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

ABSTRACT

The purpose of this paper is to perform the 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 of the honey types, 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. 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 of the fusion method improved significantly compared to single modality. Honey classification based on fusion method give 94%. Hence, the proposed methods have the ability to provide accurate and rapid classification of honey products in terms of its origin. This work presents a system which can be more applied in Malaysia honey industry and further improve the quality assessment and provide tracebility.

*Keywords***:** Honey classification, Sensors, Support Vector Machine, Data Fusion

1. INTRODUCTION

Malaysia is a rainforest country which is bless with natural forest resources such as medicinal plants and spices, exotics fruits and abandunce of wild harvest honey. These traditional foods are one of the main sources of income for the Malaysian agricultural industry. It 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 as a result to the healthy lifestyle promotions by various groups including the government. Generally, consumers classify honey in terms of its geographical distribution and different floral sources, smell, taste and their colour. As there 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 method that currently being used, unfortunately it is not easy to set up and subjective experience is often biased and inconsistent. Chemical test methods in laboratory (Serrano, Villarejo, Espejo, & Jodral, 2004), is more accurate than the sensory evolution method. However, it is quite time consuming and costly. In addition to that this method does not guarantee the safety of ordinary consumers and hinders further development of apiculture.

This researchers have suggested to use sensory systems that make use of electronic nose and electronic tongue to discriminate 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, Stripeikis, Campanella, Cucina, & Tudino, 2007). Furthermore,

identification of pollen counts for authentication has been used, although there are difficult to guarantee a precise assignment and identification of the origin (Corbella & Cozzolino, 2008; Ramsay, 2005).

There are also some researchers used colour classifications for commercialization grading. One example such as using the Pfund colour scale. The Pfund colourimeter is a simple tool which is used to compare the colour of the honey sample with a standard coloured glass (Bertoncelj, Doberšek, Jamnik, & Golob, 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 reference sample and depends heavily 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, visible and UV-Vis 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, Zhao, Liu, & Cai, 2008), milk (Balabin & Smirnov, 2011), coffee (Lv & Yang, 2011), and other materials (Sinelli, Cerretani, Di Egidio, Bendini, & Casiraghi, 2010; H Yang, Kuang, & Mouazen, 2012; Haiqing Yang, Kuang, & Mouazen, 2011). Some researchers used this technique such as Zhu et al. (Zhu et al., 2010) and Gallardo-Vel´azquez et al. (Gallardo-Velázquez, Osorio-Revilla, Zuñiga-de Loa, & Rivera-Espinoza, 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 work 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. Fifteen samples honey of different brands and origin were used. In total, 81 samples were studied. All samples were duplicated for UV-Vis and Digital Camera measurement. Different brands of honey samples are as follows: three different brands of Acacia honey, three different brands of Kelulut honey and nine different brands of Tualang honey. All honey samples (Acacia, Kelulut and Tualang) were obtained from local sources and producers, and quality of the samples were checked 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 as follows: A: Acacia, K: Kelulut and T: Tualang.

2.2. UV–Vis spectroscopy

Eighty one samples of 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◦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 40◦C until all the soluble substances fully dissolved.

Spectral scanning was conducted using a Lambda 35 UV-Vis Spectroscopy, Perkin Elmer. There reference record scans were taken per sample during the full year in each four months to observe any change in antioxidant, 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 measurement involve the use of DPPH as reported by (Estevinho, Pereira, Moreira, Dias, & Pereira, 2008). An aliquot two drop 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 101 data points were recorded with an average separation in the range of 200–310 nm wavelengths.

Table 1 Scanning distribute every four months during the year

	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

* 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, STANCIU, & Mircea, 2010; Wybranowski, Ziomkowska, & Kruszewski, 2013). The second type of antioxidant named "Pinocembrin" is only found in honey, which shows maximum absorption at 288 nm (Anti-Bacterial & Anti-Viral; Greenaway, May, Scaysbrook, & Whatley, 1991; Markham, 1982).

The two peaks which are almost similar values were obtained for Tualang and Kelulut at wavelength of 228 nm. However Acacia possess 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.

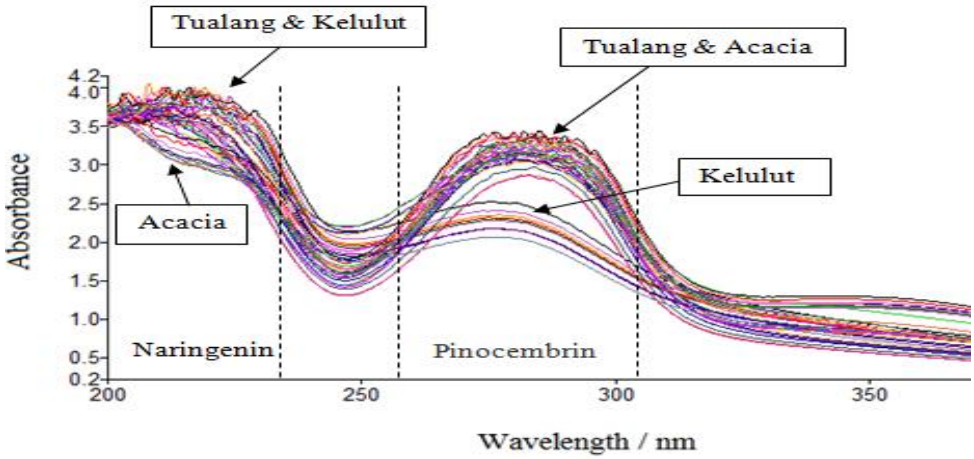


Figure 1 Absorbance spectra of honey samples.

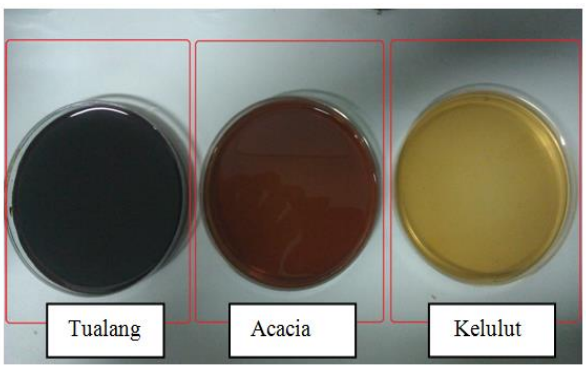
2.4. RGB Digital Camera

In this investigation, the honey samples are differentiated by the colour or hue components in a 360o-hue image. 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 x H: 15mm). 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 vs Pixel Count). Chemometric data treatment was implemented in Matlab® 6.5 software (GUI Mathworks).



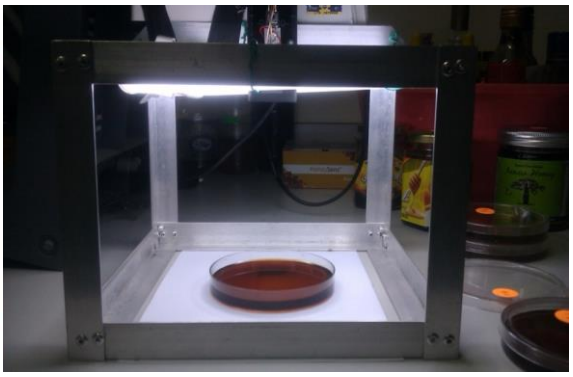


Figure 2 Digital images obtained from three different types of honey having distinctive colour.

Figure 3 Honey image capturing device without the external light shield.

2.6. Digital Image Acquisition

In order to dissolve sugar crystals, the temperature of 10 grams honey sample was raised up to 40°C. 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 card box. This is done to avoid un-wanted noise in terms of stray light. Additionally the internal walls of the box were covered with black coloured and coarse paper in order to avoid light scattering and reflection. Digital images obtained from honey samples are presented in Figure 3.

2.7. Converting RGB to Hue of Image

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

$$R = \text{value of Red} \;/\; 255 \quad (1)$$

$$G = \text{value of Green} \;/\; 255 \quad (2)$$

$$B = \text{value of Blue} \;/\; 255 \quad (3)$$

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

$$\text{Min} = 0.09 \text{ (the R value)} \quad (4)$$

$$\text{Max} = 0.46 \text{ (the B value)} \quad (5)$$

The Hue formula is depending on the max value in RGB colour channel. The three different formulas are.

If Red is max, then $\text{Hue} = (G-B)/(\text{max}-\text{min})$ (6)

If Green is max, then $\text{Hue} = 2.0 + (B-R)/(\text{max}-\text{min})$ (7)

If Blue is max, then $\text{Hue} = 4.0 + (R-G)/(\text{max}-\text{min})$ (8)

The Hue value needs to be multiplied by 60 in order to convert it to degree on the colour circle chart. If Hue becomes negative, it needs to be added with 360 to it, as a full circle has 360 degrees.

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 colour (Hue Value), and Y-axis represents the count of pixels in an image based on the tonal scale.

To analyse the images, three regions circular with diameter 10 mm was selected of every snap image, as shown in Figure 4. The histogram for each colour regions circular Hue was obtained using the selected regions and takes them the arithmetic average.
There reference record Images were taken per sample during the full year in each four months to avoid any change in colour, shown in the Table 2. Kelulut and Acasia honey samples were imaged three times and all data were used for the analysis. As for Tualang honey, the samples were relying on the average value of triplicate images. This is to ensure even distribution of data size for statistical analysis. In this work the histograms were calculated using Matlab software. Each colour component is composed of 256 tones, which can be used as analytical information.

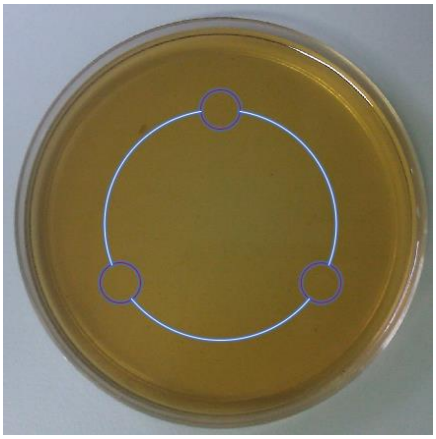


Figure 4 The three circular regions selected in the centre of the image.

Table 2 Image sampling distributes every four months during the year

	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

* Average of Hue value were used for the analysis

2.9. Image Sampling

The histograms of one selected sample from each three different type of honey are presented in Figure 5. The histogram is a representation of the intensity levels of each of the colours with respect to the number of pixels in the digital image. A histogram is a bar graph, whose X-axis represents the Hue, and Y-axis represents the count of pixels in an image in a certain area of the tonal scale. In Figure 5, Acacia Honey has more pixels count at 15^{O} Hue and sample has noticeable dark in colour as shown at 190^{O} Hue, Kelulut Honey has more pixels count at 50^{O} Hue and sample has noticeable dark in colour as shown at 220^{O} Hue and Tualang Honey has more noticeable dark at 350^{O} and 110O Hue.

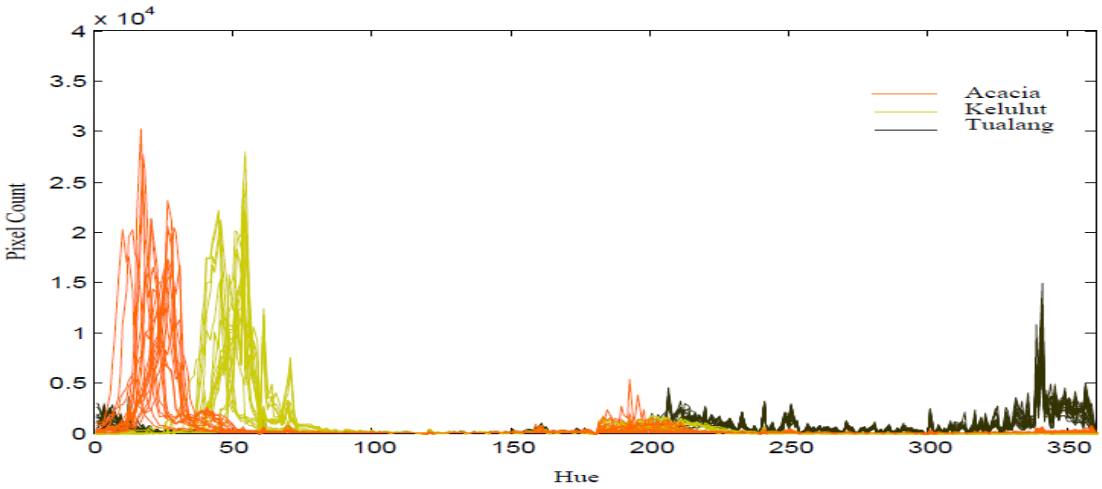


Figure 5: The histogram scans of three types honey

3. STATISTICAL ANALYSIS

The Principal Component Analysis (PCA) and Support Vector Machine (SVM) will be used after subjected to a suitable data pre-processing. The classifier will be undergo training, and classification. Both method were performed using SPSS (IBM- SPSS Statistic) and (SVM Toolbox) running under Matlab software. The SPSS was used to extract the Principal Components (PCs), and developed KM-SVM Matlab software was used for other mathematical data transformations 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, a number of adequate PCs were determined, which in this 3 PCs are found to be adequate. The amounts of percentage variance (%) of the first three principal components for each method and classification accuracy 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 variance accumulated variance in the first three principal components are accounted for more than 71%.

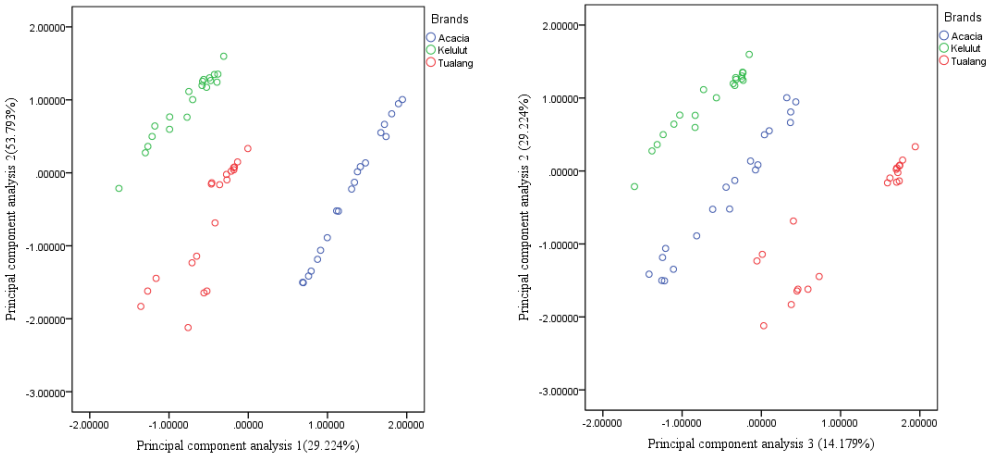
After subjected to PCA, the three PCs were selected used for classification. SVM classifier used and the results are shown also in the Table 3. SVM result from Digital camera shown 70.8% accuracy, while the SVM from UV-Vis shown the accuracy of 90.7%.

Table 3 The amount of variance (%) of the first three principal components for two different methods and the classification accuracy

	Digital Camera			UV-Vis Spectroscopy		
PCs	Variance	Cumulative	SVM	Variance	Cumulative	SVM
PC1	42.217 %	42.217 %	70.8%	53.793 %	53.793 %	90.7%
PC2	19.997 %	62.214 %		29.224 %	83.017 %	

PC3	9.411 %	71.625 %		14.179 %	97.196 %

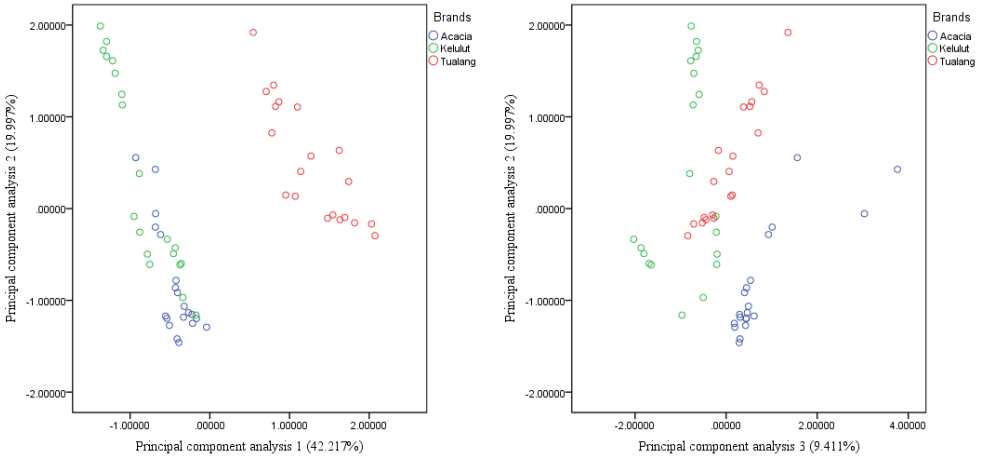
The PCA of UV-Vis is shown in Figure 6. 2D plot of PC1 with PC2 is shown in (a) and 2D plot of PC2 with PC3 is shown in (b).



(a) (b)

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

The (Hue vs pixel) 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 are shown in Figure 7. The 2D plot of PC1 with PC2 is shown in (a) and 2D plot of PC2 with PC3 is shown in (b).



(a) (b)

Figure 7 PCA plot of 81 samples of honey from three different 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, Mohamad Saleh, Md Shakaff, & Zakaria, 2012). 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 compare 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.

Table 4 classification results on fused UV and Digital Image using SVM of three different type of honey [a, c]

				Predicted Group Membership			
Group				1	2	3	Total
	Cases Selected	Count	T	17	1	1	19
			A	1	19	2	22
			K	1	2	19	22
		%	T	89.5	5.3	5.3	100.0
			A	4.5	86.4	9.1	100.0
			K	4.5	9.1	86.4	100.0
	Cross-validated[b]	Count	T	17	1	1	19
			A	1	19	2	22
			K	1	2	19	22
		%	T	89.5	5.3	5.3	100.0
			A	4.5	86.4	9.1	100.0
			K	4.5	9.1	86.4	100.0
	Cases Not Selected	Count	T	7	1	0	8
			A	0	5	0	5
			K	0	0	5	5
		%	T	87.5	12.5	.0	100.0
			A	.0	100.0	.0	100.0
			K	.0	.0	100.0	100.0

a. 92.9% of selected original grouped cases are correctly classified, and 94.4% of unselected original grouped cases are correctly classified.
b) Cross validation done only for those in the analysis. In cross validation, each case is classified by the function derived from all cases other than that case. c) 92.9% of cross-validated grouped cases correctly classified.

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 fusions of both 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 condition, 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 industrialised, the honey industry can be more cost effective, safer and competitive.

7. REFERENCES

Anti-Bacterial, Raw Honey—An, & Anti-Viral, Anti-Fungal Substance. Health Benefits.

Balabin, Roman M, & Smirnov, Sergey 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, 85*(1), 562-568.

Bertoncelj, Jasna, Doberšek, Urška, Jamnik, Mojca, & Golob, Terezija. (2007). Evaluation of the phenolic content, antioxidant activity and colour of Slovenian honey. *Food Chemistry, 105*(2), 822-828.

CheN, Quansheng, Zhao, Jiewen, Liu, Muhua, & Cai, Jianrong. (2008). Nondestructive identification of tea (Camellia sinensis L.) varieties using FT-NIR spectroscopy and pattern recognition. *Czech J Food Sci, 26*, 360-367.

Conti, Marcelo Enrique, Stripeikis, Jorge, Campanella, Luigi, Cucina, Domenico, & Tudino, Mabel Beatriz. (2007). Characterization of Italian honeys (Marche Region) on the basis of their mineral content and some typical quality parameters. *Chemistry Central Journal, 1*(14), 14.

Corbella, Eduardo, & Cozzolino, Daniel. (2008). Combining multivariate analysis and pollen count to classify honey samples accordingly to different botanical origins. *Chilean Journal of Agricultural Research, 68*(1), 102-107.

DEZMIREAN, Daniel, STANCIU, Oltica 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, 67*(1-2).

Estevinho, Letícia, Pereira, Ana Paula, Moreira, Leandro, Dias, Luís G, & Pereira, Ermelinda. (2008). Antioxidant and antimicrobial effects of phenolic compounds extracts of Northeast Portugal honey. *Food and Chemical Toxicology, 46*(12), 3774-3779.

Gallardo-Velázquez, Tzayhri, Osorio-Revilla, Guillermo, Zuñiga-de Loa, Marlene, & Rivera-Espinoza, Yadira. (2009). Application of FTIR-HATR spectroscopy and multivariate analysis to the quantification of adulterants in Mexican honeys. *Food Research International, 42*(3), 313-318.

Greenaway, W, May, J, Scaysbrook, T, & Whatley, FR. (1991). Identification by gas chromatography-mass spectrometry of 150 compounds in propolis. *Zeitschrift für Naturforschung C, 46*(1-2), 111-121.

Lv, G, & Yang, H. (2011). Discrimination of different brands of Nescafé coffee using vis-NIR spectroscopy and comparative study. *Advances in Biochemical Engineering, 1*(2), 163-166.

Markham, Kenneth R. (1982). *Techniques of flavonoid identification* (Vol. 31): Academic press London.

Ramsay, Helena. (2005). *How Bees Make Honey*: Evans Brothers.

Serrano, Salud, Villarejo, Marta, Espejo, Roberto, & Jodral, Manuela. (2004). Chemical and physical parameters of Andalusian honey: classification of Citrus and Eucalyptus honeys by discriminant analysis. *Food Chemistry, 87*(4), 619-625.

Sinelli, Nicoletta, Cerretani, Lorenzo, Di Egidio, Valentina, Bendini, Alessandra, & Casiraghi, Ernestina. (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, 43*(1), 369-375.

Subari, Norazian, Mohamad Saleh, Junita, Md Shakaff, Ali Yeon, & Zakaria, Ammar. (2012). A hybrid sensing approach for pure and adulterated honey classification. *Sensors, 12*(10), 14022-14040.

Wybranowski, Tomasz, Ziomkowska, Blanka, & Kruszewski, Stefan. (2013). Antioxidant properties of flavonoids and honeys studied by optical spectroscopy methods. *Medical and Biological Sciences, 27*(4), 53-58.

Yang, H, Kuang, B, & Mouazen, AM. (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, 63*(3), 410-420.

Yang, Haiqing, Kuang, Boyan, & Mouazen, Abdul Mounem. (2011). In Situ Determination of Growing Stages and Harvest Time of Tomato (Lycopersicon esculentum) Fruits Using Fiber-Optic Visible–Near-Infrared (Vis-NIR) Spectroscopy. *Applied spectroscopy, 65*(8), 931-938.

A biomimetic sensor for the classification of honeys of different floral origin and the detection of adulteration, No. 8, 11 7799-7822 (2011).

Zhu, Xiangrong, Li, Shuifang, Shan, Yang, Zhang, Zhuoyong, Li, Gaoyang, Su, Donglin, & Liu, Feng. (2010). Detection of adulterants such as sweeteners materials in honey using near-infrared spectroscopy and chemometrics. *Journal of Food Engineering, 101*(1), 92-97.