Dr. Glauco de Souza Rolim**

Dissertation submitted to the College of Agricultural and Veterinary Sciences - UNESP, Jaboticabal Campus, as part of the requirements for obtaining the title of MSc. in Agronomy (Soil Sciences).

**2023**

M543c      Meneses, Mariana Dias  
Coffee classification according to its detachment  
force: a decision tree-based approach / Mariana Dias  
Meneses. -- Jaboticabal, 2023  
33 f.

Dissertação (mestrado) - Universidade Estadual  
Paulista (Unesp), Faculdade de Ciências Agrárias e  
Veterinárias, Jaboticabal  
Orientador: Rouverson Pereira da Silva

1. Aprendizado do computador. 2. Café Colheita.  
3. Maturação. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp.  
Biblioteca da Faculdade de Ciências Agrárias e Veterinárias,  
Jaboticabal. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

**CERTIFICADO DE APROVAÇÃO**

TÍTULO DA DISSERTAÇÃO: COFFEE CLASSIFICATION ACCORDING TO IT'S DETACHMENT FORCE: A DECISION TREE-BASED APPROACH

**AUTORA: MARIANA DIAS MENESES**


**ORIENTADOR: ROVERSON PEREIRA DA SILVA**

**COORDENADOR: WELINGTON GONZAGA DO VALE**

**COORDENADOR: EDGARD HENRIQUE COSTA SILVA**

**COORDENADOR: GLAUCO DE SOUZA ROLIM**

Aprovada como parte das exigências para obtenção do Título de Mestra em Agronomia (Ciência do Solo), área: Agronomia (Produção Vegetal) pela Comissão Examinadora:

  
Prof. Dr. ROVERSON PEREIRA DA SILVA (Participação Presencial)  
Departamento de Engenharia e Ciências Exatas DECEX / FCAV UNESP Jaboticabal

  
Prof. Dr. ALAN RODRIGO PANOSSO (Participação Presencial)  
Departamento de Engenharia e Ciências Exatas DECEX / FCAV UNESP Jaboticabal-SP

  
Profa. Dra. NARA SILVEIRA VELLOSO (Participação Presencial)  
Departamento de Engenharia Agrícola / Universidade Federal de Sergipe UFS

Jaboticabal, 28 de fevereiro de 2023

## **AUTHOR'S CURRICULUM INFORMATION**

**MARIANA DIAS MENESES** daughter of Maria da Conceição Dias Meneses and Aguinaldo José de Meneses, born on December 12th, 1995, in Aracaju, Sergipe, Brazil. Graduated from Federal University of Sergipe in agriculture engineering in 2021. During the graduation she participated for two years in the study group in mechanization, automation, and agricultural instrumentation (GEMAIA). In 2021/2 she started her master's degree in Agronomy (Soil Sciences) and focused on machine learning techniques for improve mechanized harvest.

*“It’s a dangerous business Frodo, going out your door. You step onto the road, and if you don’t keep your feet, there’s no knowing where you might be swept off to.”*

J.R.R. Tolkien

## **ACKNOWLEDGEMENTS**

Firstly, I would like to thank my grandmother for all years encouraging me to be an independent and successful woman. Your absence is felt deeply in my heart, and I will miss you forever. Thank you for your unconditional love.

I would like to thank my family, firstly my parents Conceição and Aguinaldo for supporting my dreams and for helping me to get as far as possible. Every day you sacrificed many things so that I could conquer my dreams. You have given me strength and believed on me when I needed the most. Also, to my “sister” Marília. Thank you for your sweet smile and your soft heart, you are the purest and kindest human in this world.

To achieve this professional victory my advisors were indispensable. To Dr. Rouverson Pereira da Silva for trust in my work, for the precious talks to help me in be a better researcher and human. To my co-advisor Dr. Glauco Rolim that helped me in my data analysis.

Achieve the master’s degree would not be possible without my co-advisor and friend, Dr. Welington Gonzaga do Vale, who show me the pleasure to be a researcher and that support my decisions in pursue this career.

Finally, I would like to thank my friends Laura, Jamile, Eduarda, Vanessa, João, and Malvs for the good talks and laughs that make me fell home.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.



## SUMMARY

Chapter 1 – General introduction.....	1
1.1 Introduction .....	1
1.2 Literature rereview.....	1
1.2.1 Coffee: global trends.....	1
1.2.2 Coffee plant uneven maturation.....	2
1.2.3 Mechanized harvesting.....	3
1.2.4 Machine Learning: a subfield of Artificial Intelligence .....	5
1.2.5 Decision Tree Classifier .....	6
1.2.6 Classification performance evaluation .....	8
References.....	9
Chapter 2 – COFFEE CLASSIFICATION ACCORDING TO ITS DETACHMENT FORCE: A DECISION TREE-BASED APPROACH .....	15
1. Introduction .....	15
2. Materials and Methods .....	17
2.1 Experimental Field.....	17
2.2 Detachment Force .....	19
2.3 Descriptive Analysis.....	20
2.4 Cluster Analysis .....	20
2.5 Creation of the conditional for Machine Learning.....	21
2.6 Decision Tree Classifier and Confusion Matrix .....	21
3. Results and Discussion .....	22
3.1 Descriptive Analysis.....	22
3.2 Grouping the cultivars.....	23
3.3 Decision Tree Classifier and Confusion Matrix .....	27
4. Conclusion .....	31

References.....31

## **COFFEE CLASSIFICATION ACCORDING TO ITS DETACHMENT FORCE: A DECISION TREE-BASED APPROACH**

**ABSTRACT** – The world coffee consumption demands high-efficiency crop systems. Consumers appreciate flavor and aroma in this beverage, characteristics that are game-changing in coffee value. A key role to improve this production chain, mechanized harvesting fails in provide coffee fruits selectivity. It means that the industry receives fruits with astringent flavor or fermentation. Because coffee plant has uneven maturation, i.e., green, cherry, and dry fruits, and the harvester settings are generalist, the fruits are detached regardless their maturation stage. The use of Machine Learning techniques improves the traditional agriculture to a digital one, its use in mechanized harvesting enhances selectivity of the coffee fruits. Overall, the present study aimed to classify the coffee fruit detachment force using a Decision Tree Classifier. The experiment was conducted in two field in the Brazilian state of Minas Gerais. A dynamometer was used to measure the detachment force of 23 coffee cultivars. The cultivars were grouped using a cluster algorithm and a Decision Tree classified each group according to the detachment force. The Decision Tree obtained a mean Matthews Correlation Coefficient of 0.81, proving its efficiency in classify the detachment force. Therefore, we proved that Decision Tree can power the mechanized harvesting as a tool to more accurate decision-making settings.

**Keywords:** Machine Learning, Selective Mechanized Harvesting, Uneven Maturation

## **CHAPTER 1 – General introduction**

### **1.1 INTRODUCTION**

Coffee is one of the most famous and popular beverages in the world, brewed from roasted beans the beverage is consumed around the world and Brazil is the larger producer and exporter of the commodity. The harvest is one of the most decisive moments of the crop production and the success on this operation would ensure the quality of the beverage. Currently, mechanized harvesting relies on detaching fruits regardless of its maturity stage, due to the tree uneven maturation. Otherwise, it is imperative to detach mainly the cherry fruit to maintain the high quality of the beverage, characterizing a selective harvesting.

Because machine learning provides ways to overcome challenges by an in-depth association of multi-source data, coffee harvesting would be rethink by a robust analysis of the interaction between the plant uneven maturation and harvester mechanical dynamics. It will be the first step towards an effective yet applicable selective mechanized harvesting. Among the results, the Machine Learning methods can make it possible to: understand the physiological characteristics of the coffee tree, the influence of the environment, and the exact force needed to the green fruits stayed in the tree and the cherry fruits be detached. This type of study makes it possible the integration of automation devices to adapt the coffee harvester and make it capable of performing a selective harvest.

Therefore, our purpose was to use a Machine Learning algorithm to classify the detachment force of coffee fruits. In Chapter 1, we present a brief literature review on the global utilization of coffee, concept of coffee uneven maturity, mechanized harvesting, Machine Learning, decision tree algorithm, and measures to evaluate a classification algorithm. In Chapter 2, we describe the classification of the detachment force using a decision tree algorithm and its classification quality applied to 23 cultivars.

### **1.2 LITERATURE REVIEW**

#### **1.2.1 Coffee: global trends**

The world coffee production (Arabica and Robusta) for 2022/23 is estimated in 10,368 million tons an increase of 396.0 million tons from the last year, and its consumption is expected in 10,074 million tons powered by European Union, United

States, and Brazil demand (USDA, 2022). It is forecast that Brazil will remain the leader in production and export with estimative of 3,756 million tons and 2,202 million tons, respectively, with Minas Gerais and Espírito Santo being the larger producers and exporters of the coffee (USDA, 2022). The coffee consumers describe emotional experiences when drinking it e.g., happiness, pleasure, comfort and calm (BHUMIRATANA; ADHIKARI; CHAMBERS, 2014) yet the most prominent use is for energy and focus enhance. Coffee intake behavior can vary by country and culture, while the consumption is associated with health care e.g., weight control, prevention of cancer, Parkinson's and Alzheimer's disease (PHAN; CHAMBERS, 2016; RAMALHO; SOARES, 2018). Beyond the beverage, the coffee residues generated in the post-harvest process has multiple uses which include but are not limited to: bioethanol production, livestock feeding, antioxidants, vitamin E and organic fertilizer (ANDRADE et al., 2022).

### **1.2.2 Coffee plant uneven maturation**

Environmental specifications to ensure Arabica coffee perfectly growing involves temperatures between 18 and 22° C, rainfall of 1200 to 1800 mm, and altitudes between 1600 to 2800 m. Colombia and some Brazilian regions have this environmental characteristics (ADHIKARI et al., 2020). For these regions, the reproductive period is stimulated by short-days, and the flowering and fruiting occur irregularly over several months (RAKOCEVIC et al., 2020).

In non-equatorial regions as in Brazilian state of Minas Gerais, the coffee plant has more the one blossoming period, with different physiological development of the floral buds within each branch, as a consequence, on the same tree the emergence and development of the fruits occur at different moments, resulting in uneven fruit maturation (DAMATTA et al., 2007). In the fruit maturation, several modifications happen in the coffee plant geometrical, physiological, and chemical characteristics. Because fruit-peduncle elasticity modulus change, the green fruits have different resistance than cherry and dry ones (TINOCO et al., 2014).

As maturation progress, the increasing of enzymatic activity responsible for the cell walls degradation reduce the elasticity modulus of the peduncle structure and its resistance. Otherwise, the peduncle of green fruit has a higher elasticity modulus due

to the amount of nutrient flowing through the structure to stimulate its development (COELHO et al., 2015). This uneven maturation can be intensified by the sunlight. The incidence of light act as an activator of flowering, stimulating the transition from the vegetative to reproductive stage (KAMI et al., 2010).

### **1.2.3 Mechanized harvesting**

Mechanized harvesting it's a solution for the growers who pursue for cost reduction, time saving, and yield increasing (HOSHYARMANESH et al., 2017). The mechanized harvesting by vibration its applied in several agriculture crops e.g., olives, apricot, almond, coffee and citrus (HOSHYARMANESH et al., 2017; LI; LEE; HSU, 2011). Vibration harvesting works by applying a frequency of a mechanical movement in the plant. It movement must be adjusted close to the peduncle's natural frequency to overcome the structure resistance and detach the fruit from the plant (ZHOU et al., 2022). For coffee crop, harvesters have two cylinders with rods in their length to promote the detachment of the coffee fruits (SILVA et al., 2020). The harvester passes over the coffee plants, the rods oscillate and produce a vibration on it, causing the fruit detachment (FURRIEL et al., 2022).

In an attempt to make a selective mechanized harvesting, growers adjust the harvester according to the crop and area characteristic (DE SOUZA et al., 2020). The most common changes in harvester settings are ground speed, frequency and vibration rods amplitude (FERREIRA JÚNIOR et al., 2020). Other occasional changes depends on plant special feature e.g., for the first harvest or for smaller plants, the machine height must be decreased to enable harvesting in the lower parts of the coffee plant (SANTINATO et al., 2015). However, due to the dynamic properties of the coffee plant, the adjustments made do not ensure the selectivity (FERREIRA JÚNIOR et al., 2020).

Methods using image acquisition and machine learning techniques seek to improvement coffee mechanized harvesting. Rosas et al., (2022) used a modified MAPIR Survey 3W camera in a Phantom 4 Pro drone to capture imagens from five coffee fields with four different cultivars. The authors identify the green and cherry fruits and the suitable field for harvesting using nine vegetation indices extracted from the imagens. The indices with more relevance for the harvest were selected using a

Principal Component Analysis. As a result, the CRI, MCARI1, and MTVI1 indices were considered suitable for discriminate the plants with green fruits from the plants with cherry ones. However, the canopy volume and crop yield influenced in the results.

With a smartphone, Bazame et al., (2022) capture imagens from coffee plants to identify the green, cherry, and dry fruits. The imagens were saved with a resolution of 72 dpi. The authors used YOLO algorithms with six different network sizes to detect and classify the fruits. The best accuracy was obtained for the larger network size (800 X 800 pixels), more than 70% of correct classification. All models detected the cherry fruits easier than the green ones, the coffee leaves interfered for the detection of this last one.

The use of filming cameras are also a way to describe the coffee characteristics. Villibor et al., (2016) used a Casio, Exilim EX-FH20, and an electromagnetic shaker to identify the coffee modal parameters: damped oscillation period, damping ratio, undamped and damped natural frequencies, damping coefficient, and stiffness of the system. Using a coordinates system (s,t) in mathematical equations, the authors recorded the movement of green and cherry fruits and its fruit-peduncle behavior in different frequencies. Confusion matrixes was used to classify the operation performance. It was possible to identify which frequency cause the peduncle breaking, consequently the natural frequency of the green and cherry fruits.

Bazame et al., (2021) developed and implement an algorithm to detect and classify coffee fruits during the harvest. The authors used a video camera with 1920x1080 resolution in the spout of the coffee harvester to record the fruits harvested. A YOLO neural networks and a k-means clustering algorithm was used to detect and classify the fruits in three classes: green, cherry, and dry. The performance of the classification was higher than 80%. Despite the accuracy of the method, the tool is better used for pos harvest analysis, not for the pre-harvest decision making. Therefore, the other methods of imagen capture make possible the creation of pre-harvest maps to identify the coffee maturation variability in the field. However, the use of these studies to improve harvester settings remains in theoretical field, still not suitable to provide the selective harvesting.

#### **1.2.4 Machine Learning: a subfield of Artificial Intelligence**

Conceived as the science and engineering of making intelligent machines, the Artificial Intelligence (AI) is a tool that give the computers the ability to solve complex tasks faster and more accurately than any human being (HAMET; TREMBLAY, 2017). As a subfield of the AI, Machine Learning (ML) use mathematical models to solve multidisciplinary problems, and to provide a system the ability to learn and to perform a task by analyzing a dataset behavior (SHARMA; KUMAR, 2017). Overall, the algorithm interacts with the dataset to establish complex relations between the variables (JANIESCH; ZSCHECH; HEINRICH, 2021).

In ML the data is used as set of attributes to describe a phenomenon, regardless of being a categorical, numerical or a binary feature (LIAKOS et al., 2018). The learning process can be classified in two types: supervised, and unsupervised. In supervised learning, input variables are used as a training dataset to predict a quantitative output in a testing data. Otherwise, for the unsupervised models the algorithm must learn from a data without any previous training input (SHARMA; KUMAR, 2017). For these two types of learning, the algorithms are classified as: classification, regression, and clustering.

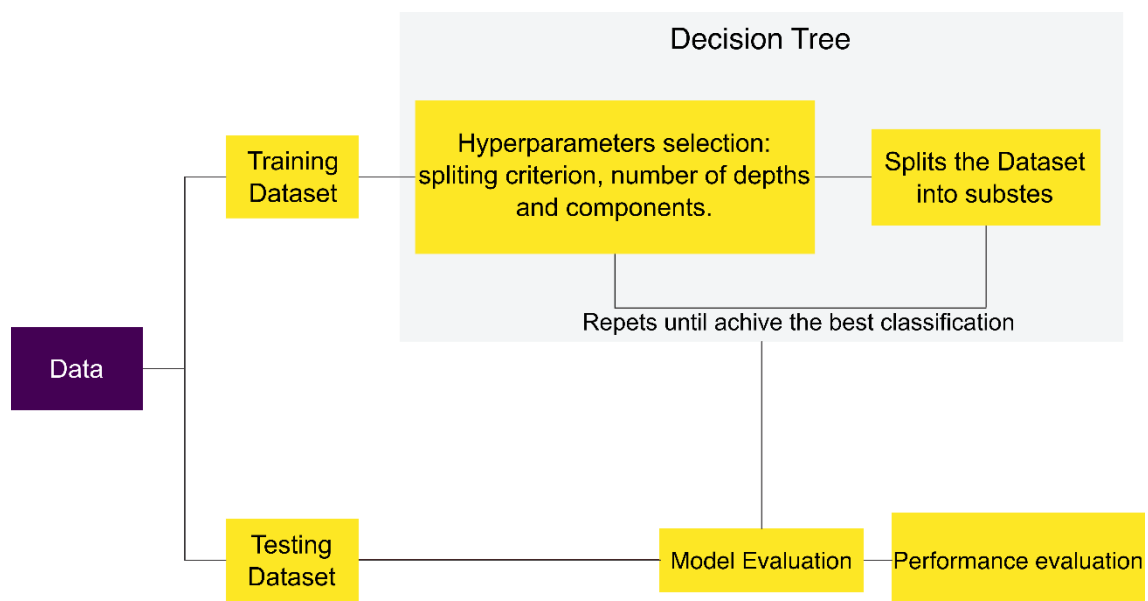
The classification models work as a supervised learning, assigning a label to the dataset whose label are unknown. The process happens in two phases: training the algorithm using a labeled dataset, then testing using a unlabeled dataset (NIKAM, 2015). The regression models are also a supervised learning that predict an output according to a input (LIAKOS et al., 2018). Otherwise, the cluster analysis is a unsupervised method that group the data in categories considering their differences or similarities (MCINTOSH; SHARPE; LAWRIE, 2010).

Farmers need to collect data of sensors, cameras, and online services to progress from the traditional agriculture to smart agriculture. For this purpose, the ML is used to process these amount of data and help to achieve the four pillars of smart farming: optimal resources management, conservation of the ecosystem, adequate service, and utilization of modern technologies (BENOS et al., 2021). The use of ML in smart farming had benefited farmers along the production chain, making the agriculture sustainable, reducing costs, optimizing the use of natural resources and inputs. There are several applications of ML in agriculture e.g., crop disease and yield

prediction, reducing of water waste, soil classification, crop monitoring and tracking, and phenotyping (PALLATHADKA et al., 2021).

### 1.2.5 Decision Tree Classifier

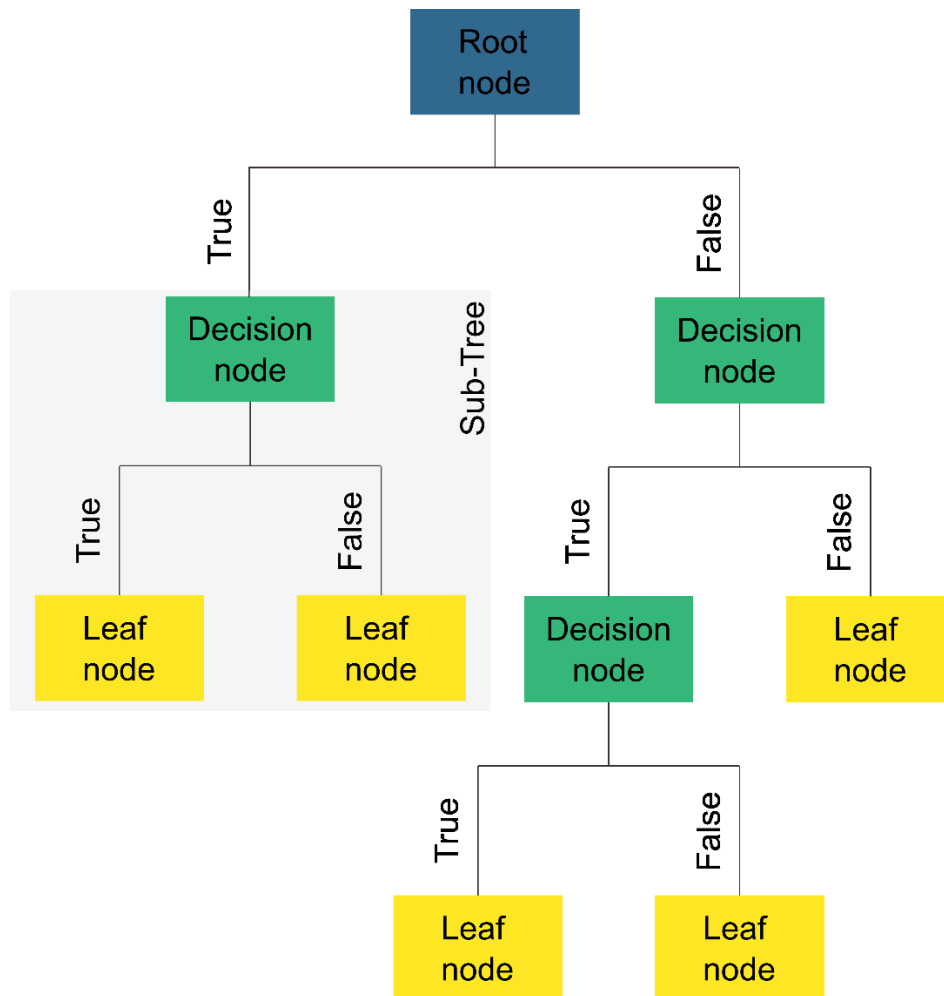
Largely used because of the easy way to understand its results, the Decision Tree Classifier (DT) relies on a series of tests that compare numeric attributes with categorical ones to label classes. The DT is a supervised ML model that split the data domain (node) in two sub nodes. They have more information than the first one with different weights to make the new decision. The ideal architecture of these models is based on splitting the tree in subsets to gain information (SUTHAHARAN, 2016). The hyperparameters are chosen and the DT starts the classification process (Figure 1). Thereafter, the performance of the classification is measured by several metrics.



**Figure 1.** Behavior of the DT for split the data and training and testing dataset.

The DT architecture is structured in nodes and branches. The tree starts with the first test made by the algorithm that splits the root node into two more, the decision nodes. They represent another two tests to split the dataset and after that the final decision generate the leaf node (Figure 2). These nodes are connected by the branches. The left ones represents the true decision while the right branches is the

false decision (SONG; LU, 2015). Each node has the specificity of the category to be classified (CHARBUTY; ABDULAZEEZ, 2021).



**Figure 2.** Decision Tree Classifier architecture. The decisions are made in the nodes, first with the root one and after the decision nodes are created until the final classification in the leaf nodes.

Despite the easy understanding and high performance, the overfitting of the DT constrains the tree applicability (GARCÍA LEIVA et al., 2019). The overfitting phenomenon is the difference among the testing and training performance of supervised algorithms, resulting in a high error because of the generalization of the data (YEOM et al., 2018). To avoid this event, the adjustments of the DT hyperparameters e.g., index, max depth, and number of components, can be a solution by limiting the tree growing (NIE; ZHU; LI, 2020; YUVARAJ et al., 2021). In agriculture, the DT algorithms can assist in coffee leaf rust disease identification, climate modelling

for greenhouse optimization, machinery classification. (CAI et al., 2022; MARIN et al., 2021; RAJESWARI; SUTHENDRAN, 2019; REHMAN et al., 2019).

### 1.2.6 Classification performance evaluation

The ability to correctly classify the class label of an unknown data is measured by accuracy models. The usual way is the Confusion Matrix (CM), a two-dimensions matrix composed by the true class of the object and the predicted one (DENG et al., 2016). The rows represent the true class labels and the columns the predicted ones (Figure 3). A testing dataset is used, and the matrix report the number of True positives (TP), False Negative (FN), False Positive (FP), and True Negative (TN) (CAELEN, 2017). A measure of the classification performance is obtained by comparing the true labels with the predicted by the algorithm (TANGIRALA, 2020).

	Predicted Label		
True Label	TP	FN	TP - Predicted = Positive True = Positive
	FP	TN	FP - Predicted = Positive True = Negative FN - Predicted = Negative True = Positive TN - Predicted = Negative True = Negative

**Figure 3.** Confusion Matrix. The rows represent the actual label of the data and the columns the label predicted by the algorithm.

For the performance evaluation several metrics can be use e.g., F1, Matthews Correlation Coefficient (MCC), and accuracy (MARKOULIDAKIS et al., 2021). The F1 and accuracy are popular metrics. However, considering an imbalanced dataset, with more positive or negative samples, these two metrics fails in consider the relation between the positive and negative cases. Otherwise, MCC does not reduce the capability to correctly identify the classes (CHICCO; TÖTSCH; JURMAN, 2021). The MCC evaluate the predicted labels according to the true ones. It range from -1 to +1,

perfect class misclassification and perfect class classification, respectively (CHICCO; JURMAN, 2020).

## REFERENCES

ADHIKARI, M. et al. A Review of Potential Impacts of Climate Change on Coffee Cultivation and Mycotoxigenic Fungi. **Microorganisms**, v. 8, n. 10, p. 1625, out. 2020.

ANDRADE, M. C. et al. Crop residues: applications of lignocellulosic biomass in the context of a biorefinery. **Frontiers in Energy**, v. 16, n. 2, p. 224–245, 1 abr. 2022.

BARROS, M. M. DE et al. Use of classifier to determine coffee harvest time by detachment force. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 22, p. 366–370, maio 2018.

BAZAME, H. C. et al. Detection, classification, and mapping of coffee fruits during harvest with computer vision. **Computers and Electronics in Agriculture**, v. 183, p. 106066, 1 abr. 2021.

BAZAME, H. C. et al. Detection of coffee fruits on tree branches using computer vision. **Scientia Agricola**, v. 80, 12 set. 2022.

BENOS, L. et al. Machine Learning in Agriculture: A Comprehensive Updated Review. **Sensors**, v. 21, n. 11, p. 3758, jan. 2021.

BHUMIRATANA, N.; ADHIKARI, K.; CHAMBERS, E. The development of an emotion lexicon for the coffee drinking experience. **Food Research International**, Coffee – Science, Technology and Impacts on Human Health. v. 61, p. 83–92, 1 jul. 2014.

BRANDÃO, I. R. et al. Physiological and ultrastructural analysis reveal the absence of a defined abscission zone in coffee fruits. **Bragantia**, v. 75, p. 386–395, 6 out. 2016.

**Brazil: Coffee Semi-annual**. Disponível em: <<https://www.fas.usda.gov/data/brazil-coffee-semi-annual-7>>. Acesso em: 17 jan. 2023.

CAELEN, O. A Bayesian interpretation of the confusion matrix. **Annals of Mathematics and Artificial Intelligence**, v. 81, n. 3, p. 429–450, 1 dez. 2017.

CAI, W. et al. A method for modelling greenhouse temperature using gradient boost decision tree. **Information Processing in Agriculture**, v. 9, n. 3, p. 343–354, 1 set. 2022.

CHARBUTY, B.; ABDULAZEEZ, A. Classification Based on Decision Tree Algorithm for Machine Learning. **Journal of Applied Science and Technology Trends**, v. 2, n. 01, p. 20–28, 24 mar. 2021.

CHICCO, D.; JURMAN, G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. **BMC Genomics**, v. 21, n. 1, p. 6, 2 jan. 2020.

CHICCO, D.; TÖTSCH, N.; JURMAN, G. The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. **BioData Mining**, v. 14, n. 1, p. 13, 4 fev. 2021.

COELHO, A. L. DE F. et al. Determinação das propriedades geométricas, físicas e mecânicas do sistema fruto-pedúnculo-ramo do cafeeiro. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 19, p. 286–292, mar. 2015.

**Coffee: World Markets and Trade.** Disponível em: <<https://www.fas.usda.gov/data/coffee-world-markets-and-trade>>. Acesso em: 5 dez. 2022.

DALMAIJER, E. S.; NORD, C. L.; ASTLE, D. E. Statistical power for cluster analysis. **BMC Bioinformatics**, v. 23, n. 1, p. 205, 31 maio 2022.

DAMATTA, F. M. et al. Ecophysiology of coffee growth and production. **Brazilian Journal of Plant Physiology**, v. 19, p. 485–510, dez. 2007.

DE SOUZA, G. S. et al. Mechanized harvesting of “Conilon” coffee clones. **Pesquisa Agropecuária Brasileira**, v. 55, 2020.

DENG, X. et al. An improved method to construct basic probability assignment based on the confusion matrix for classification problem. **Information Sciences**, v. 340–341, p. 250–261, 1 maio 2016.

FERREIRA JÚNIOR, L. DE G. et al. Characterization of the coffee fruit detachment force in crop subjected to mechanized harvesting. jan. 2018.

FERREIRA JÚNIOR, L. DE G. et al. Dynamic behavior of coffee tree branches during mechanical harvest. **Computers and Electronics in Agriculture**, v. 173, p. 105415, 1 jun. 2020.

FURRIEL, G. P. et al. Acoustics applied in the development of equipment for precision agriculture: Coffee handling and harvesting. **Computers and Electronics in Agriculture**, v. 198, p. 106981, 1 jul. 2022.

GARCÍA LEIVA, R. et al. A Novel Hyperparameter-Free Approach to Decision Tree Construction That Avoids Overfitting by Design. **IEEE Access**, v. 7, p. 99978–99987, 2019.

GODINHO, J. DE D. et al. The best moment to carry out the selective harvest of coffee fruits. **Agronomy Journal**, v. 114, n. 6, p. 3297–3305, 2022.

HAMET, P.; TREMBLAY, J. Artificial intelligence in medicine. **Metabolism, Insights Into the Future of Medicine: Technologies, Concepts, and Integration**. v. 69, p. S36–S40, 1 abr. 2017.

HOSHYARMANESH, H. et al. Numerical and experimental vibration analysis of olive tree for optimal mechanized harvesting efficiency and productivity. **Computers and Electronics in Agriculture**, v. 132, p. 34–48, 1 jan. 2017.

JANIESCH, C.; ZSCHECH, P.; HEINRICH, K. Machine learning and deep learning. **Electronic Markets**, v. 31, n. 3, p. 685–695, 1 set. 2021.

KAMI, C. et al. Chapter Two - Light-Regulated Plant Growth and Development. Em: TIMMERMANS, M. C. P. (Ed.). **Current Topics in Developmental Biology**. Plant Development. [s.l.] Academic Press, 2010. v. 91p. 29–66.

KAZAMA, E. H. et al. Methodology for selective coffee harvesting in management zones of yield and maturation. **Precision Agriculture**, v. 22, n. 3, p. 711–733, 2021.

KOTSIANTIS, S. B. Decision trees: a recent overview. **Artificial Intelligence Review**, v. 39, n. 4, p. 261–283, 1 abr. 2013.

LANGE, L.; HEDDERICH, M. A.; KLAKOW, D. **Feature-Dependent Confusion Matrices for Low-Resource NER Labeling with Noisy Labels**. arXiv, , 4 nov. 2019. Disponível em: <<http://arxiv.org/abs/1910.06061>>. Acesso em: 23 nov. 2022

LI, P.; LEE, S.; HSU, H.-Y. Review on fruit harvesting method for potential use of automatic fruit harvesting systems. **Procedia Engineering**, PEEA 2011. v. 23, p. 351–366, 1 jan. 2011.

LIAKOS, K. G. et al. Machine Learning in Agriculture: A Review. **Sensors**, v. 18, n. 8, p. 2674, ago. 2018.

MARIN, D. B. et al. Detecting coffee leaf rust with UAV-based vegetation indices and decision tree machine learning models. **Computers and Electronics in Agriculture**, v. 190, p. 106476, 1 nov. 2021.

MARKOULIDAKIS, I. et al. Multiclass Confusion Matrix Reduction Method and Its Application on Net Promoter Score Classification Problem. **Technologies**, v. 9, n. 4, p. 81, dez. 2021.

MCINTOSH, A. M.; SHARPE, M.; LAWRIE, S. M. 9 - Research methods, statistics and evidence-based practice. Em: JOHNSTONE, E. C. et al. (Eds.). **Companion to Psychiatric Studies (Eighth Edition)**. St. Louis: Churchill Livingstone, 2010. p. 157–198.

NIE, F.; ZHU, W.; LI, X. Decision Tree SVM: An extension of linear SVM for non-linear classification. **Neurocomputing**, v. 401, p. 153–159, 11 ago. 2020.

NIKAM, S. S. A Comparative Study of Classification Techniques in Data Mining Algorithms. **Oriental Journal of Computer Science and Technology**, v. 8, n. 1, p. 13–19, 30 abr. 2015.

OLIVEIRA, B. R. DE [UNESP. A luz solar e a agressividade da colheita de café afetam a qualidade da operação? 26 mar. 2021.

PALLATHADKA, H. et al. IMPACT OF MACHINE learning ON Management, healthcare AND AGRICULTURE. **Materials Today: Proceedings**, 22 jul. 2021.

PANHALKAR, A. R.; DOYE, D. D. Optimization of decision trees using modified African buffalo algorithm. **Journal of King Saud University - Computer and Information Sciences**, v. 34, n. 8, Part A, p. 4763–4772, 1 set. 2022.

PÉREZ-ORTIZ, M. et al. A Review of Classification Problems and Algorithms in Renewable Energy Applications. **Energies**, v. 9, n. 8, p. 607, ago. 2016.

PHAN, U. T. X.; CHAMBERS, E. Motivations for choosing various food groups based on individual foods. **Appetite**, v. 105, p. 204–211, 1 out. 2016.

PRIYA, R.; RAMESH, D. ML based sustainable precision agriculture: A future generation perspective. **Sustainable Computing: Informatics and Systems**, v. 28, p. 100439, 1 dez. 2020.

RAJESWARI, S.; SUTHENDRAN, K. C5.0: Advanced Decision Tree (ADT) classification model for agricultural data analysis on cloud. **Computers and Electronics in Agriculture**, v. 156, p. 530–539, 1 jan. 2019.

RAKOCEVIC, M. et al. The vegetative growth assists to reproductive responses of Arabic coffee trees in a long-term FACE experiment. **Plant Growth Regulation**, v. 91, n. 2, p. 305–316, 1 jun. 2020.

RAMALHO, M. E. O.; SOARES, N. M. CAFÉ E SEUS BENEFÍCIOS. **Revista Interface Tecnológica**, v. 15, n. 1, p. 285–292, 30 jun. 2018.

REAY, D. Climate-Smart Coffee. Em: REAY, D. (Ed.). **Climate-Smart Food**. Cham: Springer International Publishing, 2019. p. 93–104.

REHMAN, T. U. et al. Current and future applications of statistical machine learning algorithms for agricultural machine vision systems. **Computers and Electronics in Agriculture**, v. 156, p. 585–605, 1 jan. 2019.

ROSAS, J. T. F. et al. Coffee ripeness monitoring using a UAV-mounted low-cost multispectral camera. **Precision Agriculture**, v. 23, n. 1, p. 300–318, 2022.

SÁGIO, S. A. et al. Identification and expression analysis of ethylene biosynthesis and signaling genes provides insights into the early and late coffee cultivars ripening pathway. **Planta**, v. 239, n. 5, p. 951–963, 1 maio 2014.

SANTIN, M. R. et al. CICLO DE MATURAÇÃO E FORÇA DE DESPRENDIMENTO DOS FRUTOS DE CAFÉ CONILON EM CULTIVO IRRIGADO NO CERRADO. p. 5, 2015.

SANTINATO, F. et al. Colheita mecanizada do café em lavouras de primeira safra. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 19, p. 1215–1219, dez. 2015.

SENINDE, D. R.; CHAMBERS, E. Coffee Flavor: A Review. **Beverages**, v. 6, n. 3, p. 44, set. 2020.

SHARMA, D.; KUMAR, N. A Review on Machine Learning Algorithms, Tasks and Applications. v. 6, p. 2278–1323, 1 out. 2017.

SILVA, C. A. DA et al. Análise experimental em um cilindro de varetas de uma colhedora de café para diagnóstico de falha. **ForScience**, v. 8, n. 2, p. e00632–e00632, 6 out. 2020.

SOARES, L. DOS S. et al. Interaction between climate, flowering and production of dry coffee (*Coffea arabica* L.) in Minas Gerais. **Coffee Science - ISSN 1984-3909**, v. 16, p. e161786–e161786, 15 jun. 2021.

SONG, Y.; LU, Y. Decision tree methods: applications for classification and prediction. **Shanghai Archives of Psychiatry**, v. 27, n. 2, p. 130–135, 25 abr. 2015.

SOUZA, G. S. D. et al. FORÇA DE DESPRENDIMENTO DE FRUTOS DE CAFÉ CONILON. **Pensar Acadêmico**, v. 16, n. 1, p. 6, 2018.

SUTHAHARAN, S. Science of Information. Em: SUTHAHARAN, S. (Ed.). **Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning**. Integrated Series in Information Systems. Boston, MA: Springer US, 2016. p. 1–13.

TANGIRALA, S. Evaluating the Impact of GINI Index and Information Gain on Classification using Decision Tree Classifier Algorithm\*. **International Journal of Advanced Computer Science and Applications**, v. 11, n. 2, 2020.

TINOCO, H. A. et al. Finite element modal analysis of the fruit-peduncle of *Coffea arabica* L. var. Colombia estimating its geometrical and mechanical properties. **Computers and Electronics in Agriculture**, v. 108, p. 17–27, 1 out. 2014.

TINOCO, H. A.; PEÑA, F. M. Finite Element Analysis of *Coffea arabica* L. var. Colombia Fruits for Selective Detachment Using Forced Vibrations. **Vibration**, v. 1, n. 1, p. 207–219, set. 2018.

TKACZYNSKI, A. Segmentation Using Two-Step Cluster Analysis. Em: DIETRICH, T.; RUNDLE-THIELE, S.; KUBACKI, K. (Eds.). **Segmentation in Social Marketing: Process, Methods and Application**. Singapore: Springer, 2017. p. 109–125.

VILLIBOR, G. P. et al. Determination of modal properties of the coffee fruit-stem system using high speed digital video and digital image processing. **Acta Scientiarum. Technology**, v. 38, n. 1, p. 41–48, 1 jan. 2016.

VILLIBOR, G. P. et al. Dynamic behavior of coffee fruit-stem system using modeling of flexible bodies. **Computers and Electronics in Agriculture**, v. 166, p. 105009, 1 nov. 2019.

YEOH, L.; NG, K. S. Future Prospects of Spent Coffee Ground Valorisation Using a Biorefinery Approach. **Resources, Conservation and Recycling**, v. 179, p. 106123, 1 abr. 2022.

YEOM, S. et al. **Privacy Risk in Machine Learning: Analyzing the Connection to Overfitting**. 2018 IEEE 31st Computer Security Foundations Symposium (CSF). **Anais...** Em: 2018 IEEE 31ST COMPUTER SECURITY FOUNDATIONS SYMPOSIUM (CSF). jul. 2018.

YUVARAJ, N. et al. Automatic detection of cyberbullying using multi-feature based artificial intelligence with deep decision tree classification. **Computers & Electrical Engineering**, v. 92, p. 107186, 1 jun. 2021.

ZHOU, J. et al. Finite element explicit dynamics simulation of motion and shedding of jujube fruits under forced vibration. **Computers and Electronics in Agriculture**, v. 198, p. 107009, 1 jul. 2022.

## CHAPTER 2 – COFFEE CLASSIFICATION ACCORDING TO ITS DETACHMENT FORCE: A DECISION TREE-BASED APPROACH

**Abstract:** Coffee aroma and flavor dramatically depends on fruit maturity stage. Because of its climacteric nature, coffee tree has multiple maturity stages at the same time. Although high-efficient, mechanized systems promote total harvesting to reach green, cherry, and dry fruits. It reduces coffee beverage quality by mixing different fruits. Therefore, it is imperative to develop methods and tools towards selective mechanized harvesting. The aim of the present study was to classify coffee cultivars according to fruit detachment force using a decision tree-based approach. It would reduce the subjectivity of the harvester settings and make it possible to prioritize cherry fruits during harvesting. Detachment force data were collected from 23 coffee cultivars. Three stages of maturity were analyzed: green, cherry, and dry, considering the West Side (WSP) East Side (SEP) of the plant. Algorithms and data analysis were developed using Python programming language. A cluster analysis was used to reduce dimensionality and group the cultivars according to the detachment force. Then, a decision tree was adjusted to classify input data in four force classes. The IPR 100 has the cultivar with the highest detachment force in three of the four groups. The Matthews Coefficient from the decision tree classification had an average of 0.86 and 0.81 for test and training dataset, respectively. The method proves useful to classify the fruit detachment force from several cultivars and contributes for future systems embedded in mechanized harvesting.

**Keywords:** Machine Learning, Selective Mechanized Harvesting, Uneven Maturation.

### 1. Introduction

Coffee is one of the most consumed beverage in the world and its consumption tends to increase (YEOH; NG, 2022). Brazil is a leader of production among other strategic countries mostly located between the Cancer and Capricorn tropics (REAY, 2019). Although strongly promoted as an energy enhancement drink, coffee intake has health benefits, e.g., cardiovascular disease prevention, Parkinson's and Alzheimer's prevention (RAMALHO; SOARES, 2018).

The coffee beverage is highly appreciated for its aroma and flavor. However, the crop has uneven maturity due to its climacteric nature (SÁGIO et al., 2014). At harvest time, the coffee tree has fruits with different periods of maturity. Such multi-stage fruits impair the beverage quality, as harvesting green fruits will provide an astringent flavor and the senescent ones are susceptible to fermentation (SENINDE; CHAMBERS, 2020). Moreover, the maturation dynamics dramatically changes the fruit-peduncle system and, therefore, its detachment resistance. Overall, the more immature the fruit, the more force is used to harvest it (COELHO et al., 2015).

It complex, multi-factor condition challenges the mechanized harvesting. A common coffee harvester consists of detaching the fruit-peduncle by a set of ceaseless vibrating rods embedded in a roller. The vibration transmitted through the system must overcome the fruit-peduncle resistance (KAZAMA et al., 2021). However, the system promotes the total harvesting. Because the physical-chemical properties from fruit-peduncle varies for each maturity stage, the mechanized harvesting use a maximum vibration. Therefore, every fruit regardless the maturity stage is harvested. It completely conflicts with the quality standards described before.

The coffee physical characteristics can be analyzed with the help of mathematical models and simulations; however, the results do not reach the practical field and are not aimed at the use of harvesters (TINOCO et al., 2014; TINOCO; PEÑA, 2018). Moreover, it is unknown whether a large group of cultivars have the same fruit-peduncle response. The human decision is a catalyst here by wrong adjusting the amount of force in selective mechanized harvesting. A machine learning-based model would open pathways towards selective mechanized harvesting and fill this gap in the literature.

Machine learning supervised models for classification have an enormous number of applications. These algorithms establishes complex relationships between observed and predicted data to emerge a robust model able to classify a unknowing input (PÉREZ-ORTIZ et al., 2016). A decision tree algorithm is an example of a supervised machine learning method, with the advantage of an easy understanding, evaluation and a high performance with small training dataset when compare with others algorithms (PANHALKAR; DOYE, 2022). In agriculture, decision trees have

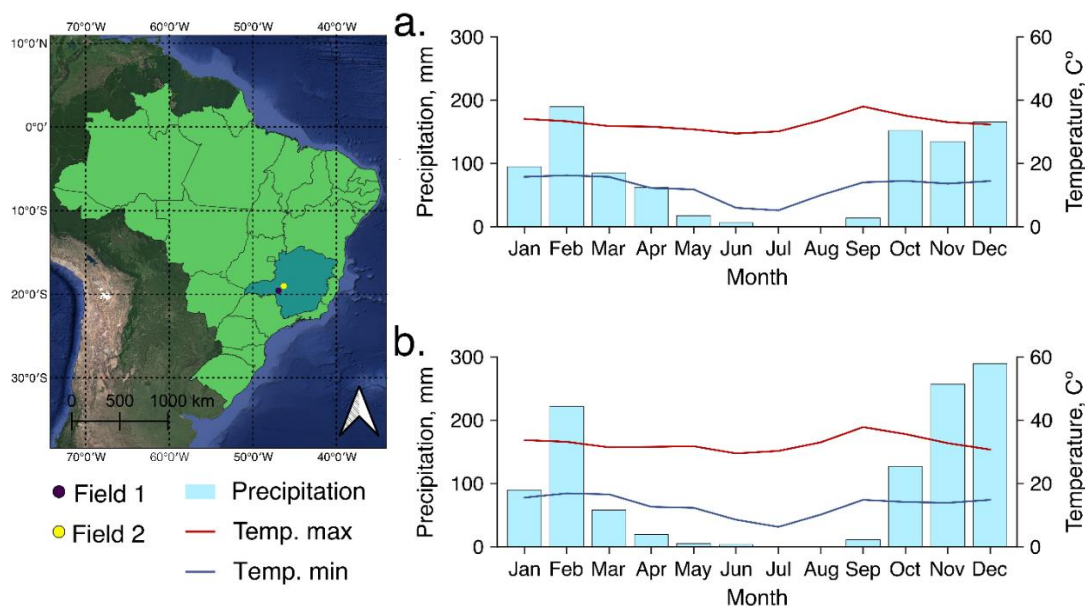
several applications, e.g., identify crop disease, crop yield calculation, market evaluation and crop performance in different environments (PRIYA; RAMESH, 2020).

Therefore, in this study the aim was to classify the detachment force of a large group of coffee cultivars using a decision tree. A framework to classify detachment force would timely support decision-making in harvester adjustments. Combined with the machine's embedded systems it would be possible to choose the speed and vibration settings to achieve the necessary force to harvest the cherry fruits of the chosen cultivar.

## **2. Material and Methods**

### **2.1 Experimental Field**

The experiment was conducted from last week of May to the first week of June 2022, in the municipalities of Araxá (Field 1) and Carmo do Paranaíba (Field 2), southwest region of Minas Gerais state (Figure 1), important coffee producing areas in Brazil. The areas had elevation of 1,001 m and 1,106 m, respectively. In both locations the cultivars were planted in January 2020. Experimental plots were distributed across the areas. The plants analyzed were arranged in a single row in north-south direction, with inter-row spacing of 4.0 meters and intra-row spacing of 0.5 meters regardless the field. Field 1 had drip irrigation with different applications during the growing period, 100% and 50% irrigation during the cycle. Otherwise, Field 2 was not irrigated (Table 1).



**Figure 1.** Geographical location of the two fields. Araxá (Field 1) is approximately 130 km away from Carmo do Paranaíba (Field 2). Monthly precipitation, maximum and minimum temperature for: a. Field 1 (Cwa) and b. Field 2 (Aw) during the year of 2021. Climate zones described according to Köppen-Geiger.

**Table 1.** Coffee cultivars present in the experimental fields and its irrigation system and maturity period.

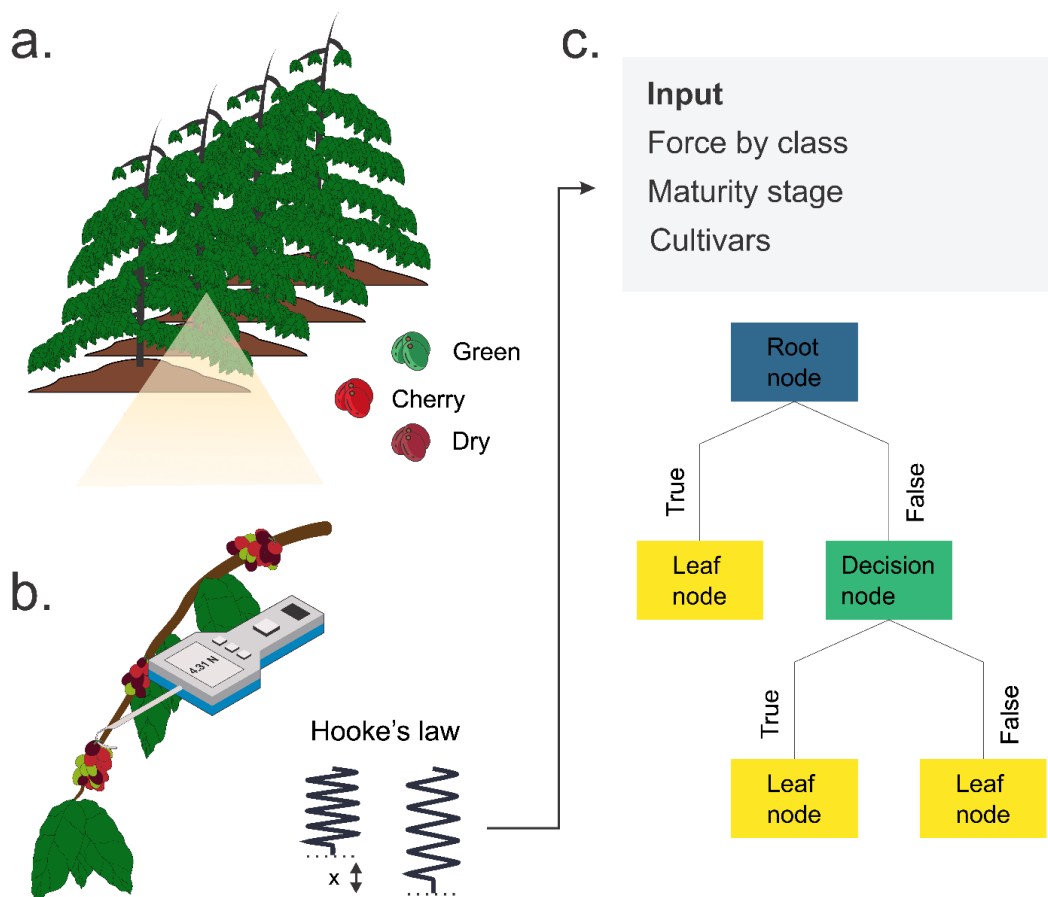
Coffee cultivar	Field		Maturity period	
	1 – Irrigation			2 - Rainfed
	Fully Irrigated	Partially irrigated		
Acauãma		X	Medium - Late	
Acauã JCG	X		Medium – Late	
Arara	X		Late	
Araçari	X		Medium	
Asa Branca	X		Medium	
Azulão	X		Early – Medium	
Beija Flor	X		Early	
Catuai Vermelho	X		Medium – Late	
Catiguá	X		Medium – Late	
IAC 125 RN			Early - Medium	
IAC 3439-4			Early – Medium	
IAC 4520			Early – Medium	
IAC 4932			Early – Medium	
IAC SH3			Early – Medium	
IPR 100	X		Late	
IPR 103	X		Late	
IPR 105	X		Late	
IPR 106		X	Late	
IPR 107		X	Early	

**Table 1.** Coffee cultivars present ... (Cont.)

Coffee cultivar	Field		Maturity period
	1 – Irrigation Fully Irrigated	2 - Rainfed Partially irrigated	
IPR 108	X		Late
Palma 2	X		Late
Palma 3	X		Late
Siriema AS 1		X	Early

## 2.2 Detachment Force

Using a dynamometer (DD-500, Instrutherm), the detachment forces ( $F$ ) of the fruit-peduncle system were measured. The device supports a maximum of 49.03 N, with an accuracy of 0.4% (+/- 1 digit). By positioning the hook on the peduncle and pulling it, the force is measured using Hooke's law (Figure 2).



**Figure 2.** Method to measure the detachment force: a. The plants were randomly chosen and the fruit in the different maturation stages, b. The dynamometer hook was positioned on the fruit and the fruit was pulled and, c. The data of the detachment

force, the force class, maturity stage and coffee cultivar were used as input in the decision tree algorithm.

For each cultivar, four random plants were chosen. Two fruits for each stage of maturity (green, cherry, and dry) were randomly selected from each side of the plant in eight replicates, totalizing 48 repetitions per cultivar. The sides were classified in West Side (WSP) and East Side (ESP) of the plant (GODINHO et al., 2022), the data was collected in the morning with sunlight exposure for ESP and the WSP was on the shade.

An equation was used to determinate the difference between the detachment force of coffee in cherry stage from the green one, used as default measure in each plant:

$$\Delta F (\%) = \left[ \frac{Fg - Fc}{Fg} \right] 100$$

Where,  $\Delta F$  is the difference of the detachment force (%),  $Fg$  is the green fruit detachment force (N),  $Fc$  is the cherry fruit detachment force (N).

### 2.3 Descriptive Analysis

A descriptive analysis was performed in Python (version 3.9.7) for  $F$  to obtain the mean, first, second and third quartiles (Q1, Q2 and Q3), minimum and maximum, difference between quartiles and difference between the means for the cherry and green maturity stages.

### 2.4 Cluster Analysis

Cluster analysis was performed for dimensionality reduction, clustering of cultivars, and group evaluation (DALMAIJER; NORD; ASTLE, 2022). Dimensionality reduction and clustering were performed using the mean force of the coffee cultivars as a parameter, regardless of their maturity or sunlight exposure face. The groups were selected from the hierarchical indentation of the cluster in which the cultivars were most similar (TKACZYNSKI, 2017). The analyses were performed using the scikit-learn library (version 1.1) in the Jupyterlab interface in Python (version 3.9.7) programming language.

## 2.5 Creation of the conditional for Machine Learning

With all force data, a conditional using the three quartile values was created to classify the coffee cultivars by force class. Four force classes were created, and the cultivars could be labeled as:

If  $F < Q1$ ; Force class I;

If  $Q1 \leq F < Q2$ ; Force class II;

If  $Q2 \leq F < Q3$ ; Force class III;

If  $F > Q3$ ; Force class IV.

## 2.6 Decision Tree Classifier and Confusion Matrix

A Decision Tree (DT) was used for classify the cultivars by previously trained classes (PÉREZ-ORTIZ et al., 2016; SHARMA; KUMAR, 2017). The DT classify the data of each group by choosing the best interaction among the attributes analyzed by the algorithm, the nodes presents in the tree structure confirm or deny the function (KOTSIANTIS, 2013). In addition to the force classes, the factors maturity and cultivars were also analyzed as categorical variables to classify the force. Because DT works with this type of variable, the classification was possible (TANGIRALA, 2020).

The scikit-learn library (version 1.1) in Python (version 3.9.7) programming language was used to build the DT algorithm. The dataset was randomly divided in training and testing parameters with 70% and 30% of the dataset of each group, with a number of samples of 46, 16, 8 and, 29 respectively. The criterion, number of decision layers (depth) and components were selected from the GridSearchCV function that generated the appropriate hyperparameters for the algorithm for each time.

To evaluate the classification quality of the training and testing plots, the Matthews Correlation Coefficient (MCC) was used. From the tests of true positives, false positives, false negatives and true negatives, the coefficient generates values from -1 to 1, perfect misclassification and perfect classification, respectively (CHICCO; JURMAN, 2020). Overall, the classification is true positive if the observed data is a positive label and the predicted is also positive. Otherwise, the classification is a false positive if the observed data is a negative label and the predicted is a positive.

Moreover, confusion matrixes were used to illustrate the algorithm's performance to classify each group.

### 3. Results and Discussion

#### 3.1 Descriptive Analysis

The average detachment force for green fruits was 25% higher than the cherry ones (Table 2). Therefore, the maturation progression is inversely proportional to the fruit detachment force, regardless of cultivars (SANTIN et al., 2015; SOUZA et al., 2018). The progression on fruit maturation incites the loss of resistance of the fruit-peduncle system, which changes the detachment force (BRANDÃO et al., 2016).

For green fruits, the amount of nutrients that pass through the peduncle to stimulate fruit development gives the structure a higher elastic modulus and increases resistance (COELHO et al., 2015). As maturation progress, the increase in enzymatic activity degrades the cell wall of the peduncle and decreases the elastic modulus, i.e., ability to deform without breaking, generating a decrease in the structure's strength (VILLIBOR et al., 2019).

**Table 2.** Descriptive analysis of the detachment forces (N) for the three periods of coffee maturity.

Maturity Stage	Mean	Min	Q1	Q2	Q3	Max	Q3-Q1	$\Delta F$
Green	4.67	0.76	2.81	3.76	5.70	15.37	2.89	25%
Cherry	3.74	0.19	2.21	3.04	4.80	16.33	2.59	-
Dry	1.18	0.03	0.62	0.95	1.49	12.96	0.87	-

Min – Minimum force; Max – Maximum force; Q1 – First quartile; Q2 – Second quartile; Q3 – Third quartile; Q3-Q1 – Difference between the third and first quartile;  $\Delta F$  – Difference between the average detachment forces of green and cherry fruits.

Solar irradiance enhances photosynthetic activity and the progression from the vegetative stage to flowering stage by acting as an activator of regulators (KAMI et al., 2010). The incidence of sunlight on different sides of the plant can intensify the irregularity in flowering, which results in different periods of fruit maturity on the coffee plant (SOARES et al., 2021). Therefore, the detachment force may vary for maturity periods and for the WSP and ESP. In cherry fruit, the detachment force was lower for ESP compared to WSP (Table 3). While cherry fruit required 28% less force than green fruit in ESP, the difference decreased to 21% in WSP, which confirms previous results that described this fruit-exposure dynamic (KAZAMA et al., 2021).

**Table 3.** Descriptive analysis of the detachment forces (N) for the three periods of maturation on the WSP and sunlight exposure ESP sides.

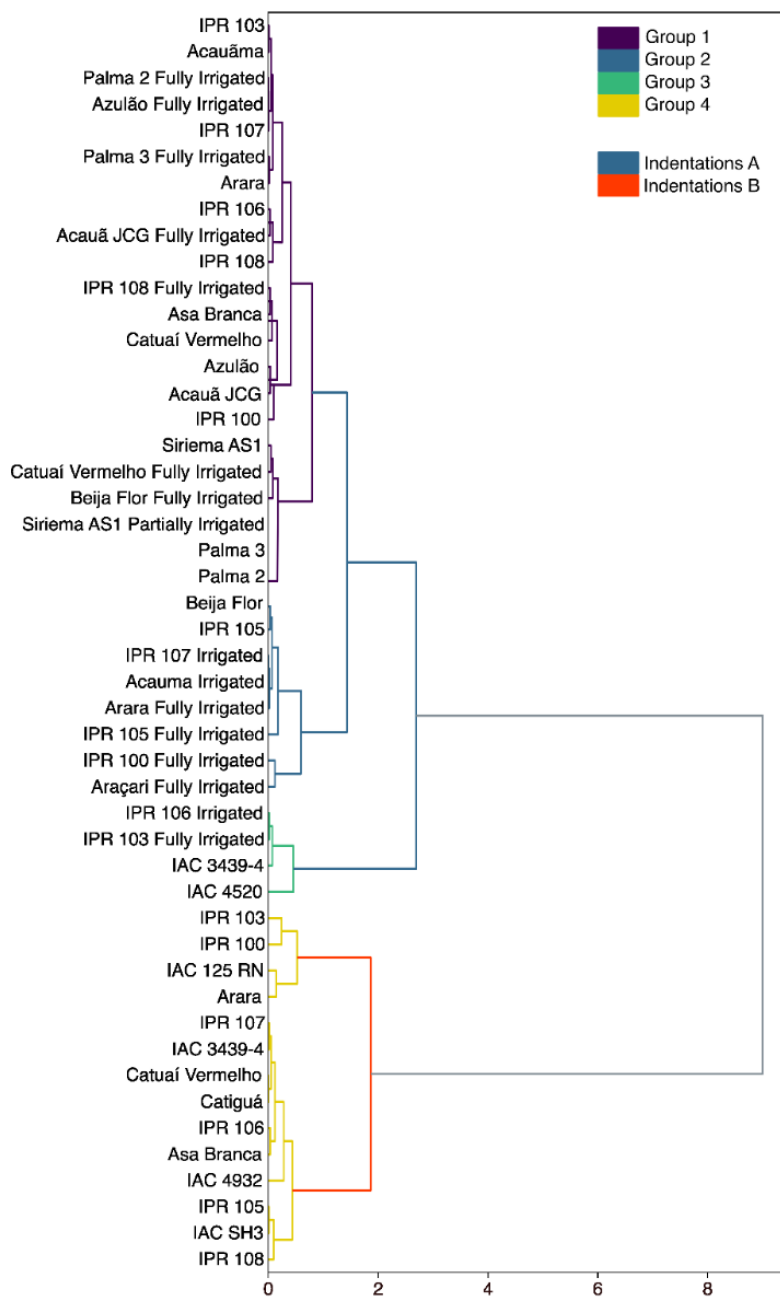
Maturity Stage	Mean	Min	Q1	Q2	Q3	Max	Q3-Q1	$\Delta F$
Green (WSP)	4.61	0.76	2.81	3.76	5.70	15.37	2.89	21%
Cherry (WSP)	3.79	0.40	2.14	3.10	4.65	16.33	2.51	
Green (ESP)	4.73	0.76	2.58	3.75	6.09	14.90	3.51	28%
Cherry (ESP)	3.68	0.19	2.25	3.00	4.80	12.33	2.56	

Min – Minimum force; Max – Maximum force; Q1 – First quartile; Q2 – Second quartile; Q3 – Third quartile; Q3-Q1 – Difference between the third and de first quartile; ESP – East Side of the Plant; WSP – West Side of the plant.

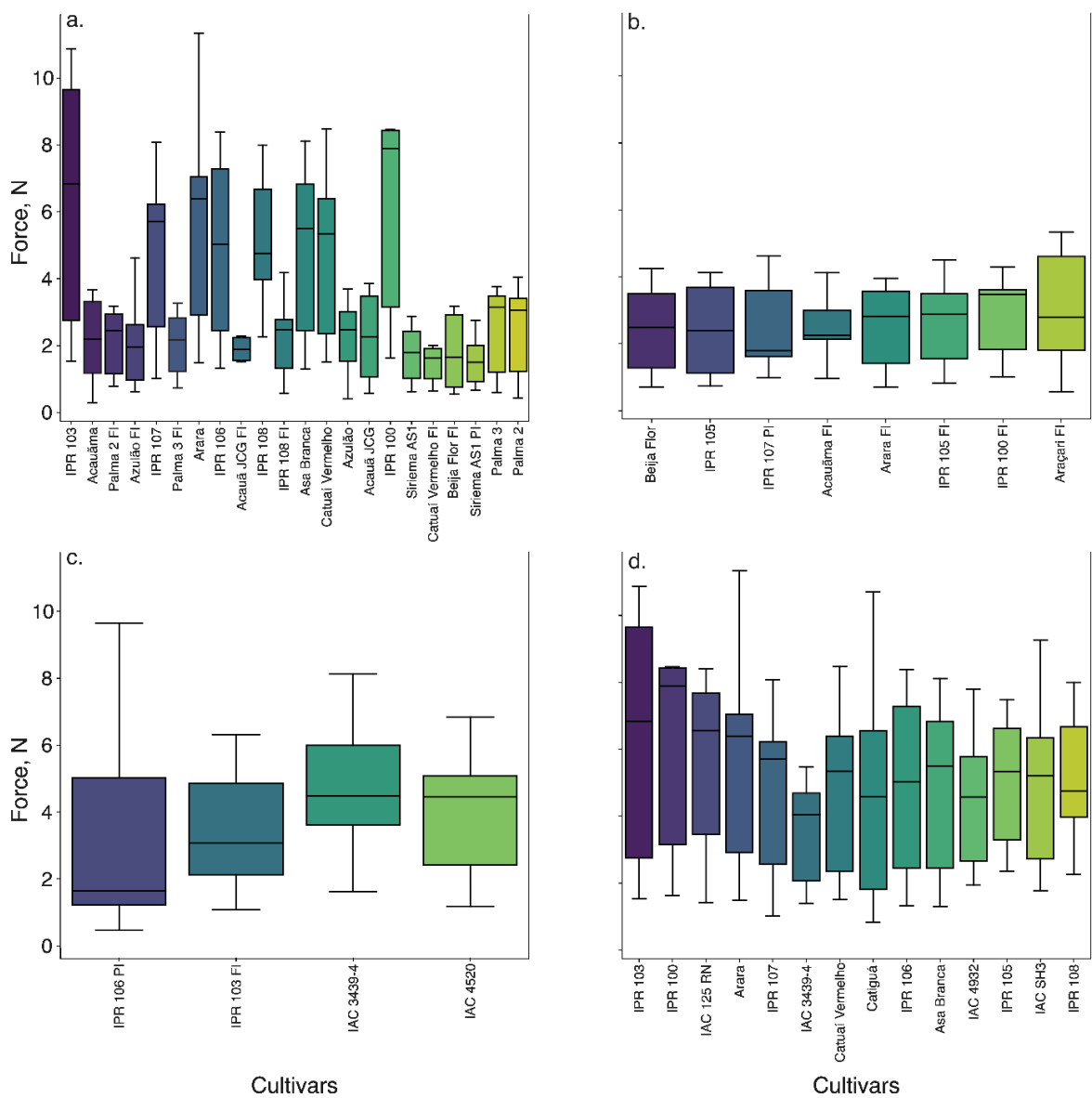
The distinction between the group's benefits mechanized harvesting. Because selective mechanized harvesting relies on picking only cherry fruits, it will open pathways to adapt the harvester according to the force of detachment. A way is the adjustment of the harvesting rods position (FERREIRA JÚNIOR et al., 2018) or changing the plant exposition time i.e., harvester ground speed. The characterization should evaluate the difference between green and cherry fruits, because the more prominent this difference, the greater the contribution for selective mechanized harvesting (Oliveira, 2021). Therefore, the condition of ESP would be more suitable for this activity.

### 3.2 Grouping the cultivars

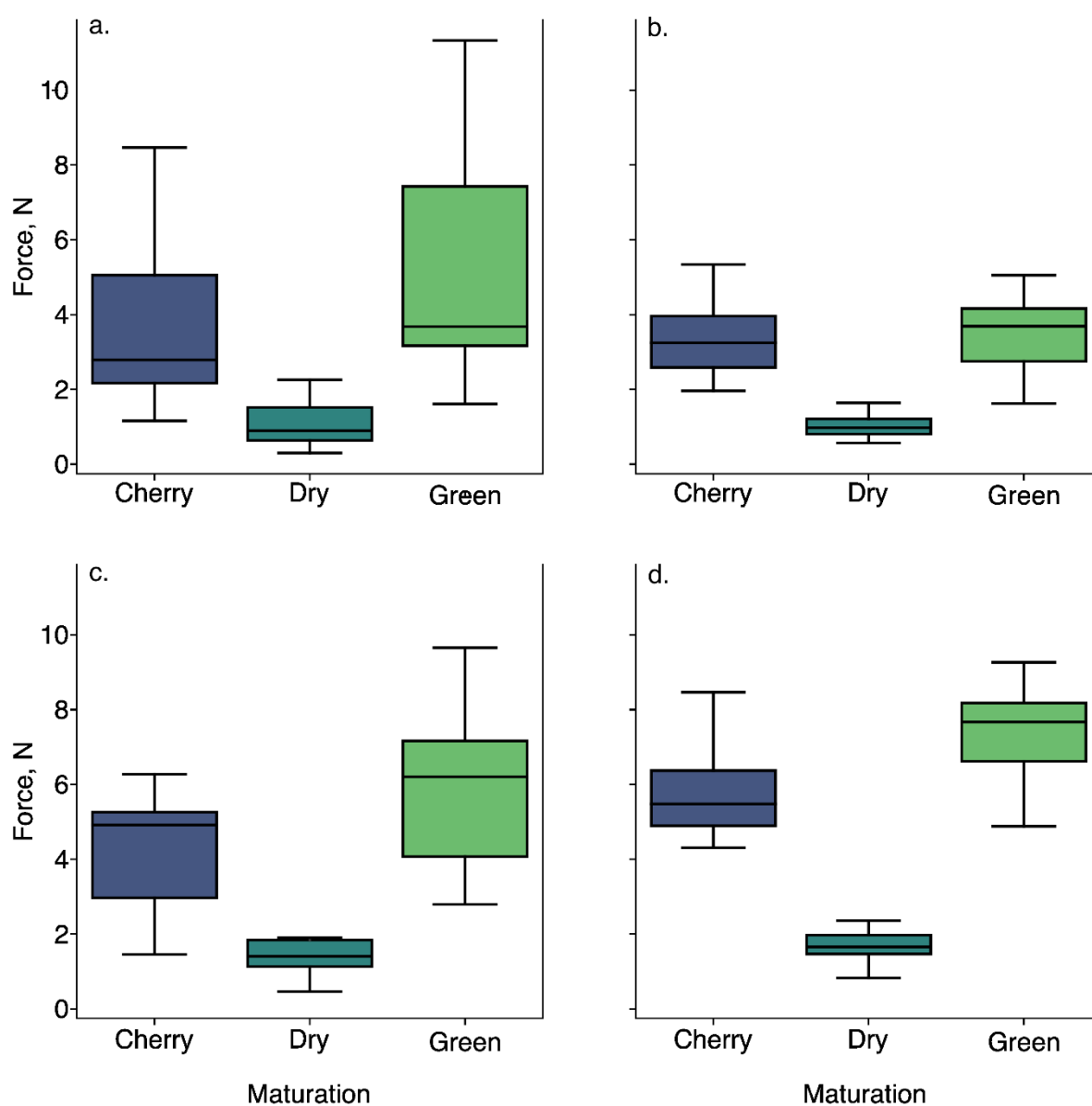
The cluster analysis resulted in two indentations, A and B, and four groups of cultivars (Figure 3). The groups were named as 1 (n = 22), 2 (n = 8), 3 (n = 4) and 4 (n = 14). The Group 2 had the lowest variation of F for each cultivar, while Group 1 had the highest variation (Figure 4). The highest F mean per group was: IPR 100 (Group 1, 2 and 4) and IAC 4520 (Group 3). Otherwise, the lowest was: Siriema AS1 (Group 1), IPR 107 (Group 2), IPR 106 (Group 3) and IAC 3439-4 (Group 4) (Figure 4). In all the groups the green fruits had the highest F, with the Group 4 with the highest F mean and Group 2 with the lowest (Figure 5).



**Figure 3.** Classification of the groups according to the Cluster Analysis.



**Figure 4.** Boxplot of the distribution of F for all cultivars in a. Group1, b. Group 2, c. Group 3 and d. Group 4.



**Figure 5.** Boxplot of the distribution of F for a. Group 1, b. Group 2, c. Group 3 and d. Group 4 determined by the cluster analysis.

The geographical location strongly conditioned the separation. Indentation A included the coffee cultivars from the Field 1, except for IAC 3439-4 and IAC 4520. The other cultivars, from the Field 2, were placed in indentation B. Although present in both areas and under the same irrigation management, cultivars such as IPR 103, IPR 107, IPR 108, Asa Branca, Arara and Catuaí Vermelho were not included in the same group, reaffirming the strong influence of the geographical location.

Clustering was also conditioned on variables intrinsic to plant development e.g., maturation stage, common ancestry, although not used as input data. Overall, cultivars that shared one or more agronomic or management characteristics were likely to be grouped. An important characteristic in each variety is the maturity period. Indentations A and B comprised mostly medium to late maturing plants, except for IPR 107 and Siriema, which are early ones.

The family of origin of the cultivars may also have contributed to the increased homogeneity in the groups. Most cultivars analyzed have a common ancestral, originated from crossbreeding Catuaí Vermelho-Icatu IAC or Catuaí Vermelho-Mundo Novo. The exceptions are Beija Flor and Azulão (originated from Catucaí) and Arara (originated from Obatã IAC 1669-20).

In Group 3, the cultivars IAC 3439-4 and IAC 4520, originated from similar crossings, just as there is similarity between the cultivars IPR 106 and IPR 103. For Group 4, the similarity of the cultivars is on crossings of origin, which are: Catuaí Vermelho-Icatu IAC, Catuaí Vermelho-Mundo Novo, excluding only the variety Arara with origin from Obatã IAC 1669-20.

### **3.3 Decision Tree Classifier and Confusion Matrix**

The DT algorithm had a different number of max depths for each group (Figure 6). The training dataset of Group 1 obtained the best MCC, while Group 3 achieved the lowest coefficients, with emphasis on the ESP standing out, however, still with a positive correlation of 0.65. It is worth emphasizing that the best coefficients were achieved in WSP, except for Group 4 where ESP had the best coefficient (Table 4).



**Table 4.** Matthews Coefficient for the training and test plots of each group for the non-exposure (NES) and sunlight exposure (SES) sides.

Group	Matthews Coefficient			
	Training		Test	
	WSP	ESP	WSP	ESP
1	1	1	0.91	0.93
2	0.85	0.78	0.86	0.72
3	0.80	0.65	0.67	0.65
4	0.88	0.92	0.87	0.88

ESP – East Side of the Plant; WSP – West Side of the plant.

For the test plots the behavior was similar, with the highest coefficients in Group 1 and the lowest in Group 3. The small number of cultivars in this group was a limiting factor for the results obtained by the MCC to be above acceptable (LANGE; HEDDERICH; KLAKOW, 2019). In general, the test pilot had a lower coefficient when compared to training, due to the classification of samples into different force classes than the true ones (Figure 7).

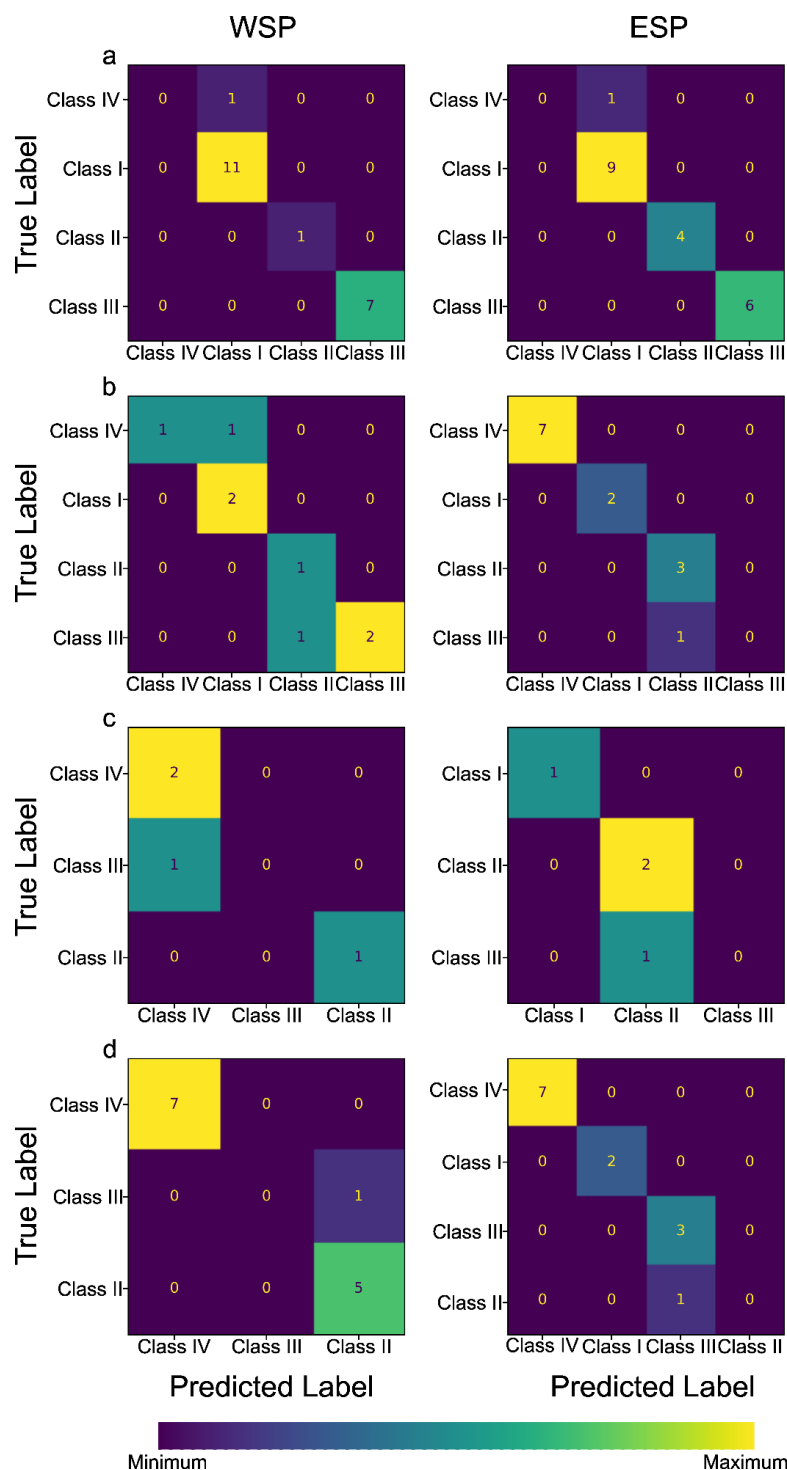


Figure 7. Confusion matrix for the test samples for the WSP and ESP for: a. Group 1, b. Group 2, c. Group 3 and, d. Group 4.

The use of the classifier algorithm allowed the classification according with the force required to detach the fruits. The quality of the classification can vary according to the discrepancy between the detachment forces of green and cherry fruits, the

greater this difference, the lower the possibility of errors in the classification (BARROS et al., 2018). In other applications of DT in coffee culture it is possible to identify leaf rust disease. Using several vegetation indices as variables, leaf samples with four levels of disease severity the DT identify the plants with the disease in the early and later stages (MARIN et al., 2021).

The DT for the F classification can be use with other methods to improve the selective mechanized harvest. With the introduction of methods that allows the harvester to apply different vibrations, the DT can provide different levels of force to be applied by the detaching rods, considering the difference of necessary force caused by de sunlight exposure. Otherwise, the present study needs to be improved with more numerical variables, number of samples per group, more maturity data e.g., percentage of green, cherry fruits and the level of maturation of each fruit.

#### 4. Conclusion

The machine learning approach successfully classify the class detachment force of coffee fruits. The lower classification quality in some groups does not invalidate the method. The insertion of new numerical variables for the factors, maturity stage and percentage of fruits in the different periods, can be used as new inputs to improve the algorithm.

#### References

- BARROS, M. M. DE et al. Use of classifier to determine coffee harvest time by detachment force. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 22, p. 366–370, maio 2018.
- BRANDÃO, I. R. et al. Physiological and ultrastructural analysis reveal the absence of a defined abscission zone in coffee fruits. **Bragantia**, v. 75, p. 386–395, 6 out. 2016.
- CHICCO, D.; JURMAN, G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. **BMC Genomics**, v. 21, n. 1, p. 6, 2 jan. 2020.
- COELHO, A. L. DE F. et al. Determinação das propriedades geométricas, físicas e mecânicas do sistema fruto-pedúnculo-ramo do cafeeiro. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 19, p. 286–292, mar. 2015.
- DALMAIJER, E. S.; NORD, C. L.; ASTLE, D. E. Statistical power for cluster analysis. **BMC Bioinformatics**, v. 23, n. 1, p. 205, 31 maio 2022.

FERREIRA JÚNIOR, L. DE G. et al. Characterization of the coffee fruit detachment force in crop subjected to mechanized harvesting. jan. 2018.

GODINHO, J. DE D. et al. The best moment to carry out the selective harvest of coffee fruits. **Agronomy Journal**, v. 114, n. 6, p. 3297–3305, 2022.

KAMI, C. et al. Chapter Two - Light-Regulated Plant Growth and Development. Em: TIMMERMANS, M. C. P. (Ed.). **Current Topics in Developmental Biology**. Plant Development. [s.l.] Academic Press, 2010. v. 91p. 29–66.

KAZAMA, E. H. et al. Methodology for selective coffee harvesting in management zones of yield and maturation. **Precision Agriculture**, v. 22, n. 3, p. 711–733, 2021.

KOTSIANTIS, S. B. Decision trees: a recent overview. **Artificial Intelligence Review**, v. 39, n. 4, p. 261–283, 1 abr. 2013.

LANGE, L.; HEDDERICH, M. A.; KLAKEW, D. **Feature-Dependent Confusion Matrices for Low-Resource NER Labeling with Noisy Labels**. arXiv, , 4 nov. 2019. Disponível em: <<http://arxiv.org/abs/1910.06061>>. Acesso em: 23 nov. 2022

MARIN, D. B. et al. Detecting coffee leaf rust with UAV-based vegetation indices and decision tree machine learning models. **Computers and Electronics in Agriculture**, v. 190, p. 106476, 1 nov. 2021.

OLIVEIRA, B. R. DE [UNESP. A luz solar e a agressividade da colheita de café afetam a qualidade da operação? 26 mar. 2021.

PANHALKAR, A. R.; DOYE, D. D. Optimization of decision trees using modified African buffalo algorithm. **Journal of King Saud University - Computer and Information Sciences**, v. 34, n. 8, Part A, p. 4763–4772, 1 set. 2022.

PÉREZ-ORTIZ, M. et al. A Review of Classification Problems and Algorithms in Renewable Energy Applications. **Energies**, v. 9, n. 8, p. 607, ago. 2016.

PRIYA, R.; RAMESH, D. ML based sustainable precision agriculture: A future generation perspective. **Sustainable Computing: Informatics and Systems**, v. 28, p. 100439, 1 dez. 2020.

RAMALHO, M. E. O.; SOARES, N. M. CAFÉ E SEUS BENEFÍCIOS. **Revista Interface Tecnológica**, v. 15, n. 1, p. 285–292, 30 jun. 2018.

REAY, D. Climate-Smart Coffee. Em: REAY, D. (Ed.). **Climate-Smart Food**. Cham: Springer International Publishing, 2019. p. 93–104.

SÁGIO, S. A. et al. Identification and expression analysis of ethylene biosynthesis and signaling genes provides insights into the early and late coffee cultivars ripening pathway. **Planta**, v. 239, n. 5, p. 951–963, 1 maio 2014.

SANTIN, M. R. et al. CICLO DE MATURAÇÃO E FORÇA DE DESPRENDIMENTO DOS FRUTOS DE CAFÉ CONILON EM CULTIVO IRRIGADO NO CERRADO. p. 5, 2015.

SENINDE, D. R.; CHAMBERS, E. Coffee Flavor: A Review. **Beverages**, v. 6, n. 3, p. 44, set. 2020.

SHARMA, D.; KUMAR, N. A Review on Machine Learning Algorithms, Tasks and Applications. v. 6, p. 2278–1323, 1 out. 2017.

SOARES, L. DOS S. et al. Interaction between climate, flowering and production of dry coffee (*Coffea arabica* L.) in Minas Gerais. **Coffee Science - ISSN 1984-3909**, v. 16, p. e161786–e161786, 15 jun. 2021.

SOUZA, G. S. D. et al. FORÇA DE DESPRENDIMENTO DE FRUTOS DE CAFÉ CONILON. **Pensar Acadêmico**, v. 16, n. 1, p. 6, 2018.

TANGIRALA, S. Evaluating the Impact of GINI Index and Information Gain on Classification using Decision Tree Classifier Algorithm\*. **International Journal of Advanced Computer Science and Applications**, v. 11, n. 2, 2020.

TINOCO, H. A. et al. Finite element modal analysis of the fruit-peduncle of *Coffea arabica* L. var. Colombia estimating its geometrical and mechanical properties. **Computers and Electronics in Agriculture**, v. 108, p. 17–27, 1 out. 2014.

TINOCO, H. A.; PEÑA, F. M. Finite Element Analysis of *Coffea arabica* L. var. Colombia Fruits for Selective Detachment Using Forced Vibrations. **Vibration**, v. 1, n. 1, p. 207–219, set. 2018.

TKACZYNSKI, A. Segmentation Using Two-Step Cluster Analysis. Em: DIETRICH, T.; RUNDLE-THIELE, S.; KUBACKI, K. (Eds.). **Segmentation in Social Marketing: Process, Methods and Application**. Singapore: Springer, 2017. p. 109–125.

VILLIBOR, G. P. et al. Dynamic behavior of coffee fruit-stem system using modeling of flexible bodies. **Computers and Electronics in Agriculture**, v. 166, p. 105009, 1 nov. 2019.

YEOH, L.; NG, K. S. Future Prospects of Spent Coffee Ground Valorisation Using a Biorefinery Approach. **Resources, Conservation and Recycling**, v. 179, p. 106123, 1 abr. 2022.