



Revista Brasileira de Computação Aplicada, July, 2020

DOI: 10.5335/rbca.v12i2.10866

Vol. 12, N⁰ 2, pp. 54-66

Homepage: seer.upf.br/index.php/rbca/index

ORIGINAL PAPER

A data mining approach for prediction of quality attributes in Palmer mango from images

Daniele Silva Reis ^{[6],1}, Rosalvo Ferreira de Oliveira Neto ^{[6],1}, Josenara Daiane de Souza Costa ^{[6],3}, Acácio Figueiredo Neto ^{[6],1}, Marylia de Sousa Costa ^{[6],2}

Received: 2020-04-17. Revised: 2020-06-09. Accepted: 2020-06-30.

Abstract

The monitoring of quality attributes such as, total soluble solids (TSS), mass, acidity and firmness are essential for a better postharvest conservation of mango. This work proposes a non destructive approach for prediction of those quality attributes using digital images. The proposed approach is composed by three stages: 1) specification of the sampling parameters of mango, 2) identification of digital images pre-processing techniques and 3) utilization of the Random Forest technique as estimator of the quality attributes. In order to validate the proposed approach, a study comparing its performance with models found in literature was carried out. The study used two metrics of performance evaluation: the correlation coefficient (R) and the root mean square error (RMSE). In order to compare the differences of performance between the proposed approach and approaches found in literature, a paired t-student's hypothesis test was carried out. Results show that the proposed approach has a superior performance regarding the existing ones, with confidence level of 95%.

Keywords: Image processing, Non-destructive methods, Random forest, Regression

Resumo

O monitoramento de atributos de qualidade tais como sólidos solúveis totais (SST), massa, acidez e firmeza são essenciais para uma melhor conservação pós-colheita da manga. Este trabalho propõe uma abordagem não-destrutiva para predição desses atributos de qualidade usando imagens digitais. A abordagem proposta é composta de três etapas: 1) especificação dos parâmetros de amostragem da manga, 2) identificação de técnicas de processamento digital de imagens e 3) utilização da técnica Random Forest como estimador dos atributos de qualidades. Para validar a abordagem proposta, foi feito um estudo comparativo entre sua performance e modelos encontrados na literatura. O estudo empregou duas métricas de avaliação de performance: o coeficiente de correlação (R) e a raiz do erro quadrático médio (RMSE). De forma a comparar a diferença de performance entre a abordagem proposta e as encontradas na literatura, foi realizado um teste t de Student pareado. Os resultados mostram que a abordagem proposta teve uma performance superior quando comparada às demais, com um nível de confiança a 95%.

Palavras-Chave: Métodos não-destrutivos; Processamento de imagens; Random Forest; Regressão

1 Introduction

Mango was classified as the most produced tropical fruit in 2017, responding for more than half of the world

production of tropical fruits in that year (Altendorf, 2019). On the next year it was among the four more exported fruits in the world, with huge demand in the two more importing markets, the United States and the

¹Federal University of São Francisco Valley, ²Federal University of Campina Grande, ³Federal Institute of Piauí

^{*}danielesreis@outlook.com; rosalvo.oliveira@univasf.edu.br; josenara.costa@gmail.com; acacio.figueiredo@univasf.edu.br; marylia.sousacosta@gmail.com

European Union, supported by the flavor and consumer preference when compared with more common fruits, such as banana and ananas (FAO, 2018). In Brazil the exports reached 179.7 thousand tons in 2017 (FAO, 2019), being in first place on the world ranking, mainly due to the capacity of perennially producing mango (Altendorf, 2019).

Production of mango is an activity of large expression in the Brazilian fruit growing scene. The greater planted area of mango in Brazil is in the region of the "Vale do São Francisco" and the advancement of the production and exportation resulted in the expansion of this area, going from 27.17 ha in 2017 to 30.30 ha in 2018 (HFBRASIL, 2018). Among the varieties cultivated in the country, the plantation of mango without fiber such as the 'Palmer', 'Keitt' and 'Kent', mainly destined for the European block was greater regarding 'Tommy Atkins' (HFBRASIL, 2019), a scene confirmed in the region of the Valley where 'Palmer' is gaining space, due to new plantations and, also, of the over-grafting in plantation of the variety 'Tommy Atkins' (Trindade et al., 2015).

Having the external market as one of the main consumers of Brazilian mango and at the same time being it a demanding and competitive market, there is a need of doing studies about the maturation process, in order to reach a quality level acceptable for the consumer and a better post-harvest conservation of the mango (Cardoso-Almeida et al., 2017). Traditional and more used methods in the determination of the maturity and quality of the fruits are based in destructive processes. Thus, the development and study of alternative techniques allowing the determination of quality attributes, in a precise and non-invasive way, is of extreme importance (Goulart et al., 2013), making the quality evaluation of the fruits faster, more economical and consistent (Donis-González et al., 2013).

The potential of non-destructive techniques such as tools for evaluation and classification of fruits is being the target of different studies. Modalities of images were investigated for the quality evaluation, from images of the near infrared (NIR) to multi and hyperspectral images, images of reflection of structured illumination, visible images based on monochrome light or black/white to images in color or RGB (red, green and blue) (Li et al., 2015). With the analysis and processing of digital images it is possible to evaluate the change of coloration of the fruits in an objective, integral and representative way, as well as to correlate with the physical-chemical attributes of the pulp (Nagle et al., 2016).

This paper aims to solve the following research question: "How to preview the quality attributes of Palmer mango?". The problem of estimating quality attributes of a mango from an image is a regression problem, since the target variables TSS, Mass, Firmness and Acidity are continuous. Thus, this work proposes a data mining solution in order to solve this problem.

The remaining of this paper is organized as follows. Section 2 presents the related works. Section 3 describes the stages of sample collection,

image acquisition and obtention of reference values. Section 4 presents the experimental results and their interpretation. Finally, Section 5 concludes this paper and proposes future works.

2 Related Works

Studies with mango from different cultivars using image are being developed, however, differences between them and also in the cultivation environment may affect the performance and consistence of maturation indexes, not existing a unanimity or standardization regarding the extracted variables and estimated quality attributes, as summarized in Table 1.

Yahaya et al. (2015) determined attributes total soluble solids, titratable acidity and firmness in mango Sala, extracting the mean values in the RGB space and using as inference technique the MLR (Multiple Linear Regression) (James et al., 2013), obtaining correlation coefficients of 0.814, 0.913 and 0.875, respectively. In studies with mango of variety Carabao, (Abarra et al., 2018) determined the attributes titratable acidity, total sugars, total starch, firmness, TSS and total of reduced sugar, without using image pre-processing. The extracted variables consisted in the mean values of the intensity of pixels in RGB, HSV and L*a*b* spaces, being used as input in models of linear regression for each quality attribute. The best achieved results were for titratable acidity and firmness, when using L* channel only, obtaining correlation coefficients equal to 0.977 and 0.968, respectively. (Khairunniza-Bejo and Kamarudin, 2011) determined TSS in Chokanan mango by the HSB color space, without using preprocessing. Each channel was used as an input in a linear regression, in order to determine the most adequate one. The best result was obtained for hue, with R equal to 0.92. For determination of the mass in mango of Chokanan variety (Teoh and Syaifudin, 2007), 100 mature and green samples were used, whose images were pre-processed by the median filter and segmentation. The extracted variable was the number of pixels corresponding to the mango, being used as input in a linear regression. The obtained value for the correlation coefficient was 0.9769.

It is possible to perceive that there is no consensus about the choice of techniques of image pre-processing and input variables, which may depend upon the analyzed variety and nature of the problem. This reinforces that for every variety the studies may result in a differentiated model, with not being found in literature related works to the 'Palmer' variety, which has a highlight in the international scene.

3 Material and Methods

On the next subsection the stages of sample collection, image acquisition and obtention of reference values are described, needed for the process of knowledge discovery.

Table 1: Developed works with image variables extraction for different varieties of mango.

Authors	Output variable	Variety		Mean differences and RGB gradient	R/G, R/B e S/H		Dominant HSV color	L*a*b* mean	Number of pixels	Fractal variables	Diameter
Teoh and Syaifudin (2007)	Mass	Chokanan							X		
Khairunniza- Bejo and Kamarudin (2011)	TSS	Chokanan				X					
Zheng and Lu (2012)	Browning degree	Sannianmang						X		X	
Nandi et al. (2014)	Maturity	Kumrapali, Amrapali, Sori, Langra, Himsagar	X	X							
Vélez-Rivera et al. (2014)	Maturity	Manila				X		X			
Pandey et al. (2014)	Size and disease	Totapuri, Badami e Neelam						X	X		X
Yahaya et al. (2015)	TSS, titratable acidity and firmness	Sala	X								
Salunkhe and Patil (2015)	Maturity	Alphonso	X		X	X					
Yossy et al. (2017)	Maturity	Gincu					X				
Abarra et al. (2018)	TSS, titrable acidity and firmness	Carabao	Х			X		X			

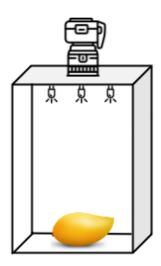


Figure 1: System of image acquisition.

3.1 Samples

Mango cv. 'Palmer' manually collected in a commercial orchard of the "Fazenda Special Fruit Importação e Exportação Ltda.", located in the city of Petrolina–Pernambuco were used, region of climate type BSwh (semiarid, type steppe, very hot, with raining season in summer), according to Koppen's classification, located in 9°18'13,5"S and 40°40'04,7"O, with approximate altitude of 380 m.

The Farm has 114 hectares planted with mango 'Palmer'. The batch being studied has 3.47 ha, with spacing of culture of 6 X 4 m and the system of irrigation with micro-aspersion is used, with daily watering shift and blade adjusted along the cycle. The mango plants received all cultural treatments according to the demands of the culture.

Thirty plants were selected, distributed in five rows of plantation from a batch of the orchard. A total of 750 fruits were collected, in different stages: 35, 50, 65, 80, 95, 110, 125, 140, 165 and 180 days after flowering (DAF), point of commercial harvest adopted by the Farm.

3.2 Image acquisition

The system of acquisition of reflectance images was constituted by the photo camera Canon T5i, box with interior matte black, adjustable power source and a control box for LED lightning, shown in Fig. 1. The illumination system consisted in 3 Solderless LEDs XPE2 of 3W from CREE, cold white 5000K to 8300K, disposed in an angular distance of 120° between them.

For the process of image acquisition one image was obtained for each side of each fruit (considering the resting position), by means of camera adjusted with manual focus, ISO-100; exposition time 1/2s; F/5.6; focal distance of 48mm. The process of image acquisition was made in partnership with the Laboratory of Energy in Agriculture (LENA) from the Federal University of the Valley of São Francisco.

3.3 Obtention of reference values

Before doing the chemical analysis, the fruits were washed in running water, one by one, and immersed in solution of 150mg of chlorine per liter of water during 15 minutes, with posterior rinsing for removal of the excess of chlorine and drying in environmental temperature. Then, the mass of the fruits was determined with the help of a semi-analytical scale with precision of 0.01 g.

Fruit firmness was determined with the help of digital penetrometer model PTR 300, with tip of 6 mm of diameter. One reading by fruit was made, in the equatorial portion, and the result expressed in Newtons (N). The Total Soluble Solids (TSS) were determined, destructively, with the filtered centrifuged pulp, using a digital refractometer (Hanna – HI 96804), results being expressed in ^oBrix. Finally, the titratable acidity was determined by titration with sodium hydroxide solution (0.1 M NaOH) with 1% of phenolphthalein as indicator, also according to the methodology of AOAC (1997).

3.4 Proposed approach

This section describes the approach proposed for the solution of the problem of previewing the quality attributes of mango Palmer. Since it is a solution of data mining, it will be described according to the stages of the process of knowledge discovery KDD (Fayyad and Stolorz, 1997).

3.4.1 Selection

This stage, also known as data sampling, is the process that defines what data will be used for building the solution. Data must represent all diversity of the population of interest. Because of that, this solution proposes to do the capture of the images of mango in each stage of maturation, more specifically, it is proposed that the sample has images of mango with 35, 50, 65, 80, 95, 110, 125, 140, 165 and 180 days after flowering.

3.4.2 Pre-processing

The stage of pre-processing the data is responsible by cleaning the data, in order to eliminate noise and irrelevant information existing in the captured images. For that, the solution used the following sequence of techniques of digital image processing:

1) Median filtering, in order to remove stain contained in the images, 2) Otsu's algorithm to remove the image background, 3) Opening operation to restore small pixels from the image that were removed by previous operations, 4) Simple thresholding for complete removal of the remaining shadows and 5) Closing Morphologic operation for filling contours removed by the previous stage. Fig. 2 illustrates the effect of each technique during the pre-processing stage.

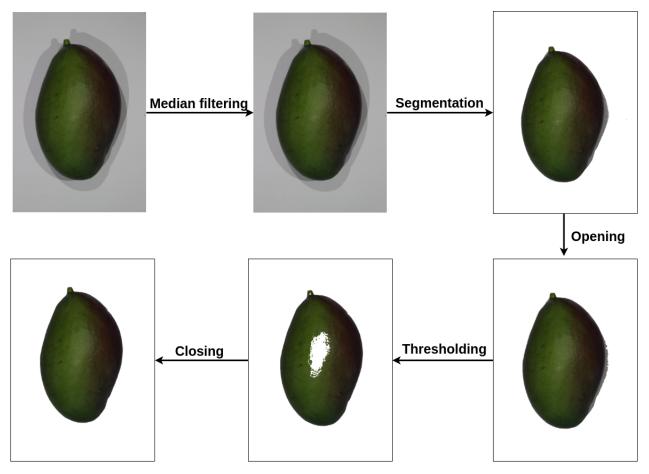


Figure 2: Pre-processing stages of images.

3.4.3 Transformation

According to Pyle (1999), besides the data cleaning, another goal of the pre-processing stage is to transform the data in a format that allows the application of an data mining algorithm. According to (Krogel, 2005), normally this task represents the process of Feature Construction. This process was responsible by building variables from the original data base. For this problem, each digital image was transformed in a vector containing variables with information about the mango. The construction of new input variables is a systematic way of embedding knowledge of the domain in a KDD project. According to Neto et al. (2017), the task of building new variables is much more dependent on the knowledge of the domain than the construction of an estimator. Because of that, domain knowledge is a requirement. Existing approaches use only a small set of input variables as can be seen in the literature review. This approach proposes to enlarge the space of input variables. The strategy of amplification of the sample space of the vector of input variables in order to reach a greater discriminating power is even more common in areas such as Image and Sound Recognition, as can be seen in Gao et al. (2008). The best performing solutions in international competitions used this strategy, as can be seen in Adeodato et al. (2008). From the literature review made, the proposed approach indicates the construction of the following input variables:

- Mean of intensity of pixels in spaces RGB, HSV, L*a*b* (Abarra et al., 2018)
- Dominant HSV color (Yossy et al., 2017)
- Mean of RGB intensities in the whole mango, in regions apex, equator and summit, difference of means and longitudinal gradient (Nandi et al., 2014)
- Number of pixels corresponding to the mango (Teoh and Syaifudin, 2007)
- Mean of intensity of pixels in space RGB (Yahaya et al., 2015)
- Mean of intensity of pixels in space L*a*b* and fractal variables (box counting dimension, dilation dimension and correlation dimension) (Zheng and Lu, 2012)
- Mean of intensity of pixels in space HSV (Khairunniza-Bejo and Kamarudin, 2011)
- Mean of intensity of pixels in spaces HSV and L*a*b* Vélez-Rivera et al. (2014)
- Mean of intensity of pixels in space RGB and HSV and rates R/G, R/B and S/H (Salunkhe and Patil, 2015)
- Mean of intensity of pixels in channel b*, number of pixels and diameter (Pandey et al., 2014)

3.4.4 Data mining

There is a huge variety of data mining techniques. Ngai et al. (2009) indicate in literature more than 30 methods frequently used in scientific papers. Because of that, the identification of a proper technique for the problem of previewing the quality attributes of a mango from an image is a relevant contribution.

The solution proposed in this paper suggests the use of the Random Forest technique as estimator. This technique is based in decision tree, which is one of the most used classifying models in literature due to the easiness of understanding its response, which is organized as a tree and from that one it is possible to easily extract rules of type "If-Then" (Polat and Gunes, 2007). A decision tree uses the strategy of divide and conquer. A problem is decomposed in subproblems and recursively the same strategy is used for each sub-problem. The simplicity of the decision tree also brings disadvantages. The main one of them is the instability caused by noise on the data (Hastie et al., 2009). The Random Forest technique improves stability and precision of the decision tree by incorporating a large number of trees in a single estimator (Breiman, 2001). Random Forest is an ensemble of decision trees, in which the variables that will be used in each tree are randomly selected. The strategy of ensemble used by Random Forest is Bagging. Bagging is an acronym for bootstrap aggregating. Its main idea is to build several individual estimators from a bootstrap sample (sample with replacement of the same size of the training set, in which each example has the same chance of being chosen). The objective is to reduce the variance error of the final estimator, calculating the response as the mean between the estimators of each tree.

3.4.5 Evaluation

This stage includes the validation of the mined content by means of metrics of performance evaluation. To achieve this, the k-fold cross validation method was used, as it is a widely accepted way to split a single sample (Jain et al. (2000)) in k statistically independent test sets, allowing the construction of confidence intervals for the evaluation metric, as recommended by traditional authors Witten and Frank (2005).

The metrics selected in this study were the correlation coefficient and the root mean square error (RMSE). Those indicators measure, respectively, the degree of dependence between input and output variables and the mean magnitude of the estimated errors, according to the following equations:

$$R = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2}}$$
(1)

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

Where x_i is the value of the input variable, \bar{x} is the mean of values of x, y_i is the real value of the output variable, \bar{y} is the mean of y values, n is the number of samples and \hat{y}_i i is the predicted value for the output variable.

3.5 Reference approaches

In order to verify the efficiency of the proposed approach, the achieved results were compared to the

-word - Quantity would wood, in pur variables and interested too mind wood and in interested.									
Quality attribute	Input variables	Inference technique	Authors						
Mass	Number of pixels	Linear regression	Teoh and Syaifudin (2007)						
TSS	Hue mean value	Linear regression	Khairunniza-Bejo and Kamarudin (2011)						
Firmness	L* mean value	Linear regression	Abarra et al. (2018)						
Titratable acidity	L* mean value	Linear regression	Abarra et al. (2018)						

Table 2: Quality attributes, input variables and inference techniques used in literature.

best results found in the literature review made in this study. Since a unique reference model was not found to estimate all quality attributes, the best models for each attribute were selected, as specified in Table 2.

3.6 Statistical analysis

In order to compare the performance of the two approaches a Student's paired t-test was used (Montgomery and Runger (2010)). In this work two hypothesis tests were made: (1) using the difference of the mean value of RMSE (2) using the difference of the mean value of R. Those metrics were obtained for each one of the 5 folds. The configurations of the tests applied were:

Test I

- Null Hypothesis: $\mu_1 = \mu_2$
- Alternative Hypothesis : $\mu_1 < \mu_2$

In which:

- μ₁ represents the RMSE mean for the 5 folds of the proposed approach;
- + μ_2 represents the RMSE mean for the 5 folds of the literature approach

Test II

- Null Hypothesis: $\mu_1 = \mu_2$
- Alternative Hypothesis : $\mu_1 > \mu_2$

In which:

- μ_1 represents the R mean for the 5 folds of the proposed approach;
- μ_2 represents the R mean for the 5 folds of the literature approach

In Fig. 3, the stages for solving the problem of prediction of the quality attributes in mango of variety Palmer are shown.

4 Results and discussion

The descriptive statistics obtained from reference analysis of mass, total soluble solids, titratable acidity and firmness are shown in Table 3. A large variability on data is perceived, insuring a greater robustness to the built models, which made predictions for 'Palmer' mango in different stages of maturation.

In Figs. 4 to 7 the metrics obtained by fold for attributes mass, TSS, firmness and titratable acidity are shown. It is noted that by means of the proposed approach it was possible to achieve more precise models for all quality attributes. The proposed models achieved

greater values of R for all folds of the four attributes. The RMSE values for the proposed approach were smaller for all folds except for the second one in the titrable acidity.

For the mass attribute, the proposed approach and the literature model had a slight difference in both metrics considering the amplitude of the real values as displayed in Table 3. Although the proposed model by the authors Khairunniza-Bejo and Kamarudin (2011) delivered good results with one size based variable, it's noticeable that the addition of color and texture based variables improved the results.

As for the TSS, firmness and titrable acidity results, it's clear that the literature models weren't able to capture the variability of the attributes. By using a linear model with only one color based variable, the authors Teoh and Syaifudin (2007) and Abarra et al. (2018) may have achieved good results for their studied varieties, but for the 'Palmer' mango the model had poor results, such as correlation coefficients close to 0.

In Table 4, the results of the hypothesis test for the problem of estimating the quality attributes mass, TSS, firmness and titratable acidity are displayed. For mass, the obtained metrics by means of the proposed approach were better than the ones in literature, with mean values of RMSE and R equal to 16.483 and 0.992, respectively, while by means of literature the values 29.847 and 0.975 were achieved. For the other attributes the proposed approach was also superior, but with differences in metrics being more discrepant. For TSS, the means of RMSE were 0.907 vs 4.479 and for correlation coefficient 0.979 vs 0.045. For titratable acidity, the means of error were 0.256 and 0.589 and the means of the correlation coefficient were 0.908 vs 0.388. As all p-values were under 0.05 it is proved that the proposed approach was superior in the prediction of mass, TSS, firmness and titratable acidity in mango of 'Palmer' variety with a confidence level at 95%.

By means of the scatter plots of Figs. 8 to 11, it is visually proved the superiority of the proposed approach regarding literature, as there is a better adjustment to the line.

The plots for the TSS, firmness and titarable acidity literature models had poor adjustments as it was expected considering the obtained metrics by fold. The mass literature model provided better results for the 'Palmer' mango than the other three literature models, but it is still inferior to the proposed approach. With the combination of size, color and texture based variables and a non-linear estimator it was possible to correlate the input variables to the quality attributes with higher accuracy.

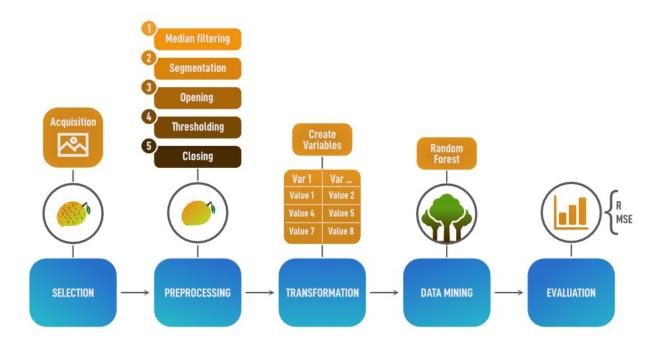


Figure 3: Applied stages for problem solution.

Table 3: Descriptive statistics of real values of mass, solids, total soluble solids, titratable acidity and firmness.

-			,				,	
	Attributes	Samples	Mean	Min	Max	Amp	SD	Var
	Mass	750	436,51	26,7	757,35	730,65	136,22	18557,30
	TSS	750	9,83	3,8	19,7	15,9	4,49	20,24
	Firmness	750	74,048	2,95	181,4	178,45	60,80	3697,50
	Titratable acidity	750	0,69	0,03	9,97	9,93	0,64	0,42

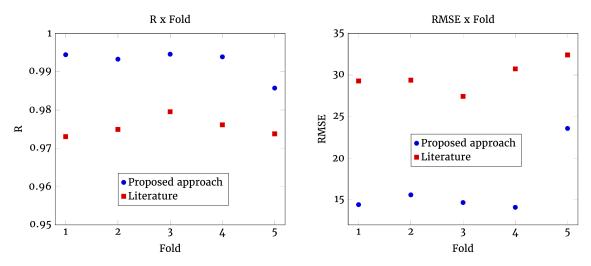


Figure 4: Obtained metrics in the prediction of mass in 'Palmer' mango (R to the left and RMSE to the right).

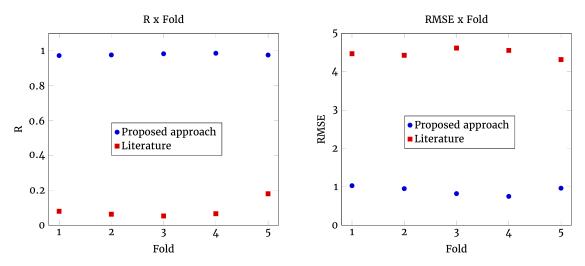


Figure 5: Obtained metrics in the prediction of TSS in 'Palmer' mango (R to the left and RMSE to the right).

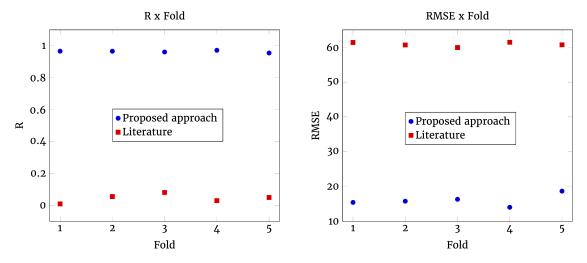


Figure 6: Obtained metrics in the prediction of firmness in 'Palmer' mango (R to the left and RMSE to the right).

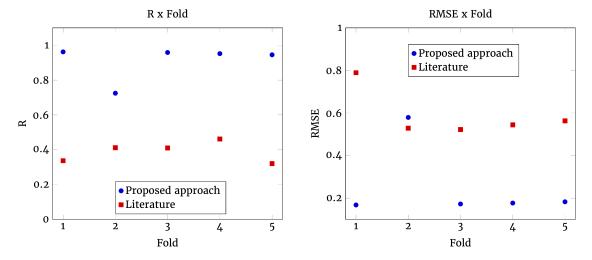


Figure 7: Obtained metrics in the prediction of titrable acidity in 'Palmer' mango (R to the left and RMSE to the right).

Table 4: Hypothesis tests summary.								
Attribute	Metric	μ 1	μ 2	μ 1 – μ 2	p-value			
Mass	RMSE	16.483	29.847	-13.364	0.000256			
171033	R	0.992	0.975	0.017	0.0001953			
TSS	RMSE	0.907	4.479	-3.572	1.449 * 10 ⁻⁶			
133	R	0.979	0.089	0.890	$1.693 * 10^{-6}$			
Firmness	RMSE	16.065	60.820	-44.754	$5.515 * 10^{-7}$			
riffilliess	R	0.964	0.045	0.919	$1.555 * 10^{-7}$			
Titrable acidity	RMSE	0.256	0.589	-0.334	0.01831			
Titrable actuity	R	0.908	0.388	0.520	0.0004158			

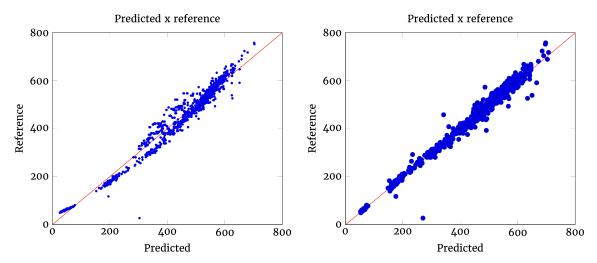


Figure 8: Scatter plots for mass in Palmer mangoes (Literature model to the left and Proposed model to the right).

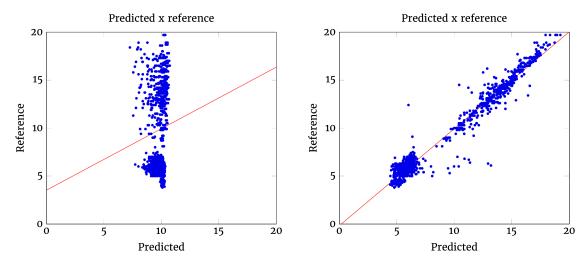


Figure 9: Scatter plots for TSS in Palmer mangoes (Literature model to the left and Proposed model to the right).

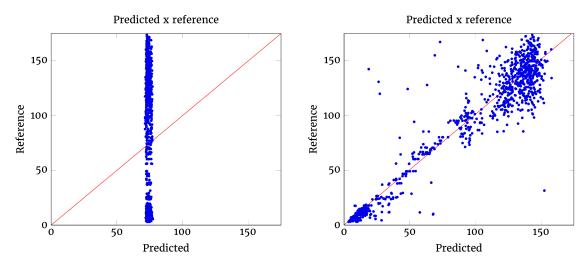


Figure 10: Scatter plots for firmness in Palmer mangoes (Literature model to the left and Proposed model to the right).

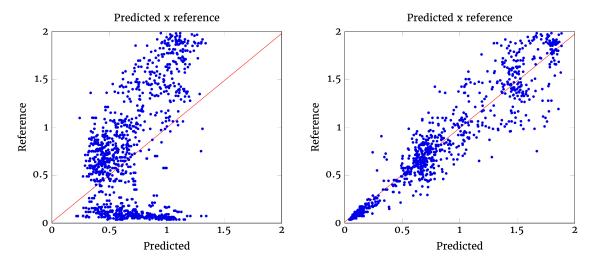


Figure 11: Scatter plots for titrable acidity in Palmer mangoes (Literature model to the left and Proposed model to the right).

5 Conclusion

The use of digital images for prediction of total soluble solids, acidity, mass and firmness associated to the proposed approach overcame the methods found in literature in a statistically significant way. Besides, the proposed approach provides a non-destructive evaluation of the quality attributes of the mango of Palmer variety. The four characteristics of this work justifying its superior performance regarding methods found in literature are: 1) sampling process allowing a great variability for the different stages of maturation, ensuring a greater robustness to the built models; 2) precise specification of pre-processing techniques of digital images; 3) enlargement of the space of input variables for the estimator, by using 40 variables identified as significant ones and 4) use of the Random Forest technique as estimator. The models used as reference use linear regression and assume a linear relationship between input variables and the quality attribute. However, the relationships between those variables may not be linear, and Random Forest could capture this type of relationship with a great number of input variables. As a future work the authors suggest that the proposed approach is evaluated in different varieties of mango.

References

- Abarra, M. S. J., Serrano, E. P., Sabularse, V. C., Mendoza, H. E. T. and del Rosario, E. J. (2018). Determination of fruit ripeness degree of 'carabao'mango (mangifera indica l.) using digital photometry, *Philippine Journal of Science* **147**(2): 249–253. Available at https://bityli.com/X642K.
- Adeodato, P. J. L., Vasconcelos, G. C., Arnaud, A. L., Cunha, R. C. L. V., Monteiro, D. S. M. P. and Neto, R. F. O. (2008). The power of sampling and stacking for the pakdd2007 cross-selling problem, *International Journal of Data Warehousing and Mining* 4: 22-31. https://doi.org/10.4018/jdwm.2008040104.
- Altendorf, S. (2019). Major tropical fruits market review 2017. Available at http://www.fao.org/fileadmin/templates/est/COMM_MARKETS_MONITORING/Tropical_Fruits/Documents/CA2895EN.pdf.
- AOAC (1997). Official methods of analysis of the Association of Official Analytical Chemists, Association of Official Analytical Chemists, Arlington.
- Breiman, L. (2001). Random forests, *Machine learning* **45**(1): 5-32. https://doi.org/10.1023/A: 1010933404324.
- Cardoso-Almeida, F. D. A., Figueiredo-Neto, A., Lucena-Cavalcante, Í. H. et al. (2017). Physical and mechanical parameters correlated to the ripening of mangoes (mangifera indica l.) cv.'tommy atkins', *Acta Agronómica* **66**(2): 186–192. http://dx.doi.org/10.15446/acag.v66n2.54757.

- Donis-González, I. R., Guyer, D. E., Leiva-Valenzuela, G. A. and Burns, J. (2013). Assessment of chestnut (castanea spp.) slice quality using color images, *Journal of Food Engineering* 115(3): 407-414. https://doi.org/10.1016/j.jfoodeng.2012.09.017.
- FAO (2018). Developments in bananas and major tropical fruits. Available at http://www.fao.org/fileadmin/templates/est/COMM_MARKETS_MONITORING/Tropical_Fruits/Documents/Developments_in_bananas_and_major_tropical_fruits.pdf.
- FAO (2019). Major tropical fruits statistical compendium 2017. Available at http://www.fao.org/fileadmin/templates/est/COMM_MARKETS_MONITORING/Tropical_Fruits/Documents/CA2909EN.pdf.
- Fayyad, U. and Stolorz, P. (1997). Data mining and kdd: Promise and challenges, Future generation computer systems 13(2-3): 99-115. https://doi.org/10.1016/S0167-739X(97)00015-0.
- Gao, X., Lu, W., Li, X. and Tao, D. (2008). Wavelet-based contourlet in quality evaluation of digital images, *Neurocomput.* **72**(1-3): 378-385. https://doi.org/10.1016/j.neucom.2007.12.031.
- Goulart, C., Loy, F. S., Galarça, S. P., Giovanaz, M. s., Malgarim, M. B. and Fachinello, J. C. (2013). Evolução do índice da e coloração da epiderme de mangas da cultivar tommy atkins, *Revista Iberoamericana de Tecnología Postcosecha* 14: 8-13. Available at https://www.redalyc.org/articulo.oa?id=81327871003.
- Hastie, T., Tibshirani, R. and Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition, Springer Series in Statistics, Springer.
- HFBRASIL (2018). Anuário 2017-2018. Available at https://www.hfbrasil.org.br/br/revista/acessar/completo/anuario-2017-2018.aspx.
- HFBRASIL (2019). Anuário 2018-2019. Available at https://www.hfbrasil.org.br/br/revista/acessar/completo/anuario-2018-2019.aspx.
- Jain, A. K., Duin, R. P. W. and Mao, J. (2000). Statistical pattern recognition: A review, IEEE Transactions on pattern analysis and machine intelligence 22(1): 4-37. https://dx.doi.org/10.1109/34.824819.
- James, G., Witten, D., Hastie, T. and Tibshirani, R. (2013). An introduction to statistical learning, Vol. 112, Springer.
- Khairunniza-Bejo, S. and Kamarudin, S. (2011). Chokanan mango sweetness determination using hsb color space, 2011 Third International Conference on Computational Intelligence, Modelling & Simulation, IEEE, pp. 216-221. https://doi.org/10.1109/CIMSim. 2011.45.
- Krogel, M.-A. (2005). On propositionalization for knowledge discovery in relational databases.

- Li, J., Huang, W. and Zhao, C. (2015). Machine vision technology for detecting the external defects of fruits—a review, *The Imaging Science Journal* 63(5): 241–251. https://doi.org/10.1179/1743131X14Y.0000000088.
- Montgomery, D. C. and Runger, G. C. (2010). Applied statistics and probability for engineers, John Wiley & Sons.
- Nagle, M., Intani, K., Romano, G., Mahayothee, B., Sardsud, V. and Müller, J. (2016). Determination of surface color of 'all yellow'mango cultivars using computer vision, *International Journal of Agricultural and Biological Engineering* **9(1)**: 42–50. https://doi.org/10.3965/j.ijabe.20160901.1861.
- Nandi, C. S., Tudu, B. and Koley, C. (2014). A machine vision-based maturity prediction system for sorting of harvested mangoes, *IEEE Transactions on Instrumentation and measurement* **63**(7): 1722–1730. https://doi.org/10.1109/TIM.2014.2299527.
- Neto, R., Adeodato, P. J. and Salgado, A. C. (2017). A framework for data transformation in credit behavioral scoring applications based on model driven development, Expert Systems with Applications 72: 293 305. https://doi.org/10.1016/j.eswa.2016.10.059.
- Ngai, E. W. T., Xiu, L. and Chau, D. C. K. (2009). Application of data mining techniques in customer relationship management: A literature review and classification, *Expert Syst. Appl.* **36**(2): 2592-2602. https://doi.org/10.1016/j.eswa.2008.02.021.
- Pandey, R., Gamit, N. and Naik, S. (2014). Non-destructive quality grading of mango (mangifera indica l) based on cielab colour model and size, 2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies, IEEE, pp. 1246–1251. https://doi.org/10.1109/ICACCCT.2014.7019298.
- Polat, K. and Gunes, S. (2007). Classification of epileptiform {EEG} using a hybrid system based on decision tree classifier and fast fourier transform, Applied Mathematics and Computation 187(2): 1017–1026. https://doi.org/10.1016/j.amc.2006.09.022.
- Pyle, D. (1999). Data preparation for data mining, morgan kaufmann.
- Salunkhe, R. P. and Patil, A. A. (2015). Image processing for mango ripening stage detection: Rgb and hsv method, 2015 Third International Conference on Image Information Processing (ICIIP), IEEE, pp. 362–365. https://doi.org/10.1109/ICIIP.2015.7414796.
- Teoh, C. and Syaifudin, A. M. (2007). Image processing and analysis techniques for estimating weight of chokanan mangoes, *Journal of Tropical Agriculture and Food Science* **35**(1): 183. Available at https://bityli.com/pP3Iu.

- Trindade, D. C. G. d. T., Lima, M. A. C. and Assis, J. S. (2015). Ação do 1-metilciclopropeno na conservação pós-colheita de manga 'palmer' em diferentes estádios de maturação, *Pesquisa Agropecuária Brasileira* **50**(9): 753-762. https://doi.org/10.1590/S0100-204X2015000900003.
- Vélez-Rivera, N., Blasco, J., Chanona-Pérez, J., Calderón-Domínguez, G., de Jesús Perea-Flores, M., Arzate-Vázquez, I., Cubero, S. and Farrera-Rebollo, R. (2014). Computer vision system applied to classification of "manila" mangoes during ripening process, Food and bioprocess technology 7(4): 1183-1194. https://doi.org/10.1007/s11947-013-1142-4.
- Witten, I. H. and Frank, E. (2005). Data mining: Practical machine learning tools and techniques, Morgan Kaufmann, San Francisco . Available at https://bityli.com/GWyxY.
- Yahaya, O. K. M., Jafri, M. Z. M., Aziz, A. A. and Omar, A. F. (2015). Determining sala mango qualities with the use of rgb images captured by a mobile phone camera, *AIP Conference Proceedings*, Vol. 1657, AIP Publishing, p. 060003. https://doi.org/10.1063/1.4915190.
- Yossy, E. H., Pranata, J., Wijaya, T., Hermawan, H. and Budiharto, W. (2017). Mango fruit sortation system using neural network and computer vision, *Procedia computer science* **116**: 596–603. https://doi.org/10.1016/j.procs.2017.10.013.
- Zheng, H. and Lu, H. (2012). A least-squares support vector machine (ls-svm) based on fractal analysis and cielab parameters for the detection of browning degree on mango (mangifera indica l.), Computers and Electronics in Agriculture 83: 47-51. https://doi.org/10.1016/j.compag.2012.01.012.