Discrimination of Lard and other Edible Fats after Heating Treatments using Partial Least Square Regression (PLSR), Principal Component Regression (PCR) and Linear Support Vector Machine Regression (SVMR).

Nor Aishah Mohd Salleh¹, Mohd Sukri Hassan¹²a

¹ Faculty of Science and Technology, Universiti Sains Islam Malaysia, Bandar Baru Nilai, 71800 Nilai, Negeri Sembilan, Malaysia.
² Institute of Halal Research and Management, Universiti Sains Islam Malaysia, Bandar Baru Nilai, 71800 Nilai, Negeri Sembilan, Malaysia

a Corresponding author: mohdsukri@usim.edu.my

Abstract. Discrimination between lard and other edibles fats is a challenging task for halal determination especially after the fats were heated at high temperature for a long period. In this study, three multivariate regression models such as partial least square regression (PLSR), principal component regression (PCR) and support vector machine regression (SVMR) were applied to evaluate the spectral data of FTIR (n=195) obtained from lard, chicken, beef, mutton and vegetable fats after heated at different conditions (120-240°C and 0.5-3 hrs). The regression of the $Y$-binary matrix was used to discriminate lard (as 1) and the others edibles fats (as 0). Kennard Stone (KS) algorithm selected a subset of the training set (n=145) and test set (n=50). The test set was used to validate the prediction ability of the suggested models. The obtained results showed the ability of the three proposed models to discriminate the heated lard simultaneously. The values of the $R^2$, adjusted $R^2$, root-mean-square error (RMSE) and root-mean-square error of validation (RMSEV) showed a good results under Basic ATR correction transformation as PLSR (0.984, 0.977, 0.052 and 0.062); PCR (0.974, 0.971, 0.067 and 0.070), and SVMR (0.971, 0.959, 0.087 and 0.102) respectively. However, when using mean square error (MSE), it gives lower prediction error for PLSR (0.006), PCR (0.007) and SVMR (0.015). The results showed that PLSR as the best model for discrimination spectral data of lard and other edible fats after heating treatments for halal determination.
1. Introduction

Lard is prohibited from being consumed by Muslim. In the food product, animal’s fat including lard, has been used widely as an additive ingredient in processed food such as bakeries, pastries and frying. Food processing involves elevated temperatures as the main parameters of cooking, baking and frying. The identification of lard spectra has been made extensively and their mixtures with other fats in the classification, authentication and prediction purposes by Fourier Transform Infrared (FTIR). Combination of the spectroscopic instrumentation and chemometrics techniques are promising feasibility study in halal analysis. Most successful identification of triacylglycerol and fatty acid fingerprints such as the calibration models developed for the binary mixtures of other fats to address the halal issues using PCA and PLS. For instance of lard research using FTIR combined with chemometrics for halal purposes are; quantitation lard in the mixtures with body fats of lamb, cow and chicken by PLS & DA [1]; the authentication lard, cod liver, canola, corn, extra virgin olive, grape seed, palm, pumpkin seed, rice bran, sesame, soybean, sunflower, walnut & virgin coconut oil by PCA and PLS [2]; the authentication lard and beef fat in meatball broth by PCA and PLS; the authentication lard adulteration in cake formulation [4].

During measurement of fats by FTIR, the undesired scattering effect and systematics variation might happen by light scatters [5]. Therefore, data transformation is an essential step to minimise the variation and systematic error before applied to chemometrics bilinear modelling. Spectral pre-processing techniques consist of mathematical methods for correcting any error during reflectance measurements. Spectroscopy instruments of offers software such as basic ATR correction to overcome the scattering effect issues and enhance the data points for challenging spectrum. There are four type basic pre-processing techniques such as smoothing, baseline removal, scaling and normalisation. The most common pre-processing techniques on spectral oils and fats are Savitzky-Golay filters and second derivative [6]-[8].

Calibration, validation and prediction of the oils and fats employed by chemometrics techniques have been enhanced through revolutionizing statistical classification and data mining techniques. Traditional statistical classification methods such as Principal Components Regression (PCR) and Partial Least Squares (PLS) have been extensively used in edibles oils and fats classification problems for which the criterion variable is to compared with non-parametric techniques [9][10]. PLS is probably the most popular use in oils and fats regression analysis using infrared data. For examples, PLSR demonstrating better prediction of lard in palm olein oil at polyunsaturated fatty acids detected by mid-infrared wavenumber yielding % RMSE of 16.03 compared to the Simple Linear Regression (SLR) and Multiple Linear Regression (MLR) [11]. Monitoring quality parameters such as the acidity of olives oils resulting values of $R^2$ at 0.96, by PLS [12]. The others data mining spectral data was a combination of PLS and Near Infrared (NIR) spectroscopy has resulted in a good predictive capability with relative RMSE values below or equal to 1.8% for determination of fatty acids and lipid classes in salmon oil [13]. More recently, researches have been constructed on the accuracy and efficiency of data mining, with classifiers like Support Vector Machines (SVM), used for oil and fats prediction and classification dichotomous [14]-[16].

In halal research, it is crucial to determine lard from other edible fats after heating treatments, and this task is challenging due to the similarity of chemical composition and thermal degradation. Regression techniques by chemometrics approach were employed by considering the classic statistical such as PLS, PCR and the recently most powerful classification, SVM. In this paper, the performances
chemometrics model on heated-edibles fats were evaluated to select the best techniques for chemometrics on spectral data. The objective of the study is to demonstrate regression analysis on spectral data of various edibles fats to discriminate lard after heating treatments. The different pre-processing or transformation techniques such as Basic ATR (BA) and Savitzky-Golay Second Derivative (SG2D) were proposed prior chemometrics techniques and the performance comparison between the models of PLSR, PCR and SVMR were evaluated.

2. Methodologies

2.1 Heating Treatments of Fats and Sample Preparation

A total of 5 types of fats, beef, chicken, lard, beef, and plant were purchased at markets in Nilai, Malaysia. Animal fats were minced and weighted at 100 g (w/w) and placed on hot digital in 250 ml quartz beakers. Heating treatment design was applied according to Salleh et al., 2018 [17]. The temperatures were designed at three levels, 120 °C, 180 °C, and 240 °C were controlled by a probe that contacted into fat tissues. The heating intervals were selected at 0.5, 1, 2 and 3 hours (hrs). Each 20 ml aliquot of heated oils were collected and cooled at room temperatures before kept into the universal bottle at 4 °C in the chiller. Fats were extracted to remove protein and tissue residuals according to the Folch, et al., 1957 [18]. The schematic of the experiment steps are shown in Fig. 1.

2.2 Acquisition of Spectra in the Infrared (FTIR)

Fats are registered on the disc by infrared spectra and data were acquired on a Perkin Elmer model Spectrum 400 FTIR Spectrometer using Attenuated Total Reflectance technique (ATR-FTIR). A range from 4,000 to 650 cm\(^{-1}\) was scanned, with a resolution of 8 cm\(^{-1}\) and 16 scans. Background of the air spectrum was taken after each scan. The spectra of each fat were acquired with triplicate.

2.3 Randomization

The Kennard-Stone (K-S) algorithm selected the training (n=145) and test sets (n=50) at 70:30 ratios. The test set was used to validate the prediction ability of the suggested models [19]. The K–S algorithm chooses a subgroup of samples from \(g_1\) that provide a demonstrative and identical coverage over the data matrix. The method initiates by discovery the two samples that are the farthermost in the ordered distance of PCA scores plot of the raw data before the transformation.

2.4 Spectra Pre-processing

The obtained spectra were transformed or pre-process before subjected to Principal Components Analysis (PCA). The original data were transformed into the two techniques; firstly, the basic BA has applied the correction to minimise scattering effect. This method maintained the original data wavelength and automatically normalised the absorbance of wavelength. The second method of transformation is the Savitzky-Golay Second Derivative (SG2D) and the second polynomial with smoothing 15 points. The derivatives techniques have capabilities to remove both additive and multiplicative effect in the spectra.
2.5 **Principal Component Analysis (PCA)**

Principal Component Analysis (PCA) is probably the most used in the linear projection method of multivariate data [20]. Its goal is to extract the most useful information from the dataset and creates new orthogonal variables (latent variables) that are linear combinations of the original $x$ variables. For that reason, a spectral data matrix can be decomposed as shown in the equation:

$$X = TP' + E$$  \hspace{1cm} (1)

$X$ is a $m \times n$ matrix of spectral data, $T$ is $m \times k$ matrix of score values for all of the spectra, and $P$ is $k \times n$ matrix of principal components (PC). The $E$ matrix contains the spectral residuals not fit by the optimal PCA and has the same dimension as $X$, $m$ is the number of samples (spectra), $n$ is the number of data points (wavelengths), and $k$ is the number of PC used reconstruct the $X$, which is $k < m$. The superscripts (') denotes matrix transpose.

2.6 **Multivariate Regression**

Principal Component Regression (PCR), Partial Least Squares Regression (PLSR) and Support Vector Machine Regression (SVMR) were selected to apply on the data matrix of heating treatment. SVM was using a linear parameter to give a fair comparison between PLS and PCR that constructed by linear principles. The discriminant of SVMR as C value: 1, Nu value: 0.5 and Rho at 0.138 were set up. Data transformation and regression method by chemometrics software UNSCRAMBLER® version 10.3 X by CAMO Software USA.

2.6.1 **Principal component regression (PCR)**

The two-step multivariate method of PCR is involving PCA of the data matrix $X$ is performed by converted and measured variables (e.g., absorbance at different wavelengths) into new latent, followed by multiple linear regression steps (MLR). MLR was applied between the scores obtained in the PCA step and the characteristic $y$ to be modelled.

Moreover, the first new variables or PCs are assumed; accounting is the most variance of the original data contains meaningful information, while the additional numbers of PC sequences are less than the first PC. Consequently, the only $r$ PCs are retained and $r < \min (n, p)$ to data simplification. After performing PCA on $X$, the second step in PCR consists of the linear regression of the scores and the $y$ property of interest. The linear model between $(n \times p)$ and $A (n \times r)$ in the following the form:

$$Y = Ab + e$$  \hspace{1cm} (2)

$A$ is the weighted normalised matrix of order $n \times p$ and will represent the new co-ordinates for the $n$ objects in the new system, and $p$ is the loading matrix, and the column vectors are called eigenvectors or loading-PCs. Where $b$ ($r \times k$) are the coefficients and $e$ is the error vector ($n \times p$). The objectives of the response $y$ is not directly correlated with $X$, but with its PC. The PC is obtained by decomposing $X$ via principal component analysis (PCA) as follows equation (1). The prediction responses on $X$ with $\hat{b}_r$ are given by:

$$\hat{y} = X\hat{b}_r$$  \hspace{1cm} (3)
2.6.2 Partial least square (PLS)

The PLS algorithm finds principal components from spectral data that are also relevant to analytic concentration [21]. PLS is a method for constructing regression models on the latent or hidden variable decomposition to relate two blocks, matrices \( X \) and \( Y \), which contain the independent, \( x \), and dependent, \( y \), variables, respectively. The objective PLS is to find the best number of latent variables, which is typically performed by using cross-validation (CV), based on the determination of minimum prediction error.

\[
X = \sum t_f \hat{p}_f + E \tag{4}
\]
\[
Y = \sum u_f \hat{q}_f + F \tag{5}
\]

In which \( T \) and \( U \) are the score matrices for data set of \( X \) in Equation (4) and \( Y \) in Equation (5) respectively; \( p \) and \( q \) are the loading matrices for \( X \) and \( Y \), respectively, \( E \) and \( F \) are the residual matrices. The two matrices are associated by the scores \( T \) and \( U \), for each latent variable, as follows.

\[
u_f = b_f + t_f \tag{6}
\]

The matrix \( Y \) can be calculated from \( u_f \), as Equation (6), and the constant of the new samples can be estimated from the new scores \( T \), which are substituted in Equation (4) and (5), leading to Equation (7)

\[
Y_{new} = TBQT \tag{7}
\]

2.6.3 Support vector machine

SVM can be applied to regression by the introduction of an alternative loss function, and the results perform to be favourable. The basic idea of SVRM is to map the data \( X \) into a higher-dimensional feature space \( F \) via a nonlinear mapping and then to do linear regression in this space. Thus, regression approximation addresses the problem of estimating a function established to a given data set \( G = \{(x_i,d_i)\}_{i=1}^{I} \) (\( x_i \) is the input vector, \( d_i \) is the chosen value). SVM approximates the function in the following the form:

\[
y = \sum_{i=1}^{I} w_i \Phi_i(w) + b \tag{8}
\]

the \( \{\Phi_i(x)\} \) is the set of mappings of input features, and \( \{(w)\}_{i=1}^{I} \cdot b \) are the coefficients.

2.7 Model evaluation

Evaluation of model accuracy by statistics has been employed such as coefficient of correlation in calibration (R squared or R\(^2\)) coefficient of correlation in prediction (adjusted R squared or adj. R\(^2\)) as per equation (9), root mean squared of error validation (RMSE) and prediction (RMSEV) and root mean square of error as per equation (10).
where $y_i$ is the calibration value, $\hat{y}$ is predicted by the multivariate techniques model and $\bar{y}$ is the average of the calibration value.

$$\sqrt{\frac{\sum_{i=1}^{n_c} (y_i - \hat{y})^2}{\sum_{i=1}^{n_c} (y_i - \bar{y})^2}}$$

(9)

where $n_c$ is the number of calibration or validation set, $\hat{y}_i$ and $y_i$ are the predicted and calibration values of the $i^{th}$ observation. The model has prediction performance capability measured by the mean squared error of the prediction (MSEP) as per the equation below.

$$\frac{1}{n_p} \sqrt{\sum_{i=1}^{n_p} (y_i - \hat{y}_i)^2}$$

(10)

Where $n_p$ is the prediction set denoted by $\hat{y}_i$ and $y_i$ are the predicted and measured values of the $i^{th}$ observation.

The smallest RMSEV value was associated with the optimal calibration and validation model. $R^2$ shows the percentage of the variance in the Y variables that represents the variable X. The values of $R^2$ nearest to the 1 is considered an excellent model performance in calibration. The RMSEV, RMSEP and MSEP values is an indicator of the reliability and predictive ability of the model and should be lower into 0.

Figure 1: Flow chart of chemometrics to evaluate the heated lard and other fats by FTIR analysis.
3. Results & Discussion

Measurement of FTIR spectral data set of n=195 (lard, chicken, beef, mutton and vegetable fats) spectra measured at 1743 different wavenumbers containing 339 885 data points. Illustration of the complex data points by line plot using UNSRAMLER® software that resemblance of the origin data. The FTIR spectra (650-4000 cm$^{-1}$) of typical fats samples are shown in Fig.2.

There were several peaks located 715-1006 cm$^{-1}$, 1166-1370 cm$^{-1}$, 1499-1746 and 2800-3025 cm$^{-1}$ [22][23]. Scattering noises could be observed at the beginning and end of spectral data due to system imperfections. Therefore, data points of the wavelength were eliminated at below 715 cm$^{-1}$. As can be seen the weak band at 3000-3400 cm$^{-1}$ assigned to amines groups (protein) contribute by mutton and beef fats that might come from inconsistent fats extraction. Thus the peaks 715-3100 cm$^{-1}$ were selected because this study only focused on the wavelength bands related to the lipid.

Figure 2. Spectral FTIR after Basic–ATR (BA) transformation.  
Figure 3. Spectral FTIR after Savitzky-Golay Second Derivative (SG2D) transformation.

3.1 Spectral pre-processing

The training set was applied to the BA and SG2D transformation and selected peak (715-3100 cm$^{-1}$) were subjected to the PCA to observe the nature of the distribution of the heated fats. BA transformation is similar from the original (not shown), the only magnitude of absorbance is standardised between (0.2 and –0.2) as per Fig. 2. SG2D enhanced the main peaks at 1166-1370 cm$^{-1}$, 1499-1746 and 2800-3025 cm$^{-1}$ by producing the sharp peaks. Prior transformation, the raw data matrix, $X$ ($m \times n$) or (195x1743) were applied to PCA in order to a subdivision of training set and test set using K-S algorithm. The calibration was produced a new coordinate, $(m \times k)$ as per equation (1) consist of 145 x $k$ for the training set. Then the data were transformed to BA and SG2D to observe the comparison the best discrimination of fats by visualization.
3.2 Principal Components Analysis

The numbers of PC, $k=2$ were chosen to observed the discrimination of lard. It can be observed that BA-PCA were performed in order to get an overall distribution of 139 training set (after removing some outliers) of heating treatment edibles fats as showed in Fig 4. The first two PC is sufficient to explain the variances as PC-1 (70%), and PC-2 (11%) and a total of variances are 81% to observe the discrimination of lard with the other fats. Lard and chicken are closed together, and some points were overlapping at the positive side of PC 1 contrast from plant fat at negative PC-1. Beef and mutton fats are positioned near the origin and mixed. Some of the points beef and mutton contributed to negative side PC1.

As can be seen at Fig. 4, the SG2D-PCA model was discriminated well between lard and chicken fats along PC1 (33%) and PC2 (17%) by accounted into 50% of the total variation. From both (Fig 4 & 5) PCA visualization data of the transformation of BA-PCA has given poor discrimination than SG2D-PCA. However, the BA line plots (Fig. 1) has a similar to the origin wavelength that important for variables interpretation upon discrimination of lard and other fats. Likewise, SG2D transformations tend to smooth out some relevant information.

The other point, PCA are not advised to be decisive of discrimination power because based on total variance captured by the first 2Pcs, BA-PCA is over than SG2D-PCA. In a sense, the multivariate regression methods are proposed to apply to the data.

![Figure 4. BA-PCA](image1)

![Figure 5. SG2D-PCA](image2)

3.3 Multivariate Regression

The other chemometrics approach was suggested to be applied dummy dependent variables to achieve the objectives on halal analysis. The regression analysis of PLSR, PCR, and SVMR were recommended in order discriminate halal and non-halal fats by using regression of the Y-binary matrix was used to discriminate lard (as 1) and the others edibles fats (as 0). The regression techniques employed the dependent variables as the dummies been done in chemometrics techniques by the other field studies [24].

After applied regression method, only PLSR and PCR can be visualised the distribution of the plot. The scores plot for PLSR displayed factors component $(X, Y)$ and PCR denotes by PC is to summarize more variation in the data than any other pair of components. BA-PLSR (Fig. 6) showed separation between lard and the other fats. The sums of explained variance $X$ of this model is 81% (72%+2%) better than SG2D-PLSR (Fig.7), 47% (19% + 28%). The sum of the explained variances for the two
components of X is large (70%-80%) than variances of Y, the plot shows a large portion of the information in the data indicated the relationships between X and Y with a high degree of certainty.

The outcomes of the BA-PCR (Fig. 8) model have the same pattern as BA-PCA (Fig. 3) because PCR utilised the same concept as PCA which are based on maximum variances plots. Plots of SGD2-PLSR and SGD2-PCR (Fig. 9) have the same visualisation results as SGD2-PCA (Fig. 3) with good separation of lard with other fats.

Figure 6. BA-PLSR

Figure 7. SGD2-PLSR

The outcomes of the BA-PCR (Fig. 8) model have the same pattern as BA-PCA (Fig. 3) because PCR utilised the same concept as PCA which are based on maximum variances plots. Plots of SGD2-PLSR and SGD2-PCR (Fig. 9) have the same visualisation results as SGD2-PCA (Fig. 3) with good separation of lard with other fats.

Figure 8. BA-PCR

Figure 9. SGD2-PCR

3.4 Model Evaluation

The evaluation of PLSR, PCR and SVMR models using techniques regression according to data transformation are summarised as Table 1. BA-PLSR showed the highest performance for calibration, validation (R²=0.984, adj. R²=0.977, RMSE=0.052). The performances values of BA-PCR are closer to BA-PLSR (Fig. 8 & Fig. 9) but from the visualization, PCR less robust than PLSR because the
separation between lard and chicken have not happened at the first 2 components. BA-SVMR model gives better performance on validation than SG2D-SVMR. Regression analysis showed that BA transformation gives better results than SG2D.

The overall model performances are satisfactory (>0.80) except SG2D-SVMR. The SG2D-SVMR showed overfitting trends based on validation or adjusted $R^2$ (0.799) higher than calibration or $R^2$ (0.893). Scores plot pattern was suggested to be influenced by the results of the evaluation. As can be seen from Fig 4, SG2D-PLSR plots were distributed in the circle's pattern compared to BA-PLS that has a linear pattern for each group of heated fats. Moreover, the parameters of SVMR applied in the study are linear based. The results may better if the non–linear parameters of SVMR were applied.

The test set (n=50) were applied to each prediction. The entire models indicated a satisfactory prediction quality except BA-SVMR and SG2D-SVMR which are produced a poor prediction performance based on MSEP (0.015 and 0.18).

Table 1. The calibration, validation (training set) and prediction (test set) of lard and edibles fats after heating treatments from FTIR data.

|       | R square ($R^2$) | Adjusted R square (adj $R^2$) | RMSEC | RMSEV | MSEP |
|-------|-----------------|-------------------------------|-------|-------|------|
| BA    |                 |                               |       |       |      |
| PLS   | 0.984           | 0.977                         | 0.052 | 0.062 | 0.006|
| PCR   | 0.974           | 0.971                         | 0.067 | 0.07  | 0.007|
| SVM   | 0.971           | 0.959                         | 0.087 | 0.102 | 0.015|
| SGD2  |                 |                               |       |       |      |
| PLS   | 0.977           | 0.974                         | 0.063 | 0.067 | 0.007|
| PCR   | 0.978           | 0.976                         | 0.060 | 0.063 | 0.006|
| SVM   | 0.799           | 0.893                         | 0.464 | 0.464 | 0.180|

4. Conclusion

This study implemented the simultaneous chemometrics technique to evaluate for halal determination on edibles fats after heating treatments by using Fourier transform infrared (FTIR). Spectra transformations such as BA and SG2D were applied for an optimised modelling band. Although the discrimination of lard by BA-PCA is not robust as SD2D-PCA, the other application of regressions is suggested as a complementary method for discrimination. BA-PLSR model gives the best superior in model performances. It also can be concluded that transformation data by maintaining resemblance original data curves are suitable for regression method employed by dependent variables into the dummies codes (0, 1).
References

[1] Rohman A, Kuwat T, Retno S, Yuni E, and Tridjoko W. 2012 Fourier Transform Infrared Spectroscopy applied for rapid analysis of lard in palm oil. *Int. Fd. Res. J.* 19 1161.

[2] Man YC, Syahariza ZA, Mirghani ME, Jinap S, and Bakar J. 2005 Analysis of potential lard adulteration in chocolate and chocolate products using Fourier transform infrared spectroscopy. *Fd. Chem.* 90 815-9.

[3] Kurniawati E, Rohman A, and Triyana K. 2014 Analysis of lard in meatball broth using Fourier transform infrared spectroscopy and chemometrics. *Meat Sci.* 96 94-8.

[4] Syahariza ZA, Man YC, Selamat J, and Bakar J. 2005 Detection of lard adulteration in cake formulation by Fourier transform infrared (FTIR) spectroscopy. *Fd Chem.* 92 365-71.

[5] Bjorsvik, H. -R., & Martens, H. 2000. Calibration of NIR instruments by PLS regression. In E. W. Ciurczak, & D. A. Burns (Eds.), *Handbook of near-infrared analysis (3rd ed.).* New York, USA: CRC Press Taylor & Francis Group, pp. 189–205.

[6] Yang M, Nie S, Li J, Xie M, Xiong H, Deng Z, Zheng W, Li L, and Zhang X. 2010 Near-Infrared Spectroscopy and Partial Least-Squares Regression for Determination of Arachidonic Acid in Powdered Oil. *Lipids.* 45 559-65.

[7] de Santana F.B, Gontijo L.C, Mitsutake H, Mazivila S.J, de Souza, L.M. and Neto, W.B., 2016. Non-destructive fraud detection in rosehip oil by MIR spectroscopy and chemometrics. *Fd Chem*, 209 228-233.

[8] Vilela J, Coelho L, and de Almeida JM. 2015 Investigation of adulteration of sunflower oil with thermally deteriorated oil using Fourier transform mid-infrared spectroscopy and chemometrics. *Cogent Fd. & Agric* 1 1020-254.

[9] da Silva CE, Filardi VL, Pepe IM, Chaves MA, and Santos CM. 2015 Classification of food vegetable oils by fluorimetry and artificial neural networks. *Fd Control.* 47 86-91.

[10] Devos O, Downey G, andDuponchel L. 2014 Simultaneous data pre-processing and SVM classification model selection based on a parallel genetic algorithm applied to spectroscopic data of olive oils. *Fd. Chem.* 148 124-30.

[11] Karunathilaka, S.R., Farris, S. and Mossoba, M.M., 2018. Rapid Prediction of Low (< 1%) trans Fat Conten in Edible Oils and Fast Food Lipid Extracts by Infrared Spectroscopy and Partial Least Squares Regression. *J. of Fd. Sci.* 83 (8), pp.2101-2108.

[12] Fernández-Espinosa, A.J., 2016. Combining PLS regression with portable NIR spectroscopy to on-line monitor quality parameters in intact olives for determining optimal harvesting time. *Talanta,* 148, pp.216-228.

[13] Cascant, M.M., Breil, C., Fabiano-Tixier, A.S., Chemat, F., Garrigues, S. and de la Guardia, M., 2018. Determination of fatty acids and lipid classes in salmon oil by near infrared spectroscopy. *Fd Chem,* 239, pp.865-871.

[14] Hou X, Wang G, Su G, Wang X, and Nie S. 2019 Rapid identification of edible oil species using supervised support vector machine based on low-field nuclear magnetic resonance relaxation features. *Fd. Chem.* 280 139-45.

[15] Shen X, Zheng X, Song Z, He D, and Qi P. 2012 Rapid identification of waste cooking oil with near infrared spectroscopy based on support vector machine. *International Conference on*
[16] Zhao M, Nian Y, Allen P, Downey G, Kerry JP, and O'Donnell CP. 2018 Application of Raman spectroscopy and chemometric techniques to assess sensory characteristics of young dairy bull beef. *Fd. Res Int.* **107** 27-40.

[17] Salleh NA, Hassan MS, Jumal J, Harun FW ,and Jaafar MZ. 2018 Differentiation of edible fats from selected sources after heating treatments using fourier transform infrared spectroscopy (FTIR) and multivariate analysis. *AIP Conf. Proc. Jun 5 1972.* (AIP Publishing) p. 030015

[18] Folch J, Lees M, and Sloane Stanley GH. 1957 A simple method for the isolation and purification of total lipides from animal tissues. *J. Biol. Chem.* **226** 497-509.

[19] Kennard R.W., Stone L.A. 1969 Computer aided design of experiments. *Technometrics*, **11**, 137–14

[20] Abdi, H. and Williams, L. J. 2010, Principal component analysis. *WIREs Comp Stat*, 2: 433- 459.

[21] Geladi P and Kowlakski BR . 1986 Partial least squares methods regression: A *Tutorial Anal. Chemica Acta*. **185**, 1-17.

[22] Guillen MD, and Cabo N. 1997 Characterization of edible oils and lard by Fourier transform infrared spectroscopy. Relationships between composition and frequency of concrete bands in the fingerprint region. *J Am Oil Chemists' Society (JAOC)* **74** 1281-6.

[23] Vlachos N, Skopelitis Y, Psaroudaki M, Konstantinidou V, Chatzilazarou A, Tegou E. 2006 Applications of Fourier transform-infrared spectroscopy to edible oils. *Anal. Chim. Acta* **573** 459-65.

[24] Zhu, Y. and Tan, T.L., 2016. Penalized discriminant analysis for the detection of wild-grown and cultivated Ganoderma lucidum using Fourier transform infrared spectroscopy. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, **159**, pp.68-77.

[25] Fu, H., Jiang, D., Zhou, R., Yang, T., Chen, F., Li, H., Yin, Q. and Fan, Y., 2017. Predicting mildew contamination and shelf-life of sunflower seeds and soybeans by fourier transform near-infrared spectroscopy and chemometric data analysis. *Fd Anal. Methods*, **10** (5), pp.1597-1608.

**Acknowledgments**

The research financial was fully supported by Ministry of Higher Education of Malaysia under Fundamental Research Grant Scheme (FRGS) No. 50715.