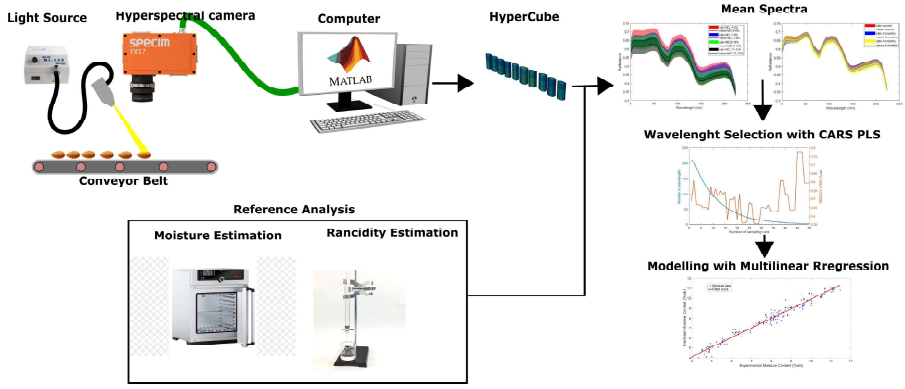
Graphical Abstract

Rancidity and Moisture Estimation in Shelled Almond Kernels using NIR Hyperspectral Imaging and Chemometric Analysis⋆

Brajesh Kumar Panda, Gayatri Mishra, Wilmer Ariza Ramirez, Hyewon Jung, Chandra B Singh, Sang-Heon Lee, Ivan Lee



Highlights

Rancidity and Moisture Estimation in Shelled Almond Kernels using NIR Hyperspectral Imaging and Chemometric Analysis

Brajesh Kumar Panda, Gayatri Mishra, Wilmer Ariza Ramirez, Hyewon Jung, Chandra B Singh, Sang-Heon Lee, Ivan Lee

- Hyperspectral imaging was used to develop non-destructive method for determination of MC, FFA and PV of almond.
- Full spectrum PLSR models predicted MC, FFA and PV with R_p^2 values of 0.957, 0.970, 0.955, respectively.
- CARS-PLS was used to select the feature wavelengths for rapid and accurate quantification.
- MLR models with feature wavelengths showed R_p^2 values of 0.941, 0.903, 0.886 for MC, FFA and PV, respectively.

Rancidity and Moisture Estimation in Shelled Almond Kernels using NIR Hyperspectral Imaging and Chemometric Analysis

Brajesh Kumar Panda[a,1], Gayatri Mishra[a,1,*], Wilmer Ariza Ramirez[a,1], Hyewon Jung[a], Chandra B Singh[a,b,*], Sang-Heon Lee[a], Ivan Lee[a]

[a]*STEM, University of South Australia, Mawson Lakes, 5095, South Australia, Australia*
[b]*Centre for Applied Research, Innovation and Entrepreneurship, Lethbridge College, Address Two, Lethbridge, T1K 1L6, Alberta, Canada*

Abstract

The current work focused on rapid and non-destructive determination of moisture content (MC), free fatty acids (FFA) and peroxide value (PV) in shelled almonds, using reflectance NIR-hyperspectral imaging (HSI). Sample set of 354 and 235 almond kernels were treated to vary the MC, rancidity (FFA and PV), respectively and used for model development. The sample set for each response was divided into calibration and testing set in the ratio of 70:30. Reference MC, FFA and PV were measured using wet anal-

*This work was supported by Cooperative Research Centres Projects program (CRCP-SIX000081), Australia and Surenut Pty Ltd. Financial support from the above sources is gratefully acknowledged. We also thank University of South Australia for providing necessary infrastructure facilities throughout the project duration.

*Corresponding author

Email addresses: Brajesh.Panda@unisa.edu.au (Brajesh Kumar Panda), Gayatri.Mishra@unisa.edu.au (Gayatri Mishra), Wilmer.Ariza@unisa.edu.au (Wilmer Ariza Ramirez), Hyewon.Jung@unisa.edu.au (Hyewon Jung), chandra.singh@lethbridgecollege.ca (Chandra B Singh), Sang-Heon.Lee@unisa.edu.au (Sang-Heon Lee), Ivan.Lee@unisa.edu.au (Ivan Lee)

[1]These authors had finish their affiliation with the University of South Australia and can be contacted at brajeshkumarpnd2@gmail.com, gayatri.mishra21@gmail.com, and wilmer.arizaramirez@utas.edu.au.

ysis methods and their association with the spectral data were modelled using partial least square regression. Superior models were obtained from the full-spectrum (900-1700 nm) data with R_p^2 values of 0.957, 0.970, 0.955 and $RMSE_P$ values of 0.41%, 0.108%, 0.453 mEq for MC, FFA and PV, respectively. Competitive adaptive reweighted sampling method was used to select the feature wavelengths for rapid quantification. Multiple linear regression models were developed using the feature wavelengths having good predictability (R_p^2 values: 0.941, 0.903, 0.886 and $RMSE_P$ values: 0.494%, 0.162%, 0.658 mEq for MC, FFA and PV, respectively). The current findings demonstrated great feasibility in industrial deployment of HSI technique for non-destructive estimation of moisture and rancidity indices in almonds and other nuts.

Keywords: Hyperspectral imaging, Moisture, Free fatty acids, Peroxide value, Non-destructive analysis

1. Introduction

Almond is the major tree nut consumed on earth, accounting nearly one third of all the tree nuts produced (inc, 2020). These are basically the fruits of almond tree (Prunus dulcis), grown for the seed purpose. Almonds are considered as superfood because of their rich nutritional profile containing high amount of mono-unsaturated fatty acids and polyunsaturated fatty acids, fibre, proteins, minerals and vitamins B and E (Yildirim et al., 2016). Owing to the growing health concern all around and nutritional advantages of almonds, consumers are now getting a range of option to enjoy almonds in various desserts like ice creams, chocolates, and sweets or in baked and

confectionery products (Cortés et al., 2018).

The USA, Spain, and Australia are the tabletop three exporters of almonds in the world (inc, 2020) hence carry the major responsibility of ensuring the nuts qualities to be exported. This leads to need for stringent regulations on accurate quality estimations which is driven by growing concern from the consumer end and the prevailing competitive market. One of the major quality concerns associated with the almond industry is rancidity caused due to lipid oxidation within the nuts (Canneddu et al., 2016). There could be multiple extrinsic and intrinsic factors causing rancidity such as high moisture content, temperature, relative humidity, insect damage, and fungal growth (Moscetti et al., 2013). However, high moisture in nuts during harvest, is reportedly one of the prime reasons behind initiating hydrolytic rancidity (Canneddu et al., 2016; Ghirardello et al., 2013). High moisture level encourages microbial growth, that produce certain enzymes responsible for hydrolyzing the fat to form free fatty acids (FFAs) (Moscetti et al., 2013; Wang et al., 2014). Development of FFAs compromise the flavor, odor, and taste of the almonds hence reduce the consumer acceptability. Along with FFAs value the other most used parameter for the estimation of rancidity in nuts is peroxide value (PV) which indicates the stability of the nuts against oxidation (Borompichaichartkul et al., 2013).

Several wet lab analytical techniques are recommended for estimation of fat quality (Association of Official Agricultural Chemists & Horwitz, 1975; Möller, 2010; ISO 17059:2019) in nuts. Although these classical techniques are quite reliable and accurate, have several downsides as well such as destructive sample preparation, involvement of hazardous chemicals, time consum-

ing, require large batches of samples, and limitation of accessing intrasample variabilities (ISO 17059:2019; Bai et al., 2018; Han et al., 2021; Carvalho et al., 2019). Therefore, these classical analytical methods are losing their relevance in food industries, and a complete automation and real-time control during process monitoring are the need of the moment. In the recent times, hyperspectral imaging (HSI) has emerged as a promising tool for non-destructive quality estimation of food commodities (Mahesh et al., 2015; Huang et al., 2014; Elmasry et al., 2012; Sricharoonratana et al., 2021; Khamsopha et al., 2021).

HSI can operate across a wide range of wavelengths (780-2500 nm) depending upon the chemical nature of the food commodity. The specific one operates in the wavelength range of 900-1700 nm termed as shortwave infrared HSI (SWIR-HSI). SWIR-HSI acquires both spatial and spectral information of the captured images and combine them to provide unique chemical fingerprint to each pixel of the captured image (Elmasry et al., 2012). This feature is certainly useful in dealing with the critical concern around food heterogeneity (Cruz-Tirado et al., 2020). Moreover, this technique is capable of non-destructive evaluation of chemical composition in very short time. Interestingly, some of the recent trials carried out for quality estimation of nuts have shown promising outcome in favor of HSI technique (Faqeerzada et al., 2020; Bilal et al., 2020; Torres et al., 2021; Moscetti et al., 2015; Jin et al., 2016). HSI-based analysis produces high dimensional data set which demands high computational complexity to carry out the analysis (Huang et al., 2014). This situation is handled with chemometrics which allows to reduce the dimensionality of the data, retain essential spectral information,

and classify or quantify important areas of an image (Gowen et al., 2007; Amigo et al., 2008; Mishra et al., 2021).

Although HSI-based analysis has predominantly been applied for quality estimation of agricultural produce such as fruits, vegetables, and meats, but for nuts quality it is relatively underexplored (Suktanarak & Teerachaichayut, 2017; Teerachaichayut & Ho, 2017). Moreover, the current systems employed for nuts sorting rely on random sampling that decides the market value of entire batch leading to huge losses (Canneddu et al., 2016). More recently, Faqeerzada et al. (2020) and Torres et al. (2021) have demonstrated applicability of HSI on quality estimation of almond powder and shelled almonds, respectively. However, no work in the past have addressed the non-destructive estimation of both rancidity and moisture content in shelled almonds, to the best of our knowledge. Hence, in the current study SWIR-HSI was used to investigate the rancidity and moisture content in shelled almonds. The major objectives formulated in this study were to:

- develop a full spectrum SWIR-HSI regression model for accurate prediction of rancidity indices (i.e., FFAs and PV) and mc (%wb),
- select the feature variables for development of multispectral regression model for rapid quantification,
- and visualize the spatial distribution of FFAs, PV, and mc in almond kernels.

2. Materials and Methods

2.1. Almond sample preparation

Freshly harvested and cleaned batch of almonds of variety "Nonpareil" were collected from the almond processing company Lindsay point (Victoria, Australia) for the experiment. Almond samples were collected in 2019-2020 season. On arrival, the samples were packed in zip lock bags and stored in refrigerator at 4°C until sample preparation, scanning and reference analysis. To develop robust calibration model for prediction of moisture and rancidity, the level of moisture content and rancidity level were varied by moisture conditioning and accelerated storage, respectively.

Initial moisture content of the fresh almonds was determined by oven drying method (AOAC, 2019) and found to be $4.1\% \pm 0.15$ (wb). One kilogram each batch of almond kernels was conditioned to four different moisture contents, viz. 6, 8, 10, and 12% (wb) by adding calculated amount of distilled water and thorough mixing. The conditioned almond kernels were kept in zip lock low density polyethylene (LDPE) bags and stored inside moisture lock compartment of refrigerator for a week in order to achieve uniform moisture distribution within the kernels. The final moisture content of the kernels was determined after one week of treatment by gravimetric analysis (AOAC, 2019). Six almond kernels of similar moisture content were placed in a line for scanning of each hyperspectral image. Seventy five images of each five moisture content levels were scanned to build image library of 375 almond images for development of the moisture regression model. During calibration of the full-spectrum moisture prediction model, 21 outliers were removed to get final image library of 354 almond images.

Rancidity in the samples was achieved through accelerated aging of the almond kernels under 75% relative humidity and 40°C temperature (Zacheo et al., 1998). The almond kernels were divided into two groups, each weighing one kilogram were stored separately inside vacuum desiccators containing saturated sodium chloride (NaCl) solutions at the base for maintaining 75% relative humidity (Young, 1967). The desiccators were stored inside an environmental chamber maintained at 40°C and samples were removed after 3 months and 4 months of storage. The rancid samples were packed in zip lock LDPE bags and kept in refrigerator for scanning and reference analysis. Six almond kernels of same storage duration were placed in a line for scanning of each hyperspectral image. Eighty images of each three storage durations were scanned to build image library of 240 almond images for development of the rancidity regression models. During calibration of the full-spectrum rancidity prediction models, 5 and 7 outliers were removed to get final image library of 235 and 233 almond images for FFA and PV prediction. Free fatty acid content and peroxide values were considered as the indicators of rancidity of the almond samples.

2.2. Hyperspectral image acquisition

The hyperspectral images of almond samples were acquired by using a short-wave infrared (SWIR) hyperspectral imaging system. The system included a InGaAs camera (Specim FX17, Spectral Imaging Ltd., Oulu, Finland) with the spectral range of 900–1700 nm with spatial resolution of 640 pixels, 224 wavebands and spectral resolution of 8 nm full width half maximum (FWHM), one 150-watt halogen light source equipment with a fiber optic line light (Illumination, Technologies Inc. USA) to illuminate the cam-

era's field of view, an one-axis linear conveyor belt, and a computer (Dell, Inter(R) Core (TM) i5-2400 CPU@ 3.10 GHz, 16GB RAM) with a software called Lumo scanner (Spectral Imaging Ltd., Oulu, Finland) which controlled the system to acquire the hyperspectral images of the almond samples. Six almond kernels at a time were placed on the conveyor belt platform to be scanned with appropriate resolution, which can be maintained by moving the belt at correct speed and by maintaining correct exposure time. The halogen light source intensity was set to 60% and the exposure time was set to 12 ms during scanning. The distance between almond kernels and the lens was about 10 cm. The moving speed of conveyor belt was maintained at 15 mm/sec to obtain a hypercube of dimension $1200 \times 640 \times 224$.

A dark current image ($\mathbf{I_D}$) and white reference image ($\mathbf{I_W}$) were obtained after scanning of each image by covering the camera lens and using a 99% reflectance standard (Spectralon™, SRS-99-10, Labsphere, Inc., North Sutton, NH, USA), respectively. Noise signals present in the raw hyperspectral images ($\mathbf{I}$) were corrected using equation (1) (Sun et al., 2020).

$$\mathbf{I} = \frac{\mathbf{I_{Raw}} - \mathbf{I_D}}{\mathbf{I_w} - \mathbf{I_D}} \tag{1}$$

2.3. Hyperspectral image segmentation and spectral data extraction

Segmentation was applied to the corrected images for removal of the background and for extraction of the Region of Interest (ROI) in each sample using MATLAB 2019b (MathWorks Inc., Natick, MA). The difference between the background and almond pixels were obtained from the HSV (hue, saturation, value) color space of the images and then a threshold value of 0.45 was applied to the resultant image. A binary mask was then created

for each sample, assigning 0 value for the background and 1 for the pixels corresponding to almonds. The mask was then applied to each hyperspectral image cube before spectral data extraction. Mean spectra of each image were extracted by averaging index values of each almond pixel, producing a total of 224 spectral data points. Final spectral matrix for development of models were obtained by concatenating the spectral data of all the almond samples. In the final spectral matrix, the row represents the number of almond images and column represents number of wavelengths. Therefore, we obtained two spectral matrices of dimension 354×224 and 235×224 for moisture and rancidity value prediction, respectively.

2.4. Spectral data pre-processing

Prior to development of the multivariate regression model, pre-processing and outlier removal techniques were performed to improve the signal to noise ratio by reducing undesirable effects such as light scattering, baseline shifts, random noise, and uncontrolled external factors (Bai et al., 2018). MATLAB 2019b (MathWorks Inc., Natick, MA) was used to pre-process all the spectral data of the scanned almond images. The signal-to-noise ratio was improved and useful information for models were extracted by pre-processing the spectral data. In this study, five different pre-processing methods, viz. standard normal variate (SNV) and Savitzky–Golay (SG) smoothing, SG+SNV, SG+first derivative, SG+second derivative were used to pre-process the raw spectral data (Mishra et al., 2020). Suitable pre-processing techniques were selected based on the full spectrum model performance. SNV was used to reduce the baseline shift produced due to the multiplicative effects of light scattering during the interaction between infrared (IR) radiation and the

sample particles. A constant offset term was removed by subtracting the mean value of the full spectrum using SNV, which brought all spectra to the same scale by subsequent division by the standard deviation of the full spectrum (Grisanti et al., 2018). Moreover, Savitzky–Golay filter was applied to smoothing the spectrum in addition with 1st and 2nd derivative, which enhanced the difference between the reflectance values.

2.5. Reference analysis

Each sample set of almonds after hyperspectral image acquisition were ground in a laboratory mixer and packed in zip lock LDPE bags, individually. The ground almond samples were used in the measurement of moisture and rancidity analysis as described below.

2.5.1. Moisture analysis

Ground almond samples were analyzed for moisture content using standard gravimetric methods described by AOAC (2019). Approximately five grams of ground almond for each sample in duplicates were placed in pre-weighed aluminium moisture box and put into a drying oven (DHG-9101-2SA; Shanghai Sanfa Scientific Instruments Co., Ltd, Shanghai, China) maintained at 130°C for 3 hours. After drying, the moisture boxes were immediately cooled in a vacuum desiccator containing allochronic silica gel for 1 hour. The weight of the ground almond samples was recorded before and after drying using weighing balance (company and sensitivity), and the moisture content was determined by the mass loss after drying. Wet basis moisture content was calculated using the following equation:

$$\text{Moisture content}(\%wb) = \frac{W_2 - W_1}{W_3 - W_1} \times 100 \tag{2}$$

where W_1 is the weight of empty moisture box, W_2 and W_3 are weights of moisture box with almond samples before and after drying, respectively. All weights are in grams.

2.5.2. Rancidity analysis

Free fatty acid (FFA) content and peroxide values (PV) were used as rancidity index for almond samples in present study. Almond oil from each set of samples were extracted using a hydraulic press and the extracted oils were immediately analyzed for the rancidity indices. For FFA analysis, approximately 2 g of oil (W) was extracted from almond samples in duplicates were placed into a 100 mL Erlenmeyer flask and mixed with 10 mL ether, 10 mL 95% ethanol, 2 to 3 drops of 1% phenolphthalein indicator. The mixture was titrated with 0.005 mol/L sodium hydroxide ethanol solution until the mixture turns pink for at least 1 min. The volume of the titrant consumed was recorded as V. FFA content was calculated using the following equation, where C is the calibrated concentration of sodium hydroxide ethanol solution and 282 is oleic acid molar mass in g/mol:

$$FFA(\%) = \frac{V \times C \times 282}{1000 \times W} \times 100 \tag{3}$$

Peroxide value was estimated by using the titration method provided in AOAC Official Method 965.33 (AOAC, 2019) with a slight modification to be used for microtitration. Approximately, 2 g of oil sample (W) in duplicates were placed into a 100 mL Erlenmeyer flask and mixed with 12 mL acetic acid–chloroform (3:2) solution, and 0.5 mL saturated potassium iodide (KI) solution. The solution was allowed stand with occasional swirling for 1 min and then 12 mL distilled water was added. After 1 min, the solution was

slowly titrated with 0.01-N sodium thiosulfate ($Na_2S_2O_3$). Continue titrating until the yellow color goes away. 0.5 mL of 1% soluble starch indicator was then added to change the color from light yellow to purple. The solution was again titrated with sodium thiosulfate ($Na_2S_2O_3$), dropwise until the purple color just disappears. The PV as mEq of oleic acid per kg of oil was calculated according to following equation:

$$PV = \frac{(V \times C \times 1000)}{W} \tag{4}$$

where V is the volume of $Na_2S_2O_3$ solution in mL, C is the concentration of the $Na_2S_2O_3$ solution in mol/L, and W is the weight of the oil sample taken in g.

2.6. Full spectrum calibration model development

PLSR model was used to establish the relationship between the spectral data and the response, viz. moisture, FFA and PV. For each parameter, SG-second derivative pre-processed spectral data ($\mathbf{X}$) and the reference response values ($\mathbf{Y}$) were used for the regression model development. The sample set for each response was divided randomly into two categories, viz. calibration set, and testing set. The raw spectra were projected onto several latent variables (LVs) for simplifying the association of spectra and the attribute matrixes. The LVs maximize the covariance between the two matrices of the original data and significantly reduce the original spectral dimension. The ideal number of LVs is decided by cross validation when the root mean square error of cross-validation ($RMSE_{CV}$) reaches a minimum. In this study, 10-fold cross-validation was conducted and the developed model was tested by the prediction set.The association between spectra matrix $\mathbf{X}$ ($n \times m$) and

quality attribute $\mathbf{Y}$ $(n \times 1)$ can be predicted using equations (5) to (7) (Wold et al., 2001; Guebel & Torres, 2013).

$$\mathbf{X_{(n;p)}} = \widehat{\mathbf{X}}_{\mathbf{(n;p)}} + \mathbf{E_{(n;p)}} = \mathbf{T_{(n;k)}} \mathbf{P^T_{(n;p)}} + \mathbf{E_{(n;p)}} \tag{5}$$

$$\mathbf{Y_{(n;m)}} = \widehat{\mathbf{Y}}_{\mathbf{(n;m)}} + \mathbf{F_{(n;m)}} = \mathbf{U_{(n;k)}} \mathbf{Q^T_{(k;m)}} + \mathbf{F_{(n;m)}} \tag{6}$$

$$\mathbf{U_{(n;k)}} = \mathbf{T_{(n;k)}} \mathbf{C_{(k;k)}} \tag{7}$$

Where $\widehat{\mathbf{X}}$ and $\widehat{\mathbf{Y}}$ give the PLS estimations of the matrices $\mathbf{X}$ and $\mathbf{Y}$ respectively, $\mathbf{E}$ and $\mathbf{F}$ are the residual error matrices, $\mathbf{T}$ and $\mathbf{U}$ are the score matrices, $\mathbf{P}$ and $\mathbf{Q}$ are the loading matrices. After this calculations the matrix B_{PLS} of regression coeffiennts can be calaculated with crefeqbeta3,eqbeta4

$$\mathbf{B_{PLS}} = \mathbf{W}(\mathbf{P^T W})^{-1} \mathbf{C Q^T} \tag{8}$$

$$\widehat{\mathbf{Y}} = \mathbf{T C Q^T} = \mathbf{X B_{PLS}} \tag{9}$$

Developed PLSR models were then tested using the mean spectra of the testing data set. The predictability of the developed models was evaluated by using the coefficient of determination (R^2) and root mean square error $(RMSE)$ using the following equations:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (Y_i - \widehat{Y})^2}{\sum_{i=1}^{n} (Y_i - \widehat{Y})^2} \tag{10}$$

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(Y_i - Y)^2}{n}} \tag{11}$$

The co-efficients of determination of calibration (R^2_c), cross-validation (R^2_{cv}) and prediction (R^2_p) and the corresponding root mean square error of

calibration ($RMSE_C$), root mean square of cross validation ($RMSE_{CV}$), and prediction ($RMSE_P$) were used for evaluating developed PLS models. Best fit models have high R_c^2, R_{cv}^2 R_p^2 and low $RMSE_C$, $RMSE_{CV}$ and $RMSE_P$ together with slight differences among them. It always expects R^2 to go to 1 and $RMSE$ to go to 0. Furthermore, the robustness of the model was evaluated by calculating the residual predictive deviation (RPD), which is the ratio between standard deviation (SD) of the population's reference value and the standard error of prediction of the cross-validated data set (Equation (12)). RPD provides standardization of prediction accuracy of the model (Mishra et al., 2018). Williams & Sobering (1996) suggested that RPD values of 2.4–3.0 are poor; $>$ 3 indicate that the model is appropriate for screening and is robust.

$$RPD = \frac{SD}{RMSE} \tag{12}$$

2.7. Feature wavelength selection

The acquired hyperspectral images contained 224 number of wavelengths, which is of high dimension and involves multi-collinearity resulting in increase in the speed of the acquisition. The high dimensional hyperspectral data is not suitable for commercial application, where the estimation speed should match with the production capacity. To improve the calculation speed and prediction accuracy, some sensitive wavelengths were selected from the full spectral matrices to compress hyperspectral data. These feature wavelengths must carry most of the useful information which could perfectly predict the moisture content and rancidity indices of the almond samples. Competitive adaptive reweighted sampling (CARS) technique was used for selection of key wavelengths. To perform CARS, MATLAB 2019b (MathWorks Inc., Natick,

MA) was used.

CARS is a useful feature selection technique, works through selecting N subsets of independent variables by N sampling runs in an iterative manner and develop a robust PLSR (Tao et al., 2019). It chooses the subset with the lowest root-mean-square error of cross validation ($RMSE_{CV}$) as the optimal subset. In this manuscript, the optimal wavelengths imply the feature wavelengths with large absolute coefficients in the developed PLSR model. CARS have five basic steps involved during each sample analysis: (1) divide the sample set into Monte Carlo technique to the sample set, (2) build PLS model at each sampling run (3) utilize exponentially decreasing function (EDF) for wavelength selection; (4) adopt reweighted sampling for wavelength selection in a competitive manner; and (5) conduct cross validation to evaluate the prediction capability of the selected wavelengths. In this study, 10-fold cross validation was used to evaluate the feature wavelengths, and the number (N) of Monte Carlo sampling runs was set to be 50. The process was repeated The PLSR model was re-developed using the reflectance value of feature wavelengths to predict the moisture and rancidity indices of the almond samples, rapidly. Then the developed model was tested by applying on the external testing tests.

2.8. Development of multiple linear regression (MLR) model using feature wavelengths

MLR models were established for MC, FFA and PV using the selected wavelengths from CARS, without any pretreatment in order to simplify prediction process and to improve calculation speed. Testing data set were used for development of calibration model and were validated using 10-fold

cross-validation process. Testing data sets were used to predict the MC, FFA and PV of almond kernels. R_c^2, R_{cv}^2, R_p^2 and the corresponding $RMSE_C$, $RMSE_{CV}$, and $RMSE_P$ values were used for evaluating developed MLR models.

2.9. Classification of almonds according to rancidity

Support vector machine classification (SVM) were performed to classify almonds based on rancidity parameters, viz. FFA value and PV using the feature wavelengths. Two class classification was done considering class A as fresh having FFA value <1% and PV<5 Meq/kg and class B as rancid having FFA value >1% and PV>5 Meq/kg. SVM is a machine learning technique used to handle nonlinear input spaces. SVM constructs a hyperplane in multidimensional space to separate different classes. SVM generates optimal hyperplane in an iterative manner, which is used to minimise an error. The core idea of SVM is to find a maximum marginal hyperplane (MMH) that best divides the dataset into classes (Cruz-Tirado et al., 2021). The calibration set was used to build the model with 10-fold internal cross-validation and prediction was performed using 30% of remaining test samples. R_c^2, R_{cv}^2, R_p^2 and the corresponding $RMSE_C$, $RMSE_{CV}$, and $RMSE_P$ values were used for evaluating developed MLR models.

2.10. Visualization of MC and rancidity distribution

Visualization of the moisture and rancidity distribution in a batch of sample, the calibration model was first built using the mean spectrum of all pixels of each seed and the reference MC obtained by the gravimetric method as described in the previous sections. Then the MC value, FFA and

PV content of each pixel were predicted by the developed PLS model. MC and rancidity distribution maps were created with a pseudo color bar and the spatial distribution of MC, FFA and PV can be assessed by observing color change in a distribution map. Pseudo color maps for quality distribution were developed using MATLAB 2019b (MathWorks Inc., Natick, MA).

3. Results and Discussion

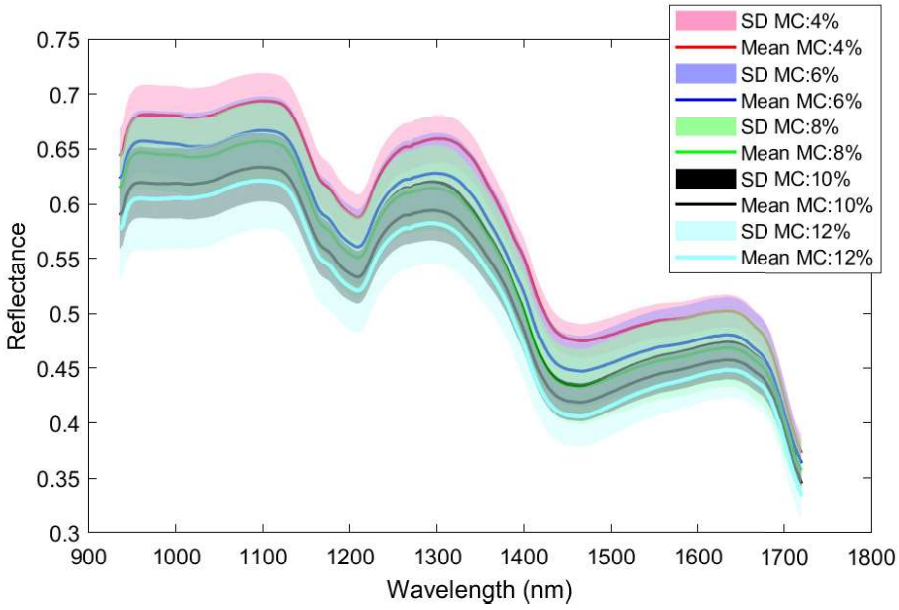
3.1. Spectral features

The reflectance spectral library obtained from almond samples with five different MC levels, viz. 4, 6, 8, 10, 12% (wb) and three different rancidity levels, viz. control, stored for 3 months and four months are shown in figure 1a and figure 1b, respectively. The most https://www.overleaf.com/project/615b465d52462e38a wavebands existing in the NIR region are the strong overtone and combination absorptions of hydrogen containing bonds O–H (found in water), C–H (found in carbohydrates & oil), and N–H (found in protein). The reflectance values gradually decreased with the increase in MC throughout the spectra (Figure 1a) (Xu et al., 2019). In figure 1a, the whole spectra are found to be varied with varying moisture content as water greatly absorbed NIR radiation and O-H bond dominated the NIR spectra of kernels. The luminous energy from the NIR electromagnetic waves is transferred to the elastic potential energy (stored in the form of bending O–H bonds) and chemical potential energy (stored in the form of stretching O–H bonds). Due to increase moisture content in the kernel, more energy from electromagnetic waves is absorbed, reflecting less energy back into the lens of the hyperspectral camera (Deng et al., 2015). Thus, it interferes with the reflectance characteristics

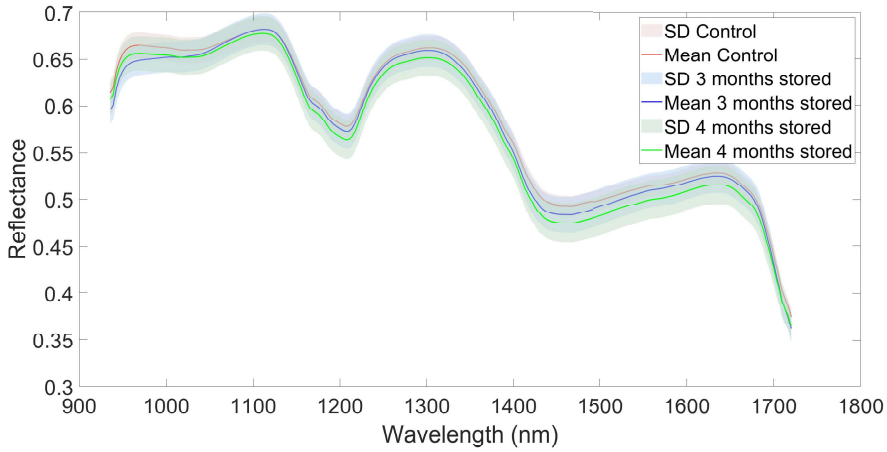
of other molecules present in the kernel. The absorption peaks at 960 nm and valleys at 1450 nm are associated with the second and first overtone of O–H stretches in water, respectively (Zhang & Guo, 2020). The reflectance peaks or valleys of rancid almond spectra increased at specific wavebands, viz. 1210, 1450, and 1638 nm with increase in the rancidity level (figure 1b), however all other peaks were not showing any relationship among them. The absorption bands at 1200-1250 nm are related to C-H stretch second overtone and the spectral region between 1600 and 1650 nm is related to the first overtone of C–H vibration, influenced by the presence of CH3 bonds of fats present in the almond kernels (Cruz-Tirado et al., 2020). The absorption peak at 920-950 nm probably related to the combination effect of third overtone of band C-H and fat (Sun et al., 2020). The change in the spectral features in the rancid almond samples were observed might be due to oxidation of the complex fat into free fatty acids and peroxides because of storing the almonds in a high humidity and temperature condition (Hong et al., 2018). Consequently, more lipid degradation reported as the storage duration increased from 3 months to 4 months. The variation in the absorption peaks of rancid almonds at 960, 1150 and valleys at 1450 nm might be due to the O-H stretching, indicating that there was a reduction in moisture content during storage. Similar spectral variation was also found by Carvalho et al. (2019) for macadamia nuts.

3.2. Full-spectrum PLS models of MC

Final moisture content obtained after conditioning were vary from 4.52±0.45% (wb) for control samples to 11.44±0.5% for samples conditioned to 12%(wb) (table 1). Based on the all the pre-processed spectral data extracted from



(a)



(b)

Figure 1: Mean reflectance spectra with standard deviation obtained from almond kernels of (a) treated moisture and (b) stored in accelerated conditions (N.B: SD=Standard deviation; MC=moisture content in % wet basis)

the hyperspectral images of almond kernels and their corresponding reference MC values, the calibration models with the whole spectral wavelengths were established using PLSR algorithms. table 2 shows the calibration, cross-validation, and prediction outcome of the above two established models. Calibration plot obtained for training MC is shown in figure 2a. The model performances were given in terms of R_c^2, R_{cv}^2, R_v^2 and corresponding $RMSE_C$, $RMSE_{CV}$, and $RMSE_P$ values. SG+1st derivative was found to be the best pre-processing techniques for prediction of MC, with seven LVs, R_p value of 0.957 and $RMSE_P$ value of 1.702%. The model developed using SG+1st derivative also had high RPD value of 5.07, showing the robustness of the model. All other pre-processing techniques showed poor performance having low RPD values as compared to the raw spectra. The reason behind poor performance on application of SNV pre-processing technique alone or together with SG filter might not properly normalise the spectra since within-class mean spectrum might be inappropriate in the particular data set. Thus, the applied pre-processing might cause interaction of the unimportant information, resulting non-orthogonal nature of the data. The prediction result obtained from the spectral data indicated that SG filter with combination of derivation pre-processing techniques improved the PLS model predictive accuracy, which might be due to smoothing of the spectra and enhancement of the difference between them. The prediction error for natural moisture content was always much below 0.3% for a range of 4-5%, whereas for laboratory treated almond kernels, the prediction error was 0.6% for a moisture range of 11-12% and thus, the average prediction error was found to be 0.410. The capability of HSI for moisture prediction is consistent with previous studies

of (Xu et al., 2019) and (Sun et al., 2020) for who reported higher prediction accuracy for cucumber seeds and peanuts with R_p^2 values of 0.917 and 0.943, respectively.

Table 1: Statistics of moisture content and rancidity indices of various sample sets

Sample set	No. of samples	Min	Max	Mean	SEM	CV
MC (% wb)						
Total	354	4.00	11.44	7.72±2.01	0.10	0.26
Calibration	250	4.00	11.44	7.66±1.99	0.12	0.25
Testing (Prediction)	104	4.40	11.44	7.89±2.05	0.20	0.27
FFA (%)						
Total	235	0.20	1.62	1.23±0.51	0.03	0.42
Calibration	165	0.20	1.62	1.26±0.53	0.04	0.44
Testing (Prediction)	70	0.20	1.61	1.14±0.48	0.05	0.37
PV (Meq OAA)						
Total	233	1.91	7.62	3.43±2.01	0.13	0.58
Calibration	165	1.91	7.62	3.62±2.00	0.15	0.59
Testing (Prediction)	68	1.91	6.99	2.97±2.03	0.24	0.58

N.B. Mean±SD indicates the average value with standard deviation among the sample set;SEM=standard error of mean; CV=Co-efficient of variation

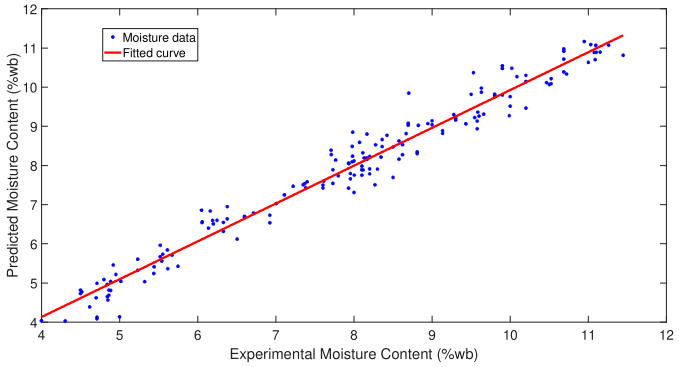
3.3. Full-spectrum PLS models of rancidity indices

Rancidity of the stored samples were calibrated according to the FFA and PV values of each storage period. FFA values of the stored samples were increased drastically after 3 months of storage, but peroxide values were

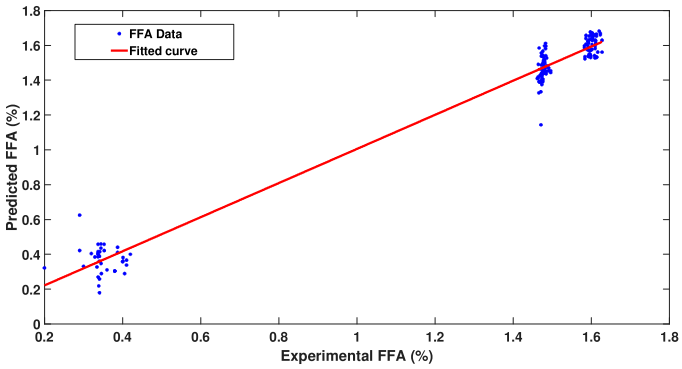
Table 2: PLS prediction results of almond MC and rancidity with different pretreatments

	Pre-Processing Methods	LVs	Calibration		Cross-validation		Prediction		RPD
			R_c^2	$RMSE_C$	R_c^2v	$RMSE_{CV}$	R_p^2	$RMSE_P$	
Moisture content (%wb)	Raw	9	0.967	0.375	0.944	0.464	0.943	0.441	4.550
	SNV	9	0.961	0.384	0.947	0.443	0.959	0.432	4.070
	SG	10	0.917	0.567	0.900	0.607	0.859	0.778	3.310
	SG_SNV	8	0.900	0.627	0.881	0.653	0.884	0.698	3.150
	SG_1st derivative	7	0.973	0.328	0.966	0.366	0.957	0.410	5.070
	SG_2nd derivative	7	0.970	0.347	0.954	0.432	0.946	0.451	4.330
FFA (%)	Raw	10	0.958	0.101	0.945	0.108	0.931	0.123	4.020
	SNV	8	0.970	0.091	0.954	0.111	0.927	0.124	4.93
	SG	9	0.975	0.082	0.972	0.083	0.970	0.053	5.600
	SG_SNV	7	0.983	0.066	0.977	0.074	0.958	0.108	5.840
	SG_1st derivative	7	0.971	0.087	0.959	0.102	0.948	0.113	5.060
	SG_2nd derivative	8	0.976	0.079	0.958	0.015	0.943	0.122	5.050
PV (Meq OAA)	Raw	10	0.943	0.463	0.927	0.508	0.934	0.535	3.940
	SNV	9	0.950	0.447	0.925	0.543	0.928	0.533	3.740
	SG	8	0.974	0.313	0.969	0.337	0.941	0.410	4.870
	SG_SNV	8	0.975	0.349	0.962	0.382	0.955	0.453	5.630
	SG_1st derivative	9	0.952	0.453	0.934	0.519	0.920	0.506	4.410
	SG_2nd derivative	8	0.973	0.328	0.949	0.439	0.907	0.615	3.840

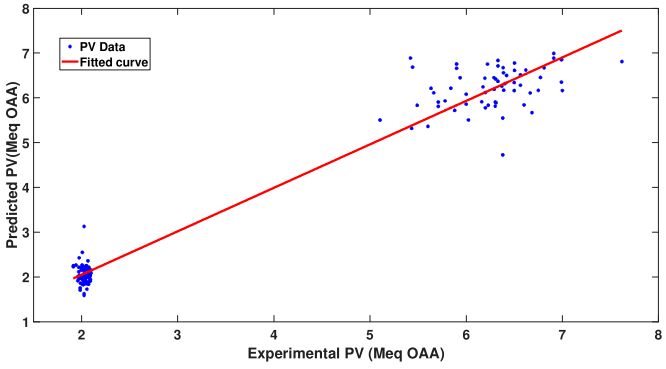
LV=latent variable; R_c^2, R_{cv}^2 and R_p^2=regression co-efficients of calibration, cross-validation and prediction, respectively; $RMSE_c$, $RMSE_cv$, $RMSE_p$=root mean square error of calibration, cross-validation and prediction, respectively; RPD=residual predictive deviation



(a)



(b)



(c)

Figure 2: Experimental vs predicted results obtained from the full spectrum PLSR models for (a) MC, (b) FFA and (c) PV

not increased significantly even after 3 months of storage; however, after 4 months of storage PV content was also increased above 5 Meq OAA. FFA value obtained after 3 and 4 months storage was about 1.47±0.006% and 1.60±0.010%, respectively. Moreover, the PV of the almonds were found to be 2.03±0.03% and 6.27±0.43%, respectively for 3 and 4 months storage samples. The results for FFA and PV model development using HSI in intact almond kernels are reported in table 2. It was found that all the pretreatment techniques improved the prediction capability of the calibration models by removing nonlinear spectroscopic artefacts, viz. drifting baselines, scattering and noise involved in the spectra. RPD values of all the models using different pre-processing techniques are found to be greater than the raw spectral model. The SG smoothing combined with SNV pre-treatment gave the best model performance for both FFA and PV due to removal of high-frequency noise during the normalisation process. Higher accuracy was found using SNV pretreatment might be due to correction of baseline shift in the dataset resulting from light scattering. Calibration plot showing experimental rancidity indices vs predicted indices, using the best pre-processing technique obtained are shown in figure 2b and figure 2c. Results reported in figure 2b and figure 2c should be considered with caution, given the lack of samples with intermediary values of reference, as this could probably caused a biased prediction model with two major regions of concentration of the reference data. The best full spectrum model for FFA resulted R^2 values of 0.983, 0.977, and 0.958 for the calibration, cross-validation, and prediction datasets, respectively. The optimized calibration model for PV resulted R^2 values of 0.975, 0.962, and 0.955 for the calibration, cross-validation, and

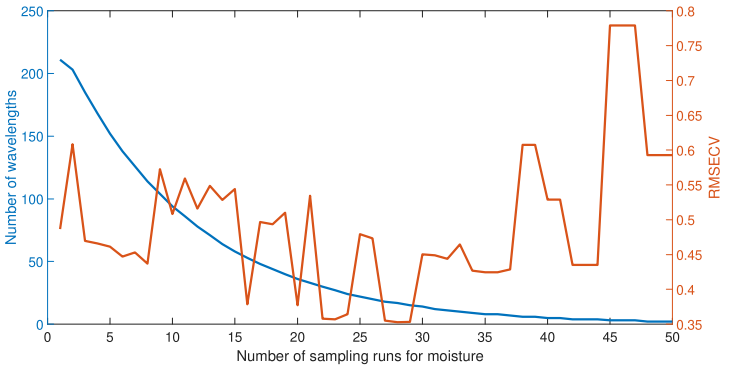
prediction datasets, respectively. The corresponding $RMSE_P$ and RPD values for both FFA and PV were found to be 0.108% and 0.453 Meq OAA, and 5.84 and 5.63, respectively indicating excellent performance of the developed PLS models. Previous research on determination of lipid oxidation using HSI also described its higher predictive capability. Han et al. (2021) reported application of HSI coupled with convolutional neural network (CNN) for PV estimation in canarium nut having highest overall accuracy of 93.48%. Xia et al. (2016) and Ma et al. (2015) also reported superior prediction and classification efficiency of multispectral imaging for butter cookies and sunflower seeds, respectively. They used selective wavelengths for determination of rancidity indices and found peroxide value prediction accuracy of 97.1% and rancid kernels classification accuracy of $> 97\%$, respectively.

3.4. Feature wavelength selection

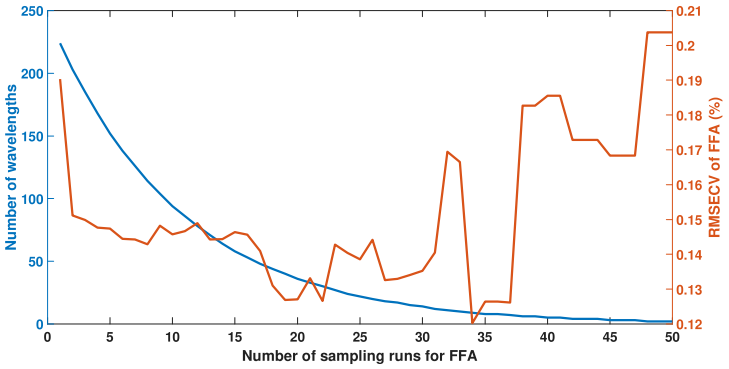
The selected wavelengths were compared with the β-coefficient values extracted from the full spectrum PLS models. Figure 3 shows the feature wavelength selection for curves for MC, FFA and PV. The blue curves in Figure 3 indicating the decrease in the number of wavelength variables as the number of sampling runs increases. The declining trend was found to be gradually slowing down, which showed the fast and the refined selection of wavelengths. The red curves in Figure 3 show the variation of $RMSE_{CV}$ values with the number of sampling. As the elimination process of redundant wavelengths was increased, $RMSE_{CV}$ values were first reduced moderately and then gradually increased with the elimination of feature wavelengths. The optimum wavelengths were chosen based on the minimum $RMSE_{CV}$ values. For moisture calibration set, the $RMSE_{CV}$ value was lowest at the

29th sampling and the corresponding number of feature wavelengths were 7. Similarly, for rancid almond sample set, least $RMSE_{CV}$ value was found at the 34th and 42nd sampling for FFA and PV, respectively and the optimum number of feature wavelengths selected were 5 for both the parameters. The feature wavelengths were selected from the optimal subset based on the highest weight coefficients applying "survival of the fittest" principle. The specific effective wavelengths selected by CARS technique were listed in table 3. The selection of the most important wavelengths was also compared with the β-coefficients of these PLS regression prediction models. Only those wavelengths with large β-coefficient values are useful for the prediction and are shown in figure S1.

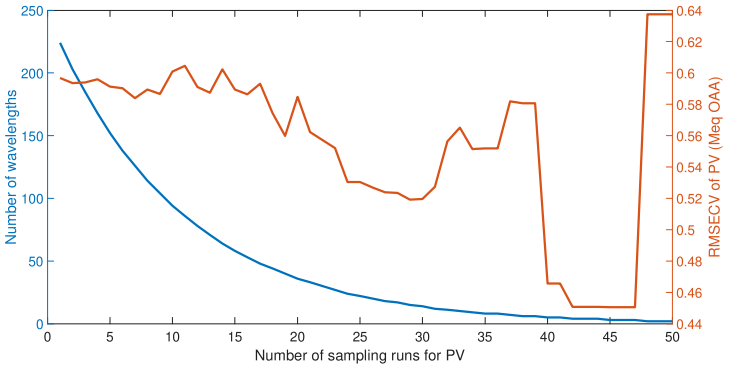
For moisture prediction, the feature wavelengths were 970, 1001, 1154, 1312, 1350, 1437, 1670 nm. Previous research on MC prediction for corn kernels described the major wavelengths at 970 nm, 1150 and 1450 nm, related to the 2nd, 3rd and 1st overtone of O-H stretching, respectively (Zhang & Guo, 2020). Our results were also comparable with the study of Liu et al. (2014), who reported 1178, 1299, 1363, 1389, 1643, 1788, 1941, 2213, 2439 nm were the important wavelengths selected for MC determination in pork samples, and their regression model showed MC prediction with a test R^2 of 0.920. The strongest feature wavelengths for prediction of FFA and PV were 945, 994, 1105, 1382, 1527 nm and 945, 1025, 1393, 1488, 1695 nm, respectively. From β coefficient graph of FFA and PV, small peaks were found at 1210 nm, which may also relate to the change in C-H bond present in the almond lipids. The results were like the findings of Caporaso et al. (2018), who reported the important peaks for lipid content prediction lie at



(a)



(b)



(c)

Figure 3: The result of variable selection by CARS for (a) MC (b) FFA and (b) PV in almond kernels (N.B: RMSECV=Root mean square error of cross-validation; MC=Moisture content; FFA=free fatty acids; PV=peroxide value; Meq OAA=Milli equivalent oleic acid)

1210, 1360, 1700 nm, mostly arise from the first and second overtone of C-H and stretching of the -CH2 groups. Khodabux et al. (2007) also reported the absorption peaks around 1160 nm was attributed to -HC=CH- bonds. Feature wavelength of 1488, 1695 nm for PV prediction might be arise due to the secondary vibration resulted due to stretching of O-H bond present in hydroperoxides in the samples.

3.5. MLR model development, validation, and testing

Table 3 shows the performance of MLR models built for MC, FFA and PV using the selected wavelengths. The use of lower number of bands led to a satisfactory performance compared to the PLS regression model reported in the previous section for moisture using the full spectra range. In fact, the moisture model had a $RMSE_C$ of 0.452% and $RMSE_P$ of 0.494%, which is comparable to the full spectrum PLS regression model, having $RMSE_P$ value of 0.410. Similarly, $RMSE_P$ values of 0.162% and 0.658 Meq OAA were obtained from MLR model of FFA and PV, which are bit high as compared to the full spectrum model, still it is acceptable provided the quick prediction ability in a less complex manner. It should be noted that this prediction is made on the reflectance data with no spectral pre-treatment, which result into a larger prediction error when using multispectral camera arrangement with few feature wavelengths. Considering the advantage of low-cost multispectral use and the rapid computational capacity required, the seven-band model for MC and five-band model for rancidity indices could be still considered as acceptable for practical applications for rapid screening of higher or lower moisture almond kernels and rancid or fresh kernel during storage and processing.

Table 3: Feature wavelengths selected by CARS PLS for MC and rancidity indices of almonds

Quality Parameters	LVs used	Selected wavelengths (nm)	Calibration		Cross-validation		Prediction	
			R_c^2	$RMSE_C$	R_{cv}^2	$RMSE_{CV}$	R_p^2	$RMSE_P$
MC	8	970, 1001, 1154, 1312, 1350, 1437, 1670	0.95	0.452	0.93	0.518	0.941	0.494
FFA	7	945, 994, 1105, 1382, 1527	0.92	0.146	0.9	0.165	0.903	0.162
PV	7	945, 1025, 1393, 1488, 1695	0.9	0.666	0.89	0.675	0.886	0.658

R_c^2, R_{cv}^2 and R_p^2=regression co-efficients of calibration, cross-validation and prediction, respectively; $RMSE_c$, $RMSE_{cv}$, $RMSE_p$=root mean square error of calibration, cross-validation and prediction, respectively

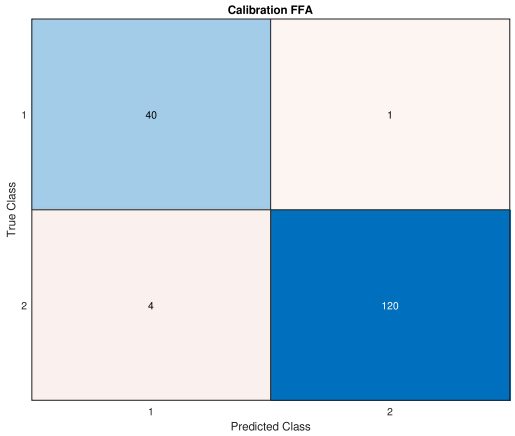
3.6. Classification of almonds according to rancidity

Figure 4 shows the confusion matrix of the two-class SVM classification model for FFA and PV value indicating whether the almonds were fresh or rancid using the feature wavelengths. SVM calibration models showed accuracy of 97% and 97.6% in classifying the rancid almonds from the fresh kernels based on FFA and PV value, respectively. The classification results of FFA model showed 4 out of 44 fresh samples were classified incorrectly and 1 out of 121 rancid samples is classified as fresh. The the validation results of PV model showed 2 out of 109 samples were classified incorrectly and 2 out of 56 rancid samples is classified as fresh. According to PV value the classification results were different because even after 3 months of storage, the PV value was not significantly increased as described earlier. Only the samples stored for 4 months were considered as rancid, in which the PV value exceeded 5 Meq OAA/kg. Samples classified as rancid according to FFA value and PV were considered as the almonds affected by oxidative rancidity and hydrolytic rancidity, respectively(Franklin & Mitchell, 2019). The goodness of fit of the developed models can be demonstrated using R_v^2 values of 0.96 and 0.98 for FFA and PV, respectively for the external test

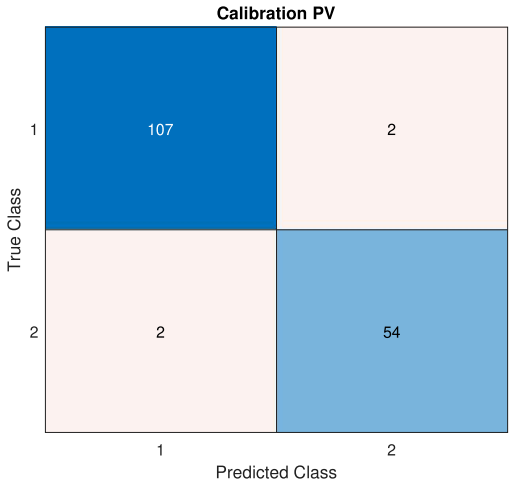
set. However, these models could suffer some biased overfitting caused by the lack of samples in intermediate regions of the results. The classification results demonstrate the potential of the multispectral imaging system with the developed models to identify rancid almonds, which is very useful for the automation of quality control system in nuts industries. The results obtained from this work were lie in line with the previous work using portable NIR spectrometer by Cruz-Tirado et al. (2021) for classification of fresh and stale eggs. They developed classification models using PLS-DA and SVM to discriminate between fresh (Haugh Unit > 60) and stale (Haugh Unit < 60) based on Haugh Unit as a freshness indicator eggs and achieved an accuracy higher than 75% of correct classification of fresh samples.

3.7. Visualization of the moisture distribution

Figure 5 shows the moisture and rancidity distribution pseudo color map generated by applying the MLR models to each pixel of the multispectral image obtained by combining the images at feature wavelengths. The purpose of showing these charts is to provide the readers an idea about intra-kernel distribution of the moisture and rancidity indices. It should be noted that the values assigned for the specific pixels are not representing the exact concentration of moisture, FFA, and PV; instead, it is showing their relative distribution across individual almond kernels.In figure 5a, the average kernel moisture contents of the almonds presented in different columns were 4, 6, 8 and 10% (wb), respectively from left to right. As the moisture content increases, the distribution of intra-kernel moisture can be easily seen. Noise involved in the moisture distribution pattern of first two columns of figure 5a could be due to several contributing factors such as inconsistent intra-cellular



(a)

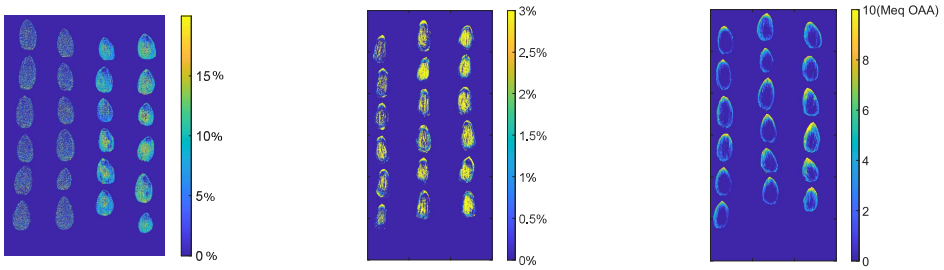


(b)

Figure 4: SVM classification results of almond rancidity based on FFA and PV: (a) confusion matrix of classification based on FFA (%) and (b) confusion matrix of classification based on PV (Meq OAA); class '1' and '2' in the figure refers to 'fresh' and 'rancid' amonds, respectively

arrangement across the kernels, variation in the surface texture, relative affinity of building components of almonds towards moisture and sensitivity of SWIR-HSI towards the moisture level. Similarly, increasing FFA distribu-

tion within the kernel can be seen across the column of figure 5b. PV was not affected until 3 months of storage, due to which first two columns of figure 5c showing low PV value. Third column of figure 5c representing the almonds stored for 4 months due to which majority of the kernels showed high PV values. From the industrial standpoint pixel wise classification may carry limited importance but it could be a good starting line for fundamental research to discover correlation between moisture distribution and initiation of hydrolytic and oxidative rancidity within the kernels.



(a) (b) (c)

Figure 5: Visualization of response distribution for (a) Moisture (%wb), (b) FFA (%) and (c) PV (Meq OAA); units of the color bars are same as that of the corresponding responses

4. Conclusion

Current study involved combination of hyperspectral data with PLSR chemometrics technique, yielded best prediction results for moisture and rancidity of shelled almond kernels. The SG-1st derivative and SG-SNV pre-processing techniques when applied to the raw spectra showed optimal prediction results for MC and rancidity, respectively. Accuracy of the full-spectrum PLSR models were very good with R^2 and $RMSE_P$ values of 0.957,

0.970, 0.955 and 0.410%, 0.053%, 0.412 mEq OAA, respectively for MC, FFA and PV. For rapid quantification and suitability of the system to use in the commercial scale, feature wavelengths were selected using CARS-PLS method without losing significant spectral information, which demonstrates the novelty of the current study. Five and seven number of feature wavelengths were selected for prediction of MC and rancidity, respectively. Simplified MLR models were developed and tested for the robustness and the results showed good predictability with R^2 values of 0.941, 0.903, 0.886 and $RMSE_P$ values of 0.494%,0.162%, 0.658 mEq for MC, FFA and PV, respectively. The overall results concluded that HSI system can be used for basic research purpose to select the useful wavelengths or lab scale prediction of the quality parameters and then multispectral imaging system with specific filters can be developed to generate monochrome images for meeting the need of rapid online prediction. The rapid analysis of all nuts would reduce the waste by early detection and removal of the rancid nuts from the batch. However, automation of the developed technology in food industry may face difficulty with the influence of physical and biological variability, which affects the robustness of the features and algorithms. Thus, future works will be focused on the development of diverse libraries considering cultivar, maturity level, origin, and harvest season to bring sufficient variation in the calibration dataset. Moreover, manufacturing of multispectral system with filter wheels and transferring the calibration method could be done for commercial use of the developed technology.

Acknowledgement

This work was supported by Cooperative Research Centres Projects pro-

gram (CRCPSIX000081), Australia and Surenut Pty Ltd. Financial support from the above sources is gratefully acknowledged. We also thank University of South Australia for providing necessary infrastructure facilities throughout the project duration.

CrediT authors statement

Brajesh Kumar Panda: Investigation (equal), data curation, writing original draft (Lead); Gayatri Mishra: Methodology, Investigation (equal), Software, validation, writing original draft (supporting); Wilmer Ariza Ramirez: Data curation, Formal analysis, validation, writing-review and editing; Hyewon Jung: Formal analysis, validation; Chandra B Singh: Conceptualisation, supervision (equal), resources, writing-review and editing; Sang-Heon Lee: Project administration, supervision (equal), writing-review and editing; Ivan Lee: supervision (equal), writing-review and editing.

References

(2020). *NUTS & DRIED FRUITS STATISTICAL YEARBOOK 2019 / 2020*. Technical Report INC.

Amigo, J. M., Cruz, J., Bautista, M., Maspoch, S., Coello, J., & Blanco, M. (2008). Study of pharmaceutical samples by nir chemical-image and multivariate analysis. *TrAC Trends in Analytical Chemistry*, *27*, 696–713.

AOAC (2019). *Official methods of analysis of AOAC International*. (21st ed.). Gaithersburg: AOAC International.

Association of Official Agricultural Chemists, & Horwitz, W. (1975). *Of-*

ficial methods of analysis volume 222. Association of Official Analytical Chemists Washington, DC.

Bai, S. H., Tahmasbian, I., Zhou, J., Nevenimo, T., Hannet, G., Walton, D., Randall, B., Gama, T., & 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.

Bilal, M., Xiaobo, Z., Arslan, M., Tahir, H. E., Azam, M., Junjun, Z., Basheer, S. et al. (2020). Rapid determination of the chemical compositions of peanut seed (arachis hypogaea.) using portable near-infrared spectroscopy. *Vibrational Spectroscopy*, *110*, 103138.

Borompichaichartkul, C., Chinprahast, N., Devahastin, S., Wiset, L., Poomsa-Ad, N., & Ratchapo, T. (2013). Multistage heat pump drying of macadamia nut under modified atmosphere. *International Food Research Journal*, *20*, 2199.

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*, C1613–C1621.

Caporaso, N., Whitworth, M. B., Grebby, S., & Fisk, I. D. (2018). Rapid prediction of single green coffee bean moisture and lipid content by hyperspectral imaging. *Journal of food engineering*, *227*, 18–29.

Carvalho, L. C., Leite, M. L., Morais, C. L., Lima, K. M., & Teixeira, G. H. (2019). Non-destructive assessment of the oxidative stability of intact

macadamia nuts during the drying process by near-infrared spectroscopy. *LWT*, *103*, 101–107.

Cortés, V., Barat, J. M., Talens, P., Blasco, J., & Lerma-García, M. J. (2018). A comparison between NIR and TR-FTIR spectroscopy for varietal differentiation of spanish intact almonds. *Food Control*, *94*, 241–248.

Cruz-Tirado, J., Pierna, J. A. F., Rogez, H., Barbin, D. F., & Baeten, V. (2020). Authentication of cocoa (theobroma cacao) bean hybrids by nir-hyperspectral imaging and chemometrics. *Food Control*, *118*, 107445.

Cruz-Tirado, J., da Silva Medeiros, M. L., & Barbin, D. F. (2021). On-line monitoring of egg freshness using a portable nir spectrometer in tandem with machine learning. *Journal of Food Engineering*, *306*, 110643.

Deng, S., Xu, Y., Li, X., & He, Y. (2015). Moisture content prediction in tealeaf with near infrared hyperspectral imaging. *Computers and Electronics in Agriculture*, *118*, 38–46.

Elmasry, G., Kamruzzaman, M., Sun, D.-W., & Allen, P. (2012). Principles and applications of hyperspectral imaging in quality evaluation of agro-food products: a review. *Critical reviews in food science and nutrition*, *52*, 999–1023.

Faqeerzada, M. A., Lohumi, S., Kim, G., Joshi, R., Lee, H., Kim, M. S., & Cho, B.-K. (2020). Hyperspectral shortwave infrared image analysis for detection of adulterants in almond powder with one-class classification method. *Sensors*, *20*, 5855.

Franklin, L. M., & Mitchell, A. E. (2019). Review of the sensory and chemical characteristics of almond (prunus dulcis) flavor. *Journal of agricultural and food chemistry*, *67*, 2743–2753.

Ghirardello, D., Contessa, C., Valentini, N., Zeppa, G., Rolle, L., Gerbi, V., & Botta, R. (2013). Effect of storage conditions on chemical and physical characteristics of hazelnut (corylus avellana l.). *Postharvest Biology and Technology*, *81*, 37–43.

Gowen, A. A., O'Donnell, C. P., Cullen, P. J., Downey, G., & Frias, J. M. (2007). Hyperspectral imaging–an emerging process analytical tool for food quality and safety control. *Trends in food science & technology*, *18*, 590–598.

Grisanti, E., Totska, M., Huber, S., Krick Calderon, C., Hohmann, M., Lingenfelser, D., & Otto, M. (2018). Dynamic localized snv, peak snv, and partial peak snv: Novel standardization methods for preprocessing of spectroscopic data used in predictive modeling. *Journal of Spectroscopy*, *2018*.

Guebel, D. V., & Torres, N. V. (2013). Partial least-squares regression (plsr). In W. Dubitzky, O. Wolkenhauer, K.-H. Cho, & H. Yokota (Eds.), *Encyclopedia of Systems Biology* (pp. 1646–1648). New York, NY: Springer New York.

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.

Hong, S.-J., Lee, A.-Y., Han, Y.-h., Park, J., So, J. D., & Kim, G. (2018). Rancidity prediction of soybean oil by using near-infrared spectroscopy techniques. *Journal of Biosystems Engineering*, *43*, 219–228.

Huang, H., Liu, L., & Ngadi, M. O. (2014). Recent developments in hyperspectral imaging for assessment of food quality and safety. *Sensors*, *14*, 7248–7276.

ISO 17059:2019 (2019). *Oilseeds — Extraction of oil and preparation of methyl esters of triglyceride fatty acids for analysis by gas chromatography (rapid method)*. Standard International Organization for Standardization Geneva, CH.

Jin, H., Ma, Y., Li, L., & Cheng, J.-H. (2016). Rapid and non-destructive determination of oil content of peanut (arachis hypogaea l.) using hyperspectral imaging analysis. *Food Analytical Methods*, *9*, 2060–2067.

Khamsopha, D., Woranitta, S., & Teerachaichayut, S. (2021). Utilizing near infrared hyperspectral imaging for quantitatively predicting adulteration in tapioca starch. *Food Control*, *123*, 107781.

Khodabux, K., L'Omelette, M. S. S., Jhaumeer-Laulloo, S., Ramasami, P., & Rondeau, P. (2007). Chemical and near-infrared determination of moisture, fat and protein in tuna fishes. *Food chemistry*, *102*, 669–675.

Liu, D., Sun, D.-W., Qu, J., Zeng, X.-A., Pu, H., & Ma, J. (2014). Feasibility of using hyperspectral imaging to predict moisture content of porcine meat during salting process. *Food chemistry*, *152*, 197–204.

Ma, F., Wang, J., Liu, C., Lu, X., Chen, W., Chen, C., Yang, J., & Zheng, L. (2015). Discrimination of kernel quality characteristics for sunflower seeds based on multispectral imaging approach. *Food analytical methods*, *8*, 1629–1636.

Mahesh, S., Jayas, D., Paliwal, J., & White, N. (2015). Hyperspectral imaging to classify and monitor quality of agricultural materials. *Journal of Stored Products Research*, *61*, 17–26.

Mishra, G., Panda, B. K., Ramirez, W. A., Jung, H., Singh, C. B., Lee, S.-H., & Lee, I. (2021). Research advancements in optical imaging and spectroscopic techniques for nondestructive detection of mold infection and mycotoxins in cereal grains and nuts. *Comprehensive Reviews in Food Science and Food Safety*, .

Mishra, G., Srivastava, S., Panda, B. K., & Mishra, H. N. (2018). Rapid assessment of quality change and insect infestation in stored wheat grain using ft-nir spectroscopy and chemometrics. *Food analytical methods*, *11*, 1189–1198.

Mishra, P., Biancolillo, A., Roger, J. M., Marini, F., & Rutledge, D. N. (2020). New data preprocessing trends based on ensemble of multiple preprocessing techniques. *TrAC Trends in Analytical Chemistry*, (p. 116045).

Möller, J. (2010). Cereals, cereals-based products and animal feeding stuffs–determination of crude fat and total fat content by the randall extraction method: a collaborative study. *Quality Assurance and Safety of Crops & Foods*, *2*, 197–202.

Moscetti, R., Haff, R. P., Aernouts, B., Saeys, W., Monarca, D., Cecchini, M., & Massantini, R. (2013). Feasibility of vis/nir spectroscopy for detection of flaws in hazelnut kernels. *Journal of Food Engineering*, *118*, 1–7.

Moscetti, R., Saeys, W., Keresztes, J. C., Goodarzi, M., Cecchini, M., Danilo, M., & Massantini, R. (2015). Hazelnut quality sorting using high dynamic range short-wave infrared hyperspectral imaging. *Food and bioprocess technology*, *8*, 1593–1604.

Sricharoonratana, M., Thompson, A. K., & Teerachaichayut, S. (2021). Use of near infrared hyperspectral imaging as a nondestructive method of determining and classifying shelf life of cakes. *LWT*, *136*, 110369.

Suktanarak, S., & Teerachaichayut, S. (2017). Non-destructive quality assessment of hens' eggs using hyperspectral images. *Journal of Food Engineering*, *215*, 97–103.

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.

Tao, F., Yao, H., Zhu, F., Hruska, Z., Liu, Y., Rajasekaran, K., & Bhatnagar, D. (2019). A rapid and nondestructive method for simultaneous determination of aflatoxigenic fungus and aflatoxin contamination on corn kernels. *Journal of agricultural and food chemistry*, *67*, 5230–5239.

Teerachaichayut, S., & Ho, H. T. (2017). Non-destructive prediction of total soluble solids, titratable acidity and maturity index of limes by near

infrared hyperspectral imaging. *Postharvest Biology and Technology*, *133*, 20–25.

Torres, I., Pérez-Marín, D., Vega-Castellote, M., & Sánchez, M.-T. (2021). Mapping of fatty acids composition in shelled almonds analysed in bulk using a hyperspectral imaging system. *LWT*, *138*, 110678.

Wang, Y., Zhang, L., Johnson, J., Gao, M., Tang, J., Powers, J. R., & Wang, S. (2014). Developing hot air-assisted radio frequency drying for in-shell macadamia nuts. *Food and Bioprocess Technology*, *7*, 278–288.

Williams, P. C., & Sobering, D. (1996). How do we do it: a brief summary of the methods we use in developing near infrared calibrations. *Near infrared spectroscopy: The future waves*, (pp. 185–188).

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

Xia, Q., Liu, C., Liu, J., Pan, W., Lu, X., Yang, J., Chen, W., & Zheng, L. (2016). Rapid and non-destructive determination of rancidity levels in butter cookies by multi-spectral imaging. *Journal of the Science of Food and Agriculture*, *96*, 1821–1827.

Xu, Y., Zhang, H., Zhang, C., Wu, P., Li, J., Xia, Y., & Fan, S. (2019). Rapid prediction and visualization of moisture content in single cucumber (cucumis sativus l.) seed using hyperspectral imaging technology. *Infrared Physics & Technology*, *102*, 103034.

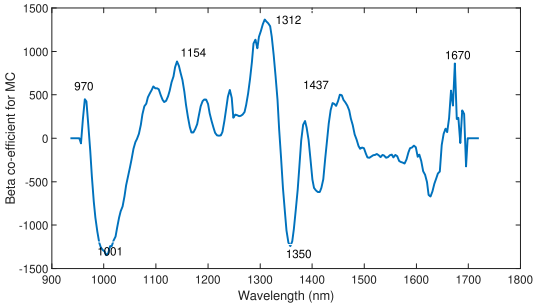
Yildirim, A. N., Akinci-Yildirim, F., Şan, B., & Sesli, Y. (2016). Total oil content and fatty acid profile of some almond (amygdalus communis l.) cultivars. *Polish journal of food and nutrition sciences*, *66*, 173–178.

Young, J. F. (1967). Humidity control in the laboratory using salt solutions-a review. *Journal of Applied Chemistry*, *17*, 241–245.

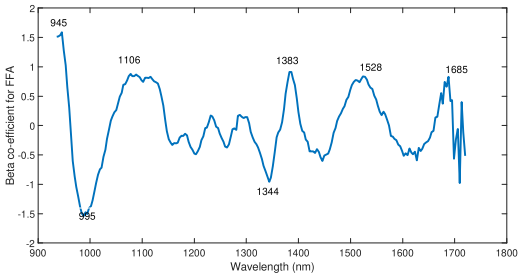
Zhang, Y., & Guo, W. (2020). Moisture content detection of maize seed based on visible/near-infrared and near-infrared hyperspectral imaging technology. *International Journal of Food Science & Technology*, *55*, 631–640.

Supplemental Materials

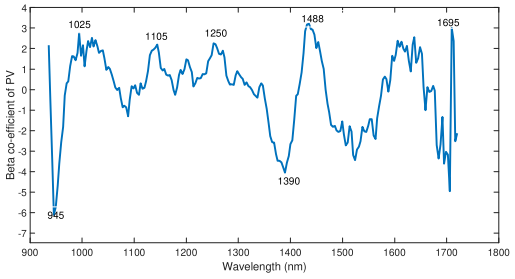
1. Beta Coefficients



(a)



(b)



(c)

Figure S1: Beta co-efficient values obtained from the PLS models developed for prediction of (a) MC (b) FFA and (c) PV of almond kernels