International Journal on Advanced Science Engineering Information Technology

Vol.12 (2022) No. 4
ISSN: 2088-5334

Characterization of Lard Profile from Different Geographical Regions of Husbandries and Body Parts of Pig using FTIR Technique Combined with Chemometrics

N.A.M Salleh [a], N. Yahaya [b,c], D. Rosli [c], M.F. Azizan [c] Bohari M Yamin [a], M.S. Hassan [a,c,*]

[a] *Programme of Industrial Chemical Technology, Faculty of Science and Technology, Universiti Sains Islam Malaysia, 71800 Bandar Baru Nilai, Negeri Sembilan, Malaysia*
[b] *Programme of Food Biotechnology, Faculty of Science and Technology, Universiti Sains Islam Malaysia, 71800 Bandar Baru Nilai, Negeri Sembilan, Malaysia*
[c] *International Fatwa and Halal (iFFAH) Centre, Universiti Sains Islam Malaysia, 71800 Bandar Baru Nilai, Negeri Sembilan, Malaysia*
Corresponding author: [*]mohdsukri@usim.edu.my

***Abstract*—In halal authentication, there is a necessity for vigorous, rapid, faultless, easy, and cost-worthwhile approaches to characterize adulteration materials in food or highly processed products. Lard adulteration in food and other consumer products has been a primary concern, especially among Muslims. This research has been conducted to characterize lard profile collected from different geographical regions, which is southern, northern, and central Malaysia peninsular, and different pig body parts such as belly fats (BL), back fats (BK), and shoulder fat (SF) using Fourier transform infrared spectroscopy (FTIR) technique combined with chemometrics. The chemometrics evaluation on lard spectral was conducted by dividing two data sets into training for the calibration and test sets for the validation model. Comparison between the Multiplicative Scattering Correction (MSC) and Second Derivative Savitzky-Golay (2nd DSG) data transformation was conducted to resolve the baseline issues. The Hotelling T Squared (T2) observed outliers' detection and determined the acceptance threshold based on the first three Principals Components (PCs). Spectral from FTIR spectroscopy showed a similar pattern of lard samples from different geographical regions and pig body parts. In this research, the PCA model of collected lards established using Hotelling T2, and the quality of the model was determined using the projection method. These outcomes of the PCA- data-driven approach in this research have resolved the issue of obtaining a valid source of raw material for preparing certified reference material (CRM) of lard.**

***Keywords*— FTIR; hotelling T2; lard; multivariate; PCA.**

Manuscript received 25 Nov. 2021; revised 26 Dec. 2021; accepted 16 Apr. 2022. Date of publication 31 Aug. 2022.

I. INTRODUCTION

The Muslim community is very concerned about issues of food mixed with haram substances such as lard, pork gelatin, and its derivatives. Lard is pig fat from the back and kidneys, representing pig's metabolites from the biological process. Detecting lards in foods is challenging in halal forensics as the certified reference material (CRM) is limited. The composition of fatty acids (FAs) in the CRM must represent the actual composition in lard. Many theoretical models have been developed to discriminate between lard and other fats and oils to establish a database for halal determination [1].

Furthermore, lard has been reported to differ in geography, which may be affected by nutritional feeds and even body parts of slaughtered pigs. FAs of lard have been reported to have variations in chemical composition depending on the diet taken by pigs. However, the studies were not comprehensive because only specific FAs were involved [2]–[4]. In addition, different diets consumed by pigs [5] and other geographic regions of pig husbandries could contribute to a different metabolite fingerprinting of lard [6]. Hence, the use of CRM for the specific fats that involve the body part and husbandry's location is required to ensure fats authenticity.

In *halal* verification studies, due to the "fingerprint technique," Fourier transform infrared spectroscopy (FTIR) in combination with chemometrics can be used for quantitative analysis and classification of lard [7]. The variety of FTIR and chemometrics could provide accurate results with low errors and the detectability percentage of 1% of lard in the mixture with other fats and oils. The correlation between specific FTIR spectral of lard and FAs chemical

composition has been established by the combination of chemometrics such as Principal Components Analysis (PCA) and Partial Least Squares (PLS) algorithm [8]. PLS can potentially evaluate the adulteration of edible oils (sunflower, canola, coconut, olive, and mustard oils) with lard through single calibration at a frequency range between 1078.01 cm^{-1} and 1246.75 cm^{-1} of the wavenumbers [9]. Another similar study by Nurulhidayah *et al* [10] exploited PLS techniques on the adulteration of lard in dairies products and highlighted oleic and linoleic acyl groups compared to butter at the frequency at 2923 cm^{-1} and 2850 cm^{-1}.

It is necessary to conduct an in-depth study to establish a lard profile database from different regions and parts of the body. Infrared high-dimensional data consists of a sample size (n) smaller than data observation (p) and requires a reduction method such as a PCA. Reduction in variables or dependents to several Principal Components (PCs) allows for hypothetical testing of lard spectra profiles to be performed on collected lard's samples based on region or body parts. To test the significant difference of lard data from various sources has been reported by univariate data by conducting a t-test and seemed impossible to test on spectral and multivariate data. However, PCA data-driven utilizes and extends Hotelling T^2 statistics based on their respective critical limits. Hoteling statistics can be utilized F-test as a t-test replacement for multivariate data. This PCA data-driven and extended Hotelling T^2 has been well used in the multivariate statistical quality control in other industries such as herbal medicine and natural beverage [11]; pharmaceutical products [12]; edibles seed oils quality assurance[13], and quality inspection of the concrete admixture [14].

This study aimed to analyze the spectral of lard from different geographical regions of husbandries and pig body parts to determine whether the null hypothesis (there is no distinction between the lard spectral) can be rejected. The chemometric approaches included transformation spectra dimension and determination confident levels to find similarities between profiles of collected lard according to different regions and body parts.

II. Materials and Methods

A. Chemicals and Materials

Chloroform, methanol, sodium chloride anhydrous, and sodium sulfate anhydrous are all analytical grade (98-99% of purity) and were purchased from Sigma Aldrich. Distilled water (H_2O) was prepared in the iFFAH, USIM laboratory.

B. Samples Collection

Lards samples from pigs were taken from belly fats (BL), back fats (BK), and shoulder fats (SF), according to[15]. All lard samples were purchased from different farms and wet markets in Penang and Perak (Northern Region), Melaka, Johor (Southern Region), Selangor, and Negeri Sembilan (Central Region). The crude lard was neatly wrapped in polystyrene containing ice to maintain freshness before being stored in the laboratory freezer at -4 °C before further steps.

C. Preparation of Lard Samples

Lards (100 g) were cleaned from blood, mucus, pieces of meat, and other tissues using a cutter. Then lards were finely chopped and dried before weighing at 10 g ± 0.05 in a 100 ml beaker.

D. Extraction of Lards Samples

The lards from samples were extracted based on gravimetric methods by non-polar solvents [16]. Lards were homogenized with chloroform: methanol (2:1) to a final volume 10 times of the crude fats sample (10 g in 100 ml of solvent mixture). After dispersion, the whole mixtures were agitated for 15-20 minutes in a shaker at room temperature. The mixture was filtered using a Whatman™ 1(11µm) filter paper to remove non-lards materials from the liquid phase. The liquid phase contained chloroform/methanol was washed with 20 ml of sodium chloride solution (0.9% w/v) and left for 30 minutes until the mixtures were separated into two layers. The upper layer (methanol) was discarded, and the lower layer (chloroform) containing lipids was kept in the flask. An additional small amount of sodium sulfate anhydrous was put into the flask to absorb any moisture from H_2O. The chloroform layer was then filtered again and dried under vacuum in a rotary evaporator (Rotavapor® R-300) at 40 ºC for 20 minutes. A final amount of total fats between 3-4 g for each sample was successfully extracted and kept in a freezer at -4 °C before FTIR analysis.

E. Samples Preparation for Fourier Transformation Infrared (FTIR) Analysis

The extracted lard was removed from the freezer and defrosted for 10 minutes in a water bath (Memmert GmbH) at 60 °C before being placed in a centrifuge tube and transferred to the FTIR room 25 ° C for Analysis.

F. FTIR Analysis

The extracted lard was placed carefully on a multibounce plate at a controlled ambient temperature (25 °C) using a Pasteur pipette in direct contact with attenuated total reflectance (ATR) crystal. FTIR spectrometer (Varian 300) is equipped with deuterated triglycine sulfate (DTGS) detector. FTIR spectra were recorded from 16 scans at a resolution of 8 cm^{-1} at 4000–650 cm^{-1}. A new reference air background spectrum was taken for each scan. The ATR plate was thoroughly washed twice with methanol, then acetone, and dried with soft tissue. The ATR plate was washed and dried after each use. to prevent contamination from previous samples.

G. Data Analysis

The output FTIR data in *.sp* files were converted into ASCII files in the x-y column. First, the profile of raw data lard samples from different regions of southern, northern and central and body parts of pig belly fats (BL), back fats (BK) and shoulder fats (SF) was observed for spectral differences. Secondly, the transformation and PCA approach for the discrimination collected samples of lard according to regions and body parts were demonstrated. Unscrambler X v.10.3 (CAMO Software, Oslo, Norway) was used for raw data transformation before classification was performed using PCA. Then the finalized transformed data-PCA model was used to establish a significant level using Hotelling's T-Square, (T^2). Finally, the projection method determined the quality of the chosen PCA model.

H. Transformation of FTIR Spectra

Prior to applying the PCA, the obtained results (raw data) were transferred into table form (Microsoft Office Excel 2013) where the data format is on the x-y axis. Sample's name of collected lard according to region and body part were inserted into the rows. The wavenumbers were positioned in columns and known as X-matrix. Line plots of the spectral were resemblance or illustrated the original data were observed before a further transformation data. Data were divided into training sets and test sets through Kennard Stones (K-S) method. The multiplicative Scattering Correction (MSC) and Second Derivative Savitzky-Golay (2^{nd} DSG) were selected to correct the baseline issues.

I. Principal Component Analysis (PCA)

The exploratory approach of Principal Components Analysis (PCA) was used to determine similarities and discrepancies between the observations and to interpret the relationship between the variables. Scores and loadings plots were used to visualize the PCA performance. In the scores plot, each observation that includes all variables simultaneously is represented by a single point, and the plot in the graph represents each variable. The loadings show how critical the individual variables are in defining the underlying dimensions. As a result, the loadings describe the model space in terms of all variables that can be projected onto new observations. The relevant variables and observations are revealed simultaneously when scores and loadings plots are interpreted.

J. Hotelling's T-Square, (T2)

The distance to the model center as spanned by the principal components (PCs) is measured by Hotelling's T^2 statistics, displaying the model space's extreme plots. As a result, it can track systemic environmental changes and identify cases in which a system operates outside of its usual parameters. The Hotelling's T^2 value for individual samples can test the similarity of the models depending on Confident Interval (CI) selection.

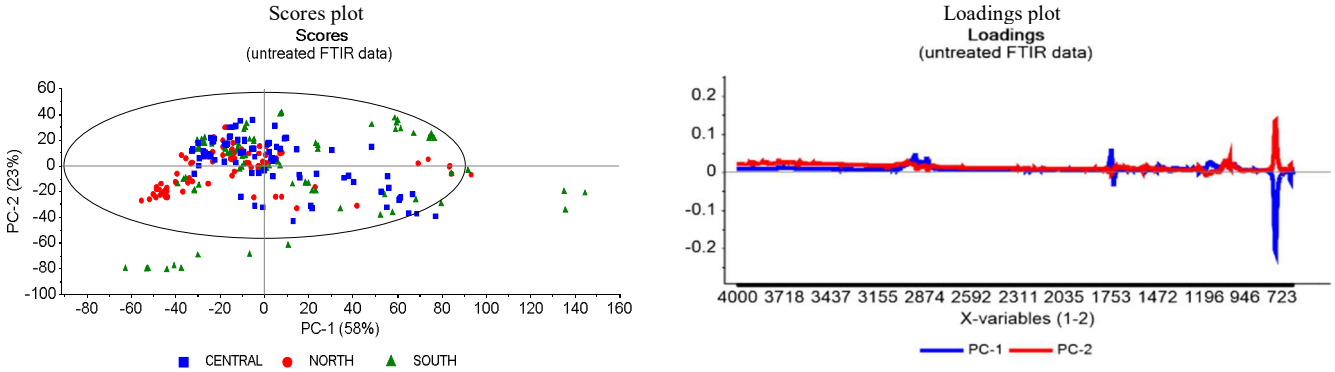
III. Results and Discussion

A total of 270 FTIR spectra were measured, and line plots illustrated lard spectra collected by body parts based on different regions; North, South and Central according to the body parts; BK (back fats), BL (belly fats) and SD (shoulder fats). A total of 180 sample training sets were divided into regions; North (60), South (61), Central (59). The remaining 90 samples of the test set are from North (30), South (29), Central (31) regions. The division of training and test set were conducted by K-S algorithm selection. After that, pre-processing or transformation was applied into the training set of divided samples to improve the information of interest and reduce/remove the noises on spectral signals before applying them to PCA.

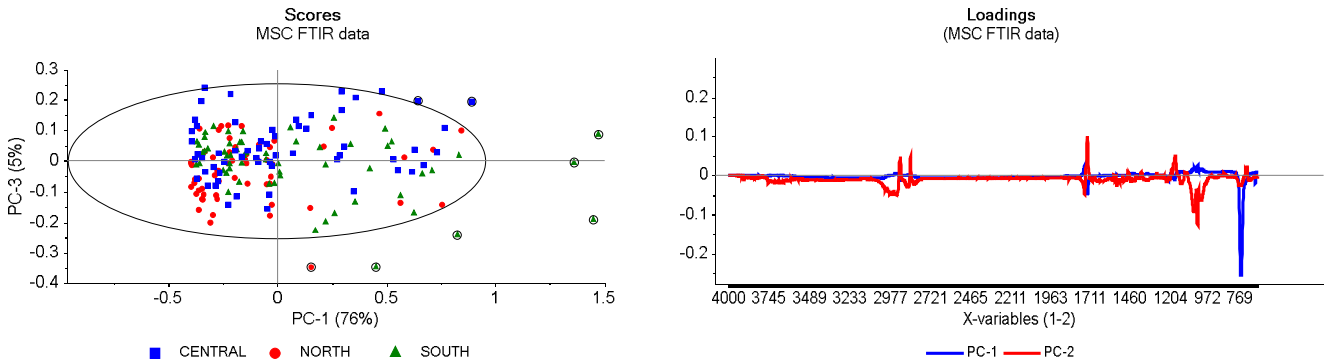
A. PCA on FTIR Spectra before and after Transformation

In this study, PCA was performed on raw FTIR spectral of lard between different regions and body parts to obtain an overview of the data and discrepancies of outliers. In the previous section, it was reported that the spectral of untreated FTIR data showed no differences of lards profiles by regions or body parts. In PCA, the important data was compressed into PC. Furthermore, FTIR data treated with MSC and 2nd DSG pre-processing were also subjected to PCA. Fig. 1 shows PCA according to the raw and transformation or pre-processing of FTIR data subdivision by as a) Untreated PCA-FTIR, b) Full spectra of MSC-PCA FTIR, c) Selected frequencies of MSC-PCA FTIR, and d) 2nd DSG-PCA-FTIR.

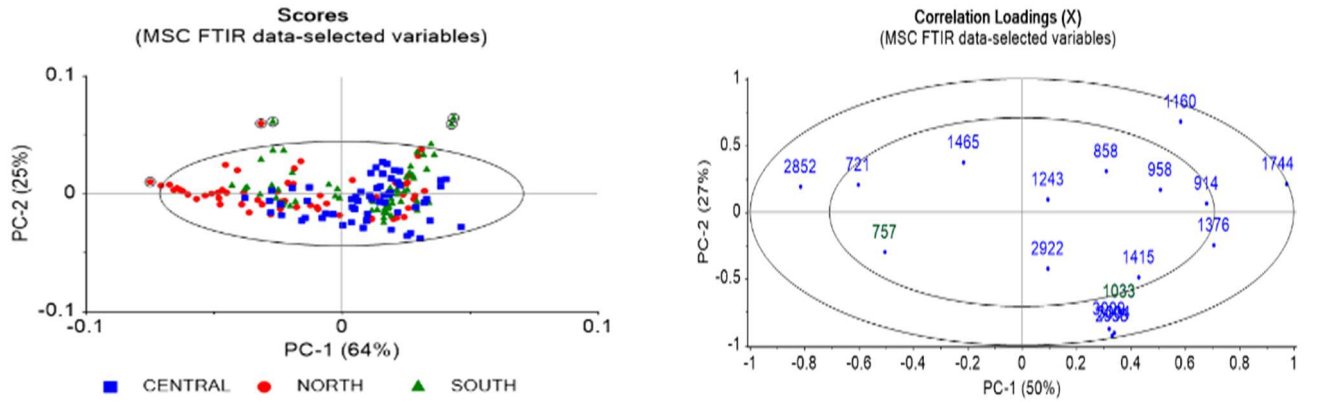
The outcome of PCA on raw FTIR data shows that the lard samples clustered in a PC-1 at the negative direction a) capture 81% at first 2PCs variances. A few samples, mainly from the south region, are outside the circle (CI at 95%). These outlier samples can be related to the sharp peak 756 cm-1 as shown in loading plots Fig. 1 (a), which are identified as chloroform residues as interference peaks. These peaks also appear in positive PC-1 plots due to the scattering baseline. Thus, pre-processing with peak data selection was evaluated in lard characterization. The line loading plot only shows a similar profile to the original data and may highlight regions of high importance. PCA on MSC pre-processing FTIR Spectra.



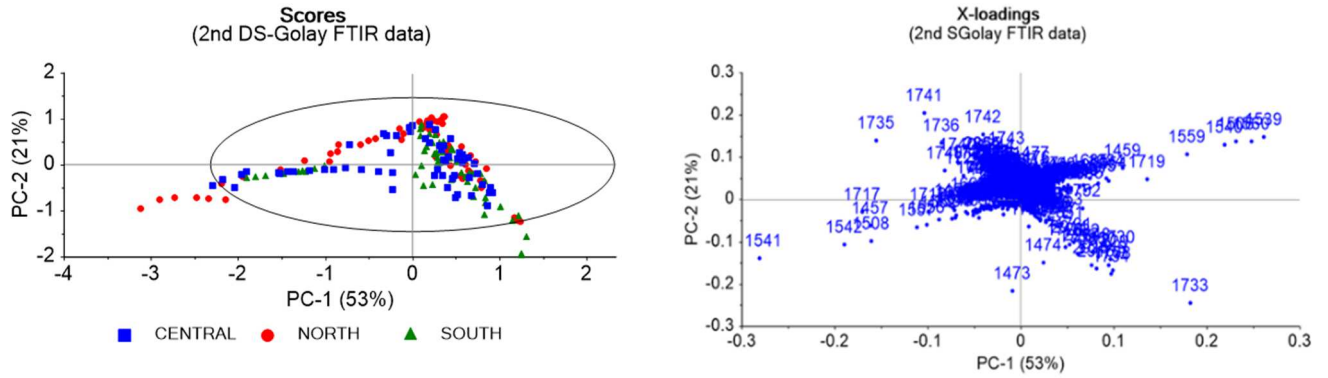
a) Full spectra of Untreated-PCA FTIR



(b) Full spectra of MSC-PCA FTIR



(c) Selected frequencies of MSC-PCA FTIR



(d) 2nd DSG-PCA-FTIR

Fig. 1 Scores and loadings plot of PCA-FTIR after pre-processing

As shown in Fig. 1 (b), the MSC-PCA full spectra showed similar trends in raw PCA and equal total variances of 81% at the first two PCs. However, scores plots are closer, and outliers (CI at 95%) are lesser than the original data. From its loading plots, the interfering peaks at 756 cm-1 only at negative PC-1. There are slight variations on PC-2 North region samples outweighed at positive PC-2 direction and South region at negative PC-2 direction. Thus, MSC-PCA variables were selected because the interferent peaks are dominant.

Fig. 1 (c) shows that only 77% of the variances are captured after variable selection on MSC-PCA. Selection of the variables was made according to the lard individual's frequencies. X variances more clearly visualize the variables selected in the correlation loadings plot in Fig. 1 (c) the loadings plots compared to the standard view loadings plot in Fig. 1 (b). PCA on FTIR Spectra after 2nd DSG transformation (2nd DSG -PCA)

The plot in correlation loadings (Fig. 1 (c)) included the two ellipses representing correlation X-loading between selected peaks. The outer ellipse is the unit-circle and indicates 51-100% variances. The inner ellipse indicates 0-50% of the variance. It can be seen that the importance of individual selected peaks with correlations at positive PC-1 : 1744 cm-1 (0.973), 1376 cm-1 (0.706), and negative PC-1: 2852 cm-1 (-0.814), 3009 cm-1 (-0.882), positive PC-2: 1160 cm-1 (0.684). The negative PC-2: 3004 cm-1 (-0.909), 2958 cm-1 (-0.928), 1160 cm-1 (0.684). Each main peaks are in agreement with the previous research [17] [18].

The other pre-processing applied is 2nd DSG. Fig. 1 (d) shows that 2nd DSG-PCA captures 74 % of the variances at the first 2PCs. Unlike MSC pre-processing, determining

loading plots from 2nd DSG-PCA is not made by selecting individuals' frequencies. The sharp peaks produced by 2nd DSG-PCA caused a discrepancy in the point data of each sample. Therefore, only interval wavenumbers selection between 3030-1120 cm-1 was used on FTIR spectra after 2nd DSG-PCA to exclude interferent peaks. From the loadings plot, the first 2PCs display noise in 1823-1696 cm-1 and 1500-680 cm-1, notably at positive PC-1. This confirms the statement that the method of 2nd DSG-PCA gives much change to the spectra of the origin [19][20].

From the three PCA models of the untreated and transformed data, no distinction grouping either by region or body parts even after interferent removal. The selected MSC-PCA was chosen as the final model for the similarity test to validate the model. Then extended PCA employed Hoteling T2 for outliers and critical limit evaluation.

B. Hotelling T2 Similarity Assessment of MSC-PCA Models

In this study, the similarity of lard spectra was determined by evaluating different percentages of α (0.5%, 0.1%, 1% & 5%). For quantitative similarity of the collected lard, the Hotelling T2 statistic was proposed to test the α of the multivariate data. The Hotelling T2 hypothesis of a multivariate normal distribution is most often used to identify outliers or abnormalities during multivariate process regulation. Hotelling's T2 statistics are used to describe a class space, and the trust ellipse is used to test a multivariate normal distribution hypothesis.

All samples within the critical line limit and Hotelling T2 eclipse remain the same spectra. In comparison, those outside the line samples are determined as extremists or outlier samples. A total of eight samples are spotted as outer ellipse at 95% CI of Hotelling T2 limits MSC-PCA model on PC-1 vs. PC-3. (Fig. 2).

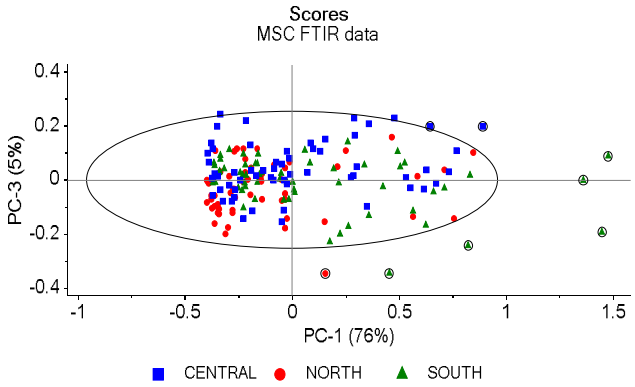


Fig. 2 Determination of outliers by MSC-PCA

Table 1 shows the comparison between four different significance levels and the outliers for the first 3 PCs after removing eight outliers at outside 95% CI. There are no outliers at the first 3 PCs at 0.1 % and 0.5% α. After removing the outliers to accommodate Type II error was chosen by reducing α. Therefore, α at 0.5% would have to determine the similarity of the lard profile. Based on the decision α at 0.5%, it can be stated that these study outcomes failed to reject the null hypothesis. The α of the MSC-PCA model of the collected lard by Hotelling T2 statistic has been determined by accepting the null hypothesis of lard profile regardless of the regions and body parts.

TABLE I
COMPARISON BETWEEN FOUR DIFFERENT SIGNIFICANCE LEVELS (α)

	Hoteling T^2 Limit			Outliers (Validation F-Test)		
(1-α) α	**PC-1**	**PC-2**	**PC-3**	**PC-1**	**PC-2**	**PC-3**
(99.9%) 0.10%	11.269	14.548	17.305	0	0	0
(99.5%) 0.50%	8.130	11.054	13.521	0	0	0
(99.0%) 1%	6.822	9.569	11.896	3	0	3
(95%) 5%	3.918	6.167	8.110	5	10	13

α = Significance Levels

First 3 PCs (total variances >80%) were retained and the threshold to an arbitrary critical value were set at 0.5% α, referring to the Hoteling T2 limits (PC-1 = 8.130, PC-2 = 11.054 & PC-3 = 13.521; df; 179, 6). These outcomes can be used as the benchmark for profiling lard at different regions and body parts. Thus, the first 3PCs of MSC-PCA are chosen to determine the similarity of lard spectra and subsequently to predict the new samples.

The results, Table 1, clearly show that the statistical distance between the corrected limits decreases as p-values (denoted by α) increase. From the findings, the similarities scores vector presents after considering removing some extreme outliers. Thus, a p-value of 0.5% α (99.5% CI) was chosen after considering the elimination of the extreme outliers. Interestingly, the generalization of Hoteling T2 statistics as a CRM benchmark was not found in lard or selected fats, especially on spectroscopic analysis. The influence plots are typically used to identify outliers or abnormalities when a procedure runs outside standard parameters. In confidence interval (CI), though, it can measure significance [21].

C. Prediction of Lard by MSC-PCA Projection

After the MSC-PCA model was finalized, the prediction process on the test set was applied. Test sets were data outside the training set randomly selected by the K-S algorithm. The two-dimensional (2D) scatter plot scores for two specified components (Factors) from Prediction results. Once the MSC-PCA model is established on a training set or calibration, new objects or variables can be added to the model, which would cause new scores, t, and loadings. In addition, the variance of the residuals, e, is calculated for each fitted component, providing a measure of similarity between the test and training sets.

The plot observed how close the new samples' predictions were to the original samples. The training set used to construct the MSC-PCA model is shown in blue green as the latest prediction samples in Fig. 3. As expected, all the scores plot of the test set (green) are located inside the eclipse. The total variance using MSC-PCA is acceptable in terms of percentage calibration (at right) and prediction (at the left); PC-1 (50%, 43%), PC-2 (27%, 33%) and PC-3 (13%, 15%).

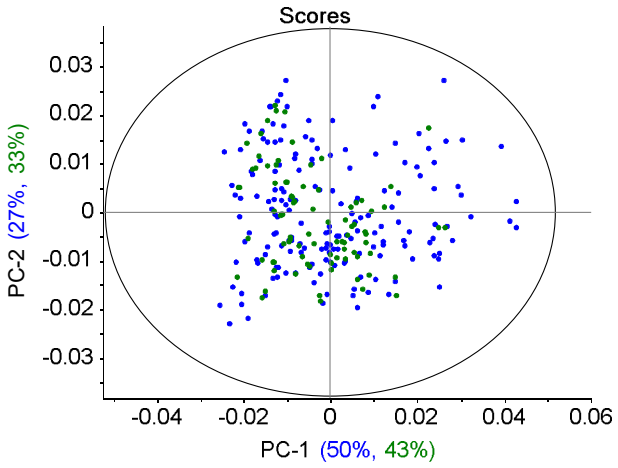
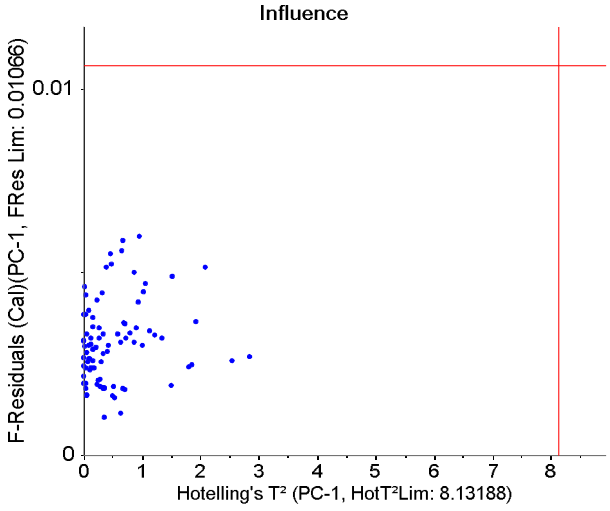


Fig. 3 Projection of test set into 2D Scores plot of MSC-PCA

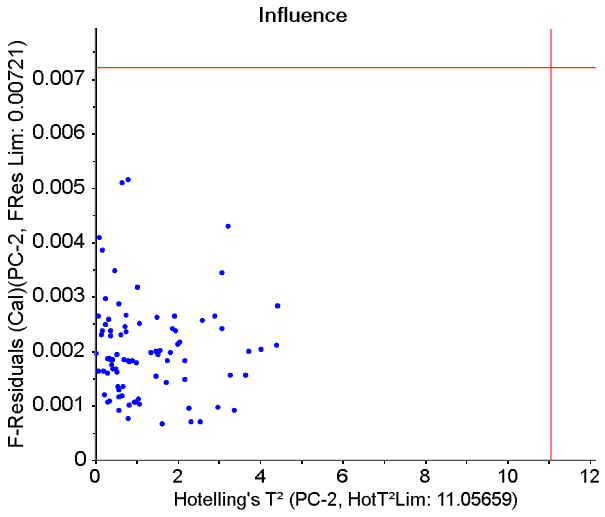
D. Prediction Validation by Total Residual Variance

The 2D graph in Fig. 3, may not fully show test sample plotting. Apart from the projected 2D Scores plot, the test set results for the first 3 PCs were observed through the Influence graph comprising F-residual value vs. Hotelling's T2 weight with their respective critical limits based on α above (Table 1). Hence, the first 3PC distribution so each PC was visualized individually by an influence graph. F-residual value vs Hotelling's T2 was set to use as boundaries to conclude the test set in belonging to the lard.

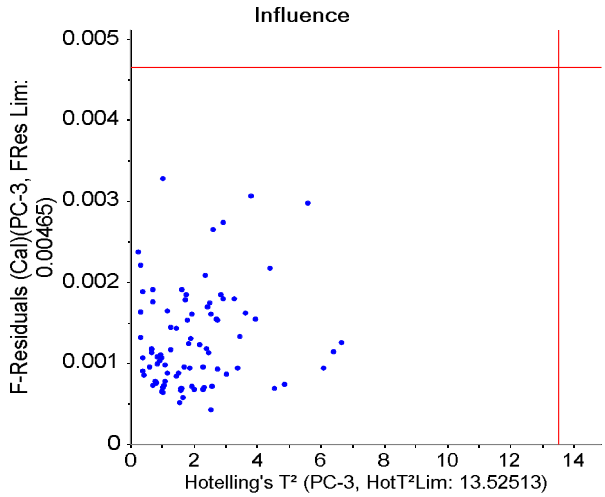
The test set samples re-assembly inside the first 3 PCs (A) boundaries in the Influence graph plot. As can be seen in Fig. 4, The F-residual value versus Hotelling's T2 in Fig. 4 (a) PC-1: 0.01, 8.130; Fig. 4(b) PC-2: 0.007, 11.054; and Fig. 4 (c) PC-3: 0.005, 13.825. All projected test samples are shown inside the red line (critical limit), and there are no extreme samples.



(a) PC-1



(b) PC-2



(c) PC-3

Fig. 4 Influence graph plots F-Residual versus Hoteling T^2

The PCA projection outcomes can also be validated as the whole first 3PCs through a residual variance graph (Fig. 5) to indicate the quality of the projected samples. The scores vector outside these boundaries indicates that test samples are not lard groups. Hotelling's T2 statistic measures the distance of a new observation to the model mean. The critical limit is based on an F-test that replaces the t-test in the univariate statistic.

The blue curve represents the residual calibration variance, which depends on fitting the calibration data to the MSC-PCA model. Cross-validation is used to calculate residual validation variance, represented by the red curve. The residual variance of the predicted samples is often seen as a green curve in the projection case (Fig. 5). The resultant MSC - PCA models produced the smallest values below zero of total residual variance, where the residual variance is dropped sharply to zero from the fewest possible components. The projected curve below the calibration and validation curve showed reliable prediction results.

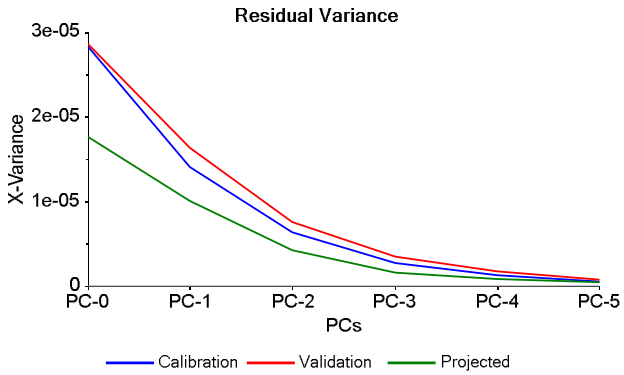


Fig. 5 MSC-PCA of Total Residual Variance

Visualization of residual distances of the X-variances can be related directly to each PC and give a robust conclusion about the prediction through MSC-PCA residual variance graph. The findings show that the first 3PCs have given a satisfactory smaller residual for the better quality of the model [22]. It can be seen from the results of the prediction curve, which are parallel to the calibration/validation curve. This means that the new projected samples (test set) belong to the calibration.

E. Transformation of FTIR Spectra

Prior work has succeeded in identifying lard between other fats using combination FTIR spectra and chemometrics at certain frequencies in the wavenumbers. The qualitative and quantitation determination of lard blended with other animal fats found almost different at 3010-3000, 1220-1095 cm-1 in the wavenumbers ranges [23], [24]; qualitatively studies of lard to differ with other edibles fats and oils at 2852.8, 2922 and 1464.7 cm-1 [25]. However, these studies relied on small sets of lard samples that may not be sufficient to represent the entire lard profile. The study was conducted to compile lard from different regions and body parts and encourage using a standard library of lard samples to improve lard research and reliability. Ideally, such a library would be composed of lard spectral and unconventional CRMs with well characterized chemometric data. Traditionally, ANOVA has been performed as a univariate statistical method in the CRM process uncertainty test assessment [26]. In this study, the high dimensions generated by FTIR and chemometrics techniques were proposed to develop the CRM. Recently, Adilson et al [27] developed chemometrics tools as an alternative to PCA and HCA to quantify CRM of inorganic material.

The chemometrics strategies have found that the baseline shift variation could be caused by different light scattering. Light scattering is present when it penetrates through the surface into thickened lard samples in semi-solid form. It can be seen by observing the baseline shift of the %T line plots of the raw data. This baseline shift was also in general issues in FTIR measurement of oils [28], [29]. Thus, the transformation of FTIR spectral data is important as a correction before chemometrics modeling. The 2nd DSG transformation can correct the multiplicative effect by removing both baseline and linear trends. However, the techniques are unsuitable for this study because some noise spectral would be inflated. Hence, the MSC transformation is suitable to be applied to conserve the spectral features of the lard. From line plots Fig.6 (a) & (b), MSC transformation and raw others not changing much differ lines plot. In looking for similarities of lard, the MSC transformation is preferable. The importance of transformation shows outliers in the ATR-FTIR spectra and irrelevant signals arising from contamination [30].

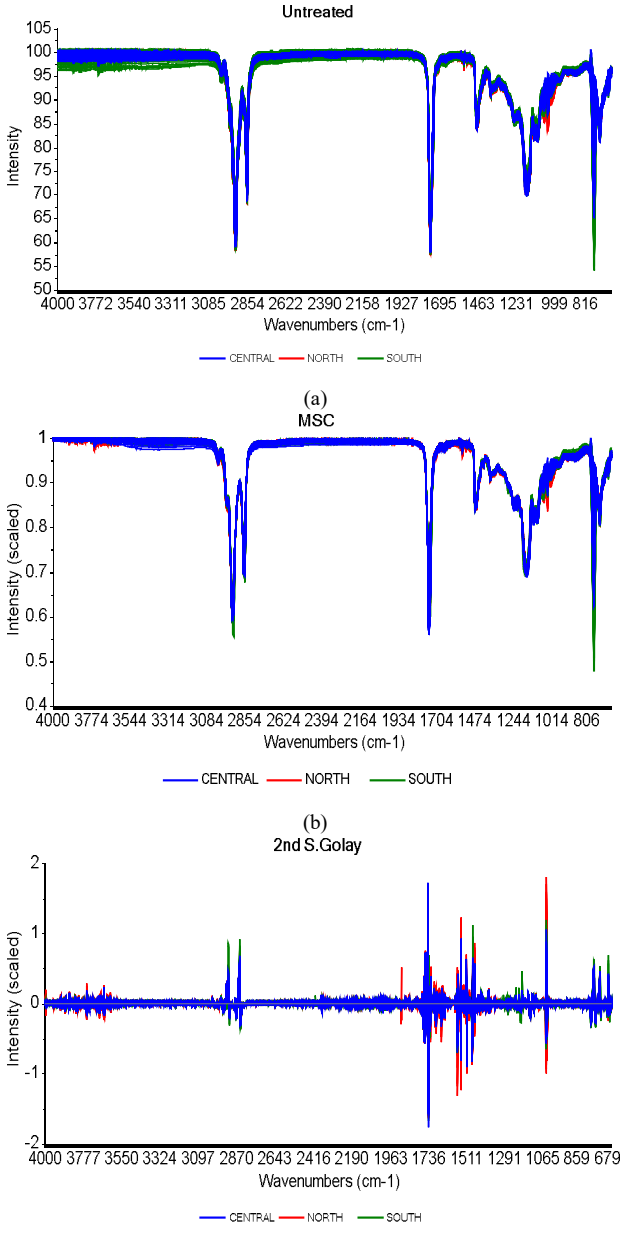


Fig. 6 FTIR Spectra of lard from Central, North, and South regions of Peninsular Malaysia

F. Principal Component Analysis (PCA) on Raw and Transformed FTIR Spectra

PCA demonstrated by the outlier's evaluation that can be removed before the model was finalized. In this study, some wavenumbers selections are important to refine data that identify interferent peaks such as solvents due to fat extraction. Before and after transformation data, the initial results could not cluster lard according to regions and body parts. The findings suggested that there is the homogeneity of

the spectral lard qualitatively. The finalized MSC-PCA calibration model compressed multivariate data into the first 3PCs to represent the lard spectrum from selected frequencies displayed by loading plots.

G. Hotelling T^2 Statistic Similarity Assessment of MSC-PCA Models

Hotelling T2 statistics was proposed to test the significance level of the multivariate data. The Hoteling T2 hypothesis of a multivariate normal distribution is most often used to identify outliers or abnormalities during multivariate process regulation Hotelling's T2 statistics are used to describe a class space, and the trust ellipse is used to test a multivariate normal distribution hypothesis. From the results, TABLE clearly shows that the statistical distance between the corrected limits decreases as p-values (denote by CI) increase. From the findings, the similarity scores vector presented by the confidence ellipse represents the 99.5% CI was determined after considering potential outliers. Thus p-value of 0.5% was chosen by elimination of the extreme outliers. Interestingly, the generalization of Hoteling T2 statistics as a CRM benchmark is almost not found in lard or other fats in the research, especially on spectroscopic analysis.

H. Prediction of Lard by MSC-PCA Projection

Projection in the PCA is equivalent to regression prediction and provides a robust statistic. Projecting the test set samples onto the existing calibration model (training set) and checking residual variances and leverages determining model validation for the test set samples simultaneously. Once the MSC-PCA model is established on the training set or calibration, new objects or variables can be added to the model, resulting in new ratings, t, and loadings p. In addition, the variance of the residuals, e, is calculated for each fitted component, providing a measure of similarity between the test and training sets as per equation (1),

$$t = xp \tag{1}$$

Test set x is projected into the A-dimensional space generated in the training set (MSC-PCA). The residuals vector e are calculated as per equation (2),

$$e = x - tp' \text{ or } e = x(I - PP') \tag{2}$$

The identity matrix is denoted as I. Since the vectors are independent, this estimation of the new scores t, or loadings P, is analogous to linear regression [31], [32]. The finding showed that PCA projection is highly recommended to validate the similarity of the new samples, primarily when the model's saturated and homogenous scores vectors exist. Otherwise, to find dissimilarity between scores plots, a combination PCA and OPLSDA had found chemical composition related to the lard feed by fish by other studies [33].

The statistical inference method applied the random effect model to determine how the built-in model reapplied. Prediction variances were slightly higher at PC-2 and PC-3 but acceptable because the matrix X is not always perfectly reconstituted after resampling into training especially involving spectral data, which is small object significant variables [34].

I. Prediction of New Samples: Hotelling's T2 Statistic Visualized by Influence Plot and Residual Validation

Results showed that the test set samples were reassembled inside the first 3 PCs (A) in the influence plot. The outcomes also can be validated as the whole first 3PCs through a residual variance graph to indicate the quality of the projected samples. The influence graph visualized the distribution scores vector of each PC to be feasible to determine outliers and the similarity. The Unscrambler software has convenient for setting up the limit outliers. This study's leverage and residual (outlier) limits are given according to the Hotelling T2 Similarity Assessment of the MSC-PCA model section.

The scores vectors located outside these boundaries indicate that test samples are not in lard groups. Hotelling's T2 statistic measures the distance of a new observation to the model mean. The critical limit is based on an F-test that replaces the t-test in univariate statistics. The influence plots are typically used to identify outliers or abnormalities when a procedure runs outside standard parameters. Though it comes to Confidence Interval (CI), it can measure significance [21].

Visualization of residual distances of the X variances can be related directly to each PC and give a robust conclusion about the MSC-PCA residual variance graph (Figure 6). The findings show first 3PCs have given a good smaller residual for the better quality of the model. It can be seen from the results of the prediction curve, which are parallel below to the calibration/validation curve.

IV. Conclusions

A novel PCA data-driven approach has been successfully demonstrated to characterize lards from different regions and body parts. The close similarity of the lard's profile from the different areas and body parts can be used as a CRM for lard. The findings exploited Hoteling T2 from the MSC-PCA model for calibration and projection in the validation method. The use of multivariate analysis for evaluating CRM has been stated in the Guide to Measuring Uncertainty in Measurement (GUM) according to the ISO/IEC 98-3 (2008)[35]. Alternative assessment corresponding to multivariate analysis trends, such as chemometrics approaches, can reapply on the new data on the proposed PCA model. This study has shown that the MSC-PCA model extended to Hotelling T2 statistic provides a robust validation through resampling by PCA projection.

Acknowledgments

The research financial was fully supported by the Universiti Sains Islam Malaysia under Centre of Excellence (COE), International Fatwa and Halal (iFFAH) Centre (PPPI /KHAS_IHRAM /04/061007/13818). The solvent dryer, Rotavapor® R-300 was provided by Kolej Genius Insan, USIM.

References

[1] N. A. M. . Salleh and M. S. Hassan, "Discrimination of Lard and other Edible Fats after Heating Treatments using Partial Least Square Regression (PLSR), Principal Component Regression (PCR) and Linear Support Vector Machine Regression (SVMR).," *J. Phys. Conf. Ser.*, vol. 1366, no. 1, 2019, doi: 10.1088/1742-6596/1366/1/012114.

[2] M. Joven, E. Pintos, M. A. Latorre, J. Suárez-Belloch, J. A. Guada, and M. Fondevila, "Effect of replacing barley by increasing levels of

olive cake in the diet of finishing pigs: Growth performances, digestibility, carcass, meat and fat quality," *Anim. Feed Sci. Technol.*, vol. 197, pp. 185–193, 2014, doi: 10.1016/j.anifeedsci.2014.08.007.

[3] Y. Y. Tian, L. M. Gong, J. X. Xue, J. Cao, and L. Y. Zhang, "Effects of Graded Levels of Chromium Methionine on Performance, Carcass Traits, Meat Quality, Fatty Acid Profiles of Fat, Tissue Chromium Concentrations, and Antioxidant Status in Growing-Finishing Pigs," *Biol. Trace Elem. Res.*, vol. 168, no. 1, pp. 110–121, 2015, doi: 10.1007/s12011-015-0352-1.

[4] T. Wandee, T. Tinnagon, and K. Nattporn, "Effects of the ratio of unsaturated fatty acid to saturated fatty acid on the growth performance, carcass and meat quality of finishing pigs," *Anim. Nutr.*, vol. 2, no. 2, pp. 79–85, 2016, doi: 10.1016/j.aninu.2016.03.004.

[5] X. Ao and I. H. Kim, "Effects of dietary dried mealworm (Ptecticus tenebrifer) larvae on growth performance and nutrient digestibility in weaning pigs," *Livest. Sci.*, vol. 230, p. 103615, 2019, doi: 10.1016/j.livsci.2019.09.031.

[6] N. Tous, R. Lizardo, B. Vilà, M. Gispert, M. Font-i-Furnols, and E. Esteve-Garcia, "Effect of a high dose of CLA in finishing pig diets on fat deposition and fatty acid composition in intramuscular fat and other fat depots," *Meat Sci.*, vol. 93, no. 3, pp. 517–524, 2013, doi: 10.1016/j.meatsci.2012.10.005.

[7] A. Rohman, T. Kuwat, S. Retno, Sismindari, E. Yuny, and W. Tridjoko, "Fourier transform infrared spectroscopy applied for rapid analysis of lard in palm oil," *Int. Food Res. J.*, vol. 19, no. 3, pp. 1161–1165, 2012.

[8] A. Rohman *et al.*, "Comprehensive Review on Application of FTIR Spectroscopy Coupled with Chemometrics for Authentication Analysis of Fats and Oils in the Food Products," *Molecules*, vol. 25, no. 22. MDPI AG, p. 5485, 2020, doi: 10.3390/molecules25225485.

[9] F. Munir, S. G. Musharraf, S. T. H. Sherazi, M. I. Malik, and M. I. Bhanger, "Detection of lard contamination in five different edible oils by FT-IR spectroscopy using a partial least squares calibration model," *Turkish J. Chem.*, vol. 43, no. 4, pp. 1098–1108, 2019, doi: 10.3906/kim-1902-17.

[10] A. F. Nurrulhidayah *et al.*, "FTIR-ATR spectroscopy based metabolite fingerprinting as a direct determination of butter adulterated with lard," *Int. J. Food Prop.*, vol. 18, no. 2, pp. 372–379, 2015, doi: 10.1080/10942912.2012.692224.

[11] H. Feng *et al.*, "Quality assessment of Succus Bambusae oral liquids based on gas chromatography/mass spectrometry fingerprints and chemometrics," *Rapid Commun. Mass Spectrom.*, vol. 35, no. 24, p. e9200, Dec. 2021, doi: 10.1002/rcm.9200.

[12] M. Mittal, K. Sharma, and A. Rathore, "Checking counterfeiting of pharmaceutical products by attenuated total reflection mid-infrared spectroscopy," *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.*, vol. 255, p. 119710, Jul. 2021, doi: 10.1016/j.saa.2021.119710.

[13] C. Lörchner, M. Horn, F. Berger, C. Fauhl-Hassek, M. A. Glomb, and S. Esslinger, "Quality control of spectroscopic data in non-targeted analysis – Development of a multivariate control chart," *Food Control*, vol. 133, p. 108601, Mar. 2022, doi: 10.1016/j.foodcont.2021.108601.

[14] Z. Li, B. Li, Z. Zhao, W. Ma, W. Li, and J. Wang, "Rapid Identification and Quantitative Analysis of Polycarboxylate Superplasticizers Using ATR-FTIR Spectroscopy Combined with Chemometric Methods," *Math. Probl. Eng.*, vol. 2021, 2021, doi: 10.1155/2021/6613382.

[15] D. Franco, J. A. Vazquez, and J. M. Lorenzo, "Growth performance, carcass and meat quality of the Celta pig crossbred with Duroc and Landrance genotypes," *Meat Sci.*, vol. 96, no. 1, pp. 195–202, 2014, doi: 10.1016/j.meatsci.2013.06.024.

[16] J. Folch, M. Lees, and G. H. S. Stanley, "A simple method for the isolation and purification of total lipides from animal tissues," *J. Biol. Chem.*, vol. 226, no. 1, pp. 497–509, 1957, doi: 10.1007/s10858-011-9570-9.

[17] A. N. Nina Naquiah, J. M. N. Marikkar, M. E. S. Mirghani, A. F. Nurrulhidayah, and N. A. M. Yanty, "Differentiation of fractionated components of lard from other animal fats using different analytical techniques," *Sains Malaysiana*, vol. 46, no. 2, pp. 209–216, 2017, doi: 10.17576/jsm-2017-4602-04.

[18] S. F. Sim, M. X. L. Chai, and A. L. Jeffrey Kimura, "Prediction of lard in palm olein oil using simple linear regression (SLR), multiple linear regression (MLR), and partial least squares regression (PLSR) based on fourier-transform infrared (FTIR)," *J. Chem.*, vol. 2018, Nov. 2018, doi: 10.1155/2018/7182801.

[19] W. Jiang, C. Lu, Y. Zhang, W. Ju, J. Wang, and M. Xiao, "Molecular spectroscopic wavelength selection using combined interval partial least squares and correlation coefficient optimization," *Anal. Methods*, vol. 11, no. 24, pp. 3108–3116, 2019, doi: 10.1039/c9ay00898e.

[20] V. Panchuk, V. Semenov, A. Legin, and D. Kirsanov, "Signal Smoothing with PLS Regression," *Anal. Chem.*, vol. 90, no. 9, pp. 5959–5964, 2018, doi: 10.1021/acs.analchem.8b01194.

[21] B. Efron, "Estimation and Accuracy After Model Selection," *J. Am. Stat. Assoc.*, vol. 109, no. 507, pp. 991–1007, Jul. 2014, doi: 10.1080/01621459.2013.823775.

[22] CAMO, "The Unscrambler Tutorials By CAMO Process AS," 2015.

[23] Y. B. Che Man and M. E. S. Mirghani, "Detection of lard mixed with body fats of chicken, lamb, and cow by Fourier Transform Infrared Spectroscopy," *JAOCS, J. Am. Oil Chem. Soc.*, vol. 78, no. 7, pp. 753–761, Jul. 2001, doi: 10.1007/s11746-001-0338-4.

[24] I. Jaswir, M. E. Mirghani Saeed, T. H. Hassan, and M. Z. M. Said, "Determination of Lard in Mixture of Body Fats of Mutton and Cow by Fourier Transform Infrared Spectroscopy," *J. Oleo Sci.*, vol. 52, no. 12, pp. 633–638, 2003.

[25] Y. B. Che Man, A. Rohman, and T. S. T. Mansor, "Differentiation of lard from other edible fats and oils by means of Fourier transform infrared spectroscopy and chemometrics," *JAOCS, J. Am. Oil Chem. Soc.*, vol. 88, no. 2, pp. 187–192, Feb. 2011, doi: 10.1007/s11746-010-1659-x.

[26] 35 ISO Guide, "Reference materials – guidance for characterization and assessment of homogeneity and stability," Geneva, Switzerland, 2017.

[27] S. Adilson, O. Liz, F. Ivanic, L. T. Ricardo, and S. L. C. Ferreira, "Application of chemometric tools for homogeneity and stability evaluation during the preparation of a powdered milk laboratory reference material for inorganic analysis," *Anal. Methods*, vol. 12, no. 8, pp. 1055–1063, 2020, doi: 10.1039/c9ay02184a.

[28] S. Daoud, E. Bou-maroun, L. Dujourdy, G. Waschatko, N. Billecke, and P. Cayot, "Fast and direct analysis of oxidation levels of oil-in-water emulsions using ATR-FTIR," *Food Chem.*, vol. 293, pp. 307–314, Sep. 2019, doi: 10.1016/j.foodchem.2019.05.005.

[29] S. R. Karunathilaka *et al.*, "Rapid classification and quantification of marine oil omega-3 supplements using ATR-FTIR, FT-NIR and chemometrics," *J. Food Compos. Anal.*, vol. 77, pp. 9–19, Apr. 2019, doi: 10.1016/j.jfca.2018.12.009.

[30] L. C. Lee, C. Y. Liong, and A. A. Jemain, "Effects of data pre-processing methods on classification of ATR-FTIR spectra of pen inks using partial least squares-discriminant analysis (PLS-DA)," *Chemom. Intell. Lab. Syst.*, vol. 182, no. December 2017, pp. 90–100, Nov. 2018, doi: 10.1016/j.chemolab.2018.09.001.

[31] S. Wold, M. Sjöström, and L. Eriksson, "PLS-regression: A basic tool of chemometrics," *Chemom. Intell. Lab. Syst.*, vol. 58, no. 2, pp. 109–130, 2001, doi: 10.1016/S0169-7439(01)00155-1.

[32] L. Eriksson, J. Trygg, and S. Wold, "A chemometrics toolbox based on projections and latent variables," *J. Chemom.*, vol. 28, no. 5, pp. 332–346, May 2014, doi: 10.1002/cem.2581.

[33] M. H. Hanafy Idris *et al.*, "A conjunction of sn-2 fatty acids and overall fatty acid composition combined with chemometric techniques increase the effectiveness of lard detection in fish feed," *Chemom. Intell. Lab. Syst.*, vol. 213, no. March, p. 104308, Jun. 2021, doi: 10.1016/j.chemolab.2021.104308.

[34] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interdiscip. Rev. Comput. Stat.*, vol. 2, no. 4, pp. 433–459, 2010, doi: 10.1002/wics.101.

[35] 98-3 ISO/IEC, "Evaluation of measurement data — Guide to the expression of uncertainty in measurement," 2008.