UNIVERSITI TEKNOLOGI MARA

CHEMOMETRICS METHOD FOR CLASSIFICATION OF COOKING OILS FROM FT-MIR AND FT-NIR SPECTRA

MAS EZATUL NADIA BINTI MOHD RUAH

Thesis submitted in fulfilment
of the requirements for the degree of
Master of Science

Faculty of Applied Sciences

November 2017

ABSTRACT

Traditional analyses for cooking oil authentication which are based on chemical and physical methods have several drawbacks such as slow result outcome, necessity for pre-treatments, the needs for highly skilled personnel to handle instrument and test samples destruction. A combination of Fourier Transform Mid Infrared (FT-MIR) and Near Infrared (FT-NIR) spectroscopy with chemometrics has proven to be a successful analytical method for variety of food products. These techniques particularly may possess certain advantages such as rapid measurement, moderate instrument cost and relative ease of sample preparation. The aims of this study are to classify cooking oil into two batches of group which are Batch 1 (palm oil and non-palm oil) and Batch 2 (unused palm oil and used palm oil). The samples were analyzed by FT-MIR and FT-NIR spectroscopy and processed using classification methods; linear discriminant analysis (LDA), learning vector quantization (LVQ) and support vector machines (SVM), quadratic discriminant analysis (QDA), and euclidean distance centroid (EDC). The classification model was built using FT-MIR and FT-NIR cooking oil spectra datasets in absorbance mode in the range of 650 to 4000 cm^{-1} and 4000 - 14000 cm^{-1}, respectively. Savitzky Golay derivative was applied before developing the classification model. Then, the data was separated into two sets (training set and test set) by using Duplex method and 100 split random methods. The best variable selection method known as *t*-statistic was applied to the datasets in order to find the most significant variable. Column standardisation is the best data pre-processing method for both spectroscopy methods. The evaluation of data pre-processing was evaluated by using modified silhouette width (mSW). Then, it was followed with finding the value of percentage correctly classified (%CC) for every model in order to show the performance of developed classification models. The combination of FT-MIR and FT-NIR spectroscopy with chemometrics method showed the ability of classifying the sample into the interest groups of sample which are palm oil, non-palm oil, used palm oil and unused palm oil with using 2 principal components (PC's).

ACKNOWLEDGMENT

First and foremost, I would like to express my sincere gratitude to my supervisor Dr. Mohd Zuli Jaafar for his continuous support for my Msc. study and research as well as his patience, motivation, enthusiasm, and immense knowledge. His guidance has helped me during the time of researching and writing of this thesis. I could not imagine having a better advisor and mentor for my Msc. study.

Beside my supervisor, I would like to thank the rest of my co-supervisor, Dr. Sim Siong Fong from Universiti Malaysia Sarawak (UNIMAS) for her encouragement and insightful comments and guidance in understanding the precious MATLAB coding.

A special mention on the support given by the collaborator, Pn Zalhan Md Yusoff and her team from Photonics Laboratory, MIMOS, Technology Park Malaysia (TPM), Bukit Jalil, Kuala Lumpur to complete part of this thesis. The gratitude was giving to my research mate, for her productive idea's and discussions. Also, many thanks to my friends in Universiti Teknologi MARA, Kuala Pilah, for their endless support as well as special acknowledgement to all Laboratory Assistants for their kindness in helping me to complete my lab work.

Finally, I would like to thank my family, my parents En. Mohd Ruah Mohd Dali and , for their love and supports. Not forgetting the monetary support from Kementerian Pelajaran Tinggi and Majlis Amanah Rakyat.

TABLE OF CONTENTS

CHAPTER ONE
INTRODUCTION

1.1 BACKGROUND OF STUDY

1.1.1 Chemometrics Method

Chemometrics, or multivariate data analysis, is the science which applies optimal mathematical and statistical methods to process data. Chemometrics includes the design of experiments upstream and the analysis of data to get valuable information after measurements have been taken. The need for chemometrics tools mainly comes from the development of analytical instruments providing large amounts of increasingly complex data (Van Den Berg et al., 2013) .

The discipline of chemometrics originates in chemistry. Typical applications of chemometric methods are the development of quantitative structure activity relationships or the evaluation of analytical–chemical data. A Chemometric method in analytics is the discipline that uses mathematical and statistical methods to obtain relevant information on material systems.

Recently, the combination of chemometrics methods with FT-MIR and FT-NIR spectroscopy analysis for the evaluation of the quality of cooking oil is become increasingly used by researcher. For example, Fourier Transform Mid Infrared Spectroscopy (FT-MIR) combined with multivariate data analysis such as Principal Component Analysis (PCA), Fisher Discriminate Analysis (FDA), Linear Discriminate Analysis (LDA), Support Vector Machine (SVM), k-Nearest Neighbour (k-NN) and partial least squares have all been successfully used to monitor the quality parameters or authenticity of cooking oil (Gori et al., 2012). Looking on spectra alone cannot be really reliable because there would be some hidden information that cannot be seen by naked eye. In general, chemometrics have two types of methods which are regression and classification. Multivariate regression methods in chemometrics are capable in predicting the chemical of interest by using data from chemical instruments.

Another method is classification method which is function to classify the unknown (Duda et al., 2001).