

# Principal Component Analysis – A Realization of Classification Success in Multi Sensor Data Fusion

Maz Jamilah Masnan, Ammar Zakaria, Ali Yeon Md. Shakaff,  
Nor Idayu Mahat, Hashibah Hamid, Norazian Subari  
and Junita Mohamad Saleh

*Universiti Malaysia Perlis, Universiti Utara Malaysia & Universiti Sains Malaysia  
Malaysia*

## 1. Introduction

The field of measurement technology in the sensors domain is rapidly changing due to the availability of statistical tools to handle many variables simultaneously. The phenomenon has led to a change in the approach of generating dataset from sensors. Nowadays, multiple sensors, or more specifically multi sensor data fusion (MSDF) are more favourable than a single sensor due to significant advantages over single source data and has better presentation of real cases. MSDF is an evolving technique related to the problem for combining data systematically from one or multiple (and possibly diverse) sensors in order to make inferences about a physical event, activity or situation. Mitchell (2007) defined MSDF as the theory, techniques, and tools which are used for combining sensor data, or data derived from sensory data into a common representational format. The definition also includes multiple measurements produced at different time instants by a single sensor as described by (Smith & Erickson, 1991).

Although the concept of MSDF was first introduced in the 1960s and implemented in the 1970s in the robotic and defense application, today, the application of MSDF has proliferated into various nonmilitary applications. However the method is still disparate, where it is impossible to create a one-fits-all data fusion framework. The applications of MSDF are now multidisciplinary in nature. Some specific applications of MSDF include multimodal biometric systems using face and palm-print (Raghavendra et al., 2011); renewable energy system (Li et al., 2010); color texture analysis (Wu et al., 2007); face and voice outdoor multi-biometric system (Vajaria et al., 2007); medical decision making (Harper, 2005); image recognition (Sun et al., 2005), road traffic accidents (Sohn et al., 2003); and personal authentication (Duc et al., 1997; Kumar et al., 2006).

MSDF technique has become as a prominent tool in food quality assessment. Quality assessment in food processing industries aims to guarantee the standard and safety control of food products. Traditional approach of exploiting trained human panels to evaluate quality parameters can be replaced by artificial sensors. An example of artificial sensor receiving great interest from researcher in these industries is the electronic nose (i.e. e-nose)

sensor that mimics the function of human smell. In the context of MSDF, usually e-nose is applied with another sensor called electronic tongue (i.e. e-tongue) which imitates the human taste function. Several applications of e-nose and e-tongue in food research include flavor sensing system (Cole et al., 2011); honey classification (Zakaria et al., 2011); classification of *orthosiphon stamineus* (Zakaria et al., 2010); detection of polluted food (Maciejak et al., 2003); discrimination of standard fruit solutions (Boilot et al., 2003); quality control of yoghurt fermentation (Cimander et al., 2002); and discrimination of several types of fruit juices (Winqvist et al., 1999).

It is believed that the application of MSDF such as the fusion of e-nose and e-tongue, may overcome some drawbacks of using trained human panels especially for on-line food production. The use of artificial sensors is capable of overcoming human exhaustion and stress, minimize between-panels variability, and obviously human panels are not suitable for online measurements. Thus, this chapter focuses on the application of Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) in MSDF. Two models of MSDF proposed by Hall (1992) namely low level data fusion and intermediate level data fusion are proposed in order to identify and classify different types of pure honey, beet sugar, cane sugar and adulterated samples (i.e. mixtures of pure honey with cane sugar and beet sugar). This chapter also aim to provide a concept to the constructive and lists some advantageousness of PCA in the application of MSDF especially in the analysis of multivariate data.

### 1.1 The fusion of artificial sensors

The appreciation of food is basically based on the combination of many human senses including sight, touch, sound, taste and smell. However, due to the expensive cost of having panels of trained expert to evaluate food quality parameters, a more rapid technique for objective measurement of food products in a consistent and cost-effective manner is highly needed in the food industry (Winqvist et al., 2003). Two human senses that are believed to be closely correlated in the perception of flavour are the sense of smell and taste. The e-nose and e-tongue have been defined as the artificial sensing systems capable of producing a digital fingerprint of a given chemical ambient (D'Amico, 2000). Both devices consist of chemical sensor arrays coupled with an appropriate pattern recognition system capable of extracting information from complex signals (Buratti et al., 2004).

Basically, an e-nose is formed by having an array of gas sensors with different selectivity, a signal collecting unit and suitable pattern recognition software, all controlled and executed by a computer. The principle of e-tongue is similar to that of the e-nose, except for the array of sensors designed for liquids (Cosio et al., 2007). The ultimate task of these sensors is to collect the digital fingerprint or signals that would be further interpreted using multivariate statistical tools before the objective of the fusion approach is attained. One of the most popular exploratory data analyses in chemical sensors is PCA (Di Natale et al., 2006). PCA is a procedure that permits to extract useful information from the data, to explore the data structure, the relationship between the objects and features, and the global correlation of the features. Further details of PCA are described in Section 2. The selected principal components based on certain criteria will be used as an input for classification procedure using linear discriminant analysis (LDA). Further descriptions of this technique are illustrated in section 3 of this chapter.

The selected architecture of MSDF in this research focuses on the approach of identity fusion. Identity fusion is a fusion of parametric data to determine the identity of an observed object. Our interest is to convert multiple sensor observations of a target attributes (such as e-nose and e-tongue responses) to a joint declaration of target identity. One of the key issues in developing an MSDF system is to determine the stage or phase in the data flow to combine or fuse the data (Hall & Llinas, 1997). In an identity fusion, Hall (1992) suggested three frameworks to be applied; (i) low level data fusion (or data level fusion); (ii) intermediate level data fusion (or feature level fusion); and (iii) high level data fusion (or decision level fusion). However, for the purpose of this discussion only data level and feature level fusion are discussed.

### 1.1.1 Low level data fusion

In low level data fusion, the e-nose and e-tongue sensors observe the target objects independently, and later the raw sensor data (i.e. original data collected from each sensor) are combined. In order to fuse raw sensor data, the original sensor data must be commensurate i.e. must be observations of similar physical quantities (Hall et al., 1997). Sometimes, the number of features recorded by the e-nose and e-tongue are different, but the raw sensor data can still be fused if both datasets are of the same sample size (equal  $n$ ). It is important to ensure the new dataset is formed from the original non-normalized data. A framework of low level data fusion is illustrated in Fig. 1.

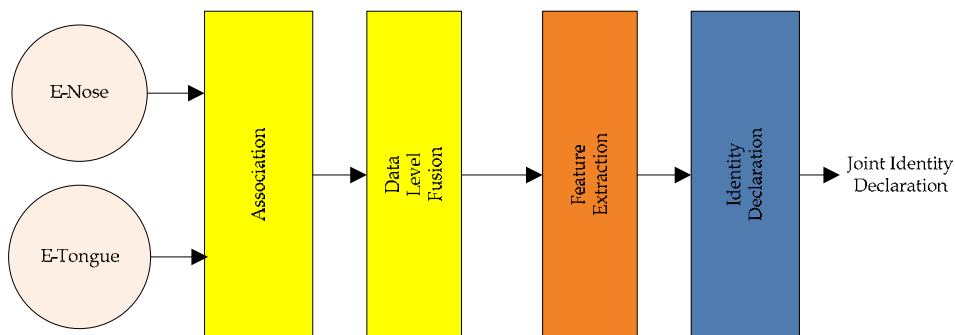


Fig. 1. Framework of low level data fusion by Hall (1992)

It is believed that the low level data fusion in identity fusion provides the most accurate result (Hall et al., 1997). This may be due to the fact that the originality information from each sensor is maintained and used in further processes. Thus, low level data fusion is potentially more accurate than the other two fusion methods. However, the difficulties in the application of low level data fusion method are due to the noise that frequently occurs in the sensor data and redundant data, which have an adverse effect on the classification results.

### 1.1.2 Intermediate level data fusion

This approach consists of extracting features from the signals of each sensor to yield feature vectors. Then, the feature vectors are fused and identity declaration is made based on the joint feature vectors. The identity declaration process includes techniques such as

knowledge-based approach that includes expert system and fuzzy logic, or training-based approach like discriminant analysis, neural network, Bayesian technique, k-nearest neighbors and centre mobile algorithms. Fig. 2 illustrates the framework of the intermediate level data fusion.

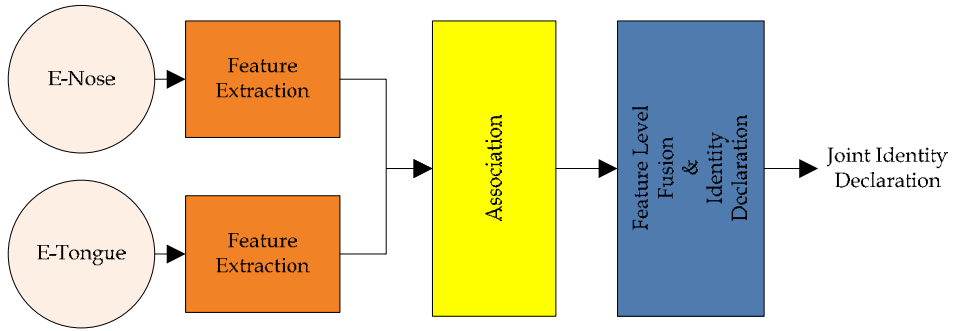


Fig. 2. Framework of intermediate level data fusion by Hall (1992)

It is important to note that both low and intermediate level data fusion apply feature extraction in transforming the raw signals provided by the sensor into a reduced vector of features describing parsimoniously the original information. Then, in the identity declaration, a quality class is assigned to the signals based on the feature extraction result.

## 2. Principal component analysis

Principal component analysis (PCA) was first described by Karl Pearson in 1901. A description of practical computing methods came much later from Harold Hotelling in 1933 (Manly, 2004). The idea of PCA is to keep the variation of the number of  $p$  original features into a fewer number of  $k$  unobservable variables ( $k \leq p$ ), which is termed as principal components, as maximum as possible. Let Table 1 below describes the original data of a sensor data set with  $n$  objects each was observed with  $p$  features.

Case	$X_1$	$X_2$	...	$X_p$
1	$X_{11}$	$X_{12}$	.	$X_{1p}$
2	$X_{21}$	$X_{22}$	.	$X_{2p}$
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
$n$	$X_{n1}$	$X_{n2}$	...	$X_{np}$

Table 1. The form of data for a principal component analysis with  $p$  features on  $n$  cases

The aim of PCA is to find a new set of variables, say  $Z_1, Z_2, \dots, Z_i$  in a form of a linear combination of  $X$ 's which is  $Z = \alpha^T X$ . Here,  $Z = (Z_1, Z_2, \dots, Z_p)$  is a vector of principal components and  $\alpha^T$  is a matrix of coefficients  $\alpha_{ij}$  for  $i, j = 1, 2, \dots, p$ .

The first principal component ( $Z_1$ ) is the linear combination of the original features which mathematically written as  $Z_1 = \alpha_{11}X_1 + \alpha_{12}X_2 + \dots + \alpha_{1p}X_p$ , assemble as the largest as possible of variance of  $p$  features subject to the condition that  $\alpha_{11}^2 + \alpha_{12}^2 + \dots + \alpha_{1p}^2 = 1$ . Then, the second principal component ( $Z_2$ ) is chosen to have the property of having the second largest possible variance of  $X_1, X_2, \dots, X_p$  while being uncorrelated with the first component ( $Z_1$ ). The remaining principal components are defined similarly, with the  $j$ th principal component having the largest possible variance given that it is uncorrelated with the  $i$ th principal component for  $i < j$ . Let  $\lambda_i$  be the variance (eigenvalues) of  $Z_i$ , and  $\alpha_{ij}$  be the eigenvectors of  $Z_i$  where  $i, j = 1, 2, \dots, p$ , then these conditions hold for the eigenvalues and eigenvectors:

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_i \geq 0 \quad (1)$$

$$\alpha_i^T \alpha_i = 1 \quad (2)$$

$$\alpha_i^T \alpha_j = 0 \quad \text{where } i \neq j \quad (3)$$

Before we proceed to discuss on the issue of reducing the dimension intended for further analysis, it is a need to understand which matrix of information should be used, either a correlation matrix or a covariance matrix to allow for a computation of principal components. One should clearly understand when to use either one of the input matrix as often the results of these two are different. The next sections 2.1 and 2.2 briefly discuss the guidelines.

## 2.1 Information matrix for principal component analysis

### 2.1.1 Principal component using covariance matrix

An implicit assumption when using covariance matrix as an input is that the features should not have grossly different variances. Such differences in variance might arise because of different scales of measurements, different magnitude of measurements, or some combination of the two factors (Krzanowski, 2000). If they do, then the first few principal components will be pulled toward those features with the larger variances (Dillon & Goldstein, 1984).

In such cases, the data should be standardized and it means the correlation matrix is used in the PCA. As a general guideline, it would seem sensible to standardize first whenever the measured features show differences in variances, or whenever the user is concerned with very different measured entities or units (Krzanowski, 2000). However, transformation on the original data would result PC scores of a different meaning (Martinez & A.R. Martinez, 2001). Obviously, the big drawback of PCA based on covariance matrix is the sensitivity of the PCs to the units of measurement used for each element of  $X$  (Jolliffe, 2002).

### 2.1.2 Principal component analysis using correlation matrix

PCA aims to create linear combination of new variables that are uncorrelated to each other, thus, if the correlation matrix portrays nearly small correlation, then there is probably not much point in carrying PCA (Chatfield & Collins, 1980). PCA calculation based on correlation matrix is suitable for features with unequal scales of measure. One way to trace unequal scales is through wide differing variances among the features. In computing a correlation coefficient between two features, differences due to the mean and the dispersion of the features are removed (Dillon & Goldstein, 1984). This is recommended as the original features are all standardized to unit variance (Borgognone et al., 2001).

Therefore, data that is used to calculate PCA for correlation input does not need any transformation as it is applied automatically in the correlation computation. However, a disadvantage in using correlation matrix to calculate the principal components are that they give coefficients for standardized variables and are therefore less easy to interpret directly. Thus, to interpret the principal components in terms of the original variables, each coefficient must be divided by the standard deviation of the corresponding variables (Jolliffe, 2002).

### 2.2 Deciding the number of components to retain

Mathematically, the choice of values for coefficients  $\alpha$  is subjected to the restrictions given in equations (2) and (3). Thus, the obtained principal components are in decreasing order of variance,  $\text{var}(Z_1) \geq \text{var}(Z_2) \geq \dots \geq \text{var}(Z_p) = \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ . In practice, only the first  $k$  numbers of principal components account for most of the variability of the original data, thus keeping all the  $p$  principal components sound impractical. This mean, only the first  $k$  principal components will be used in further analysis while the  $p-k$  principal components will be ignored. However, there is no universally accepted method to do so because the decision is largely judgemental and a matter of taste (Dillon & Goldstein, 1984). A number of procedures to determine  $k$  have been suggested. Among the most common procedures are as follows.

#### 2.2.1 Average eigenvalue

The most common criterion to determine the number of informative principal components in PCA is the Guttman-Kaiser criterion (Jackson, 1993). Principal components associated with eigenvalues ( $\lambda$ ) derived from a covariance matrix which are larger in magnitude than the average of the eigenvalues, are retained. In the case of eigenvalues derived from a correlation matrix, the average is 1.0 for the variables to retain. Therefore, any principal component associated with an eigenvalue whose magnitude is greater than or equal to 1.0 is chosen for further analysis. However, Rencher (1998) warned that this method works well in practice but when it identifies wrongly, it is likely to retain too many components. It is well known as simple and the most suitable criterion to be applied especially when confronted with numerous variables.

#### 2.2.2 Proportion of total variance explained

In a PCA model, each eigenvalue represents the level of variation of the original features explained by the associated principal components. Another popular decision criterion is

based on the proportion of the total variance explained by the principal components retained in the model. If  $k$ -components are retained, then we may represent the cumulative variance explained by the first  $k$  principal components by,

$$t_k = \frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^p \lambda_i} \quad (4)$$

Often, the researcher decides on a satisfactory value for  $t_k$  and then determines  $k$  accordingly. The obvious problem with the technique is to decide on an appropriate  $t_k$ . In practice, it is common to select from 70% to 90% (Jolliffe, 2002). Because of such obviously arbitrary, this approach has sometimes been criticized for its subjectivity (Kim & Mueller, 1978). While Jackson (1993) strongly argues against the use of this method except possibly for exploratory purposes when little are known about the population of the data.

### 2.2.3 Scree plot

Perhaps much easier decision on  $k$  can be made based on graphical approaches as suggested by Cattell (1966) called the scree plot. A scree plot is a plot of the eigenvalues versus the index of the eigenvalue. With this approach, the eigenvalues of each component are plotted in successive order of their extraction, and then an elbow in the curve is identified by applying a straightedge to the bottom portion of the eigenvalues to see where they form an approximate straight line (Dillon & Goldstein, 1984).

The value of  $k$  is given by the point at which the components curve above the straight line formed by the smaller eigenvalues. Fig. 3 shows a case in which  $k$  is equal to three and the straight line (shallow) begins at the fourth until the last component. As we can observe from Fig. 3, the third component is marked exactly at eigenvalue is equal to 1. Dillon and Goldstein (1984) argue that this method is inconclusive when there is no obvious break or there may be several breaks. And it become more troublesome when two breaks occur among the first half of the eigenvalues, since it will be difficult to decide which of the breaks reflect the correct number of components.

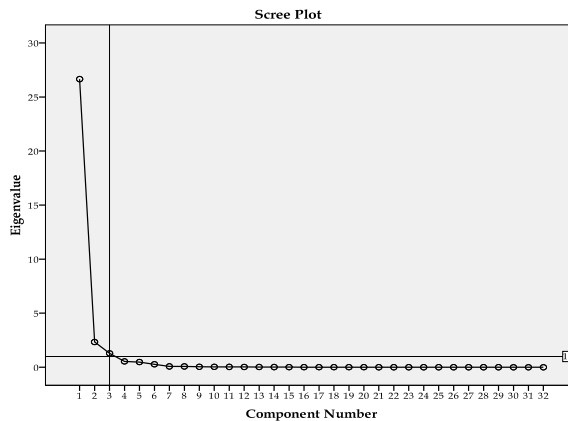


Fig. 3. Illustration of the scree plot.

### 3. Linear discriminant analysis

Linear discriminant analysis or discriminant function analysis or in short discriminant analysis is a supervised technique for classifying objects into two or more groups, given the measurements for these objects is from several features (i.e. sensor responses). It involves deriving linear combinations of the independent features that will discriminate between the *a priori* defined groups in such a way that the misclassification error are minimized (Dillon & Goldstein, 1984). The discrimination can be accomplished by maximizing the between group variance relative to the within-group variance. The basic discriminant analysis is the one that involves only two-group problem which was first suggested by R. A. Fisher (1936). In the two-group problem, the aim is to find a single linear composite of the predictor features that could discriminate between the two groups. The linear composite then acts as a new axis along which the groups were maximally separated.

In reality, we may encounter discrimination problems of more than two groups which require an extension of the basic discriminant analysis called the multiple discriminant analysis. The goal in multiple discriminant analysis is much similar with discriminant analysis for two groups. Dillon and Goldstein (1984) describe in general, with  $k$  groups and  $p$  predictor features, there are in total,  $\min(p, k-1)$  possible discriminant functions (i.e. linear composites). In most applications, since the number of features ( $p$ ) is exceeding the number of groups ( $k$ ), at most  $k-1$  discriminant functions will be considered. However, not all of these functions show statistically significant variation among the groups, and fewer than  $k-1$  discriminant functions is actually needed. Likewise in forming principal components in PCA, discriminant functions are generated so that the scores of each new discriminant function are uncorrelated with the scores of previously obtained discriminant function. Thus, each linear composite is the new single function that maximizes the ratio of the between-groups to within-groups variability, accordingly. Besides, the discriminant functions are extracted in a decreasing order of accounted variation.

There are assumptions that need to be considered by researchers for obtaining optimal procedure in the sense of producing smallest misclassification error rate. According to Dillon and Goldstein (1984), for optimality, we assume (i) multivariate normality of the  $p$  predictor features, and (ii) equal variance-covariance matrices in each of the  $k$  groups. They added that the objectives of multiple discriminant analysis are for the most part is the generalizations of those of the two-group problem. Among others it includes:

- i. To find the linear composites with as large as possible between-groups variability subject to each uncovered linear composites being uncorrelated with previously extracted composites. The accounted variations for all linear composites are in decreasing order.
- ii. To determine whether the group centroids are statistically different.
- iii. To determine the number of discriminant functions that is statistically significant.
- iv. To successfully assign new signal or observation to one of the several groups.
- v. To determine the predictor features that contributes most for discrimination among groups.

The goal in constructing classification rules is to minimize the mistakes in assigning new signals to its groups. Less mistakes means less error for the classification rules to correctly allocate the signals. In real problem, often one has a set of data to be discriminated



accordingly to  $g$  groups. However, using the same data for constructing a rule and evaluating a rule is biased. As the matter of fact, it does not mimic the real use of discrimination rule to classify a future object where the rule is constructed based on the existing data. There are some techniques that can be considered in an attempt to avoid such bias. Some of the techniques are re-substitution method, cross validation method which is also known as sample-splitting method and leave-one-out method. Lachenbruch and Mickey (1968) in (Krzanowski, 2000) proposed the leave-one-out method that was believed to be able to overcome most problems inherent in the previous two methods. The technique consists of determining the allocation rule using the sample data minus one observation and then using the subsequent rule to classify the omitted observation. Repeating this procedure by omitting each of the individuals in the two training set in turn yields, an estimate of the error rates, the proportions of misclassified signals in the two training sets.

#### 4. Materials and methods

The experiment was implemented in the Sensor Laboratory, Centre of Excellence for Advanced Sensor Technology, University Malaysia Perlis. The aim is to identify and classify different types of pure honey, beet sugar, cane sugar and adulterated samples (i.e. mixtures of pure honey with cane sugar and beet sugar) by applying the low level data fusion and intermediate level data fusion. PCA was employed to reduce the data dimension and further classification was fulfilled by LDA.

##### 4.1 Sample selection and preparation

In this experiment, 10 different brands of *Tualang* honey were purchased from the local market with three different batches of each particular honey. While for the adulteration purposes, two types of sugar solution namely beet sugar and cane sugar were imported from Germany and United Kingdom respectively. Display of pure honey and sugar are illustrated in Fig. 4 and all honey and sugar samples are summarized in Table 2.

Item	Descriptions	Group
AG	Agromas	1
AS	As-Syifa	1
ST	Syair Timur	1
T3	Tualang 3	1
TB	Tayyibah	1
TK	Tualang King	1
TLH	Tualang TLH	1
TN	Tualang Napis	1
WT	Wild Tualang	1
YB	Yubalam Bahtera	1
BS	Beet Sugar	2
CS	Cane Sugar	3
XXBS	Pure Honey + BS	4 (20%), 5 (40%)
XXCS	Pure Honey + CS	6 (20%), 7 (40%)

Table 2. Description and abbreviation of honey samples, sugar and adulterated samples used in the experiments

Based on the three different batches of each pure honey, three samples of 5ml was prepared for further measurement. For adulteration samples, each pure honey was mixed with sugar of different concentration (i.e. 20% and 40%) as shown in Table 3. Each pure sugar was also measured. Each sampling of pure honey, sugar and adulterated were repeated ten times. In total there were about 172 samples of pure honey, pure sugar and adulterated mixtures.

Percentage of Pure Honey	Descriptions
20% pure honey	1:4 (ratio of pure honey /sugar solution)
40% pure honey	2:3 (ratio of pure honey /sugar solution)

Table 3. Description of mixture for different samples of honey and sugar



Fig. 4. Display of different samples of honey and sugar

#### 4.2 Electronic nose setup and measurement

The e-nose used was Cyranose320 from Smith Detection™, consists of 32 non-selective sensors of different types of polymer matrix, blended with carbon black composite, configured as an array. It can be trained to analyze both simple and complex vapor mixtures with equal ease. When the sensors are exposed to vapors or aromatic volatile compounds they swell, changing the conductivity of the carbon pathways and causing an increase in the resistance value that is monitored as the sensor signal. The resistance changes across the array are captured as a digital pattern i.e. representative of the test smell (Dutta et al., 2006).

The e-nose setup for this experiment is illustrated in Fig. 5 and the setting of the sniffing cycle is also indicated in Table 4. Each sample was drawn from the bottle using 10ml syringe and kept in a 13 x 100 mm test tube and seal with a silicone stopper. Each sample was replicated ten times. Before measurement, each sample was placed in a heater block and heat up for 10 minutes to generate sufficient headspace volatiles. The temperature of sample was controlled at  $50 \pm ^\circ\text{C}$  during the headspace collection.

Preliminary experiments were performed to determine the optimal experimental setup for the purging, baseline purge and sample draw durations. Ten seconds baseline purge with 40 seconds sample draw produced an optimal result (result is not shown). Baseline purge was set longer to ensure residual gases were properly removed since all the samples are in a liquid form and contains moisture. The pump setting was set to medium speed during

sample draw. The filter used is made up of activated carbon granules and has large surface area which is effective to remove a wide range of volatile organic compounds and moisture in the ambient air. The experiment was carried out using e-nose for a variety of honey samples followed by sugar and adulterated samples.

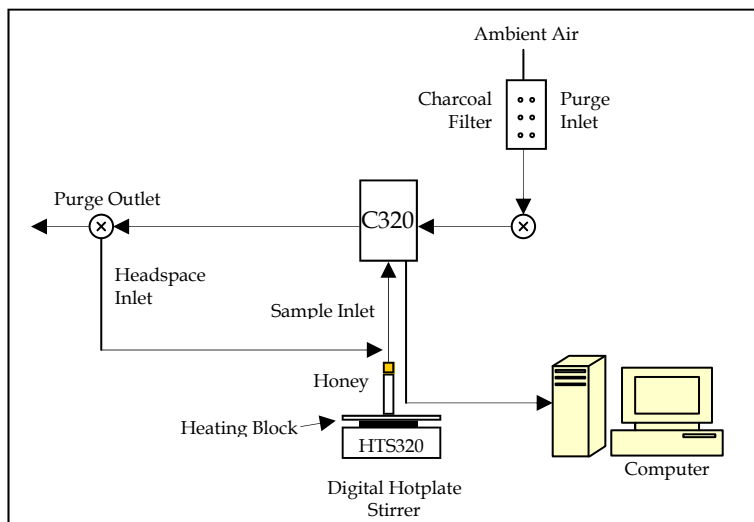


Fig. 5. E-nose setup for headspace evaluation of honey, sugar concentration and adulteration sample

	Cycle	Time (s)	Pump Speed
Sampling Setting	Baseline Purge	10	120 mL/min
	Sample Draw	40	120 mL/min
	Idle Time	3	-
	Air Intake Purge	40	120mL/min

Table 4. E-nose parameter setting for honey, sugar and adulterated samples assessment

### 4.3 Electronic tongue setup and measurement

The chalcogenide-based potentiometric e-tongue was made up of eleven distinct ion-selective sensors from Sensor Systems (St. Petersburg, Russia). The e-tongue system shown in Figure 6 was implemented by arranging an array of potentiometric sensors around the reference probe. Table 5 describes the potentiometric sensors used in this experiment. Each sensor output was connected to the analogue input of a data acquisition board (NI USB-6008) from National Instruments (Austin TX, USA).

A 10% (w/v) solution of honey in distilled water was prepared and stirred for 3 minutes at 1000rpm before making any measurements. Each sample was replicated ten times. For each measurement, the e-tongue was steeped simultaneously and left for two minutes, and the potential readings were recorded for the whole duration. After each sampling, the e-tongue was rinsed twice using distilled water (stirred at 400rpm for two minutes) to remove any

sticky residues from previous sample sticking on the sensor surface to avoid contaminating of the next sample.

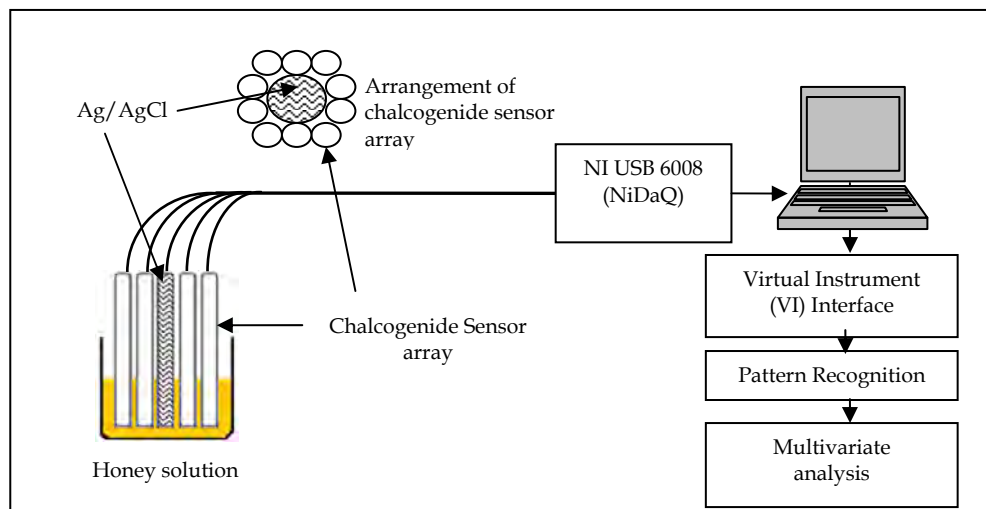


Fig. 6. E-tongue setup for headspace evaluation of honey, sugar concentration and adulterated sample

Sensor Label	Description
Fe <sup>3+</sup>	Ion-selective sensor for Iron ions
Cd <sup>2+</sup>	Ion-selective sensor for Cadmium ions
Cu <sup>2+</sup>	Ion-selective sensor for Copper ions
Hg <sup>2+</sup>	Ion-selective sensor for Mercury ions
Ti <sup>+</sup>	Ion-selective sensor for Titanium ions
S <sup>2-</sup>	Ion-selective sensor for Sulfur ions
Cr(VI)	Ion-selective sensor for Chromium ions
Ag <sup>+</sup>	Ion-selective sensor for Argentum ions
Pb <sup>2+</sup>	Ion-selective sensor for Plumbum ions
HI 5311	pH sensor
HI 2111	Reference probe using Ag/AgCl electrode

Table 5. Chalcogenide-based potentiometric electrodes used in the e-tongue.

#### 4.4 Data preprocessing

The fractional measurement method is essential when using a multi-modalities sensor fusion. This technique is often known as baseline manipulation and was applied to preprocess the data of both modalities (Gardner & Bartlett, 1999). The maximum sensor response,  $S_i$  is subtracted from the baseline,  $S_0$  and then divided again by the  $S_0$ . The formula for this dimensionless and normalized  $S_{frac}$  is determined as follows:

$$S_{frac} = [S_t - S_0]/S_0 \quad (5)$$

This gives a unit response for each sensor array output with respect to the baseline, which compensates for sensors that have intrinsically large varying response levels. It can also further minimize the effect of temperature, humidity and temporal drifts (Gardner & Bartlett, 1999).

The data from different modalities were processed separately and all sensors were used in this analysis. In the case of the e-nose,  $S_0$  is the minimum value taken during the baseline purge with ambient air and  $S_t$  was measured during the sample draw. Each sampling cycle was repeated three times and the average was obtained for each of ten replicated samples. For the e-tongue measurements,  $S_0$  (baseline reading) is the average reading of distilled water, while  $S_t$  is the sensor reading when steeped in the solution. The steeping cycle was repeated three times for each sample and the average was obtained for each ten of the replicated samples. Each  $S_{frac}$  data point from each e-nose and e-tongue sensor formed the  $S_{frac}$  matrix for further analyses.

#### 4.4.1 Low level data fusion

For the purpose of low level data fusion, measurements recorded from both sensors were fused during the data level. For the e-nose data, there were 720 observations with 32 features from 16 different honey, sugar and adulterated samples. Likewise for the e-tongue data, 720 observations with 11 features from 16 different honey, sugar and adulterated samples were recorded. As a result, a new dimension for the fused data was represented by 720 observations with 43 features. At this stage, the original data from both measurements is formed in a data matrix, and is described in Fig. 7 as follows. No transformation is being applied at this stage.

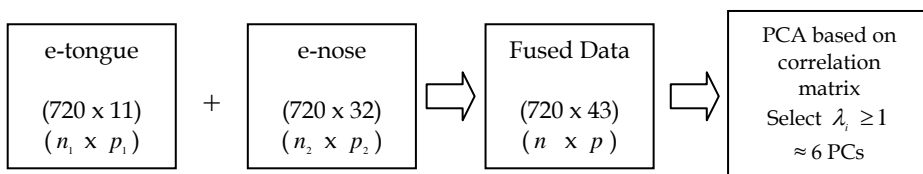


Fig. 7. Illustration of fusing data in low level data fusion

The correlation input matrix from the fused data was proceeded for the PCA calculation. For the purpose of classification in LDA, the reduced number of principal components was selected based on magnitude eigenvalues greater or equal to 1 ( $\lambda_i \geq 1$ ). The result from the scree plot is also applied for comparison and confirmation purposes.

#### 4.4.2 Intermediate level data fusion

In this framework, fusion was applied after feature extraction process. For that purpose, PCA was calculated based on the correlation matrix from both datasets. The number of principal components to retain is decided based on the associated eigenvalues with magnitude greater than or equal to 1.0 ( $\lambda_i \geq 1$ ). The results were double checked using the

scree plot of each dataset. Fig. 8 illustrates the related processes. The resulting principal components from each sensor which is three principal components were then combined before the classification using LDA is performed.

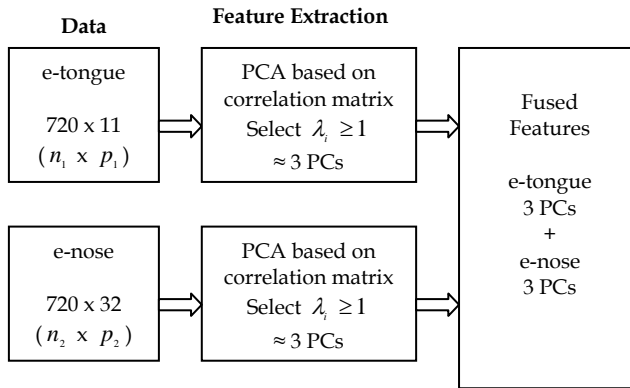


Fig. 8. Illustration of fusing extracted features in intermediate level data fusion

## 5. Results and discussion

Before the analyses of PCA was continued, a thorough study on each and every selected principal components (i.e. at low level data fusion) considered for classification using LDA was performed and the resulting classification error rate for each case are highlighted in Fig. 9. Comparisons and evaluations of classification error rate were performed differently based on correlation or covariance input matrix, procedure to evaluate performance of leave-one-out approach and the elimination of the least important of principal components (i.e. elimination begin with principal components of the smallest eigenvalue). Table 6 shows the total of variance explained using the correlation and covariance matrix input for the low level data fusion.

Number of Retained Principal Component	Correlation Matrix		Covariance Matrix	
	Total Eigenvalue (%)	Error Rate (Leave-one-out)	Total Eigenvalue (%)	Error Rate (Leave-one-out)
2	72.070	0.546	77.277	0.546
3	77.836	0.500	86.640	0.500
⋮	⋮	⋮	⋮	⋮
20	99.581	0.151	99.999	0.153
21	99.649	0.151	100.000	0.150
⋮	⋮	⋮	⋮	⋮
42	99.998	0.142	100.000	0.142
43	100.000	0.144	100.000	0.144

Table 6. Total variance explained for low level data fusion

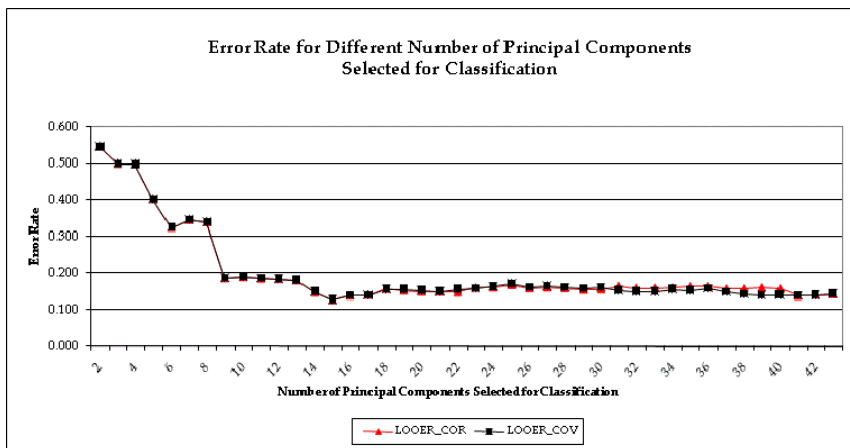


Fig. 9. Different classification performance for correlation and covariance input matrix with leave-one-out approach.

Fig. 9 clearly reveals similar classification performance of correlation and covariance input matrix with a leave-one-out approach for the low level data fusion. It should be highlighted here that the performance of classification for the correlation and covariance input is not much differ because the standard deviations for each features in the fused dataset is slightly small.

In reality, good classification performance is not determined by the greater number of features included in data. What we need is features with the most discriminative effect which often measured by the error rate. In the case of low level data fusion, the PCA based on the correlation matrix of fused data was used to extract the most important features in a linear combination form. Table 7 displays the total of variance explained for the principal components of low level data fusion. Six principal components with eigenvalues greater than or equal to 1.0 were retained to be the input for classification using LDA. It can be seen that with only six linear combinations of the original features out of 43-principal-component, we only loose about 9.3% of information to proceed with classification task. The scree plot in Fig. 10 also shows that six principal components should be retained.

Component	Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %
1	27.167	63.179	63.179
2	3.823	8.891	72.070
3	2.480	5.767	77.836
4	2.223	5.169	83.005
5	1.966	4.572	87.577
6	1.356	3.153	90.730
⋮	⋮	⋮	⋮

Table 7. Total variance explained for low level data fusion

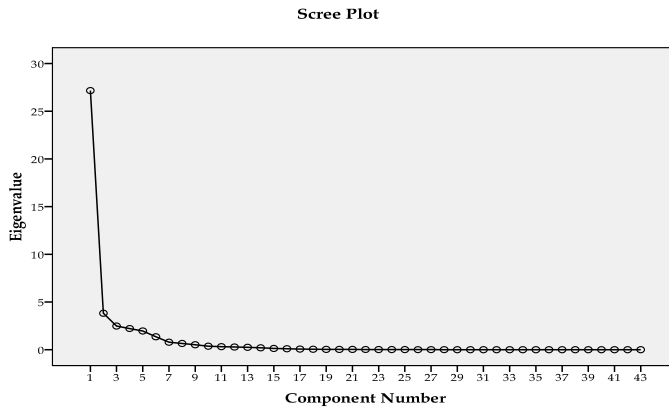


Fig. 10. Scree plot for the low level data fusion

Component	Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %
1	4.023	36.573	36.573
2	2.232	20.289	56.862
3	1.930	17.549	74.411
⋮	⋮	⋮	⋮

Table 8. Total variance explained of e-tongue data for intermediate level data fusion

Component	Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %
1	26.652	83.287	83.287
2	2.336	7.300	90.587
3	1.287	4.023	94.610
⋮	⋮	⋮	⋮

Table 9. Total variance explained of e-nose data for intermediate level data fusion

Table 8 and 9 display the total of variance explained for the principal components of intermediate level data fusion. Based on the eigenvalues greater than or equal to 1.0 from both e-tongue and e-nose data, three principal components each were retained to be the input for classification using LDA. With the three principal components selected from e-tongue and e-nose data, we loose about 31% of information which is quite high compared to the low level data fusion. The scree plot in Fig. 11 seems agrees that three principal components are adequate to represent the original features.



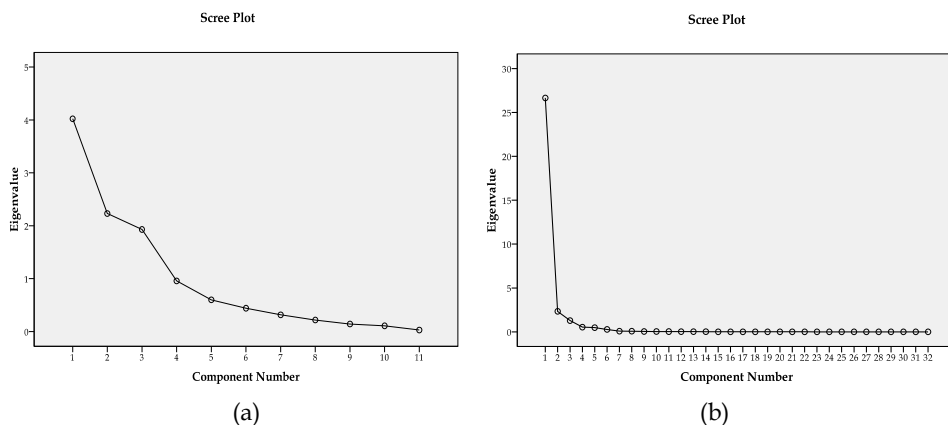


Fig. 11. Scree plot for (a) e-tongue data and (b) e-nose data low level data fusion

The selected principal components for low and intermediate level data fusion are further analyzed. The classification and prediction of the class of different types of pure honey, sugar, and adulterated samples were carried out using LDA with leave-one-out procedure. Table 10 indicates the significant differences in means of the predictors (i.e. the selected principal components) between the seven groups for both fused models. The results indirectly show the importance of the principal component to the discrimination function. Based on the Wilk’s Lambda, principal component with smaller value means it is an important predictor. The most important principal components to the least important were arranged according to the italic number. Note in contrast, the bigger the Wilk’s Lambda, the smaller the F values. Besides knowing the important predictors for the discrimination function, it is worth to investigate whether the assumption of homogeneity of covariance matrices is met. Table 11 displays the Box’s M test for both data fusion models. The significant values of both data fusion models indicate that the covariance matrices are not similar for the seven groups.

Tests of Equality of Group Means							
Low Level Data Fusion				Intermediate Level Data Fusion			
	Wilks' Lambda	F	Sig.		Wilks' Lambda	F	Sig.
PC1	.777 <sup>5</sup>	34.109	.000	PC1_EN	.794 <sup>5</sup>	30.742	.000
PC2	.686 <sup>2</sup>	54.404	.000	PC2_EN	.612 <sup>2</sup>	75.467	.000
PC3	.741 <sup>4</sup>	41.578	.000	PC3_EN	.928 <sup>6</sup>	9.206	.000
PC4	.739 <sup>3</sup>	42.005	.000	PC1_ET	.718 <sup>4</sup>	46.707	.000
PC5	.399 <sup>1</sup>	178.960	.000	PC2_ET	.676 <sup>3</sup>	56.940	.000
PC6	.921 <sup>6</sup>	10.183	.000	PC3_ET	.423 <sup>1</sup>	162.029	.000

Table 10. Test of equality of group means to identify the important variable to the discrimination function

Test Results					
Low Level			Intermediate Level		
Box's M		3194.447	Box's M		3450.654
F	Approx.	22.505	F	Approx.	24.310
	df1	126		df1	126
	df2	6677.884		df2	6677.884
	Sig.	.000		Sig.	.000

Table 11. Test null hypothesis of equal population covariance matrices.

Based on Table 12 and Table 13, all the first five discriminant functions for low and intermediate level data fusion are able to explain 100% of the total variance. However, the canonical correlation values greater than 0.5 reveal that only the first two discriminant functions from both fusion model describe strong relationship.

Eigenvalues				
Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	7.151	75.1	75.1	.937
2	2.177	22.9	98.0	.828
3	.106	1.1	99.1	.309
4	.076	.8	99.9	.267
5	.008	.1	100.0	.090
6	.000	.0	100.0	.000

Table 12. Percentage of variance explained for each discrimination function for low level data fusion.

Eigenvalues				
Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	6.365	74.3	74.3	.930
2	2.015	23.5	97.8	.818
3	.105	1.2	99.1	.309
4	.074	.9	99.9	.263
5	.006	.1	100.0	.077
6	.000	.0	100.0	.002

Table 13. Percentage of variance explained for each discrimination function for intermediate level data fusion.

The best predictors in predicting the types of honey, sugar, and adulterated samples from the respective discrimination functions of each data fusion model are marked italic in Table 14. The highest value in each function (column) marks as the best predictor. For example, the best predictor for the first discriminant function of the low level data fusion is the third principal components (PC3).

Standardized Canonical Discriminant Function Coefficients						
Function (Low Level Data Fusion)						
	1	2	3	4	5	6
PC1	-0.420	.758	.277	.583	.404	-.165
PC2	-.342	.902	.353	-.500	-.018	.265
PC3	1.299	-.128	.139	.275	.057	.661
PC4	1.097	-.452	.352	-.312	.497	-.281
PC5	-1.236	-.580	.171	-.008	.179	.218
PC6	.115	-.317	.716	.193	-.591	-.207
Standardized Canonical Discriminant Function Coefficients (cont'd)						
Function (Intermediate Level Data Fusion)						
	1	2	3	4	5	6
PC1_T	1.238	-.117	.161	.340	-.089	.628
PC2_T	.968	-.363	.384	-.144	.621	-.166
PC3_T	-1.263	-.512	.146	.074	.177	.258
PC1_N	-.084	.615	-.153	.562	.636	-.027
PC2_N	-.020	.923	.434	-.320	.119	.275
PC3_N	.005	.032	.741	.551	-.287	-.297

Table 14. Indication of relative importance of the independent variables in predicting the groups for both data fusion models.

Graphical representations of the classification for low level data fusion and intermediate level data fusion are as of Fig. 12 and Fig. 13 respectively. Table 15 and 16 describes in detail the classification results for each fusion model. It seems that the classification of several types of pure honey (group 1), beet sugar (group 2) and cane sugar (group 3) were very

good. Confusions occur a lot for adulterated samples of group 4, 5, 6 and 7. As we can see the classification performance of the intermediate level data fusion based on the leave-one-out approach is slightly better than the classification performance of the same approach of low level data fusion with 73.5% and 71.5% correct classification respectively.

Cross-validated Classification Results of Leave-One-Out Procedure									
	Group	Predicted Group Membership							Total
		1	2	3	4	5	6	7	
Count	1	300	0	0	0	0	0	0	300
	2	0	10	0	0	0	0	0	10
	3	0	0	10	0	0	0	0	10
	4	6	10	4	49	0	31	0	100
	5	0	0	8	0	41	4	47	100
	6	1	6	0	35	0	58	0	100
	7	0	0	0	0	53	0	47	100

Table 15. Classification performance for low level data fusion

Cross-validated Classification Results of Leave-One-Out Procedure									
	Group	Predicted Group Membership							Total
		1	2	3	4	5	6	7	
Count	1	300	0	0	0	0	0	0	300
	2	0	10	0	0	0	0	0	10
	3	0	0	10	0	0	0	0	10
	4	7	9	4	46	0	34	0	100
	5	0	0	10	1	45	5	39	100
	6	2	6	0	23	4	65	0	100
	7	0	0	0	0	47	0	53	100

Table 16. Classification performance for intermediate level data fusion

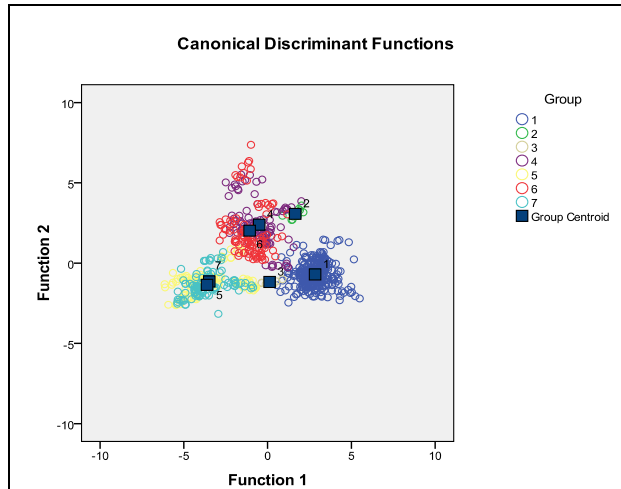


Fig. 12. Seven groups discriminating plot for low level data fusion

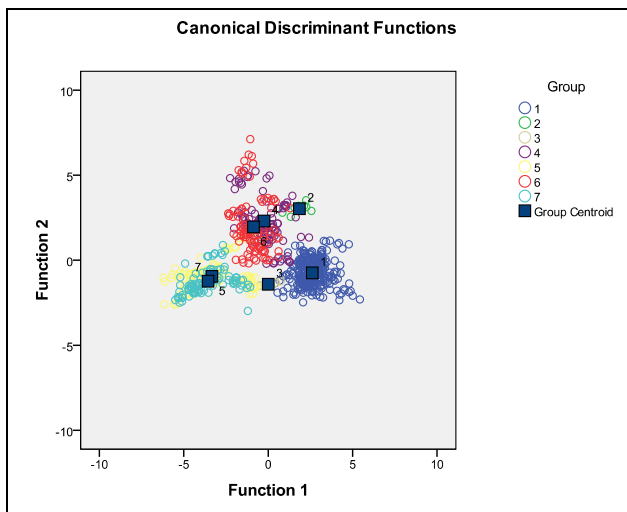


Fig. 13. Seven groups discriminating plot for intermediate level data fusion

## 6. Conclusions

This study focuses on the application of PCA in reducing the dimension of fused data from e-tongue and e-nose at low level and intermediate level data fusion. Previous studies on PCA have proven that this method is strongly advisable to be applied before performing any classification. In this study, we have shown the ability of PCA to create new variables in the form of principal components of the original features. Even though with some loss of information, special characteristics preserved in the selected principal components have made the new variables as reliable predictors in the discrimination and classification

process. In order to improve the classification performance of the multi sensor data fusion models in this study, there are two special attentions that should be given. Firstly, to fulfil the discriminant analysis assumption on the homogeneity of covariance for each group, and secondly to study and overcome the violation effect to discriminant analysis method caused by the existence of outliers. In future, we will attempt to solve these problems.

## 7. Acknowledgement

The equipment used in this project was provided by the Universiti Malaysia Perlis (UniMAP). This project is also funded by the Fundamental Research Grants Scheme (9003-00250), Ministry of Higher Education Malaysia (MOHE) and Short Term Grant (2011), Universiti Sains Malaysia (USM). The authors take this opportunity to express their sincere gratitude to Prof. Mohd Noor Ahmad (UniMAP), and Assoc. Prof. Abdul Hamid Adom (UniMAP) for their support. The authors acknowledge the financial sponsorship provided by UniMAP and MOHE, under the Academic Staff Training Scheme.

## 8. References

- Afifi, A., A. Clark, V., & May, S. (2004). *Computer-Aided Multivariate Analysis*. Chapman & Hall, ISBN 1-58488-308-1, Boca Raton, Florida
- Berrueta, L.A. & Alonso-Slases, R.M., Heberger, K. (2007). Supervised pattern recognition in food analysis. *J. Chrom: A*, Vol. 1158, pp. 196-214
- Boilot, P.; Hines, E. L.; Gongora, M.A. & Folland, R. S. (2003). Electronic Noses Inter-Comparison, Data Fusion and Sensor Selection in Discrimination of Standard Fruit Solutions. *Sensors and Actuators*, Vol. B 88, pp. 80-88
- Borgognone, M. G.; Bussi, J. & Hough, G. (2001). Principal Component Analysis: Covariance or Correlation Matrix. *Food Quality and Preference*, Vol. 12, pp. 323-326
- Buratti, S.; Benedetti, S.; Scampicchio, M. & Pangerod, E. C., (2004). Characterization and Classification of Italian Barbera Wines by Using an Electronic Nose and an Amperometric Electronic Tongue. *Analytica Chimica Acta*, Vol. 525, September 2004, pp. 133-139
- Cattell, R. B. (1966). The scree test for the number of factors. *Multiv.Behav. Res.*, Vol. 1, pp. 245-276.
- Chatfield, C. & Collins, A. J. (1980). *Introduction to multivariate analysis*; Chapman and Hall, ISBN 0-412-16030-7, Great Britain
- Cimander, C.; Carlsson, M. & Mandenius, C. (2002). Sensor Fusion for On-Line Monitoring of Yoghurt Fermentation. *Journal of Biotechnology*, Vol. 99, pp. 237-248.
- Cole, M.; Covington, J. A. & Gardner, J. W. (2011). Combined Electronic Nose and Electronic Tongue for a Flavor Sensing System. *Sensors and Actuators B: Chemical*, Vol. 156, Issue 2, pp. 832-839
- Cosio, M. S.; Ballbio, D.; Benedetti, S. & Gigliotti, C. (2007). Evaluation of Different Conditions of Extra Virgin Olive Oils with an Innovative Recognition Tool Built by Means of Electronic Nose and Electronic Tongue. *Food Chemistry*, Vol. 101, February 2006, pp. 485-491
- D'Amico, A.; Di Natale, C. & Paolesse, R. (2000). Portraits of Gasses and Liquids by Arrays of Nonspecific Chemical Sensors: Trends and Perspectives. *Sensors and Actuators B*, Vol. 68, 2000, pp. 324-330

- Di Natale, C.; Martinelli, E.; Pennazza, G.; Orsini, A. & Santonico, M. (2006). Data Analysis for Chemical Sensor Array. *Advances in Sensing with Security Applications*, pp.147-169.
- Dillon, W. R. & Goldstein, M. (1984). *Multivariate analysis, methods and applications*. John Wiley & Sons, Inc., ISBN 0-471-08317- 8, New York, USA
- Duc, B.; Bigun, E. S.; Bigun, J.; Maitre, G. & Fischer, S. (1997). Fusion of Audio and Video Information for Multi Modal Person Authentication. *Pattern Recognition letters*, Vol. 18, pp. 835-843
- Dutta, R.; Das, A.; Stocks, N.A.; Morgan, D. (2006). Stochastic Resonance-based Electronic Nose: A Novel Way to Classify Bacteria. *Sensors and Actuators B*, Vol. 115, pp. 17-27
- Gardner, J.W. & Bartlett, P.N. (1999). *Electronic Noses: Principals and Applications*. Oxford University Press: Oxford, 0-19-855955-0, UK
- Gnanadesikan, R. (1997). *Methods for statistical data analysis of multivariate observations*. John Wiley and Sons, Inc., ISBN 0-471-16119-5, New Jersey, USA
- Hall, D. L. (1992). *Mathematical techniques in Multisensor Data Fusion*. Artec House Inc., ISBN 0-89006-558-6, Boston, London
- Hall., D. L. & Llinas, J. (1997). An Introduction to Multisensor Data Fusion. *Proceedings of the IEEE*, 58, 6-22
- Hardle, W. & Simar, L. (2007). *Applied multivariate statistical analysis*; Springer, ISBN 3-540-72243-4, Berlin
- Harper, P. R. (2005). A Review and Comparison of Classification Algorithms for Medical Decision Making. *Health Policy*, Vol. 71, pp. 315-331
- Jackson, D.A. (1993). Stopping Rules in Components Analysis: A Comparison of Heuristical and Statistical Approaches. *Ecology*, Vol. 74, pp. 2204-2214
- Jolliffe, I. T. (2002). *Principal Component Analysis*. 2<sup>nd</sup>. Ed. Springer, ISBN 0-387-95442-2, New York, USA
- Kim, J. O. & Mueller, C. W. (1978). *Factor Analysis: Statistical Methods and Practical Issues*. Sage, ISBN 9780803911666, Beverly Hill, CA
- Krzanowski, W. J. (2000). *Principal of Multivariate Analysis, A User's Perspective*. Oxford, ISBN 0-19-850708-9, New York, USA
- Kumar, A.; Wong, D. C. M.; Shen, H. C. & Jain, A. K. (2006). Personal Authentication using Hand Images. *Pattern Recognition Letters*, Vol. 27, pp. 1478-1486
- Li, J.; Luo, S. & Jin, J. S. (2010). Sensor Data Fusion for Acurate Cloud Presence Prediction using Dempster-Shafer Evidence Theory. *Sensors*, Vol. 10, pp. 9384-9396.
- Maciejak, T. R.; Kukawska-Tarnawska, B.; Tyszkiewicz, J. & Tyszkiewicz, S. (2003). Multi-Sensor Odour Detection and Measurement of Polluted Food. *Polish Journal of Food and nutrition Sciences*, Vol. 12/53, pp. 45-48
- Manly, B. F. J. (2004). *Multivariate Statistical Methods: a Primer*. Chapman & Hall, ISBN 1-58488-414-2, Boca Raton, Florida
- Martinez, W. L. & Martinez, A. R. (2001). *Computational Statistics Handbook with Matlab*. Chapman & Hall/CRC, ISBN 1-58488-229-8, London, UK
- Mitchell, H.B. (2007). *Multi-Sensor Data Fusion, an Introduction*. Springer, ISBN 978-3-540-71463-7, Heidelberg, Berlin
- Persaud, K.; Dodd, G. (1982). Analysis of discrimination mechanisms in the mammalian olfactory system using a model nose. *Nature*, Vol. 299, pp. 352-355

- Raghavendra, R.; Dorizzi, B.; Rao, A., & Kumar, G. H. (2011). Designing Efficient Fusion Schemes for Multimodal Biometric Systems using Face and Palmprint. *Pattern Recognition*, Vol. 44, pp. 1076-1088
- Rencher, A. C. (1998). *Multivariate Statistical Inference and Applications*. Wiley, ISBN 0-471-57151-2, New York
- Smith, C. R. & Erickson, G. J. (1991). Multisensor Data Fusion: Concepts and Principals. *IEEE Pacific Rim Conference on Communications, Computers and Signal Processing*, pp. 235-237
- Sohn, S. Y. & Lee, S. H. (2003). Data Fusion, Ensemble and Clustering to Improve the Classification Accuracy for the Severity of Road Traffic Accidents in Korea. *Safety Science*, Vol. 41, pp. 1-14
- Sun, Q.; Zeng, S.; Liu, Y.; Heng, P. & Xia, D. (2005). A New Method of Feature Fusion and its Application in Image Recognition. *Pattern Recognition*, Vol. 38, pp. 2437-2448
- Vajaria, H.; Islam, T.; Mohanty, P.; Sarkar, S.; Sarkar, R. & Kasturi, R. (2007). Evaluation and Analysis of a Face and Voice Outdoor Multi-Biometric System. *Pattern Recognition Letters*, Vol. 28, pp. 1572-1580
- Winqvist, F.; Krantz-Rülcker, C. & Lundström, I., (2003). Electronic Tongues and Combinations of Artificial Senses, In: *Sensors Update*, Vol. II, Baltes, H.; Fedder, G. K. & Korvink, J. G., pp. 279-306, Wiley-VCH, ISBN 3-527-30601-3, Germany
- Winqvist, F.; Lundström, I. & Wide, P. (1999). The Combination of an Electronic Tongue and Electronic Nose. *Sensors and Actuators B*, Vol. 58, pp. 512-517
- Wu, Y.; Li, M. & Liao, G. (2007). Multiple Features Data Fusion Method in Color Texture Analysis. *Applied Mathematics and Computation*, Vol. 185, pp. 784-797
- Zakaria, A.; Shakaff, A. Y. M.; Adom, A. H.; Ahmad, M. N.; Masnan, M. J.; Aziz, A. H. A.; Fikri, N. A.; Abdullah, A. H. & Kamarudin, L. M. (2010). Improved Classification of *Orthosiphon stamineus* by Data Fusion of Electronic Nose and Tongue Sensors, *Sensors*, Vol. 10, pp. 8782-8796, ISSN 1424-8220
- Zakaria, A.; Shakaff, A. Y. M.; Masnan, M.J.; Ahmad, M. N.; Adom, A. H.; Jaafar, M. N.; A. Ghani, S., Abdullah, A. H.; Aziz, A. H. A.; Kamarudin, L. M.; Subari, N. & Fikri, N. A. (2011). A Biomimetic Sensor for the Classification of Honeys of Different Floral Origin and the Detection of Adulteration. *Sensors*, Vol. 11, pp. 7799-7822, ISSN 1424-8220