CHAPTER 4

RESULTS AND DISCUSSION

4.1 Profiling of Lard from Collected Pig Samples at Northern, Central, and Southern Malaysia using FTIR combined with Principal Component Analysis (PCA)

4.1.1 Introduction

A total of 270 FTIR spectra were measured, and line plots illustrated lard spectra collected by body parts; BK (back fats), BL (belly fats), and SD (shoulder fats) from different regions; north, south, and central. The total of 180 training sets samples were divided into three regions; central (60), north (60), and south (60). The remaining 90 samples of the test set are from the central (30), north (30), and south (30) regions. First, the training and test set division were conducted by K-S algorithm selection. Afterward, pre-processings or transformations were applied to the training set of divided samples to improve the information of interest and reduce or remove the noise on spectral signals in Section 4.1.2. Then, PCA was performed on each transformed data in Section 4.1.3 to select and observe outliers. Next, the selected PCA model was assessed in Section 4.1.4 using Hotelling T^2. In the last part, the selected PCA model was validated by test set projection in Section 4.1.5.

4.1.2 Data Pre-processing

This section discusses the raw or untreated and pre-processing spectra to identify any spectral discrepancies. As a whole or globally, the spectra of fats collected from various regions cannot be distinguished significantly by the naked eye even though the data has been transformed, as shown in Figure 4.1 (a) – (c).

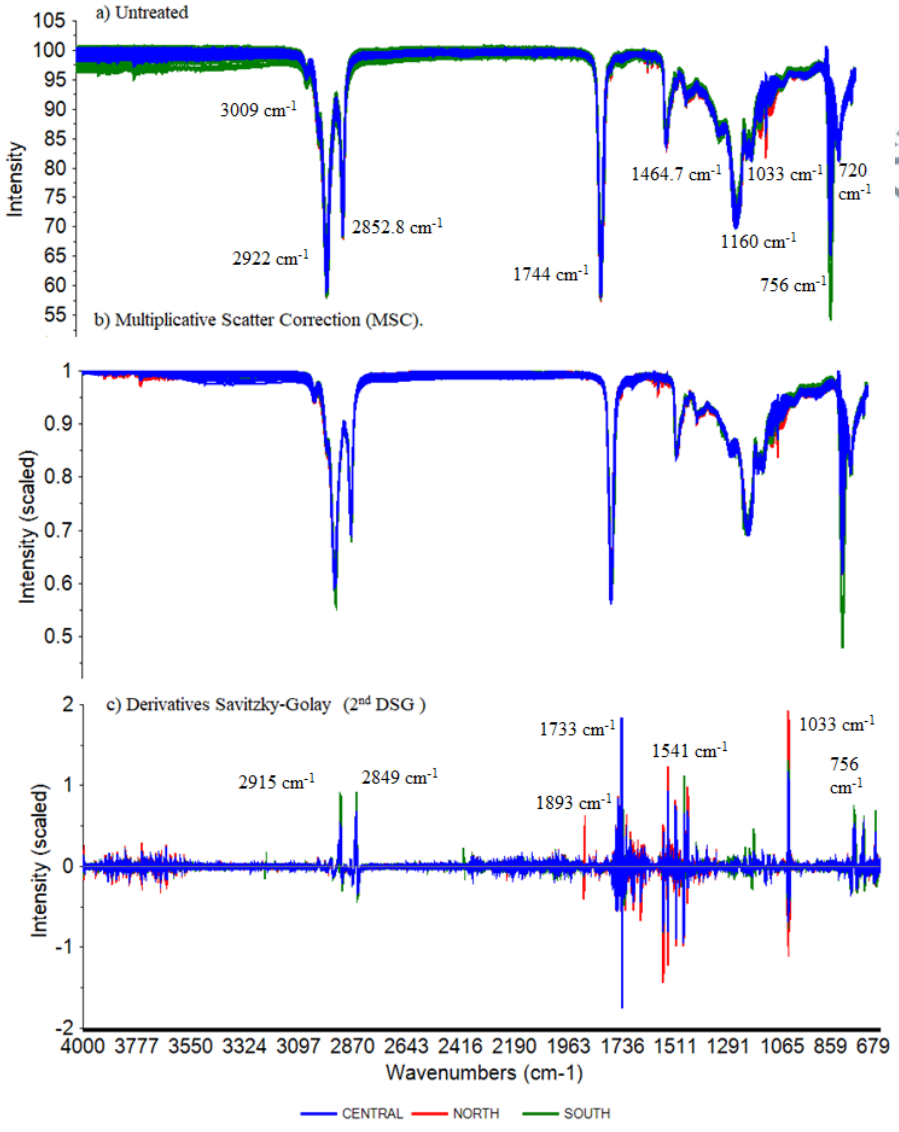


Figure 4.1: Comparison of the Untreated and Pre-processing FTIR Spectra.

First, an in-depth analysis of untreated FTIR spectra can be seen in Figure 4.1 (a) shows significant additive offset multiplicative scaling effects and light scattering. It can be noted that some south region samples (green lines) seem uplifting from the others between wavenumbers 4000-3085 cm^{-1}. Three peaks regions around 3059-2762 cm^{-1}, 1823-1696 $cm^{-1,}$ and 1500-680 cm^{-1} dominated the whole spectra. The first overtone and combination area of hydroxyl (-OH) stretching and bending vibrations can be observed at 3900-3100 $cm^{-1.}$. The sharp peak at 756 cm^{-1} is identified as from chloroform residues during the extraction process of crude lards. The oils and fats have a peak around 720 cm^{-1} which is attributed to the carbonyl (C=O) branch.

Previous works have attempted to identify lard between selected fats using a combination of FTIR spectra and chemometrics at the specific frequency wavenumbers. For the qualitative and quantitation determination of lard blended with other animal fats, it was found that to be almost different were observed at wavenumber regions 3010-3000 cm^{-1}, 1220-1095 cm^{-1} (Jaswir et al., 2003; Che Man & Mirghani, 2001); qualitatively studies of lard to differ with other edibles fats and oils at 2852.8 cm^{-1}, 2922 cm^{-1} and 1464.7 cm^{-1} (Che Man et al., 2011a). However, the finding of this study showed that the wavenumbers from spectra of lard to represent the entire lard profiles from various sources has similar wavenumbers with the previous works.

FTIR spectra were pre-processed or transformed by multiplicative scatter correction (MSC). As can be seen in Figure 4.1 (b), both MSC and untreated spectra show a similar pattern according to the regions (north, central & south). MSC corrections were demonstrated to solve baseline offset problems. After pre-processing, the baseline between 4000-3010 cm^{-1}

becomes tighter and closer to each other than the original. It can be seen that the uplifted samples from the original line plots of the raw data are unseen. Still, both regions and body parts samples showed a similar pattern to the raw data.

Other than MSC, the second choice and most common pre-processing is the 2nd DSG (Derivatives Savitzky-Golay) pre-processing to solve the multiple scattering effects. After the 2nd DSG pre-processing, the spectra of lard origin have also turned into some sharp peaks (Figure 4.1 (c)). Interference peaks clearly showed chloroform and methanol's sharp peaks at 1033 cm^{-1} and 756 cm^{-1}, respectively. In addition, multiple noise peaks were seen from the broadening regions (1823-1696 cm^{-1}, 1590-1480 $cm^{-1,}$ and 1500-680 cm^{-1}). In contrast, spectra lard such as peaks 3003 cm^{-1}, 3010 $cm^{-1,}$ and small peaks around 1500-680 cm^{-1} cannot be identified and were lost after 2nd DSG pre-processing (Eid et al., 2020).

The chemometrics strategies have found that a difference in light scattering could cause the variation of baseline shift. 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 an issue in the FTIR measurement of oils (Daoud et al., 2019; Karunathilaka et al., 2019). Thus, the pre-processing of FTIR spectra is necessary as a correction part before chemometrics modelling.

4.1.3 Principal Component Analysis (PCA)

Figure 4.2 compares PCA according to the raw and pre-processing of FTIR data as; a) Untreated PCA-FTIR, b) Full spectra of MSC-PCA, c) Selected frequencies of MSC-PCA and d) 2nd DSG-PCA.

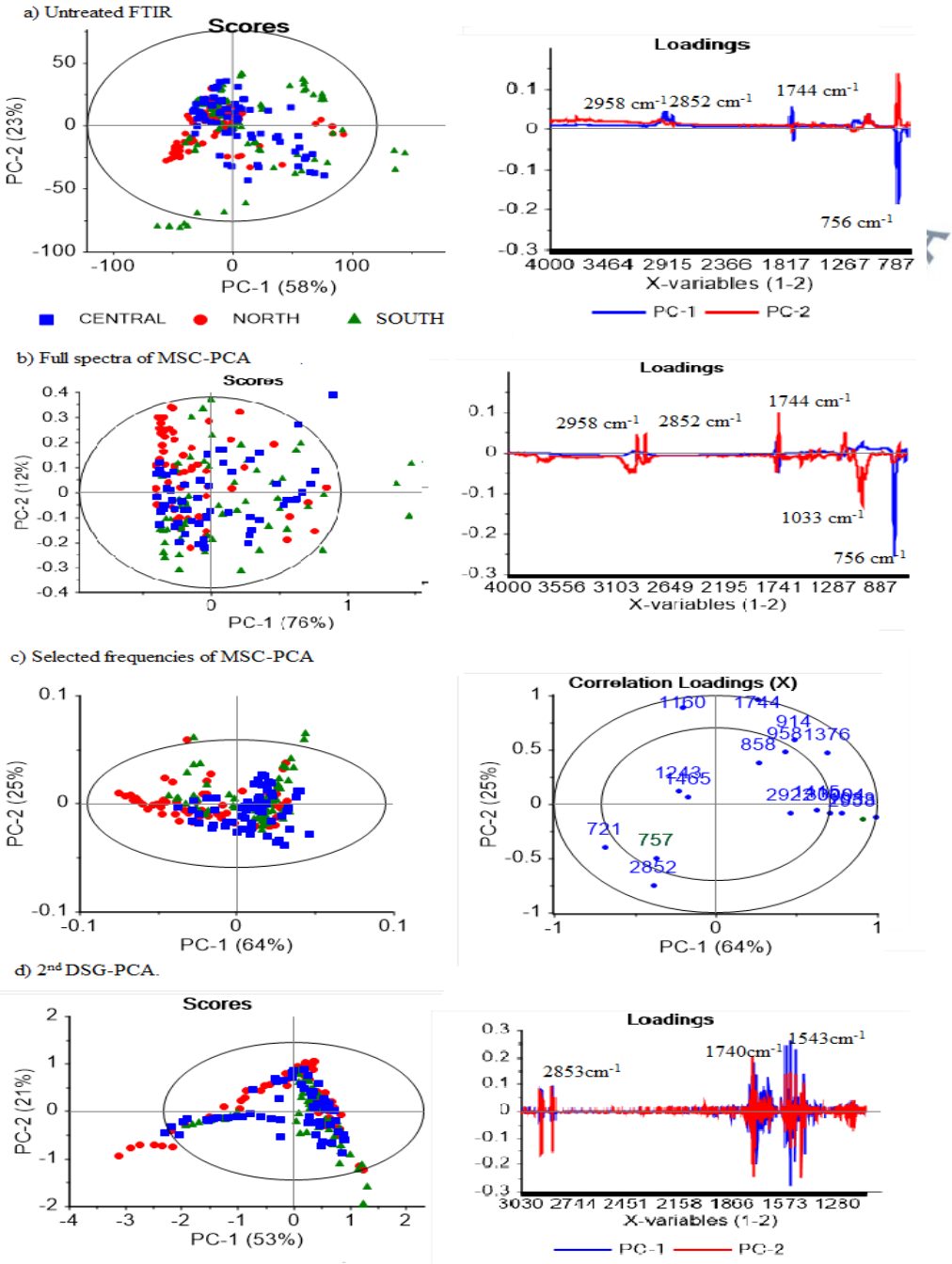


Figure 4.2: Comparison of the FTIR-PCA Before and After Pre-processing.

In this study, PCA was performed on raw FTIR spectra 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 spectra of untreated FTIR data showed no differences in lards profiles by regions or body parts. In PCA, the important data was compressed into the PC. Furthermore, FTIR data that were treated with MSC and 2nd DSG pre-processing were also subjected to PCA.

The outcome of PCA on raw FTIR data shows that the lard samples explained a total of 81% variances by the first two Principal Components (PCs). 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} and 1033 cm^{-1} as can be seen in Figure 4.2 (a), which are identified as chloroform and methanol 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 of original data and may highlight regions of high importance.

Figure 4.2 (b), shows the MSC-PCA full spectra has similar trends in raw PCA and similar total variances of as much as 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 in PC-2 north region samples outweighed at positive PC-2 direction and south region at negative PC-2 direction.

As can be seen in Figure 4.2 (c), only 89% of the variances are captured after variables selection on MSC-PCA. This strategy is because variables were selected according to the lard individual's frequencies to overcome the interferents peaks issues. Therefore, variables (frequencies) were selected in the correlation loadings plot in Figure 4.2 (c) of the loadings plots compared to the standard view loadings plot in Figure 4.2 (b) for better frequency visualisation.

The plot in correlation loadings (Figure 4.2 (c)) included the two ellipses or circles representing correlation *X*-loadings 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 individually 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 prominent peaks related to lard are in agreement with the previous research (Naquiah et al., 2017; Sim et al., 2018).

The others pre-processing applied is 2nd DSG. Figure 4.2 (d) shows that 2nd DSG-PCA captures 74 % of the variances at the first two PCs. Unlike MSC pre-processing, loading plots from the 2nd DSG-PCA were not determined by selecting individuals' frequencies. The sharp peaks produced by the 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 two PCs display noise in 1823-1696 cm^{-1} and

1500-680 $cm^{-1,}$ notably at positive PC-1. This outcome confirms the statement that the method of 2nd DSG-PCA gives much change to the spectra of the origin (Jiang et al., 2019; Panchuk et al., 2018).

The 2nd DSG pre-processing can correct the multiplicative effect by removing both baseline and linear trends. However, the techniques are not helpful for this study because some noise spectra would be inflated. Hence, the MSC pre-processing is proper to be applied to conserve the spectral features of the lard.

From the three PCA models of the untreated and transformed data, there are no distinction grouping either by region or body parts, even after the interferent removal. Therefore, the selected MSC-PCA was chosen as the final model for the similarity test and to validate the model. Then extended PCA employed Hoteling T^2 for outliers and critical limit evaluation.

Prior PCA, MSC pre-processing, and untreated lines plot spectra are not changing much differs. Therefore, the MSC pre-processing is appropriate in looking for similarities of lard. The importance of pre-processing shows that outliers in the ATR-FTIR spectra and removal of irrelevant signals come from sample contamination (Lee et al., 2018; Mishra et al., 2020), and the finding of this PCA evaluation study also enhanced the visualisation of the important wavenumbers.

PCA proved that the outliers' evaluation is essential before the finalised model. In this view, some wavenumber selections are necessary to refine data identifying the interfering peak, such as solvents inherited from fats extraction. Furthermore, untreated and transformed FTIR data could not cluster lard according to the regions and body parts. These findings suggested

that there is a homogeneity of the spectral lard qualitatively. The finalised MSC-PCA calibration model compressed multivariate data into the first three PCs to represent lard spectra from selected frequencies revealed by loadings plot.

4.1.4 Hotelling T^2 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, Hotelling T^2 statistic was proposed to test the α of the multivariate data. The Hotelling T^2 hypothesis of a multivariate normal distribution is often used to identify outliers or abnormalities during multivariate process regulation (Rinnan et al., 2009).

In Figure 4.3, a total of eight samples are spotted as outer ellipse at 95% CI of Hotelling T^2 limits MSC-PCA model on PC-1 *vs* PC-3.

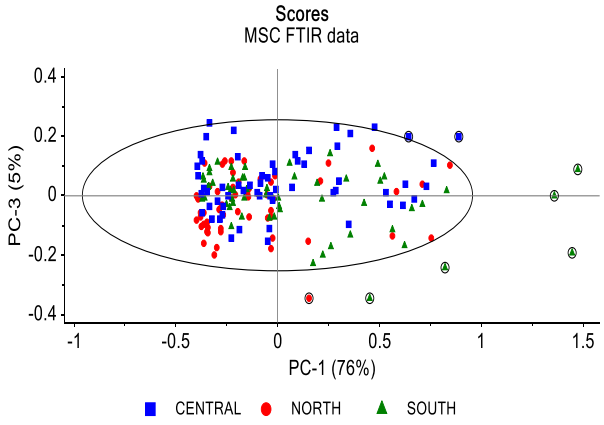


Figure 4.3: Determination of Outliers by MSC-PCA.

Hotelling's T^2 statistics 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 T^2 eclipse remain the same spectra. In

comparison, those outsides of the line samples are determined as extremists or outliers. Unscrambler® X software removed the outliers from the MSC-PCA by recalculating this model and updating the statistical distances, and the results presented in Table 4.1.

Table 4.1: Comparison between Four Different Significance Levels (α).

	Hoteling T 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

Comparison of four different significance levels and the outliers for the first 3 PCs after removing eight outliers outside the 95% CI. There are no outliers at the first 3 PCs at 0.1 % and 0.5% α, after removing the outliers. In order to accommodate Type II error, a longer distances Hoteling T Limit 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. In the future, the α of the MSC-PCA model of the collected lard by Hotelling T^2 statistic has been determined by accepting the null hypothesis of the lard profile regardless of the regions and body parts.

The first 3 PCs (total variances >80%) were retained, and the threshold to an arbitrary critical value was set at 0.5% α, referring to the Hoteling T^2 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 3 PCs of MSC-PCA were chosen to determine the similarity of lard spectra and subsequently to predict the new samples.

The results in Table 4.1 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 extremes outliers. Thus, a *p*-value at 0.5% α (99.5% CI) was chosen after considering the elimination of the extreme outliers. Interestingly, the generalisation of Hoteling T^2 statistics as a certified reference material (CRM) benchmark was not found in lard or selected fats, especially in the spectroscopic analysis. The influence plots are typically used to identify outliers or abnormalities when a procedure runs outside standard parameters. The statistical distance can be used to measure the significance level in the multivariate data (Efron, 2014).

4.1.5 Prediction of Lard by MSC-PCA Projection

After the MSC-PCA model was finalised, 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 ratings, t, and loadings p (*See* Equation 2.2). 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 plots observed how close the new samples were predicted according to the distances of the original samples. The training datasets were used to construct the MSC-PCA model and denoted as blue, whereas green was the test set's latest prediction, are presented in Figure 4.4.

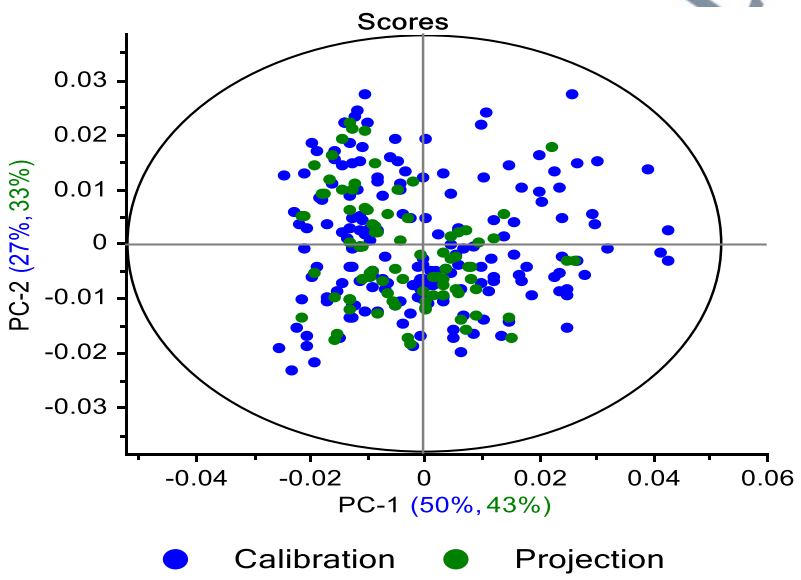


Figure 4.4: Projection of the Test Set into MSC-PCA.

As expected, all the scores plot of the test set (green) are located inside the eclipse. Therefore, MSC-PCA produces an acceptable percentage calibration (at the right) and prediction (at the left); PC-1 (50%, 43%), PC-2 (27%, 33%), and PC-3 (13%, 15%).

The 2D graph in Figure 4.4 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 T^2 weighted with their respective critical limits based on α above (Table 4.1). The test set samples re-assembly inside the first 3 PCs (*A*)

boundaries in the Influence graph plot. As can be seen in Figure 4.5, The F-Residual value versus Hotelling's T^2 are presented in Figure 4.5 (a) PC-1: 0.01, 8.130; Figure 4.5 (b) PC-2: 0.007, 11.054; and Figure 4.5 (c) PC-3: 0.005, 13.825.

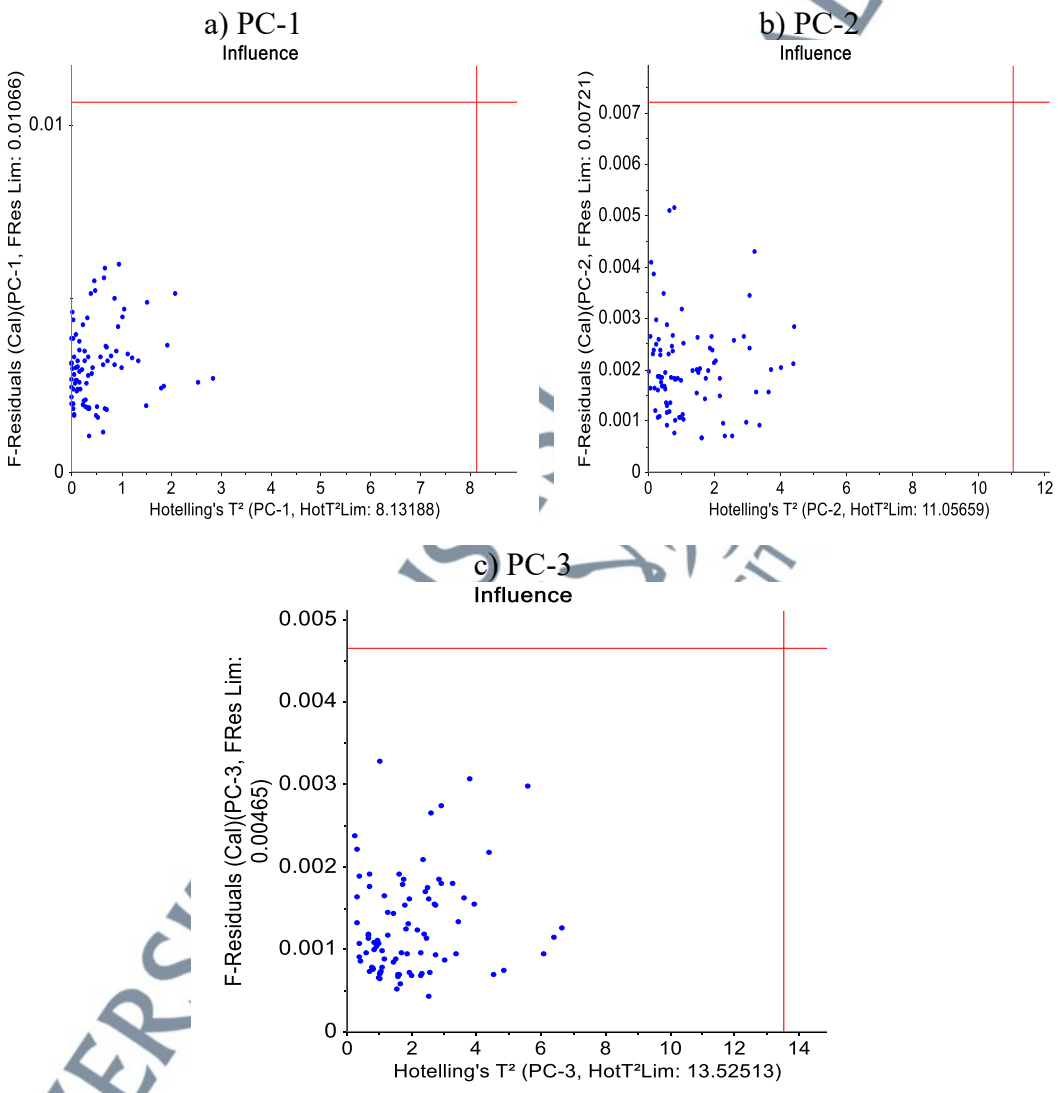


Figure 4.5: Influence Graph Plots F-Residual *vs* Hoteling T^2.

Hence, the first 3 PCs distribution so each PC were visualized individually by influence graph. Finally, F-Residual value *vs* Hotelling's T^2 was set to use as boundaries to conclude the test set in belonging to the lard.

All projected test samples were shown inside the red line (critical limit), and there were no extreme samples. Hotelling's T^2 has been used as a parameter of the data quality in the multivariate study. For instance, biodiesel samples' model assessment and error analysis quality by spectrophotometry data were performed using the Hotelling T^2 statistic and Q-Residual index to replace F-Residual (Tehranizadeh & Ghazanfari, 2020). The Influence graph was also conducted in chemometric modelling to UV-Vis spectroscopy (Angheluta et al., 2020).

The PCA projection outcomes can also be validated as the first 3 PCs through a residual variance graph in Figure 4.6 to indicate the quality of the projected samples.

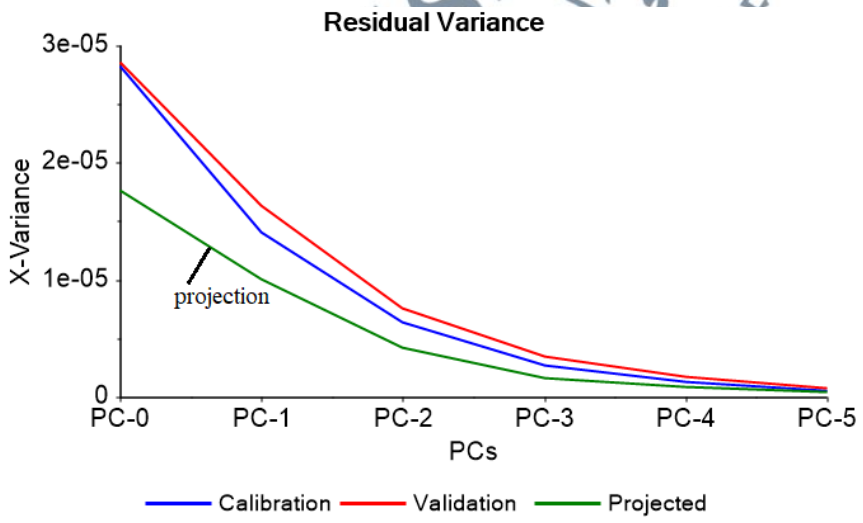


Figure 4.6: Total Residual Variance of MSC-PCA.

The scores vector outside these boundaries indicates that test samples were not in lard groups. Hotelling's T^2 statistic measures the distance of a new observation to the model mean. The critical limit is based on an *F*-test, which 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, which is represented by the red curve. The residual variance of the predicted samples is often seen as a green curve in the projection case (Figure 4.6). The resultant MSC-PCA models produced the smallest values below zero of total residual variance, where the residual variance dropped sharply to zero from the fewest possible components. The projected curve below the calibration and validation curve showed reliable prediction results.

Visualisation 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 3 PCs have given a satisfactory smaller residual for the better quality of the model. It can be seen from the results of the prediction curve, which are parallel below the calibration/validation curve. This outcome means the new projected samples (test set) belong to the calibration.

The statistical inference method was used by applying a random effect model to determine the extent to which the built-in model can be re-applied. Prediction variances were slightly higher at PC-2 and PC-3 but acceptable because the matrix X is not always perfectly reconstituted after re-sampling into the training set especially involving spectral data, which are typically inherited by the small number of objects, n and large numbers of the variables, p (Abdi & Williams, 2010).

The finding shows that PCA projection with other Influence and Residual Variances graphs are highly recommended to validate the similarity

of the new samples in case the model's saturated or homogenous scores vectors exist. Otherwise, to find dissimilarity between scores plots, a combination of PCA and OPLSDA has been found to discriminate lard in fish lipids successfully (Hanafy et al., 2021).

The profiling of lard and its reliability have to be improved. Thus, the study was conducted to compile lard from different regions and body parts and to encourage using a standard library of lard samples. Ideally, such a library would be composed of lard spectral and unconventional CRMs with well-characterized chemometric data. Traditionally, ANOVA has been implemented as a univariate statistical method in the process of the uncertainty test assessment of the CRM (ISO Guide, 2017). On the other hand, chemometrics techniques such as PCA and HCA have been used to quantify the CRM of inorganic material and alternative tools in multivariate data (Adilson et al., 2020). Thus, this study proposes the high-dimensional data generated by FTIR and chemometrics techniques to develop the CRM.

4.2 Discrimination of Pigs (Lard), Chicken, Beef, Mutton, and Plant Fats After Heating-Process using FTIR, ^{1}H-NMR, and ^{13}C-NMR with Chemometrics Techniques

4.2.1 Introduction

In this study, fats subjected to the heating-process were measured using FTIR, ^{1}H-NMR, and ^{13}C-NMR combined with chemometric techniques such as PCA, multivariate classification, and multivariate regression.

The FTIR and ^{1}H-NMR findings were outlined according to the chemometrics evaluation: Data pre-processing, PCA, multivariate classification, and multivariate regression, and then determined the chemical features that contributed to the lard using OSC-PLSR. Multivariate classifications were performed to classify lard and selected fat according to the respective groups. Multivariate classifications employed were LDA, MDA, QDA, and SVMDA. Meanwhile, multivariate regression employed the OSC-PLSR, PCR, and PLSR to discriminate lard and other fats by the dummy binary (0/1).

A continuation study of ^{13}C-NMR analysis was conducted on lard and chicken fat overlapped with the ^{1}H-NMR-PCA scores plot. The chemometrics evaluation on ^{13}C-NMR data was repeated from pre-processing, PCA and heatmap for biomarkers of lard discovery.

4.2.2 Profiling of Lard and Selected Edible Fats After the Heating-Process.

4.2.2.1 FTIR

FTIR data was obtained from the range 4000-650 cm^{-1} as the starting point for spectral acquisition. The spectra of the FTIR dataset are measured and contain 339885 data points (195 samples x 1743 wavenumbers). Figure 4.7 illustrates the total of samples spectra simultaneously according to the five types of edible fats after the heating-process denoted by; (a) chicken, (b) lard, (c) beef, (d) mutton, and (e) plant fats.

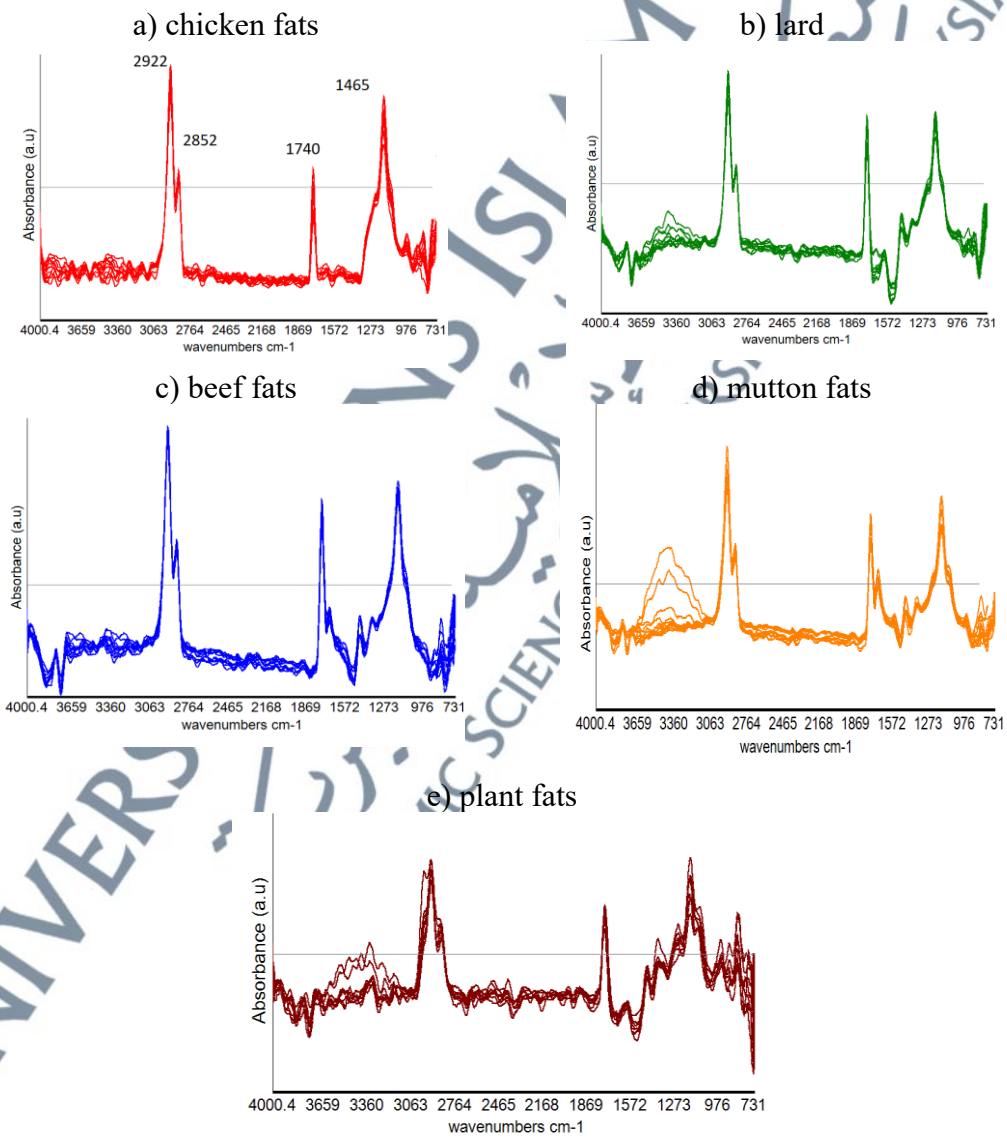


Figure 4.7: Comparison of FTIR Spectra.

Spectra in the analysis of fats subject to the heating-process of each figure are presented globally or simultaneously to align with the chemometrics evaluation concept. Each spectral was imposed according to the fats type and heating-process design 240 ºC; and temperature (0.5, 1, 2 & 3 hrs). The extreme intensities at lower bands between 650-709 cm^{-1} were considered to be eliminated and known as interferent peaks. The interferent peaks can mask the other important peaks of fats, which can be observed through the loadings plot of PCA (Appendix 1).

The scattering or high magnitude at the lowest bands may come from light dispersed from the semiliquid fats. Three sharp peaks are identified as important characteristics of fats. The typical spectra of edible fats for typical TAG are shown at 2922 cm^{-1} and 2852 cm^{-1}. These two peaks can be associated with the stretching vibrations of methylene ($-CH_2$) and methyl ($-CH_3-$) groups, respectively. Due to bending vibrations, the methylene group shows bands at 1465 cm^{-1}. The sharp C=O absorption peak of ester linkage at 1740 cm^{-1} was clearly observed for all animal fats.

Bands in the range 900-1800 cm^{-1} have been shown to differ from animal and plant fat. Chicken fat and lard have very similar characteristics at a glance under the naked eye (Figure 4.7 (a) & (b)). While beef and mutton have similar characteristics with slight variations. Beef and mutton fats are shown some different spectra around 2850-2950 cm^{-1} at 240 °C (Figure 4.7 (c) & (d)). The broad bands 3100-4000 cm^{-1} were identified as moisture from water residues in some lard, beef, and plant fats. Plant fats have little variation 1500-731 cm^{-1} compared to animal fat (Figure 4.7 (e)).

The main indicator is that the 2922 cm^{-1} can be associated with the asymmetrical stretching vibrations of methylene ($-CH_2$), while the 2852 cm^{-1} methyl absorbance was the result of a symmetrical stretching band of ($-CH_3$-) groups, and sharp carbonyl (C=O) of triglycerides absorption peak approved at 1740 cm^{-1} as in the structure is shown in Figure 4.8. The TAG, commonly called triglycerides (known as lipids in general), are formed by combining glycerol with three FA molecules at different numbers of carbon chains denoted as *R*. The chain lengths and saturation of the FAs in natural TAG vary depending on their sources, e.g., plants, animals, or bacteria, sometimes at minimal variances. FTIR is remarkable for identifying lard if combined with chemometric techniques on lard analysis, especially when involving massive data.

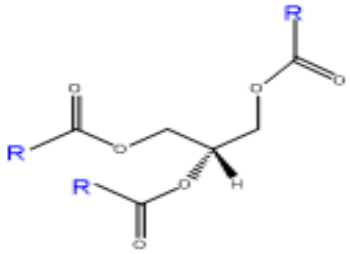


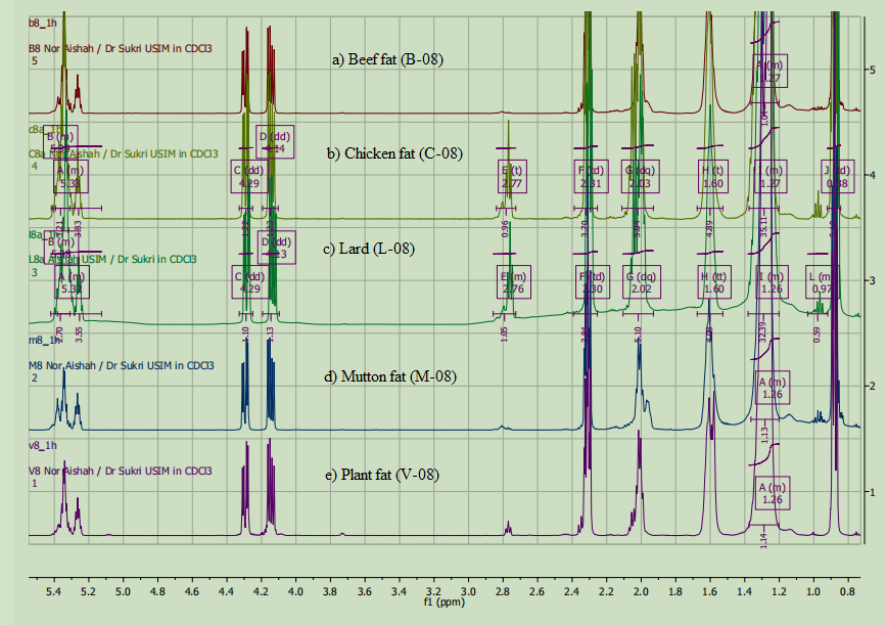
Figure 4.8: The Basic Structure of TAG.

4.2.2.2 ^{1}H-NMR

The results of ^{1}H-NMR spectroscopy produced a total of 136 spectra (i.e., 65 samples were duplicated; additional triplicate on lard at untreated and 120 °C) from lard, chicken fats, beef, mutton, and plant fats subjected to heating-process and labelling (*See* at Table 3.4). Each peak value of chemical shift (δ), denoted in ppm units, was recorded. The δ were utilized at ranges, as can be seen in Appendix 4; each peak of δ has been assigned to the corresponding functional group (Lia et al., 2020; Sacchi, 1997).

Figure 4.9 (a) shows the ^{1}H-NMR spectra of the five fat types after the heating-process. The spectra of various edible fats heated at 180 ºC for 3 hrs were stacked to observe the specific signal/marker to differentiate among fats of different types. Some signals have varying intensities at δ 0.95 ppm and δ 2.76 ppm, especially by lard and chicken. However, intense signals of TAG in the aliphatic region (δ 1.2 - 1.4 ppm) dominate the spectra.

a) Comparison of Individuals ^{1}H-NMR Spectra.



b) An Overlay A Total of 136 ^{1}H-NMR Spectra.

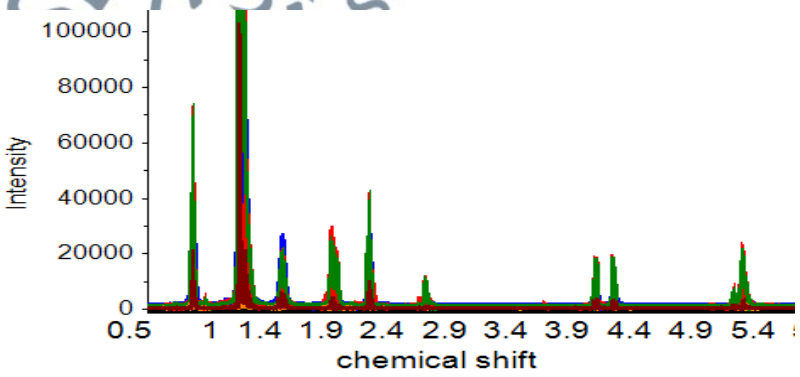


Figure 4.9: ^{1}H-NMR Spectra.

It could be time-consuming to observe 136 samples individually. Thus, the spectra were superimposed for further multivariate analysis. The data from MestReNova software were binned and transferred into ASCII files before being imported to the Unscrambler® X software, is shown in Figure 4.9 (b). There was no specific signal/marker to differentiate between fat varieties. Some signals have little varying intensities on each resonance. For example, the concentrations of methyl groups of the acyl chains (in the δ 1.26 ppm) are dominated with large intensities, but the others δ could not discriminate against fats varieties.

4.2.3 Data Pre-processing

4.2.3.1 FTIR

The different of the spectra after two data pre-processings as; a) normalisation and (b) 2nd DSG, are illustrated in Figure 4.10.

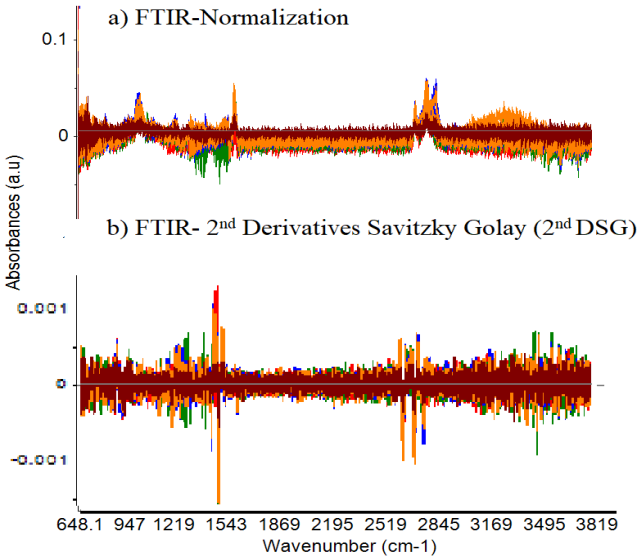


Figure 4.10: Comparison of Normalised *vs* 2nd DSG FTIR Spectra.

Before pre-processing, the raw FTIR data matrix, X, samples (m), and wavenumbers (n) or (m=195, n=1743) were divided by training (145) samples and test sets (50 test) by the K-S algorithm. First, the training set was applied to the pre-processing at selected peaks, 715-3100 cm^{-1}. Next, the identified interferent peaks at wavenumbers ranging from 650 -709 cm^{-1} and 3101-4000 cm^{-1}, were excluded.

The selection of two types of pre-processing are considered to improve the baseline of raw FTIR spectra data (*See* Figure 4.7). The raw spectra apparently displayed relatively high background noise and scattering effects. According to the literature, normalisation and 2nd DSG pre-processing are suitable to overcome the scattering effect (Rinnan et al., 2009).

Normalisation is spotted differently from 2nd DSG in the following two plots aspect. First, as can be seen in Figure 4.10 (a), normalisation pre-processing is similar to the originals spectra features (*See* Figure 4.7 (a)-(e)). The only magnitude of absorbance is standardized to be between 0.2 and (-) 0.2 a.u. Meanwhile, 2nd DSG (Figure 4.10 (b)) has centralised intensities to the zero baselines. Second, peaks lipids region; 1098-1235 cm^{-1}, 1375-1740 cm^{-1}, and 2852-2922 cm^{-1} remained after normalisation pre-processing and allowed spectra to represent lipids composition. Noted that previous studies indicated three-band regions could be notably lipids or fat (Arslan & Çağlar, 2019; Guillén & Cabo, 1997; Lerma-García et al., 2010). On the contrary, by producing sharper peaks, the 2nd DSG enhanced the main peaks at 1166-1370 cm^{-1}, 1499-1746 cm^{-1}, and 2800-3025 cm^{-1}.

4.2.3.2 ^{1}H-NMR

^{1}H-NMR profiling on untreated spectra from previous Subsection 4.2.2.2 (*See* Figure 4.9 (b)) encountered drifting issues on spectra because many samples were conducted simultaneously. Although the analyse were appropriately referenced, chemical shift drift was inevitable. The δ drift may happen due to many factors, such as compounding, changes in pH samples, ionic strength, instrumental factor, and level of compound dilution. Thus, the spectra are required to be pre-processed for comparability.

Figure 4.11 shows that combination pre-processing techniques were applied to the ^{1}H-NMR spectra simultaneously, according to the a) Correlation Optimized Warping (COW), b) COW-Normalise, c) COW-Normalise -Baseline.

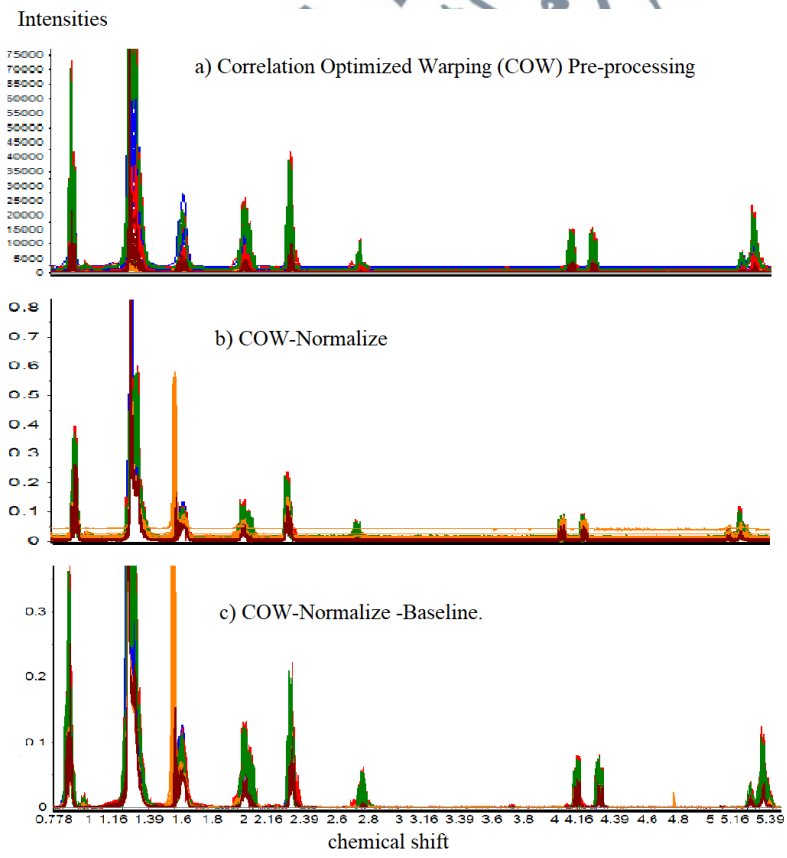


Figure 4.11: ^{1}H-NMR Spectra After Pre-processing.

The region of *d*-TMS and *d*-CHCl3 (0.0-0.5 ppm & 6.1-8.0 ppm) were firstly removed, and then combination pre-processing was applied due to the δ drift of the binned spectra. Some baseline offsets seemed obvious and required baseline correction on the spectra. These popular alignment spectra, such as the COW method, need referenced spectra representing the entire samples. After COW pre-processing of spectra, the unit of intensities 0-100 x 10^3 is required to be scaled (Figure 4.11 (a)). The normalisation by full spectra to a unit area (0-1) should enable an unbiased classification of the fat samples based on their compositions, is shown in Figure 4.11 (b). Normalisation is a method of good change to the differ ^{1}H-NMR spectral variation due to intensity or concentration variation. However, some extreme values and scattering effects are noticeably uplifting after normalisation. Figure 4.11 (b) shows the spectral display systematic baseline variations after COW-Normalise pre-processing was applied. Such problems were suggested to be corrected by using baseline pre-processing techniques to subtract the baseline offset (Figure 4.11 (c)).

The alignment of the ^{1}H-NMR spectra was satisfactory using the COW-Normalise-Baseline pre-processing algorithm. Hence, a thorough assessment of the sample set was carried out to verify the occurrence of spectral outliers. In this context, PCA was performed.

4.2.4 PCA

4.2.4.1 PCA of FTIR data

The normalised-PCA is shown in Figure 4.12; plant fats and the first two PCs separated animal fats with a total variance of 80%.

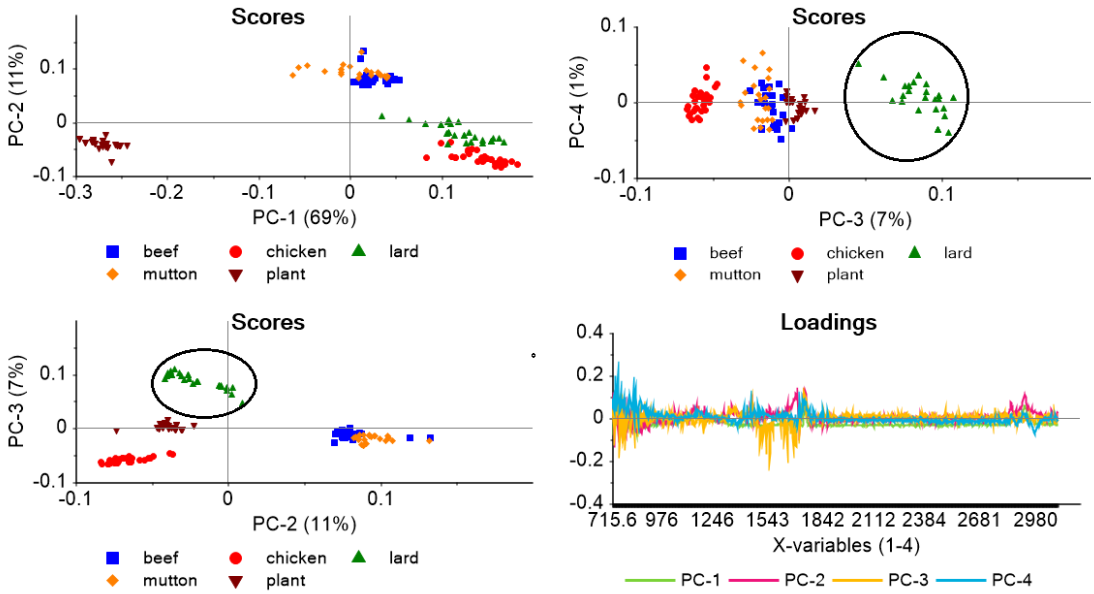


Figure 4.12: The Normalised-PCA Overview of the FTIR Data.

It can be seen that the highest variances in mutton/beef fats and chicken/lard overlapped and were anti-correlated with plant fats. PC-1 separated well between animal and plant fats into two groups: mutton and chicken. PC-2 separated mutton and beef samples from the rest using the first two PCs.

Beef-mutton is located separately from the selected fats in PC-2 *vs* PC-3, accounting for 18% total variance. Lard is located positive at PC-3 and anti-correlated with chicken fats (in circled), as can be seen in PC-3 *vs* PC-4, which accounted for only 8% of the total variance. At the same axis, beef-mutton and plant fats cluster each other in the middle of the axis as average plots. Therefore, the PC-3 *vs* PC-4 can be defined as lard and chicken fats as their clusters separation from the selected fats, but the total variances explained are too low for the defensive results.

Normalised-PCA loadings plots have defined the wavenumbers, displayed poorly, and required in-depth investigation. The alternative to

define wavenumbers or frequencies could be applied by scores plot selection according to the distinctively each fats clustering, as shown in Appendix 2. Each PC represents the wavenumbers contributed by the loadings plot. The wavenumbers ranges are addressed as band regions in the data FTIR. For example, the PC-1 *vs* PC-2 plot shows that the separation of the plant fats was mostly contributed by the band region 721-854 cm^{-1}. The PC-2 *vs* PC-3 plot shows that band region 1680-1751 cm^{-1} significantly separated the cluster mutton and beef fats from the rest.

As can be seen in Figure 4.13, the 2nd DSG pre-processing of PCA shows almost similar clustering to the normalised-PCA.

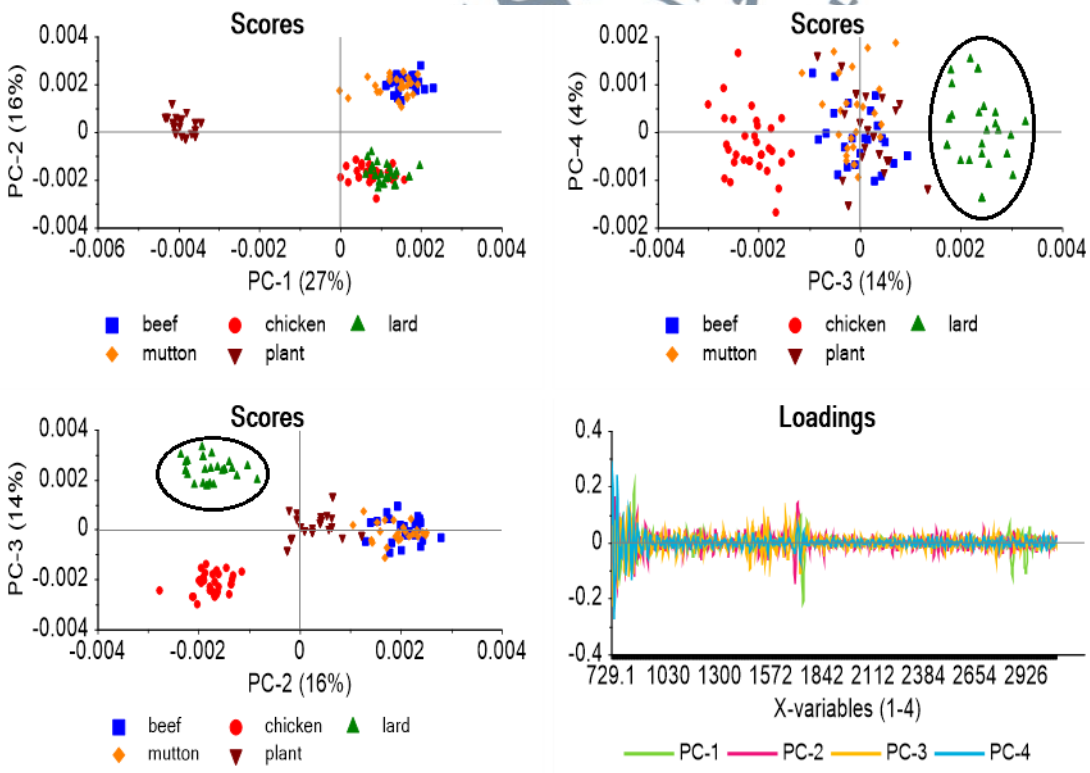


Figure 4.13: The 2nd DSG-PCA Overview of the FTIR Data.

The first two PCs (total variance 43%) separated plant fats into a tight cluster and animal fats into two tight clusters: mutton - beef fats and chicken fats and lard clusters. The clustering patterns of scores plots are very similar to normalise-PCA, where plant fats most contributed differently at PC-1 *vs* PC-2, accounting for 43%.

Separation fats, especially mutton-beef, could be better visualised on PC-2 *vs* PC-3 (in circled), accounting for total explained variances of 30%. Mutton and beef fats were slightly overlapped as circles. A cluster of lard is located opposite the chicken fats by PC-3, as shown in the scores plot of PC-2 *vs* PC-3 (in circled). Animal fats contributed by beef, and mutton closed together with plant fats in the middle of the axis of both PC-3 *vs* PC-4. The contribution of wavenumbers through the loading plot could not be observed well because 2nd DSG-PCA mostly enhanced spectra unrelated to the fats.

Plant fats are differed from animal fats by the first two PCs as shown in the PCA scores plot, the plant fats clustered among themselves tightly. This outcome could suggest those plant fats with palm oil hydrogenated (shortening) are very resistant to heat compared to the other fats, i.e., animal fats.

X-loadings plots can observe the correlation between each fat and wavenumbers according to the most scores plot on the PCs presented in Appendix 2. Both PC-2 *vs* PC-3 and PC-3 *vs* PC-4 separated lard from others fat, especially chicken fats. But the PC-3 *vs* PC-4 was chosen as the PCA model of lard separation because lard and chicken fats cluster are located at opposite PC axis. Thus, it is easier to identify wavenumber related to the lard from chicken fats. From the 1D graph in Appendix 2, lard separated from

chicken fat by negative PC-3, which explained 8% of the total variances using the third and fourth PCs. It was found that the region's band at 1225-1468 cm $^{-1}$ and 1726-1766 cm^{-1} are attributes of lard.

The pre-processing prior to PCA also demonstrated the importance of these techniques on FTIR data. Previous studies on 2nd DSG are useful in solving the baseline problem of spectra, but in some cases, it is important to consider the other pre-processing techniques as a comparison (Rinnan et al., 2009). As revealed in this study, normalisation has been found to be more appropriate and the same pattern clustering pattern as in the first two PCs of the 2nd DSG.

The covariances between lard and chicken, beef and mutton, are obviously in the first two PCs or the most variances. The normalised-PCA model was therefore chosen as an indicator as plant fats that are anti-correlated with animal fats at the most variances vectors. The unique lard cluster can only be observed on more minor variances at PC-3 *vs* PC-4 through at a much lower % of explained variance. This outcome also indicates that lard and selected fats have highly similar spectral profiles.

A previous study by Che Man et al., (2011a) focused on the first component that has the most variances contributed to the separation of lard between plant fats at a selected individual frequency in 2853 cm^{-1} region 2922 cm^{-1} and 1465 cm^{-1}. Rather than pointing out the main component (first two PCs), the new aspect revealed in this study indicated that the scores plot contributed to the lard are from the following component, which is of minor variances such as PC-3 *vs* PC-4 but could give important information of lard after heating-process.

4.2.4.2 PCA of ^{1}H-NMR data

The PCA was performed using a spectral set containing 136 samples in Figure 4.14 (a) Untreated samples, shows scattered into the two clusters. The two PCs explained total variances of 97%, but fats are not clustering according to their groups. Figure 4.14 (b) shows after pre-processing the ^{1}H-NMR data using the COW-Normalise-Baseline.

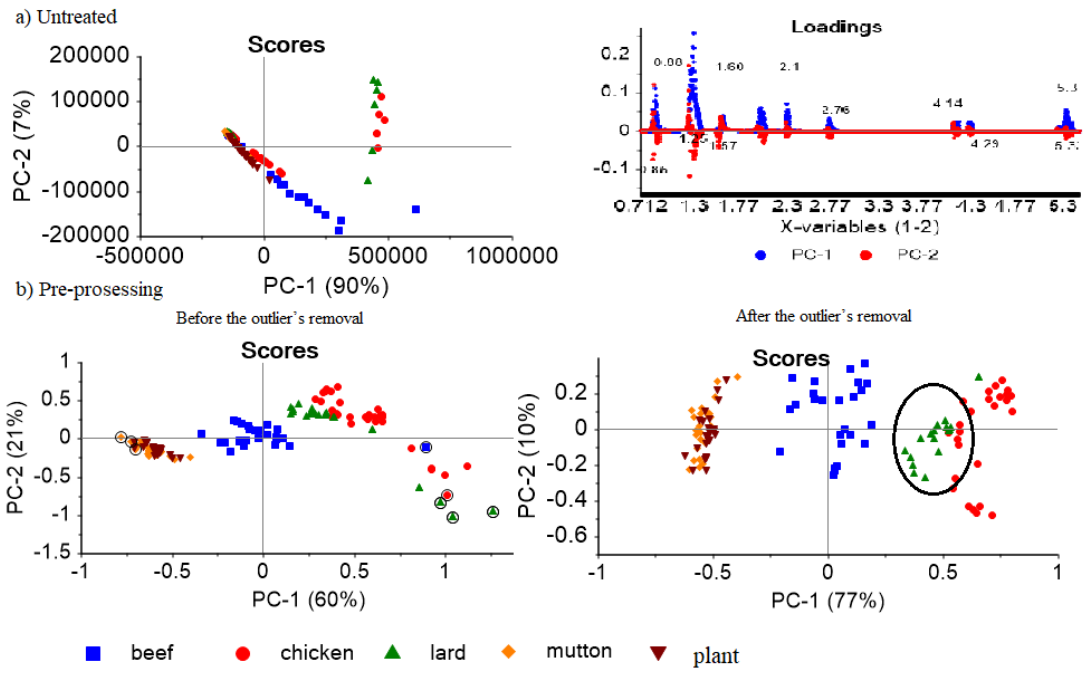


Figure 4.14: The ^{1}H-NMR-PCA of Untreated and After Pre-processing.

PCA of COW-Normalise-Baseline pre-processing shows some samples in circled, identified as outliers (at confident level 95%) of the first two PCs explained by 71% variance. Then, outliers' removal improved the clustering of each fat in the model in the scores plots making the variances of the first

two PCs slightly increase to 87%. However, the untreated and PCA pre-processed data loadings plots were fairly similar.

The plot of the beef fats is isolated from other fats in the middle of the axis. Chicken fats were almost mixed with lard clusters and closed together at positive PC-1 (in circled). However, mutton, fats and plant fat were tightly clustered by PC-1 but slightly spread by PC-2 and displayed similar clusters at negative PC-1.

The loading plots indicated a resemblance with the original data. The major δ by peak were labelled by looking into the PC-1 and PC-2 loading plots. Lard and chicken fats were separated from other fats at positive PC-1 as opposed to mutton-plant fats being far separated at negative PC-1. The beef fats were located at the middle of the axis, and this can be explained that the beef fats have an average δ. The δ at 1.22-1.30 ppm dominated the loadings plot and was mainly contributed by lard and chicken fats. This δ could be assigned to the acyl chain groups.

For the evaluation, discrimination of lard from other fats ^{1}H-NMR data using the chemometrics techniques has been demonstrated to be satisfactory to some extent, but it has been found that ^{1}H-NMR data of lard and chicken fats have high similarity, and this becomes challenging to distinguish between them using the experimental parameters.

The results obtained from untreated spectra show that ^{1}H-NMR has drifted during measurement, and some chemometrics strategies could be applied to overcome the issue. This drifting issue has been found on ^{1}H-NMR and explained by other researchers as well (Khakimov et al., 2020; Lefort et al., 2021; Savorani et al., 2010), but upon alignment pre-processing, the

results of the five types of edible fats used in this study demonstrated that PCA could distinguish lard from other fats, especially by PC-1. In addition, the novel concept of integrating several ways to baseline correction demonstrates that the various strategies used in this study could provide complementary information on edible fats.

Numerous studies to identify FAs using ^{1}H-NMR have been reported extensively, especially on plant oils (Barison et al., 2010; Roberta et al., 2020). However, research on identifying lard and other edible fats or oils using ^{1}H-NMR was limited. In addition, PCA and PLS-DA models using ^{1}H-NMR data have also been reported to be inferior in terms of sensitivity and specificity for classification when compared to GC-MS (Fang et al., 2013).

In this study, the ^{1}H-NMR data were extended to evaluate its capability to differentiate lard from selected fat after the heating process, which has shown satisfactory results. In addition, the suitability of ^{1}H-NMR and PLS algorithm for determining butter that has been adulterated with lard, using all proton-bearing components, was reported to be a good approach (Fadzillah et al., 2015).

4.2.5 Multivariate Classification (MVC)

The probability for MVC model is reliable for testing model robustness in FTIR and ^{1}H-NMR data. Classifications in each different fat were tabulated in Table 4.2 (a) for class sensitivity (*Sens*) and (b) Specificity (*Spec*), and (c) for MCC. *Sens* and *Spec* calculations were explained in Subsection 2.1.2.3 (*See* Equations 2.7 & 2.8). Overall, each model's performance was measured by MCC (Table 4.2 (c)). Then, the values of each

measurement were calculated (*See* Equation 2.9). Finally, all calculations were converted into percentages (%).

Table 4.2: Multivariate Classification on FTIR and ^{1}H-NMR.

	(%)	a) *Sens*					b) *Spec*					c) MCC
		C	L	B	M	V	C	L	B	M	V	
FTIR												
LDA	PR	100	100	88	83	100	100	100	97	97	100	98
	pred	100	85	92	69	100	100	100	90	90	100	94
QDA	PR	100	100	92	88	100	100	100	97	98	100	98
	pred	100	100	92	92	100	100	100	97	96	100	98
MDA	PR	100	100	88	96	100	100	100	99	97	100	98
	pred	100	100	92	92	100	100	100	97	96	100	98
***C-1*-SVMDA**	PR	100	96	0	50	100	100	100	83	99	100	78
	pred	100	82	0	52	100	100	100	76	73	100	48
***C-3*-SVMDA**	PR	100	100	0	50	100	100	100	83	100	100	79
	pred	100	91	0	52	100	100	100	74	100	100	67
***υ -0.5*-SVMDA**	PR	100	100	100	100	100	100	100	100	100	100	100
	pred	100	91	100	100	100	100	100	97	100	100	97
^{1}H-NMR												
LDA	PR	94	100	100	100	100	100	99	100	100	100	98
	pred	78	80	100	100	100	96	93	100	100	100	89
QDA	PR	100	100	100	100	100	100	100	100	100	100	100
	pred	100	100	100	100	100	100	100	100	100	100	100
MDA	PR	100	100	100	100	100	100	100	100	100	100	100
	pred	100	100	100	100	100	100	100	100	100	100	100
***C-1*-SVMDA**	PR	100	85	100	100	100	97	100	100	100	100	97
	pred	100	40	100	100	89	88	100	100	96	100	85
***C-3*-SVMDA**	PR	100	100	100	100	100	100	100	100	100	100	100
	pred	100	100	100	100	100	100	100	100	100	100	100
***υ -0.5*-SVMDA**	PR	100	100	100	100	100	100	100	100	100	100	100
	pred	100	100	100	100	100	100	100	100	100	100	100

Sens = Sensitivity; *Spec* = Specificity; MCC = Matthews Correlation Coefficient; PR = pattern recognition; *pred* = prediction; C = chicken fats, L = lard, B = Beef fats, M = mutton fats, & V = plant fats.

Other common classification methods, linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), Mahalanobis discriminant analysis (MDA), and support vector machine (SVM), were reported to have

been used to analyze the FTIR and ^{1}H-NMR data by selected intervals obtained from PCA. According to Vaclavik et al., (2011), the percentage (%) of those successfully classified in the training set is presented as recognition ability or pattern recognition (PR). Meanwhile, the percentages of successfully classified in training developed models' test set are presented as prediction (*pred*).

In this study, three linear SVMDA were evaluated using parameter *C*. First, the least values are 1, *C-1*-SVMDA, and bigger to value 3, *C-3*-SVMDA. Second, parameter *Nu* (*υ*) and the smallest value are 0.5; *υ-0.5*-SVMDA. These parameters were selected for this FTIR spectra study because SVM is a complex algorithm.

4.2.5.1 Class Sensitivity

As can be seen in Table 4.2 (a), chicken, lard & plant fats showed excellent sensitivity in MVC-FTIR (≥80%) of the LDA, MDA, and QDA. Mutton and beef show inconsistent results in PR and *pred* FTIR-LDA of beef fats (88 %, 92 %) and mutton fats (83%, 69%). FTIR-QDA mutton fats (83%, 69%) and FTIR-MDA of beef fats (0%) are worst in class sensitivity. Chicken, lard & plant fats showed excellent sensitivity in *υ-0.5*-SVMDA classification (≥ 80%). On the contrary, beef shows the worst PR, and the *pred* is beef fats (0). Mutton fats show in *C-1* & *C-3*-SVMDA outputs at boundaries (≥ 50 %) as output values (50 %, 52 %).

MVC-^{1}H-NMR of all fats showed excellent sensitivity in PR and *pred* (≥ 80%). ^{1}H-NMR-MDA & ^{1}H-NMR-QDA showed the best with full achievement in classifying all fats (100%). Contrarily, ^{1}H-NMR-LDA lard

was not entirely correctly classified. All fats showed excellent sensitivity in PR of the ^{1}H-NMR-SVMDA (≥ 80%). The *C-3* & *υ-0.5*-SVMDA showed the best full achievement in classifying all fats (100%). Lard exhibited the worst prediction (40%) of ^{1}H-NMR-*C-1*-SVMDA. Beef and mutton were correctly classified as values (100%) for calibration and prediction in ^{1}H-NMR-SVMDA models, in contrast to the results of FTIR-SVMDA models.

4.2.5.2 Class Specificity

As can be seen in Table 4.2 (b), LDA, MDA, and QDA showed excellent specificity PR (≥ 80%) of the FTIR data. Chicken, lard, and plant fats showed excellent PR outputs (100%) on sensitivity. However, mutton and beef showed inconsistent results for PR and *pred C-1* SVMDA between (≤ 80% & ≥ 50%).

For the ^{1}H-NMR, MDA & QDA showed the best full achievement in classifying all types of fats (100 %). It was found that all fats showed sensitivity to LDA (≥ 80 %). *C-3*-SVMDA & *υ-0.5*-SVMDA have shown the best full achievement in PR all fats (100 %). Beef and mutton are correctly classified as values (100 %) in ^{1}H-NMR-SVMDA models, in contrast to the results of FTIR-SVMDA. Lard and chicken fats were found to be overlapped, thus showing that the MVC-^{1}H-NMR lard and chicken fats were less specific than the selected fats beef, mutton, and plant.

PR could refer to calibration, and *pred* could refer to validation to simplify the results. While prior studies have examined FTIR data by LDA analysis (Jamwal et al., 2020a; Han et al., 2020; Mousa et al., 2022), it may be preferable to contemplate the other techniques using QDA and MDA on

FTIR analysis. In addition, pattern recognitions involved in multiclass of oils and fats in literature are limited. Furthermore, LDA analysis on FTIR data for oil and fat research has been most used in the literature compared to the QDA and MDA. However, in this study, FTIR-LDA was found to be less sensitive than FTIR-MDA and FTIR-QDA to predict the lard, mutton, and beef according to their classes 100%. Nevertheless, the findings were similar to other studies, such as by Akin et al., (2019), who demonstrated ATR–MIR data on LDA and QDA models gave accuracy values of 94.44 % and 100 %, respectively, which were constructed with six edibles oils grape seed, sunflower, hazelnut, soybean, olive, and cottonseed oils.

The selection parameter of SVMDA type *C* was less sensitive than *υ-0.5*-SVMDA on FTIR data. However, the sensitivity and specificity results show that the LDA, QDA, MDA, and *υ-0.5*-SVMDA were also capable methods which could be better than *C-1*-SVMDA & *C-3*-SVMDA for the discrimination beef and mutton fats samples after the heating-process. In another study, the radial basis function (RBF) was used as the kernel function of FTIR-SVM to classify lard from beef, chicken, and lamb fats in meat mixtures where overall prediction accuracies for pure samples were demonstrated. Pure samples adulterated samples at 81.25 % and 72.2 %, respectively, were successfully predicted (Siddiqui et al., 2021).

In MVC-^{1}H-NMR, lard and chicken fats' sensitivity and specificity classes were less accurate. The worst validation of lard was by ^{1}H-NMR-*C-1*-SVMDA at only 40 %. All other fats showed excellent results, especially beef and mutton fat which gave 100% of the class sensitivity. The performance of ^{1}H-NMR-MDA, QDA, *C-3*-SVMDA, and *υ-0.5*-SVMDA showed perfect,

i.e., 100 % on all fats' classes sensitivity and specificity. By comparison, Wang et al., (2020) identified pure olive oils from blended products at 84.92 % classification accuracy when the adulteration ratio was above 10 %, using C-support vector classification (C-SVC) with RBF-SVM combined with low-field NMR data.

4.2.5.3 Classification Measures of the MCC.

As can be seen in Table 4.2 (c), all FTIR-MCC have shown excellent classification (≥ 80 %) except *C-1* & *C-3*-SVMDA with only satisfactory results but was found unstable in terms of *pred* for beef and mutton fats. The FTIR-MCC results show that *υ-0.5*-SVMDA was more potent than the other fat classification methods. Otherwise, the *C-1*-SVMDA and C-3-SVMDA showed less calibration and prediction ability (78 %, 48 % & 79 %, 67 %), as shown in MCC. The MDA and QDA were found more reliable because they gave accurate results in both calibrations, with a prediction of 98 % for both abilities. The selection performances of the FTIR-SVMDA can be summarised as *C-1*-SVM > *C-3*-SVM > *υ-0.5*-SVMDA in increasing order.

It was found that ^{1}H-NMR-MCC outperformed the FTIR-MCC and showed full achievement in classifying all fats (100 %) except in LDA and *C-1*-SVMDA. The ^{1}H-NMR-MCC of the training results showed that MDA, QDA, *C-3*-SVMDA, and υ-0.5-SVMDA were more powerful than LDA and *C-1*-SVMDA classification methods for fats. Brereton (2021), suggested that MCC has to be a better evaluation measure for multiclass classification than others, such as the proportion of samples predicted (PRE), correctly classified rate (CCR), and the F1-score.

4.2.6 Multivariate Regression (MVR)

The MVR was performed to investigate the relationship between different classes of lard and non-lard. Scores plot of each model's multivariate regression observed the lard separation plots according to Factor 1 variances by the 2D graph. Two values were noted inside the brackets (*X*, *Y*), representing variances by each Factor or PC. The "*X*" denoted spectral variance, and "*Y*" is class variance which is (0/1) that was captured by Factor or PCs. The rules that are opposed in 2D multivariate regression followed the Factors, or PCs, which contribute a little to class (*Y*) if denoted by the percentage of class variance by < 50% (CAMO, 2015).

4.2.6.1 FTIR

The MVR of FTIR models are presented in Figure 4.15 (a) FTIR-OSC-PLSR, (b) FTIR-PCR, and (c) FTIR-PLSR in the score plots graph.

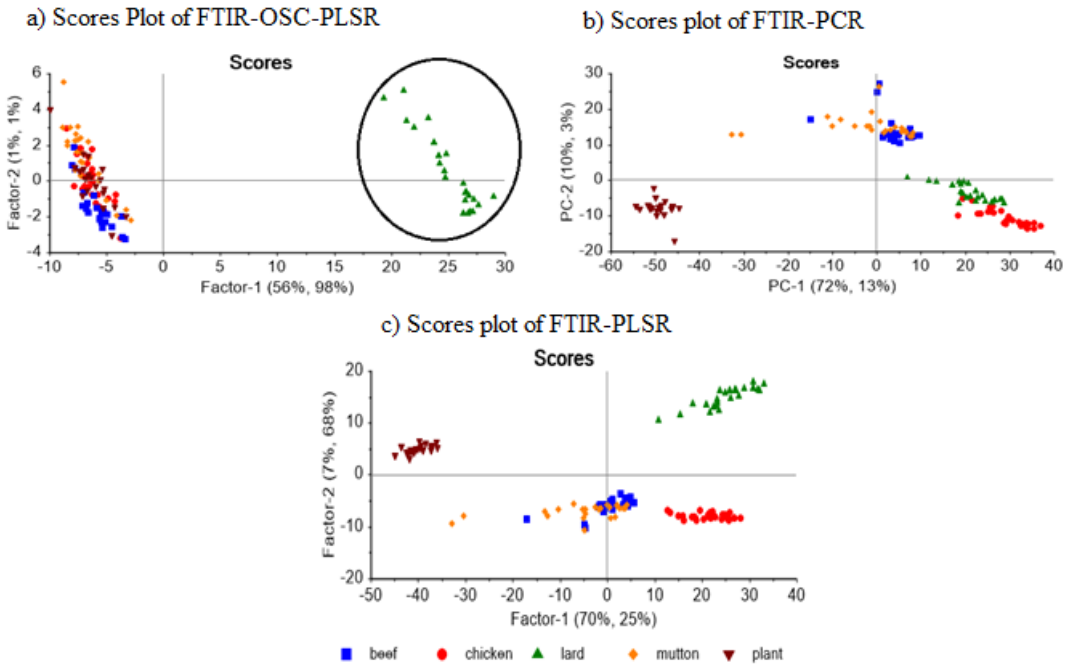


Figure 4.15: Multivariate Regression on FTIR data.

In the model of FTIR-OSC-PLSR, the fitted first two Factor models considered extracted Factor-1 as a predictive component and Factor-2 as an orthogonal component. As can be seen in Figure 4.15 (a), all the score plots of lard, coded as number 1, line up at the Factor-1 axis anti-correlated with the other fats and contributed 56% of spectral variances. The total class variance is 98%, which showed that the separation of lard from other fats was excellent using FTIR-OSC-PLSR.

Figure 4.15 (b) shows the model of FTIR-PCR could not to separate lard and chicken clusters. Therefore, Factor-1 was only given 13% separation between the class of lard and non-lard. In the model of FTIR-PLSR the first Factors were considered to be able to differentiate lard, and selected fats are shown in Figure 4.15 (c). All the scores plots of the lard were clustered in positive variances mixed with chicken fats; some beef and mutton fats at Factor-1. The percentage class separation between lard and non-lard was about 25%.

Prior works have documented that selected fats can differentiate lard using FTIR, PCA, and PLS techniques (Rohman & Che Man, 2010; Rohman & Che Man 2011; Al-Kahtani et al., 2017). Although much research has been devoted to study lard in the percentages of mixtures and binary concentration aspect, more attention is needed to determine the degradation of lard after the heating process. From this study, it is possible to attempt more than one algorithm in multivariate regression techniques to classify lard from other fats after the heating- process.

In this study, the binary dummies 0/1 allowed visualisation scores plot distribution. In addition, this study extended the last multivariate calibration

of the FTIR spectra after the heating- process (Salleh & Hassan, 2019) that has found additional information by exploiting different chemometrics techniques such as PCR and SVMR rather than PLSR alone in the heating-process of lard.

4.2.6.2 ^{1}H-NMR

The MVR of ^{1}H-NMR models are shown in Figure 4.16 (a) ^{1}H-NMR-OSC-PLSR, (b) ^{1}H-NMR-PCR, and (c) ^{1}H-NMR-PLSR.

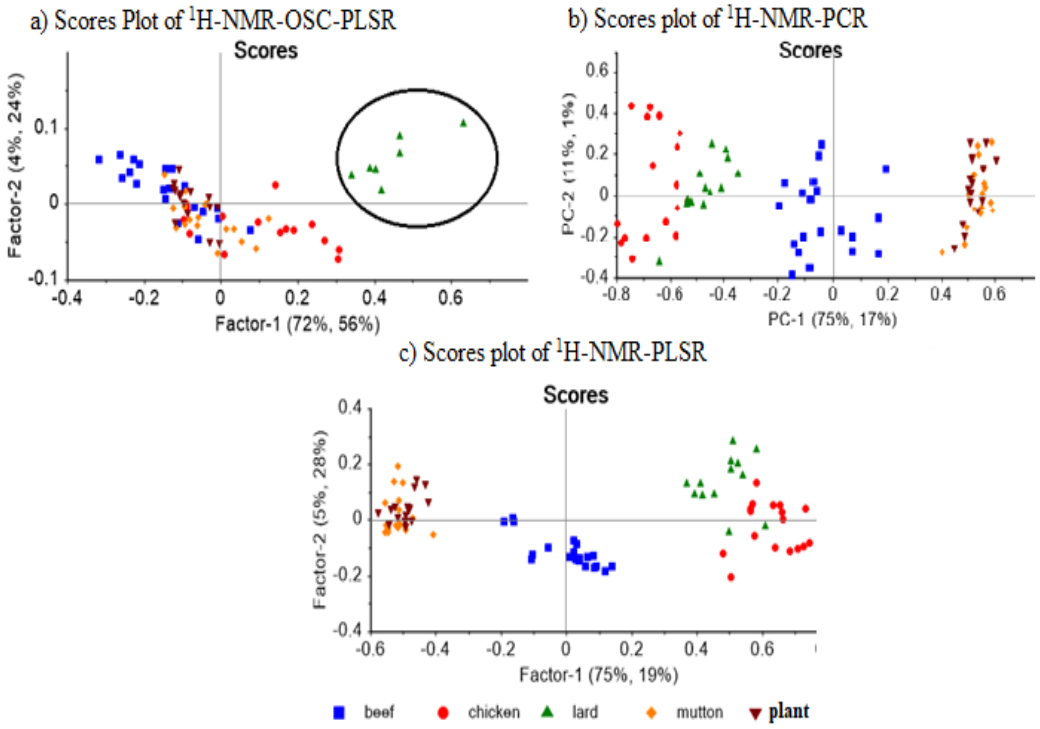


Figure 4.16: Multivariate Regression on ^{1}H-NMR.

In Figure 4.16 (a) ^{1}H-NMR-OSC-PLSR class separation, only 56% can be considered as borderline values (< 50%). In addition, lard and chicken fats were found to be slightly correlated in Factor-1. In Figure 4.16 (b), the model of ^{1}H-NMR-PCR, the 1st and 2nd components could not separate lard because

the cluster lard and chicken are correlated with each other at negative PC-1 by the class variances at 17%.

In the model of ^{1}H-NMR-PLSR (Figure 4.16 (c)), the first two Factors explained little about lard because scores plot of the clusters lard and chicken were also highly correlated at positive PC-1 but anti-correlated with the other fats in the class variances is 19%.

4.2.6.3 Comparison between FTIR and ^{1}H-NMR Multivariate Regression

In this study, only the first component (Factor-1 & PC-1) are accounted for in the statistics. Therefore, a comparison of the multivariate regression of FTIR and ^{1}H-NMR is shown in Table 4.3.

Table 4.3: Statistics for Multivariate Regression.

	R^2	*adj.* R^2	RMSEC	RMSEV	MSEP
FTIR					
OSC-PLSR	0.985	0.984	0.049	0.051	0.056
PCR	0.126	0.117	0.374	0.380	0.442
PLSR	0.249	0.238	0.347	0.351	0.405
^{1}H-NMR					
OSC-PLSR	0.561	0.548	0.193	0.201	0.203
PCR	0.165	0.134	0.33	0.341	0.290
PLSR	0.194	0.162	0.325	0.337	0.114

In addition, each model prediction performance was evaluated using a test set given by MSEP. Overall, both FTIR and ^{1}H-NMR using OSC-PLSR perform better than PLSR and PCR. As expected, the FTIR-OSC-PLSR data was outperformed and more robust than the other algorithm with R^2 = 0.985, *adj*. R^2 = 0.984, RMSEC = 0.049, RMSEV = 0.05, and MSEP = 0.054. In comparison, the ^{1}H-NMR-OSC-PLSR showed satisfactory in predictive results.

The first component, the most variances principle, was less effective in discriminating lard against others selected fats, as shown in PCR and PLSR outputs. The principal works based on most variances, but PCR produced worst results than PLSR. The similar finding in comparative studies, PLSR outperformed PCR in the authentication of extra virgin olive oil (Rohman et al., 2017); adulteration analysis of virgin coconut oil with grape seed oil and soybean oil (Rohman et al., 2019). However, both studies by Rohman et al., used more than two components.

Although PLSR finds the direction of the first latent factor in the fat class variances, the outcomes were unsatisfactory ($R^2 < 0.5$). This strategy may be the reason some publications have taken accounts for more than one factor to suit the R^2. Noted that both PCR and PLSR have different principles than OSC-PLSR.

OSC-PLSR functions as systematic variations split into two parts, i.e., one is linearly related to the response, and the second is linearly uncorrelated to the response (orthogonal). Since the OSC-PLSR is forcing separation, the scores plot of lard is separated well in FTIR-OSC-PLSR data but not in ^{1}H-NMR-OSC-PLSR data. The explanation might be that the wavelet type of FTIR data works on OSC-PLSR but not on the sharp peaks of NMR. In contrast, the ^{1}H-NMR-OSC-PLSR in some metabolomic publications gave good separation because the researchers used selected individual chemical shifts that produced discreet data types.

The second explanation for the observation may be that fats were directly diluted in *d*-$CHCl_3$, which might not be effective in dispersing the chemical composition. Thus, additional isolation works, such as TLC before

solvent dilution, could help to disperse the fatty matrix. However, the isolation works are time-consuming and not practical for large samples.

4.2.7 Correlation Chemical Features of Lard after Heating-Process by OSC-PLSR

4.2.7.1 FTIR

The weighted regression coefficients revealed the importance of the *X*-variables. The *X*-variables with a large regression coefficient play an important role in the correlation of chemical features. The plots were extracted from the *X*-loadings related to the models' most contributed lard scores plots. The predictor with a higher coefficient defined the important variables are shown in Appendix 3. Conversely, predictors (*X*) with a small coefficient are negligible. The important chemical features of FTIR-OSC-PLSR (Appendix 3) are tabulated in Table 4.4.

From weighted regression coefficients (*B*), aldehydes and ketones in saturated and aromatic chains are demonstrated. These secondary oxidation products are exhibited more on variables after the heating-process to defer lard and selected fats. Identifying wavenumbers can be related to the lard requires prior knowledge about the oxidation of fats after the heating-process. This is the first study that led to an understanding of the FTIR spectra of lard and selected fats after the heating-process.

Table 4.4: Wavenumbers Contributed to the Lard.

Wavenumbers range (cm^{-1})	Vibrations		Groups	Chains
742.6-769.6	C-H	bending	NR	NR
827.5-839.0	C-H	bending	NR	NR
864.1-866.0	C-H	bending	NR	NR
1062.8-1085.9	C-O	stretching	ester	NR
1124.5-1136.1	C-O	stretching	ester	NR
1163.1-1193.9	C=O	stretching	aldehyde	aromatic
1163.1-1193.9	C-O	stretching	ester	NR
1217.1-1238.3	C-C-C	stretching	ketones	aromatic
1504.5-1510.3	NA	NA	NA	NA
1516.0-1523.8	NA	NA	NA	NA
1533.4-1545.0	NA	NA	NA	NA
1554.6-1575.8	NA	NA	NA	NA
1633.7-1649.1	C=C	stretching	NR	NR
1651.1-1701.2	C=O	stretching	ketones	aromatic
1735.9-1755.2	C=O	stretching	ester	NR
2854.6-2864.3	$CHCH_2$	asymmetric bending	NR	NR
2885.5-2912.5	$CHCH_2$	asymmetric bending	NR	NR
2922.1-2939.5	C-H	bending	NR	NR

NR = Not related, NA = Not Applicable in the literatures.

4.2.7.2 ^{1}H-NMR

OSC-PLSR-^{1}H-NMR enables the analysis to enhance interpretability between predictive and orthogonality. The values of explained variation of scores vectors were about 56%, and the algorithm's performances are on the boundary performances value. Although the chicken fats and lard scores plot are found to be somehow correlated in Factor-1 as indicated in previous Figure 4.16 (a), the cluster of chicken fats appeared slightly lower than lard, subsequently weakening the predictive scores of lard to relate chemical features. Thus, the alternative solution to derive lard-associated features is plotting the regression coefficient in Appendix 5 into a probability (p) graph or S-graph, are presented in Figure 4.17.

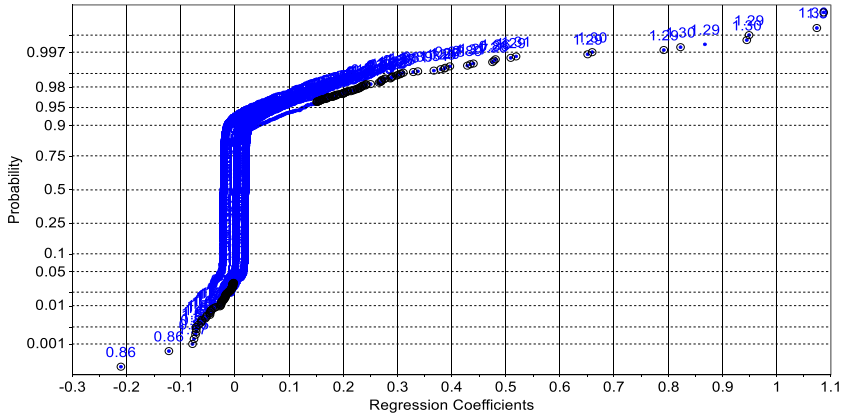


Figure 4.17: Variables Selection by *p*-Correlation.

The information of δ is extracted by plotting it into *S*-graph. Then, observing the *B* plot is a good way to identify potential important chemical shifts to differentiate lard and chicken fats. The marked signals on the *X*-loadings regression coefficient plot revealed a down-regulated lard profile based on Factor-1. The important and significant spectra contributing to the lard profile are shown in Table 4.5, derived from regression coefficient (*B*) marked in dark colour circles.

Table 4.5 summarises the chemical shifts (δ) from the regression coefficient (*B*) and correlation loadings plot significant at *p*-values < 0.05 generated from the *S*-graph. The loadings plots that influenced most in the separation of lard were located around δ 0.85-0.92 ppm (assigned to acyl groups except linolenyl), δ 1.26-1.36 (all acyl groups), δ 1.61-1.68 (all acyl groups), δ 1.99-2.06 (oleyl, linoleyl and linolenyl groups), δ 2.30-2.32 (all acyl groups) and δ 2.74-2.86 (linolenyl group). Therefore, it can be suggested that lard contains more linoleic and linolenic FAs. The elucidation of molecules based on the chemical shift in the literature reports is shown in Appendix 4.

Table 4.5: Summary of the δ by ^{1}H-NMR-OCR-PLS.

δ	Description
0.85-0.92	-CH_2-CH_2-CH_2-CH_3 (all acyl groups except linolenyl)
1.26-1.36	-(CH_2)n- (all acyl groups)
1.61-1.68	-OCO-CH_2-CH_2- (all acyl groups)
1.99-2.06	-CH_2-CH-CH-(oleyl, linoleyl & linolenyl groups)
2.30-2.32	-OCO-CH_2-(all acyl groups)
2.74-2.86	CH-CH_2-CH (linolenyl group)

The coefficient correlation graph is well known as the *S*-plot or *S*-graph in the SIMCA software package, but The Unscrambler® X also can plot the graph of the same function in the probability plot. Results show that δ 0.88-0.89, 1.26-1.31, 2.05-2.06 resonance derived from regression coefficient of lard clusters, interpreted as leading to proton signal of FAs such as palmitic (C16:0), stearic (C18:0) and oleic (C18:1). These FAs were known to be predominantly present in oil and fats. Thus, the use of ^{1}H-NMR is suggested to extend by ^{13}C-NMR analysis for further investigation to get more chemical information on lard after the heating-process.

In this study, multivariate regression FTIR and ^{1}H-NMR using OSC-PLSR, PCR, and PLSR have been found useful for evaluating lard and non-lard spectra. The multivariate regression analysis was compared to evaluate the best wavenumbers and δ responsible for the differentiation of lard from selected fats. It shows that FTIR-OSC-PLSR provides better differentiation of lard and selected fats than ^{1}H-NMR-OSC-PLSR according to the separation scores plot and statistics for multivariate regression. The algorithm of PCR and PLSR does not help separate lard and other fats by binary classification. The findings from the loadings plot of the FTIR-OSC-PLSR denoted the secondary oxidation products that can discriminate lard from selected fats, thus providing better predictive outcomes than ^{1}H-NMR-OSC-PLSR.

4.2.8 Profiling of ^{13}C-NMR Lard *vs* Chicken fats

The descriptive analysis PCA, multivariate classification and regression outputs on FTIR and ^{1}H-NMR demonstrated that lard and chicken fats have high similarity based on their scores proximity. Selected overlapped samples of lard and chicken fats from the scores plot of ^{1}H-NMR-PCA (*See* Figure 4.14) were subsequently analysed by ^{13}C-NMR. Their peak values were noted and integrated by MestReNova version 11.0 to compare the critical variances between lard and chicken fats spectra by peak assignments visualisation.

These experiments resulted from a total peak value of 15 spectra recorded in ppm units. Then the spectra were aligned to determine dissimilarities between lard and chicken fats subjected to a multilevel the heating-process. For all peaks in the ^{13}C-NMR spectra, the assignment of chemical structures for each sample utilising the chemical shift ranges is shown in Appendix 6.

The ^{13}C-NMR profiles are presented in Figure 4.18 reveal that each fat is owing by the same structures. Therefore, the overlay or stacking of ^{13}C-NMR spectra of lard and chicken fats were visually undifferentiable without the assistance of chemometrics techniques.

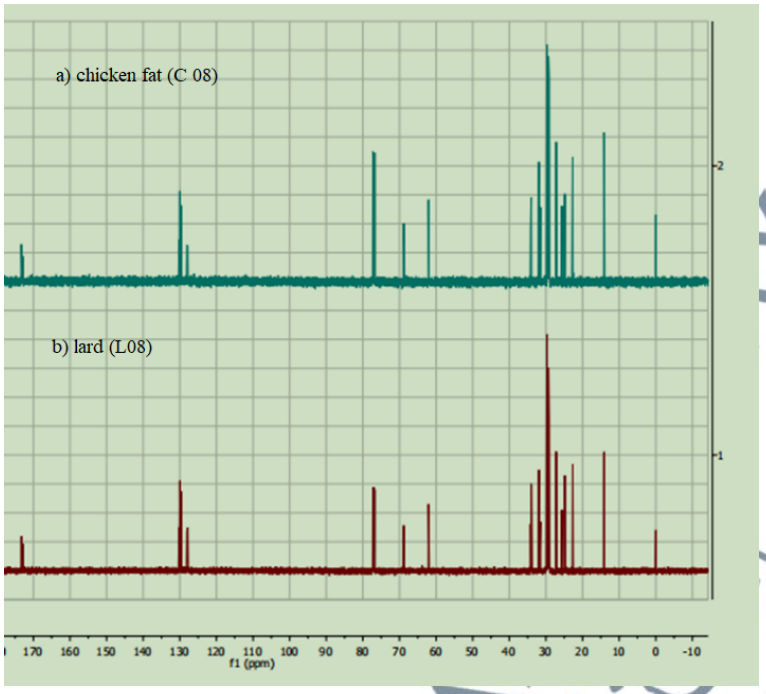


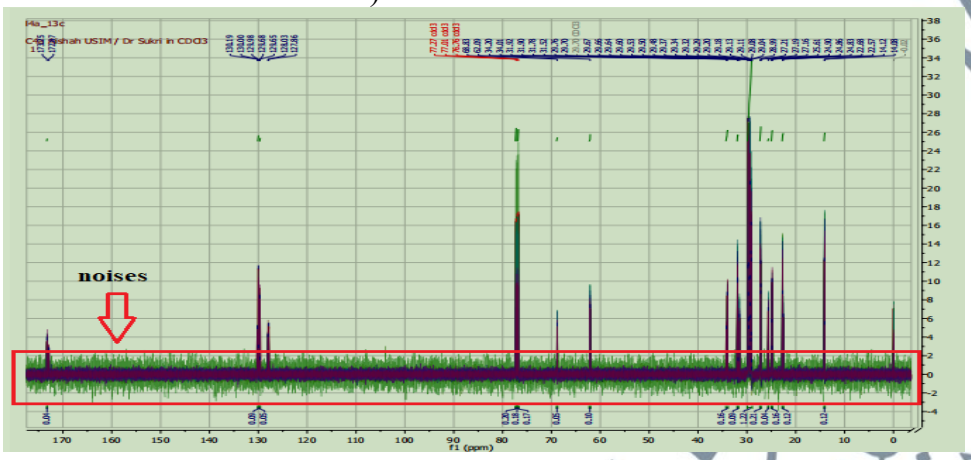
Figure 4.18: Stacked individual ^{13}C-NMR Spectra of Lard *vs* Chicken Fats.

Previous research has described that ^{1}H-NMR is less sensitive to identified lard and chicken fats (Fang et al., 2013, Fadzillah et al., 2015), and some studies on ^{1}H-NMR have suggested the use of the ^{13}C-NMR spectroscopy (Erich et al., 2015; Truzzi et al., 2021). However, due to the close similarity of the δ caused by overlapping ^{13}C-NMR in the fine spectra peak, thus, pre-processing of spectra has to be conducted to remove noise and additive baselines.

4.2.9 Data Pre-processing of ^{13}C-NMR Spectra

A total of 15 ^{13}C-NMR spectra were superimposed, as can be seen in Figure 4.19 (a) Before correction & (b) After correction.

a) Before correction



b) After correction

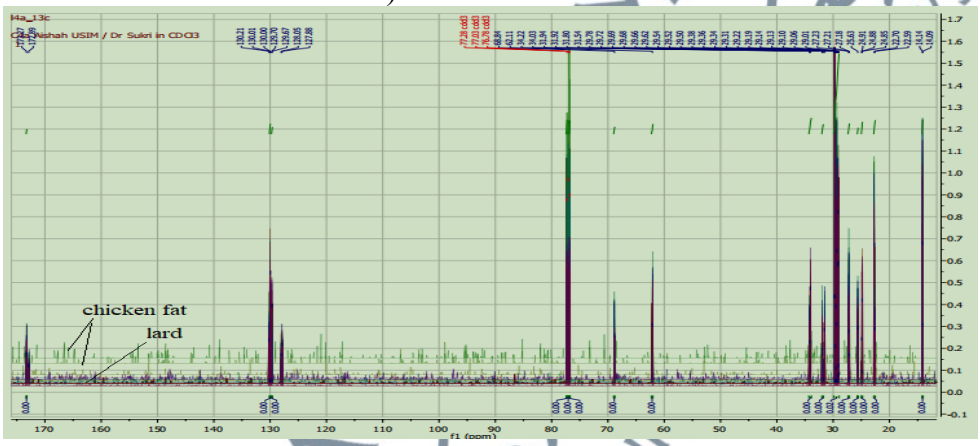


Figure 4.19: ^{13}C-NMR Spectra of the Lard & Chicken Fats.

\

The raw data contained visible noises from the peaks of fats that were aggregated together and formed continuously along the baseline, especially some of the chicken samples (green). The noise happened because a longer acquisition time could lead to more intense noise and low s/n (signal-to-noise ratio), as evident in Figure 4.19 (a).

The denoises were applied by Whittaker Smoother correction, and the results are shown in Figure 4.19 (b). The two samples of chicken fats were found uplifting from baselines after denoises correction and required further

pre-processing. Thus, the data were transferred into ASCII files for further pre-processing by Unscrambler® X software. Figure 4.20 demonstrates the signal of ^{13}C-NMR after transferring to the Unscrambler® X. After subjecting to baseline pre-processing, the line plots resembled the origin spectra.

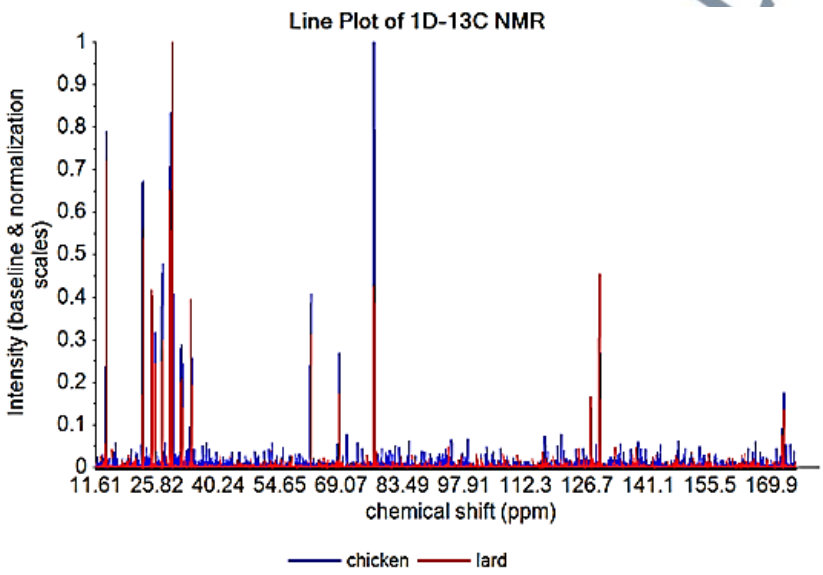


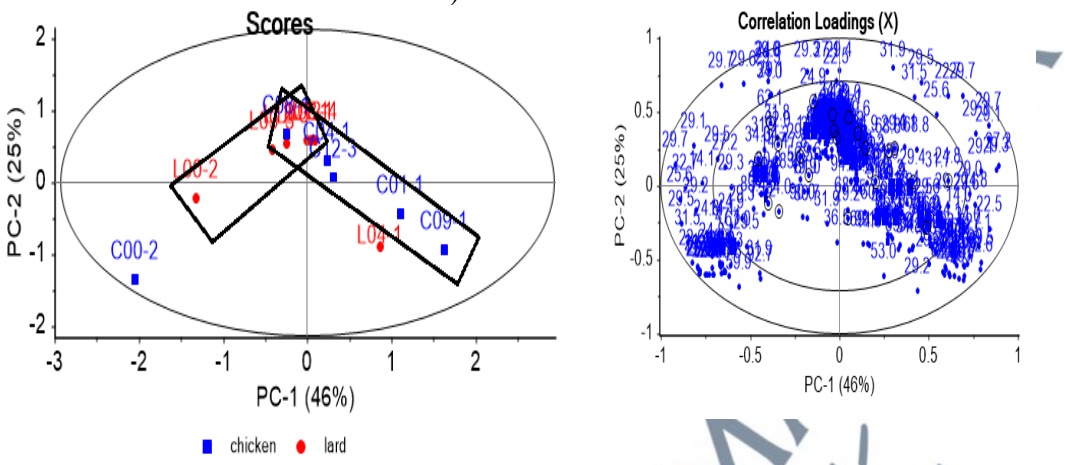
Figure 4.20: ^{13}C-NMR Spectra After Baseline Pre-Processing.

4.2.10 PCA of ^{13}C-NMR Lard *vs* Chicken fats

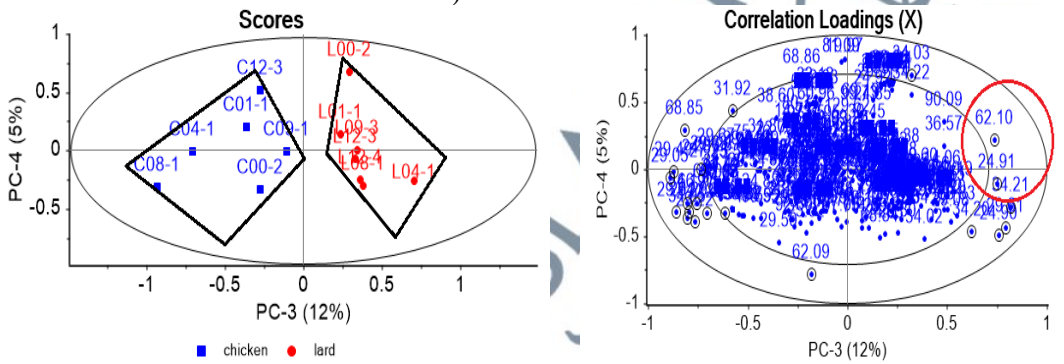
In the study, a total of 15 overlapped lard and chicken fats samples in the ^{1}H-NMR scores plot results were re-analysed using ^{13}C-NMR to be applied PCA. Eight lard samples and seven chicken samples total up to 15 overlapped samples. The spectra of the δ from 11-74 ppm and 121-170 ppm were selected and subject to PCA. The ranges of δ 0-10 ppm were excluded because of the internal standard (*d*-TMS) and 75–80 ppm region of *d*-$CHCl_3$.

In the first model, as can be seen in Figure 4.21 (a), the scores plot of PC-1 *vs* PC-2 showed no difference between lard and chicken fats. The first two PCs are explained about 70% of total variances.

a) PC-1 *vs* PC-2



b) PC-3 *vs* PC-4

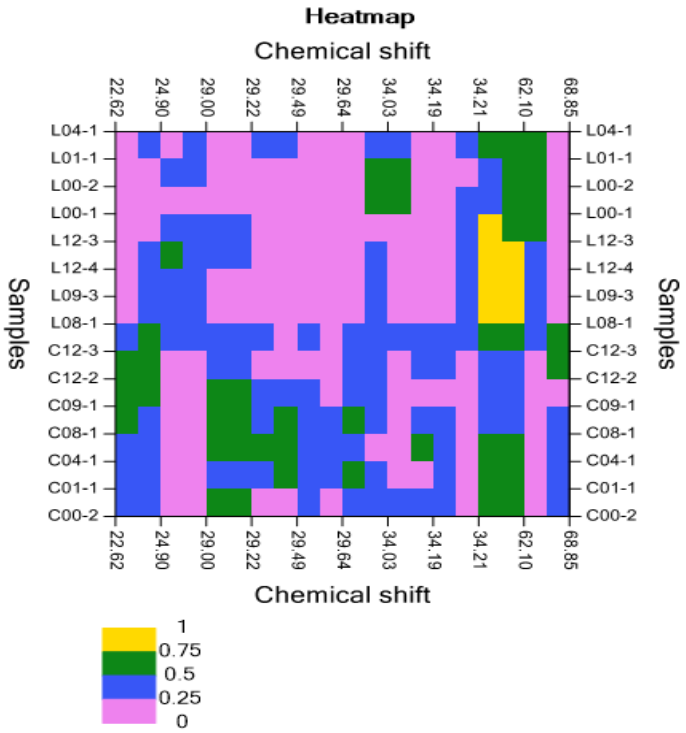


Fats : **C** = chicken fats, **L** = lard, Temperatures: **120 °C**= 01, 02, 03 & 04; **180 °C**= 05, 06, 07 & 08; **240° C**= 09. 10, 11 & 12 Duration of heating: **0.5 hr** = 01,05 & 09; **1 hr** = 02, 06 &10; **2 hrs** = 03, 07 & 11; **3 hrs** = 04, 08 & 12

Figure 4.21: ^{13}C-NMR-PCA of the Lard *vs* Chicken Fats.

The chicken fats and lard overlapped each other. The discrimination between lard and chicken fats is described in Figure 4.21 (b) at total variances of 17% of PC-3 *vs* PC-4. Scores plots are distinguished between two fats at positive values of PC-3 contributed by C-2 denoted by δ 34.21 and 62.10, marked by a red circle in correlation loadings, and the structures of elucidated TAG molecules are shown in Appendix 7. These resonances can be associated with the correlation loadings in Figure 4.21 (b) are challenging to visualise. Therefore, heatmap is a complementary approach to highlight the

contrast intensities of the δ between chicken fats and lard as can be seen in Figure 4.22.



Fats: **C** = chicken fats, **L** = lard, Temperatures: **120 °C**= 01, 02, 03 & 04; **180 °C**= 05, 06, 07 & 08; **240° C**= 09. 10, 11 & 12. Unheated = 00. Duration of heating: **0.5 hr** = 01,05 & 09; **1 hr** = 02, 06 &10; **2 hrs** = 03, 07 & 11; **3 hrs** = 04, 08 & 12.

Figure 4.22: Heatmap of ^{13}C-NMR-PCA.

The dark green and yellow colour is >50% input representing the significant compound depicted from *X*-correlation loading plots. The intensity was scaled from 0 to 1, denoted by proportions 0-100% by colour; 0-24.9 % (pink), 25.0-49.9 % (blue), 50.0-74.9 % (dark green), and 75.0-100 % (yellow).

The heatmap showed that PC-3 contributed by δ 34.21 and 62.10 to different lard and chicken fats. The heatmap has the advantage of generating an obvious additional identification of the carbon information of δ. The two groups can separate the two groups of lard: lower sum-heat (L04-1 = 120 °C/

3 hrs & L01-1 = 120 °C/ 0.5 hr), unheated (L00-2 & L00-1), and high sum-heat (L12-3 = 240 °C/ 3 hrs, L12-4 = 240 °C/ 3 hrs, L09-3 = 240 °C/ 0.5 hrs & L08-1 = 180 °C/ 3 hrs). The distinctions between these groups were contributed by δ 24.87- 24-90 ppm at a higher heat sum.

The findings show that the chemical shift, δ 62.10, relates to the lard to discriminate from other edible fats. The PCA combined ^{13}C-NMR lard and chicken fats have found δ ascribed to the TAG isomer, in line with the latest research that identified the regioisomer FA that contributes most to the lard characterization by GC-MS (Hanafy et al., 2021). Another related study successfully used ^{13}C-NMR spectroscopy in quantifying the TAG isomer from porcine and bovine sources (Kildahl et al., 2020).

The examples of TAG isomers are shown in Figure 4.23 (Beppu et al., 2017). The *sn*-1 position of TAG's isomer was separated using Thin Layer Chromatography (TLC) plate before being determined by GC. This separation process is very tedious and requires chemicals and a skilled person. However, this technique can be suggested by the ^{13}C-NMR analysis, which uses less solvent. Furthermore, many studies established on TAG isomer elucidation on ^{1}H-NMR and ^{13}C-NMR can be used as references (DiPietro et al., 2020; Erich et al., 2015; Guillén & Ruiz, 2003; Gunstone, 1994; Lie et al., 1997; Truzzi et al., 2021; Vigli et al., 2003).

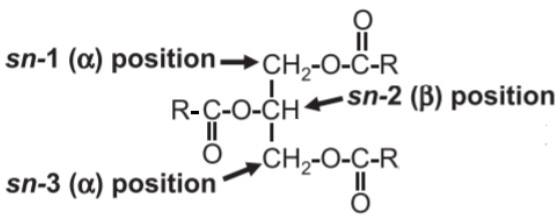


Figure 4.23: TAG Isomers.

4.3 Evaluation on Fatty Acids (FAs) of Lard and Selected Fats after Heating-Process using GC-FID and LC-MS/MS combined with Principal Component Analysis (PCA)

4.3.1 Gas Chromatography with Flame Ionization Detection (GC-FID)

4.3.1.1 Introduction

GC-FID was employed on the methylated fats after the heating process to evaluate the degradation of specific FAs. FAs were identified using FAMEs standard (Supelco ®) and automated integration using Chemstation software (Appendix 8). The typical chromatogram of lard is shown in Appendix 9. The result of 60 samples from lard, chicken, beef, mutton, and plant fats subjected to 12 heating protocols in rows and FAs that calculated area percentages of the peaks as per columns.

4.3.1.2 Fatty Acids (FAs) Profiling

Data matrix (*X*) of GC-FID, samples *vs* FAs; (60 x 21) were compressed into the box plot and the results are shown in Figure 4.24 to describe the range of the FAs distribution over the whole sample set.

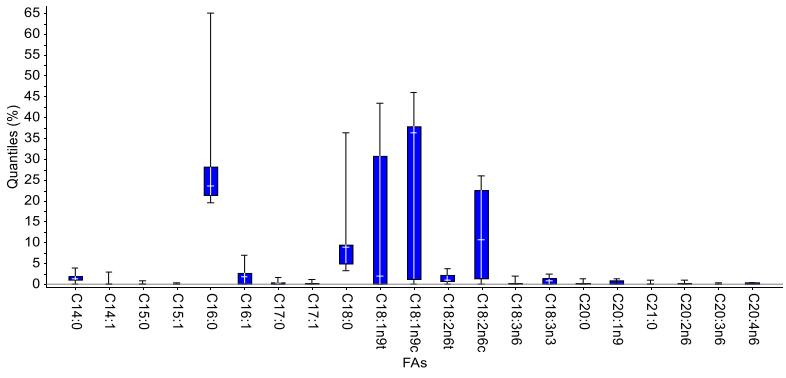


Figure 4.24: Box Plot of FAs

The box plot revealed nine detectable FAs (max >1.00) with different variations at ranges of max-min values; C14:0 (0-3.88), C16:0 (19.43-64.98), C16:1 (0-6.93), C18:0 (3.99-36.30), C18:1n9t (0-4.33), C18:1n9c (0-37.86), C18:2n6t (0-2.68), C18:2n6c (0-23.88) and C18:3n3 (0-3.44). In general, it can be divided into three classes of FAs according to the 25% percentile, Q1 (lower quartile), the median (middle), the 75% percentile, and Q3 (upper quartile). The first-class shows the Q1 ≥ 1.00 and Q3 ≥ 5.00 on FAs C16:0, C18:0, C18:1n9c, and C18:2n9c. The second class is Q1 ≥ 0.00 and Q3 ≥ 1.00 on FAs C14:0, C16:1, C18:1n9t, and C18:3n3 showed small values of Q1 but as important FAs. The rest were classed as average FAs.

The extreme maximum values showed at FA of C16:0 and C18:0 box plots represent higher variation ranges far from zero, indicating that all samples contain these FAs. The moderate distances between Q1-Q3 quantiles of C16:0 and C18:0 showed a small range of variation than the other FAs. The double bonds of the C18 hydrocarbon chain, which is MUFA (single of the double chain bonds; i.e., C18:1n9c & C18:1n9t) and PUFA (≥ 2 of the double chain bonds, i.e., C18:2n6c), show the largest distances between Q1 and Q3 than the other FAs. On the other hand, the average FAs indicated small values of variation.

This study aimed to evaluate FAs degradation of various edible fats after the heating-process. The evaluation of heating-process parameters involved the type of fats and heat parameters in determining lard and other fats using PCA, especially on GC-FID data found limited in the research.

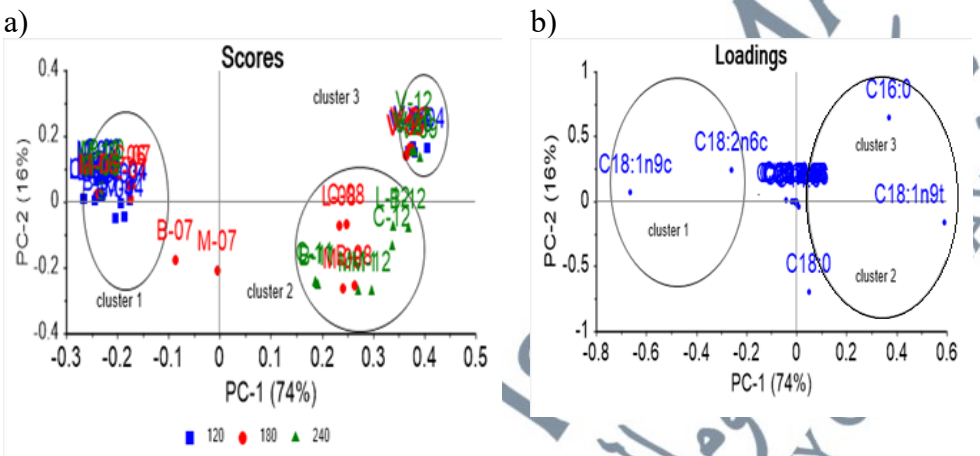
FAs compositions were compressed using a box plot based on the raw data, indicating the percentile FAs with the most variance are C16:0, C18:0, C18:1n9c, and C18:2n6c. In the studies, it has been found that raw data of lard and animal fats of the heated fats have carbon chains C16:0, C18:1, and C18:2 that present most predominantly. Such FAs have been demonstrated to be major components in lard by GC-FID (Azir et al., 2017; Nurjuliana et al., 2011) and GC-MS (Naquiah, et al., 2013). FA such as, C18:3n6 have been found to play an important role in profiling lard (Azizan et al., 2021).

Previous research found that lard could be differentiated from cocoa butter, having C16:1 (Azir et al., 2017; Kurniawati et al., 2014), and the most abundant compound was C18:2 in lard as compared to other animals fats (Nurjuliana et al., 2011). However, the basic statistic description could not explain the distribution of FAs according to the samples. Thus multivariate analysis, PCA was employed to correlate samples and FAs distribution. It is suggested that the data to be transformed by area normalisation prior to PCA. Normalisation selection generally aims to organize the FAs by minimizing data redundancy of the small values such as C14:0-15:1 and C18:3n6-C20:4n6 FAs.

4.3.1.3 PCA of GC-FID

The data matrix (60 x 21) of the heated edible fats and FAs was subjected to PCA to reduce the number of FAs determined by PCs. As a result, five types of FAs of the raw data are shown to be prominent in Appendix 10.

Figure 4.25 (a) & (b), represent the two graphs sample scores and loadings plots in the first two principal components (PC-1 *vs* PC-2). Figure 4.25 (a) shows the scores plot of the sample clustering. Figure 4.25, (b) represents FAs; Palmitic (C16:0), Stearic (C18:0). However, the integration plots are among the highest (240 °C) temperatures with the lowest (120 °C) scores plot.



Fats: **C** = chicken fats, **L** = lard, **B** = beef fats, **M** = mutton fats, & **V** = plant fats
Temperatures: **120 °C**= 1, 2, 3 & 4; **180 °C**= 5, 6, 7 & 8; **240° C**= 9,10, 11 & 12. Duration of heating: **0.5 hr** = 1,5 & 9; **1 hr** = 2, 6 & 10; **2 hrs** = 3, 7 & 11; **3 hrs** = 4, 8 & 12.

Figure 4.25: GC-FID-PCA.

Three clusters could be identified from the scores plot in Figure 4.25 (a). The first cluster is located at negative PC-1 inversely to the second at positive PC-1. Two ambiguous samples, M7 and B7, belong to cluster 1. The third cluster consists of plant fats only. Clusters 2 and 3 were located at positive PC-1 but anti-correlated to cluster 1. It is apparent cluster 2 were samples from those only temperatures 180 °C & 240 °C dominated only by animal fats. Meanwhile, only plant fats closed together at all temperatures in cluster 3.

As can be seen from Figure 4.25 (b), the FAs are represented by loadings plots. Two clusters of FAs mapping to correlate the scores plot. The FAs of C18:1n9c (Oleic) and C18:2n6c (Linoleic) are located on the positive side of PC-2. Both FAs are in unsaturated fatty acids groups with *cis* isomer represented by cluster 1. The FAs of C18:0 (Stearic), C18:1n9t (Elaidic), and C16:0 (Palmitic) are located at positives PC-1 and can be represented by clusters 2 & 3. FAs in the middle of loading plots at the nearest to 0 or centre are considered not important FAs.

From the GC-FID-PCA, the score plot of samples, lard, and other selected fats could not be differentiated. The most predominant fats contributed to the cluster mixing with different fats except plant fats clusters. The reason for this might not be clear, but heated fats' methylation process could be attributed to it. Some studies have stated that heated fats produce secondary oxidation products, such as polymeric TAG, affecting the methylation process (Dobarganes et al., 2006). The conventional separation methods include TLC as prior methylation and GC analysis for frying or heated fats at high temperatures. However, the separation process is laborious and time-consuming when it involves many samples, such as this study.

Table 4.6 shows the most significant PCs generated from the FAs; their loadings plot values ($\geq \pm 0.5$) were highlighted in bold.

Table 4.6: *X*-correlation Loadings Plot of the GC-FID-PCA.

FAs	PC-1	PC-2
C16:0	**0.735**	**0.612**
C16:1	**-0.517**	0.026
C18:0	0.121	**-0.910**
C18:1n9t	**0.969**	-0.117
C18:1n9c	**-0.984**	0.070
C18:2n6c	**-0.692**	0.315

The negative PC-1 strongly correlates with C18:1n9c (-0.984), followed by C18:2n6c (-0.692). The negative values of C16:1 and C18:2n6t were at the boundaries. The anti-correlated of C18:1n9t (0.735) and C16:0 (0.735) at positive PC-1. Meanwhile, negative PC-2 most contribute to a lower sum of heat and have strongly correlated at C18:0 (0.910) anti-correlated positive PC-2 at C16:0 (0.612).

Samples of heated fats in PCA were clustered close to each other, and the details only can be identified by observing the scores plot table (Appendix 11). Interestingly, the trend of distribution of these scores plots are according to the cluster's sum of heat (sum-heat). The details of the sum-heat can be referred to the Chapter 3 in Table 3.4.

The clusters according to the sum-heat are illustrated in Figure 4.26. Cluster 1, starting from No. 1-36, shows all of the animal fats at the range of sum-heat values at 60-360, representing the lowest heat treatment level in PCA models.

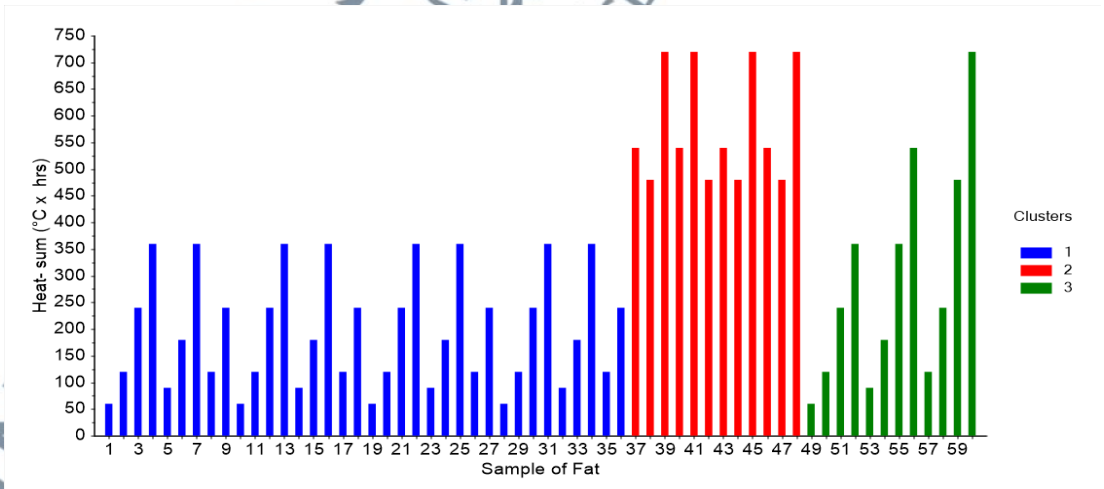


Figure 4.26: Bar Plot Sum of Heat Values.

Meanwhile, cluster 2, starting from No. 37-48, shows the whole species of animal fats at the range of sum-heat values at 480-720, representing the highest level of heat treatment. Cluster 3 are represented all samples from plant fats (V) in one group (No. 49-60).

According to the FAs pattern in PCA, there are clusters of animal fats from lower to higher levels of heat-sum. It is a reasonable outcome because each sample was treated at different heating durations and temperature interactions. Therefore, sum-heat values by multiplying the temperatures and duration of heating factors (°C) x (hrs) are the best parameter to represent this research's heating- process or heating treatment.

These findings suggested sum-heat levels were described by loadings plot of *cis*-isomer FAs C18:1n9c, C18:2n6c from lower-level sum-heat may have been degraded into *trans*-isomer; C18:1n9t and C18:2n6t for animal fats. The degradation process of *cis* and the increase of *trans* geometric isomer *trans* after prolonged and high-fat temperatures were consistent with the study by Bhardwaj et al., (2016). In this author's study, six edible fats were subjected to heating at 180 °C and 220 °C, which demonstrated *trans* isomer of the SFA increased, but the *cis* isomer of PUFA decreased. Meanwhile, the C16:0, which is mainly contributed by commercial plant fats and was suggested to have higher stability against heat, could have closer to each other. This finding is consistent with those other studies that the stability of shortening has correlated to refined oil (Yu et al., 2018).

4.3.2 Liquid Chromatography with Tandem Mass Spectrometry (LC-MS/MS)

4.3.2.1 Introduction

The study also evaluates another strategy for discriminating animal fats lipids after heating with LC-MS/MS followed by PCA. The lipid signals with an automated mass spectrometry identified and characterized the lipid classes by database searching of MS/MS fragments ions automated by LipidView Software™. The characteristic Glycerolipids (GL) and Glycerophospholipids (GPPL) of animal fats and plant fats could be observed according to the hydrocarbon chain, saturated fatty acid (SFA), monounsaturated fatty acid (MUFA), and polyunsaturated fatty acid (PUFA).

4.3.2.2 Profiling of Lard and Selected Fats by LC-MS/MS

The resulting lipid group profiles are illustrated in Figure 4.27, and each class of fats were represented according to the mass-to-charge ratio (*m/z*) *vs* abundance (intensities). It shows that lard (green diamond) mostly dominated between 250 *m/z* - 802 *m/z*. The distribution of the sample groups has more variances from 852 *m/z* to 934.5 *m/z*. The lipid groups of *m/z* according to GL and GPPL are shown in Appendix 12. Filtering unique compound techniques could be done by identifying lipid species. In total, 127 GL and 79 GPPL molecular species at different structures of the FAs chain were identified simultaneously by LipidView™ .

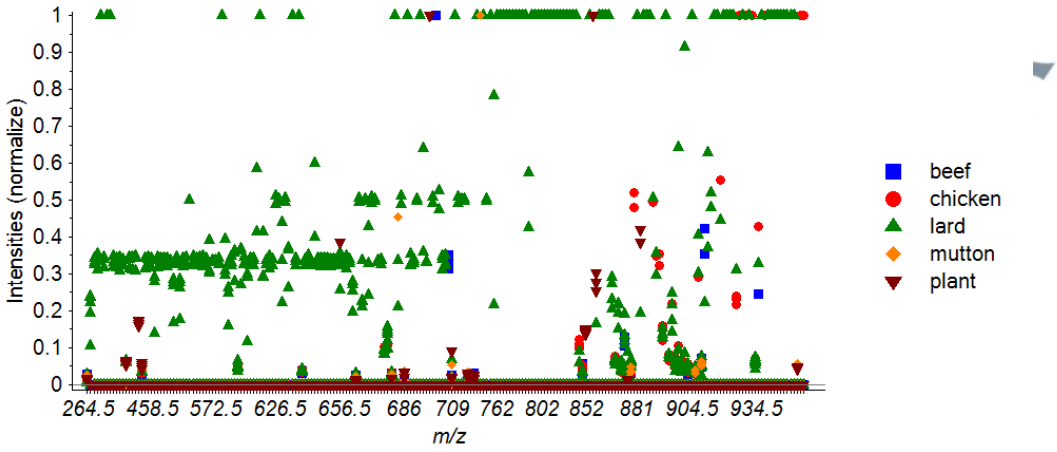
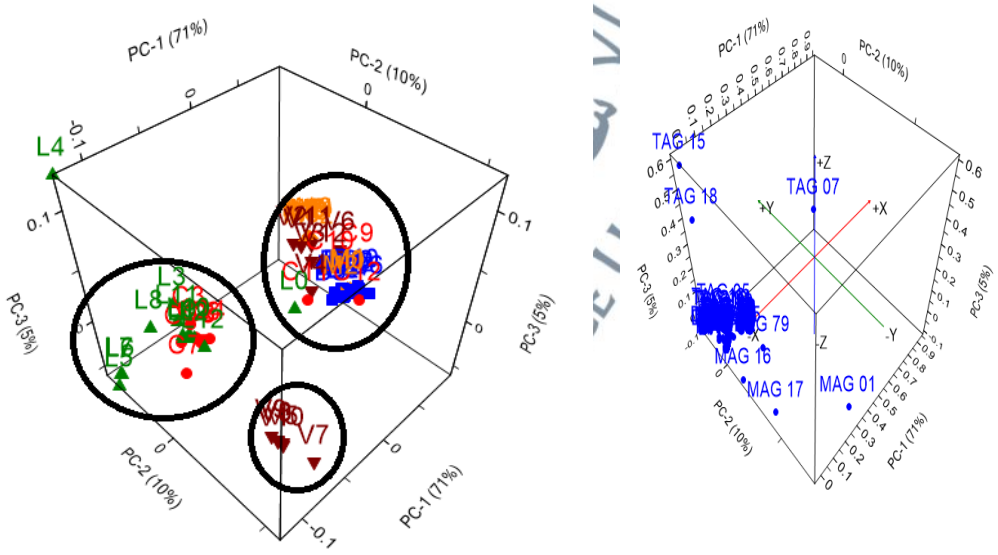


Figure 4.27: The Profiles of Fats by LC-MS/MS.

4.3.2.3 Glycerolipids of the PCA by LC-MS/MS (LC-MS/MS-GL-PCA)

The PCA scores plot in the 3D graph using the first three principal components for better visualisation is shown in Figure 4.28.

a) b)



Fats: **C** = chicken fats, **L** = lard, **B** = beef fats, **M** = mutton fats, & **V** = plant fats. Temperatures: **120 °C**= 1, 2, 3 & 4; **180 °C**= 5, 6, 7 & 8; **240° C**= 9,10, 11 & 12. Unheated = 0. Duration of heating: **0.5 hr** = 1,5 & 9; **1 hr** = 2, 6 & 10; **2 hrs** = 3, 7 & 11; **3 hrs** = 4, 8 & 12.

Figure 4.28: GL-PCA on LC-MS/MS Data of All Fats.

The first three PCs were observed the scores plot accounted for a total variance of 86% from PC-1 (71%), PC-2 (10%), and PC-3 (5%). As can be seen in Figure 4.28 (a), all fat samples can be grouped into three main clusters.

First, cluster along positive PC-1 explained most fat mixtures of beef and mutton in all parameters of the heat process: lard, chicken, and plant fats with lower heat at negative PC-1. Second, plant fat groups formed the cluster at negative PC-2. Third, one group of lard samples labelled L4, L5, L6 & L7 at negative values. The 3D graph could explain clusters better than the 2D graph since the data of LC-MS/MS contains some zero values at the particular column, giving an imbalance of the distribution of the scores.

Since the loadings plots in Figure 4.18 (b) are saturated with the individual lipids and impossible to observe with the naked eye, *X*-correlation was tabulated into Table 4.7 to identify lipids.

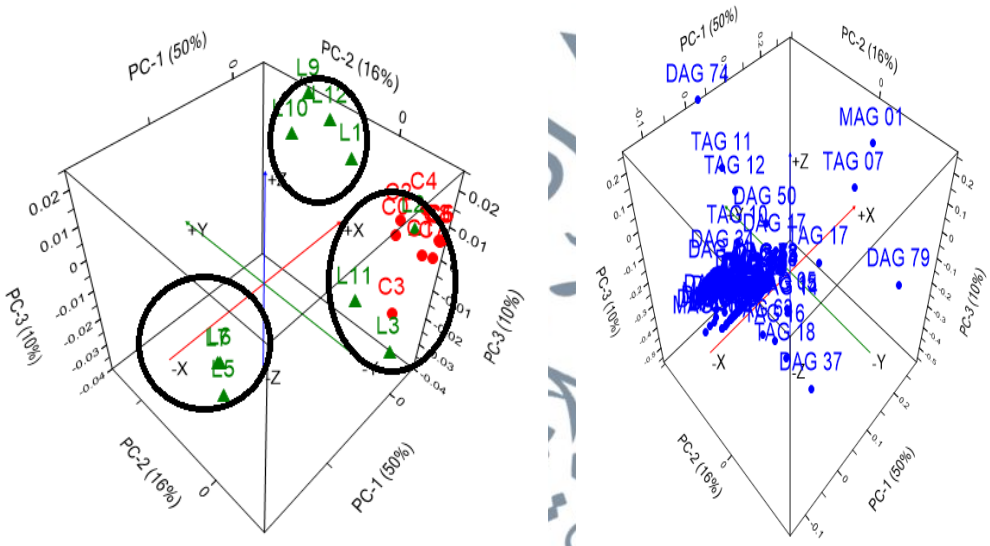
Table 4.7: *X*-correlation Loadings Plot of All Fats (LC-MS/MS-GL-PCA).

id	*m/z*	name	group	PC-1	PC-2	PC-3
MAG 01	264.5	MAG 10:0+NH4	SFA	**0.567**	**-0.786**	0.044
MAG 17	430.0	MAG 22:1+NH4	MUFA	0.030	**-0.781**	-0.162
TAG 07	708.5	TAG 41:2+NH4	PUFA	**0.996**	0.092	-0.005
TAG 15	900.5	TAG 54:3+NH4	PUFA	-0.220	0.264	**0.776**
TAG 18	902.5	TAG 55:11+NH4	PUFA	-0.306	0.307	**0.771**

The first clusters are beef, mutton, and plant fats, with the highest correlation are 0.99 on TAG 07 (*m/z*: 708.5, TAG 41.2 +NH4). The second clusters are plant fats contributed by MAG 01 (*m/z*: 264.5, MAG 10:0+NH4) at the value of -0.786; and MAG 17 (*m/z*: 430.0, MAG 22:1+NH4) at the value of -0.781 in correlation. The third clusters are mixed chicken fats

mapping by TAG 15 (*m/z*: 900.5, TAG 54:3+NH4) at the value of 0.776, and TAG 18 (*m/z*: 902, TAG 55:11+NH4) at the value of 0.771 in correlation. In order to investigate lard characters, samples of lard *vs* chicken clusters are enlarged to investigate further information on lard by excluding the dominated individual lard samples, L4 (120º C, 3 hrs), to expand the small intensities of the other lard samples in Figure 4.29. A 3D graph shows the scores plotted (a) represent a total variance of 76%, explained by PC-1 (50%), PC-2 (16%), and PC-3 (10%).

a) b)



Fats: **C** = chicken fats, **L** = lard, **B** = beef fats, **M** = mutton fats, & **V** = plant fats
Temperatures: **120 °C**= 1, 2, 3 & 4; **180 °C**= 5, 6, 7 & 8; **240° C**= 9,10, 11 & 12. Unheated = 0.
Duration of heating: **0.5 hr** = 1,5 & 9; **1 hr** = 2, 6 & 10; **2 hrs** = 3, 7 & 11; **3 hrs** = 4, 8 & 12.

Figure 4.29: GL-PCA on LC-MS/MS of Lard *vs* Chicken.

Three major clusters were shown in the GL- PCA. First clusters are lard heated at 180 °C (0.5, 1 & 2 hrs) and labelled as L5, L6 and L7. The second clusters are lard (L9, L10 & L12) at higher temperatures at 240 °C (0.5, 1 & 3). The third cluster formed from a combination of lard-chicken fats (L2, L3, L11 & C1 - C8).

From Table 4.8, which is extracted from loading plots of the PC-1 in Figure 4.29 (b), the first cluster of lard is contributed at DAG 10 (*m/z*: 586.5, DAG 32:0+NH4) at the highest value -0.993 and MAG 10 (*m/z*: 460.5, MAG 24:0+NH4) at the second highest value of -0.863. The typical spectrum for both compounds is shown in Appendix 13.

The second clusters of lard are contributed by DAG 74 (*m/z*: DAG 40:9+NH4) at a value of 0.858 and TAG 11 (*m/z*: TAG 52:3+NH4) at a value of 0.858 in correlation. Finally, the third cluster of chicken fats is contributed by TAG 07 (*m/z*: 722.0, TAG 41:2+NH4) at a value of 0.772 and lard are contributed by TAG 18 (*m/z*: 900.5 TAG 55:11+NH4) at a value of -0.903 in correlation.

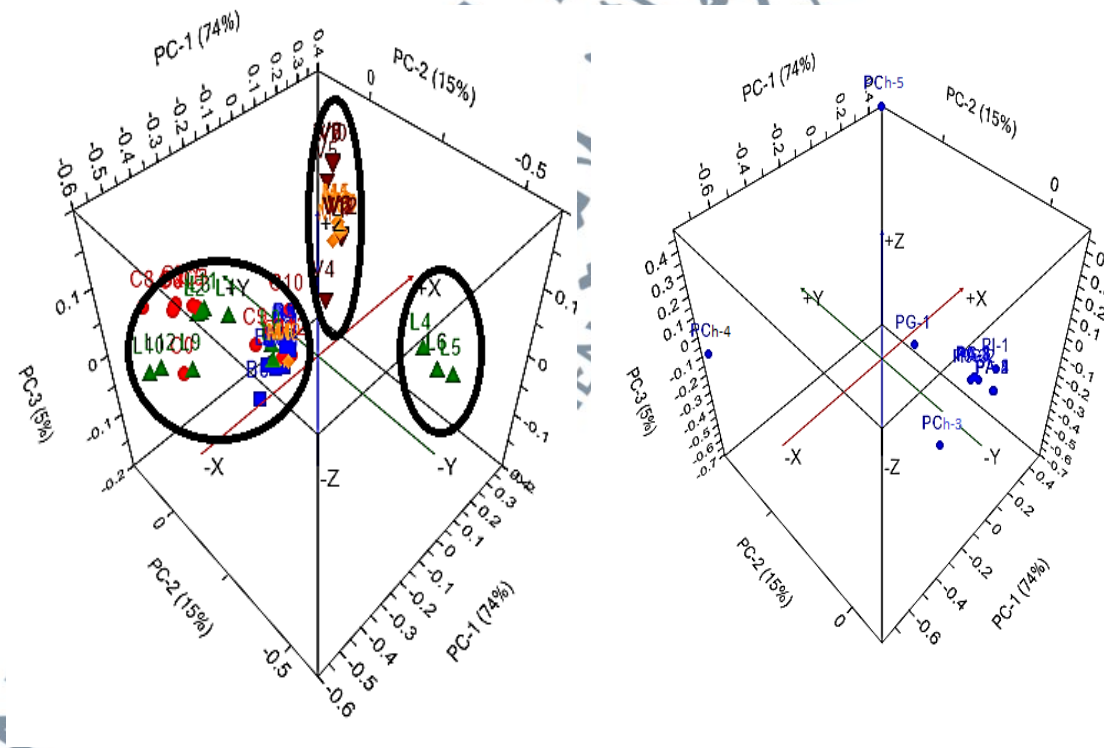
Table 4.8: *X*-correlation Loadings Plot (LC-MS/MS-GL-PCA).

id	*m/z*	Name	Group	PC-1	PC-2	PC-3
DAG 04	502.5	DAG 26:0+NH4	SFA	**-0.795**	0.358	-0.260
DAG 10	586.5	DAG 32:0+NH4	SFA	**-0.993**	-0.058	0.021
DAG 24	542.5	DAG 29:1+NH4	MUFA	**-0.795**	-0.058	0.002
DAG 37	620.5	DAG 33:5+NH4	PUFA	-0.025	-0.494	**-0.575**
DAG 60	662.5	DAG 38:4+NH4	PUFA	**-0.729**	0.406	-0.290
DAG 74	680.5	DAG 40:9+NH4	PUFA	0.090	**0.858**	0.352
DAG 79	694.5	DAG 42:2+NH4	PUFA	0.479	**-0.683**	-0.004
MAG 02	723.0	MAG 11:0+NH4	SFA	**-0.683**	0.432	-0.309
MAG 10	460.5	MAG 24:0+NH4	SFA	**-0.863**	0.297	-0.220
TAG 07	722.0	TAG 41:2+NH4	PUFA	**0.772**	0.004	0.031
TAG 11	874.5	TAG 52:3+NH4	PUFA	0.156	**0.858**	0.115
TAG 12	873.0	TAG 52:4+NH4	PUFA	0.139	**0.799**	0.125
TAG 16	898.5	TAG 54:5+NH4	PUFA	0.243	0.333	**-0.780**
TAG 18	900.5	TAG 55:11+NH4	PUFA	0.203	0.125	**-0.903**

4.3.2.4 Glycerophospholipids of the PCA by LC-MS/MS (LC-MS/MS-GPPL-PCA)

For the GPPL groups, the first 3 PCs were considered to evaluate GPPL-PCA. As can be seen in Figure 4.30 (a), the first clusters are mixed animal fats. chicken fats (C0 - C8) and lard (LI - L3) mixed together while (B1-B12), chicken fats (C9-C12) and lard (L0) at negative values of PC-1. The second clusters are mutton (M3 - M12) and plant fats (V1 - V12) at positive values of PC-1. The third clusters are lard L4, L5, L6, and L7 at negative values of PC-2. The loadings plot (Figure 4.30 (b)) shows that 3 clusters contribute lipids. The details loadings plot can be observed in Table 4.9.

a) b)



Fats: **C** = chicken fats, **L** = lard, **B** = beef fats, **M** = mutton fats, & **V** = plant fats. Temperatures: **120 °C**= 1, 2, 3 & 4; **180 °C**= 5, 6, 7 & 8; **240° C**= 9,10, 11 & 12. Unheated = 0. Duration of heating: **0.5 hr** = 1,5 & 9; **1 hr** = 2, 6 & 10; **2 hrs** = 3, 7 & 11; **3 hrs** = 4, 8 & 12.

Figure 4.30: GPPL-PCA on LC-MS/MS Data of All Fats.

Table 4.9: *X*-correlation Loadings Plot (LC-MS/MS-GPPL-PCA).

id	m/z	Name	Group	PC-1	PC-2	PC-3
PA-1	481.5	PA 20:0	SFA	0.054	**-0.605**	0.058
PA-2	495.5	PA 21:0	SFA	0.054	**-0.609**	0.060
PA-4	523.5	PA 23:0	SFA	0.054	**-0.605**	0.058
PA-5	551.5	PA 25:0	SFA	0.054	**-0.609**	0.060
PCh-3	722.0	PC 32:6	PUFA	**0.517**	0.241	**-0.678**
PCh-4	874.5	PC 42:0	SFA	**-0.945**	0.326	-0.008
PCh-5	907.0	PC 45:5	PUFA	**0.837**	0.496	0.212
PG-1	877.5	PG 43:0	SFA	**0.779**	0.425	-0.285
PI-1	951.5	PI 42:0	SFA	**-0.602**	-0.289	0.328

In Table 4.9, the first clusters of mixed fats are contributed by PCh-04 (*m/z*: 874.5; PCh 42:0) at the value of -0.945 in correlation. The second clusters are mutton & plant fats clusters contributed by PCh-05 (*m/z*: 874.5, PCh 42:0) at values of 0.837 in correlation. Third clusters are Lard (L4 - L7) contributed by PA - 1 (*m/z*: 481.5, PA 20:0) at a value of -0.605, PA - 2 (*m/z*: 495.5, PA 21:0) at a value of - 0.609, PA - 4 (*m/z*: 523.5, PA 23:0) at a value of -0.605, PA - 5 (*m/z*: 551.5, PA 25:0) at a value of - 0.609. These lard samples are far from others but contributed by phosphatidic acid lipid classes and almost average correlation values. From the hydrocarbon chain view, lard differed from the others and contributed by SFA groups in GL and GPPL lipid classes.

In general, lipids are a large and very complex group. Lipids exhibit a high degree of structural diversity and complexity, with over 37,000 distinct structures now maintained in LIPID MAPS®, the world's most comprehensive lipid structure database (Fahy et al., 2011). Therefore, the study targeted GL and GPPL in the analysis of lipid classes. GL is divided into three groups, DAG, MAG, and TAG. While the lipids classes, GPPL

consisted of phosphatidic acids (PA), Phosphatidylcholines (PCh), phosphatidyl-glycerols (PG), and phosphatidylinositol (PI). The group is analyzed using LC-MS/MS, which requires a minimum sample provision. This study used IPA to dissolve total fats before being injected into LC-MS/MS equipment. Technological advances LC-MS/MS allowed assisted by LipidView™ software to produce quick and reliable results.

Numerous studies have investigated MAG, DAG, and TAG in *halal* analysis to discriminate lard. For example, PCA and GC-TOF-MS successfully compared the FAs composition and thermal profiles of MAG and DAG of six commercial emulsifiers to distinctly identify partial acyl-glycerols of lard from those of sunflower, corn oil, butter, and palm oils (Nasyrah et al., 2014). It was reported that the δ of ^{13}C values of MAG and DAG of lard differed significantly ($p<0.05$). It also has been demonstrated that the MAG and DAG derived from chicken fat, beef fat, and mutton fat contributed by stearic, oleic, and linoleic using the EA/IRMS combined with PCA (Naquiah et al., 2016).

Most comparative studies are conducted between lard and plant fats or oils. The characterization of lard by MAG and DAG (C18 & C16) GC×GC-TOF-MS in the glycerolysis products derived from five different lipids, including lard, sunflower seed oil, corn oil, butter, and palm oil, was performed earlier by Indrasti et al., (2010). However, there are limited comparative studies between lard and other animal fats. Both lard and other animal fats (i.e., poultry and cattle) have a similar chemical composition that is challenging to differentiate. Therefore, the study of lard is opposed to other edible fats, i.e., animal fats using LC-MS/MS after the heating-process is also

limited. Nevertheless, from this study, lard has been found pronouncedly different at temperatures 180 ºC at 0.5, 1, and 2 hrs, contributed by DAG 32:0 and MAG 24:0.

Glycerophospholipids (GPPL) are mainly found in vegetable oils' saponifiable portion. However, due to a lack of sensitive and selective analytical tools, GPPL has received little attention in edible fats, as commonly as in lard analysis. GPPL is composed primarily of *sn*-1,2-DAG with a phosphate residue in the *sn*-3 position coupled to a simple organic molecule known as the head group.

Depending on the identity of the head group, GPPL can be divided into several classes. PCh, PE, PS, PG, PI, and PA, are the main groups of GPPL found in mammalian cell membranes (Han & Gross, 2005). From these studies, lard was found to be distinctively different at temperatures 180 ºC at 0.5, 1, and 2 hrs, contributed by PA 20:0, PA 21:0, PA 23:0, and PA 25:0 (Appendix 14). The LC-MS/MS-GL-PCA is nearly identical to LC-MS/MS-GPPL-PCA model, which are showed the heated at 180 °C (0.5, 1 & 2 hrs) and labelled as L5, L6 and L7 of the lard samples have more variation in lipid compounds than the other samples.

Identifying GLs and GPPL compounds using a combination of LC-MS/MS-PCA of such samples is a novel discovery of differentiation of lard and other fats after the heating-process. The dramatically different profiles of lard according to the heating-process were identified using LC-MS/MS techniques with more advantages than GC-FID.