DEVELOPMENT OF RAPID AND ACCURATE METHOD TO CLASSIFY MALAYSIAN HONEY SAMPLES USING UV AND COLOUR IMAGE

Abd Alazeez Almaleeh^{1*}, Abdul Hamid Adom^{1,2}, Ammar Zakaria^{1,2}

¹School of Mechatronic Engineering, University Malaysia Perlis, Kampus Pauh Putra 02600 Arau, Perlis, Malaysia
²Centre of Excellence for Advanced Sensor Technology, University Malaysia Perlis, Kampus Pauh Putra 02600 Arau, Perlis, Malaysia

(Received: May 2016 / Revised: February 2017 / Accepted: April 2017)

ABSTRACT

The purpose of this paper is to classification of three main types of Malaysian honey (Acacia, Kelulut and Tualang) according to their botanical origin using UV–Vis Spectroscopy and digital camera. This paper presented the classification of the honey based on two characteristics from three (3) types of local honey, namely the antioxidant contents and colour variations. The former uses the UV spectroscopy of selected wavelength range, and the latter using RGB digital camera. Principal Component Analysis (PCA) was used for both methods to reduce the dimension of extracted data. The Support Vector Machine (SVM) was used for the classification of honey. The assessment was done separately for each of the methods, and also on the fusion of both data after features extraction and association. This paper shows that classification based on the fusion method improved significantly compared to single modality Honey classification based on the fusion method was able to achieve 94% accuracy. Hence, the proposed methods have the ability to provide accurate and rapid classification of honey industry and further improve the quality assessment and provide traceability.

Keywords: Data fusion; Honey classification; Sensors; Support Vector Machine

1. INTRODUCTION

Malaysia is a tropical country, rich in natural forest resources such as herbs, medicinal plants, spices and honey. These traditional foods are one of the main sources of income for the Malaysian agricultural industry. This opens up more demand and the need for better and more to improve those products as" accurate" classification. Moreover, the products such as honey in Malaysia are gaining popularity, days by days. Generally, consumers classify honey in terms of its geographical distribution and different floral sources by smell, taste and their color. As there are different types of honey available in Malaysia, the identification of the botanical origin is important to classify it accurately. There are few existing methods to classify honey. Sensory evaluation is one of methods that currently being used; however, sensory evaluation is often inconsistent and varies from person to person. Chemical test methods in laboratory more accurate than the sensory evaluation method (Serrano et al., 2004). However, they are time consuming and costly. In addition to that, this method does not guarantee the safety of consumers and hinders further development of apiculture.

^{*}Corresponding author's email: abdoo_i1986@yahoo.com, Tel. +604-988-5166, Fax. +604-988-5167 Permalink/DOI: https://doi.org/10.14716/ijtech.v8i3.4265

Recently there are researchers have suggested using artificial sensors, that make use of an electronic nose and an electronic tongue to discriminate between types of bee honey (Zakaria et al., 2011). along with the application of physicochemical criteria such as: enzymatic activity, sugar content, pH, water content, electrical conductivity and so on to analyze different types of Uni-floral honey (Conti et al., 2007). Furthermore, identification of pollen counts for authentication has been used, although there are difficulties to provide guarantee a precise assignment of the origin (Ramsay, 2005; Corbella & Cozzolino, 2008).

There are also some researchers, who have used colour classifications for commercial grading purposes. One example uses the Pfund colour scale. The Pfund colorimeter is a simple tool which is used to compare the colour of the honey sample with a standard coloured glass (Bertoncelj et al., 2007). The reference unit is on the Pfund scale (0 to 140), starts with very light-coloured honey and rising up to the darkest honey. However, this method requires very large amounts of sample and depends on the person performing the analysis since different observers lead to different measurements. However, these drawbacks are still remained in these researches. Thus, it is essential to find out a rapid and accurate way of honey discrimination. Recently, UV-Visible spectroscopy has received wide range of attention of the researchers as it is suitable for non-destructive analysis of biological and biomedical materials. For example, the UV-Vis spectroscopy can be used for discriminating tea beverages (Chen et al., 2008), milk (Balabin & Smirnov, 2011), coffee (Lv & Yang, 2011), and other materials (Sinelli et al., 2010; Yang et al., 2011; Yang et al., 2012). Some researchers used this technique such as (Zhu et al., 2010, Gallardo-Velázquez et al., 2009) to classify adulterants in some local origins of honey. Although these researches achieved high accuracy for honey discrimination, the calibration models were developed using full range of wavelengths, which brought about high complexity in computation and cause difficulty in practical applications.

This thesis presents two methods for the classification of honey in terms of Botanical origin. Based on UV-Vis spectroscopy and Digital Camera, the difference in antioxidants and their hue colour were measured and later classified by the developed SVM classifier.

2. MATERIALS AND METHODS

2.1. Sample Selection

In this investigation, three different types of honey were selected from various brands. The honey samples were selected based on the most popular and highly demanded. Altogether, a total of fifteen brands of honey of different origins were used for this research. The three different types of honey consisted of several different brands and are as follows: Three (3) different brands of Acacia honey, Three (3) different brands of Kelulut honey, and Nine (9) different brands of Tualang honey. All honey samples (Acacia, Kelulut and Tualang) were obtained from local sources and producers, and the condition of honey samples were checked visually before performing the quantification of the different antioxidant properties. All samples were kept at room temperature under dry condition prior to analyses. All samples were labelled by its type as follows: A: Acacia, K: Kelulut and T: Tualang.

2.2. UV–Vis Spectroscopy

Eighty-one samples from three different botanical sources of honey (Acacia, Kelulut and Tualang) were purchased locally. The samples as stated on the product label from different botanical origins and sealed at room temperature of $24-26^{\circ}$ C. The number of samples that have been collected: three Acacia, three Kelulut and nine Tualang samples. Before spectral measurement, all samples were placed in a water container at $35-40^{\circ}$ C until all the soluble substances fully dissolved (Ulloa et al., 2013).

Spectral scanning was conducted using a Lambda 35 UV-Vis Spectroscopy. The reference record scans were taken per sample shown in the Table 1. Kelulut and Acasia honey samples were scanned three times and all data were used for the analysis. As for Tualang honey, the samples used were the average value of three scans performed. This is to ensure even distribution of data size for statistical analysis. The antioxidant in the honey was studied using 2,2-diphenyl-1-picrylhydrazyl (DPPH) as reported by (Estevinho et al., 2008). An aliquot of two drops around 3.00 mL of test honey sample with 0.3mL DPPH and distilled deionized water 200 mL was placed in a 3 cm quartz cell (Cuvette UV–Visible Spectroscopy). Each mixture was shaken vigorously and left to stand in the dark until a stable absorption value and distilled water was used as a blank sample. A total of 110 data points were recorded with an average separation in the range of 200–310 nm wavelengths.

Honey	No. of Samples	Scan for each sample	Every four months	Total Scan
Acacia	3	3	3	27
Kelulut	3	3	3	27
Tualang*	9	3	3	27
Total	15			81

Table 1 Scanning distribute every four months during the year

*Average scanned value were used for the analysis

2.3. UV Scanning

Figure 1 shows the absorbance spectra of the fifteen samples of honey with wavelengths ranging from 200 nm to 310 nm. Two types of antioxidants have been observed in the scanning range. The first antioxidant named "Naringenin" found in honey shows maximum absorption at 228 nm (Dezmirean et al., 2010, Wybranowski et al., 2013). The second type of antioxidant named "Pinocembrin" is only found in honey, which shows maximum absorption at 288 nm (Greenaway et al., 1991; Markham, 1982). The two peaks which are almost similar values were obtained for Tualang and Kelulut at wavelength of 228 nm. However, Acacia possesses different characteristics at this point. It can also be found that almost similar values are obtained for Tualang and Acacia at the wavelength of 288 nm. However, Kelulut honey samples possess different characteristics at this point. Low absorbance is obtained at 250 nm and above 300 nm. Almost all over the spectra similarities in the shape of spectral absorbance have been obtained except at few wavelengths. Thus, it is necessary to apply a multi-variable analysis using appropriate ways as mentioned earlier to build calibration models to discriminate honey quality.



Wavelength / nm

Figure 1 Absorbance spectra of honey samples

2.4. RGB Digital Camera

In this investigation, the honey samples are differentiated by the colour or hue components in a 360°- Hue saturation value. The term "value" here refers to the pixel count of each hue of the image.

Test honey samples were placed and secured in Sterile Glass Petri Dishes (D: 100 mm \times H: 15 mm). As well as all the honey samples were stored at room temperature for a day to ensure there is no air bubbles presence in the test samples. Honey samples with the different colour were captured by the camera is shown in Figure 2.

2.5. Apparatus and Software

The digital images were obtained using a Phillips RGB Digital Camera. These digital images were then extracted and processed to obtain the corresponding colour histogram with the properties of (Hue saturation value vs Pixel Count). Chemometric data treatment was implemented in MATLAB software (GUI Mathworks).



Figure 2 Digital images obtained from three different types of honey having distinctive colour

2.6. Digital Image Acquisition

In order to dissolve sugar crystals, the temperature of 10 grams honey sample was raised up to 35–40°C (Hamdan, 2010). The images can only be taken when all air bubbles were removed. The honey sample was placed in a mini Glass Petri dish. The images were obtained using a Digital camera. The procedure was performed in triplicate for each honey sample. The digital camera was placed in a fixed position in the centre of a circular daylight fluorescent lamp of 22W having a temperature colour of 6400K, over the honey sample. The illumination and the camera distance from the sample were kept fixed throughout the experiments. In order to shield the samples from external light, the whole system was placed in a sealed box. This is done to avoid unwanted noise in terms of stray light. Additionally, the internal walls of the box were covered with black paper in order to avoid light scattering as shown in Figure 3.



Figure 3 Honey image capturing device without the external light shield

2.7. Converting RGB to Hue Saturation Value of the Image

The digital camera captured the images of the honey colour and save the image in the value of main "RGB" colours. The Hue saturation value of the Image can be derived from the RGB values to the range from 0 to 1. This process can be implemented by dividing the RGB value by 255 for 8-bit colour depth. In this case, the 8-bit colour depth is adequate.

$$R = value \ of \ Red \ / \ 255 \tag{1}$$

$$G = value \ of \ Green \ / \ 255 \tag{2}$$

$$B = value \ of \ Blue \ / \ 255 \tag{3}$$

The minimum and maximum values of R, G and B are as follows:

$$Min = min (R, G, B) \tag{4}$$

$$Max = max (R, G, B) \tag{5}$$

The Hue formula is depending on the max value in RGB colour channel (Jeon, 2013).

$$Hue = \begin{cases} 0 , if max = Min \\ 60^{\circ} \times \left[\frac{G-B}{Max-Min}\right] , if max = R \\ 60^{\circ} \times \left[\frac{B-R}{Max-Min} + 2\right] , if max = G \\ 60^{\circ} \times \left[\frac{R-G}{Max-Min} + 4\right] , if max = B \end{cases}$$
(6)

.

Saturation calculation:

$$S = \begin{cases} 0 & , max = 0\\ \frac{Max - Min}{Max} & , max \neq 0 \end{cases}$$
(7)

Value calculation:

$$V = max \tag{8}$$

2.8. Histograms

The histogram is a representation of the intensity levels of colours with respect to the number of pixels in a digital image. A histogram plot is a bar graph projection, whereby the X-axis represents the tonal scale of Rainbow colours (Hue), and Y-axis represents the count of pixels in an image in a certain area of the tonal scale.

One of the issues to be considered for the selection of the sections of the image to be used for the classification. The sections of the image selected must be not affected by ambient light and not contain dark areas. According to Yimyam et al. (2005), area in the middle of the image was considered to avoid the noise that may be introduced by the ambient light, or other sources of light near the subject. In this research, three regions of circular shape with diameters of 10 mm were considered at random, away from the edge of the image as shown in Figure 4. The initial analyses conducted on the selected regions were able to give consistent results.

Another issue to be considered in the data collection is the deterioration of the product with time, and whether this will affect the data collected. To enable this aspect of the product to be analysed, the data of the honey samples at different ages of shelf storage were recorded. The honey samples aged zero months (just purchased), four and eight months after purchased were recorded. This will enable the changes in colour with age, if applicable; will be considered in the classification. The data collected with the different shelf age is shown in Table 2.

The image of the Kelulut and Acasia honey samples were recorded three times and all data were used for the analysis. As for Tualang honey, the image samples used were the average value of triplicate images. This is to ensure even distribution of data size for statistical analysis. In this work, the histograms were obtained using Matlab software. Each colour component is composed of 256 tones, which are used as analytical information.



Figure 4 The three circular colour regions were selected in the centre of the image

Honey	No. of Samples	Image for each sample	Every four months	Total Scan
Acacia	3	3	3	27
Kelulut	3	3	3	27
Tualang*	9	3	3	27
Total				81

Table 2 Image sampling distributes every four months during the year

*Average of Hue value were used for the analysis

2.9. Image Sampling

The histograms of one selected sample of each three different type of honey are presented in Figure 5. Acacia Honey has a high Pixel Count at 15° Hue saturation value and the sample is noticeably darker in colour as shown at a 190° Hue saturation value. Kelulut Honey has a high Pixel Count at 50° Hue saturation value and the sample is noticeably darker in colour, as shown at a 220° Hue saturation value. Tualang Honey has a noticeably darker colour at 350° and 110° Hue saturation values, respectively.



Figure 5 The histogram scans for the three types of honey

3. STATISTICAL ANALYSIS

The Principal Component Analysis (PCA) and Support Vector Machine (SVM), the measured data will be subjected to a suitable data pre-processing. The classification would be the classification and prediction. The steps were performed with the SPSS (IBM- SPSS Statistic) and (SVM Toolbox) running under Matlab software. The SPSS was used to extract the Principal Components (PCs), and SVM MATLAB software was used for classification using written functions. All 81 sample datasets were analyzed prior in performing PCA on the entire spectra. Later, the SVM was used to analyse the PCs data to find the accuracy of testing matrix, training matrix and confusion matrix.

3.1. Separate Analysis using PCA and SVM

Each modality (Image and UV data) was processed separately. Prior to PCA Plot Plot (Dillon and Goldstein, 1984), a number of adequate PCs were determined, which are 3 PCs components to be used in the analysis for UV spectroscopy and Digital Image. The amount of percentage variances (%) of the first three principal components for each method are shown in Table 3. In UV-Vis spectrums, the amount of accumulated variance in the first three principal components accounted for more than 97%. This suggests that only the first three PCs should be considered or adequate enough for further analysis. As for the digital images, the amounts of accumulated variance in the first three principal components have accounted for more than 71%. After subjected to PCA, the three PCs were selected for classification. SVM classifier used and the results are shown also in the Table 3. SVM results for the digital camera 70.8% accuracy, while the SVM from UV-Vis 90.7% accuracy.

PCs	Digital Camera			UV-Vis Spectroscopy		
1 0.5	Variance	Cumulative	SVM	Variance	Cumulative	SVM
PC1	42.217 %	42.217 %		53.793 %	53.793 %	
PC2	19.997 %	62.214 %	70.8 %	29.224 %	83.017 %	90.7 %
PC3	9.411 %	71.625 %		14.179 %	97.196 %	



Figure 6. PCA plot of 81 samples of honey from three types using UV-Vis spectroscopy

The PCA of UV-Vis is shown in Figure 6. 2D plot of PC1 with PC2 is shown in Figure 6a, and 2D plot of PC2 with PC3 is shown in Figure 6b.

The (Hue vs Pixel) Count of the honey from 81 different samples was measured using image processing and projected using PCA plot. The SVM accuracy gives 70.8% correct classification. The PCA of the images shown in Figure 7. The 2D plot of PC1 with PC2 is shown in Figure 7a, and 2D plot of PC2 with PC3 is shown in Figure 7b.



Figure 7 PCA plot of 81 samples of honey from three types using Hue image

3.2. Data Fusion

The intermediate level fusion has been used in this research to evaluate the concept of fusion whereby the combined information from different modality should perhaps produce better results (Subari et al., 2012).

Group			P	Predicted Group Membership			
			1	2	3	_	
			Т	17	1	1	19
		Count	А	1	19	2	22
	Tasiains		Κ	1	2	19	$\begin{array}{cccccccccccccccccccccccccccccccccccc$
	Training		Т	89.5	5.3	5.3	100.0
		%	А	4.5	86.4	9.1	100.0
			Κ	4.5	9.1	86.4	100.0
			Т	17	1	1	19
т		Count	А	1	19	2	22
	Training-validated		Κ	1	2	19	22
			Т	89.5	5.3	5.3	100.0
		%	А	4.5	86.4	9.1	100.0
			Κ	4.5	9.1	86.4	100.0
		Count	Т	7	1	0	8
			А	0	5	0	5
т	Testina		Κ	0	0	5	5
	resting		Т	87.5	12.5	.0	100.0
		%	А	.0	100.0	.0	100.0
			Κ	.0	.0	100.0	100.0

Table 4 Classification results on fused UV and Digital Image using SVM for three different honey types

PCA technique was performed before and after fusion to evaluate the grouping behaviour. In total, there are 81 data collected from all honey samples for fusion experiments. This is to evaluate whether the fusion method is able to classify with better performance compared to single method.

SVM on the fusion method shows a better performance. The SVM model was tested and 94.4% of original grouped cases correctly classified and 92.9% of cross-validated group cases correctly classified. The classification results are shown in Table 4. The SVM of the fusion method was expected to give better classification, which outperforms the single method classification performance.

From Table 4 The SVM model was tested and 94.4% of original grouped cases correctly classified and 92.9% of training-validated grouped cases correctly classified. The SVM classification was performed as expected.

4. CONCLUSION

This work has successfully demonstrated two different methods for classification of Malaysian honey based on their botanical origins, namely Acacia, Kelulut and Tualang. Also, the fusion of the two methods has been conducted to further improve the classification accuracy. UV and digital images was further reduced before subjected to SVM Classifier. The data fusion of both the UV and digital camera have resulted in improved classification accuracy of up to 94%.

One aspect of the work that needs careful consideration is the sample preparation whereby we need to ensure that both measurements are subjected to the same sample under the same conditions and then this fusion method can be implemented.

The two methods introduced in this work have been demonstrated to be able to provide rapid classification of different honey samples by their botanical origins. It has also been demonstrated that the fusion of the two methods yield a better accuracy compared to single measurement. This means that if this approach can be industrialized for other materials such as vegetable oils, juices and coffee, the honey industry can be more cost effective, safer and competitive.

5. **REFERENCES**

- Balabin, R.M., Smirnov, S.V., 2011. Melamine Detection by Mid-and Near-infrared (MIR/NIR) Spectroscopy: A Quick and Sensitive Method for Dairy Products Analysis Including Liquid Milk, Infant Formula, and Milk Powder. *Talanta*, Volume 85, pp. 562–568
- Bertoncelj, J., Doberšek, U., Jamnik, M., Golob, T., 2007. Evaluation of the Phenolic Content, Antioxidant Activity and Colour of Slovenian Honey. *Food Chemistry*, Volume 105, pp. 822–828
- Chen, Q., Zhao, J., Liu, M., Cai, J., 2008. Nondestructive Identification of Tea (Camellia Sinensis L.) Varieties using FT-NIR Spectroscopy and Pattern Recognition. *Czech J Food Sci*, Volume 26, pp. 360–367
- Conti, M.E., Stripeikis, J., Campanella, L., Cucina, D., Tudino, M.B., 2007. Characterization of Italian Honeys (Marche Region) on the Basis of their Mineral Content and Some Typical Quality Parameters. *Chemistry Central Journal*, Volume 1, pp. 1–10
- Corbella, E., Cozzolino, D., 2008. Combining Multivariate Analysis and Pollen Count to Classify Honey Samples Accordingly to Different Botanical Origins. *Chilean Journal of Agricultural Research*, Volume 68, pp 102–107

Dezmirean, D., Stanciu, O.G., Mircea, T., 2010. Spectrophotometric Evaluation of Flavonoid Content of Propolis and Poplar Buds by ZrOCl₂ Reagent. *Bulletin of University of Agricultural Sciences and Veterinary Medicine Cluj-Napoca. Animal Science and Biotechnologies*, Volume 67(1-2), pp. 85–90

Dillon, W.R., Goldstein, M., 1984. Multivariate Analysis: Methods and Applications

- Estevinho, L., Pereira, A.P., Moreira, L., Dias, L.G., Pereira, E., 2008. Antioxidant and Antimicrobial Effects of Phenolic Compounds Extracts of Northeast Portugal Honey. *Food and Chemical Toxicology*, Volume 46, pp. 3774–3779
- Gallardo-Velázquez, T., Osorio-Revilla, G., Zuñiga-De Loa, M., Rivera-Espinoza, Y., 2009. Application of FTIR-HATR Spectroscopy and Multivariate Analysis to the Quantification of Adulterants in Mexican Honeys. *Food Research International*, Volume 42, pp. 313–318
- Greenaway, W., May, J., Scaysbrook, T., Whatley, F., 1991. Identification by Gas Chromatography-mass Spectrometry of 150 Compounds in Propolis. *Zeitschrift für Naturforschung C.*, Volume 46, pp. 111–121
- Hamdan, K., 2010. Crystallization of Honey. Bee World, Volume 87, pp. 71-74
- Jeon, G., 2013. Measuring and Comparison of Edge Detectors in Color Spaces. *International Journal of Control and Automation*, Volume 6, pp. 21–30
- Lv, G., Yang, H., 2011. Discrimination of Different Brands of Nescafé Coffee using Vis-NIR Spectroscopy and Comparative Study. Advances in Biochemical Engineering – Proceedings of 2011 International Conference on Agricultural and Biosystems Engineering (ICABE). Volume 1. pp 163–166
- Markham, K.R., 1982. Techniques of Flavonoid Identification. Academic press London
- Ramsay, H., 2005. How Bees Make Honey. Evans Brothers
- Serrano, S., Villarejo, M., Espejo, R., Jodral, M., 2004. Chemical and Physical Parameters of Andalusian Honey: Classification of Citrus and Eucalyptus Honeys by Discriminant Analysis. *Food Chemistry*, Volume 87, pp. 619–625
- Sinelli, N., Cerretani, L., Di Egidio, V., Bendini, A., CasiraghI, E., 2010. Application of Near (NIR) Infrared and Mid (MIR) Infrared Spectroscopy as a Rapid Tool to Classify Extra Virgin Olive Oil on the Basis of Fruity Attribute Intensity. *Food Research International*, Volume 43, pp. 369–375
- Subari, N., Mohamad Saleh, J., Md Shakaff, A.Y., Zakaria, A., 2012. A Hybrid Sensing Approach for Pure and Adulterated Honey Classification. *Sensors*, Volume 12, pp. 14022–14040
- Ulloa, P.A., Guerra, R., Cavaco, A.M., Da Costa, A.M.R., Figueira, A.C., Brigas, A.F., 2013. Determination of the Botanical Origin of Honey by Sensor Fusion of Impedance E-tongue and Optical Spectroscopy. *Computers and Electronics in Agriculture*, Volume 94, pp. 1–11
- Wybranowski, T., Ziomkowska, B., Kruszewski, S., 2013. Antioxidant Properties of Flavonoids and Honeys Studied by Optical Spectroscopy Methods. *Medical and Biological Sciences*, Volume 27, pp. 53–58
- Yang, H., Kuang, B., Mouazen, A., 2012. Quantitative Analysis of Soil Nitrogen and Carbon at a Farm Scale using Visible and Near Infrared Spectroscopy Coupled with Wavelength Reduction. *European Journal of Soil Science*, Volume 63, pp. 410–420
- Yang, H., Kuang, B., Mouazen, A.M., 2011. In Situ Determination of Growing Stages and Harvest Time of Tomato (Lycopersicon esculentum) Fruits using Fiber-optic Visible-nearinfrared (Vis-NIR) Spectroscopy. *Applied Spectroscopy*, Volume 65, pp. 931–938
- Yimyam, P., Chalidabhongse, T., Sirisomboon, P., Boonmung, S., 2005. Physical Properties Analysis of Mango using Computer Vision. *In:* Proceeding of ICCAS. Citeseer
- Zakaria, A., Shakaff, A.Y.M., Masnan, M.J., Ahmad, M.N., Adom, A.H., Jaafar, M.N., Ghani, S.A., Abdullah, A.H., Aziz, A.H.A., Kamarudin, L.M., 2011. A Biomimetic Sensor for the

Classification of Honeys of Different Floral Origin and the Detection of Adulteration. *Sensors*, Volume 11, pp. 7799–7822

Zhu, X., Li, S., Shan, Y., Zhang, Z., Li, G., Su, D., Liu, F., 2010. Detection of Adulterants such as Sweeteners Materials in Honey using Near-infrared Spectroscopy and Chemometrics. *Journal of Food Engineering*, Volume 101, pp. 92–97