



Contents lists available at ScienceDirect

LWT

journal homepage: [www.elsevier.com/locate/lwt](http://www.elsevier.com/locate/lwt)

# An automated non-destructive prediction of peroxide value and free fatty acid level in mixed nut samples

Iman Tahmasbian<sup>a,\*</sup>, Helen M. Wallace<sup>b</sup>, Tsvakai Gama<sup>c</sup>, Shahla Hosseini Bai<sup>b</sup>

<sup>a</sup> Department of Agriculture and Fisheries, Queensland Government, Toowoomba, QLD, 4350, Australia

<sup>b</sup> Centre for Planetary Health and Food Security, School of Environment and Science, Griffith University, Nathan, QLD, 4111, Australia

<sup>c</sup> Genecology, Faculty of Science, Health, Education and Engineering, University of the Sunshine Coast, Maroochydore DC, Qld, 4558, Australia

## ARTICLE INFO

### Keywords:

Food quality control  
Free fatty acid  
Hyperspectral imaging  
Oleic acid  
Peroxide value  
SVM

## ABSTRACT

This study aimed to develop an automated technique, which is rapid, non-destructive and inexpensive, to test for rancidity of nuts. A visible to near infrared benchtop hyperspectral camera was used to capture images from blanched canarium, unblanched canarium and macadamia samples. Support vector machine classification (SVC) and PLSR models were developed to segregate the pooled spectra of the nuts and predict their peroxide values (PV) and free fatty acid (FFA) concentrations. The SVC and PLSR models were then used in a hierarchical model to develop an automated system for predicting PV and FFA. The automated model was then tested using a test set providing classification accuracy of 87% and  $R^2$  between 0.60 and 0.76 and RPD between 1.6 and 2.7 for PV and FFA prediction. Overall, the automated system has the potential commercial application in nut processing to detect rancidity of mixed nut samples non-destructively and in real-time. It is suggested to train other machine learning models with more samples to improve the accuracy of predictions.

## 1. Introduction

Nuts are important source of protein, nutrients, carbohydrates, unsaturated fat, fibre and vitamins, and are therefore recommended for daily consumption as a part of a healthy diet (Brufau et al., 2006; Ros, 2010). Unsaturated fats are, however, susceptible to oxidation when exposed to light, oxygen, heat, moisture and postharvest processing (Hosseini Bai et al., 2019a, 2019b). Nut oil oxidation accelerates rancidity and shortens shelf-life of nuts (Bai et al., 2019; Shahidi & John, 2013). Nut rancidity is usually estimated by measuring peroxide value (PV) and free fatty acid (FFA) concentrations (Franklin et al., 2017; Yang et al., 2013). Analyses of PV and FFA in nuts using methods such as titration, mass spectrometry and instruments such as an OxiTester are, however, expensive and time-consuming. Furthermore, these methods are destructive and do not allow continuous monitoring of the nut quality during processing and storage. Rancid nuts usually remain undetected until they have an off-flavour and have contaminated other nuts in the same batch. New methods are required that can monitor the quality of the nuts during storage in real time, and are inexpensive, rapid and non-destructive.

Laboratory-based hyperspectral imaging (HSI) is a non-destructive technology that measures hundreds of narrow wavelengths and gives

spectral information over spatial dimensions (Manley, 2014). The correlations between the reflectance/absorbance from the spectral dimension (wavelengths) with the concentration of the targeted material are used to predict concentrations of various chemicals (Sun, 2010). The spatial dimension gives HSI an advantage over visible-to near-infrared (VNIR) spectroscopy since HSI can produce a chemical map/images for non-homogenous surfaces (Manley, 2014). The spatial dimension in HSI is also an advantage for homogeneous surfaces since it uses the average reflectance of several pixels and therefore has higher accuracy and reproducibility compared with the single points used in VNIR spectroscopy (Manley, 2014). Moreover, HSI has the potential for simultaneous measurement of multiple quality attributes in a single operation (Sun, 2010).

In the recent years, HSI has been successfully adopted in the VNIR region (400–1700 nm) for food quality assessment, food safety control and classification (Bai et al., 2018; Calvini et al., 2015; Gowen et al., 2007; Kämper et al., 2020; Qin et al., 2020). HSI has also been successfully used in nut quality assessment (Cheng et al., 2018; Gama et al., 2018; Jiang et al., 2016; Yang et al., 2013). For example, HSI has been used for analysing moisture content and textural characteristics of pistachio (Mohammadi-Moghaddam et al., 2018), detecting aflatoxin and deoxynivalenol in hazelnuts and wheat kernels (Kalkan et al., 2011;

\* Corresponding author. Department of Agriculture and Fisheries, Queensland Government, Toowoomba, QLD, 4350, Australia.

E-mail address: [iman.tahmasbian@daf.qld.gov.au](mailto:iman.tahmasbian@daf.qld.gov.au) (I. Tahmasbian).

<https://doi.org/10.1016/j.lwt.2021.110893>

Received 2 June 2020; Received in revised form 20 November 2020; Accepted 10 January 2021

Available online 12 January 2021

0023-6438/Crown Copyright © 2021 Published by Elsevier Ltd. All rights reserved.

Liang et al., 2020) and classifying internally damaged almond nuts (Nakariyakul & Casasent, 2011). HSI has been used to investigate protein, oil concentration, nutrient concentration, moisture content and fungal contamination in peanut, almond and canarium samples (Cheng et al., 2018; Gama et al., 2018; Jiang et al., 2016; Jin et al., 2015; Qiao et al., 2017; Sun et al., 2020). In our complementary study, the concentrations of PV in canarium nuts (blanched) have been successfully predicted using HSI and partial least squares regression (PLSR) models (Bai et al., 2018). However, nuts are often packaged and sold in mixtures. Therefore, identifying and segregating the spectral signatures of various nut species are important for predicting rancidity of each nut where the samples are mixed. Nut identification using HSI can then be used to develop an automated system for predicting nut rancidity. To the best of our knowledge, this is the first study using HSI for simultaneous classification and prediction of PV and FFA in a pooled spectra of nuts (blanched and unblanched canarium and macadamia) using an automated system.

This study aimed to 1) investigate the capabilities of a VNIR HSI system for segregating the pooled spectra of mixed samples including blanched canarium, unblanched canarium and macadamia samples; 2) assess the potential of the HSI for predicting PV and FFA in the blanched and unblanched samples; and also 3) develop an automated system that can predict both PV and FFA using the pooled spectra of nut samples. We hypothesised that the differences among HSI reflectance data from various nuts samples could be used to segregate the pooled spectra. We also hypothesised that the HSI reflectance data could be used to predict PV and FFA in the classified samples using PLSR models correlating the spectral data with the PV and FFA concentrations.

## 2. Material and methods

### 2.1. Sample collection and preparation

The canarium fruits were sourced from a canarium factory in Keravat, Papua New Guinea. The canarium factory received freshly collected fruits from various villages and plantations across East New Britain over a fruit season in 2017. The fruits were carefully assessed for high quality at the factory gate to ensure all fruits were fresh. The fruits were then pooled and mixed in the factory before processing to ensure samples collected for this experiment would be randomised. The fruit pulps were removed manually from the canarium shells, before cracking (Walton et al., 2017). The canarium shells were cracked and samples were then blanched in hot water (100 °C) for 90 s. The testa of the blanched samples was then manually removed. The remaining samples were kept as unblanched canarium. All canarium samples were oven dried at 40 °C until the moisture content of the samples reached 4.50%. In total, 190 blanched and 70 unblanched canarium samples were prepared (Table 1). Macadamia samples were sourced from two orchards located in Queensland and New South Wales. The macadamia fruits were dehusked and placed in ovens to dry within the first 24 h of collecting as described in Gama et al. (2020) to ensure the initial quality of kernels were not compromised. In total, 130 macadamia samples were prepared (Table 1).

Both canarium and macadamia samples were subject to accelerate

ageing at 45 °C for 24 days to increase the variation in PV and FFA data (Bai et al., 2018; Gama et al., 2020). Each nut group was divided into two sub-groups. One sub-group of samples was stored at room temperature and one sub-group of samples was stored at 45 °C for 24 days.

### 2.2. PV and FFA determination

The samples were then crushed using a garlic crusher and added to 80 mL of pentane and stirred for 20 min. The pentane was then removed from the oil using an air-tight vacuum rotator, (BÜCHI Labortechnik AG, Switzerland). The extracted oil was collected and stored at 4 °C for further analyses.

An OxiTester Touch Analyser (Olive OxiTester, Sw version 1.22, CDR, Ginestra, Fiorentina, Florence, Italy) was used to measure PV and FFA at day 24 following the incubation. 5 µL and 2.5 µL oil were placed in glass cuvettes containing relevant reagents to test PV and FFA, respectively. The intensity of the colour developed after adding oils to reagents was measured at 505 nm and 630 nm for PV and FFA samples, respectively. The PV and FFA were expressed as meq O<sub>2</sub>/kg and percentage of oleic acid, respectively.

### 2.3. HSI system and image acquisition process

Hyperspectral images of the samples were acquired using a laboratory-based VNIR hyperspectral camera in the spectral region of 400–1000 nm in an illumination-controlled condition at the University of the Sunshine Coast, Australia. The HSI system comprised a 12-bit push-broom camera (Pika XC2, USA) with the spectral resolution of ~1.3 nm producing 462 spectral wavelength bands, four current-controlled wide spectrum quartz halogen light powered by a continuous current supply to avoid light flickering, a linear translation stage and a data acquisition software (SpectronPro 2.94, Resonon, USA).

The samples were placed on the moving stage and moved to the camera's field of view. The exposure time and speed of the moving stage were set at 27.31 ms and 147 pps, respectively. The region of interest (ROI), which was the area covered by all the samples used in each replicate, were selected and the average signal intensity ( $R_0$ ) of all the ROI pixels (ca. 200,000 pixels) was extracted using the SpectronPro software. To transform the signal intensities obtained from the hyperspectral camera to the reflectance values, image correction was conducted using the dark and white reference images using Eq. (1) (Tahmasbian, Xu, et al., 2018; Tahmasbian, Bia, et al., 2018; Han, Liu, Khoshelham, & Bai, 2021).

$$R = (R_0 - D)/(W - D) \quad (1)$$

where R is the reflectance at each wavelength,  $R_0$  is the average signal intensity of ROI at the corresponding wavelength and D and W are the signal intensity of the dark and white reference images, respectively. The dark image was captured when the lens of the camera were completely closed with its opaque cap and the white image was captured from a uniform and high white reflectance (99.9% diffuse reflectance).

**Table 1**  
Descriptive analysis of canarium and macadamia kernels in both calibration and test sets.

		Calibration set				Test set							
		No.	Mean	SD	CV	Min	Max	No.	Mean	SD	CV	Min	Max
Blanched canarium	PV	65	2.11	1.31	0.62	1.03	6.5	13	2.37	1.11	0.47	1.18	5.2
Unblanched canarium	PV	70	1.94	1.41	0.72	0.3	5	14	2.19	1.7	0.77	0.3	5
Blanched canarium	FFA	190	0.11	0.06	0.57	0.02	0.27	38	0.10	0.06	0.55	0.02	0.27
Unblanched canarium	FFA	70	0.08	0.03	0.40	0.03	0.16	14	0.07	0.03	0.38	0.03	0.13
Macadamia	FFA	130	0.11	0.05	0.44	0.02	0.02	26	0.11	0.05	0.48	0.02	0.2

SD: standard deviation, CV: coefficient of variations, Min: minimum, Max: maximum, PV: peroxide value (meq O<sub>2</sub>/kg) and FFA: free fatty acid (%).

## 2.4. Data analysis and model development

We first developed a supervised support vector machine classification (SVC) model to segregate the spectral reflectance of the different samples. Then, PLSR models were developed to predict PV and FFA separately in blanched canarium and unblanched canarium and to predict FFA in macadamia samples. The SVC and the PLSR models were finally combined using a hierarchical model to develop an automated system which classifies the pooled spectral reflectance data of a test dataset into blanched canarium, unblanched canarium and macadamia and predict PV and FFA in each category (details below).

### 2.4.1. Data cleaning, division and transformation

Hottelling's  $T^2$  test within 99% of confidence level was performed to identify the spectral outliers in the datasets (Tahmasbian, Xu, et al., 2018). Then, the dataset was randomly divided into two separate datasets using 80% of the data (calibration set), to develop the models, and 20% of the data, that were not used to develop the model (test set), to test the developed models (Tahmasbian, Bia, et al., 2018). See Table 1 for number of samples. There were no significant differences between the means in the calibration and test using a  $t$ -test ( $p > 0.05$ ).

The calibration data were then subjected to different data transformations including standard normal variate (SNV), first derivative (1dv), second derivative (2dv), de-trend and multiple scatter correction (MSC) prior to SVC and PLSR modelling for reducing the artefacts and increase the signal-to noise-ratio. All the data transformations were conducted using the Unscrambler software package (version10.5.1, CAMO Software Inc., Trondheim, Norway).

### 2.4.2. SVC model development

The SVC is an emerging data classification technique which has been of great interest to classify high-dimensional data such as hyperspectral data (Xu et al., 2017). In the present study, different kernel types of SVC including linear, polynomial, radial bases function and sigmoid were used on both C-SVC and  $\nu$ -SVC. A systematic grid search method was used to selected the optimal values of C,  $\nu$  and  $\gamma$  parameters (Xu et al., 2017). Various types of data transformation were conducted prior to the SVC. The accuracy of the classification was defined by the percentage of the correctly classified spectra. The most accurate SVC was then selected for use in the hierarchical model. The SVC models were developed using the Unscrambler software package (version10.5.1, CAMO Software Inc., Trondheim, Norway).

### 2.4.3. PLSR model development

Five separate PLSR models were developed to correlate the spectra of samples to their corresponding value of (1) PV in blanched canarium, (2) PV in unblanched canarium, (3) FFA in blanched canarium, (4) FFA in unblanched canarium and (5) FFA in macadamia samples. The number of latent variables of each PLSR models was set to the lowest predicted residual error sum of square (PRESS) calculated using Eq. (2) (Malmir et al., 2019, 2020).

$$PRESS = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (2)$$

where  $Y_i$  and  $\hat{Y}_i$  are the reference and predicted values of each target variable in the  $i^{\text{th}}$  sample, respectively.

The initial PLSR models were developed using the differently transformed calibration data. The models providing the highest accuracy of prediction (high  $R^2$  and low RMSE) in the calibration and cross-validation sets were selected for further processes.

The wavelengths with the lowest  $\beta$  coefficient in each selected initial model were removed and the models were re-developed using the remaining (informative) wavelengths (Kamruzzaman et al., 2016). The removed wavelengths were returned to the models in the case where the accuracy of the models reduced after wavelength selection. The process

was repeated until the highest accuracy of the PLSR models were achieved (Bai et al., 2018; Malmir et al., 2019; Tahmasbian et al., 2017). The wavelength selection was conducted in an attempt to simplify the calculations and to reduce the dimensionality of the hyperspectral data (Wold et al., 2001). The models developed using informative wavelengths were selected as the final models and were used in the hierarchical model. The variable importance in projection (VIP) was used to select the most important wavelengths in the final models for further analysis. VIP was calculated using Eq. 3. PLSR modelling was conducted using the Unscrambler software package (version10.5.1, CAMO Software Inc., Trondheim, Norway).

$$VIP_j = \sqrt{\frac{\sum_{f=1}^F w_{jf}^2 \cdot SSY_f \cdot J}{SSY_t \cdot F}} \quad (3)$$

where,  $VIP_j$  was the importance of  $j^{\text{th}}$  wavelength in the model with  $F$  number of components,  $w_{jf}$  was the loading weight of the corresponding wavelength in the  $f^{\text{th}}$  component,  $SSY_f$  was the explained sum of squares of the target variable in the  $f^{\text{th}}$  component,  $SSY_t$  was the total sum of squares of the target variable and  $J$  was total number of wavelengths. The most common threshold set for VIP is 1, above which the wavelengths are considered important (Oussama et al., 2012; Williams et al., 2012). However, we used a more strict threshold of  $VIP > 1.5$  as the most important wavelengths to simplify the discussion due to the large number of the selected bands in the current study (Kandpal et al., 2016).

## 2.5. Model evaluation

The classification accuracy was evaluated by calculating the percentage of the correctly classified samples and the prediction accuracies were evaluated by  $R^2$  and RMSE in both calibration and cross-validation sets. The cross-validation method was a k-fold ( $k = 10$ ) method (Kohavi, 1995). The ratio of performance to deviation (RPD) was also calculated using Eq. 4 to further evaluate the reliability of the predictions (Bellon-Maurel et al., 2010). The models with  $1.4 < RPD < 2$  were considered as reliable and those with  $RPD > 2$  were considered as highly reliable (Bellon-Maurel et al., 2010). The RPD was calculated using Microsoft Excel 365 ProPlus.

$$RPD = SD_t / RMSE_t \quad (4)$$

where  $SD_t$  and  $RMSE_t$  are the standard deviation of the reference and root mean square error of prediction in the test set, respectively.

## 3. Results

### 3.1. Descriptive analysis

The spectra of the blanched canarium and macadamia showed an approximately similar trend in the studied region of 400–1000 nm (Fig. 1). The average reflectance of unblanched canarium was lower than those of blanched canarium and macadamia. There was an obvious difference in the spectral signature of the unblanched canarium compared with those of blanched canarium and macadamia. However, spectra of the all three types of kernels overlapped in the spectral region of 950–1000 nm (Fig. 1).

### 3.2. SVC descriptions

The best classification (90% accuracy in calibration and cross-validation sets) was obtained using a  $\nu$ -SVC with a 3-degree polynomial kernel,  $\nu$  of 0.44 and  $\gamma$  of 0.01, developed using the MSC + 1dv transformed data. The MSC + 1dv improved the accuracy of prediction by 4% compared with that of raw data (86% of accuracy in the cross-validation set). The confusion matrix showed that most of the misclassifications occurred between blanched canarium and macadamia,

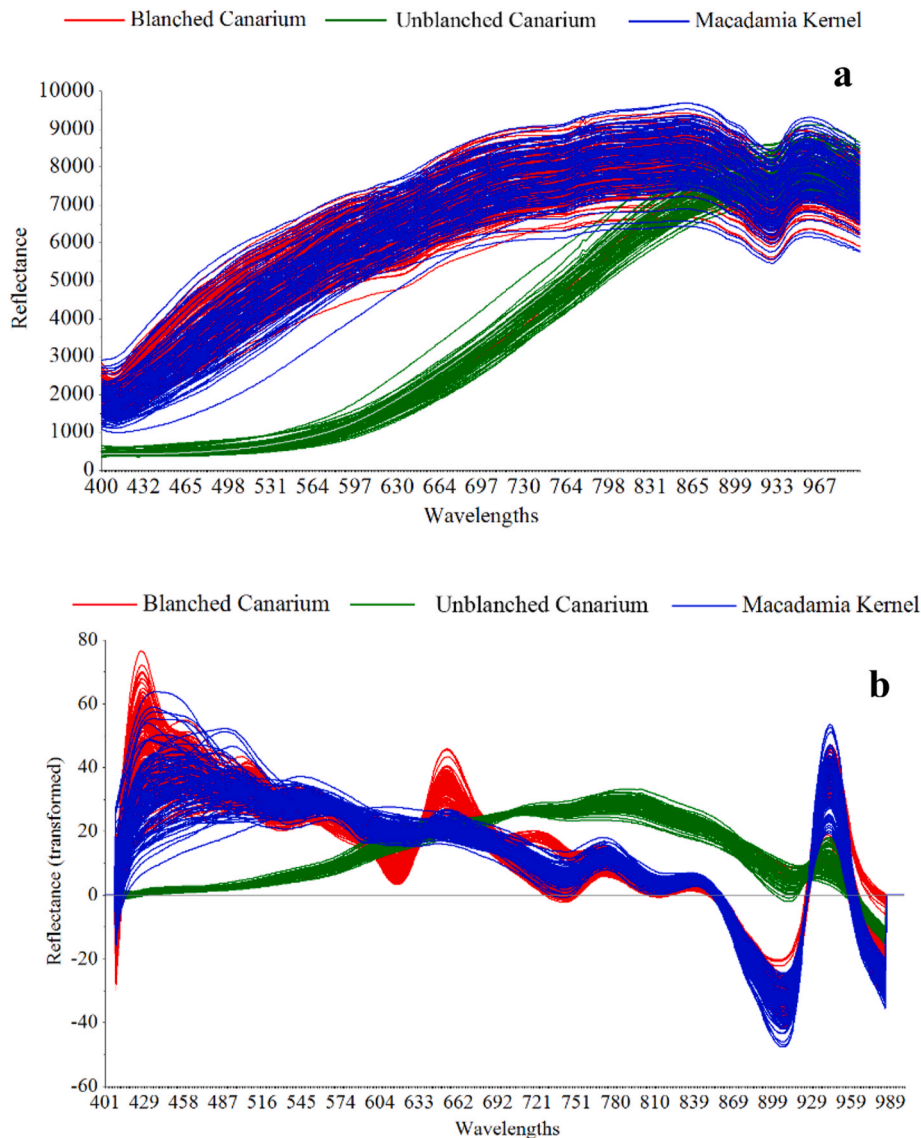


Fig. 1. The (a) raw spectra and (b) transformed spectra of blanched canarium, unblanched canarium and macadamia.

resulting in 80% accuracy, whereas all the unblanched canarium and

Table 2

Confusion matrix of support vector machine classification (SVC) representing the number of nuts predicted correctly versus those predicted incorrectly using SVC.

		Blanched canarium (Reference)	Unblanched canarium (Reference)	Macadamia (Reference)
Calibration	Blanched canarium	152	0	1
	Unblanched canarium	5	70	0
	Macadamia	33	0	129
Test set	Blanched canarium	28	0	0
	Unblanched canarium	1	14	0
	Macadamia	9	0	26

Sum of the numbers in each column represents the actual number of samples of that specific nut (in calibration or test sets), while the number in each row represents the number of classified nuts for that specific class. Higher number in the matched column and row shows better prediction.

macadamia samples were classified correctly—100% accuracy (Table 2). The transformation reduced the overlaps in the spectra between the raw and transformed spectral signature of the blanched canarium and macadamia (Fig. 1a vs. Fig. 1b).

### 3.3. PLSR model descriptions

The best initial models for PV in blanched canarium ( $R_c^2 = 0.76$  and  $R_{cv}^2 = 0.58$ ) and unblanched canarium ( $R_c^2 = 0.94$  and  $R_{cv}^2 = 0.89$ ) were developed using 2dv-transformed data and raw data, respectively (Table 3). The best initial models for FFA in blanched canarium ( $R_c^2 = 0.68$  and  $R_{cv}^2 = 0.55$ ), unblanched canarium ( $R_c^2 = 0.64$  and  $R_{cv}^2 = 0.40$ ) and macadamia ( $R_c^2 = 0.50$  and  $R_{cv}^2 = 0.33$ ) were developed using OSC-transformed data, 2dv-transformed data and raw data, respectively (Table 3). The prediction accuracy of the models improved after removing the uninformative wavelengths from the models. The accuracy of the final models (with only informative wavelengths) were  $R_c^2 = 0.76$  and  $R_{cv}^2 = 0.60$  for PV prediction in blanched canarium,  $R_c^2 = 0.95$  and  $R_{cv}^2 = 0.89$  for PV in unblanched canarium,  $R_c^2 = 0.75$  and  $R_{cv}^2 = 0.67$  for FFA prediction in blanched canarium,  $R_c^2 = 0.72$  and  $R_{cv}^2 = 0.54$  for FFA in unblanched canarium and  $R_c^2 = 0.88$  and  $R_{cv}^2 = 0.70$  for FFA in

**Table 3**  
Initial and final partial least squares regression (PLSR) models properties.

			TF	WL No.	LV No.	R <sup>2</sup> <sub>c</sub>	R <sup>2</sup> <sub>cv</sub>	RMSE <sub>c</sub>	RMSE <sub>cv</sub>
Initial models	Blanched canarium	PV	2dv	462	6	0.76	0.58	0.64	0.89
	Unblanched canarium	PV	Raw	462	11	0.94	0.89	0.32	0.47
	Blanched canarium	FFA	OSC	462	4	0.68	0.55	0.04	0.04
	Unblanched canarium	FFA	2dv	462	5	0.64	0.40	0.02	0.03
	Macadamia	FFA	Raw	462	12	0.50	0.33	0.04	0.05
Final models	Blanched canarium	PV	2dv	292	6	0.76	0.60	0.64	0.87
	Unblanched canarium	PV	Raw	377	11	0.95	0.89	0.32	0.47
	Blanched canarium	FFA	OSC	247	4	0.75	0.67	0.03	0.04
	Unblanched canarium	FFA	2dv	440	5	0.72	0.54	0.02	0.02
	Macadamia	FFA	Raw	331	15	0.88	0.70	0.09	0.14

PV: peroxide value, FFA: free fatty acid, TF: transformation, WL: wavelengths, LV: latent variable, 2dv: second derivative, OSC: orthogonal signal correction. R<sup>2</sup><sub>c</sub>: coefficient of determination for calibration set, R<sup>2</sup><sub>cv</sub>: the coefficient of determination for cross-validation set, RMSE<sub>c</sub>: root mean square error for calibration (meq O<sub>2</sub>/kg for PV and % for FFA), RMSE<sub>cv</sub>: root mean square error for cross-validation (meq O<sub>2</sub>/kg for PV and % for FFA). The initial models are the models developed using full number of wavelengths and the final models were developed using informative wavelengths.

macadamia (Table 3). The final models were included in the hierarchical models to automatically predict the PV and FFA values using the spectra of blanched canarium, unblanched canarium and macadamia classified using SVC.

### 3.4. Selecting the informative and the most important wavelengths

The number of informative wavelengths selected for predicting PV in blanched canarium, PV in unblanched canarium, FFA in blanched canarium, FFA in unblanched canarium and FFA in macadamia samples were 292, 377, 241, 440 and 321, respectively. The number of most important wavelengths were 12, 6, 25, 14 and 3 for PV in blanched canarium, PV in unblanched canarium, FFA in blanched canarium, FFA in unblanched canarium and FFA in macadamia samples, respectively (Fig. 2). The most important wavelengths for predicting PV in blanched canarium were located in the spectral regions of 400–410 nm and 925–935 nm and those of PV in unblanched canarium were located in the region of 773–781 nm (Fig. 2). The most important wavelengths for the prediction of FFA in blanched canarium were located in the spectral region of 737–805 nm, those for prediction of FFA in unblanched canarium were located in the region of 403–415 and 763–764 nm in addition to 791 nm (Fig. 2). For the prediction of FFA in macadamia, the most important wavelengths were 774 nm, 777 nm and 998 nm (Fig. 2).

### 3.5. Evaluating the automated classification and prediction performance of the models using the test set

The hierarchical model was tested using the test set comprising blanched canarium, unblanched canarium and macadamia samples. The results indicated that, overall, 87% of the samples were classified correctly. All of the unblanched canarium and macadamia were classified correctly, 100% accurate classification (Table 2). However, there were some misclassifications for blanched canarium which were classified as macadamia, 73.6% accurate classification (Table 2).

The correctly classified samples were entered the PLSR models developed for prediction of PV and FFA in each specific class. The PVs were predicted in blanched canarium and unblanched canarium in the test set with R<sup>2</sup><sub>t</sub> of 0.60, RMSE<sub>t</sub> of 0.90 meq O<sub>2</sub>/kg and RPD of 1.6 and R<sup>2</sup><sub>t</sub> of 0.76, RMSE<sub>2</sub> of 0.81 meq O<sub>2</sub>/kg and RPD of 2.0, respectively (Fig. 3a and b). The FFA values were predicted in blanched canarium, unblanched canarium and macadamia with R<sup>2</sup><sub>t</sub> of 0.70, RMSE<sub>2</sub> of 0.03% and RPD of 2.0 and R<sup>2</sup><sub>t</sub> of 0.76, RMSE<sub>2</sub> of 0.01% and RPD of 2.7 and R<sup>2</sup><sub>t</sub> of 0.71, RMSE<sub>2</sub> of 0.16% and RPD of 1.8, respectively (Fig. 2c–e).

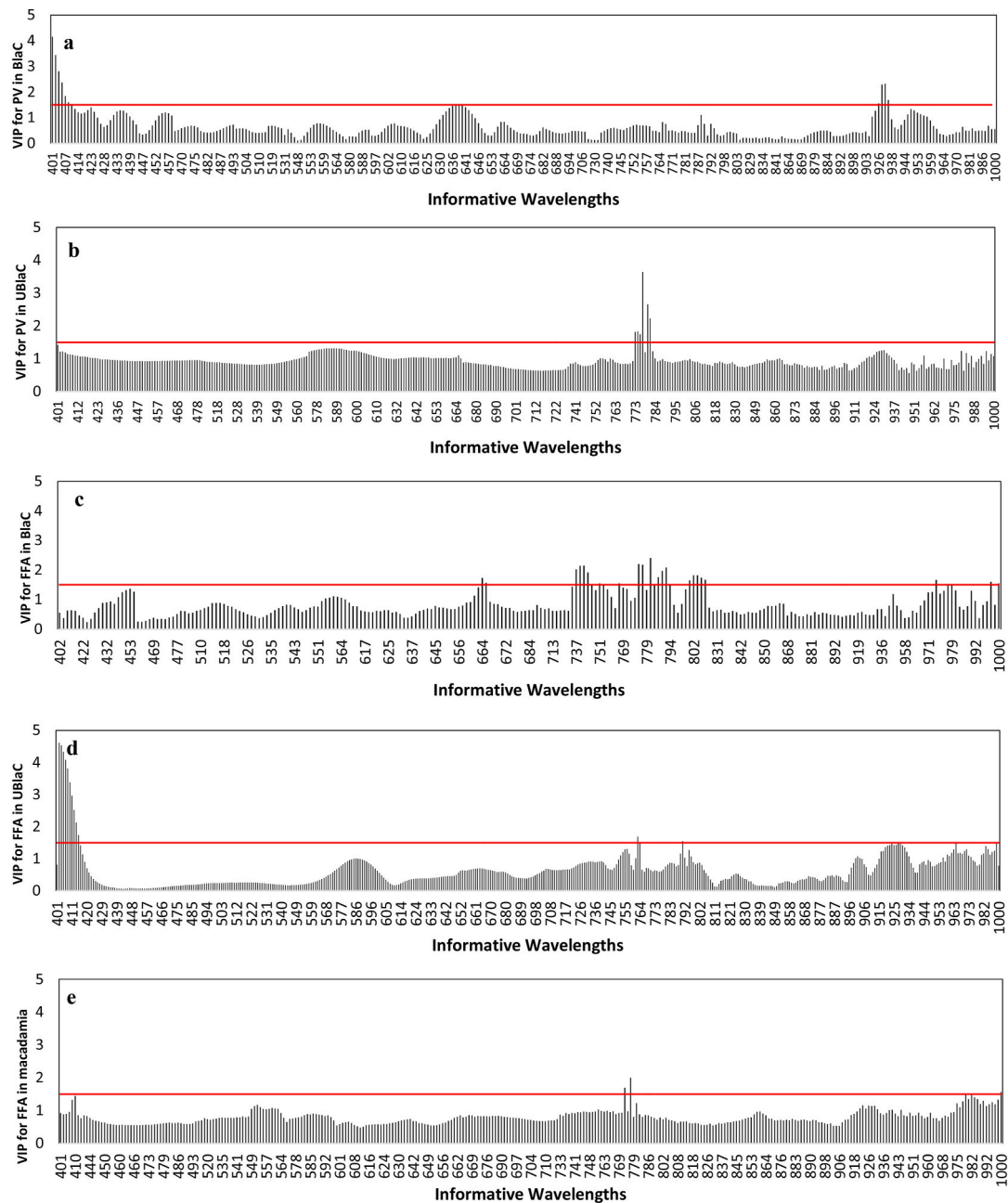
## 4. Discussion

Our results showed that the automated system successfully segregated the pooled spectral reflectance of the nut samples and predicted

PV and FFA in unblanched canarium, blanched canarium and macadamia with high classification and prediction accuracy. The automatic detection of the nuts' spectra in the pooled samples gives the HSI a potential for identifying the targeted nuts and removing impurities. The HSI has the capacity to analyse bulk samples in real time compared with destructive samples analysing only one subsample in long period of time (less representative results). The rapid analysis of all nuts would also reduce the waste by early detection and removal of the rancid nuts. Thus, HSI in combination with SVC and PLSR modelling has potential commercial applications in nut processing as an automated system that can detect and remove rancid nuts in mixed samples.

This is the first study that used a VNIR (400–1000 nm) HSI system for classification of pooled nuts spectra and automatically predicted PV and FFA in the nut samples. In other studies, the concentrations of PV and FFA in corn kernels, sunflower seeds and macadamia single nuts have been predicted using HSI/VNIR spectroscopy operated in various spectral regions from 950 nm to 2500 nm (Canneddu et al., 2016; Cantarelli et al., 2009; Weinstock et al., 2006). In our study, however, HSI (400–1000 nm) was the system used to predict PV and FFA in canarium and macadamia samples. The HSI system used in this study included a spatial dimension, using the average of several pixels (ca. 200,000 pixels) compared with NIR spectroscopy where only one or limited pixel is used (Manley, 2014). Using the average of several pixels would result in an increased accuracy, reliability and reproducibility of the prediction due to increasing signal to noise ratio, compared with NIR spectroscopy (Manley, 2014). Hence, using the average of several pixels of each hyperspectral image captured with a high resolution HSI camera may explain why we were able to predict PV and FFA in the nuts in a smaller range of spectral reflectance (400–1000 nm).

Classification of the nuts, however, was more challenging in the spectral range of 400–1000 nm, due to the complete overlap of the raw spectra measured from blanched canarium and macadamia samples in this spectral range. The blanched canarium and macadamia kernels have similar light cream colour. The similar light colour might explain the overlapped spectral signature of blanched canarium and macadamia kernels. The data transformation, however, reduced the overlaps between blanched canarium and macadamia spectra, in the most part of the spectra. The effects of transformation might be attributed to removing baseline using MSC and 1dv, which resulted in higher signal to noise ratio (Manley, 2014). Other studies have also used data transformation techniques to distinguish the nuts having almost similar spectra. For example, a combination of SNV and 2dv has been used prior to separating the marketable shelled macadamia nuts from those defected with citrus fruit borer (*Ecdyolopha aurantiana*) using NIR spectroscopy (Canneddu et al., 2016). The 1dv and 2dv have also been used to distinguish insect-damaged and fungal-infected chestnuts and almonds using NIR spectroscopy (Liang et al., 2015; Moscetti et al., 2014). Therefore, specific data transformations are required to be



**Fig. 2.** The most important wavelengths for the prediction of a: PV in blanched canarium (Blac), b: PV in unblanched canarium (UBlaC), c: FFA in Blac, d: FFA in UBlaC and e: FFA in macadamia. VIP is variable importance in projection. Wavelengths with VIP larger than 1.5 (red line) are the most important wavelengths. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

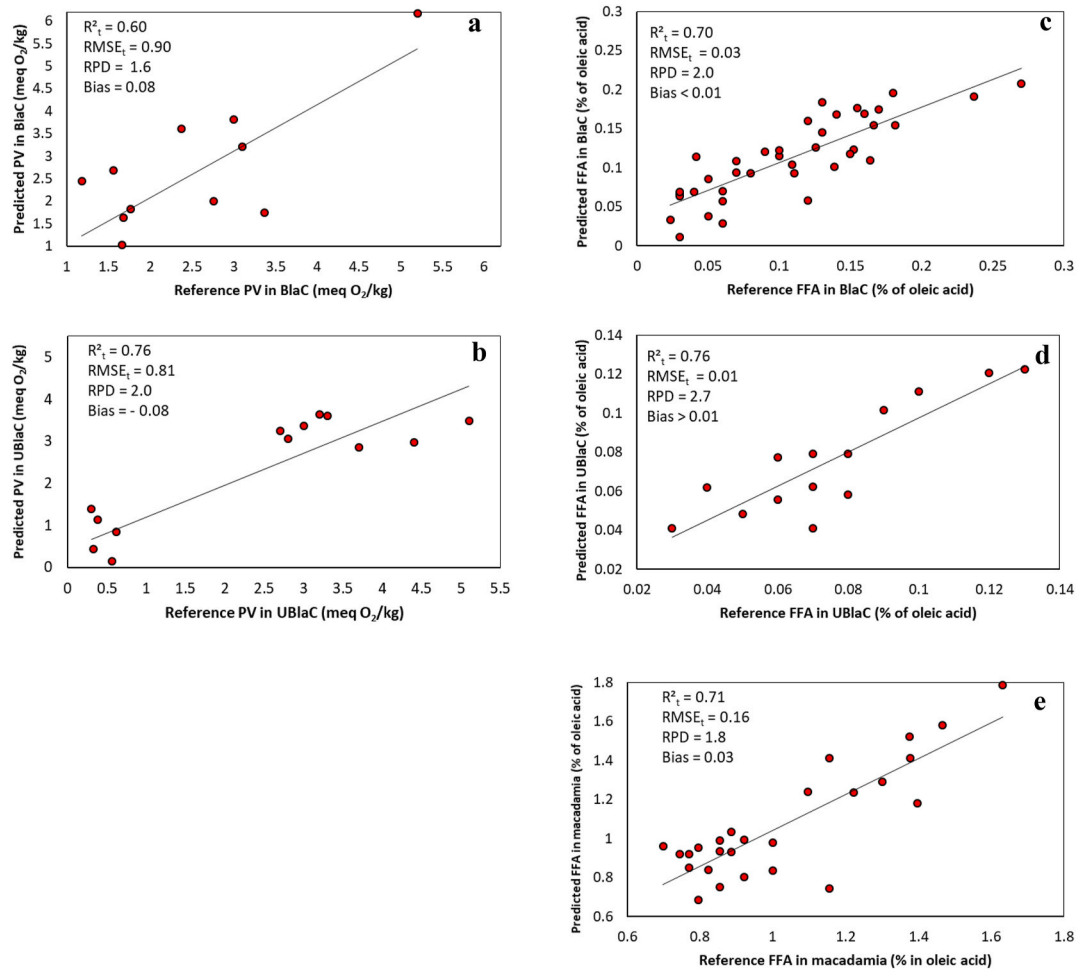
considered to decrease the overlap among spectra of various nut species when models are developed in pool dataset.

We were able to predict PV and FFA in blanched and unblanched canarium and FFA in macadamia, automatically after classification, with  $R_t^2$  between 0.60 and 0.76 and RPD between 1.6 and 2.7. The RPD over 1.4 shows reliable to highly reliable prediction accuracies (Belton-Maurel et al., 2010). Little published information is available to compare these results with. In our complementary study, we have compared a push broom and a staring camera (400–1000 nm) for predicting PV in canarium nuts (blanched only) providing prediction accuracies within the same range as this study (Bai et al., 2018). PV and FFA in unblanched canarium and macadamia have not been investigated in the study undertaken by Bai et al. (2018). In both the current and complementary studies, the accuracy of the predictions were improved

after removing the uninformative wavelengths. The improved accuracies after removing the uninformative wavelengths might be due to the reduced inter-correlation of the wavelengths (Wold et al., 1996).

The most important wavelengths for the prediction of PV and FFA in the blanched and unblanched canarium and in macadamia samples were observed in the spectral regions of 400–415 nm, 730–790 nm and 926–1000 nm, while the regions of 500–700 nm and 800–900 nm were less important (still used in the models as informative wavelengths).

It should be noted that the main reasons to explain why PV and FFA concentrations were predicted in VNIR range remain uncertain. However, previous studies have attributed the reflectance/absorbance in 910–990 nm to C–H third overtone associated with protein and oil contents and O–H first and second overtones related to water and starch contents (Curran 1989). Both PV and FFA have O–H and C–H groups and



**Fig. 3.** The predicted vs. reference values of a: PV in blanched canarium (BlaC), b: PV in unblanched canarium (UBlaC), c: FFA in BlaC, d: FFA in UBlaC and e: FFA in macadamia.

are associated with oil oxidation which may explain the mechanisms behind the results achieved in this study. Nut rancidity which occurs with PV and FFA accumulation may lead to subtle nut colour changes which may not be visible. The variations in the nut colour, affected by varying concentrations of PV and FFA, might be another reason for the successful prediction of PV and FFA in VNIR range. Previous studies measured oil and fatty acid concentrations using NIR spectroscopy, in sunflower, perilla and Brassica seeds, reported that the best wavelengths are located in 1600 nm–1800 nm and around 2310 nm (Kaur et al., 2016; Kim et al., 2007; Sen et al., 2018; Velasco et al., 1999). Therefore, more studies are recommended to investigate the mechanisms whereby the predictions of PV and FFA in canarium and macadamia were possible using HSI in VNIR spectral range.

Prediction of the PV and FFA contents of canarium and macadamia in the VNIR range would help to reduce the cost of equipment as HSI cameras operated in the VNIR range are more affordable than SWIR (1000–2500 nm) HSI cameras. Further, selecting the important spectral regions for prediction of PV and FFA enhances the producing/selecting lower cost multispectral cameras that cover the region of interest only and produces less inter-correlated wavelengths. The HSI system used in this study also has potential to be calibrated for rapid and inexpensive monitoring the quality of other nuts and foods.

## 5. Conclusions

An automated system was designed and used to segregate the pooled spectra of different nut samples, including blanched and unblanched

canarium and macadamia, and to predict PV and FFA in the nuts classified nuts. The SVC classification of the spectra provided 90% and 87% accuracy in the calibration and the test sets respectively. The PLSR models showed acceptable predictions with  $R^2$  between 0.6 and 0.76 and RPD between 1.6 and 2.7. The results indicated that the HSI combined with SVC and PLSR can be used to detect rancidity of mixed nut samples rapidly and non-destructively and has potential commercial applications in nut storage and processing. There is also potential application for the HSI system developed in this study to be calibrated for identifying/classifying other nuts/foods and predicting various properties. More studies are recommended to improve the calibration using more samples and training other machine learning algorithms and to investigate the mechanisms enabled prediction of PV and FFA in VNIR spectral range.

## CRedit authorship contribution statement

**Iman Tahmasbian:** Formal analysis, Data curation, Formal data analysis and, Visualization, Writing - original draft, preparation. **Helen M. Wallace:** Funding acquisition, Funding, Supervision, Writing - review & editing. **Tsvakai Gama:** Data curation, Data collection and laboratory, Formal analysis, analysis. **Shahla Hosseini Bai:** Funding acquisition, Funding, Supervision, Writing - review & editing.

## Acknowledgments

This study has been partly funded by the Australian Centre for

International Agricultural Research (project FST/2014/099). We thank Ms. Hannet and Mr. Hannet from National Agriculture Research Institute (NARI) in Kerevat, New East Britain for providing canarium nuts.

## References

- Bai, S. H., Brooks, P., Gama, R., Nevenimo, T., Hannet, G., Hannet, D., ... Wallace, H. M. (2019). Nutritional quality of almond, canarium, cashew and pistachio and their oil photooxidative stability. *Journal of Food Science & Technology*, *56*, 792–798.
- Bai, S. H., Tahmasbian, I., Zhou, J., Nevenimo, T., Hannet, G., Walton, D., ... Wallace, H. M. (2018). A non-destructive determination of peroxide values, total nitrogen and mineral nutrients in an edible tree nut using hyperspectral imaging. *Computers and Electronics in Agriculture*, *151*, 492–500. <https://doi.org/10.1016/j.compag.2018.06.029>
- Bellon-Maurel, V., Fernandez-Ahumada, E., Palagos, B., Roger, J.-M., & McBratney, A. (2010). Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. *TRAC Trends in Analytical Chemistry*, *29*(9), 1073–1081.
- Brufau, G., Boatella, J., & Rafecas, M. (2006). Nuts: Source of energy and macronutrients. *British Journal of Nutrition*, *96*(S2), S24–S28.
- Calvini, R., Ulrici, A., & Amigo, J. M. (2015). Practical comparison of sparse methods for classification of Arabica and Robusta coffee species using near infrared hyperspectral imaging. *Chemometrics and Intelligent Laboratory Systems*, *146*, 503–511.
- Canneddu, G., Júnior, L. C. C., & de Almeida Teixeira, G. H. (2016). Quality evaluation of shelled and unshelled macadamia nuts by means of near-infrared spectroscopy (NIR). *Journal of Food Science*, *81*(7), C1613–C1621.
- Cantarelli, M. A., Punes, I. G., Marchevsky, E. J., & Camiña, J. M. (2009). Determination of oleic acid in sunflower seeds by infrared spectroscopy and multivariate calibration method. *Talanta*, *80*(2), 489–492. <https://doi.org/10.1016/j.talanta.2009.07.004>
- Cheng, J.-H., Jin, H., & Liu, Z. (2018). Developing a NIR multispectral imaging for prediction and visualization of peanut protein content using variable selection algorithms. *Infrared Physics & Technology*, *88*, 92–96. <https://doi.org/10.1016/j.infrared.2017.11.018>
- Curran, P. J. (1989). Remote sensing of foliar chemistry. *Remote Sensing of Environment*, *30*(3), 271–278.
- Franklin, L. M., Chapman, D. M., King, E. S., Mau, M., Huang, G., & Mitchell, A. E. (2017). Chemical and sensory characterization of oxidative changes in roasted almonds undergoing accelerated shelf life. *Journal of Agricultural and Food Chemistry*, *65*(12), 2549–2563.
- Gama, T., Wallace, H. M., Trueman, S. J., Jones, K., & Hosseini-Bai, S. (2020). Late-dropping macadamia nuts have reduced shelf life. *Scientia Horticulturae*, *268*, Article 109378. <https://doi.org/10.1016/j.scienta.2020.109378>
- Gama, T., Wallace, H., Trueman, S., Tahmasbian, I., & Hosseini Bai, S. (2018). Hyperspectral imaging for non-destructive prediction of total nitrogen concentration in almond kernels. *Acta Horticulturae*, *1219*, 259–264. <https://doi.org/10.17660/ActaHortic.2018.1219.40>
- Gowen, A., O'Donnell, C., Cullen, P., Downey, G., & Frias, J. (2007). Hyperspectral imaging—an emerging process analytical tool for food quality and safety control. *Trends in Food Science & Technology*, *18*(12), 590–598.
- Han, Y., Liu, Z., Khoshelham, K., & Bai, S. H. (2021). Quality estimation of nuts using deep learning classification of hyperspectral imagery. *Computers and Electronics in Agriculture*, *180*, 105868. <https://doi.org/10.1016/j.compag.2020.105868>
- Hosseini Bai, S., Nevenimo, T., Hannet, G., Hannet, D., Jones, K., Trueman, S., ... Wallace, H. (2019). Freezing, roasting and salt dipping impacts on peroxide value, free fatty acid and fatty acid concentrations of nut kernels. *Acta Horticulturae*, *1256*, 71–75.
- Hosseini Bai, S., Trueman, S., Gama, T., Jones, K., Walton, D., Randall, B., & Wallace, H. (2019). Shelf life of macadamia kernels of different origin. *Acta Horticulturae*, *1256*, 375–378. <https://doi.org/10.17660/ActaHortic.2019.1256.53>
- Jiang, J., Qiao, X., & He, R. (2016). Use of Near-Infrared hyperspectral images to identify moldy peanuts. *Journal of Food Engineering*, *169*, 284–290. <https://doi.org/10.1016/j.jfoodeng.2015.09.013>
- Jin, H., Li, L., & Cheng, J. (2015). Rapid and non-destructive determination of moisture content of peanut kernels using hyperspectral imaging technique. *Food Analytical Methods*, *8*(10), 2524–2532. <https://doi.org/10.1007/s12161-015-0147-1>
- Kalkan, H., Beriat, P., Yardimci, Y., & Pearson, T. C. (2011). Detection of contaminated hazelnuts and ground red chili pepper flakes by multispectral imaging. *Computers and Electronics in Agriculture*, *77*(1), 28–34. <https://doi.org/10.1016/j.compag.2011.03.005>
- Kämper, W., Trueman, S. J., Tahmasbian, I., & Bai, S. H. (2020). Rapid determination of nutrient concentrations in Hass avocado fruit by Vis/NIR hyperspectral imaging of flesh or skin. *Remote Sensing*, *12*(20), 3409.
- Kamruzzaman, M., Makino, Y., & Oshita, S. (2016). Parsimonious model development for real-time monitoring of moisture in red meat using hyperspectral imaging. *Food Chemistry*, *196*, 1084–1091.
- Kandpal, L. M., Lohumi, S., Kim, M. S., Kang, J.-S., & Cho, B.-K. (2016). Near-infrared hyperspectral imaging system coupled with multivariate methods to predict viability and vigor in muskmelon seeds. *Sensors and Actuators B: Chemical*, *229*, 534–544. <https://doi.org/10.1016/j.snb.2016.02.015>
- Kaur, B., Sangha, M. K., & Kaur, G. (2016). Calibration of NIRS for the estimation of fatty acids in Brassica juncea. *Journal of the American Oil Chemists' Society*, *93*(5), 673–680.
- Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *International Joint Conference on Artificial Intelligence*, *14*, 1137–1145 (Canada).
- Liang, K., Jiani, H., Ruiyin, H., Qiujin, W., Yinyin, C., & Mingxia, S. (2020). Comparison of Vis-NIR and SWIR hyperspectral imaging for the non-destructive detection of DON levels in Fusarium head blight wheat kernels and wheat flour. *Infrared Physics & Technology*, Article 103281.
- Liang, P.-S., Slaughter, D. C., Ortega-Beltran, A., & Michailides, T. J. (2015). Detection of fungal infection in almond kernels using near-infrared reflectance spectroscopy. *Biosystems Engineering*, *137*, 64–72. <https://doi.org/10.1016/j.biosystemseng.2015.07.010>
- Malmir, M., Tahmasbian, I., Xu, Z., Farrar, M. B., & Bai, S. H. (2019). Prediction of soil macro- and micro-elements in sieved and ground air-dried soils using laboratory-based hyperspectral imaging technique. *Geoderma*, *340*, 70–80. <https://doi.org/10.1016/j.geoderma.2018.12.049>
- Malmir, M., Tahmasbian, I., Xu, Z., Farrar, M. B., & Bai, S. H. (2020). Prediction of macronutrients in plant leaves using chemometric analysis and wavelength selection. *Journal of Soils and Sediments*, *20*(1), 249–259.
- Manley, M. (2014). Near-infrared spectroscopy and hyperspectral imaging: Non-destructive analysis of biological materials. *Chemical Society Reviews*, *43*(24), 8200–8214.
- Mohammadi-Moghaddam, T., Razavi, S. M., Taghizadeh, M., Pradhan, B., Sazgarnia, A., & Shaker-Ardekani, A. (2018). Hyperspectral imaging as an effective tool for prediction the moisture content and textural characteristics of roasted pistachio kernels. *Journal of Food Measurement and Characterization*, 1–10.
- Moscetti, R., Haff, R. P., Saranwong, S., Monarca, D., Cecchini, M., & Massantini, R. (2014). Nondestructive detection of insect infested chestnuts based on NIR spectroscopy. *Postharvest Biology and Technology*, *87*, 88–94. <https://doi.org/10.1016/j.postharvbio.2013.08.010>
- Nakariyakul, S., & Casasent, D. P. (2011). Classification of internally damaged almond nuts using hyperspectral imagery. *Journal of Food Engineering*, *103*(1), 62–67. <https://doi.org/10.1016/j.jfoodeng.2010.09.020>
- Oussama, A., Elabadi, F., Platikanov, S., Kzaiber, F., & Tauler, R. (2012). Detection of olive oil adulteration using FT-IR spectroscopy and PLS with variable importance of projection (VIP) scores. *Journal of the American Oil Chemists' Society*, *89*(10), 1807–1812.
- Qiao, X., Jiang, J., Qi, X., Guo, H., & Yuan, D. (2017). Utilization of spectral-spatial characteristics in shortwave infrared hyperspectral images to classify and identify fungi-contaminated peanuts. *Food Chemistry*, *220*, 393–399. <https://doi.org/10.1016/j.foodchem.2016.09.119>
- Qin, J., Vasefi, F., Hellberg, R. S., Akhbardeh, A., Isaacs, R. B., Yilmaz, A. G., Hwang, C., Baek, I., Schmidt, W. F., & Kim, M. S. (2020). Detection of fish fillet substitution and mislabeling using multimode hyperspectral imaging techniques. *Food Control*, *114*, 107234.
- Ros, E. (2010). Health benefits of nut consumption. *Nutrients*, *2*(7), 652–682.
- Sen, R., Sharma, S., Kaur, G., & Banga, S. S. (2018). Near-infrared reflectance spectroscopy calibrations for assessment of oil, phenols, glucosinolates and fatty acid content in the intact seeds of oilseed Brassica species. *Journal of the Science of Food and Agriculture*, *98*(11), 4050–4057.
- Shahidi, F., & John, J. (2013). Oxidative rancidity in nuts. In *Improving the safety and quality of nuts* (pp. 198–229). Elsevier.
- Sun, D.-W. (2010). *Hyperspectral imaging for food quality analysis and control*. Elsevier.
- Sun, J., Wang, G., Zhang, H., Xia, L., Zhao, W., Guo, Y., & Sun, X. (2020). Detection of fat content in peanut kernels based on chemometrics and hyperspectral imaging technology. *Infrared Physics & Technology*, *105*, 103226. <https://doi.org/10.1016/j.infrared.2020.103226>
- Tahmasbian, I., Xu, Z., Abdullah, K., Zhou, J., Esmailiani, R., Nguyen, T. T. N., & Hosseini Bai, S. (2017). The potential of hyperspectral images and partial least square regression for predicting total carbon, total nitrogen and their isotope composition in forest litterfall samples. *Journal of Soils and Sediments*, *17*(8), 2091–2103. <https://doi.org/10.1007/s11368-017-1751-z>
- Tahmasbian, I., Bai, S. H., Wang, Y., Boyd, S., Zhou, J., Esmailiani, R., & Xu, Z. (2018). Using laboratory-based hyperspectral imaging method to determine carbon functional group distributions in decomposing forest litterfall. *Catena*, *167*, 18–27.
- Tahmasbian, I., Xu, Z., Boyd, S., Zhou, J., Esmailiani, R., Che, R., & Bai, S. H. (2018). Laboratory-based hyperspectral image analysis for predicting soil carbon, nitrogen and their isotopic compositions. *Geoderma*, *330*, 254–263.
- Walton, D. A., Randall, B. W., Poienou, M., Nevenimo, T., Moxon, J., & Wallace, H. M. (2017). Shelf life of tropical Canarium nut stored under ambient conditions. *Horticulturae*, *3*(1), 24.
- Weinstock, B. A., Janni, J., Hagen, L., & Wright, S. (2006). Prediction of oil and oleic acid concentrations in individual corn (Zea mays L.) kernels using near-infrared reflectance hyperspectral imaging and multivariate analysis. *Applied Spectroscopy*, *60* (1), 9–16.
- Williams, P. J., Geladi, P., Britz, T. J., & Manley, M. (2012). Investigation of fungal development in maize kernels using NIR hyperspectral imaging and multivariate data analysis. *Journal of Cereal Science*, *55*(3), 272–278. <https://doi.org/10.1016/j.jcs.2011.12.003>
- Wold, J. P., Jakobsen, T., & Krane, L. (1996). Atlantic salmon average fat content estimated by near-infrared transmittance spectroscopy. *Journal of Food Science*, *61* (1), 74–77. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-0002825661&partnerID=40&md5=1be165060d2799c8b5d659af18f2d>



Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: A basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58(2), 109–130.

Xu, S., Zhao, Y., Wang, M., & Shi, X. (2017). Determination of rice root density from Vis–NIR spectroscopy by support vector machine regression and spectral variable selection techniques. *Catena*, 157, 12–23.

Yang, J., Pan, Z., Takeoka, G., Mackey, B., Bingol, G., Brandl, M. T., ... Wang, H. (2013). Shelf-life of infrared dry-roasted almonds. *Food Chemistry*, 138(1), 671–678.