CrossMark

Using Intact Nuts and Near Infrared Spectroscopy to Classify Macadamia Cultivars

Lívia C. Carvalho[1] · Camilo L. M. Morais[2] · Kássio M. G. Lima[2] · Gustavo W. P. Leite[3] ·
Gabriele S. Oliveira[3] · Izabela P. Casagrande[3] · João P. Santos Neto[3] ·
Gustavo H. A. Teixeira[3]

Abstract Macadamia nut industry is increasingly gaining more space in the food market and the success of the industry and the quality are largely due to the selection of cultivars through macadamia nut breeding programs. Thus, the objective of this study was to investigate the feasibility NIRS coupled to chemometric classification methods, to build a rapid and non-invasive analytical procedure to classify different macadamia cultivars based on intact nuts. Intact nuts of five different macadamia cultivars (HAES 246, IAC 4-20, IAC 2-23, IAC 5-10, and IAC 8-17) were harvested in 2017. Two NIR reflectance spectra were collected per nut, and the mean spectra were used to chemometrics analysis. Principal component analysis-linear discriminant analysis (PCA-LDA) and genetic algorithm-linear discriminant analysis (GA-LDA) were used to develop the classifications models. The GA-LDA approach resulted in accuracy higher than 94.44%, with spectra preprocessed with Savitzky-Golay smoothing. Thus, this approach can be implemented in the macadamia industry, allowing the selection of cultivars based on intact nuts. However, it is recommended that more experimentation to include more data variability in order to increase the classification accuracy to 100%.

Introduction

Macadamia nut industry is increasingly gaining more space in the food market, particularly in Australia, South Africa, and Hawaii, which together account for more than 70% of world macadamia nut production (Piza and Moriya 2014). In Brazil, according to Piza and Moriya (2014), this market is represented by three main exporting producers, which in 2012 were responsible for the processing of 79% of the Brazilian crop, which is more than 700 tons of the 1100 produced in the same year.

The success of the industry and the quality of macadamia are largely due to the selection of cultivars through macadamia nut breeding programs, whose goal has been to develop high yielding cultivars that are adapted to specific locations and produce top-quality kernels (Wall 2013). Cultivars can be selected based on growth, resistance to diseases and insects, kernel mass and uniformity, bark thickness, percentage of kernel recovery, and percentage of kernels (Lee 1996).

Each main producer country has developed cultivars adapted to their agroecosystem and many cultivars were developed by the Hawaii Agricultural Experiment Station (HAES) and/or using such cultivars in breeding programs (Sobierajski et al. 2008). In Brazil, the Agronomic Institute of Campinas (IAC) has developed 17 cultivars of macadamia nuts adapted to Brazilian climatic conditions, such as weather, soil type, density, topography, and orchard management (Dierberger and Netto 1985). However, the Hawaiian varieties

✉ Gustavo H. A. Teixeira
   gustavo@fcav.unesp.br

[1] Faculdade de Ciências Farmacêuticas (FCFAR), Campus de Araraquara, Departamento de Alimentos e Nutrição, Universidade Estadual Paulista (UNESP), Rodovia Araraquara-Jaú, km 1 – CP 502, Araraquara, São Paulo 14800-903, Brazil

[2] Instituto de Química, Química Biológica e Quimiometria, Universidade Federal do Rio Grande do Norte (UFRN), Avenida Senador Salgado Filho, no. 3000, Bairro de Lagoa Nova, Natal, Rio Grande do Norte, Brazil

[3] Faculdade de Ciências Agrárias e Veterinárias (FCAV), Campus de Jaboticabal, Departamento de Produção Vegetal, Universidade Estadual Paulista (UNESP), Via de Acesso Prof. Paulo Donato Castellane s/n, Jaboticabal, São Paulo 14884-900, Brazil

developed by HAES are the most cultivated in Brazil, accounting for 80% of the planted area (Sobierajski et al. 2008), and the main cultivars are 'Keauhou,' 'Keaau,' 'Kau,' and 'Mauka' (Barbosa et al. 2003).

Different macadamia cultivars can be identified based on the shape and size of the tree, size, width, presence of spines and tip type of the leaves, shell shape, and distinctive shell features (Vock and Bell 1999). However, Vock and Bell (1999) report that identifying cultivars is a common problem among macadamia nut producers and this is important because each cultivar has distinct characteristics that can influence the type of management, nutrition, pest management, diseases, and the type of crop. Cultivars can present a broad difference in terms of growth, disease and insect resistance, nut mass and uniformity, shell thickness, kernel mass, percent kernel recovery, percent whole kernels, processing characteristics, and kernel quality (Hamilton and Ito 1976). For example, 'Keauhou' (HAES 246) has a low incidence of gray kernel disease, and 'Makai' (HAES 800) and 'Ikaika' (HAES 333) have less damage from tropical nut borer and southern green stinkbug compared to other selections (Nishijima et al. 2007). Cultivars also exhibit differences in oil content, fatty acid profiles, sugar concentrations, antioxidant contents, and oxidative stability and identification of cultivars with superior oxidative stability is relevant to macadamia industry, because lipid oxidation imparts off flavors and aromas to nuts and compromises nutritional quality (Kaijser et al. 2000; Wall 2010).

Because it is not recommended to plant only one macadamia nut cultivar, because even when self-pollination of plants occurs, cross pollination between cultivars is more effective for the formation of fruit, increasing the number of first quality kernels, the percentage of recovery of the kernels, and the size and number of nuts (O'Hare et al. 2004). In this way, there is no planting of only one cultivar, but of different cultivars to favor cross-pollination. Because of this, the identification of the cultivars becomes important because several cultivars are harvested together, leading to a mixture of physical and chemical properties of the nuts, consequently the quality of the macadamia nuts arriving at the processor can be highly variable and this may influence the formula by which payment is determined.

In this regard, non-destructive methods as quick tools to classify macadamia nuts according to cultivars can be useful, since it can help the macadamia nut industry to identify the cultivars, helping producers to choose good harvest practices and high-quality nuts. Among the non-destructive techniques, near infrared spectroscopy (NIRS) has presented great potential for evaluating the quality of various food products, since it is a non-destructive technique, fast, safe operation, relatively inexpensive, and provides process automation quality control in agribusiness products (Pasquini 2003).

The NIRS has been successfully applied to develop models for macadamia nut classification. Canneddu et al. (2016) classified macadamia nuts in marketable and non-marketable models constructed using partial least squares-discriminant analysis (PLS-DA) reporting 93.2% correctly classified macadamia nuts. Guthrie et al. (2004) also discriminated macadamia nuts according to their defects, such as rancidity, brown centers, discoloration, mold growth, germination, and decomposition, using principal components analysis (PCA). In other types of nuts, the NIRS tool was also effective, as in pine nuts, that the classification in different geographical origins using discriminant partial least square (DPLS) obtained an accuracy of 87.8% (Lowe et al. 2017). In pistachio nuts, NIRS was used for origin authentication and the accuracy results were higher than 90% with the soft independent modeling of class analogy (SIMCA) and PLS-DA techniques (Vitale et al. 2013).

Currently, there is no information regarding the development of classification models using NIRS and intact macadamia nuts to discriminate macadamia cultivars associated with chemometric techniques. Thus, the objective of this study was to investigate the feasibility NIRS coupled to chemometric classification methods, such as principal component analysis-linear discriminant analysis (PCA-LDA) and genetic algorithm-linear discriminant analysis (GA-LDA) to build a rapid and non-invasive analytical procedure to classify different macadamia cultivars based on intact nuts.

Material and Methods

Plant Material

Intact nuts of five different macadamia cultivars (*Macadamia integrifolia* Maiden and Betche) were harvested in 2017, as such: 'HAES 246' or 'Keauhou' ($n = 160$), 'IAC 4-20' or 'Keaumi' ($n = 160$), 'IAC 2-23' or 'Keaudo' ($n = 80$), 'IAC 5-10' or 'Kakere' ($n = 120$), and 'IAC 8-17' or 'Waiaré' ($n = 160$). The nuts were collected on the tree in Jaboticabal, São Paulo State, Brazil (latitude 21° 15′ S, longitude 48° 18′ W). Macadamia were dehusked within a 24-h period and each intact nut was weighted to determine the fresh weight and oven dried for 48 h as recommended by O'Hare et al. (2004), and moisture and dry matter were determined.

Nut Color

The color measurement was individually performed at two sites of each macadamia nut using a Minolta CR-400 colorimeter (Minolta Corp., Japan), which measures the color according to the CIE system (L*, a*, b*) according to McGuire (1992). Where L* corresponds to the luminosity which ranges from black = 0 to white = 100, the positive a* indicates red

and negative a* green. A positive b* indicates yellow and negative b* blue (McGuire 1992).

NIR Spectra Acquistion

For each macadamia nut, two NIR reflectance spectra (11,544–3952 cm^{-1}, nm, resolution of 16 cm^{-1}, and 64 scans) were collected using a Bruker NIR spectrometer (Tango, Ettlingen, Germany) after temperature stabilization at ~ 25 °C. The NIR spectra were averaged, and the mean spectra were used to chemometrics analysis.

Chemometrics

All NIR spectra were processed using MATLAB® software R2012b (MathWorks, USA) with PLS Toolbox version 7.9.3 (Eigenvector Research, Inc., USA). Prior to spectral pretreatment, all NIR spectra were analyzed to identify and eliminate defective spectral outliers, based on Hotelling T^2 and Q-residuals test (Bro and Smilde 2014).

For the construction of the models, the data were pretreated by the Savitzky-Golay smoothing (15-point window), multiplicative dispersion correction (MSC), and the first and second derivative Savitzky-Golay. Savitzky-Golay smoothing removes noises from the spectra and maintains its original shape; MSC removes effects of light scattering caused by non-homogenous particle size; and the first and second Savitzky-Golay derivatives correct baseline as well as improve the resolution of the spectral bands, while the Savitzky-Golay is applied to remove noises before differentiation (Hibbert 2016; Li et al. 2011; Susi and Byler 1983).

The classic Kennard-Stone (KS) uniform sampling algorithm (Kennard and Stone 1969) was adopted to divide the available samples into training, validation and test sets for construction, and validation of the PCA-LDA and GA-LDA models (Table 1). PCA-LDA and GA-LDA algorithms were applied using lab-made codes within a MATLAB® environment.

The training set was used for model construction and the validation set for internal model optimization; and the test set was used to evaluate the final classification performance of the model towards external samples.

Principal Component Analysis-Linear Discriminant Analysis

Principal component analysis-linear discriminant analysis (PCA-LDA) is based into a principal component analysis (PCA) reduction followed by a linear discriminant analysis (LDA). PCA is an exploratory analysis technique that reduces the original data into a few number of orthogonal variables that explain most of the original information.

Table 1 NIR spectra division into training, validation, and test set

Levels	Training	Validation	Test
IAC 4-20	111	24	24
IAC 5-10	84	18	18
ICA 8-17	112	24	24
'Keaudo'	55	12	12
'Keauhou'	112	24	24
Total	474	102	102

These variables, called principal components (PCs), are arranged in order that the first PC explains the most of the original variance, followed by the second PC, and so on. Each PC is a linear combination of a score (projections of the samples on the PC direction) and a loading (angle cosines of the variables projected on the PC direction) (Bro and Smilde 2014; Geladi and Kowalski 1986), where the scores are utilized as discriminant variables for LDA. LDA is a supervised classification technique that maximizes the between-class (level) variance over the within-class (level) variance (Kelly et al. 2011), in order to create a linear decision boundary between them (Dixon and Brereton 2009).

Genetic Algorithm with Linear Discriminant Analysis

Genetic algorithm with linear discriminant analysis (GA-LDA) uses a genetic algorithm (GA) as a variable selection technique followed by LDA. GA reduces the original data into a few number of variables according to an evolutionary process (McCall 2005). These variables are in the same scale as the original data and are selected according to the lowest risk of miss classification G. The G is calculated in the validation set as (Baia et al. 2016):

$$G = \frac{1}{N_V} \sum_{n=1}^{N_V} g_n, \tag{1}$$

where N_V is the number of validation samples; and g_n is defined as:

$$g_n = \frac{r^2\left(x_n, m_{I(n)}\right)}{\min_{I(m) \neq I(n)} r^2\left(x_n, m_{I(m)}\right)} \tag{2}$$

where the numerator is the squared Mahalanobis distance between object x_n (of class and/or level index I_n) and the sample mean $m_{I(n)}$ of its true class (level); whereas the denominator is the squared Mahalanobis distance between object x_n and the mean $m_{I(m)}$ of the closest wrong class (level). GA was performed through 80 generations, having 160 chromosomes each. Crossover and mutation probability were set to 60 and 10%, respectively. The algorithm was repeated three times, and the best result was chosen.

Figures of Merit

The classification performance of each model was evaluated using the correct classification rate, sensitivity, and specificity. The accuracy represents the number of samples correctly classified for each data set; and sensitivity and specificity measure the proportion of positives and negatives that are correctly classified. These figures of merit are calculated as follows (BAIA et al. 2016; CARVALHO et al. 2016):

$$Accuracy(\%) = 100 - \left(\frac{1}{N}\sum_{h=1}^{H}y_h^*\right) \times 100 \qquad (3)$$

$$Sensitivity\ (\%) = \frac{TP}{TP + FN} \times 100 \qquad (4)$$

$$Specificity\ (\%) = \frac{TN}{TN + FP} \times 100 \qquad (5)$$

where N is the total number of samples; H is the total number of classes (levels); y_h^* is the number of samples incorrectly classified in the h class (level); TP is the number of true positives; TN is the number of true negatives; FP is the number of false positives; and FN is the number of false negatives.

Results and Discussion

Quality Parameters

Macadamia nuts from all cultivars did not present significant difference ($p < 0.05$) for fresh weight (Table 2). The nut weight ranged from 6.33 g (IAC 2–23) to 5.16 g (5–10). Penoni et al. (2011) also did not observe differences in weight of macadamia nuts from the cultivars 'IAC 4-20' and 'Keauhou'. On the other hand, Dierberger and Netto (1985), evaluating the main characteristics of nuts, such as diameter, nut and kernel weight, percentage of kernel, and bark hardness, reported higher nut weight values for 'IAC 4-20' (8.5 g), 'IAC 5-10' (9.5 g), 'IAC 8-17' (6.5 g), and 'IAC 2-23' (8.0 g). This difference can be related to the variation of the cultivars in relation to environmental conditions, such as climate, soil, different management conditions that directly affect the quality of macadamia nuts (Rebouças 1991).

Significant differences in moisture (average of 7.70%) and dry matter (DM) content (average 92.29%) between nuts from all cultivars (Table 2) were not observed. These values were in accordance to the quality parameter stated by the Australian Macadamia Industry Code of Sound Orchard Practices (O'Hare et al. 2004) and United Nations Economic Commission for Europe (UNECE STANDARD 2010) that recommend that macadamia nuts must have 8–10% moisture and 90% DM.

Regarding macadamia nut color, the brown color that characterizes the macadamia nut shell is produced by the combination of red (a*—positive value) and yellow (b*—positive value) CIE parameter from which the hue angle is calculated (Birch et al. 2009). The a* values of the macadamia nuts varied according to the cultivars with higher and positive a* values on the shell of 'IAC 5-10' (10.46) than those of 'IAC 4-20' (9.64) cultivars, indicating higher red color intensity (Table 2). Significant difference in the L* was not observed, indicating that there is little luminosity differences between macadamia nut cultivars.

NIR Spectra

The pre-treated NIR spectra of all macadamia nuts from the five cultivars can be seen in Fig. 1. Overall, there was no difference of NIR spectra according to macadamia cultivar and the observed peaks were related to the presence of the main components of macadamia nut shell, which are lignin (47.6%), cellulose (25.8%), and hemicellulose (11.7%) (Toles et al. 1998). For cellulose and hemicellulose, the most important spectral variables were observed at a broadband between 6800 and 6400 cm^{-1}, related to the first OH stretching overtone and the peak between 5000 and 4500 cm^{-1} corresponds to the first CH, CH$_2$ overtone region, and CH, CH$_2$, CH$_3$ combination band region (Guimarães et al. 2014; Canneddu et al. 2016). For lignin, the peaks are at about 5230 cm^{-1}, first O-H stretching overtone of aromatics, 4415 cm^{-1}, O-H combination bands, and C-O stretches (Guimarães et al. 2014).

To carry out the discriminant analysis, the NIR spectra were subjected to different pre-treatments, such as Savitzky-Golay (SG) smoothing (Fig. 1a), SG plus MSC (Fig. 1b), first derivative Savitzky-Golay (Fig. 1c), and second derivative Savitzky-Golay (Fig. 1d).

Classification of Macadamia Cultivars

To discriminate between the five different levels of macadamia cultivars, PLS-DA and GA-LDA were carried out. The best PCA-LDA result was achieved by using 10 principal components (PCs) and NIR spectra pre-treated with the second derivative Savitzky-Golay, which accounted for 98.1% of total variance (Fig. 2). However, the results were not satisfactory, 58.33% accuracy for the cultivar 'Keaudo', even using pre-processing (Table 3). A clear separation between macadamia cultivars with overlapping among samples (Fig. 2) was not observed. This might be related to the similarities between cultivars and influenced by climatic factors or that they have the same pollination, that is, the cultivars can be of the same crossing, since they were developed in the same place (Sobierajski 2012).

Regarding the GA-LDA, the discriminant models were developed selecting 27 wavenumbers (Tables 4 and 5) and NIR spectra pre-treated with SG smoothing (a), SG plus multiplicative scatter correction (b), SG plus first derivative Savitzky-

Table 2 Average nut mass, humidity, dry matter, color of five different macadamia cultivars

Effects	Mass (g)	Moisture (%)	DM (%)	L	a*	b*
Cultivars (C)						
IAC 4-20	6.23	8.21	91.79	42.19	9.64[b]	13.14
IAC 5-10	5.16	7.58	92.42	40.92	10.46[a]	13.10
IAC 8-17	5.28	7.38	92.62	40.38	10.24[a,b]	12.87
Keaudo	6.33	7.77	92.23	42.54	9.75[a,b]	13.78
Keauhou	6.13	7.57	92.43	41.83	9.72[a,b]	13.34
F test	2.19	0.92	0.92	3.00	4.48*	1.00
Day (D)						
1	5.99[a]	12.59[a]	87.42[c]	41.26[b,c]	10.12[a]	13.15[a,b,c]
2	5.89[a,b]	12.75[a]	87.25[c]	40.71[c]	10.02[a]	12.84[c]
3	5.91[a]	7.46[b]	92.54[b]	41.85[a,b]	10.07[a]	13.38[a,b]
4	5.99[a]	6.21[b,c]	93.79[a,b]	41.42[a,b,c]	10.02[a]	13,18[a,b]
5	6.06[a]	4.74[c]	95.26[a]	42.10[a.]	9.90[a,b]	13.46[a]
6	5.78[a,b]	6.91[b,c]	93.09[a,b]	41.36[a,b,c]	9.88[a,b]	13.14[a,b]
7	5.50[a,b]	5.81[b,c]	94.19[a,b]	41.56[a,b]	9.67[b]	13.04[c]
8	5.35[b]	5.13[b,c]	94.87[a,b]	41.70[a,b]	10.03[a]	13.36[a,b]
F test	2.97*	28.85*	29.61*	5.12*	5.45*	6.85*
Interactions						
C × D	*	NS	NS	NS	NS	NS

Averages followed by the same superscripted letter in the column are significantly different according to Tukey's test ($p < 0.05$)

DM dry matter, *NS* not significant

* 0,022

Golay (c), and SG plus second derivative Savitzky-Golay (d) (Fig. 3).

The segregation between macadamia cultivar improved when compared to that of PCA-LDA (Table 3), reaching

Fig. 1 NIR spectra with Savitzky-Golay (SG) smoothing (**a**); SG plus MSC (**b**); SG plus first derivative Savitzky-Golay (**c**); and SG plus second derivative Savitzky-Golay (**d**)
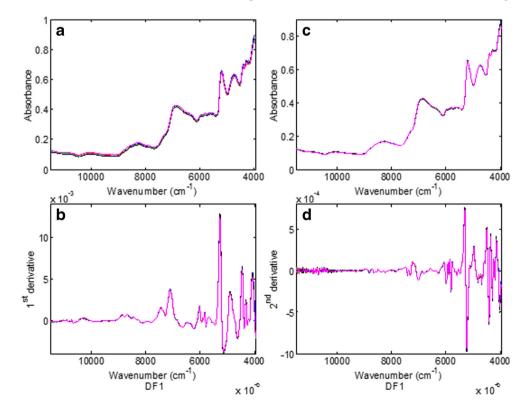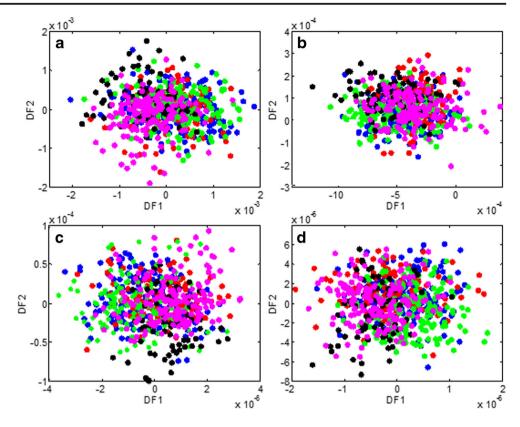
Fig. 2 DF plot of PCA-LDA with Savitzky-Golay (SG) smoothing (**a**); SG plus MSC (**b**); SG plus first derivative Savitzky-Golay (**c**); and SG plus second derivative (**d**)

Table 3 Classification and prediction performance of PCA-LDA and GA-LDA on macadamia (*Macadamia integrifolia* Maiden and Betche) nuts according to five cultivars

Pre-processing	Cultivar	PCA-LDA (%)			GA-LDA (%)		
		Training	Validation	Test	Training	Validation	Test
SG	IAC 4-20	28.83	16.67	16.67	64.86	58.33	45.83
	IAC 5-10	23.81	38.89	22.22	73.81	83.33	94.44
	IAC 8-17	51.79	54.l7	50.00	73.21	83.33	83.33
	Keaudo	49.09	58.33	58.33	83.64	83.33	83.33
	Keauhou	44.64	16.67	29.17	58.04	50.00	62.50
SG + MSC	IAC 4-20	18.92	4.17	4.17	63.06	66.67	37.50
	IAC 5-10	34.52	44.44	44.44	75.00	100	83.33
	IAC 8-17	53.57	50.00	45.83	67.86	70.83	83.33
	Keaudo	56.36	66.67	58.33	85.45	100	91.67
	Keauhou	63.39	20.83	33.33	51.79	50.00	50.00
SG + 1D	IAC 4-20	42.34	29.17	41.67	65.77	66.67	50.00
	IAC 5-10	35.71	38.89	38.89	73.81	83.33	83.33
	IAC 8-17	36.61	12.50	25.00	69.64	79.17	79.17
	Keaudo	45.45	16.67	16.67	87.27	91.67	75.00
	Keauhou	58.93	29.17	33.33	64.29	50.00	58.33
SG + 2D	IAC 4-20	41.44	45.83	29.17	64.86	66.67	62.50
	IAC 5-10	55.95	44.44	38.89	69.05	83.33	77.78
	IAC 8-17	53.57	29.17	50.00	67.86	66.67	75.00
	Keaudo	40.00	25.00	16.67	83.64	91.67	75.00
	Keauhou	49.11	29.17	33.33	62.50	45.83	58.33

SG Savitzky-Golay smoothing, *MSC* multiplicative scatter correction, *1D* first derivative Savitzky-Golay, *2D* second derivative Savitzky-Golay

Table 4 Values of sensitivity and specificity from two classification methods (PCA-LDA, GA-LDA) by NIR spectroscopy of macadamia (*Macadamia integrifolia* Maiden and Betche) nuts using different pre-processing

Pre-processing	Cultivar	PCA-LDA (%)		GA-LDA (%)	
		Sensitivity	Specificity	Sensitivity	Specificity
SG	IAC 4-20	16.67	89.74	45.83	97.44
	IAC 5-10	22.22	89.29	94.44	90.48
	IAC 8-17	50.00	66.67	83.33	87.18
	Keaudo	58.33	77.78	83.33	96.67
	Keauhou	29.17	93.59	62.50	92.31
SG + MSC	IAC 4-20	4.17	94.87	37.50	93.59
	IAC 5-10	44.44	79.76	83.33	83.33
	IAC 8-17	45.83	76.92	83.33	89.74
	Keaudo	58.33	78.89	91.67	95.56
	Keauhou	33.33	88.46	50.00	94.87
SG + 1D	IAC 4-20	41.67	83.33	50.00	94.87
	IAC 5-10	38.89	90.48	83.33	91.67
	IAC 8-17	25.00	80.77	79.17	89.74
	Keaudo	16.67	71.11	75.00	94.44
	Keauhou	33.33	91.03	58.33	88.46
SG + 2D	IAC 4–20	29.17	83.33	62.50	94.87
	IAC 5-10	38.89	85.71	77.78	89.29
	IAC 8-17	50.00	83.33	75.00	89.74
	Keaudo	16.67	77.78	75.00	96.67
	Keauhou	33.33	89.74	58.33	89.74

SG Savitzky-Golay smoothing, *MSC* multiplicative scatter correction, *1D* first derivative Savitzky-Golay, *2D* second derivative Savitzky-Golay

94.44% accuracy in test samples for the cultivar 'IAC 5-10' (Fig. 3). The better performance of GA-LDA can be due to that GA algorithm selects several wavenumbers in a single band, due to collinearity problems and this fact may have contributed to the fact that the GA-LDA technique has discriminated better when compared to PCA-LDA.
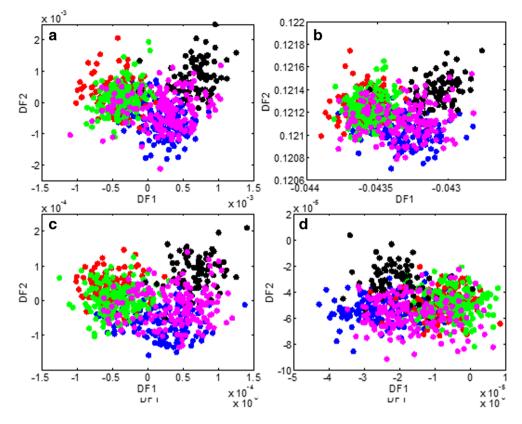
All in all, it was observed that the four developed models with the different pretreatment were a trend along the DF 1 axis (Fig. 3). The cultivars 'IAC 4-20,' 'Keauhou,' and 'Keaudo' tend to get together and that is because the cultivars 'IAC 4-20' and 'Keaudo' are progenies of the 'Keauhou' cultivar, which originated from the Hawaii Agricultural Experiment Station (HAES). On the other hand, the cultivars 'IAC 5-10' and 'IAC 8-17,' which are located in the first quadrant of DF 1, are 'Kakea' (HAES 508) and 'Waialua' (HAES 475) progeny, respectively (Sobierajski 2012). The 'Keauhou' and 'IAC 4-20' cultivars also present similarities in carpel and nut shell diameter (Penoni et al. 2011), and in the composition of oil extract, protein, and crude fiber (Maro et al. 2012), which might have contributed to the observed clustering among cultivars.

Table 5 Selected variables for GA-LDA to classify cultivars of macadamia (*Macadamia integrifolia* Maiden and Betche) nuts using different pre-processing

Pre-processing	Selected variables (cm^{-1})
SG	4320; 4368; 4712; 5472; 5588; 5708; 6052; 6076; 6296; 6696; 7032; 7232; 7940; 8204; 8620; 8648; 8852; 8944; 8948; 8964; 9132; 9300; 10,256; 10,524; 10,636; 10,880; 11,492
SG + MSC	4048; 4060; 4776; 4932; 5468; 5480; 6004; 6996; 7836; 7916; 7980; 9672; 9820; 9920; 10,712; 10,888; 11,268; 11,288; 11,320; 11,532
SG + 1D	4112; 4352; 4476; 4944; 5164; 5392; 5484; 5780; 6048; 6224; 6700; 6804; 6896; 7148; 7528; 8856; 9168; 9320; 9488; 9644; 10,192; 10,208; 10,952; 11,288
SG + 2D	4128; 4620; 4644; 5228; 5240; 5392; 5844; 5936; 6072; 6148; 7504; 7560; 7656; 9112; 9388; 10,100; 10,264; 10,604; 10,608; 10,952; 11,232; 11,472

SG Savitzky-Golay smoothing, *MSC* multiplicative scatter correction, *1D* first derivative Savitzky-Golay, *2D* second derivative Savitzky-Golay

Fig. 3 DF plot of GA-LDA with Savitzky-Golay (SG) smoothing (**a**); SG plus MSC (**b**); SG plus first derivative Savitzky-Golay (**c**); and SG plus second derivative Savitzky-Golay (**d**)

The classification models were then evaluated for sensitivity for each cultivar, which is the probability that the sample possessing the desired characteristic gives a positive test result, and specificity for each cultivar, which is the probability that the sample without the desired characteristic gives a negative test result (Amodio et al. 2017). From the results of sensitivity shown in Table 4, it is possible to verify that the GA-LDA classification models achieved a value of 94.44% for sensitivity for 'IAC 5-10' cultivar, showing that this model is able to correctly classify 'IAC 5-10' cultivar of macadamia nuts based on intact nuts.

All cultivars of macadamia nuts were correctly discriminated with higher 87% specificity, and these results demonstrate the percentage of objects foreign to the category classified as foreign.

A number of literature studies have investigated different methodologies for the rapid determination of food classification and are receiving growing interest in the scientific community and in food industry, especially Vitale et al. (2013), which constructed classification models according to the origin of pistachio, using techniques such as SIMCA and PLS-DA. In this approach, it was possible to achieve a sensitivity of 90% using PLS-DA and specificity of 97% using SIMCA, demonstrating the applicability of the technique.

Concerning the classification of macadamia cultivars using NIR spectroscopy, from the results of sensitivity and specificity shown in Table 4, it is possible to verify that this

methodology can be used to classify different cultivars of macadamia nuts.

Conclusion

Near infrared spectroscopy (NIRS) and supervised pattern recognition techniques (PCA-LDA and GA-LDA) were effective to classify macadamia cultivars based on intact nuts. Using the GA-LDA technique with pre-treated spectra was possible to reach a value of 94.44% accuracy of macadamia nut cultivars. However, it is recommended that more experimentation, to include more data variability in order to improve robustness and increase the classification accuracy to 100%.

Compliance with Ethical Standards

Ethical Approval   This article does not contain any studies with human participants or animals performed by any of the authors.

Informed Consent   This article does not contain any studies with human participants performed by any of the authors.

References

Amodio ML, Ceglie F, Chaudhry MM, Piazzolla F, Colelli G (2017) Potential of NIR spectroscopy for predicting internal quality and discriminating among strawberry fruits from different production systems. Postharvest Biol Technol 125:112–121. https://doi.org/10.1016/j.postharvbio.2016.11.013

Baia TC, Gama RA, Lima LS, Lima KMG (2016) FTIR microspectroscopy coupled with variable selection methods for the identification of flunitrazepam in necrophagous flies. Anal Methods 8:968–972. https://doi.org/10.1039/C5AY02342D

Barbosa W, Pommer CV, Ribeiro MD, Veiga RFA, Costa AA (2003) Distribuição geográfica e diversidade varietal de frutíferas e nozes de clima temperado no Estado de São Paulo. Rev Bras Frutic 25: 341–344. https://doi.org/10.1590/S0100-29452003000200042

Birch EJ, Yap K, Silcook P (2009) Compositional analysis and roasting behavior of gevuina and macadamia nuts. J Food Sci 45:81–86. https://doi.org/10.1111/j.1365-2621.2009.02106.x

Bro R, Smilde AK (2014) Principal component analysis. Anal Methods 6: 2812–2831. https://doi.org/10.1039/C3AY41907J

Canneddu G, Cunha Júnior LC, Teixeira GHA (2016) Quality evaluation of shelled and unshelled macadamia nuts by means of near-infrared spectroscopy (NIR). J Food Sci 81:1613–1621. https://doi.org/10.1111/1750-3841.13343

Carvalho LC, Morais CL, Lima KMG, Cunha Júnior LC, Nascimento PA, Faria JB, Teixeira GHA (2016) Determination of the geographical origin and ethanol content of Brazilian sugarcane spirit using near-infrared spectroscopy coupled with discriminant analysis. Anal Methods 8:5658–5666. https://doi.org/10.1039/C6AY01325B

Dierberger JE, Netto LM (1985) Noz macadâmia: uma nova opção para a fruticultura brasileira. Nobel, São Paulo

Dixon SJ, Brereton RG (2009) Comparison of performance of five common classifiers represented as boundary methods: euclidean distance to centroids, linear discriminant analysis, quadratic discriminant analysis, learning vector quantization and support vector machines, as dependent on data structure. Chemometr Intell Lab Syst 95:1–17. https://doi.org/10.1016/j.chemolab.2008.07.010

Geladi P, Kowalski BR (1986) Partial least-squares regression: a tutorial. Anal Chim Acta 185:1–17. https://doi.org/10.1016/0003-2670(86)80028-9

Guimarães CC, Simeone ML, Parrella RAC, Sena MM (2014) Use of NIRS to predict composition and bioethanol yield from cell wall structural components of sweet sorghum biomass. Microchem J 117:194–201. https://doi.org/10.1016/j.microc.2014.06.029

Guthrie J, Greensill C, Bowden R, Walsh K (2004) Assessment of quality defects in macadamia kernels using NIR spectroscopy. Aust J Agric Res 55:471–476. https://doi.org/10.1071/AR03179

Hamilton RA, Ito PJ (1976) Development of macadamia nut cultivars in Hawaii. California Macadamia Society Yearbook 22:94–100

Hibbert DB (2016) Vocabulary of concepts and terms in chemometrics (IUPAC recommendations). Pure Appl Chem 88:407. https://doi.org/10.1515/pac-2015-0605

Kaijser A, Dutta P, Savage G (2000) Oxidative stability and lipid composition of macadamia nuts grown in New Zealand. Food Chem 71: 67–70

Kelly JG, Trevisan J, Scott AD, Carmichael PL, Pollock HM, Martin-Hirsch PL, Martin FL (2011) Biospectroscopy to metabolically profile biomolecular structure: a multistage approach linking computational analysis with biomarkers. J Proteome Res 10:1437–1448. https://doi.org/10.1021/pr101067

Kennard RW, Stone LA (1969) Computer aided design of experiments. Technometrics 11:137–148. https://doi.org/10.2307/1266770

Lee PFW (1996) The horticultural challenges for macadamia research. Challenges for horticulture in the tropics: Proceedings of the Third Australian Society of Horticultural Science and the First Australian Macadamia Society Research Conference, p. 258–253

Li N, Li XY, Zou ZX, Lin LR, Li YQ (2011) A novel baseline-correction method for standard addition based derivative spectra and its application to quantitative analysis of benzo(a)pyrene in vegetable oil samples. Analyst 136:2802–2810. https://doi.org/10.1039/c0an00751j

Lowe V, Cerrilo RMN, Garcia-Olmo J, Riccioli C, Sanchez-Cuesta R (2017) Discriminant analysis of Mediterranean pine nuts (*Pinus pinea* L.) from Chilean plantations by near infrared spectroscopy (NIRS). Food Control 73:634–643. https://doi.org/10.1016/j.foodcont.2016.09.012

Maro LAC, Pio R, Penoni ES, Oliveira MC, Prates FC, Lima LCO, Cardoso MG (2012) Caracterização química e perfil de ácidos graxos em cultivares de nogueira- macadâmia. Ciênc Rural 42: 2166–2171. https://doi.org/10.1590/S0103-4782012005000117

McCall J (2005) Genetic algorithms for modelling and optimization. Appl Math Comput – Special Issue: Math Appl Immunol 184: 205–222. https://doi.org/10.1016/j.cam.2004.07.034

McGuire RG (1992) Reporting of objective color measurements. Hortscience 27:1254–1255

Nishijima KA, Wall MM, Siderhurst MS (2007) Demonstrating pathogenicity of on Macadamia and identifying associated volatiles of gray kernel of macadamia in Hawaii. Plant Dis 91(10):1221–1228

O'Hare P, Hidden DV, Burton D, Salmon T (2004) Australia macadamia industry code of sound orchard practices. Australian Macadamia Society, Lismore

Pasquini C (2003) Near infrared spectroscopy: fundamentals, practical aspects and analytical applications. J Braz Chem Soc 14:198–219

Penoni ES, Pio R, Rodrigues FA, Maro LAC, Costa FC (2011) Análise de frutos e nozes de cultivares de nogueira-macadâmia. Ciênc Rural 41: 2080–2083. https://doi.org/10.1590/S0103-84782011001200007

Piza PLBT, Moriya LM (2014) Cultivo da macadâmia no Brasil. Rev Bras Frutic 36:39–45. https://doi.org/10.1590/S0103-50532003000200006

Rebouças A (1991) Macadâmia: tecnologia de produção e comercialização. DFZ/UESB, São Paulo

Sobierajski GR (2012) Desenvolvimento e uso de marcadores SSR e DArT para estudos de diversidade genética em macadâmia (*Macadamia integrifolia*). Dissertation, Universidade São Paulo, Piracicaba

Sobierajski GR, Jungmann L, Souza AP, Garcia AFF (2008) Desenvolvimento de marcadores microssatélites para *Macadamia integrifolia* a partir de bibliotecas enriquecidas. In: 54° Congresso Brasileiro de Genética. Anais. Salvador: SBG. 1 CD-ROM. 20p

Susi H, Byler DM (1983) Protein structure by Fourier transform infrared spectroscopy: Second derivative spectra. Biochem Biophys Res Commun 115(1):391–397

Toles C, Marshall W, Johns M (1998) Phosphoric acid activation of nutshells for metals and organic remediation: process optimization. J Chem Technol Biotechnol 72:255–263. https://doi.org/10.1002/(SICI)1097-4660(199807)72:3<255::AID-JCTB890>3.0.CO;2-P

UNECE STANDARD (2010) DDP-23 concerning the marketing and commercial quality control of macadamia kernels. United Nations, New York and Geneva

Vitale R, Bevilacqua M, Bucci R, Magri AD, Magri AL, Ma F (2013) A rapid and non-invasive method for authenticating the origin of pistachio samples by NIR spectroscopy and chemometrics. Chemom Intell Lab Sys 121:90–99. https://doi.org/10.1016/j.chemolab.2012.11.019

Vock B, Bell D (1999) Macadamia variety identifier. first ed Agrilink, Queensland

Wall MM (2010) Functional lipid characteristics, oxidative stability, and antioxidant activity of macadamia nut (*Macadamia integrifolia*) cultivars. Food Chem 121:1103–1108

Wall MM (2013) Improving the quality and safety of macadamia nuts. In: Harris LJ (Ed) Improving the quality and safety of nuts Cambridge: Woodhead Publishing Series in Food Sci. Technol Nutr. 250:274–296