VOL. 11, NO. 6, MARCH 2016 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

TOWARDS AUTHENTICATION OF BEEF, CHICKEN AND LARD USING MICRO NEAR-INFRARED SPECTROMETER BASED ON SUPPORT VECTOR MACHINE CLASSIFICATION

Ibrahim J. Alfar[1], Aida Khorshidtalab[1], Rini Akmeliawati[1], Salmiah Ahmad[1] and Irwandi Jaswir[2]
[1]Department of Mechatronics Engineering, Malaysia
[2]International Institute for Halal Research and Training International Islamic University Malaysia, Gombak, Malaysia
E-Mail: ibrahim.alfar@gmail.com

ABSTRACT

NIR (near infrared) spectrometer utilized a quick reliable mean of molecular chemical detection. In this paper, we propose a method on authenticating fats originated from beef, chicken and lard. These compositions can be identified by NIR spectrometers through qualitative and quantitative analysis. Yet most of the analysis lack the capacity to find a pattern in the spectrums to be used in classification or regression models. The disadvantage of spectrum after all is the inability to show the concentration of fatty acids, yet these fatty acid components are shared by all kinds of fat/oil. Therefore, a new method is proposed to create a clear and a distinguishable pattern for the classification. The spectrum of each group (beef fat, chicken fat and pig fat "lard") of samples were acquired using a readymade micro-NIR spectrometer. The raw data required further processing before using it in the classifier. These processes including standard normal variant and Savitsky-Golay smoothing. Furthermore, the processed data was classified using support vector machine (SVM) with polynomial kernel. The trained SVM model showed 98.33% accuracy for 10-fold cross validation and 86.67% accuracy for unseen/testing data. For each individual kind of fat the model was able to identify the unseen data satisfactorily as follows lard with 100% accuracy and combined, chicken and beef showed 80% accuracy.

Keywords: NIR, spectrometer, SVM, classification, fatty acids, fats/oils.

1. INTRODUCTION

Differentiation of oils/fats can be non-trivial to our body sensory systems, such as using the naked eye, the smell and taste. Here is where technology takes over the authenticity of food content more precisely. There are various approaches to detect and quantify the level of certain food substance percentage in a product. These techniques vary in method of detection, detection time, physical properties (e.g. size, weight, and so on.) and their accuracy. The most frequently used methods include the Fourier Transform Infrared (FTIR) Spectroscopy, Polymerase Chain Reaction (PCR), gas or liquid chromatography, and electronic nose. Still the portability of these technologies remains a great problem as they require a high power supply, they are available in bulky sizes, the testing has to be done in a controlled environment (laboratory) and above all, they are expensive. The present study was conducted to investigate the possibility of authentication of lard from other animal fats such as chicken and cow body fats using a portable mean of detection, by utilizing a micro, near infrared spectrometer, to eliminate the excessive sample preparation commonly used.

The basic operation of food analysis devices is mostly to stimulate the chemical compound/compounds of interest, then to read/analyse their responses. The near infrared (NIR) spectroscopy technique provides assistance for rapid process monitoring. Since organic substance such as food is the most absorption-sensitive in the range of 1100 to 2300 nm [1]. Near infrared region has the most appropriate energy (700-2500 nm) to excite single covalent molecular bonds as these bonds greatly differ from one substance to another.

Food and any other organic substances mainly consist of Carbohydrates (sugars, starch, and cellulose), Lipids (Glycerol, Triglyceride and Fatty acids), Nucleic acids (DNA and RNA) and/or Protein. Selecting one of these components as a fingerprint for authentication purposes depends on two complementary elements; the sample molecular structure and the technology to detect it. This fact concurs with the study of A. Rohman [2] which states that analysis of fats/oils is usually carried out by only focusing on lipids components.

NIR spectroscopy offers a rapid analytical method and quantifications of substance which are useful for on-line and in-line monitoring of food industry such as meat, fruit, grain dairy products, beverages and many other areas [3]. The spectrum can provide information about the chemical structure of the sample by identifying the peaks' location and their intensity. To locate the functional group and the molecular bond absorption on the wavelength, the chemical bonds are treated as a mechanical spring connecting two weights (molecular). Subsequently, the frequency can be determined with the help of Hook's law. Therefore, qualified analysis is conducted to identify the sample nature whether it is fat or carbohydrate for example, based on the representation of molecular peaks. The drawback of this method is that the spectrum does not give the concentration of each compound present in the sample. In another words, based

on its peaks' location, the functional group of the sample can be identified, however, the exact type of the content is still unknown. Furthermore, if two different kinds of fat spectra are presented, they will have an overlapping spectrum with a small variation due to the fact that fats share the same fatty acid compounds but different concentration [4]. A comparison study conducted by Che Man [5] to find a differentiation regression model between palm oil and lard successfully achieved R^2= 0.998 validity of all data. Another example on the work of Man Che [6] that classified lard from lamb fat is only capable to classify a mixture of lard and LBF (Lamb body fat) up to 30% adulteration. Briefly this method is a common practice between researchers in finding patterns in the data. Regression model is useful where the data assignment is in numbers. On the other hand, authentication process demands a group classification with labels like True/False or chicken/pig. As mentioned, the raw data obtained from fat samples scanned by IR spectrometer are similar to each other.

In chemometrics the information extracted is either quantified or qualified. Quantifying the data means to estimate the concentration of each chemical compound in the mixture, whereas data qualitative analysis explains and identifies the percentage of certain chemical compounds in the mixture [7].

In this paper, our preliminary analysis of fat type authentication by using support vector machine (SVM) is presented. The paper is organized as follows. Section 2 presents a brief overview of SVM. Section 3 presents the proposed method in details. Section 4 discusses the results. Finally, Section 5 concludes and summarises the findings.

2. OVERVIEW OF SUPPORT VECTOR MACHINE

Most machine learning algorithms have been developed and statistically validated for linearly separable data. Popular examples are linear classifiers like Support Vector Machines (SVMs) or the (standard) Principal Component Analysis (PCA) for dimensionality reduction. Nonetheless, most real world data require nonlinear methods in order to perform tasks that involve the analysis and discovery of patterns successfully. SVM technique has improved throughout the years by introducing kernel trick. Kernel trick basically maps the input data to higher dimensions in order to find a pattern in non-linear separable data. Some of the common example of kernels functions are the inner-product based linear, polynomial kernels and the Euclidean distance based Gaussian (Radial Basis Function (RBF)) kernel [8]. These attributes allows SVM to handle featureless and complex data ideally, e.g. spectroscopy data modelling, X-ray signals [9] [10].

3. MATERIAL AND METHOD

Heat treatment

In analytical chemistry the samples' temperatures play a crucial role in changing the chemical structure. Organic substance like fat is highly sensitive to temperature. Where temperature can alter the physical and chemical properties of the samples. The chemical properties will change according to the radical reaction triggered by breaking and forming new bonds. Radical reaction undergoes three stages namely initiation, propagation and termination. These stages depend on each molecular chemical concentration in the mixture (samples). According to the findings of Hashim [4] where an oil-profile comparision of different fats proved that the all four fats namely beef, chicken, lamb and lard shared the same fatty acid components with different concentration. In addition, the concentration determines the type of fat and its chemical and physical properties. Fatty acid compounds can be classified into saturated (long carbon-hydrogyn single bond chain) and unsaturated (same as saturated by with few carbon-hydrogyn double bond).

On the other hand, IR spectrum cannot detect the concentation of these fatty acids componds, it only detects the functional group of these componds. Consequently, if the samples of different fats/oils are to be extracted at diffferent rendereing tempereature they will show different levels of intensity and vice versa. Refering to the study done by Bereton [11], where four different oils are prepared under the same temperature showed a featureless overlapping spectrums. On the other hand, Heyes [12] proved that samples scanned at different temperatures will show changes in the peak levels. These findings concur with the research of Yuliansyah [13] where by the use of FTIR it was concluded that the water molecular in the palm oil was in fact expelled under heat treatment by observing the shifting in peaks representing OH functional group$\approx 3500\ cm^{-1}$. Therefore, fat samples rendered at different temperature will show different levels of absorbance but will maintain the numbers and locations of peaks.

Sample preparation

The sample preparation intended to be simple and without the addition of any unnecessary chemical substance. Each fat group namely beef, chicken and pig were extracted by rendering them at different temperatures ($50C^o, 60C^o, 80C^o, 100C^o$) in an oven. Thus, a total of 12 samples for all group were available.

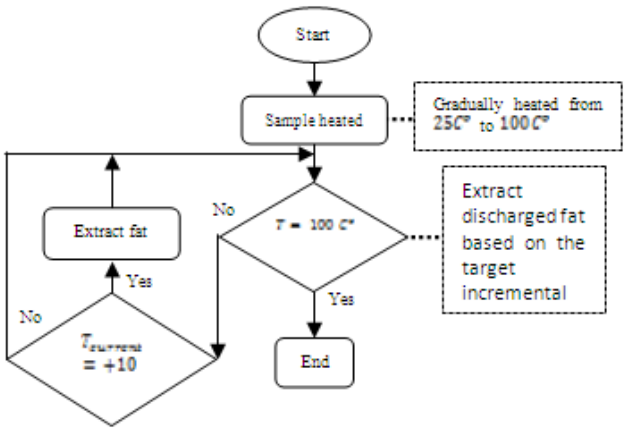


Figure-1. Fat extraction process.

One more step was needed to solidify the sample using KBr (Potassium bromide). KBr has no absorbance or reflective properties in the range of ultraviolet to far-infrared wavelength [14]. Moreover, KBr comes in a form of powder where it's grinded, mixed with samples and pressed into a disk shape. This step has no effects on the samples' chemical structure. In addition, a set of similar samples are also prepared for testing and model validation purposes.

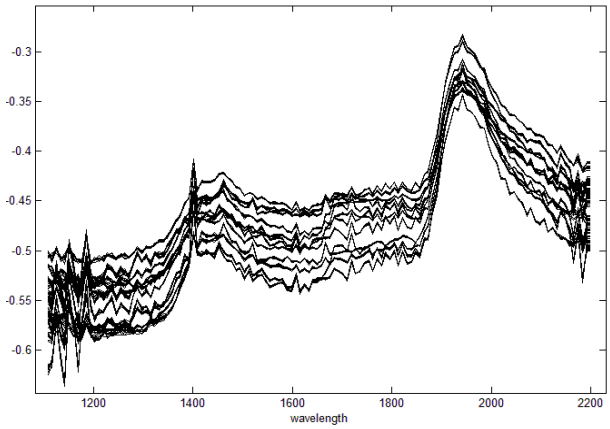
Beer Lambert law states that the absorbance of a sample is directly proportional to its thickness. Thus all samples must maintain the same disk volume otherwise the disks thickness are not equal. Therefore, the absorbance level will not be stable.

Spectrum acquisition

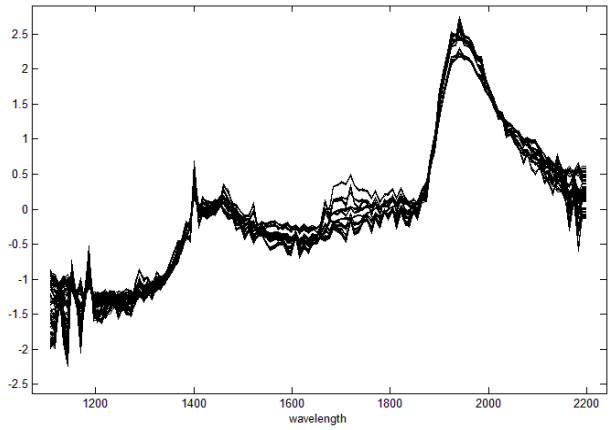
The instrument used in this study was JDSU, a ready-made micro NIR spectrometer. The device exports the captured spectrum sampled to 128 absorption values span over 900 to 1500 nm wavelength. For instance, beef was rendered at $50C^o$ then 10 spectrum were captured at the same temperature to insure there is no major variation between each spectrum scanned. This step was necessary to eliminate any uncontrollable environmental factors acting on the device and the sample.

Spectral data pre-processing

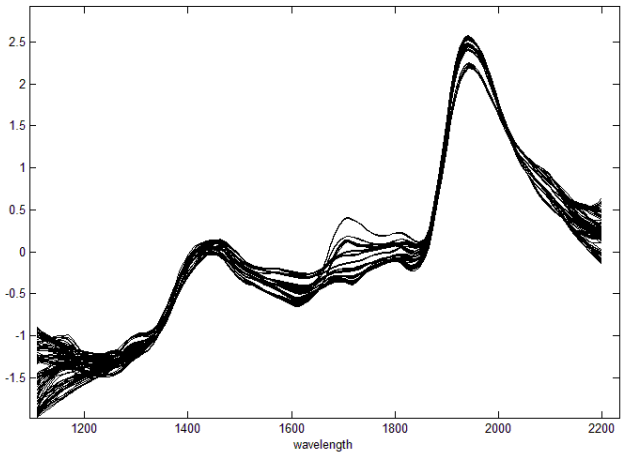
As explained by the company JDSU the devices act more as a sensor than a spectrometer. Therefore, the generated data is not processed and might pick a lot of noises. As a matter of fact, the data needed a few processing steps before the classification is performed. Figure 2 shows all spectrum for all temperature from each group (beef, chicken and lard) a total of 120 spectra. Moreover, the spectra obtained contained noise-spikes (raw data) Figure-2(a). First, a standard normal variant (SNV) is applied to remove the slope variation of each spectra and are normally distributed with zero mean. It is noticeable in Figure-2(b) that the spectra have small variant comparing to the raw data. The variant in absorbance is greatly affected by the light scattering effect and this application of SNV was necessary [15]. Thereafter, two smoothing functions were applied to the data. A robust smoothing function was required to eliminate the noise spikes as the first stage smoothing. This function assigns lower weight to outliers in the regression. Even after the first smoothing, Figure-2(c) shows a smooth trend compared to Figure-2(b) but some spectra still have double peaks. The second stage smoothing or filtering is performed by using the Savitsky-Golay algorithm. These techniques use a number of points and decide about the required variation for each point and the order of polynomial function to be used then the process should be repeated to the next set of points. It is important that the degree of the polynomial is properly selected due to the fact that when the degree of polynomial increases, the signal to noise ratio decreases. Figure-2(d) delineates that all 120 spectra show a smooth trending with the same number and location of peaks. Eventually, it is evident that the wavelength region from 1100 to 1300 contained featureless data. As a consequent, the data were windowed from 1300 to 2200 wavelength for classification.



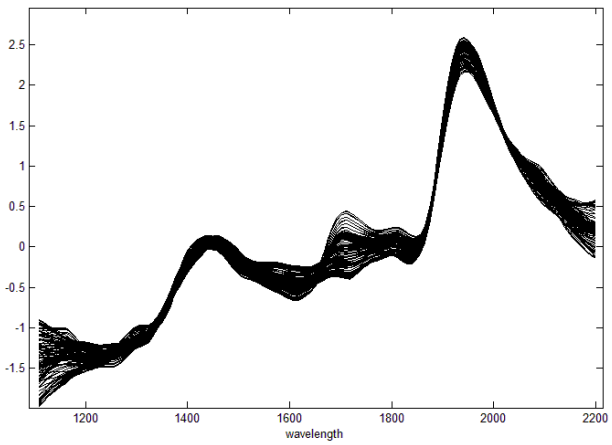
(a)



(b)

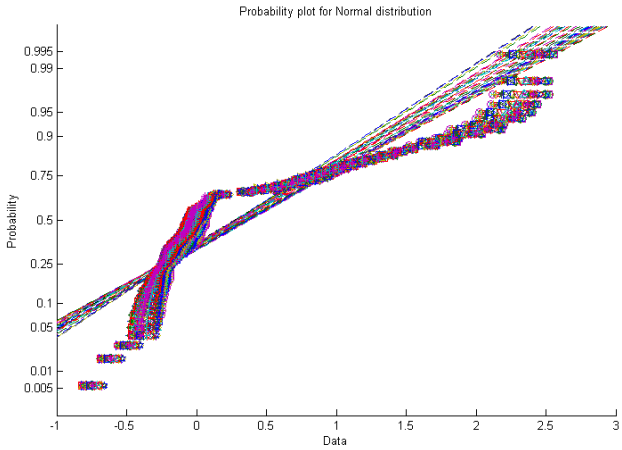


(c)

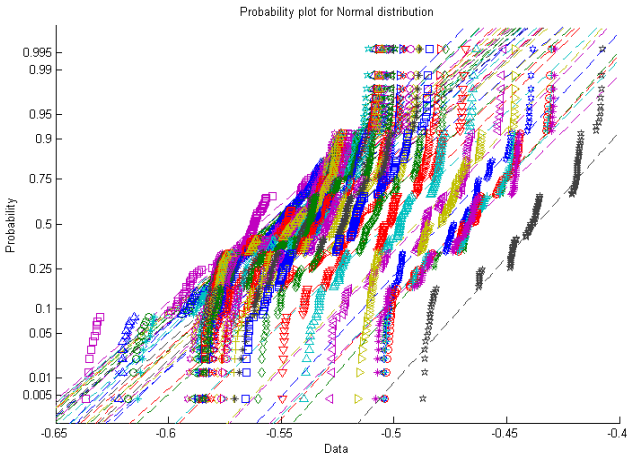


(d)

Figure-2. Data-processing stages (a) Raw data, (b) SNV, (c) Smoothing and (d) Savitsky-Golay filter.



(a)



(b)

Figure-3. Normal probability plot (a) processed data with mean P-value = 0.817 (b) raw data with P-value = 0.47.

4. RESULTS AND DISCUSSIONS

Data validation

Raw data might contain unnecessary spectrums, which only confuse the classification model. Tested data should be normally distributed to determine the variation of the data from one another [16]. Graphically, the normal probability plot used to illustrate each data feature variation, the data should form approximately a straight line if they are normal. The boundary of this line is statistically referred to as a significant level or commonly known as the P-value. The P-value ranges from 0 (no chance) to 1 (absolute certainty), e.g. 0.5 means 50 %. Figure-3 graphically elucidate the normal distribution of processed and raw data respectively. It is worthy to mention that, the processed data lines are more closely bounded comparing to raw data probability plot. Numerically the P-value was significantly improved from 0.47 to 0.817.

Model validation

SVM is commonly validated using a technique referred to as n-fold cross-validation. The technique divides the training set to n-1 number of subsets and leave-one-out set is left for validation [17]. The subsets are used to train the SVM model to be validated it using the leave-one-out set. This procedure is repeated n time. Eventually, the error of each n model is averaged to determine the validation of the model. The SVM model showed a 98.33 % accuracy for 10-fold cross validation. Whereas the model indicates 86.67 % accuracy or unseen data.

Additionally precision and recall decides how much the model evaluates an unseen data. Precision is the fraction of predicted data to the relevant data. Recall is the fraction of relevant data that are successfully predicted. The precision and recall of the model trained and tested listed in Table-1. The tested model evaluation is listed in Table-2.

Table-1. Trained model precision and recall evaluation.

	Precision	Recall
Beef	**1**	**0.95**
Chicken	0.952	1
Lard	1	1
Average	0.984	0.983

Table-2. Tested model precision and recall evaluation.

	Precision	Recall
Beef	**0.963**	**0.65**
Chicken	0.731	0.95
Lard	0.976	1
Average	0.984	0.983

Confusion matrix [k-class X k-class] adds an extra validation step to the model. The matrix projects true data against predicted. In fact, the built model is tested with unseen data, these data are mapped into the model and the results projected based on the prediction. The matrix 1 illustrated the 10-fold confusion matrix for the trained data where the $(a, b \text{ and } c)$ row is the tested set and according to the model prediction the $(a, b \text{ and } c)$ column lays the number of predicted set to its respective class. In other words, in an ideal case the confusion matric is a diagonal matrix. Conversely matrix 2, shows the unseen data prediction.

$$\begin{bmatrix} a & b & c & \leftarrow & Classified\ as \\ 38 & 2 & 0 & a = & beef \\ 1 & 40 & 0 & b = & chicken \\ 0 & 0 & 40 & c = & lard \end{bmatrix} \quad (1)$$

$$\begin{bmatrix} a & b & c & \leftarrow & Classified\ as \\ 26 & 14 & 0 & a = & beef \\ 1 & 38 & 1 & b = & chicken \\ 0 & 0 & 40 & c = & lard \end{bmatrix} \quad (2)$$

5. DISCUSSIONS

The prepared samples of beef, chicken and lard are grouped according to their rendering temperature. Despite the temperature changes all group portrayed a similar trending spectrum with the same number and peaks location. From Figure-2(d) it is evident that, there are three major peaks and they are located at $\approx 1450, 1750$ and $1950\ cm^{-1}$ respectively. From the chemical point of view, the change in absorbance intensity indicates the gradual weakening of the chemical bond due to the temperature changes [18]. The bond responsible for those peaks are O-H stretch, N-H stretch and C=O carbonyl group respectively. With the C=O bond and O-H bond presented we can conculde that an unsaturated fatty acids composition precent in the samples. Therefore, the sample are chemically verified as fats/oils.

The collected raw data exhibits a high curve variant from on another. The high variant is visually projected in the normal propaboity distribution plot Figure-3b with a less significant P value $= 0.817$. This means that the raw data is statisitally non-significant and not normally distributed. On the other hand, the processed data showed a successful transformtion of the data variance. Thus all the spetrums were normalized and standarized. This step demonstrates its effectiveness when a new batch of spectrums follows the same processing method is successfully tested for 86.67 % prediction rate.

The selected features for the classifier were the spectrual absorbance level of each temperature from each group (beef, chicken and lard). The close visual inspection, Figure-4, suggestes that peaks are in fact not overlapping instead they shift along both axes. For instance, the chicken and beef overlap from 0 to -0.4 yet the peaks maxima (labeled in circle) of each group differ in their wavelength locations. This close demonestration helps in assessing the intended training for the classification model.

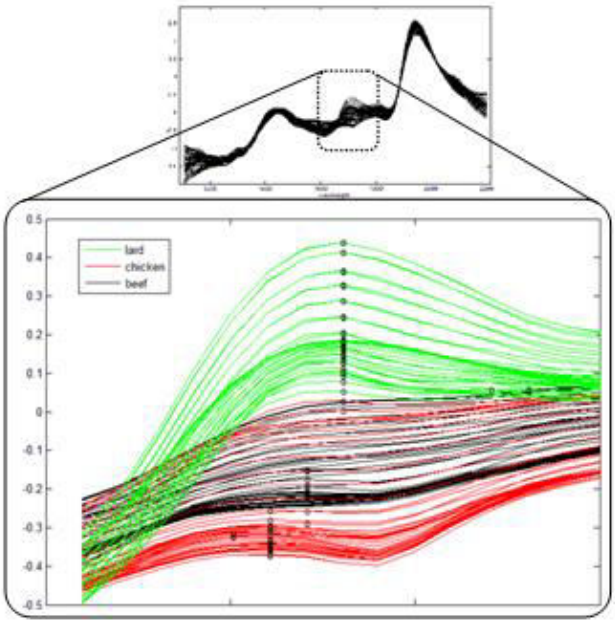


Figure-4. Peak number 2 close analysis.

Model validation underwent two stages namely self-validation and unseen validation. The self-validation is performed using 10-fold cross validation, where the results were satisfactory. As expected the cross validation misclassified two spectrums. However, the model accurately classified the other 118 spectrums with 98.33 % predication rate. More importantly, the model was able to predict 86.67 % of unseen data correctly. The majority of the beef and chicken samples were in fact correctly classified. The 12 unseen sample (4 temperature: 50 60 80

100 per fat-group) are originally extracted from three meats. The 120 spectrums (ten scans per unseen samples) were fed to the model. The model predicted 26 instance correctly and 14 instance misclassified. Primarily, all the lard samples were classified correctly.

6. CONCLUSIONS

This study attempts to examine the feasibility of distinguishing various fats from beef, chicken and lard by a portable, micro near infrared spectrometer, JDSU. The modified absorption values of different samples of lard, chicken and cow fat are fed to the SVM classifier for recognition. The classifier was expected to perform the task of recognition unseen fat in two different scenarios. In the first scenario, the classifier was supposed to recognize lard from other fats (chicken and beef both), which was able to do it with no error. In the second scenario, the classifier was supposed to recognize lard, chicken and beef fat based on the provided training data. An accuracy of 86.67% was obtained. Interestingly, the misclassified samples were beef and chicken samples only. Lard samples could be recognized flawlessly. Future study could be an attempt to focus on data processing and feature extraction to provide less confusing data for the applied classifiers.

ACKNOWLEDGEMENT

We thank Irwan Saputra for preparing the samples and Dr. G. Witjaksono from MIMOS for providing valuable suggestions and assistance in the experimental set-up.

REFERENCES

[1] B. Liebmann, A. Friedl and K. Varmuza, "Applicability of near-infrared spectroscopy for process," Biochemical Engineering Journal, pp. 187-193, 2010.

[2] A. Rohman and Y. Che Man, "Analysis of pig derivatives for halal authentication studies," Food Review international, pp. 97-112, 2012.

[3] [3] R. Akmeliawati, N. Asyikeen and I. Jaswir, "Design and Development of Alcohol and Lard Detector in," International Seminar on Science and Technology Innovation, pp. 170-173, 2-4 Oct 2012.

[4] M. Nurjuliana, Y. Che Man and M. Hashim, "Analysis of Lard's Aroma by an Electronic Nose for Rapid Halal Authentication," J Am Oil Chem Soc, pp. 75-82, 2011.

[5] Y. Che Man, A. M. Marina, A. Rohman, H. A. Al-Kahtani and O. Norazura, "A Fourier Transform Infrared Spectroscopy Method for Analysis of Palm Oil Adulterated with Lard in Pre-Fried French fries," International Journal of Food Properties, pp. 354-362, 2013.

[6] C. Man and M. Mirghani, "Detection of Lard Mixed with Body Fats," Journal of the American Oil Chemists' Society, pp. 753-761, 2001.

[7] J. Holler, D. Skoog and D. West, Fundamentals of analytical chemistry, Philadelphia: Saunders College Pub, 1996.

[8] B. Ustun, W. Melssen and L. Buydens, "Facilitating the application of Support Vector Regression by using a universal Pearson VII function based kernel," Chemometrics and Intelligent Laboratory Systems, pp. 29-40, 2006.

[9] A. d. Weijer and C. Lucasius, "Curve fitting using natural computation," Analytical Chemistry, pp. 23-31, 1994.

[10] A. Khorshidtalab, M. J. E. Salami and M. Hamedi, "Evaluating the effectiveness of time-domain features for motor imagery movements using SVM," Computer and Communication Engineering (ICCCE), pp. 909-913, 2012.

[11] R. Brereton and G. Lloyd, "Support Vector Machines for classification and regression," Analyst, pp. 230-267, 2010.

[12] C. Heyes, J. Wang, L. Sanii and M. A. El-Sayed, "Fourier Transform Infrared Study of the Effect of Different Cations on Bacteriorhodopsin Protein Thermal Stability," Biophysical Journal, pp. 1598-1606, 2002.

[13] A. Yuliansyah and T. Hirajima, "Efficacy of Hydrothermal Treatment for Production of Solid Fuel from Oil Palm Wastes," in Resource Management for Sustainable Agriculture, Japan, CC BY 3.0 license, 2012, pp. 1-20.

[14] W. Reusch, "Infrared Spectroscopy". Virtual Text of Organic Chemistry, 2007.

[15] D. Alomar, R. Fuchslocher, J. Cuevas, R. Mardones and E. Cuevas, "Prediction of the Composition of Fresh Pastures by Near Infrared Reflectance or Interactance-Reflectance Spectroscopy," Chilean journal of agricultural research, pp. 198-206, 2009.

[16] D. Montgomery and G. Runger, Applied statistics and probability for engineers, 2007.

[17] G. Cawley and N. Talbot, "Fast exact leave-one-out cross-validation of sparse least-squares support vector machines," Neural Networks, pp. 1467-1475, 2004.

[18] R. Ryu, I. Noda and Y. Jung, "Relationship between Infrared Peak Maximum Position and Molecular Interactions," Bull. Korean Chem. Soc, 2011.