LWT - Food Science and Technology 152 (2021) 112372

Contents lists available at ScienceDirect

LWT

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

Edible seeds clustering based on phenolics and antioxidant activity using multivariate analysis

Hanna Elisia Araújo de Barros [a], Ana Cláudia Silveira Alexandre [a], Gabriela Aguiar Campolina [a], Gabriela Fontes Alvarenga [a], Lara Maria dos Santos Ferraz e Silva [a], Caio Vinicius Lima Natarelli [b], Elisângela Elena Nunes Carvalho [a], Eduardo Valério de Barros Vilas Boas [a,*]

[a] Food Science Department, Federal University of Lavras, 37200-900, Lavras, MG, Brazil

[b] Graduate Program in Materials Science and Engineering, Federal University of São Carlos, 13565-905, São Carlos, SP, Brazil

ARTICLE INFO

Keywords:
Bioactive compounds
Nuts
Principal component analysis
Hierarchical clusters analysis
Artificial neural network

ABSTRACT

Edible seeds, especially those known by the population as nuts, have their consumption associated with functional appeal. The present study aimed to compare and group nine different seeds, traditional and regional, according to their similarities, in terms of moisture, total phenolic compounds (TPC) and antioxidant activity, through multivariate analyses. All results were submitted to Principal Component Analysis (PCA), Hierarchical Clusters (HCA) and Kohonen's self-organizing maps (ANN/KSOM). The seeds differed in terms of moisture content, TPC and antioxidant activity. The walnut butterfly stood out with the highest levels of TPC and antioxidant activity. In the multivariate analyses application, three groups were formed: i) hazel, baru, Brazil, macadamia, almond and cashew; ii) pequi and marolo; iii) walnut butterfly. It is concluded that the seeds can be separated into three groups, with ANN/KSOMs being the most self-explanatory analysis and that regional seeds are nutritionally similar to those traditionally consumed.

1. Introduction

The bioactive compounds present in foods are substances capable of promoting beneficial physiological effects to the human health, such as the prevention or delay of chronic-degenerative diseases, reducing cholesterol levels in the blood, the risk of coronary heart disease and type 2 diabetes (Carocho et al. 2018; Silva et al. 2019). Antioxidant compounds, fatty acids, essential amino acids and fibers can be cited as examples of bioactive compounds (Giaconia et al., 2020). Due to these effects, the search for food sources of these substances is increasing.

Seeds of several fruits, including those commonly known as nuts, are widely consumed as sources of minerals, proteins, vitamins and fatty acids, although they also have bioactive compounds with antioxidant capacity (Moreira et al., 2019). Among the oil seeds traditionally consumed worldwide, cashew nuts (*Anacardium occidentale* L.), Brazil nuts (*Bertholletia excelsa*), macadamia (*Macadamia integrifolia*), almonds (*Prunus dulcis*), walnut butterfly (*Juglans regia* L.) and hazel (*Corylus avellana*) nuts stand out for having in their composition dietary fibers, phenolic compounds (phenolic acids, flavonoids, tannins and anthocyanins), phospholipids, sphingolipids, sterols and tocopherols, in addition to several vitamins and minerals, such as selenium, iron, copper and manganese (Rengel et al., 2015; Rico et al., 2016; Ros et al., 2010; Trox et al., 2010; Viera-Alcaide et al., 2019).

Seeds of native species from Brazil, such as pequi seed (*Caryocar brasiliense* Camb.), baru nut (*Dipteryx alata*) and marolo seed (*Annona crasssiflora*), are alternatives that can be consumed due to their functional appeal (Reis & Schmiele, 2019). According to Lima et al. (2007), the pequi seed presents in its composition a high content of phenolic and carotenoid compounds. Baru nuts contain omega 6 and 9 polyunsaturated fatty acids and phenolic compounds (Santiago et al., 2018). Marolo seed contains phytosterols, tocopherols and unsaturated fatty acids (Luzia & Jorge, 2013).

The interpretation of chemical data from food analysis through the use of chemometrics has recently gained prominence, since it can be used to evaluate similarities/differences between various foods, or even to project objects (samples) on a two-dimensional factorial plane based on different characteristics using different mathematical and statistical methods. Chemometrics is an area that involves the analysis of chemical

* Corresponding author.
E-mail address: evbvboas@ufla.br (E.V.B. Vilas Boas).

https://doi.org/10.1016/j.lwt.2021.112372
Received 3 August 2020; Received in revised form 23 August 2021; Accepted 26 August 2021
Available online 27 August 2021

data of a multivariate nature and can be divided into calibration, classification and exploratory methods (Granato et al., 2018). Among the various existing exploratory methods, there are Principal Component Analysis (PCA) and Hierarchical Cluster Analysis (HCA), techniques well known for meeting the needs of researchers and due to the ease of interpretation of the results generated. Another existing method known as Artificial Neural Networks (ANNs), has been showing itself as a trend for applications involving technologies combined with food quality, optimization and standardization, as seen in the study of de Barros et al. (2020) in which Kohonen's Self-Organizing Maps (KSOMs) were applied to cluster cookies according to the similarities of their nutritional properties. ANNs assumes the simulating of the human brain behavior (Grekousis, 2019; Moreira et al., 2019) and learning takes place through simple algorithms capable of organizing dimensionally complex data in clusters according to their similarities (Kohonen, 1982).

The new demands of the consumer market are inclined towards the search for foods that have functional health properties, especially those that contain high antioxidant capacity (Carocho et al., 2018). Considering that most of the works in the literature that study nuts revolve around the mineral composition and fatty acid profile, the opportunity arises to study the presence of bioactive compounds and antioxidant activity, aiming to present a food with functional appeal. It is also emphasized the importance of studying alternative seeds or nuts, which can be consumed. To the best of our knowledge, no study has been conducted and published about the use of multivariate tools to evaluate edible seeds with functional appeal. Given the above, the present study aimed to determine the moisture content, total phenolic compounds and the antioxidant activity of seeds, traditionally and regionally consumed, and to evaluate the results using three multivariate tools (PCA, HCA and ANNs), aiming to compare them and group them according to the similarities.

2. Materials and methods

2.1. Raw material

Seeds (hazel, almonds, baru, walnut butterfly, cashews, macadamia nuts and Brazil nuts) were purchased from retailers in the municipality of Lavras – MG, Brazil. Pequi and Marolo seeds were taken from fruits acquired from producers in the municipalities of Montes Claros - MG and Curvelo - MG, Brazil, respectively. Processing was carried out at the Fruit and Vegetable Post-Harvest Laboratory at the Federal University of Lavras, Lavras - MG, Brazil, using a multipurpose mill (Tecnal TE-631/4, Brazil) and an 80 mesh sieve, until obtaining of fine powder.

2.2. Analyses

2.2.1. Determination of moisture

Moisture was determined by the gravimetric technique using heat in a ventilated oven at 65 °C until constant weight, according to AOAC International (2019) official method number 925.40.

2.2.2. Total phenolics compounds (TPC) and antioxidant activity

2.2.2.1. Preparation of extracts. The extracts were prepared according to the methodology described by Larrauri et al. (1997) with adaptations. Five grams of crushed samples and 20 mL of methanol 50% were placed in centrifuge tubes, vortexed for 1 min and after an hour of rest in a dark environment, the homogenate was centrifuged (Beckman GS-15R, USA) at 4 °C for 15 min at 21952 g. The supernatant was reserved and re-extraction was carried out, under the same conditions, with 20 mL of acetone 70%. Subsequently, the two supernatants were combined, filtered on a Whatman paper No. 1 and the volume was made up to 50 mL. The extracts, obtained in triplicate, were stored in amber glass bottles at −18 °C.

2.2.2.2. Determination of TPC. The content of TPC was determined using the method adapted from Folin-Ciocalteau (Waterhouse, 2002). In test tubes, 0.5 mL aliquots of the extracts, 2.5 mL of the 10% Folin Ciocalteau reagent and 2 mL of 4 g 100 mL^{-1} sodium carbonate were combined and stirred. Absorbances were measured on a spectrophotometer (Varian Cary 50, USA) at 750 nm, after 2 h of rest. The results were expressed in mg gallic acid equivalent (GAE) 100 g^{-1} of sample.

2.2.2.3. Determination of antioxidant activity: ABTS^{*+}, β-carotene/linoleic acid and FRAP assays. The antioxidant activities were determined following Rufino et al. (2010). The results of the 2,2′-azino-bis (3-ethylbenzothiazoline-6-sulfonic acid) - ABTS^{*+} methodology were expressed in μmol Trolox g^{-1} of sample, β-carotene/linoleic acid were presented in percentage of oxidation inhibition (% of protection) and Ferric Reducing Antioxidant Power (FRAP) in μmol ferrous sulphate g^{-1} of sample.

2.3. Statistics

The experiment was conducted using a completely randomized design (CRD), with 9 treatments, that were the different seeds. The analytical determinations (moisture content, total phenolic compounds and antioxidant activity) were performed in 3 repetitions and in triplicate.

The R version 4.0.2 program together with the FactoMineR and factoextra packages was used to apply Principal Component Analysis (PCA) and Hierarchical Cluster Analysis (HCA) to the data obtained. For the HCA, the data were scaled between 0 and 100 using the normalize function of the BBmisc package. In addition, Kohonen's self-organizing maps were applied to cluster the samples, according to the similarity of their properties. The SOM Toolbox 2.1 package (Vatanen et al., 2015) was used in the Matlab R2015a program, together with the necessary modifications to improve the obtaining and validation of the clusters, through the use of the k-means algorithm together with the Davies-Bouldin and Silhouette indexes. The training was realized using a batch algorithm that uses a Gaussian neighborhood function and two training phases: first, a rough phase with an initial radius of 2 and a training length of 53, and, lastly, a fine-tuning phase with an initial radius of 1 and training length of 212.

3. Results and discussions

3.1. Moisture content, TPC and antioxidant activity

The averages of moisture content, total phenolic compounds (TPC) and antioxidant activity (ABTS^{*+}, β-carotene/linoleic acid and FRAP assays) of the seeds are shown in Table 1. In this table, it was observed that there was a significant difference between the samples ($p < 0.05$) by the Tukey's test, for all determinations performed.

The moisture content of the seeds ranged from 2.23 g 100 g^{-1} (macadamia) to 38.94 g 100 g^{-1} (pequi) (Table 1). It is noteworthy that the seeds acquired in the market presented moisture varying from 2.23 g 100 g^{-1} (macadamia) to 4.77 g 100 g^{-1} (hazel), since they were subjected to dehydration aiming at greater conservation and safety in commercialization. The pequi and marolo seeds, which did not undergo dehydration, presented moisture of 26.23 g 100 g^{-1} and 38.94 g 100 g^{-1}, respectively.

The moisture values determined are within the ranges mentioned by Ferreira et al. (2016) for almond (3.10–9.50 g 100 g^{-1}), hazel (4.00–5.30 g 100 g^{-1}), walnut butterfly (2.70–4.70 g 100 g^{-1}), Brazil nut (3.10–3.50 g 100 g^{-1}) and macadamia (1.40–2.10 g 100 g^{-1}). The moisture averages found for baru, cashew nuts and pequi seed are close to those reported in the literature by Sousa et al. (2011) (3.49 g 100 g^{-1}), Lima et al. (2004) (3.29 g 100 g^{-1}) and Araújo et al. (2018) (36.20 g 100 g^{-1}), respectively. The average moisture observed in the marolo seed

Table 1
Moisture content, total phenolic compounds (TPC) and antioxidant activity (ABTS*+, β-Carotene/linoleic acid and FRAP) of the nine edible seeds. TPC and antioxidant activity done in dry weight (DW).

Nuts and seeds	Moisture (g 100 g^{-1})	TPC (mg GAE 100 g^{-1})	ABTS*+ (μmol Trolox g^{-1})	β-Carotene/linoleic acid (% of protection)	FRAP (μmol ferrous sulphate g^{-1})
Almond	4.82 ± 0.15[c]	54.2 ± 1.1[bc]	194 ± 29.5[c]	49.1 ± 4.6[c]	441 ± 17.5[c]
Hazel	4.77 ± 0.23[c]	74.5 ± 0.6[bc]	86.8 ± 27.6[c]	67.8 ± 1.5[b]	1265 ± 121[b]
Baru	3.41 ± 0.20[d]	49.2 ± 0.3[bc]	473 ± 20.0[bc]	66.5 ± 1.9[b]	357 ± 21.5[c]
Walnut Butterfly	3.14 ± 0.83[de]	573 ± 51.1[a]	11102 ± 867[a]	67.6 ± 1.4[b]	28305 ± 586[a]
Brazil	2.98 ± 0.16[de]	34.9 ± 3.4[c]	1297 ± 127[b]	67.8 ± 1.1[b]	126 ± 7.1[c]
Cashew	2.85 ± 0.17[de]	34.4 ± 1.1[c]	421 ± 131[c]	49.5 ± 2.3[c]	204 ± 4.7[c]
Macadâmia	2.23 ± 0.07[e]	35.0 ± 1.8[c]	169 ± 10.6[c]	61.9 ± 3.1[b]	227 ± 34.3[c]
Marolo	26.23 ± 0.13[b]	96.1 ± 0.6[b]	205 ± 16.0[c]	90.7 ± 1.8[a]	659 ± 9.8[c]
Pequi	38.94 ± 0.3[a]	39.0 ± 1.8[c]	362 ± 34.1[c]	89.4 ± 4.9[a]	260 ± 6.3[c]

Means ± standard deviations followed by the same letter in the same column do not differ from each other by the Tukey's test ($p < 0.05$).

was higher than that reported by Barros et al. (2018) (10.10 g 100 g^{-1}). The moisture content of the nuts is normally directly related to their oxidative stability, as well as their microbiological safety. As long as the food is not frozen, the higher the moisture in the food, the greater its susceptibility to microbiological deterioration. On the other hand, the lower the water content of the food, the greater its susceptibility to lipid oxidation, although intrinsic antioxidant compounds can mitigate it. Thus, the moisture content must be monitored, as well as the bioactive compounds, in order to extend the useful life of the nuts (Carocho et al., 2018).

The TPC contents of the seeds varied from 573 (Walnut Butterfly) to 34.4 (Cashew) mg gallic acid equivalent (GAE) 100 g^{-1}. The walnut butterfly stood out, with about 8–24 times more TPC than the other seeds. In the present study, regionally consumed seeds (pequi, marolo and baru) presented similar TPC to most traditionally consumed seeds, as shown in Table 1. The variations on the phenolic content of the analysed seeds are justified, since the species, the fruit's maturation stage, the soil and the environmental conditions under which the mother plant and the seed were subjected influence the TPC levels (Herbello-Hermelo et al., 2018). In addition, Dolatabadi et al. (2015) reported that the time and storage conditions can cause changes in the levels of TPC in chestnuts.

With the exception of the walnut butterfly (573 mg gallic acid equivalent (GAE) 100 g^{-1}), the seeds analysed in the present study had lower TPC contents than those obtained by Herbello-Hermelo et al. (2018), who analysed 17 samples of different edible seeds, which presented TPC contents between 122 and 543 mg gallic acid equivalent (GAE) 100 g^{-1}. In the study by Oliveira et al. (2018), who analysed 8 Portuguese chestnut cultivars, the phenolic content was close to that of this study, with a maximum of 347 mg gallic acid equivalent (GAE) 100 g^{-1} and a minimum of 32 mg gallic acid equivalent (GAE) 100 g^{-1}.

Intrinsic antioxidants can prevent the self-oxidation of compounds present in food, scavenging free radicals from some oxidative reactions. Based on this assumption, such antioxidant compounds have the potential to preserve the sensory and nutritional properties of the food, thus prolonging its storage time (Carocho et al., 2018). Therefore, the antioxidant activity of the seeds was evaluated using three different methods (ABTS*+, β-carotene/linoleic acid and FRAP), considering that they present different mechanisms of action in the most varied food matrices (Choi et al., 2002).

The ABTS*+ test measures the relative ability of antioxidant substances to eliminate the ABTS*+ compared to a standard concentration of Trolox, a water-soluble compound, analogous to α-tocopherol. The absorbance of the ABTS*+ decreases proportionally with the ease of reduction of this radical by electron donor species (Shaikh et al., 2019). The extracts derived from the seeds showed antioxidant activity by ABTS*+ ranging from 86.8 (hazel) to 11102 μmol Trolox g^{-1} (walnut butterfly). As observed for TPC, the walnut butterfly nut showed the highest antioxidant activity ABTS*+, about 23–128 times higher than that of the other seeds.

All seeds showed a protective capacity against oxidation of β-carotene. Few variations were observed in the percentage of protection between them, with the highest percentage of protection observed in marolo and pequi seed, with an average of 90.0% protection, approximately. The lowest percentages were observed in almonds and cashews nuts, on an average of 49.0% protection, approximately. Rufino et al. (2010) classify foods in categories of high, intermediate and low antioxidant capacity, when they present protection >70%, 40–70% and <40%, respectively. Therefore, it is possible to classify the analysed seeds as having an intermediate ability to protect against oxidation of β-carotene, which is caused by linoleic acid, with the exception of marolo and pequi seeds, which have a high protective capacity for β-carotene. The difference between the protective capacities of the seeds can be attributed to the presence of different antioxidant species, which interact with each other, being able to act in a synergistic or inhibitory way (Salcedo et al., 2010). In addition, the difference in the proportion of hydrophobic and hydrophilic compounds also contributes to this variation, since the β-carotene/linoleic acid model is similar to an oil-in-water emulsion system, whose hydrophobic compounds tend to be more effective in preventing oxidation, as they are oriented in the oil and in the oil-water interface (Wijeratne et al., 2006).

The FRAP method is based on the reduction of the ferric-tripyridyltriazine complex (Fe III-TPTZ) to ferrous complex (Fe II-TPTZ) in the presence of antioxidants (Blomhoff et al., 2006). The antioxidant activity evaluated by the FRAP method varied between 126 (Brazil nuts) and 28305 (walnut butterfly) μmol ferrous sulphate g^{-1}. Once again, the walnut butterfly stood out, averaging 22 to 224 times higher than other seeds, followed by hazel.

Blomhoff et al. (2006), when evaluated the antioxidant activity of different species of nuts (almond, Brazil, cashew, hazel, macadamia and walnut butterfly), by the FRAP method, obtained results close to those presented in Table 1. The high antioxidant activity of the walnut butterfly, when compared to the others, is probably due to its high concentration of polyphenols and the presence of the film that makes up this nut. In this film, most of the antioxidants are located, with only 10% located in the nuts when the film is removed (Blomhoff et al., 2006). Ozer (2017) when comparing the antioxidant activity of walnut butterfly, almonds and peanuts by the FRAP method also obtained higher values for the walnut butterfly, 22.64 ± 1.1 mmol ferrous sulphate g^{-1}, a value lower than that shown in Table 1.

3.2. Multivariate analyses

3.2.1. Principal component analysis (PCA)

PCA is a statistical test that belongs to a group of factor analysis, which aims to represent the variation present in the data set using a small number of factors (Granato et al., 2018). The loadings of the variables, as well as the percentage of data variation explained and the accumulated variation for the extracted main components, are described in Table 2. The first two components (PC1 and PC2) explain 96.57% of the data variation, where PC1 and PC2 represent 60.75% and 35.82% of the total variation, respectively. The variables responsible for the

Table 2
Loadings of the variables together with the percentage of variance and accumulated variance for each main component of the Principal Component Analysis (PCA).

	PC1	PC2
TPC	0.983	0.164
ABTS*+	0.989	0.108
FRAP	0.991	0.121
Moisture	−0.314	0.907
β-carotene	−0.108	0.956
Variance (%)	60.75	35.82
Cumulative variance (%)	60.75	99.57

separations in PC1 include TPC, ABTS*+ and FRAP, as their loadings are 0.983, 0.989 and 0.991, respectively, thus showing a positive correlation between these variables. As for PC2, the moisture and β-carotene variables, with loadings equals to 0.907 and 0.956, respectively, are responsible for the separations that occurred in this component, and also showed a positive correlation between these variables.

For the visual analysis, a two-dimensional projection of the loads and scores in the respective dimensional spaces is presented (Fig. 1). Based on PC1 and PC2, there is a tendency for the samples to get closer, forming three groups. A first group consisting of hazel, baru, Brazil, macadamia, almond and cashew nuts, a second group of pequi and marolo, and another group formed by the walnut butterfly. The main factor responsible for the separation of the groups is the difference in the levels of TPC and antioxidant activity (FRAP and ABTS*+) between the samples (Table 2), with a notable separation between the walnut butterfly, with the highest averages of these variables, and the others, with the lowest values (Fig. 1). Also, the moisture and β-carotene were responsible for the separation of the pequi and marolo seeds from the rest of the samples. It is noteworthy that the seeds tend to be organized in groups, regardless of whether they are of traditional or regional consumption.

3.2.2. Hierarchical clusters analysis (HCA)

The data from the analyses performed on the seeds were submitted to the HCA, whose samples were grouped according to their similarities, as shown in the resulting dendrogram (Fig. 2). It is noteworthy that the shorter the line connecting two groups, the greater the affinity between them. It is noted that the HCA corroborates the PCA, with the gathering of seeds in three groups, similar to those observed in the PCA, regardless of whether they are of traditional or regional consumption. However, despite the HCA being prepared with data from the analyses carried out in the present study, allowing to present the groups formed, it is not possible to infer, based on Fig. 2, which parameters are responsible for the groupings of the samples. In addition, the HCA analysis showed that the group of the hazel, baru, Brazil, macadamia, almond and cashew nuts may be formed by two sub-groups, one with almond and cashew, and another with macadamia, Brazil, hazel and baru.

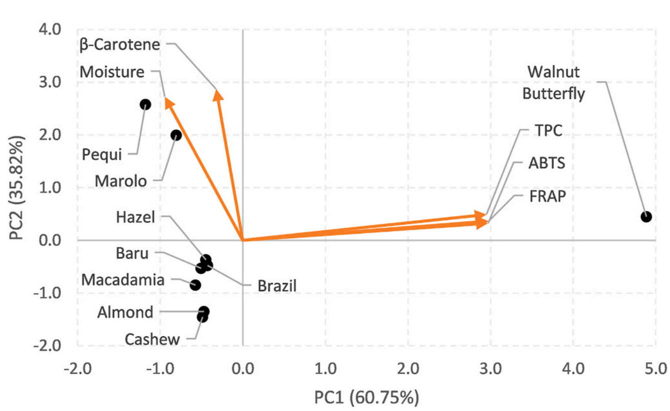

Fig. 1. Principal component analysis (PCA) for the edible seeds samples.

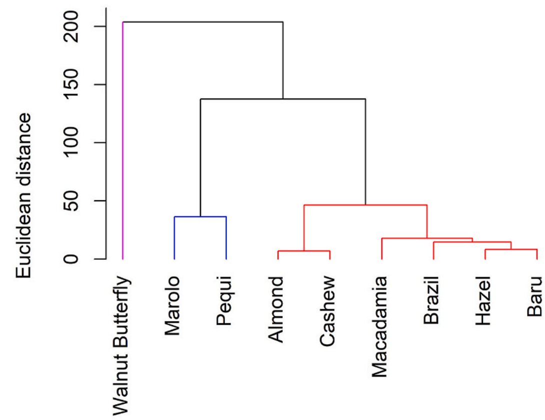

Fig. 2. Hierarchical cluster analysis (HCA) for the edible seeds samples.

3.2.3. Kohonen self-organizing maps (ANN/KSOM)

In the neural network (ANN/KSOM), maps with different dimensions were generated in order to obtain the arrangement that has the best cluster validation indexes (Davies-Bouldin and Silhouette indexes), together with lower measurement errors (quantization and topographical errors). Thus, the Davies-Bouldin index (DB) represents the degree of similarity between the clusters, regardless of the number of clusters and the partition method used. The closer to zero, the better the data partition. The Silhouette index (S) defines the quality of the group based on the proximity between the objects of a certain group and the proximity of these objects to the closest group. Therefore, the closer it is to 1, the more effective the grouping performed was (Davies & Bouldin, 1979; Moreira et al., 2019; Rousseeuw, 1987). Based on this assumption, the hexagonal 13 × 11 was chosen, which had the Davies-Bouldin index equal to 0.342 and Silhouette equal to 0.900. In addition, the chosen map had the quantization error (QE) of 0.010, topographic error (TE) of 0 and combined error of 0.028, which are measures of accuracy, continuity and the two combined measures, respectively (Serrien et al., 2017). According to Silva et al. (2019), the closer it is to zero, the better the measurement errors, despite not having a standard value.

The two-dimensional cluster neural map (or topological map of the network), the component maps of each analysis performed and the U-matrix are shown in Fig. 3A, considering the nine seeds and three repetitions. The variation of the results obtained in the analytical determinations can be observed by the color gradient (right end of each map), except for the U-matrix, in which the scale indicates the distance between the adjacent neurons, with the dark red color indicating a large distance between the neurons and the dark blue color indicating a smaller distance.

The analysis of the U-matrix shows the existence of two dark blue regions, one in the top right portion of the map and another in the bottom right portion, indicating the presence of at least two clusters. The use of the k-means algorithm with the Davies-Bouldin and Silhouette indexes generates the two-dimensional cluster neural map (Fig. 3B), that show the division of the samples, again, into three groups. In the right part of the map, there are two groupings, as shown in the U-matrix, with one at the top formed by the walnut butterfly and another at the bottom formed by the pequi and marolo seeds. On the left part of the map, the other samples (almond, cashew, Brazil, macadamia, baru and hazel) forms another cluster. All groupings performed were based on the similarities between the samples, indicated by their proximity. The observed clusters corroborate those obtained by PCA and HCA.

As the position occupied by a sample in the neural map corresponds to the same position in the component map for each analysis performed (Fig. 3A), the values found for the five variables analysed in every seed and their repetitions can be found by locating them in the same

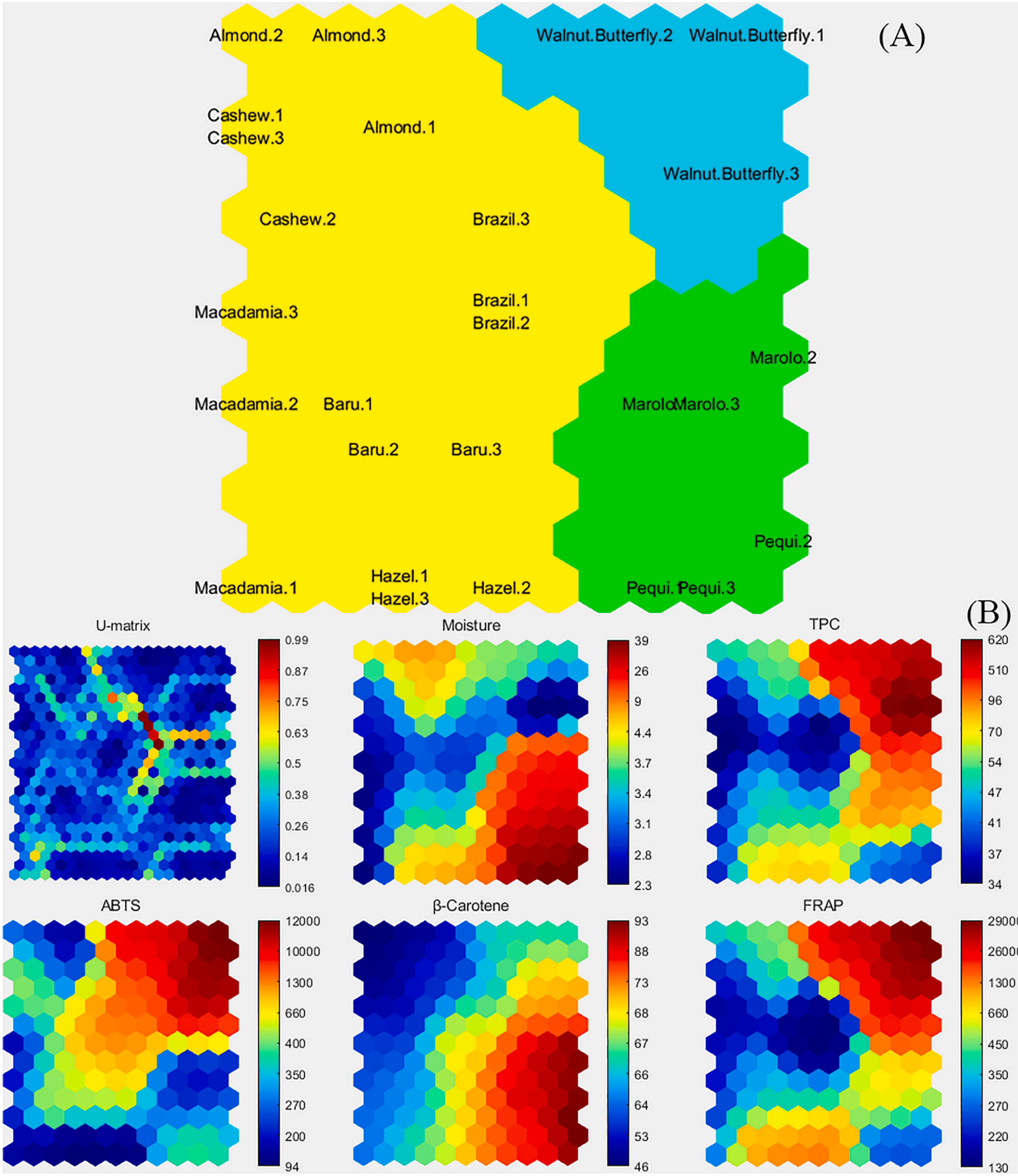

Fig. 3. (A) Two-dimensional cluster neural map showing the formation of five groups with their respective samples; (B) Component maps and distance matrix (U matrix) for data generated through the analysis of moisture, total phenolic compounds (TPC), ABTS^{*+}, β-carotene/linoleic acid and FRAP of the edible seed samples, expressed in g 100 g^{-1}, mg GAE 100 g^{-1}, μmol Trolox g^{-1}, % of protection and μmol ferrous sulphate g^{-1}, respectively.

corresponding neuron in each component map. It allows the identification of which variables were responsible for the grouping and separation of the seeds. So, the component maps show the walnut butterfly with the highest TPC content and the highest antioxidant activity measured by ABTS*+ and FRAP methods. The pequi and marolo seeds form a group that stands out for the greater antioxidant activity measured by the β-carotene/linoleic acid method, as well as the greater moisture among the samples. The lower and intermediate values of the variables induced the formation of a cluster by the other samples (almond, cashew, Brazil, macadamia, baru and hazel).

Accordingly, the results obtained in ANN/KSOM (Fig. 3) were consistent with the data in Table 1 and the PCA and HCA analyses discussed earlier, allowing the grouping of seeds according to their similarities, regardless of whether they are traditional or regional. Furthermore, ANN/KSOM allows for a more intuitive and visual interpretation when compared to other conventional statistical methods.

4. Conclusions

The nine seeds evaluated showed a high content of TPC and antioxidant activity, especially the walnut butterfly. The multivariate analyses applied (PCA, HCA and ANN/KSOM) corroborated each other, allowing the formation of three groups: i) hazel, baru, Brazil, macadamia, almond and cashew; ii) pequi and marolo; iii) walnut butterfly. Although HCA and PCA make it possible to reveal relevant information from the data set, ANN/KSOM stood out for being a more complete, visual and self-explanatory statistical method, in addition to enabling greater minimization of modeling errors, leading to a more accurate description of the results behavior. Regional seeds (Pequi and Marolo) and the baru nuts are similar to those traditionally consumed, in terms of TPC and antioxidant activity, thus configuring new alternatives for the food consumer market with antioxidant activities and functional appeal.

Funding

The authors are grateful for the financial support of Conselho Nacional de Desenvolvimento Científico e Tecnológico (Bolsas de Produtividade em Pesquisa 302259/2018–0, 302699/2019–8, CNPq, Brazil), Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG PPM-00458-15), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES: 88881.068456/2014–01) and Federal University of Lavras (UFLA, Lavras, Minas Gerais, Brazil).

CRediT authorship contribution statement

Hanna Elisia Araújo de Barros: Conceptualization, Methodology, Investigation, Formal analysis, Writing – original draft. **Ana Cláudia Silveira Alexandre:** Conceptualization, Investigation, Writing – original draft. **Gabriela Aguiar Campolina:** Investigation, Writing – original draft. **Gabriela Fontes Alvarenga:** Investigation, Writing – original draft. **Lara Maria dos Santos Ferraz e Silva:** Investigation, Writing – original draft. **Caio Vinicius Lima Natarelli:** Data curation, Formal analysis, Writing – original draft. **Elisângela Elena Nunes Carvalho:** Funding acquisition, Project administration, Supervision, Writing – review & editing. **Eduardo Valério de Barros Vilas Boas:** Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that there are no conflicts of interest.

References

Araújo, A. C. M. A., Oliveira, É. R. de, Menezes, E. G. T., Dias, B. O., Terra, A. W. C., & Queiroz, F. (2018). Solvent effect on the extraction of soluble solids from murici and pequi seeds. *Journal of Food Process Engineering, 41*(6), Article e12813. https://doi.org/10.1111/jfpe.12813

Barros, A. P. G., Dos Santos, L. A., Santos, A. K. do C. F. dos, Rosa e Silva, P. S. da, De Araújo, C., & Pimenta, A. C. (2018). Characterization of fruits and seeds of araticum (*Annona crassiflora* mart.) native to the cerrado of mato grosso. *Agropecuária Científica No Semiárido - ACSA, 14*(4), 280. https://doi.org/10.30969/acsa.v14i4.1001

Blomhoff, R., Carlsen, M. H., Andersen, L. F., & Jacobs, D. R. (2006). Health benefits of nuts: Potential role of antioxidants. *British Journal of Nutrition, 96*(S2), S52–S60. https://doi.org/10.1017/BJN20061864

Carocho, M., Morales, P., & Ferreira, I. C. F. R. (2018). Antioxidants: Reviewing the chemistry, food applications, legislation and role as preservatives. *Trends in Food Science & Technology, 71*, 107–120. https://doi.org/10.1016/j.tifs.2017.11.008

Choi, C. W., Kim, S. C., Hwang, S. S., Choi, B. K., Ahn, H. J., Lee, M. Y., Park, S. H., & Kim, S. K. (2002). Antioxidant activity and free radical scavenging capacity between Korean medicinal plants and flavonoids by assay-guided comparison. *Plant Science, 163*(6), 1161–1168. https://doi.org/10.1016/S0168-9452(02)00332-1

Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI, 1*(2), 224–227. https://doi.org/10.1109/TPAMI.1979.4766909

de Barros, H. E. A, Natarelli, C. V. L., Tavares, I. M. de C., Oliveira, A. L. M. de, Araújo, A. B. S., Pereira, J., … Franco, M. (2020). Nutritional clustering of cookies developed with cocoa shell, soy, and green banana flours using exploratory methods. *Food and Bioprocess Technology, 13*(9), 1566–1578. https://doi.org/10.1007/s11947-020-02495-w

Dolatabadi, K. S. M., Dehghan, G., Hosseini, S., & Esfahlan, A. J. (2015). Effect of five year storage on total phenolic content and antioxidant capacity of almond (*Amygdalus communis* L.) hull and shell from different genotypes. *Avicenna Journal of Phytomedicine, 5*(1), 26–33. https://doi.org/10.22038/ajp.2014.3512

Ferreira, I., Morales, P., & Barros, L. (2016). *Wild plants, mushrooms and nuts: Functional food properties and applications*. John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118944653

Giaconia, M. A., Ramos, S. dos P., Pereira, C. F., Lemes, A. C., De Rosso, V. V., & Braga, A. R. C. (2020). Overcoming restrictions of bioactive compounds biological effects in food using nanometer-sized structures. *Food Hydrocolloids, 107*, 105939. https://doi.org/10.1016/j.foodhyd.2020.105939

Granato, D., Santos, J. S., Escher, G. B., Ferreira, B. L., & Maggio, R. M. (2018). Use of principal component analysis (PCA) and hierarchical cluster analysis (HCA) for multivariate association between bioactive compounds and functional properties in foods: A critical perspective. *Trends in Food Science & Technology, 72*, 83–90. https://doi.org/10.1016/j.tifs.2017.12.006

Grekousis, G. (2019). Artificial neural networks and deep learning in urban geography: A systematic review and meta-analysis. *Computers, Environment and Urban Systems, 74*, 244–256. https://doi.org/10.1016/j.compenvurbsys.2018.10.008

Herbello-Hermelo, P., Lamas, J. P., Lores, M., Domínguez-González, R., Bermejo-Barrera, P., & Moreda-Piñeiro, A. (2018). Polyphenol bioavailability in nuts and seeds by an in vitro dialyzability approach. *Food Chemistry, 254*, 20–25. https://doi.org/10.1016/j.foodchfem.2018.01.183

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics, 43*(1), 59–69. https://doi.org/10.1007/BF00337288

Larrauri, J. A., Rupérez, P., & Saura-Calixto, F. (1997). Effect of drying temperature on the stability of polyphenols and antioxidant activity of red grape pomace peels. *Journal of Agricultural and Food Chemistry, 45*(4), 1390–1393. https://doi.org/10.1021/jf960282f

Lima, A. C., García, N. H. P., & Lima, J. R. (2004). Obtention and characterization of the main cashew products. *Boletim do Centro de Pesquisa de Processamento de Alimentos, 22*(1). https://doi.org/10.5380/cep.v22i1.1185

Lima, A. de, Silva, A. M., de, O. e, Trindade, R. A., Torres, R. P., & Mancini-Filho, J. (2007). Chemical composition and bioactive compounds in the pulp and almond of pequi fruit (*Caryocar brasiliense*, Camb.). *Revista Brasileira de Fruticultura, 29*(3), 695–698. https://doi.org/10.1590/S0100-29452007000300052

Luzia, D. M. M., & Jorge, N. (2013). Bioactive substance contents and antioxidant capacity of the lipid fraction of *Annona crassiflora* Mart. seeds. *Industrial Crops and Products, 42*, 231–235. https://doi.org/10.1016/j.indcrop.2012.05.027

Moreira, L. S., Chagas, B. C., Pacheco, C. S. V., Santos, H. M., de Menezes, L. H. S., Nascimento, M. M., Batista, M. A. S., de Jesus, R. M., Amorim, F. A. C., Santos, L. N., & da Silva, E. G. P. (2019). Development of procedure for sample preparation of cashew nuts using mixture design and evaluation of nutrient profiles by Kohonen neural network. *Food Chemistry, 273*, 136–143. https://doi.org/10.1016/j.foodchem.2018.01.050

Official Methods of Analysis Aoac International. (2019). *21th ed., Gaithersburg, MD, USA*. Official Method 925.40.

Oliveira, I., Meyer, A., Afonso, S., Ribeiro, C., & Gonçalves, B. (2018). Morphological, mechanical and antioxidant properties of Portuguese almond cultivars. *Journal of Food Science & Technology, 55*(2), 467–478. https://doi.org/10.1007/s13197-017-2955-3

Ozer, H. K. (2017). Phenolic compositions and antioxidant activities of Maya nut (*Brosimum alicastrum*): Comparison with commercial nuts. *International Journal of Food Properties, 20*(11), 2772–2781. https://doi.org/10.1080/10942912.2016.1252389

Reis, A. F., & Schmiele, M. (2019). Características e potencialidades dos frutos do Cerrado na indústria de alimentos. *Brazilian Journal of Food Technology, 22*. https://doi.org/10.1590/1981-6723.15017

Rengel, A., Pérez, E., Piombo, G., Ricci, J., Servent, A., Tapia, M. S., Gibert, O., & Montet, D. (2015). Lipid profile and antioxidant activity of macadamia nuts (*Macadamia integrifolia*) cultivated in Venezuela. *Natural Science, 7*(12), 535–547. https://doi.org/10.4236/ns.2015.712054

Rico, R., Bulló, M., & Salas-Salvadó, J. (2016). Nutritional composition of raw fresh cashew (*Anacardium occidentale* L.) kernels from different origin. *Food Sciences and Nutrition, 4*(2), 329–338. https://doi.org/10.1002/fsn3.294

Ros, E., Tapsell, L. C., & Sabaté, J. (2010). Nuts and berries for heart health. *Current Atherosclerosis Reports, 12*(6), 397–406. https://doi.org/10.1007/s11883-010-0132-5

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics, 20*, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Rufino, M. do S. M., Alves, R. E., de Brito, E. S., Pérez-Jiménez, J., Saura-Calixto, F., & Mancini-Filho, J. (2010). Bioactive compounds and antioxidant capacities of 18 non-traditional tropical fruits from Brazil. *Food Chemistry, 121*(4), 996–1002. https://doi.org/10.1016/j.foodchem.2010.01.037

Salcedo, C. L., López de Mishima, B. A., & Nazareno, M. A. (2010). Walnuts and almonds as model systems of foods constituted by oxidisable, pro-oxidant and antioxidant factors. *Food Research International, 43*(4), 1187–1197. https://doi.org/10.1016/j.foodres.2010.02.016

Santiago, G. de L., de Oliveira, I. G., Horst, M. A., Naves, M. M. V., & Silva, M. R. (2018). Peel and pulp of baru (*Dipteryx Alata* Vog.) provide high fiber, phenolic content and antioxidant capacity. *Food Science and Technology, 38*(2), 244–249. https://doi.org/10.1590/1678-457X.36416

Serrien, B., Goossens, M., & Baeyens, J.-P. (2017). Issues in using self-organizing maps in human movement and sport science. *International Journal of Computer Science in Sport, 16*(1), 1–17. https://doi.org/10.1515/ijcss-2017-0001

Shaikh, S. A. M., Singh, B. G., Barik, A., Balaji, N. V., Subbaraju, G. V., Naik, D. B., & Priyadarsini, K. I. (2019). Unravelling the effect of β-diketo group modification on the antioxidant mechanism of curcumin derivatives: A combined experimental and DFT approach. *Journal of Molecular Structure, 1193*, 166–176. https://doi.org/10.1016/j.molstruc.2019.05.029

Silva, E. dos S., da Silva, E. G. P., Silva, D. dos S., Novaes, C. G., Amorim, F. A. C., dos Santos, M. J. S., & Bezerra, M. A. (2019). Evaluation of macro and micronutrient elements content from soft drinks using principal component analysis and Kohonen self-organizing maps. *Food Chemistry, 273*, 9–14. https://doi.org/10.1016/j.foodchem.2018.06.021

Silva, L. B. A. R., Pinheiro-Castro, N., Novaes, G. M., Pascoal, G. de F. L., & Ong, T. P. (2019). Bioactive food compounds, epigenetics and chronic disease prevention: Focus on early-life interventions with polyphenols. *Food Research International, 125*, 108646. https://doi.org/10.1016/j.foodres.2019.108646

Sousa, A. G. de O., Fernandes, D. C., Alves, A. M., de Freitas, J. B., & Naves, M. M. V. (2011). Nutritional quality and protein value of exotic almonds and nut from the Brazilian Savanna compared to peanut. *Food Research International, 44*(7), 2319–2325. https://doi.org/10.1016/j.foodres.2011.02.013

Trox, J., Vadivel, V., Vetter, W., Stuetz, W., Scherbaum, V., Gola, U., Nohr, D., & Biesalski, H. K. (2010). Bioactive compounds in cashew nut (*Anacardium occidentale* L.) kernels: Effect of different shelling methods. *Journal of Agricultural and Food Chemistry, 58*(9), 5341–5346. https://doi.org/10.1021/jf904580k

Vatanen, T., Osmala, M., Raiko, T., Lagus, K., Sysi-Aho, M., Orešič, M., Honkela, T., & Lähdesmäki, H. (2015). Self-organization and missing values in SOM and GTM. *Neurocomputing, 147*(1), 60–70. https://doi.org/10.1016/j.neucom.2014.02.061

Viera-Alcaide, I., Hamdi, A., Jiménez-Araujo, A., Rodríguez-Arcos, R., Espejo-Calvo, J. A., & Guillén-Bejarano, R. (2019). Nutritional composition and antioxidant activity of different walnut varieties (*Juglans regia* L.) from Nerpio (Spain) in comparison to commercial varieties. *Grasas Y Aceites, 70*(3), 310. https://doi.org/10.3989/gya.0932182

Waterhouse, A. L. (2002). Determination of total phenolics. In R. E. Wrolstad (Ed.), *Current protocols in food analytical chemistry*. John Wiley & Sons, Inc.

Wijeratne, S. S. K., Amarowicz, R., & Shahidi, F. (2006). Antioxidant activity of almonds and their by-products in food model systems. *Journal of the American Oil Chemists' Society, 83*(3), 223. https://doi.org/10.1007/s11746-006-1197-8