RESEARCH ARTICLE

Chia (Salvia hispanica) seeds degradation studied by fuzzy-c mean (FCM) and hyperspectral imaging and chemometrics - fatty acids quantification

J. P. Cruz-Tirado, Pedro Renann Lopes de França, Douglas Fernandes Barbin

1 Department of Food Engineering and Technology, School of Food Engineering, University of Campinas (UNICAMP), Monteiro Lobato St. 80, 13083-862, Campinas, São Paulo, Brazil.

* Corresponding author: dfbarbin@unicamp.br (D. F. Barbin).

Received: 23 March 2022. Accepted: 18 June 2022. Published: 4 July 2022.

Abstract

Chia seeds are nutritious food because they have a high content of protein, polyunsaturated fatty acids (omega-3 and omega-6) and phenolic compounds. During storage, fatty acids are degraded, by oxidative and hydrolytic reactions, forming free fatty acids (FFA). In this work, we used Near Infrared Hyperspectral Imaging (NIR- HSI) and chemometrics to predict FFA acid value and fatty acids concentrations in chia seeds during storage. First, we explore the hyperspectral images by Fuzzy c-means (FCM), where it is possible to observe as chemical compounds are formed or degraded during storage. Second, PLSR models were developed to predict FFA value and fatty acids concentration. RPD values reached values higher then 2.0, indicating a goodability to estimate these chemical compounds, especially polyunsaturated fatty acids omega-3 and omega-6. Finally, NIR-hyperspectral imaging coupled with chemometrics allowed us to show the chemical degradation process of chia seeds during storage, mainly associated with polyunsaturated fatty acids degradation. Besides NIR- HSI showed to be a powerful technique to quantify the main fatty acids with high accuracy.

Keywords: polyunsaturated fatty acids; machine learning; fuzzy c-means; oleaginous seeds.

DOI: https://doi.org/10.17268/sci.agropecu.2022.015

Cite this article:


1. Introduction

Chia (Salvia hispanica) seeds are valuable super food, because seeds have a high content of polyunsaturated fatty acids (PUFAs) (mainly α-linolenic acid (ω-3), 59.9 – 63.2%) (Oliveira-Alves et al., 2017; Salvatieri-Pajuelo et al., 2019), protein (19.0 – 26.5%) (Grancieri et al., 2019), dietary fibre (47.1 – 59.8%), phenolic compounds (mainly caffeic, rosmarinic and gallic acids) (De Falco et al., 2017; Oliveira-Alves et al., 2017) and mucilage (Muñoz et al., 2012). Consequently, chia seeds crop, originating in southern Mexico and northern Guatemala, has spread to countries such as Peru, Colombia, Argentina, Ecuador, Bolivia, Paraguay and Australia (Busilacchi et al., 2013). Fatty acid profile of chia seeds is a routine analysis to determine the quality of seeds and to drive seeds for specific applications in food and pharmaceuticals (Giaretta et al., 2018; Knez Hrnčič et al., 2019). For example, it is necessary to quantify the amount of omega-3 and omega-6 present in chia seeds, because its positive effect on human body by lowering the content of triglycerides and cholesterol levels, have anti-inflammatory, cardioprotective and hepatoprotective activity, antidiabetic action, and offers protection against cancer, arthritis and autoimmune diseases (de Falco et al., 2018; Grancieri et al., 2019; Mohd Ali et al., 2012; Ullah et al., 2016). Traditionally, gas chromatography (GC) alone or coupled to mass spectrometry (GC- MS) is used to quantify fatty acids in oils. Lipids are extracted from seeds by Bligh - Dyer method using methanol/chloroform mixture (Hartman, 1973). Next, lipids are esterified, added with internal standard, then injected into a GC or GC-MS (Joseph & Ackman, 1992). In addition, the free fatty acids (FFA) value is a quality parameter that allows for determining if samples were submitted to abusive storage conditions (i.e. high temperatures), which could cause nutritional and energy losses (Mata et al., 2017). For quantifying FFA in oils, samples are titrated by KOH (AOCS, 2003). Despite the high precision of GC to quantify fatty acids and the AOCS method to access FFA value, these methods are expensive and/or generate polluting chemical residues.
Currently, the food industry is modernizing, and implementing spectroscopy and imaging systems for better control of its processes and products, moving towards the well-known Industry 4.0. In this context, hyperspectral imaging (HSI) shows up with a powerful imaging technique for on-line, in-line and at-line food quality control (Caballero et al., 2020). Near-infrared hyperspectral imaging (NIR-HSI) allows simultaneously collected spectral (related to chemical composition) and spatial (related to physical properties) samples (Cruz-Tirado et al., 2020; Mendez et al., 2019). Coupled with chemometrics, NIR-HSI can be used to develop predictive models to authenticate seeds (Choi et al., 2021; Yu et al., 2021) and to estimate seed composition (da Silva Medeiros et al., 2022; Wang et al., 2021). In our previous work, we used NIR-HSI to estimate kinetic parameters and shelf-life of chia seeds by applying Multivariate Accelerated Shelf Life Testing (MASLT) (Cruz-Tirado et al., 2021). Results showed that NIR-HSI collected relevant spectral information about chemical compounds’ degradation/transformation over time under controlled storage temperature. For this reason, it is possible to estimate that NIR-HSI could be used to develop predictive models to estimate fatty acid concentrations and FFA value in chia seeds. Thus, this work aimed to develop prediction models to quantify FFA value and fatty acids concentrations directly in chia seeds by hyperspectral imaging coupled with chemometrics. This work is a complement to our previous work, so the FFA and fatty acid profile results can be accessed in Cruz-Tirado et al. (2021), while here we basically present the predictive quantification models for FFA and fatty acids.

2. Materials and methods

2.1. Samples preparation

Chia (Salvia hispanica) seeds were donated by R&S BLUMOS Industrial e Comercial Ltda (São Paulo, Brazil) in 2019. The proximal composition declared by the producer was 31 g lipids (3.4 g saturated lipids, 19.7 g afa-linolenic acid and 5.4 g linolenic acid), 21 g proteins, 38 g dietary fiber, < 1 g carbohydrates and 9 g moisture100 g of seeds. 250 g of chia seeds were stored at 25, 35 and 45 °C for 180 days in a BOD TECNAL TE-371 climatic chamber, without light incidence on the samples (Cruz-Tirado et al., 2021).

2.2. Fatty acid profile and FFA analysis

Approximately 10 g of chia seeds stored at days 0, 30, 60, 90 and 180 days were collected for fatty acid profile and FFA analysis. Chia seed lipids were extracted by Bligh-Dyer method (Hartman, 1973). Then, lipids were esterified as reported by Joseph & Ackman (1992), and the chromatographic conditions were based on Ballus et al. (2014). The separation of methyl esters was performed on a 7890A gas chromatograph (GC-Agilent, Germany) equipped with a flame ionization detector (FID). Correction factors and fatty acid concentration (mg/g oil) were calculated according to Joseph & Ackman (1992). For more details on equipment and methodology, the readers can access Cruz-Tirado et al. (2021). All samples were analyzed in triplicate.

2.3. Hyperspectral imaging acquisition

Chia seeds samples collected at 0, 30, 60, 90 and 180 days of storage were dispersed on Teflon plate (~5 mm thickness) (Figure 1). Then, hyperspectral images were acquired using a SiuCHEMA SWIR hyperspectral camera (Specim Spectral Imaging Ltd, FIN-90571 Oulu, Finland), at 6.23 nm intervals in the NIR range 928–2524 nm (256 waves) with a spatial resolution of 320 pixels per line scan and spectral resolution of 10 nm. Image correction was performed using two-dimensional reference images: the dark (0% reflectance) and the white (~99% reflectance). Then, the Evince software (UnBio AB, Sweden) automatically subtracted the white and dark references from subsequently acquired images (Cruz-Tirado et al., 2021). Six samples for each time storage at each storage temperature were scanned by hyperspectral imaging (90 images). For extracted mean spectra, a self-developed code in Matlab R2016a software (The Mathworks Inc., Natick, MA, USA) was used. Standard Normal Variate (SNV) was used to correct the base line drift and light scattering, producing one mean spectrum per replicate (one spectrum per image).

2.4. Data processing

2.4.1. Fuzzy c-means analysis (FCM)

The fuzzy c-means (FCM) algorithm is well-known centroid based clustering algorithm (Bezdek, 2013), which is a generalization of hard c-means (HCM) algorithm (Hartigan, 1975). Unlike HCM, which assigns each data (each individual pixel) to a single group, FCM can assign each pixel to different groups with a certain degree of belongings called membership. FCM defines an objective function, whose objective is to minimize the objective function by minimizing the closeness between the clusters, taking into account the distance to the cluster’s centroid and the similarity of the data (spectrum of each pixel) to define the degree of membership (Verma et al., 2021). For this reason, FCM is a powerful alternative to pattern recognition, machine learning, and data mining for image analysis. In this case, we applied FCM to chia seeds hyperspectral images to explore the degradation/transformation of chemical compounds in chia seeds during storage. Chia seeds hyperspectral images were analyzed by FCM algorithm was carried out in Hypertools v.3.0 free software (Mobarak & Amigo, 2018), available at www.hypertools.org, installed in MATLAB 2020a (Mathworks;Natick, USA) environment.

2.4.2. Partial Least Square Regression (PLSR)

PLSR was used to develop regression models to estimate FFA value and fatty acids concentrations in chia seeds hyperspectral images. The spectra of chia seeds samples were used as training set. Internal cross-validation was carried out using a leave-one-out methodology. PLSR models were developed in full spectral range (928 – 2524 nm) and in reduced spectral range based on important variables. For variable selection, interval partial least squares (iPLS) were used, which divides the spectra into variable intervals, then build
PLS model for each interval, selecting the best interval (Cruz-Tirado et al., 2021; Nørgaard et al., 2000). PLSR performance was evaluated by the coefficient of determination (R²), and root mean square error (RMSE) for calibration (RMSEC) and cross-validation (RMSECV). In addition, the ratio of the standard deviation of the prediction data set to RMSECV (RPD, Residual Prediction Deviation) was evaluated. Predictive ability can be evaluated by RPD values: when RPD < 1.5 the model is not suitable for prediction, 1.5 – 2.0 means that the model can distinguish between high and low values, 2.0 – 2.5 means a model with approximate prediction ability, 2.5 – 3.0 indicates a good model for prediction and RPD > 3 suggest an excellent predictive ability (Brasil et al., 2022; Saeys et al., 2005). PLSR analysis was carried out using the PLS Toolbox from Eigenvector Research, Inc. (Manson, WA, USA), installed in MATLAB 2020a (Mathworks, Natick, USA) environment.

2.4.3. Chemical maps
To visualize the distribution of the studied chemical components (fatty acids and FFA) using the PLSR models, as well as the chemical evolution of chia seeds during storage using FCM, distribution maps or chemical maps were created. The split images were used for pixel-by-pixel prediction using the FCM and PLSR models. Then, the hyperspectral images were reconstructed to show the distribution of the components. For the prediction of fatty acids and FFA, the mean of the prediction of each pixel contained in the image was calculated.

Figure 1. Experimental scheme. I) NIR-HSI image acquisition and data structure; II) chemometric analysis by Fuzzy-c mean and PLSR.
3. Results and discussion

3.1. FCM results

FCM analysis assigned one pixel for diverse clusters (Figure 2A) based on the similarity of spectra and the distance to the center of the cluster (Figure 2B). In Figure 1, we can see the surface for the three first clusters obtained from FCM analysis. One image for each time and temperature of storage was analyzed by FCM, and the results showed the variability in pixel-to-pixel influenced by storage conditions. It is possible to observe how some chemical components degrade or transform during storage. This mainly occurs when the storage temperature is 45 °C.

In Figure 1B, the center of cluster 1 shows a peak in the spectral region 1200 - 1250 nm, related to C-H stretching second overtone (Osborne, 2006), probably associated with the presence of lipid components. Surface scores for cluster 1 show how rapidly the seeds degrade while giving way to new chemical components (Figure 1A). The components that are formed are probably glycerol and free fatty acids, which are the result of oxidation and hydrolysis reactions of polyunsaturated fatty acids (omega-3 and omega-6) (1700 – 1800 nm) (Wang et al., 2012). In the center of clusters 2 and 3, the peak around 1000 nm, associated with O–H stretching second overtone (ArOH) (Osborne, 2006), is probably associated with the degradation of phenolic components in chia seeds (caffeic, rosmarinic and gallic acids) (de Falco et al., 2018; Oliveira-Alves et al., 2017), as well as tocopherols, carotenoids and chlorophyll consumed to counteracting the oxidative reactions motivated by temperature and time storage (Cruz-Tirado et al., 2021; Mannucci et al., 2019).

Finally, during storage, the transpiration process motivates the volatilization of free water and the availability of water for biological degradation/transformation processes. The high temperatures motivate the separation of the bound water, motivating greater mobility of the chemical components, favoring pro-oxidant reactions (Escalona-Garcia et al., 2016).

This is observed in cluster 1, where it is possible to observe how oxidizing chemical components appear rapidly at 45 °C (red pixels), while the water volatilizes (blue pixels). In addition, it is possible to observe that the pixels assigned to cluster 1 were not assigned to cluster 2, which indicates that each cluster is capturing relevant information associated with different chemical components (cluster 1 >> pro-oxidant components; cluster 2 >> phenolic components).

![Figure 2. FCM cluster analysis: A) surface for the three first FCM cluster for chia seedshyperspectral images; B) Centre for clusters.](image-url)
Figure 3. PLSR result and chemical maps for FFA value estimation.

Figure 4. PLSR results and chemical maps for A) palmitic acid, B) stearic acid, C) oleic acid, D) linoleic acid, E) α-linolenic acid.
### Table 1

Statistics of PLSR models to estimate FFA value and fatty acids concentrations

<table>
<thead>
<tr>
<th>Property</th>
<th>Model</th>
<th># Var</th>
<th>Lv</th>
<th>(R^2_C)</th>
<th>RMSEC</th>
<th>(R^2_{CV})</th>
<th>RMSECV</th>
<th>RPD</th>
</tr>
</thead>
<tbody>
<tr>
<td>FFA</td>
<td>Full</td>
<td>256</td>
<td>5</td>
<td>0.922</td>
<td>0.039</td>
<td>0.906</td>
<td>0.043</td>
<td>3.268</td>
</tr>
<tr>
<td></td>
<td>Reduced</td>
<td>135</td>
<td>7</td>
<td>0.932</td>
<td>0.037</td>
<td>0.917</td>
<td>0.040</td>
<td>3.488</td>
</tr>
<tr>
<td>Palmitic acid</td>
<td>Full</td>
<td>256</td>
<td>8</td>
<td>0.907</td>
<td>1.358</td>
<td>0.836</td>
<td>1.817</td>
<td>2.710</td>
</tr>
<tr>
<td></td>
<td>Reduced</td>
<td>40</td>
<td>6</td>
<td>0.901</td>
<td>1.403</td>
<td>0.855</td>
<td>1.702</td>
<td>2.893</td>
</tr>
<tr>
<td>Stearic acid</td>
<td>Full</td>
<td>256</td>
<td>5</td>
<td>0.713</td>
<td>1.825</td>
<td>0.618</td>
<td>2.117</td>
<td>1.760</td>
</tr>
<tr>
<td></td>
<td>Reduced</td>
<td>30</td>
<td>8</td>
<td>0.7835</td>
<td>1.594</td>
<td>0.6983</td>
<td>1.896</td>
<td>1.966</td>
</tr>
<tr>
<td>Oleic acid</td>
<td>Full</td>
<td>256</td>
<td>7</td>
<td>0.980</td>
<td>0.915</td>
<td>0.965</td>
<td>1.201</td>
<td>6.053</td>
</tr>
<tr>
<td></td>
<td>Reduced</td>
<td>22</td>
<td>9</td>
<td>0.883</td>
<td>0.850</td>
<td>0.974</td>
<td>1.033</td>
<td>7.039</td>
</tr>
<tr>
<td>Linoleic acid</td>
<td>Full</td>
<td>256</td>
<td>6</td>
<td>0.973</td>
<td>4.425</td>
<td>0.958</td>
<td>5.51</td>
<td>5.956</td>
</tr>
<tr>
<td></td>
<td>Reduced</td>
<td>27</td>
<td>7</td>
<td>0.980</td>
<td>3.832</td>
<td>0.971</td>
<td>4.616</td>
<td>7.110</td>
</tr>
<tr>
<td>(\alpha)-linolenic acid</td>
<td>Full</td>
<td>256</td>
<td>7</td>
<td>0.954</td>
<td>17.022</td>
<td>0.918</td>
<td>22.816</td>
<td>3.854</td>
</tr>
<tr>
<td></td>
<td>Reduced</td>
<td>14</td>
<td>9</td>
<td>0.923</td>
<td>24.285</td>
<td>0.887</td>
<td>29.515</td>
<td>2.979</td>
</tr>
<tr>
<td>(\gamma)-linolenic acid</td>
<td>Full</td>
<td>256</td>
<td>6</td>
<td>0.891</td>
<td>0.177</td>
<td>0.837</td>
<td>0.217</td>
<td>2.461</td>
</tr>
<tr>
<td></td>
<td>Reduced</td>
<td>16</td>
<td>9</td>
<td>0.861</td>
<td>0.198</td>
<td>0.829</td>
<td>0.221</td>
<td>2.418</td>
</tr>
<tr>
<td>Arachidic acid</td>
<td>Full</td>
<td>256</td>
<td>6</td>
<td>0.773</td>
<td>0.249</td>
<td>0.655</td>
<td>0.308</td>
<td>2.025</td>
</tr>
<tr>
<td></td>
<td>Reduced</td>
<td>20</td>
<td>10</td>
<td>0.787</td>
<td>0.288</td>
<td>0.670</td>
<td>0.365</td>
<td>1.710</td>
</tr>
<tr>
<td>Gadoleic acid</td>
<td>Full</td>
<td>256</td>
<td>7</td>
<td>0.853</td>
<td>0.076</td>
<td>0.914</td>
<td>0.100</td>
<td>2.863</td>
</tr>
<tr>
<td></td>
<td>Reduced</td>
<td>24</td>
<td>10</td>
<td>0.923</td>
<td>0.076</td>
<td>0.872</td>
<td>0.098</td>
<td>2.919</td>
</tr>
</tbody>
</table>

### 3.2. Regression models

Based on spectral data, PLSR models to predict FFA value and fatty acids concentrations were calibrated and internally validated, and the results are shown in Table 1 and Figure 3 and Figure 4.

Table 1 shows the statistics for PLSR models developed to quantify FFA value and fatty acids concentrations using full spectral range and selected variables by iPLS. The performance of PLSR model to predict FFA value reached > 0.9 and RPD > 3, indicating a good ability to predict FFA (Saeyes et al., 2005). The chemical maps showed as FFA increased (Figure 3) in chia seeds during storage, due to oxidative reactions (Mannucci et al., 2019). In a similar way, PLSR models to predict oleic acid, linoleic acid and linolenic acid showed the best performance in both the full and reduced model. By applying iPLS, variable reduction reached 91, 89 and 94%, respectively. For oleic acid, linoleic acid (omega-6) and \(\alpha\)-linolenic (omega-3), the PLSR models showed the highest RPD value, reflecting the great potential of hyperspectral imaging to estimate these fatty acids’ composition. Likely, this behavior is due to the high content of oleic, linoleic and \(\alpha\)-linolenic acid in chia seeds (Cruz-Tirado et al., 2021; De Falco et al., 2017). In addition, chemical maps (Figure 4C-E) show as these fatty acids are degraded during storage by oxidative and hydrolytic reactions (Escalona-García et al., 2016).

For palmitic acid, stearic acid, and gadoleic acid, the PLSR models based on selected variables were higher than models based on the full spectra range (Table 1). The variable reduction and the increase in RPD value were 84% and 6.8%, 88% and 11.7% and 91% and 2% for palmitic acid, stearic acid, and gadoleic acid, respectively. Based on the RPD scale proposed by Saeyes et al. (2005), PLSR model on the selected variable is a good prediction of palmitic acid, stearic acid, and gadoleic acid concentration. The chemical maps for the distribution of palmitic acid and stearic acid in chia seeds are displayed in Figures 4A and 4B. Finally, the prediction of \(\gamma\)-linolenic acid and arachidic acid concentrations required a full spectra range (Table 1). The PLSR models reached RPD values between 2.0 – 2.5, indicating an approximate prediction ability.

### 4. Conclusions

Chia seeds are a valuable superfood because they have high concentrations of polyunsaturated fatty acids omega-3 and omega-6. We developed a new methodology for fast quantification of free fatty acids (FFA) and fatty acids concentration by using NIR-hyperspectral imaging and chemometrics. Fuzzy c-means (FCM) assigned the pixels in hyperspectral images to different clusters and displayed the chemical compound variations in chia seeds during storage. It was possible to observe how chemical components are degraded (i.e. phenolic compounds) and others are generated (i.e. free fatty acids) during the storage of chia seeds at different temperatures. PLSR models satisfactory estimate FFA value (= 0.917) and oleic acid (= 0.974), linoleic acid (= 0.971) and \(\alpha\)-linolenic acid (= 0.887) concentration, reached RPD values higher than 2.0. PLSR models to predict oleic acid, linoleic acid and \(\alpha\)-linolenic acid had the best performance. Finally, we demonstrated that NIR-hyperspectral imaging can be used as fast technology to predict FFA and fatty acids concentrations in chia seeds. Future works could study other components such as amino acids and phenolics of chia seeds during storage to verify the results obtained using NIR-HSI.

### Acknowledgment

J. P. Cruz-Tirado acknowledges scholarship funding from FAPESP, grant no 2018/02500-4 and 2020/09198-1. Pedro Renann Lopes de França acknowledges scholarship funding from FAPESP, grant no 2021/06606-4. Prof. Douglas Fernandes Barbin is CNPq research fellow (308260/2021-0).

### ORCID

J. P. Cruz-Tirado [https://orcid.org/0000-0002-1963-4963](https://orcid.org/0000-0002-1963-4963)

P. R. Lopes de França [https://orcid.org/0000-0002-0817-7156](https://orcid.org/0000-0002-0817-7156)

D. F. Barbin [https://orcid.org/0000-0001-9767-8130](https://orcid.org/0000-0001-9767-8130)
References


