Microchemical Journal 196 (2024) 109604

Contents lists available at ScienceDirect

Microchemical Journal

journal homepage: www.elsevier.com/locate/microc

Digital image processing combined with machine learning: A new strategy for brown sugar classification

Vandressa Alves [a,*], Jeferson M. dos Santos [a], Edgar Pinto [d,e], Isabel M.P.L.V.O. Ferreira [d], Vanderlei Aparecido Lima [b], Maria L. Felsner [a,c]

[a] Department of Chemistry, State University of Midwestern at Paraná (UNICENTRO), Vila Carli, Zip Code 85040-080 Guarapuava City, Paraná, Brazil
[b] Department of Chemistry, Federal University of Technology – Paraná (UTFPR), Zip Code 85503-390 Pato Branco City, Paraná, Brazil
[c] Department of Chemistry, State University of Londrina (UEL), Zip Code 86057-970 Londrina City, Paraná, Brazil
[d] LAQV/REQUIMTE, Chemical Sciences Department, Faculty of Pharmacy, University of Porto, Zip Code 4050-313 Porto, Portugal
[e] Department of Environmental Health, School of Health of the Polytechnic Institute of Porto (ESS - P. Porto), CISA/Research Centre in Health, and Environment, Zip Code 4200-072 Porto, Portugal

ARTICLE INFO

Keywords:
Brown Sugar
Identity and Quality Standards
Sugar composition
Classification
Digital Image Processing
Color analysis

ABSTRACT

The coloring of foods is one of the main attributes of importance for consumers and it can be decisive for a consumer to accept or reject the product. Models that explore brown sugar coloring are scarce in scientific research. So, a new strategy for brown sugar classification through the combination of digital image processing, machine learning and physicochemical composition data was proposed. RGB channel intensities and color histogram data, obtained from digital image processing, in combination with some physicochemical characteristics (sucrose, Ca, Fe, ICUMSA color and total phenolic compounds (TPC)) were used as training and external validation datasets in the creation of classification models by RF algorithm. Excellent performance of classification models was observed by high overall accuracy rates for ICUMSA color (92.6 %), Ca and sucrose (100 %), Fe (94.9 %), and TPC (97.6 %). Thus, classifying brown sugar based on its color can be a valuable strategy for the beverage and food industries, allowing for greater diversification and meeting consumer needs while enhancing the quality and consistency of products.

1. Introduction

Refined white sugar is the most widely consumed sugar globally. However, most of the sugarcane's constituents in white sugar are partially or entirely lost during the purification stages, resulting in a product with high purity but low in bioactive compounds and minerals [1]. As a result, less processed sugars such as brown sugar are gaining consumers preference, as they are perceived as functional foods with high nutritional value and composition rich in essential nutrients and minerals, vitamins, amino acids, organic acids and phenolic compounds. Brown sugar offers several health benefits contributing to the immune system with cytoprotective and detoxifying effects, as well as aiding in the prevention of hypertension and diabetes [2 –6]. It is commonly consumed as a sweetener or incorporated as an ingredient in the beverage and confectionery industries, accounting for nearly 60.0 % of the sugar consumed by these sectors [2,7].

Brown sugar's composition typically includes around 85.0 to 90.0 % sucrose, 2.0 to 5.0 % reducing sugars, 0.5 to 4.0 % moisture, and 1.0 to 3.0 % ash. Besides this, the values of these physicochemical characteristics may vary based on the concentration of sugarcane juice [8]. Regarding the bioactive compounds present in sugarcane, they can have a direct influence on the color formation in sugar [9] and its nutraceutical properties [3]. So, having well-defined identity and quality standards for this sweetener is essential. Manufacturers must ensure that the product meets consumer preferences and industrial specifications, while also guaranteeing its integrity and food safety. This helps to prevent potential contaminations and adulterations that could pose health risks [5,7]. To achieve these objectives, the evaluation of physicochemical characteristics of sugar such as purity, color, contents of moisture, insoluble solids and minerals and phenolic compound levels, among others, need to be conducted [3].

Despite Brazil being one of the world's largest producers of brown

* Corresponding author.
E-mail addresses: alvesvandressa@gmail.com (V. Alves), isabel.ferreira@ff.up.pt (I.M.P.L.V.O. Ferreira), valima@utfpr.edu.br (V.A. Lima), felsner@unicentro.br (M.L. Felsner).

https://doi.org/10.1016/j.microc.2023.109604
Received 1 October 2023; Received in revised form 30 October 2023; Accepted 1 November 2023
Available online 3 November 2023

sugar, accounting for 4.0 % of global production of 11.35 million tons in 2020 [10], there is no specific regulation regarding the production and characterization of this product in the country. As a result, manufacturers establish their own identity and quality standards. The situation becomes even more critical when it is noted that Brazilian brown sugar, as described in CL 2015/19-CS by the Codex Alimentarius Commission [11], does not align with the specifications of international standard-setting bodies like the USDA (The U.S. Department of Agriculture) [12] and the EAC (East African Community) [13]. These international standards describe Brazilian brown sugar with characteristics similar to those of non-centrifugal sugar, another byproduct obtained from sugarcane. As a result, Brazilian brown sugar exhibits a different chemical composition compared to other non-centrifuged sugars, making the implementation in each country's international regulations for this product challenging.

The lack of regulation of identity and quality standards for brown sugar by Brazilian regulatory and control agencies contributes to the availability of non-standardized products, as brown sugar is primarily an artisanal product and is usually produced by family agroindustry [14]. Verruma-Bernardi et al. [15], Durán Rojas et al. [16] and Cifuentes et al. [17] emphasize that the absence of standardization in brown sugar production can lead to undesirable characteristics such as color and sensory variability and varying levels of bioactive molecules in the product, due to a high percentage of moisture and reducing sugars. The resulting organoleptic changes arising from these physicochemical alterations hinder the production of consistent and high-quality products, leading to greater consumer rejection and consequently impairing its marketability within the food industry.

Due to the significant variability of brown sugar composition, a classification strategy based on differences in color and physicochemical attributes of this sweetener, including the concentration of bioactive compounds, is crucial for providing a more consistent product to consumers and expanding its industrial applicability. This strategy would enable the offering of brown sugar with varying quality levels, meeting the requirements of each sector that utilizes it as an ingredient. As a result, ensuring the implementation of a harmonized and well-defined production process could lead to diverse uses for brown sugar, such as an ingredient for formulating products in the food, beverage, and cosmetic industries [18]. Alternative methodologies based on the principles of Green Analytical Chemistry, such as Digital Image Processing (DIP) combined with Machine Learning (ML) algorithms, have been successfully employed for classification purposes in these sectors [19–22].

Recently, these data analysis tools have gained widespread use in the literature due to their innovative approaches compared to conventional methods in microbiological, chemical, and nutritional analyses. They also play a role in detecting foods fraud and adulteration, while offering new alternatives for industrial procedures within the food industry [23]. Additionally, they can assist regulatory agencies in establishing identity and quality standards to maintain quality control of foods such as brown sugar on an industrial scale.

Digital image processing (DIP) is employed to extract the most significant features from images, which are then used as input data to machine learning (ML) algorithms such as the Random Forest algorithm, which classify them based on various criteria [23]. Thus, Random Forest algorithm (RF) has gained prominence for modeling or regressing large datasets for classification purposes [24,25]. It is a set of tree predictors where each tree depends on a random vector that is independent and with a uniform distribution for all trees in the forest [26]. There are two methods for classification trees: Boosting and Bagging. Among them, Bagging stands out, where trees are created independently from the dataset using bootstrap sampling and data prediction is done by a multiclass system, resulting in a single process [27]. In addition to using different bootstrap samples for each tree, Random Forests also change the way that the trees are created. According to [26], in classification and regression trees (CART), each node is partitioned based on the best partition among all variables. In a Random Forest, each node is split based on the best split from a random subset of predictors at that node. This leads to very good performance compared to other classifiers such as linear discriminant analysis (LDA), artificial neural networks (ANN) and support vector machines (SVM) [20,21]. It is worth noting that, despite their numerous advantages, the application of digital image processing combined with machine learning algorithms is still relatively limited in classifying different types of sugars [21,22,28].

To date, there are no reports in the literature on classification models for brown sugar that correlate physicochemical characteristics and bioactive compounds with color variability in this sweetener. Therefore, additional studies are needed to explore the potential of these tools for assessing the quality of brown sugars. Analytical methods typically used for characterization of brown sugar composition often involve time-consuming techniques with high reagent consumption, generation of environmentally harmful wastes and high costs [1,29,30]. Thus, this work proposes classification models for Brazilian brown sugar, considering the diversity of its chemical composition, color, and lack of national standardization, aiming to assist in establishing national identity and quality standards for this type of sugar. To achieve this, a three-level class discrimination and modeling method was applied based on the Random Forest (RF) machine learning algorithm, using RGB channel intensities and histogram color data obtained from digital images of brown sugar samples, as well as chemical composition data.

2. Material and methods

2.1. Sampling

Thirty-four commercial brown sugar samples from different regions of Brazil were acquired between 2017 and 2018. The samples were stored at room temperature in their respective packaging until analysis. They were coded according to the abbreviations of the states of origin, followed by Arabic numbering. One brown sugar sample belonged to the Northern region (Acre State (AC)), four samples were from the Central-West region (Distrito Federal State (DF), Goiás State (GO) and Mato Grosso State (MT)), nine samples from the Southeast region, (Minas Gerais State (MG), São Paulo State (SP) and Rio de Janeiro State (RJ)), nineteen samples from the Southern region (Paraná State (PR), Santa Catarina State (SC) and Rio Grande do Sul State (RS)) and one sample from the Northeast region (Pernambuco State (PE)).

2.2. Physicochemical analysis

2.2.1. ICUMSA color

The color of brown sugar samples was determined and expressed in terms of the International Color Unit for Sugared Products following the standard protocol of ICUMSA (International Commission for Uniform Methods of Sugar Analysis) with some modifications [31]. A mass of 0.5 g of the sample was dissolved in 0.1 mol/L triethanolamine buffer. The mixture was vacuum-filtered using 0.45 μm hydrophilic and glass fiber membranes, followed by sonication in an ultrasonic bath for two minutes and its absorbance was measured at 420 nm using a SP2000 UV–VIS spectrophotometer from Spectrum. The color was calculated using the ICUMSA color unit equation or $IU = absorbance \times 1000/(b \times C)$, where b is the cuvette thickness (cm), and C is the sugar solution concentration (g mL^{-1}). All assays were conducted in duplicate.

2.2.2. Sugar analysis by HPLC-IR

Sucrose contents in the brown sugar samples were determined using a high-performance liquid chromatography (HPLC) method adapted from Santos et al. [32] with some modifications. A mass of 100 mg from was dissolved in 10 mL of acetonitrile (75.0 %, v/v), and then the resulting solution was centrifuged for 10 min and filtered through a 0.22 μm PTFE membrane filter. For sucrose separation from other sugars, an ultrapure Tracer Excel 120 APS silica column (5 μm, 250 mm

$\times$ 4.6 mm) was used. The chromatographic system was equipped with a quaternary pump and consisted of a refractive index detector (132 Gilson). The pump operated in an isocratic mode with a mobile phase of acetonitrile: water 75.0:25.0 (v/v), and the flow rate was set at 1.0 mL min^{-1}. Sucrose concentration was obtained by plotting the peak area against the concentrations for their respective solutions standards and it was expressed in g 100 g^{-1}. All assays were conducted in triplicate.

2.2.3. Ca and Fe analysis by FAAS

The contents of Fe and Ca were determined using Flame Atomic Absorption Spectrometry (FAAS) according to the methodology described by dos Santos et al. [33]. All assays were conducted in triplicate.

2.3. Digital image processing (DIP) combined with Machine learning (ML)

2.3.1. Acquisition and digital processing of images

Brown sugar samples of, approximately 100 g each, were placed in a transparent plastic container. Three photos were acquired for each sample, totaling 102 digital images. An 8-megapixel (MP) smartphone with a resolution of 1,280 $\times$ 720 pixels was used to capture the digital images from a distance of approximately 0.3 m from the sample area, under ambient light. The digital images were then saved in PNG format (with a spatial resolution of 4,608 x 3,456 pixels). The region of interest (ROI) was digitally cropped to a size of 200 x 200 pixels using Gimp® v. 2.10.32 free software, without any additional image manipulation. The ROIs were loaded into Chemostat® v.2. free software to separate and quantify the color intensities of the RGB and L channels and of the HSV color space for generating the color index histogram. The color datasets were subsequently employed to develop classification models (being target classes: light, pattern, and dark colorations of brown sugar samples) considering sucrose, ICUMSA color, total phenolic compounds (TPC), Ca and Fe as variables. To minimize the influence of multicollinearity in the data a Variance Inflation Factor (VIF) analysis was performed using the R programming language v. 4.2.1. Variables related to the RGB and L channel intensities and of the HSV color space that exhibited VIF values greater than 10.0 were sequentially excluded from the datasets before modeling. Schematic representation of the approach adopted for the acquisition and digital processing of images of brown sugar samples and data modeling using the Forest Random algorithm is illustrated in Fig. 1.

2.3.2. Random Forest algorithm (RF)

The classification models for the brown sugar samples based on their coloration (light, pattern, and dark) and chemical composition (sucrose, Fe, Ca, ICUMSA color, and TPC) were implemented using the R programming language, v. 4.2.1, and the Random Forest (RF) algorithm, executed through the WEKA code packages, v. 3.8.6. To assess the models generated by the RF algorithm, two datasets were created: one for building the classification model (60.0 % of the data) and another for external validation of the model (40.0 % of the data) for the variables previously defined by the VIF analysis (HSV color space). The presence of outliers in the dataset was assessed prior to modeling by applying a *Grubbs test* for each response (physicochemical characteristics) at a 95.0 % confidence level using Minitab software v. 16.2.1. Brown sugar samples with outlier values were removed from the dataset. To select the best attributes from the dataset, a genetic algorithm was adopted, which initially choose 7 out of 81 attributes for analysis in classifying the brown sugar samples into three coloration levels (light, pattern and dark).

2.3.3. Performance of classification models

Performance of the classification models was evaluated using the confusion matrix data, which quantifies classes based on the scores of TP (true positives), FP (false positives), TN (true negatives), and FN (false negatives). Additionally, the performance metrics (accuracy, precision, sensitivity, specificity, and *F1-score*; Equations 1—5) were calculated,

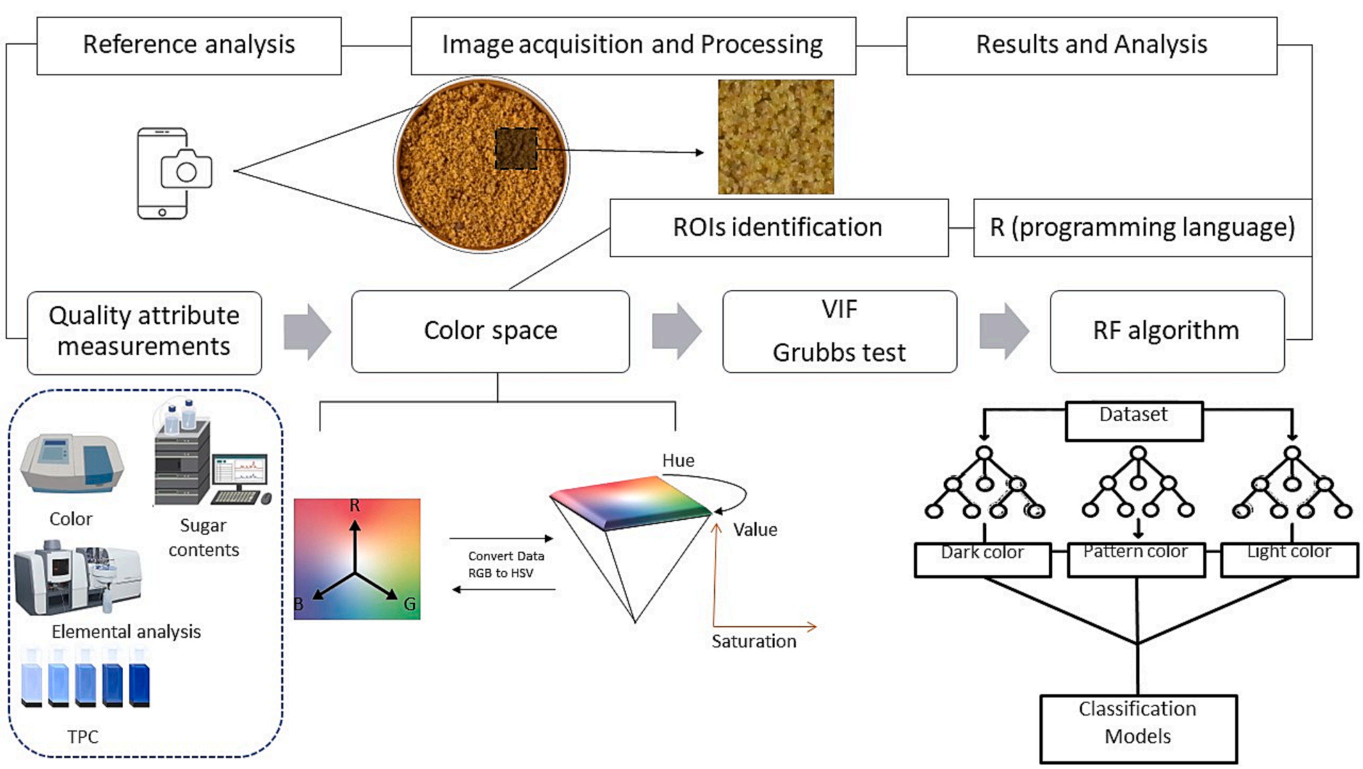

Fig. 1. Schematic representation of the approach adopted for the acquisition and digital processing of images of brown sugar samples and data modeling using the Forest Random algorithm.

Fig. 2. On the right, ROIs (Region of Interest) of Brazilian brown sugar samples with their respective encoding. On the left, an illustration of the ROI with dimensions of 200 x 200 pixels obtained by cropping the original image.

as suggested by [34].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \times 100 \tag{2}$$

$$Sensitivity = \frac{TP}{TP + FN} \times 100 \tag{3}$$

$$F1 - score = 2 \times \frac{Precision \times Sensitivity}{Precision + Sensitivity} \times 100 \tag{4}$$

$$Especificity = \frac{TN}{TN + FP} \times 100 \tag{5}$$

3. Results and discussion

3.1. Physicochemical and elemental characterization of brown sugar samples

Brown sugar displays a broad spectrum of colors and textures, ranging in color from yellow to amber – yellow to a darker brown, as illustrated in Fig. 2. Considering the various physicochemical characteristics that could be chosen for the generation of classification models, those typically adopted in national and international regulations for sugar quality control, such as sucrose and ICUMSA color and others related to production and processing stages such as Ca, Fe, and TPC, were selected, as they are correlated with the final color of this natural sweetener. Therefore, three classes were established for the coloration of brown sugar samples: (1) light coloration, (2) pattern coloration and (3) dark coloration. The average values, standard deviations, and ranges of physicochemical characteristics, as well as the established limits for these physicochemical characteristics for each color class of brown sugar, can be seen in Table 1.

Table 1
Average values, standard deviations, and range of variation for physicochemical characteristics and color classes of brown sugar samples. n = number of replicates.

Physicochemical characteristics	*Levels*	*Color Classes*
Sucrose (g 100 g^{-1}; n = 3)	*Average ± SD:* 85.4 ± 6.5 *Range:* 60.1 – 94.3	*light:* > 88 *pattern:* 85 – 88 *dark:* < 85
ICUMSA color (I.U.; n = 2)	*Average ± SD:* 38,141 ± 20,022 *Range:* 8,031 – 96,488	*light:* < 25,000 *pattern:* 25,000 – 46,000 *dark:* > 46,000
Total phenolic compounds (TPC) (mg GAE 100 g^{-1}; n = 3)	*Average ± SD:* 220.0 ± 117.5 *Range:* 2.60 – 434.7	*light:* < 205 *pattern:* 205 – 280 *dark:* > 280
Fe (mg 100 g^{-1}; n = 3)	*Average ± SD:* 4.7 ± 2.7 *Range:* 1.6 – 12.6	*light:* < 3.0 *pattern:* 3.0 – 6.0 *dark:* > 6.0
Ca (mg 100 g^{-1}; n = 3)	*Average ± SD:* 104 ± 74.0 *Range:*11.0 – 300	*light:* < 70 *pattern:* 70 – 110 *dark:* > 110

According to the [11], the variability in the color of brown sugar can result from the variety of sugarcane, agroecological conditions during harvesting, and/or the processing applied during its production, among other factors. In light of these facts, the categorization and standardization of the color of brown sugar is crucial, as color directly influences its sensory characteristics perceived by consumers, as well as its industrial application as an ingredient [35].

Assessing the data presented in Table 1, several pertinent considerations can be made regarding the association between the physicochemical characteristics (Table 1) and the observed coloration in brown sugar samples (Fig. 2). This served as the foundation for establishing the limits of the physicochemical characteristics for three color classes and subsequently generating the classification models.

The sucrose contents in the brown sugar samples ranged from 60.1 to 94.3 g per 100 g. It is worth noting that for most samples, the sucrose levels are within the ranges reported by other authors [36–38], especially for those classified as of pattern and dark colors. The main difference lies in the upper limit of sucrose contents, which are higher for Brazilian brown sugar than what has been observed in the literature.

Regarding the regulatory bodies for international standardization of brown sugar, the sucrose contents should be at a minimum of 86.0 % according to the USDA [12]. However, the Codex Alimentarius Commission [16] specifies a maximum sucrose content of 90.0 % (w/w), while the EAC [13] requires a minimum apparent sucrose content, determined by polarization, of 99.0 %. When comparing these standards to our data, it's clear that many of them, especially those belonging the dark coloration class, do not meet these international standards. This highlights the need for the establishment of national standards in Brazil for both identity and quality of brown sugar, taking into consideration not only its physicochemical composition but also its coloration.

In terms of color, the brown sugar samples had ICUMSA color values ranging from 8,031 to 96,488 units. It was observed that the Brazilian brown sugar samples exhibited higher ICUMSA color values compared to those reported in the literature [38,39] with a predominance of samples classified as having pattern and dark colors. This wide variability in ICUMSA color values can be attributed to various processes that are precursors to color formation in brown sugar [8,40].

The total phenolic compounds (TPC) levels in the brown sugar samples ranged from 2.6 to 433.7 mg (GAE) per 100 g. When comparing these TPC values to those reported in the literature, it is noted that they were similar to the values determined for the three coloration classes of brown sugar established in our work [9,41].

The concentrations of the metallic ions Ca and Fe in the Brazilian brown sugar samples showed a wide range of variation, ranging from 1.6 to 12.6 mg per 100 g for Fe and from 11 to 300 mg per 100 g for Ca. Analyzing the concentration ranges of metallic ions reported by other authors, it is noted that they encompass the three coloration classes of brown sugar established in our study [36 –38]. The wide variability in Ca levels can be explained by the fact that the processing of brown sugar is often done in an artisanal setting and includes a clarification step in that calcium hydroxide is added, with the purpose of neutralizing organic acids and inorganic ions such as phosphates and sulfates [9]. Although sugarcane contains these metallic ions in its composition, the indiscriminate use of calcium hydroxide for pH correction can lead to contamination, increasing Ca concentrations in the brown sugar. For Fe, similar concentration ranges were observed between our data and those reported in the literature. However, high levels of Fe in brown sugar samples may indicate a lack of maintenance and cleanliness in the production machinery, which can result in the incorporation of this element to the final product [14].

To better understand how the physicochemical and elemental composition can influence the coloration of brown sugar, linear correlation analyses were conducted. Significant linear correlations were observed between ICUMSA color and Ca (r = 0.462; p = 0.007) and between ICUMSA color and Fe (r = 0.541; p = 0.001). The observed correlations suggest that these metallic ions play an important role in various chemical reactions that affect the coloration of brown sugar, influencing its appearance and hue [9,42].

A significant linear correlation between total phenolic compounds (TPC) and Fe (r = 0.494; p = 0.003) was also observed. This correlation can be attributed to the enzymatic oxidation of phenolic compounds catalyzed by the presence of Fe or to complexation reactions with melanoidins, often in acidic pH conditions. These reactions result in the formation of highly colored polymers responsible for the darkening of sugar [43,44]. It is important to consider that the colored pigments responsible for the coloration of sugar come not only from plant bioactive compounds but also from other substances produced during the sugar production process [3,39,42]. This is evident from the significant correlations between sucrose contents and ICUMSA color (r = -0.345, p = 0.050) and between sucrose contents and TPC (r = -0.391, p = 0.024).

Therefore, the variability in the physicochemical characteristics of brown sugar poses a challenge for the sugar industry, as coloration can result from different complex reactions between sugarcane constituents or throughout the sugar production and processing stages. Developing classification models for brown sugar based on its coloration can be an interesting strategy to expand its application in various industrial sectors and also provide consumers with a higher quality and food-safe product.

3.2. Evaluation of classification models for brown sugar

Classification models were developed based on some physicochemical characteristics responsible for the coloration in brown sugar samples (sucrose, ICUMSA color, TPC, Fe and Ca), employing the Random Forest (RF) algorithm. The target classes were previously defined as three levels of coloration: light, pattern and dark (Table 1), which corresponded to 60.0 % of the training dataset. Subsequently, a subset of test samples was assigned or not to the target classes, constituting 40.0 % of the data.

To select within the color dataset, the most important variables, a Variance Inflation Factor (VIF) analysis was applied to all the data. VIF values $<$ 10.0 help select which variables from color dataset are significant and do not exhibit multicollinearity [45]. Based on these explanations the HSV color space data were selected for generating the brown sugar classification models.

Histogram graphs in the HSV (Hue, Saturation and Value) color space, showed in Fig. 3, illustrating the importance of evaluating the importance of variables in relation to the target classes of brown sugar. It is evident that the ICUMSA color, sucrose contents and total phenolic compounds (TPC) levels, as well as the levels of Ca and Fe, are factors strongly associated with the coloration of brown sugar samples, as previously discussed in section 3.1.

It is possible to observe in Fig. 3 that, following the physicochemical characteristics, the attribute H (Hue), which corresponds to the hue, showed the highest importance in the color pattern of the brown sugar samples when compared to the attributes S (saturation) and V (brightness). Hue typically defines, at a specific wavelength, the spectral radiance of a color; saturation characterizes the purity of the color and brightness represents the intensity of light present in the described color [46]. Decision tree (DT) algorithms, which include Random Forest (RF), have a greater advantage regarding the distribution of samples in the HSV color space [47].

The confusion matrix containing the classification data of brown sugar samples based on the test data for the target classes with respect to the ICUMSA color, sucrose, Fe, Ca, and TPC levels and the performance metrics of the classification models are displayed in Table 2.

The metrics of the generated models indicated that the classification success rate specific to the three levels of target classes (light, pattern and dark colors) resulted in precision values ranging from 83.33 % to 100 % for the ICUMSA color, 100 % for both Ca and sucrose, 88.9 % to 100 % for Fe and of 91.7 % to 100 % for TPC. Furthermore, the models exhibited excellent overall accuracy for classification with values of

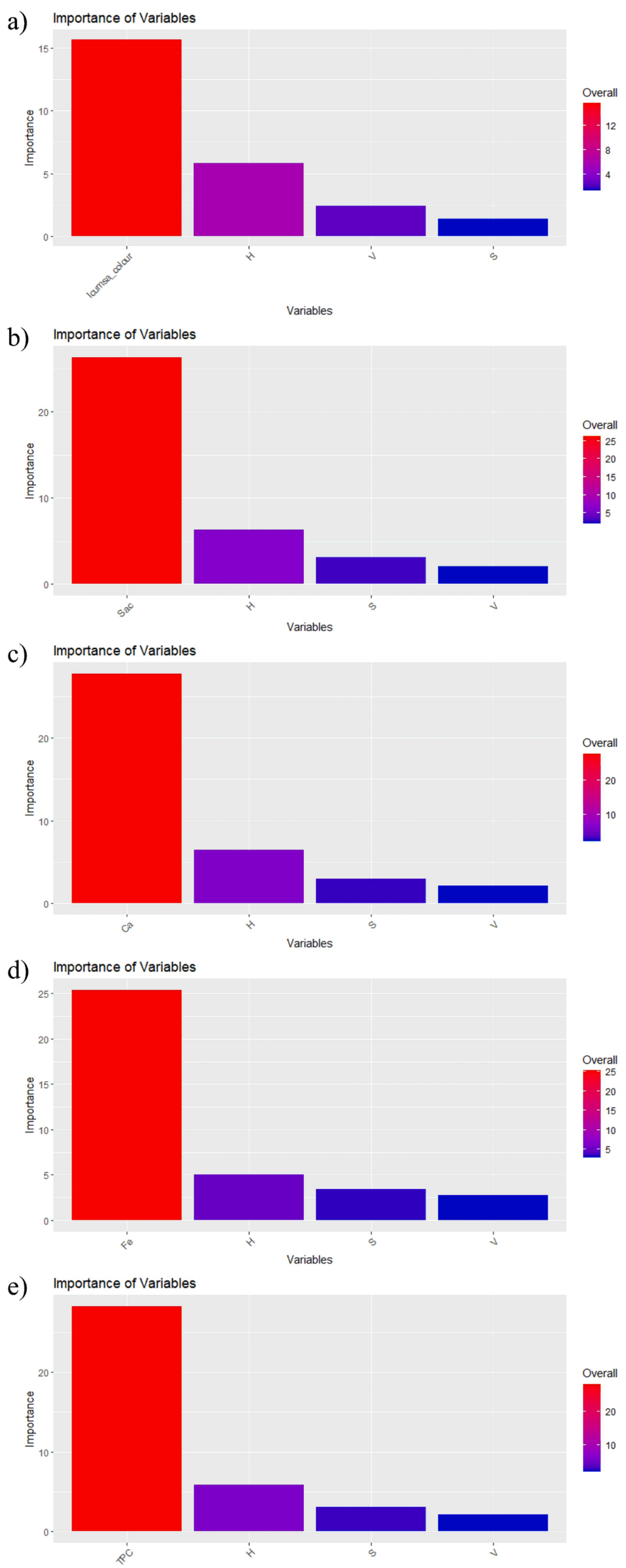

Fig. 3. Histograms of classes importance relative to variables in the HSV color space. The graphs correspond to (a) ICUMSA color; (b) sucrose; (c) Ca; (d) Fe; and (e) TPC, respectively.

Table 2
Confusion matrix and performance metrics generated by the RF algorithm of brown sugar staining classification models for each physicochemical feature evaluated. Data from the modeling phase related to external validation.

Physicochemical characteristics	Confusion matrix		Classes with reference values			Performance Metrics				
			light color	*pattern color*	*dark color*	*Sensitivity*	*Specificity*	*F1-escore*	*Precision*	*Accuracy*
ICUMSA Color	*Classes with prediction values*	*light color*	9	0	0	100	93.7	100	100	96.8
		pattern color	0	11	1	83.3	95.2	83.3	83.3	89.3
		dark color	1	0	5	90.0	100	90.0	90.0	95.0
Sucrose		*light color*	11	0	0	100	100	100	100	100
		pattern color	0	11	0	100	100	100	100	100
		dark color	0	0	18	100	100	100	100	100
Fe		*light color*	8	0	0	91.7	100	91.7	91.7	95.8
		pattern color	0	11	0	100	90.4	100	100	95.2
		dark color	1	1	18	88.8	100	88.8	88.8	94.4
Ca		*light color*	10	0	0	100	100	100	100	100
		pattern color	0	13	0	100	100	100	100	100
		dark color	0	0	5	100	100	100	100	100
Total phenolic compounds (TPC)		*light color*	11	0	0	91.7	100	91.6	91.6	95.8
		pattern color	0	10	0	100	95.4	100	100	97.7
		dark color	1	0	19	100	100	100	100	100

92.6 % for ICUMSA color, 100 % for Ca and sucrose, 94.9 % for Fe, and 97.6 % for TPC. Among the evaluated physicochemical characteristics, it was observed that Ca and sucrose yielded the best results in terms of classification performance for each target class in the *F1*-score metric (Table 2). On the other hand, lower values for performance metrics were observed for ICUMSA color and Fe, as the presence of moisture in the sample, as well as variable pH and acidity levels, can systematically interfere with the concentrations of these physicochemical characteristics, making their classification less precise [9].

In the literature, the use of the Random Forest (RF) algorithm is reported to exhibit excellent performance for classification purposes based on the accuracy, sensitivity and specificity of the method applied. The success in solving classification problems can be attributed to the characteristics exhibited by this type of machine learning algorithm. The RF algorithm is suitable for handling unstable models or class imbalance problems, as it reduces variance and minimizes the risk of overfitting. Furthermore, it offers an advantage compared to other classification methods such as the ability to directly discriminate among a set of samples/objects in a multi-class system in a single process [48]. Additionally, as it operates with only two configuration parameters (the number of variables in the random subset at each node and the number of trees in the forest), it allows for accurate classification of large datasets [27]. This machine learning algorithm had showed good performance in the literature for classification problems such as of ripening stage of papaya samples [24], for authentication of geographical origin of *Panax notoginseng* [49], for authentication of evening primrose oil using FTIR-HATR and ground nutmeg samples [50]. These studies, including the one conducted in this work, highlight the importance of the Random Forest algorithm as a classification method in data preprocessing and feature extraction from food samples. It is also worth noting that the results of this study demonstrate the contribution of the generated models to the standardization and classification of brown sugars, as there were no literature findings using the Random Forest (RF) algorithm as a classification method across a wide range of sugar samples.

4. Conclusions

Changes in the organoleptic properties and coloration of brown sugar due to variations in its chemical composition pose significant challenges in the production of consistent and high-quality products. Therefore, establishing identity and quality standards for this sweetener is crucial, as manufacturers need to ensure that the final product meets consumer preferences and industry specifications.

The creation of classification models that combine digital image processing with class modeling through the Random Forest (RF) algorithm has proven to be effective in assessing the quality parameters of brown sugar, allowing for the prediction of classification into different target categories with excellent performance metrics. It is worth noting that parameters such as sucrose and Ca achieved a global accuracy of 100 %, while the TPC reached an accuracy of 97.6 %. Thus, the application of digital image processing combined with machine learning offers several advantages over conventional methodologies, including speed, increased safety for analysts, reduced cost and a smaller environmental impact, as it does not require the use of toxic reagents or solvents.

Classification strategies like the one proposed in this study are innovative and can be highly valuable to the sugar industry in standardizing the color and composition of brown sugar destined for industrial use as well as ensuring the integrity and food safety of this sweetener for consumers. In addition, they can allow the market segmentation in that manufacturers can be offer different classes of this sugar type to several market segments thus expanding its commercialization and exploring new applications in food industries. Besides this, brown sugar with different shades can more accurately meet consumer's preferences. These characteristics helps prevent potential frauds and adulterations, contributing to the production of high-quality and consistent sugars.

CRediT authorship contribution statement

Vandressa Alves: Investigation, Methodology, Writing – original draft, Writing – review & editing. **Jeferson M. dos Santos:** Formal analysis, Investigation, Methodology. **Edgar Pinto:** Investigation, Methodology. **Isabel M.P.L.V.O. Ferreira:** Funding acquisition, Resources, Supervision. **Vanderlei Aparecido Lima:** Conceptualization, Investigation, Methodology, Writing – review & editing. **Maria L. Felsner:** Conceptualization, Funding acquisition, Resources, Supervision, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

The authors thank the Brazilian Council for Scientific and Technological Development (CNPq) and the Brazilian Coordination of Superior Level Staff Improvement (CAPES) for their financial support. The present work was carried out with support from the Coordination of Superior Level Staff Improvement – Brazil (CAPES). Alves, V. thanks CAPES for the Doctoral scholarship.

References

[1] M.R.F. Sampaio, M.C. Machado, M.T. Lisboa, M.A. Vieira, T.B.R. Zimmer, D. M. Otero, R.C. Zambiazi, Physicochemical Characterization and Antioxidant Activity of Refined and Unrefined Sugarcane Products from Southern Brazil, Sugar Tech (2022), https://doi.org/10.1007/s12355-022-01146-y.

[2] L. Cervera-Chiner, C. Barrera, N. Betoret, L. Seguí, Impact of sugar replacement by non-centrifugal sugar on physicochemical, antioxidant and sensory properties of strawberry and kiwifruit functional jams, Heliyon 7 (2021), https://doi.org/10.1016/j.heliyon.2021.e05963.

[3] W.R. Jaffé, Nutritional and functional components of non centrifugal cane sugar: A compilation of the data from the analytical literature, J. Food Compos. Anal. 43 (2015), https://doi.org/10.1016/j.jfca.2015.06.007.

[4] J.S. Lee, S. Ramalingam, I.G. Jo, Y.S. Kwon, A. Bahuguna, Y.S. Oh, O.J. Kwon, M. Kim, Comparative study of the physicochemical, nutritional, and antioxidant properties of some commercial refined and non-centrifugal sugars, Food Res. Int. 109 (2018), https://doi.org/10.1016/j.foodres.2018.04.047.

[5] F. Velásquez, J. Espitia, O. Mendieta, S. Escobar, J. Rodríguez, Non-centrifugal cane sugar processing: A review on recent advances and the influence of process variables on qualities attributes of final products, J. Food Eng. 255 (2019), https://doi.org/10.1016/j.jfoodeng.2019.03.009.

[6] Z. Zhu, C. Xie, W. Li, F. Hang, K. Li, C. Shi, W.O.S. Doherty, Nutritional and antioxidant properties of non-centrifugal cane sugar derived from membrane clarified juice, LWT 131 (2020), https://doi.org/10.1016/j.lwt.2020.109717.

[7] N. Mohan, P. Singh, Sugar and Sugar Derivatives: Changing Consumer Preferences. Singapore, Springer Singapore (2020), doi: 10.1007/978-981-15-6663-9.

[8] G. Eggleston, G. Aita, A. Triplett, Circular Sustainability of Sugarcane: Natural, Nutritious, and Functional Unrefined Sweeteners That Meet New Consumer Demands, Sugar Tech. 23 (2021), https://doi.org/10.1007/s12355-021-00994-4.

[9] L.S.M. Bento, Colorants through cane sugar production and refining, Sugar Ind. 134 (2009).

[10] FAO. Food and Agriculture Organization of the United Nations, Definition and Classification of Commodities, in Sugar Crops and Sweeteners and Derived Products, 2016. https://www.fao.org/home/en/.

[11] Codex Alimentarius Commission, Codex Committee on Sugars (CCS), 2019.

[12] USDA. The U.S. Department of Agriculture, Commercial item description sugar, white, refined, and sugar, brown, 2009.

[13] EAC. East African Community, Brown sugars - Specification, 2010.

[14] M.L. de Mello, N.Z. Barros, M.A. Sperança, F.M.V. Pereira, Impurities in Raw Sugarcane Before and After Biorefinery Processing, Food Anal. Methods 15 (2022), https://doi.org/10.1007/s12161-021-02105-1.

[15] S.R. Verruma-Bernardi, M.R. Borges, M.T.M.R. Lopes, C.H. Della-Modesta, R. C. Ceccato-Antonini, Avaliação Microbiológica, Físico-Química e Sensorial de Açúcares Mascavos Comercializados na Cidade de São Carlos, SP, Brazilian J. Food Technol. 10 (2007).
[16] O. Durán Rojas, E. Pérez, R. Cardoso, W. Pérez, A Colorimetria e aceitação de açúcar mascavo, Temas Agrários 17 (2012).
[17] J. Cifuentes, V.A. Salazar, M. Cuellar, M.C. Castellanos, J. Rodríguez, J.C. Cruz, C. Muñoz-Camargo, Antioxidant and Neuroprotective Properties of Non-Centrifugal Cane Sugar and Other Sugarcane Derivatives in an In Vitro Induced Parkinson's Model, Antioxidants 10 (2021), https://doi.org/10.3390/antiox10071040.
[18] D.H. Flórez-Martínez, C.A. Contreras-Pedraza, J. Rodríguez, A systematic analysis of non-centrifugal sugar cane processing: Research and new trends, Trends Food Sci. Technol. 107 (2021), https://doi.org/10.1016/j.tifs.2020.11.011.
[19] R.R. de Souza, D.D.S. Fernandes, P.H.G.D. Diniz, J.Y. Chen, X.W. Chen, Y.Y. Lin, G. C. Yen, J.A. Lin, M. Gavahian, A. Mousavi Khaneghah, Authentication of dark brown sugars from different processing using three-dimensional fluorescence spectroscopy, LWT 150 (2021), https://doi.org/10.1016/j.lwt.2021.111959.
[20] H.L. Gope, H. Fukai, Peaberry and normal coffee bean classification using CNN, SVM, and KNN: Their implementation in and the limitations of Raspberry Pi 3, AIMS Agric. Food. 7 (2022), https://doi.org/10.3934/agrfood.2022010.
[21] C. Hortinela, J.R. Balbin, J. Fausto, F.L. Valiente, J.C. Venturina, J.A.M. Mercado, M. Bryan, Classification of Cane Sugar Based on Physical Characteristics Using SVM, IEEE 11th Int. Conf. Humanoid, Nanotechnology, Inf. Technol. Commun. Control. Environ. Manag. HNICEM (2019), doi: 10.1109/HNICEM48295.2019.9072699.
[22] M. Bahramia, M.E. Honarvar, Measurement of Morphological Characteristics of Raw Cane Sugar Crystals Using Digital Image Analysis, J. Food Biosci. Technol. 5 (2015).
[23] M. Meenu, C. Kurade, B.C. Neelapu, S. Kalra, H.S. Ramaswamy, Y. Yu, A concise review on food quality assessment using digital image processing, Trends Food Sci. Technol. 118 (2021), https://doi.org/10.1016/j.tifs.2021.09.014.
[24] L.F. Santos Pereira, S. Barbon, N.A. Valous, D.F. Barbin, Predicting the ripening of papaya fruit with digital imaging and Random Forests, Comput. Electron. Agric. 145 (2018), https://doi.org/10.1016/j.compag.2017.12.029.
[25] F.B. de Santana, W. Borges Neto, R.J. Poppi, Random Forest as one-class classifier and infrared spectroscopy for food adulteration detection, Food Chem. 293 (2019), https://doi.org/10.1016/j.foodchem.2019.04.073.
[26] L. Breiman, Random forests, Mach. Learn. 45 (2001), https://doi.org/10.1023/A:1010933404324/METRICS.
[27] L. Wiener, A. Matthew, Classification and Regression by randomForest, R News 2 (2002).
[28] J.Y. Chen, X.W. Chen, Y.Y. Lin, G.C. Yen, J.A. Lin, Authentication of dark brown sugars from different processing using three-dimensional fluorescence spectroscopy, LWT 50 (2021), https://doi.org/10.1016/j.lwt.2021.111959.
[29] K.N. Galvis-Arias, L.D. Hidrobo-Pedroza, M.C. García-Muñoz, O.A. Mendieta-Menjura, M.P. Tarazona-Díaz, Effect of processing technology (traditional and ward furnace) on the physicochemical properties of non-centrifugal cane sugar (NCS), Rev. Fac. Ing. Univ. Antioquia (2019), https://doi.org/10.17533/udea.redin.20190839.
[30] T. Vera-Gutiérrez, M.C. García-Muñoz, A.M. Otálvaro-Alvarez, O. Mendieta-Menjura, Effect of processing technology and sugarcane varieties on the quality properties of unrefined non-centrifugal sugar, Heliyon 5 (2019), https://doi.org/10.1016/j.heliyon.2019.e02667.
[31] I.A.L. Instituto Adolfo Lutz, Métodos Físico-Químicos para Análise de Alimentos, 4. ed., São Paulo, 2008.
[32] J.R. Santos, O. Viegas, R.N.M.J. Páscoa, I.M.P.L.V.O. Ferreira, A.O.S.S. Rangel, J. A. Lopes, In-line monitoring of the coffee roasting process with near infrared spectroscopy: Measurement of sucrose and colour, Food Chem. 208 (2016), https://doi.org/10.1016/j.foodchem.2016.03.114.
[33] J.M. dos Santos, J.K. de Andrade, F. Galvão, M.L. Felsner, Optimization and validation of ultrasound-assisted extraction for the determination of micro and macro minerals in non-centrifugal sugar by F AAS, Food Chem. 292 (2019), https://doi.org/10.1016/j.foodchem.2019.04.037.
[34] R. Butuner, I. Cinar, Y.S. Taspinar, R. Kursun, M.H. Calp, M. Koklu, Classification of deep image features of lentil varieties with machine learning techniques, Eur. Food Res. Technol. 249 (2023), https://doi.org/10.1007/s00217-023-04214-z.
[35] S.I.F. Martins, W.M. Jongen, M.A.J. van Boekel, A review of Maillard reaction in food and implications to kinetic modelling, Trends Food Sci. Technol. 11 (2000), https://doi.org/10.1016/S0924-2244(01)00022-X.
[36] Y. Asikin, W. Takahara, M. Takahashi, N. Hirose, S. Ito, K. Wada, Compositional and Electronic Discrimination Analyses of Taste and Aroma Profiles of Non-Centrifugal Cane Brown Sugars, Food Anal. Methods 10 (2017), https://doi.org/10.1007/s12161-016-0746-5.
[37] L. Vargas Valencia, M. Hernández-Carrión, F. Velasquez, J. Espitia, J. Rodriguez Cortina, Functional and physicochemical properties of non-centrifugal cane sugar obtained by three concentration technologies, LWT 168 (2022), https://doi.org/10.1016/j.lwt.2022.113897.
[38] M. Weerawatanakorn, Y. Asikin, M. Takahashi, H. Tamaki, K. Wada, C.T. Ho, R. Chuekittisak, Physico-chemical properties, wax composition, aroma profiles, and antioxidant activity of granulated non-centrifugal sugars from sugarcane cultivars of Thailand, J. Food Sci. Technol. 53 (2016), https://doi.org/10.1007/s13197-016-2415-5.
[39] Y. Asikin, A. Kamiya, M. Mizu, K. Takara, H. Tamaki, K. Wada, Changes in the physicochemical characteristics, including flavour components and Maillard reaction products, of non-centrifugal cane brown sugar during storage, Food Chem. 149 (2014), https://doi.org/10.1016/j.foodchem.2013.10.089.
[40] M.G. Lindeman, P.F. O'Shea, Colorant removal during clarification and decolourisation processes, Proc. Aust. Soc. Sugar Cane Technol. 26 (2004).
[41] M.A. Clarke, M.A., Godshall, The nature of colorants in sugarcane and beet sugar manufacture, in Chemistry and Processing of Sugarbeet and Sugarcane, E.S. Publishers, Ed. Amsterdam, 1988.
[42] R. Riffer, The Nature of Colorants in Sugarcane and Cane Sugar Manufacture, in: M. A. Clarke, M.A. Godshall (Eds.), Sugar Series, Elsevier 9 (1988), doi: 10.1016/B978-0-444-43020-5.50019-9.
[43] N.P. Rodrigues, B. Brochier, J.K. de Medeiros, L.D.F. Marczak, G.D. Mercali, Phenolic profile of sugarcane juice: Effects of harvest season and processing by ohmic heating and ultrasound, Food Chem. 347 (2021), https://doi.org/10.1016/j.foodchem.2021.129058.
[44] H. Li, A. Guo, H. Wang, Mechanisms of oxidative browning of wine, Food Chem. 108 (2008), https://doi.org/10.1016/j.foodchem.2007.10.065.
[45] J. Cheng, J. Sun, K. Yao, M. Xu, Y. Cao, A variable selection method based on mutual information and variance inflation factor, Spectrochim, Acta Part A Mol. Biomol. Spectrosc. 268 (2022), https://doi.org/10.1016/j.saa.2021.120652.
[46] G. Paschos, Perceptually uniform color spaces for color texture analysis: an empirical evaluation, IEEE Trans. Image Process. 10 (2001), https://doi.org/10.1109/83.923289.
[47] Y. Fan, J. Li, Y. Guo, L. Xie, G. Zhang, Digital image colorimetry on smartphone for chemical analysis: A review, Measurement 171 (2021), https://doi.org/10.1016/j.measurement.2020.108829.
[48] A.M. Jiménez-Carvelo, A. González-Casado, M.G. Bagur-González, L. Cuadros-Rodríguez, Alternative data mining/machine learning methods for the analytical evaluation of food quality and authenticity – A review, Food Res. Int. 122 (2019), https://doi.org/10.1016/j.foodres.2019.03.063.
[49] Y. Li, J.-Y. Zhang, Y.-Z. Wang, FT-MIR and NIR spectral data fusion: a synergetic strategy for the geographical traceability of Panax notoginseng, Anal. Bioanal. Chem. 410 (2018), https://doi.org/10.1007/s00216-017-0692-0.
[50] F.B. de Santana, W. Borges Neto, R.J. Poppi, Random forest as one-class classifier and infrared spectroscopy for food adulteration detection, Food Chem. 293 (2019), https://doi.org/10.1016/j.foodchem.2019.04.073.