Discrimination of vegetable oil types using Fourier transforms near infrared spectroscopy coupled with pattern recognition techniques

IOP Conf. Series: Earth and Environmental Science **301** (2019) 012067 doi:10.1088/1755-1315/301/1/012067

Discrimination of vegetable oil types using Fourier transforms near infrared spectroscopy coupled with pattern recognition techniques

R Lapcharoensuk[1, 2], A Malithong[1], D Thappho[1] and P Phonpho[1]

[1]Department of Agricultural Engineering, Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand
[2]Post Harvest Innovation Research and Development Laboratory, Department of Agricultural Engineering, Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand

E-mail: ravipat.la@kmitl.ac.th

Abstract. The aim of this research was to investigate the potential of near infrared spectroscopy (NIRs) coupled with pattern recognition techniques for discriminating of vegetable oil types (i.e. coconut oil, olive oil, rice bran oil, sesame oil, soybean oil and sun flower oil). Principle component analysis (PCA) was performed for clustering vegetable oil types. Five of supervised pattern recognition techniques such as soft independent modelling of class analogies (SIMCA), Partial least squares-discriminant analysis (PLS-DA), k-nearest neighbor (k-NN), support vector machine (SVM) and artificial neural network (ANN) were used to identify vegetable oil types. The PCA model could separate coconut oil from other vegetable oils. Two PLS-DA and SVM models showed 100% of precision, recall F-measure and accuracy for all vegetable oil whilst remainder techniques achieved a satisfactory classified performance. All supervised models could discriminate coconut oil from other oils with precision, recall F-Measure and accuracy of 100%. It seems that NIRs technique coupled with pattern recognition techniques is possible for discriminating vegetable oil types.

1. Introduction

Vegetable oils are obtained from seeds or beans. The important chemical composition of vegetable oils are triglycerides (95-98%) and the remainders (2-5%) are complex mixtures chemical compounds [1, 2]. Vegetable oils are widely used for food cooking, food frying either in household, restaurants or in the food industry. In Thailand, popular vegetable oils are coconut oil, olive oil, rice bran oil, sesame oil, soybean oil and sun flower oil. Nowadays, there is high demand for consumption of high-quality vegetable oils. The adulteration in high quality and expensive vegetable oils with cheap or inferior quality oils may quite prevalent. Thus, rapid and simple method for discriminating vegetable oil types is requested by the market.

Several methods such as high-performance liquid chromatography (HPLC) and gas chromatography–mass spectrometry (GC-MS) have been used for authentication of agricultural materials. However, they are expensive, long time-preforming and required skilled chemical analysts. The perfect analysis technique, which is non-destructive, rapid, reproducible and low-cost, has been required for discriminating vegetable oil types. One of interesting technique is near infrared spectroscopy (NIRs). This technique has been applied for discriminating authentic and adulterated product in several researches [3-9]. However, we cannot simply visually examine on NIR spectra to identify vegetable oil types because NIR absorbance bands are often wide and lack detailed structure required for examination. Typically, Chemometric method has been coupled with NIR spectra for qualitative and quantitative analysis.

Pattern recognition is an important sub-branch in Chemometrics which purpose is to classify objects into classes according to the type of object. This technique is closely correlated to artificial intelligence and machine learning. Supervised pattern recognition is more popularly used than the unsupervised for grouping the objects [10]. In recent years, NIR spectroscopy combined with pattern recognition such as principle component analysis (PCA), partial least squares discriminant analysis (PLS-DA), soft independent modelling of class analogy (SIMCA), support vector machine (SVM), k-nearest neighbour (k-NN) and artificial neural networks (ANN) showed successful results in cultivar classification [11, 12], origin discrimination [13, 14] , adulteration detection [15, 16] and so on.

The aim of this research was to investigate the potential of near infrared spectroscopy coupled with pattern recognition techniques for discrimination of popular vegetable oil types in Thailand. The outcome of this research will be an information which can represent the possibility of NIR spectroscopy modelling for detecting adulterated inferior quality oils in high quality vegetable oil in the future.

2. Methodology

2.1. Sample preparation

Six types of vegetable oil were collected from local supermarkets including coconut oil, olive oil, rice bran oil, sesame oil, soybean oil and sunflower oil. Each vegetable oil type contains 3 different brands of oil product. Immediately the samples were delivered to the laboratory and kept at room temperature (25 °C) for 60 minutes before NIR scanning.

2.2. NIR Spectra collection

NIR spectra was recorded on the vegetable oil sample contained in the vial (20 mm diameter and 43 mm height). The sample was pushed with aluminium transflectance probe (path length 0.35 mm). A Fourier transform (FT) -NIR spectrometer (MPA, Bruker Ltd, Germany) was used to collect spectra in the range 12,500-4,000 cm^{-1} at a resolution of 8 cm^{-1}. Each sample was obtained from averaging across 32 scans. In this research, 180 spectra were obtained from scanning 10 spectra per each brand of vegetable oil.

2.3. Precision tests

The precision of the experiment was reported on repeatability and reproducibility. Repeatability is the variation of measurement under the same conditions. In this work, repeatability was tested with 10 replicates of scanning on the soybean oil sample. The average and standard deviation (SD) of NIR absorbance value at 5800 cm^{-1} (absorbance peaks of glycerides [5]) was calculated for reporting the repeatability. Reproducibility evaluates as in an experiment can be reproduced in its completeness. Reproducibility was analyzed as scanning on 10 difference part of each sample. The average and SD values of NIR absorbance at the same wavenumber were determined to explain reproducibility.

2.4. Discrimination models

Principal component analysis (PCA) is an unsupervised pattern recognition technique which was used for clustering type of vegetable oil on 180 spectra. Meanwhile supervised discrimination models were developed from five pattern recognition techniques (i.e. soft independent modelling of class analogies (SIMCA), partial least squares-discriminant analysis (PLS-DA), k-nearest neighbour (k-NN), support vector machine (SVM) and artificial neural network (ANN). The Unscrambler v10.1 software package (CAMO AS, Trondheim, Norway) was used for performing PCA, SIMCA and PLS-DA models. Remainder methods (i.e. k-NN, SVM and ANN) were analysed by open source flamework of the RapidMiner Studio (version 9.1, Education Edition). For supervised learning algorithms, about 80% of spectra (144 spectra) were used to develop discrimination models. After model development, performance of the models was tested by classification of the remainder vegetable oil samples (36 samples). Calibration and test sets were defined by a random method on each vegetable oil brand by a ratio of 8:2 (8 samples were calibration set and remanded samples were test set). Potential of discrimination models were decided in term of precision, recall (sensitivity), f-measure and accuracy which are defined as:

$$\text{precision}=\frac{TP}{TP+FP} \tag{1}$$

$$\text{recall}=\frac{TP}{TP+FN} \tag{2}$$

$$\text{f-measure}=\frac{2\times\text{precision}\times\text{recall}}{\text{precision}+\text{recall}} \tag{3}$$

$$\text{accuracy}=\frac{TP+TN}{TP+TN+FP+FN} \tag{4}$$

where TP, FN, TN and FP are a number of true positives, false negatives, true negatives and false positives, respectively. The TP explain a result where the model correctly identifies the self-type of vegetable oil. The TF is the correct identification result of other type of vegetable oil. On the other hand, FP is an outcome where the model incorrectly predicts the self-type of vegetable oil. And a FN is an incorrect identification result of other type of vegetable oil. Recall is the proportion of actual positives that are correctly identified. The F-measure is total efficiency, which measures with precision and recall value.

3. Results and discussion

3.1. Spectra analysis

Raw and second derivative NIR spectra of all vegetable oils are shown in Figure 1. The obvious absorbance peaks obtain on wavenumber of 8250, 7100, 5800, 5670 and 4650 cm^{-1} (1212, 1408, 1724, 1764 and 2150 nm). The three peaks around 8250, 5800 and 5670 cm^{-1} are the first and second overtone of CH stretching of CH, CH_2, CH_3, and CH=CH of glycerides [5, 17]. Glyceride molecules compose from fatty acid, which is a main chemical composition of vegetable oils [5]. All above information indicated that possibility discrimination of vegetable oil types using NIR spectroscopy technique. Another permanent peak obtained at 7100 and 4650 cm^{-1} are a vibration band of O-H stretching first overtone of ROH and 2Xamide I + amide III of $CONH_2$ respectively.

3.2. Precision tests

The results of experiment precision tests are shown in the Table 1. Precision tests were explained with repeatability and reproducibility at the absorbance values of 5800 cm^{-1}. Repeatability of FT-NIR spectrometer was 0.71±0.002 whilst Reproducibility of all vegetable oil were between 0.68±0.03 and 0.73±0.03. These results indicate that measurement of vegetable oils under same and different condition was excellent precision.

3.3. Discrimination models

The Figure 2 presents the PCA score plot of PC1 and PC2 from NIR spectra of six vegetable oils. The score plot explained variation in NIR spectra of vegetable oil samples of 58% and 35% for PC1 and PC2 respectively. The PCA technique could separate coconut oil from other vegetable oils. However, other kinds could not clearly separate with the PCA score plot of PC1 and PC2 and some overlap was obtained between these groups. The loading plots of PC1 and PC2 from PCA model shows in the Figure 3. Both loading plots of PC1 and PC2 also observed the permanent peak in 8250, 7100, 5800, and 4300 cm^{-1} (1212, 1408, 1724 and 2325 mm). Three observed absorbance bonds are CH stretching first and second overtone of glycerides [5, 17].

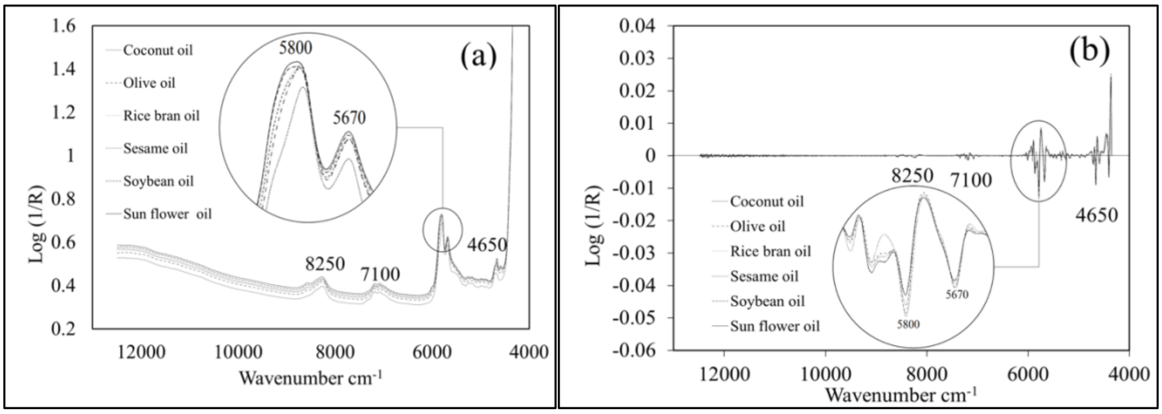

Figure 1. Near infrared spectra of six vegetable oils. (a) raw spectra; (b) second derivative spectra.

Table 1. Results of precision tests.

Precision tests	Repeatability	Reproducibility					
	Soy bean oil	Coconut oil	Olive oil	Rice bean oil	Sesame oil	Soy bean oil	Sun flower oil
Absorbance values at 5800 cm^{-1}	0.71±0.002	0.68±0.03	0.72±0.03	0.73±0.02	0.72±0.03	0.72±0.02	0.73±0.03

The results of discrimination of vegetable oil types using supervised pattern recognition models present in the Table 2. The discrimination results obtained from PLS-DA (number of factors were 10 factors) and SVM modelling presented the precision, recall F-Measure and accuracy of 100% for all vegetable oil. The results achieved in this research indicate that NIR spectra coupled with PLS-DA and SVM techniques could classify six types of vegetable oil. All supervised models could separate coconut oil from other oil which these models showed precision, recall F-Measure and accuracy of 100%. This finding point shows that NIR spectra combined with supervised pattern recognition models (i.e. PLS-DA, SIMCA, SVM, k-NN and ANN) could apply to separate coconut oil from olive oil, rice bean oil sesame oil soybean oil and sunflower oil. Accuracy of SIMCA discrimination models were between 63.9% and 100%. The k-NN models showed accuracy of 100.0%, 100.0%, 97.2, 91.7, 86.1 and 97.2% for coconut oil, olive oil, rice bean oil, sesame oil, soy bean oil and sun flower oil respectively. The ANN models showed high performance of discrimination with accuracy from 94.4% to 100%.

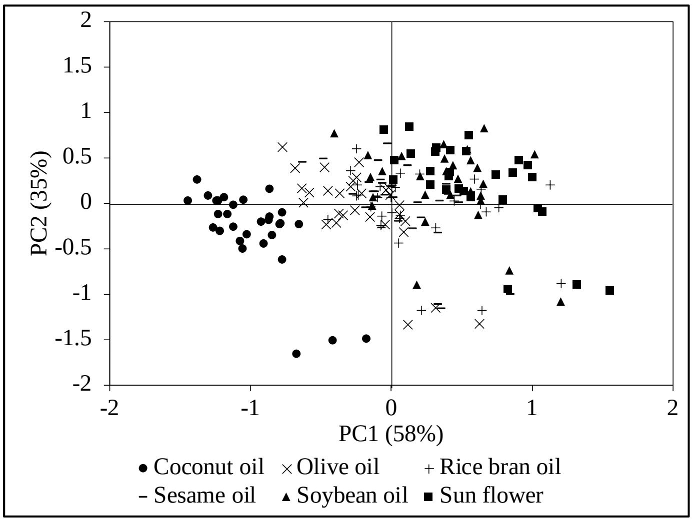

Figure 2. PCA score plot of PC1 and PC2 from NIR spectra of six vegetable oils.

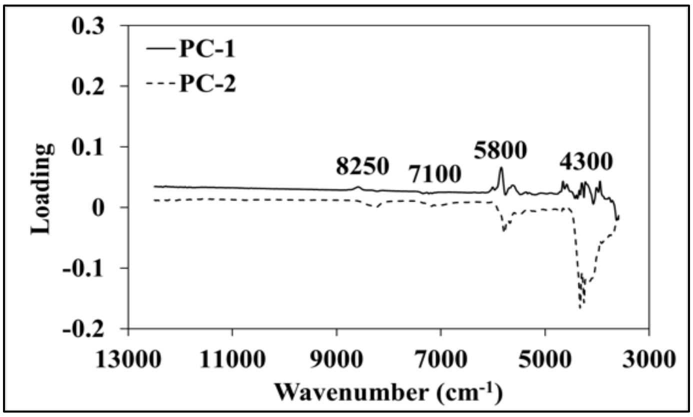

Figure 3. loading plots of PC1 and PC2 from PCA model.

Table 2. the results of discrimination of vegetable oil types using supervised pattern recognition model.

Models	Confusion matrix results				Discrimination results			
	True positive	True negative	False positive	False negative	Precision	Recall	F-Measure	Accuracy
PLS-DA (factor 10)								
Coconut oil	6	30	0	0	100.0	100.0	100.0	100.0
Olive oil	6	30	0	0	100.0	100.0	100.0	100.0
Rice bean oil	6	30	0	0	100.0	100.0	100.0	100.0
Sesame oil	6	30	0	0	100.0	100.0	100.0	100.0
Soy bean oil	6	30	0	0	100.0	100.0	100.0	100.0
Sun flower oil	6	30	0	0	100.0	100.0	100.0	100.0
SIMCA								
Coconut oil	6	30	0	0	100.0	100.0	100.0	100.0
Olive oil	6	28	0	2	100.0	75.0	85.7	94.4
Rice bean oil	6	22	0	8	100.0	42.9	60.0	77.8
Sesame oil	6	17	0	13	100.0	31.6	48.0	63.9
Soy bean oil	4	14	2	16	66.7	20.0	30.8	50.0
Sun flower oil	6	25	0	5	100.0	54.5	70.6	86.1
SVM								
Coconut oil	6	30	0	0	100.0	100.0	100.0	100.0
Olive oil	6	30	0	0	100.0	100.0	100.0	100.0
Rice bean oil	6	30	0	0	100.0	100.0	100.0	100.0
Sesame oil	6	30	0	0	100.0	100.0	100.0	100.0
Soy bean oil	6	30	0	0	100.0	100.0	100.0	100.0
Sun flower oil	6	30	0	0	100.0	100.0	100.0	100.0
k-NN								
Coconut oil	6	30	0	0	100.0	100.0	100.0	100.0
Olive oil	6	30	0	0	100.0	100.0	100.0	100.0
Rice bean oil	5	30	1	0	83.3	100.0	90.9	97.2
Sesame oil	3	30	3	0	50.0	100.0	66.7	91.7
Soy bean oil	1	30	5	0	16.7	100.0	28.6	86.1
Sun flower oil	5	30	1	0	83.3	100.0	90.9	97.2
ANN								
Coconut oil	6	30	0	0	100.0	100.0	100.0	100.0
Olive oil	5	30	1	0	83.3	100.0	90.9	97.2
Rice bean oil	4	30	2	0	66.7	100.0	80.0	94.4
Sesame oil	6	30	0	0	100.0	100.0	100.0	100.0
Soy bean oil	6	30	0	0	100.0	100.0	100.0	100.0
Sun flower oil	6	30	0	0	100.0	100.0	100.0	100.0

4. Conclusion

The near infrared spectroscopy coupled with pattern recognition techniques is proposed to discriminate the vegetable oil types. In this research, unsupervised model created by PCA could separate coconut oil from other vegetable oils. Two supervised models performed with PLS-DA and SVM algorithms showed 100% of precision, recall, F-measure and accuracy for all vegetable oil. The remainder techniques (i.e. SIMCA, k-NN and ANN) achieved a satisfactory classified performance. All supervised models could discriminate coconut oil from other oils with precision, recall F-Measure and accuracy of 100%. It seems that NIRs technique coupled with pattern recognition techniques is possible for discriminating vegetable oil types. This information is propitious that the NIRs technique may be used to detect adulteration of high quality and expensive vegetable oil with cheap or inferior quality oils in the next research.

References

[1] Kongbonga Y G M, Ghalila H, Onana M B, Majdi Y, Lakhdar Z B, Mezlini H and Ghalila S S 2011 *Food Sci. Nutr.* **2** 692-699
[2] Sikorska E, Romaniuk A, Khmelinskii I V, Herance R, Bourdelande J L, Sikorski M and Koziol J 2004 *J. Fluoresc.* **14(1)** 25-35
[3] Jiménez-Carvelo A M, Lozano V A and Olivieri A C 2019 *Food Control* **96** 22-28
[4] Chen H, Tan C, Lin Z and Li H 2019 *Spectrochim. Acta. A Mol. Biomol. Spectrosc.* **211** 280-286
[5] Chen H, Lin Z and Tan C 2018 *Vib. Spectrosc.* **99** 178-183.
[6] Karunathilaka S R Yakes B J He K, Chung J K and Mossoba M 2018 *Heliyon*, **4**, Article e00806
[7] Lapcharoensuk R. and Nakawajana N 2018 *J. Innov. Opt. Health Sci.* **11(02)** 1750019
[8] Basri K N, Hussain M N, Bakar J, Sharif Z,. Khir M F A and Zoolfakar A S 2017 *Spectrochim. Acta. A Mol. Biomol. Spectrosc.* **173** 335-342
[9] Ma H L, Wang J W, Chen Y J, Cheng J L and Lai Z T 2017 *Food Chem.* **215** 108-115
[10] Tang G, Tian K, Song X, Xiong Y and Min S 2014 *Spectrochim. Acta. A Mol. Biomol. Spectrosc.* **121** 678-684
[11] Eisenstecken D, Stürz B, Robatscher P, Lozano L, Zanella A and Oberhuber M 2019 *Postharvest Biol. Tec.* **147**, 123-131.
[12] Fuentes S, Montes E H, Escalona J M, Bota J, Viejo C G, Poblete-Echeverría C, Tongson E and Medrano H 2018 *Comput. Electron. Agric.* **151** 311-318
[13] Jintao X, Yongli S, Liming Y, Quanwei Y, Chunyan L, Xingyi C and Yun J 2018 *Spectrochim. Acta. A Mol. Biomol. Spectrosc*. **205**, 419-427.
[14] Yang Y, Wu Y, Li W, Liu X, Zheng J, Zhang W and Chen Y 2018 *Spectrochim. Acta. A Mol. Biomol. Spectrosc.* **191** 233-240
[15] Lobato K B S, Alamar P D, Caramês E T S and Pallone J A L 2018 *J. Food Eng.* **224** 105-111
[16] Quelal-Vásconez M A, Pérez-Esteve E, Arnau-Bonachera A, Barat J M and Talens P 2018 *Food Control* **92** 183-189
[17] Osborne B G and Fearn T 1993 *Theory of near infrared spectrophotometry. In Near Infrared Spectroscopy in Food Analysis* (London: Longman Science & Technical) pp 13-35

Acknowledgement

The authors gratefully acknowledge the instrument and software support provided by the NIRS Research Center for Agricultural Product and Food, Department of Agricultural Engineering, Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand.