ISSN 1807-1929
Revista Brasileira de Engenharia Agrícola e Ambiental
Brazilian Journal of Agricultural and Environmental Engineering
v.26, n.2, p.142-148, 2022
Campina Grande, PB – http://www.agriambi.com.br – http://www.scielo.br/rbeaa

DOI: http://dx.doi.org/10.1590/1807-1929/agriambi.v26n2p142-148

Estimation of percentage of impurities in coffee using a computer vision system[1]

Estimativa do percentual de impureza em café por meio de um sistema de visão computacional

Anderson G. Costa[2*], Eudócio R. O. da Silva[2], Murilo M. de Barros[2] & Jonatthan A. Fagundes[2]

[1] Research developed at Universidade Federal Rural do Rio de Janeiro, Instituto de Tecnologia, Departamento de Engenharia, Seropédica, RJ, Brazil
[2] Universidade Federal Rural do Rio de Janeiro/Instituto de Tecnologia/Pós-Graduação em Engenharia Agrícola e Ambiental, Seropédica, RJ, Brazil

HIGHLIGHTS:
It is possible to accurately estimate the percentage of impurities in ground coffee using digital images.
Colorimetric indices are descriptors with the greatest potential to estimate the percentage of impurities in ground coffee.
The results can be applied to check the quality of coffee.

ABSTRACT: The quality and price of coffee drinks can be affected by contamination with impurities during roasting and grinding. Methods that enable quality control of marketed products are important to meet the standards required by consumers and the industry. The purpose of this study was to estimate the percentage of impurities contained in coffee using textural and colorimetric descriptors obtained from digital images. Arabica coffee beans (*Coffea arabica* L.) at 100% purity were subjected to roasting and grinding processes, and the initially pure ground coffee was gradually contaminated with impurities. Digital images were collected from coffee samples with 0, 10, 30, 50, and 70% impurities. From the images, textural descriptors of the histograms (mean, standard deviation, entropy, uniformity, and third moment) and colorimetric descriptors (RGB color space and HSI color space) were obtained. The principal component regression (PCR) method was applied to the data group of textural and colorimetric descriptors for the development of linear models to estimate coffee impurities. The selected models for the textural descriptors data group and the colorimetric descriptors data group were composed of two and three principal components, respectively. The model from the colorimetric descriptors showed a greater capacity to estimate the percentage of impurities in coffee when compared to the model from the textural descriptors.

Key words: coffee quality, postharvest, principal component regression, image descriptors, non-destructive method

RESUMO: A qualidade e o valor da bebida do café podem ser afetados pela contaminação de impurezas durante a torra e moagem. Métodos que permitem o controle de qualidade do produto comercializado são importantes para atender aos padrões exigidos pelos consumidores e pela indústria. Objetivou-se neste estudo estimar o percentual de impurezas contidas no café usando descritores texturais e colorimétricos obtidos a partir de imagens digitais. Os grãos de café arabica (*Coffea arabica* L.) com 100% de pureza foram submetidos a processos de torra e moagem, e o café moído inicialmente puro foi gradativamente contaminado por impurezas. As imagens digitais foram coletadas em amostras de café com 0, 10, 30, 50 e 70% de impurezas. A partir das imagens, foram obtidos os descritores texturais de histogramas (média, desvio padrão, entropia, uniformidade e terceiro momento) e os descritores colorimétricos (espaço de cores RGB e espaço de cores HSI). O método de regressão de componente principal (PCR) foi aplicado ao grupo de dados de descritores texturais e colorimétricos para o desenvolvimento de modelos lineares para estimativa de impurezas do café. Os modelos selecionados para o grupo de dados dos descritores texturais e o grupo de dados dos descritores colorimétricos foram compostos por dois e três componentes principais, respectivamente. O modelo dos descritores colorimétricos apresentou maior capacidade de estimativa da porcentagem de impurezas no café quando comparado com o modelo dos descritores texturais.

Palavras-chave: qualidade do café, pós-colheita, regressão por componente principal, descritores de imagens, método não destrutivo

• Ref. 248495 – Received 07 Feb, 2021
* Corresponding author - E-mail: andersongc7@gmail.com
• Accepted 01 Aug, 2021 • Published 01 Sept, 2021
Edited by: Carlos Alberto Vieira de Azevedo

Introduction

Compliance with the quality standards of a product is paramount for its acceptance by consumers. Coffee consumption represents a large world-wide market and is one of the largest global commodities (Garcia et al., 2019). Coffee is valued according to qualitative variables (Monteiro et al., 2010).

During harvest and postharvest stages, coffee is subjected to various processes such as selection of beans, roasting, and grinding. In these last stages, impurities and defects related to coffee quality can go unnoticed through visual evaluation (Santos et al., 2016). Purity determination is generally carried out using conventional methods, in a laboratory, involving the use of gas chromatography–mass spectrometry, high-performance liquid chromatography, electron microscopy, and visual analysis on microscopic slides. These methods are costly, destructive, and require analysis time.

The application of techniques that allow the evaluation of agricultural products using non-destructive methods has been successful in product analysis and selection based on quality standards. This is so in the cases of Khoje & Bodhe (2015), who determined the extent of damage to guava skin, Costa et al. (2018), who determined the time of harvest of macaw palm fruits, and Leme et al. (2019), who analyzed the degree of coffee roasting. Systems based on image processing and analysis have successfully evaluated characteristic patterns and predicted phenomena that would be difficult to perceive through visual analysis (Cruz & Silupu, 2014; Bello et al., 2020; Silva et al., 2021).

Computer vision systems have already been applied to analyze the quality of coffee beans (Garcia et al., 2019) and degree of roasting (Leme et al., 2019), and it is possible that image descriptors could also be employed to estimate the percentage of impurities in coffee after grinding. Therefore, this study aimed to evaluate the ability to estimate impurities in ground coffee using textural and colorimetric descriptors obtained from digital images.

Material and Methods

For these experiments, coffee (*Coffea arabica* L.) with 100% Arabica beans without impurities was used. The coffee beans were obtained from a commercial farm located in the city of Dourado, São Paulo, Brazil (22° 03’ 45” S and 48° 15’ 30” W) with an average altitude of 719 m. The coffee beans were subjected to normal roasting and grinding processes and then divided into 20.5 × 15.5 × 2.1 cm trays. Impurities were gradually added to the samples to obtain images of the coffee after the roasting and grinding processes. The percentages of impurities added were 0, 10, 30, 50, and 70% of the total weight of the samples deposited in the trays. For each impurity percentage, 20 images were acquired, totaling 100 images analyzed in the experiment.

The impurities inserted in the coffee came from corn bran, previously submitted to the roasting and grinding processes, aiming to simulate what happens in typical agroindustry and rural properties. According to the Permanent Program of Coffee Purity Control in Brazil, the most common contaminants found in Brazilian coffee are barley and corn (ABIC, 2020).

Images of coffee samples were obtained using a NIKON (Tokyo, Japan) camera model COOLPIX L820 with 16 megapixels, with 2304 × 3456 resolution and visible spectrum, positioned perpendicularly at a height of 0.15 m in relation to the coffee samples. The camera was configured with an ISO 125 shutter and minimum zoom. Camera calibration was performed using the white balance function with a white template to standardize the color intensity throughout the experiments.

The ambient lighting consisted of white LED lamps at a temperature of 6,500 K. Based on the specifications, the lighting system was able to provide a total of 1,080 lumens m^{-1} under constant light flux, corresponding to 3,240 lumens in 3 m of LED (Figure 1).

All images were subjected to a cutout to select the region of interest (ROI) in order to reduce processing time and acquire an image pattern to obtain textural and colorimetric descriptors (Böck et al., 2015). The ROI had a dimension of 2,296 × 3,456 pixels, representing an area of 105.92 cm^2 in the sample.

MATLAB software (Mathworks, United States) was used to obtain textural descriptors of the histograms and colorimetric descriptors. The following textural descriptors were evaluated: mean m (Eq. 1), standard deviation σ (Eq. 2), entropy e(z) (Eq. 3), uniformity, and U(z) (Eq. 4), and the third moment $\mu_3(z)$ (Eq. 5).

$$m = \sum_{i=0}^{L-1} z_i p(z_i) \tag{1}$$

$$\sigma = \sqrt{\sum_{i=0}^{L-1} (z_i - m)^2 p(z_i)} \tag{2}$$

$$e(z) = -\sum_{i=0}^{L-1} p(z_i) \log_2 p(z_i) \tag{3}$$

$$U(z) = \sum_{i=0}^{L-1} p^2(z_i) \tag{4}$$

$$\mu_3(z) = \sum_{i=0}^{L-1} (z_i - m)^3 p(z_i) \tag{5}$$

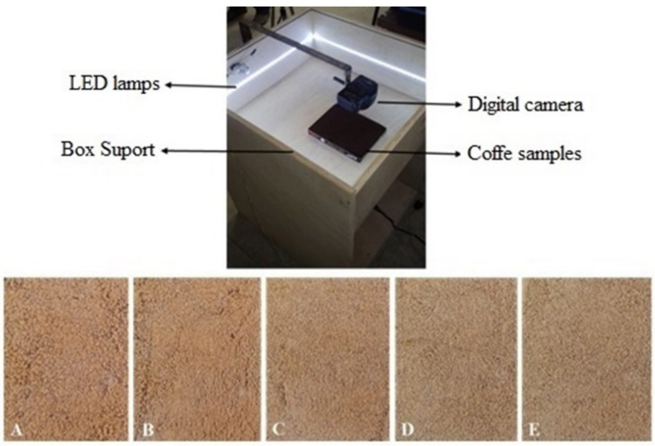


Figure 1. Image acquisition system and coffee samples in five impurities percentages: 0 (A), 10 (B), 30 (C), 50 (D) and 70% (E)

where:

z_i - variable that denotes intensity i;

$p(z_i)$ - probability density function of the corresponding histogram; and

L - number of different levels of gray intensity.

Colorimetric descriptors were obtained based on the average intensity of the decomposition bands of red (R), green (G), and blue (B), in addition to the average value of the RGB image (RGBm). The RGBm values were determined by the average of the intensity of each band (R, G, and B). From the average values of the RGB bands, the values of hue H (Eq. 6), saturation, S (Eq. 7) and intensity I (Eq. 8) were calculated (Gonzalez & Woods, 2017).

$$H = \begin{cases} \theta \text{ if } B \leq G \\ 360 - \theta \text{ if } B > G \end{cases} \qquad (6)$$

In which,

$$\theta = \left\{ \frac{\frac{1}{2}\left[(R-G)+(R-B)\right]}{\left[(R-G)^2 + (R-B)(G-B)^{\frac{1}{2}}\right]} \right\}$$

$$S = 1 - \frac{3}{(R+G+B)}\left[\min(R,G,B)\right] \qquad (7)$$

$$I = \frac{1}{3}(R+G+B) \qquad (8)$$

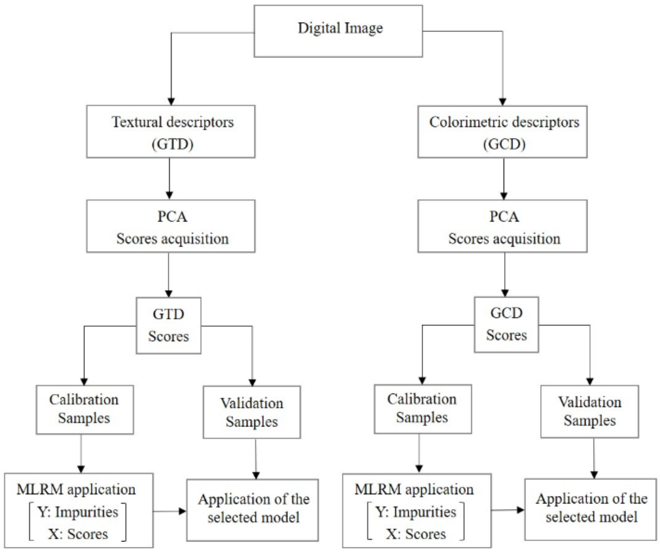
From the obtained results, the data set formed by the group of textural descriptors (GTD) and the group of colorimetric descriptors (GCD) were subjected to descriptive analysis of the coefficient of variation (CV%), standard deviation, minimum, average, and maximum values. Outlier detection was performed using boxplot diagrams. After the descriptive analysis, correlations between the descriptors were assessed using Pearson's linear correlation coefficient ($p \leq 0.01$).

PAST v.3.5 (University of Oslo, Oslo, Norway) software was used to apply multivariate analysis and obtain model parameters. The principal component regression (PCR) method was applied separately to the GTD and GCD data groups.

Initially, the data for each group were normalized through the relationship between the mean and standard deviation to remove the effect of the different units of measurement. After normalization, principal component analysis (PCA) was carried out (Mrówczyńska et al., 2020) from the covariance matrix to obtain the contribution of each principal component and the scores and weights associated with each descriptor.

To obtain the models, scores associated with each principal component were used as independent variables, while the percentage of impurities was used as a dependent variable. The data for each group of descriptors were divided into samples for calibrating the models (scores associated with 70 samples from each group of descriptors) and samples for validating the models (scores associated with 30 samples from each group of descriptors).

From this arrangement, a multiple linear regression was applied to the calibration samples for each data group, generating models for estimating the impurity of coffee samples based on the parameters obtained from the digital images. The best models for each group were selected based on the highest determination coefficient and the lowest standard error for the adjustment for multiple linear regression. Student's t-test at $p \leq 0.05$, was used to select representative coefficients of the best models in each group (Figure 2). The selected models were applied to the set of validation samples (cross-validation applied to 30% of total samples), and the coefficient of determination (R^2) was used to assess relationships between the observed and estimated percentages of coffee impurities.



MLRM - Multiple linear regression model; GTD - Group of textural descriptors; GCD - Group of colorimetric descriptors; PCA - Principal component analysis

Figure 2. Steps used to obtain regression models using principal components for textural and colorimetric descriptors

Results and Discussion

The descriptive results for the GTD (Table 1) demonstrated that the standard deviation, uniformity, and third moment presented the highest coefficients of variation. On the other hand, the descriptors of the GCD presented low coefficients of variation, and the greatest variations were found for the descriptors associated with the RGB color space. In general, it was found that increases in the percentage of impurities in coffee had a greater influence on the repetition of textural patterns of images.

Textural descriptors in digital images represent a region in which structures are formed by the repetition of patterns, in which the elements are arranged according to a composition rule (Armi & Fekri-Ershad, 2019). While textural descriptors such as mean, standard deviation, and third moment tend to quantify variations in patterns, entropy and uniformity tend to scale the degree of organization of pixels in the image, which justifies the low standard deviations observed in Table 1.

A previous analysis of correlations involving the descriptors of each group was carried out to select the parameters that

Table 1. Descriptive statistics of data referring to the group of textural descriptors (GTD) and to the group of colorimetric descriptors (GCD)

Group of textural descriptors (GTD)					
Variable	Mean	Standard deviation	Entropy	Uniformity	Third moment
Minimum	134.984	18.970	6.118	0.006	3.599
Maximum	176.056	47.260	7.545	0.019	22.335
Mean	145.781	31.111	6.906	0.010	10.310
SD	6.976	7.983	0.395	0.003	5.066
CV (%)	4.786	25.660	5.718	31.309	49.139

Group of colorimetric descriptors (GCD)							
Variable	R	G	B	RGBm	H	S	I
Minimum	159.883	126.777	93.673	130.384	22.171	0.153	0.511
Maximum	199.284	170.193	145.267	171.581	30.103	0.294	0.673
Mean	173.351	138.680	109.911	140.647	27.006	0.219	0.552
SD	7.980	7.072	8.679	7.110	1.574	0.032	0.028
CV (%)	4.604	5.099	7.897	5.055	5.830	14.442	5.054

SD - Standard deviation; CV - Coefficient of variation; R - Intensity of red; G - Intensity of green; B - Intensity of blue; RGB – Mean; H – Hue; S – Saturation; I - Intensity

participated in the development of the models (Table 2). All descriptors in the GTD showed a significant correlation with each other, which indicated that they could be used in multivariate regression, except for the mean, which presented a low correlation and could be discarded. When analyzing the descriptors of the GCD, it was observed that hue (H) and saturation (S) showed significant correlation with few variables, in addition to low correlation, indicating that they could also be discarded.

In PCA, correlations involving the original variables represent an important assessment tool for obtaining the best regression model. This analysis allows the selection of those principal components that are highly correlated with the original variables evaluated. Unrepresentative principal components (low correlation with the original variables) may be discarded, generating a simplified regression model.

It is interesting to note that uniformity exhibited a high level of negative correlation with the variables standard deviation, entropy, and the third moment, showing that the greater the uniformity of the samples, the lower the values of these variables. These image patterns indicated a smoother texture of the obtained images.

The description of images in patterns using textural descriptors has been successful in regions of discrimination. Guijarro et al. (2011) used textural descriptors to segment crop areas, soil, and sky regions in an image. From the textural descriptors, classification based on fuzzy clustering showed an average error of between 8.31 and 18.23%, depending on the combination of the image bands.

Different approaches can still be applied to this study to improve the predictive capacity of the models. The use of textural descriptors based on the co-occurrence matrix, such as Haralick's textural descriptors (Gonzalez & Woods, 2017), also allows the evaluation of descriptors obtained from interactions with neighboring pixels. Haralick's textural descriptors have been successfully applied to different areas of knowledge, such as tumor classification, skin texture analysis, land use and forest type classification, and plant leaf classification (Löfstedt et al., 2019).

In addition, an approach based on the fusion of textural descriptors extracted from a co-occurrence matrix with colorimetric properties obtained from digital images, as performed by Mishra et al. (2019) for the classification and sensory evaluation of green tea leaves obtained from different geographical origins using textural and spectral information of their leaves, can improve the accuracy of estimating impurities and classifying coffee quality.

Regarding colorimetric analysis, the descriptors associated with the RGB color space showed the highest correlation values, which demonstrated that this color space is decisive in the development of the models.

Despite the applicability of the RGB color space for the experiments reported here, the use of colorimetric

Table 2. Pearson's correlation coefficients between descriptors in each group

Group of textural descriptors (GTD)					
Variable	Mean	Standard deviation	Entropy	Uniformity	Third moment
Mean	1	-	-	-	-
Standard deviation	-0.350**	1	-	-	-
Entropy	-0.352**	0.991**	1	-	-
Uniformity	0.335**	-0.966**	-0.990**	1	-
Third moment	-0.345**	0.995**	0.974**	-0.941**	1

Group of colorimetric descriptors (GCD)							
	R	G	B	RGBm	H	S	I
R	1	-	-	-	-	-	-
G	0.752**	1	-	-	-	-	-
B	0.491**	0.920*	1	-	-	-	-
RGBm	0.823**	0.988**	0.896**	1	-	-	-
H	-0.410**	-0.018	-0.025	-0.170	1	-	-
S	0.086	-0.548**	-0.821**	-0.483**	-0.166	1	-
I	0.823**	0.988**	0.896**	1.000**	-0.170	-0.483**	1

R - Intensity of red; G - Intensity of green; B - Intensity of blue; RGB - Color space formed by red (R); green (G), and blue (B) intensities; RGBm, RGB mean; H - Hue; S - Saturation; I - Intensity; ** - Significant at p ≤ 0.01 by Student's t-test characteristics from other color spaces, such as CIELAB, can also be evaluated in order to improve the predictive capacity of the model, as observed by Cruz & Silupu (2014), Monje et al. (2018), and Leme et al. (2019).

From the application of PCA to the GTD and GCD data, it was found that it was possible to reduce the number of variables that explain the variation in the parameters obtained in the images owing to the different percentages of impurities in coffee. In Table 3, it is possible to notice that the first components (PC1) for the two data groups present a high percentage of explanation of the variation, indicating that application of PCR can result in simpler mathematical models (with lower numbers of coefficients) to estimate coffee impurities.

The resulting correlations between the principal components and the original variables allow the assessment of the variables that most influence the principal components in the model. In a study by Jolliffe (1972), different methods were presented for the disposal of low influence variables in the analysis of principal components. The author demonstrated that an evaluation of the correlation allows the detection of redundant variables that do not influence the principal component.

Cavalcante et al. (2019) used PCA to assess milk quality and reported that highly correlated original variables tend to influence the same principal component, demonstrating the relevance of this principal component in the generated model.

Table 4 shows that the variables in both groups were predominantly highly correlated with PC1, demonstrating that this principal component has great relevance for the model generated. The mean, hue, and saturation variables were discarded for the generation of the models by presenting low correlations.

The results presented in Table 5 demonstrate that for the GTD, the best-fit calibrated model was composed of the coefficients associated with PC1, PC4, and the constant value, which were significant by Student's t-test ($p \leq 0.05$). For the GCD, the best-fit calibrated model was composed of the coefficients associated with PC1, PC2, PC3, and the constant value, which were significant by Student's t-test ($p \leq 0.05$).

Table 3. Percentage of the total variation explained by each principal component for data from the group of textural descriptors (GTD) and from the group of colorimetric descriptors (GCD)

PC	GTD			GCD		
	Eigenvalue	%E	%AE	Eigenvalue	%E	%AE
PC1	3.929	98.218	98.218	4.456	89.125	89.125
PC2	0.068	1.709	99.927	0.524	10.474	99.599
PC3	0.002	0.060	99.986	0.020	0.401	99.994
PC4	0.001	0.014	100.000	4.80E^{-06}	9.59E^{-05}	99.999
PC5	--	--	--	8.94E^{-20}	1.79E^{-18}	100.000

PC - Principal component; %E - Explained variation; %AE - Accumulated explained variation

Table 4. Correlations between principal components and the original variables for data from the group of textural descriptors (GTD) and from the group of colorimetric descriptors (GCD)

Original variables	PC 1	PC 2	PC 3	PC 4	PC 5
Standard deviation	0.997	0.074	0.016	-0.018	--
Entropy	0.998	-0.057	0.032	0.012	--
Uniformity	-0.983	0.181	0.025	0.003	--
Third moment	0.986	0.163	-0.023	0.009	--
R	0.819	0.572	0.038	0.001	9.36E^{-11}
G	0.989	-0.094	-0.114	0.000	8.29E^{-11}
B	0.898	-0.433	0.075	0.001	1.02E^{-10}
RGBm	1.000	0.007	0.007	0.001	-2.50E-^{10}
I	1.000	0.007	0.007	-0.002	2.16E^{-18}

PC - Principal component; R - Intensity of red; G - Intensity of green; B - Intensity of blue; RGB - Color space formed by red (R), green (G), and blue (B) intensities

The application of artificial vision systems has been successful in assessing the quality of various agricultural products. However, the application of optical techniques in environments that are difficult to control, such as crops and industry, can generate controversial results. Santos et al. (2016) stated that manual manipulation of samples can influence variations in patterns, impacting the reliability of textural analysis of images. The repeatability of readings or use of batches of samples can guarantee a more general perception of the analysis performed.

When comparing adjustment of the calibration models to the data selected for the GTD and GCD, it can be seen that colorimetric descriptors were the most suitable ($R^2 = 0.91$) for such analysis when compared to the textural descriptors ($R^2 = 0.50$). When applying the calibration model to the validation data, it was confirmed that the model generated from the colorimetric descriptors was the most adequate for estimating the percentage of impurities in coffee (Figures 3A and B).

The use of other regression methods can be found in studies that estimated coffee impurities. Ebrahimi-Najafabadi et al. (2012) obtained models with excellent predictive capacity for the amount of barley residue in coffee samples after roasting and grinding at levels of impurities between 2 and 20%. Reis et al. (2013) also applied the partial least squares (PLS) regression method for data obtained from diffuse infrared in coffee samples with levels of impurities from corn and barley residues ranging from 1 to 66%. In both studies, high determination coefficients were obtained for the models ($R^2 > 0.90$), as well as those obtained in this study.

Some authors, such as Winkler-Moser et al. (2015), in addition to the application of PLS in data obtained from

Table 5. Coefficients of calibration models for estimating the coffee impurity percentage for the data from the group of textural descriptors (GTD) and from the group of colorimetric descriptors (GCD)

	GTD				GCD			
	Coeff.	Std. Err.	t	p	Coeff.	Std. Err.	t	p
Constant	32.852	2.340	14.040	2.4E^{-21}*	31.691	0.981	32.308	2.43E^{-41}*
CP 1	5.195	1.203	4.318	5.5E^{-05}*	-1.321	0.437	-3.020	0.0036*
CP 2	1.267	10.177	0.124	0.901	-28.869	1.259	-22.924	1.46E^{-32}*
CP 3	13.585	56.087	0.242	0.809	-88.438	7.035	-12.572	6.03E^{-19}*
CP 4	623.520	109.440	5.697	3.2E^{-07}*	619.980	369.680	1.677	0.0984
CP 5	--	--	--	--	-1.30E^{+08}	3.30E^{+09}	-0.039	0.9687

Coeff. - Coefficient; Std. Err. - Standard prediction error; t-test; * - Significant at $p \leq 0.05$

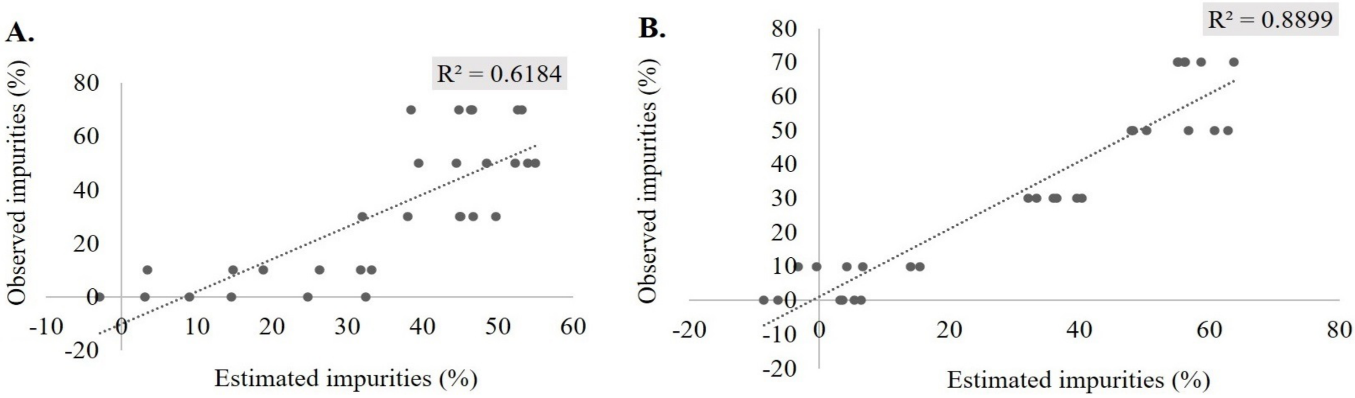


Figure 3. Observed and estimated values of coffee impurities in validation data of models generated from the principal component regression and applied to the group of textural descriptors (GTD) (A) and the group of colorimetric descriptors (GCD) (B)

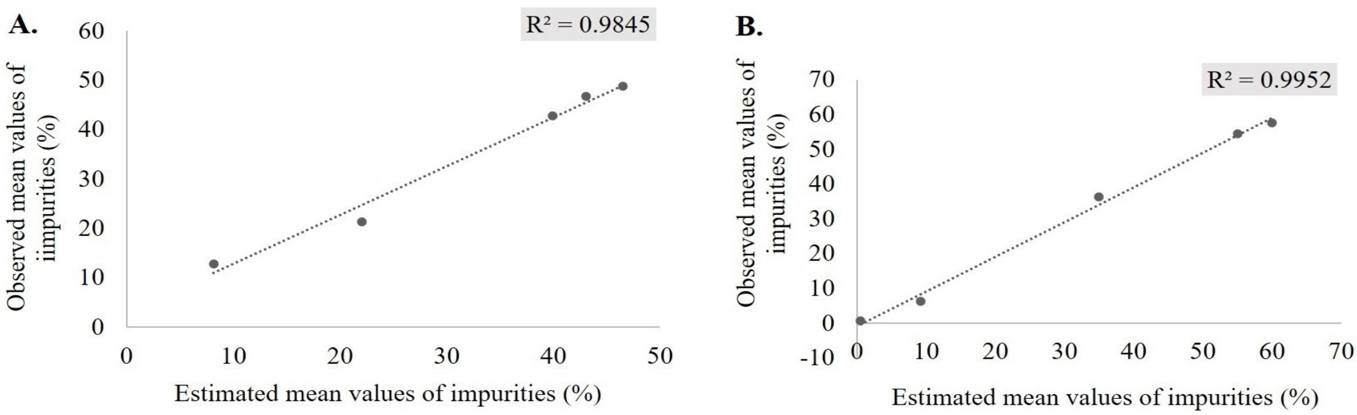


Figure 4. Observed and estimated mean values of coffee impurities according to principal component regression for the group of textural descriptors (GTD) (A) and group of colorimetric descriptors (GCD) (B)

near infrared, used the multiple linear regression model (MLRM), in which the contamination of coffee by corn was evaluated at impurity levels of 0, 1, 5, 10, 15, and 20%. These authors obtained predictive models sensitive enough to detect adulterated samples at levels lower than 5%. The sensitivity of the model was comparable to that of the GCD data group (R^2 = 0.9119).

It is noteworthy that the present study allows the estimation of a wide range of contamination levels of coffee, from the absence of impurity to extreme values (70% of impurities). In the Brazilian domestic market, for example, coffees with up to 40% impurities are commercialized (Martins et al., 2018).

As shown in Figure 4, the average values of the PCR/GTD and PCR/GCD models show R^2 = 0.985 and R^2 = 0.995, respectively, generating estimates of values close to the real values for all levels of impurities in coffee.

The results obtained indicate a promising tool for the evaluation of the quality of commercialized coffee. In the state-of-the-art, the aim is to develop a practical tool for direct measurement that allows the use of a camera in the visible spectrum, and which is capable of quantifying the purity level of the coffee after the roasting and grinding processes, providing coffee farmers an inexpensive and rapid analysis tool.

The lack of investment in modernizing crops and operational costs in the coffee sector pose serious risks to the future of coffee supply (ICO, 2020). Thus, low-cost and instant analysis tools that allow coffee growers to control the quality of their products are fundamental, especially for small coffee producers.

Conclusions

1. The use of an artificial vision system associated with models generated by PCR allowed the measurement of impurity percentages in coffee with high estimation accuracy, in real-time, and non-invasively.

2. The model generated from colorimetric descriptors exhibited a greater capacity to estimate the percentage of impurities in coffee when compared to the model generated from textural descriptors.

Acknowledgements

This study was supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES), the Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) and the Universidade Federal Rural do Rio de Janeiro.

Literature Cited

Armi, L.; Fekri-Ershad, S. Texture image analysis and texture classification methods - A review. International Online Journal of Image Processing and Pattern Recognition, v.2, p.1-29, 2019.

ABIC - Associação Brasileira da Indústria do Café.Programa Permanente de Controle da Pureza do Café: Regulamento. Rio de Janeiro: ABIC, 2020. 10p.

Bello, T. B.; Costa, A. G.; Silva, T. R. da; Paes, J. L.; Oliveira, M. V. M. de. Tomato quality based on colorimetric characteristics of digital images. Revista Brasileira de Engenharia Agrícola e Ambiental, v.24, p.567-572, 2020. https://doi.org/10.1590/1807-1929/agriambi.v24n8p567-572

Böck, F. C.; Assmann, D., Helfer; G. A.; Costa, A. B. A didactic proposal in teaching exploratory data analysis with MDF pictures. Revista Virtual de Química, v.7, p.2475-2486, 2015. https://doi.org/10.5935/1984-6835.20150147

Cavalcante, P. O. S.; Pessoa, R. A. S., Arandas, J. K. G.; Rocha, L. L. da; Ribeiro, M. N. Multivariate approach in the evaluation of the production and composition of buffalo milk in Nordestine Semiarid. Revista Caatinga, v.32, p.1077-1086, 2019. https://doi.org/10.1590/1983-21252019v32n424rc

Costa, A. G.; Pinto, F. de A. de C.; Motoike, S. Y.; Braga Júnior, R. A.; Gracia, L. M. N. Classification of Macaw Palm fruits from colorimetric properties for determining the harvest moment. Engenharia Agrícola, v.38, p.634-641, 2018. https://doi.org/10.1590/1809-4430-eng.agric.v38n4p634-641/2018

Cruz, J. O.; Silupu, W. C. Computer Vision System for the optimization of the color generated by the coffee roasting process according to time, temperature and mesh size. Ingeniería y Universidad, v.18, p.355-368, 2014. https://doi.org/10.11144/Javeriana.IYU18-2.cvso

Ebrahimi-Najafabadi, H.; Leardi, R.; Oliveri, P.; Casolino, M. C.; Jalali-Heravi, M.; Lanteri, S. Detection of addition of barley to coffee using near infrared spectroscopy and chemometric techniques. Talanta, v.99, p.175-179, 2012. https://doi.org/10.1016/j.talanta.2012.05.036

Garcia, M.; Candelo-Becerra, J. E.; Hoyos, F. E. Quality and defect inspection of Green Coffee Beans using a Computer Vision System. Applied Sciences, v.9, p.1-18, 2019. https://doi.org/10.3390/app9194195

Gonzalez, R. C.; Woods, R. E. Digital image processing. 4.ed. New York: Pearson, 2017, 1192p.

Guijarro, M.; Pajares, G.; Riomoros, I.; Herrera, P. J.; Burgos-Artizzu, X. P.; Ribeiro, A. Automatic segmentation of relevant textures in agricultural images. Computers and Electronics in Agriculture, v.75, p.75-83, 2011. https://doi.org/10.1016/j.compag.2010.09.013

ICO - International Coffee Organization. Impact of covid-19 on the global coffee sector: the demand side: Série Coffee Break. London: ICO, 2020. 10p.

Jolliffe, I. T. Discarding variables in a principal component analysis. I: Artificial data. Journal of the Royal Statistical Society: Series C (Applied Statistics) v.21, p.160-173, 1972. https://doi.org/10.2307/2346488

Khoje, S.; Bodhe, S. Comparative performance evaluation of fast discrete curvelet transform and colour texture moments as texture features for fruit skin damage detection. Journal of Food Science and Technology, v.52, p.6914-6926, 2015. https://doi.org/10.1007/s13197-015-1794-3

Leme, D. S.; Silva, S. A. da; Barbosa, B. H. G.; Borém, F. M.; Pereira, R. G. F. A. Recognition of coffee roasting degree using a computer vision system. Computers and Electronics in Agriculture, v.156, p.312-317, 2019. https://doi.org/10.1016/j.compag.2018.11.029

Löfstedt, T.; Brynolfsson, P.; Asklund, T.; Nyholm, T.; Garpebring, A. Gray-level invariant Haralick texture features. Plos One, v.14, p.1-18, 2019. https://doi.org/10.1371/journal.pone.0212110

Martins, V. de C.; Godoy, R. L. de O.; Gouvêa, A. C. M. S.; Santiago, M. C. P. de A.; Borguini, R. G.; Braga, E. C. de O.; Pacheco, S.; Nascimento, L. da S. de M. Fraud investigation in commercial coffee by chromatography. Food Quality and Safety, v.3, p.121-133, 2018. https://doi.org/10.1093/fqsafe/fyy017

Mishra, P.; Nordona, A.; Asaarib, M. S. M.; Lianc, G.; Redfernc, S. Fusing spectral and textural information in near-infrared hyperspectral imaging to improve green tea classification modelling. Journal of Food Engineering, v.249, p.40-47, 2019. https://doi.org/10.1016/j.jfoodeng.2019.01.009

Monje, A. F. B.; Parrado, L. X.; Gutiérrez-Guzmán, N. ATR-FTIR for discrimination of Espresso and Americano Coffee pods. Coffee Science, v.13, p.550-558, 2018. https://doi.org/10.25186/cs.v13i4.1499

Monteiro, M. A. M.; Minim, V. P. R.; Silva, A. F. da; Chaves, J. B. P. Influência da torra sobre a aceitação da bebida café. Revista Ceres, v.57, p.145-150, 2010. https://doi.org/10.1590/S0034-737X2010000200002

Mrówczyńska, M.; Sztubecki, J.; Greinert, A. Compression of results of geodetic displacement measurements using the PCA method and neural networks. Measurement, v.158, p.1-12, 2020. https://doi.org/10.1016/j.measurement.2020.107693

Reis, N.; Franca, A. S.; Oliveira, L. S. Quantitative evaluation of multiple adulterants in roasted coffee by Diffuse Reflectance Infrared Fourier Transform Spectroscopy (DRIFTS) and chemometrics. Talanta, v.115, p.715-722, 2013. https://doi.org/10.1016/j.talanta.2013.06.004

Santos, J. C. F.; Alkimim Filho, J. F. de; Silva, D. V. R. da; Mendes, L. C.; Corrêa, J. A.; Jesus, L. N.; Dibai, W. L. S. Validation of a method to determine the impurities husks and straw in roast ground coffee. Brazilian Journal of Food Technology, v.19, p.1-7, 2016.

Silva, T. M. de A.; Valente, D. M.; Pinto, F. de A. de C.; Queiroz, D. M. de; Santos N. T. Sensor system for acquisition of vegetation indexes. Revista Brasileira de Engenharia Agrícola e Ambiental, v.25, p.264-269, 2021. https://doi.org/10.1590/1807-1929/agriambi.v25n4p264-269

Winkler-Moser, J. K.; Singh, M.; Rennick, K. A.; Bakota, E. L.; Jham, G.; Liu, S. X.; Vaughn, S. F. Detection of corn adulteration in Brazilian Coffee (*Coffea arabica*) by Tocopherol Profiling and Near-Infrared (NIR) Spectroscopy. Journal of Agricultural and Food Chemistry, v.63, p.10662-10668, 2015. https://doi.org/10.1021/acs.jafc.5b04777