DISCRIMINANT ANALYSIS OF MULTI SENSOR DATA FUSION BASED ON PERCENTILE FORWARD FEATURE SELECTION

MAZ JAMILAH BINTI MASNAN

DOCTOR OF PHILOSOPHY
UNIVERSITI UTARA MALAYSIA
2017

Awang Had Salleh
Graduate School
of Arts And Sciences

Universiti Utara Malaysia

PERAKUAN KERJA TESIS / DISERTASI
(Certification of thesis / dissertation)

Kami, yang bertandatangan, memperakukan bahawa
(We, the undersigned, certify that)

MAZ JAMILAH MASNAN

calon untuk Ijazah
(candidate for the degree of) PhD

telah mengemukakan tesis / disertasi yang bertajuk:
(has presented his/her thesis / dissertation of the following title):

"DISCRIMINANT ANALYSIS OF MULTI SENSOR DATA FUSION BASED ON PERCENTILE FORWARD FEATURE SELECTION"

seperti yang tercatat di muka surat tajuk dan kulit tesis / disertasi.
(as it appears on the title page and front cover of the thesis / dissertation).

Bahawa tesis/disertasi tersebut boleh diterima dari segi bentuk serta kandungan dan meliputi bidang ilmu dengan memuaskan, sebagaimana yang ditunjukkan oleh calon dalam ujian lisan yang diadakan pada : *27 Ogos 2017.*
That the said thesis/dissertation is acceptable in form and content and displays a satisfactory knowledge of the field of study as demonstrated by the candidate through an oral examination held on: ***August 27, 2017.***

Pengerusi Viva: *(Chairman for VIVA)*	Prof. Dr. Sharipah Soaad Syed Yahaya	Tandatangan *(Signature)*
Pemeriksa Luar: *(External Examiner)*	Assoc. Prof. Dr. Liong Choong Yeun	Tandatangan *(Signature)*
Pemeriksa Dalam: *(Internal Examiner)*	Dr. Shamshuritawati Sharif	Tandatangan *(Signature)*
Nama Penyelia/Penyelia-penyelia: *(Name of Supervisor/Supervisors)*	Assoc. Prof. Dr. Nor Idayu Mahat	Tandatangan *(Signature)*
Nama Penyelia/Penyelia-penyelia: *(Name of Supervisor/Supervisors)*	Prof. Dato' Dr. Ali Yeon Md Shakaff	Tandatangan *(Signature)*

Tarikh:
(Date) ***August 27, 2017***

Permission to Use

In presenting this thesis in fulfilment of the requirements for a postgraduate degree from Universiti Utara Malaysia, I agree that the Universiti Library may make it freely available for inspection. I further agree that permission for the copying of this thesis in any manner, in whole or in part, for scholarly purpose may be granted by my supervisor(s) or, in their absence, by the Dean of Awang Had Salleh Graduate School of Arts and Sciences. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to Universiti Utara Malaysia for any scholarly use which may be made of any material from my thesis.

Requests for permission to copy or to make other use of materials in this thesis, in whole or in part, should be addressed to:

Dean of Awang Had Salleh Graduate School of Arts and Sciences
UUM College of Arts and Sciences
Universiti Utara Malaysia
06010 UUM Sintok

Abstrak

Penyarian fitur ialah satu kaedah yang digunakan secara meluas untuk mengekstrak fitur yang signifikan dalam masalah gabungan data pelbagai penderia. Namun demikian, penyarian fitur mempunyai beberapa kelemahan. Masalah utamanya ialah kegagalan untuk mengenal pasti fitur diskriminatif dalam data multi kumpulan. Justeru, kajian ini mencadangkan satu analisis diskriminan gabungan data pelbagai penderia yang baharu menggunakan jarak Mahalanobis tak terbatas dan terbatas untuk menggantikan kaedah penyarian fitur dalam gabungan data pelbagai penderia peringkat rendah dan pertengahan. Kajian ini juga turut membina kaedah pemilihan fitur persentil kehadapan (PFPK) untuk mengenal pasti fitur diskriminatif tersaur untuk pengelasan data penderia. Prosedur cadangan pengelasan diskriminasi bermula dengan pengiraan purata jarak antara multi kumpulan menggunakan jarak tak terbatas dan terbatas. Kemudian, pemilihan fitur dimulakan dengan memberi pangkat kepada gabungan fitur dalam peringkat rendah dan pertengahan berdasarkan jarak yang dikira. Subset fitur telah dipilih menggunakan PFPK. Peraturan pengelasan yang dibina diukur menggunakan ukuran kejituan pengelasan. Keseluruhan penyiasatan telah dijalankan ke atas sepuluh data penderia *e-nose* dan *e-tongue*. Dapatan menunjukkan bahawa jarak Mahalanobis terbatas lebih superior dalam memilih fitur yang penting dengan bilangan fitur yang sedikit berbanding kriterium jarak tak terbatas. Tambahan pula, dengan pendekatan jarak terbatas, pemilihan fitur menggunakan PFPK memperolehi kejituan pengkelasan yang tinggi. Keseluruhan prosedur yang dicadangkan didapati sesuai untuk menggantikan analisis diskriminan gabungan data pelbagai penderia tradisional berdasarkan kuasa diskriminatif yang besar dan kadar penumpuan yang pantas pada kejituan pengelasan yang tinggi. Kesimpulannya, pemilihan fitur boleh menyelesaikan masalah penyarian fitur. Kemudian, PFPK yang dicadangkan terbukti efektif dalam memilih subset fitur dengan kejituan yang tinggi serta pengiraan pantas. Kajian ini juga menunjukkan kelebihan jarak Mahalanobis tak terbatas dan terbatas dalam pemilihan fitur bagi data berdimensi tinggi yang bermanfaat kepada kedua-dua jurutera dan ahli statistik dalam teknologi penderia.

Kata Kunci : Analisis Diskriminan, Gabungan Data Pelbagai Penderia, Jarak Mahalanobis Tak terbatas, Jarak Mahalanobis Terbatas, Pemilihan Fitur Persentil Kehadapan

Abstract

Feature extraction is a widely used approach to extract significant features in multi sensor data fusion. However, feature extraction suffers from some drawbacks. The biggest problem is the failure to identify discriminative features within multi-group data. Thus, this study proposed a new discriminant analysis of multi sensor data fusion using feature selection based on the unbounded and bounded Mahalanobis distance to replace the feature extraction approach in low and intermediate levels data fusion. This study also developed percentile forward feature selection (PFFS) to identify discriminative features feasible for sensor data classification. The proposed discriminant procedure begins by computing the average distance between multi-group using the unbounded and bounded distances. Then, the selection of features started by ranking the fused features in low and intermediate levels based on the computed distances. The feature subsets were selected using the PFFS. The constructed classification rules were measured using classification accuracy measure. The whole investigations were carried out on ten e-nose and e-tongue sensor data. The findings indicated that the bounded Mahalanobis distance is superior in selecting important features with fewer features than the unbounded criterion. Moreover, with the bounded distance approach, the feature selection using the PFFS obtained higher classification accuracy. The overall proposed procedure is found fit to replace the traditional discriminant analysis of multi sensor data fusion due to greater discriminative power and faster convergence rate of higher accuracy. As conclusion, the feature selection can solve the problem of feature extraction. Next, the proposed PFFS has been proved to be effective in selecting subsets of features of higher accuracy with faster computation. The study also specified the advantage of the unbounded and bounded Mahalanobis distance in feature selection of high dimensional data which benefit both engineers and statisticians in sensor technology.

Keywords : Bounded Mahalanobis Distance, Discriminant Analysis, Multi Sensor Data Fusion, Percentile Forward Feature Selection, Unbounded Mahalanobis Distance

Acknowledgement

My utmost gratitude goes to my Creator Ya Wakil Ya Hakim Ya Wahhab – for all the experiences, lessons and gifts in completing my PhD journey. Million thanks to my supervisors, Associate Prof. Dr. Nor Idayu Mahat and Dato‟ Prof. Dr. Ali Yeon Md Shakaff from the Centre of Excellence for Advanced Sensor Technology (CEASTech), who have provided me with endless support, guidance and advice throughout my study.

My sincere thanks to the Dean of Institute of Engineering Mathematics (IMK), Dr. Muhammad Zaini Ahmad as well as Prof. Dr. Amran Ahmed, Associate Prof. Dr. Abdul Wahab Jusoh and Associate Prof. Abdull Halim Abdul as the ex-deans of IMK for the continuous support. Not to forget the Vice Chancellor of Universiti Malaysia Perlis (UniMAP), Dato‟ Prof. Dr. Zul Azhar Zahid Jamal for the precious opportunity to complete my study. This study would not have been possible without the financial support and opportunity from the Ministry of Higher Education as well as UniMAP. To all members of IMK, School of Quantitative Sciences UUM-CAS, and Awang Had Salleh Graduate School of Arts and Sciences, thank you very much for everything. My appreciation goes to all researchers at CEASTech especially Dr. Ammar Zakaria and Associate Prof. Dr. Abu Hassan Abdullah for the useful and helpful assistances.

I am forever indebted to my beloved parents (Masnan Pardi and Zainab Mohamad) and parents-in-law (late Mohd Isa Mohd Noh and Fatimah Zaharah Abu Hassan) for their continuous encouragement and du‟a. My humble thanks to all my family members and in-laws for the assistances throughout the years. Not to foget, my thanks to those who have contributed directly or indirectly to the thesis making. Finally, my deepest appreciation and thanks is dedicated to my husband Mohd Faizal Mohd Isa and my angels Mohd Fathurrahman, Mirrah Nashihin, Mirrah Nabihah and Muhammad Ukail Fikri for your sacrifies, understanding, du‟a and nerver-ending loves. I hope this tiny masterpiece would instigate more significance researches for the goodness of mankind. May Allah accept this work as good-deed.

Table of Contents

List of Tables

List of Figures

List of Appendices

Glossary of Terms

Gustatory – relates to the sensations that arise from the stimulator of taste receptor cells found throughout the mouth or easily known as sense of taste.

Olfactory – the sense of smell mediated by specialized sensory cells of the nasal cavity of vertebrates.

Sensor data – the signals from specific sensor that has been preprocessed according to some suitable preferred methods.

Array sensor – a combination of sensors arranged in an array to overcome the problem of poor sensitivity and poor selectivity.

Features – or sometimes known as variables referring to the dimension of sensor data. Easily determined as the number of array sensors attached in a sensor

Group – or category is defined as a grouping of samples characterized by the same value of discrete variables or by contiguous values of continuous variables.

Non-selectivity – a situation where the qualitative and quantitative information are combined and the sensor response become highly ambiguous which makes the sensor unusable in real conditions when sensors are exposed to more than one analyte species.

Redundancy – occurrs as a consequence of the non-selectivity state where sensors are measuring the same response which makes the related sensors highly correlated

Low level data fusion – a state of combining different sensor data at the data level

Intermediate level data fusion – a state of combining different features of different sensor data at the feature level

High level data fusion – a state of combining the decisions of different sensors at the decision level

Classifier – or sometimes called as classification function is the rule used to allocate future object with an aim to minimize the misclassification rate over all possible allocations.

Training data set – is an independent data set used to train the classifier.

Test data set – is an independent data set used to evaluate training bias and estimate real performance of the constructed classifier.

List of Abbreviations

LLDF – Low Level Data Fusion

ILDF – Intermediate Level Data Fusion

HLDF – High Level Data Fusion

LDA – Linear Discriminant Analysis

QDA – Quadratic Discriminant Analysis

kNN – *k* Nearest Neighbor

ANN – Artificial Neural Network

PCA – Principal Component Analysis

PFFS – Percentile Forward Feature Selection

CHAPTER ONE
INTRODUCTION

1.1 Introduction

Discriminant analysis is a multivariate technique that explains the group membership as a function of multiple independent variables. The group membership is the dependent variable often appears as categorical value (nominal), while the independent variables which are often called as *discriminators* are usually in continuous form (interval or ratio). Wood, Jolliffe, and Horgan (2005) described discriminant analysis as a statistical technique that assigns observations to one of several distinct populations based on measurements made on the observations, or variables derived from the measurements. The process of allocating observations to their specific groups based on the constructed *discriminant rules* is called *classification*. The concept of discriminant analysis is rather exploratory in nature whereas the classification procedures are less exploratory, but leads to well-defined rules to allocate new observations.

The notion of discriminant analysis was introduced by Sir Ronald A. Fisher in the mid of 1930s. Then, it became an area of interest to other researchers in various disciplines in the 1950s and 1960s. Some researchers break up discriminant analysis into two parts; *predictive* discriminant analysis and *descriptive* discriminant analysis. Predictive discriminant analysis focuses on the prediction of group membership based on a subset of variables selected using certain criteria which are eventually assessed by the classification accuracy. On the contrary, descriptive discriminant analysis deals with assessing the independents variables that best explain the group separation which reflects the importance. Concisely, this work adapts both concepts

where the prediction of group membership is being implemented using the most important variables measured by largest group separation. For simplicity, throughout this thesis, discriminant analysis is used as it explains proposed procedure in attempt to classify objects into some predetermined groups based on some measured variables.

The concept and usefulness of discriminant analysis in diverse fields which include the physical, biological, social sciences, engineering, and medicine are discussed through inconsiderable number of literatures. One of the subdomains in engineering that manipulate the benefit of this concept is the *multi sensor data fusion*. It has been extensively employed in the applications where multiple sources of data are required for various pattern recognition and classification research such as in sensor network, robotics, video and image processing, intelligent system design as well as in food production. Two types of artificial sensors mainly applied in the food research are the *electronic nose* (*e-nose*) and *electronic tongue* (*e-tongue*). These electronic sensors have the capability of imitating the human senses (smell and taste) using sensor arrays and pattern recognition system. Main exploitation of these sensors is to fulfill a number of research interests such as food quality assessment, food authenticity estimation, food freshness evaluation, food shelf-life investigation as well as food process monitoring.

Traditional approach in the food production related researches are highly dependent on trained human panels that solely relies on their olfactory (sense of smell) and gustatory (sense of taste) systems. Generally the manipulation of trained human panels involves lengthy and expensive methodology, which may initiate inconsistencies due to exhaustion and stress. Consequently, different analyses and

assessment results may be produced. Therefore, complementary yet reliable artificial sensors to mimic the trained human panel‟s taste and smell system are required as the conventional methods suffers from some drawbacks. The invention of e-nose and e-tongue sensors in the multi sensor data fusion framework is the key to conquer the drawbacks. Figure 1.1 shows two artificial sensors that are believed to compliment human‟s smell and taste senses.

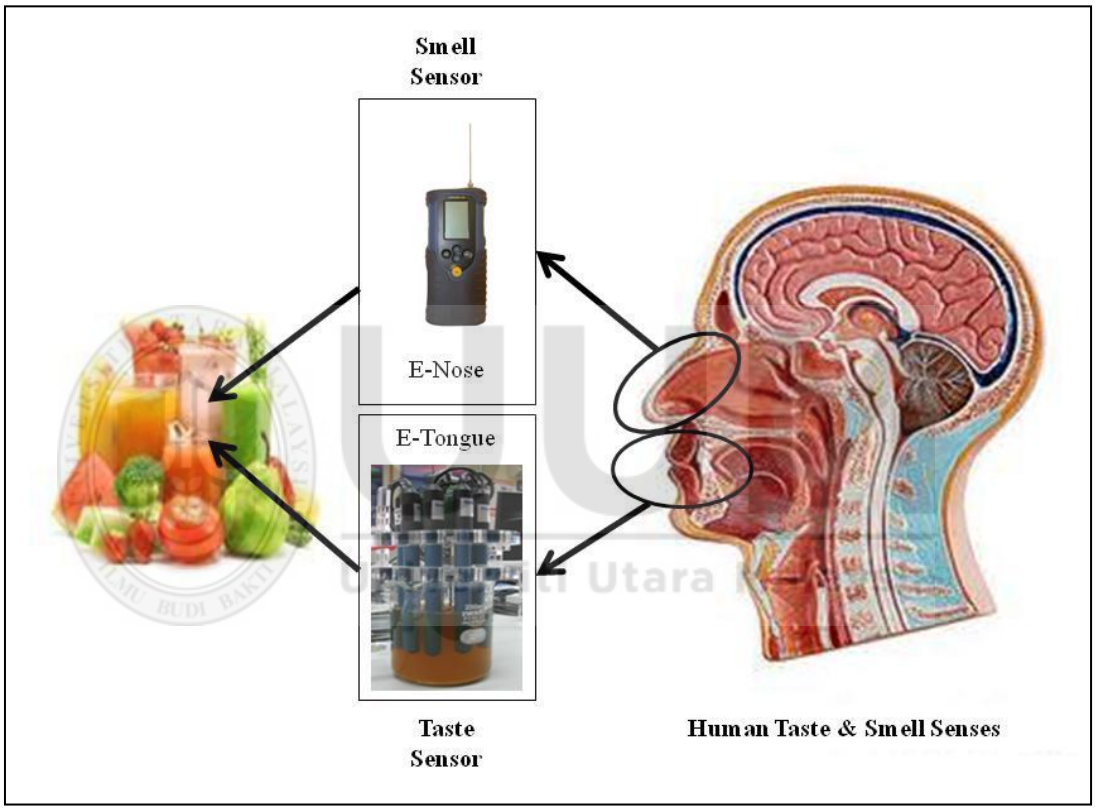

Figure 1.1. Illustration of Artificial Sensors that Imitate Human Basic Senses

The array of sensors equipped in the e-nose and e-tongue act as the detection system whenever they react to volatile compounds and chemical compounds, respectively. Figures 1.2 and 1.3 illustrate the array of sensors attached in the e-nose and e-tongue. These arrays of sensors are later referred as *features* among practitioners or commonly termed as variables among statisticians. Usually, these sensor devices are applied independently during the experiment. Thus, in order for these sensors to

work in such a way as human senses behave, the sensors are manipulated in a multi sensor data fusion framework. Presently, the application of multi sensor data fusion is burdened with abundant variables from different employed sensors, where these variables may inherit similar and/or dissimilar characteristics. Dealing with similar sensor devices may not be a big problem, but exploiting different sensor devices is a real challenge. To ensure consistency in discussion and to address both understanding on statistics and application of discriminant analysis in sensor data, this thesis will use the termed *features* to refer to measured *variables*.

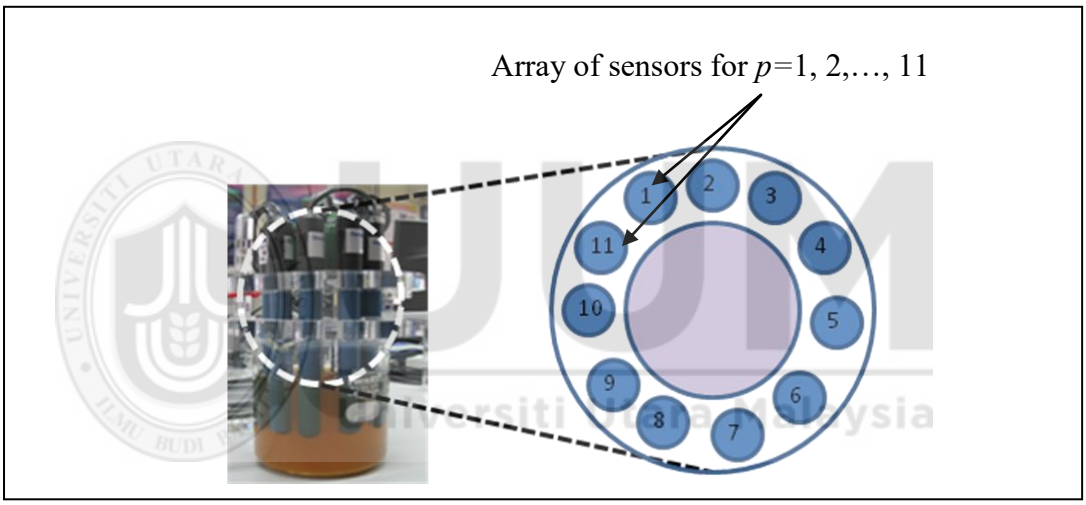

Figure 1.2. Illustration for Array of Sensors Attached in an E-Tongue (11-array)

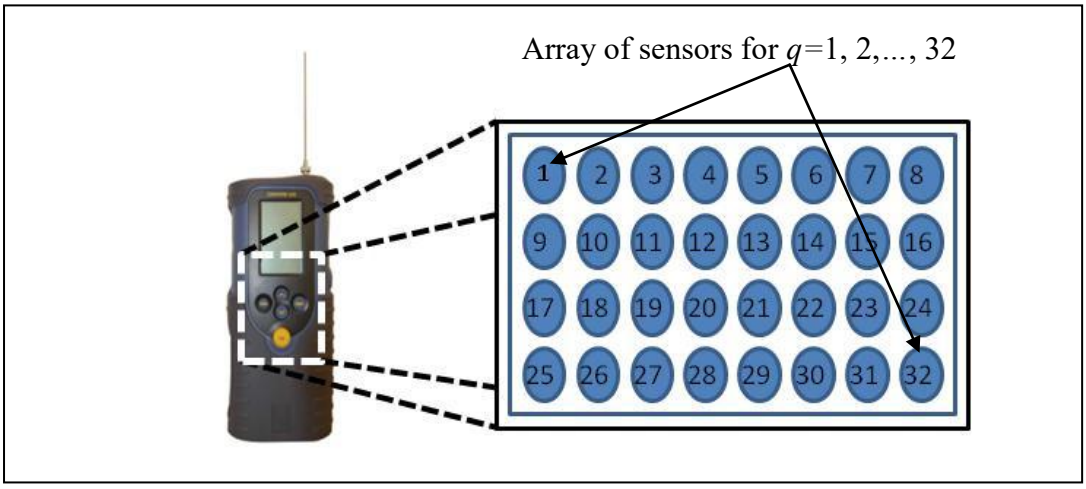

Figure 1.3. Illustration for Array of Sensors Attached in an E-Nose (32-array)

Basically, flavor is derived from the combination of the senses of taste and smell. Previous studies done by Woods (1998) and Wide, Winquist, Bergsten and Petriu (1998) suggested that the fusion of e-nose and e-tongue has the potential to mimic the human flavor panels since measurement data from both sensors are manipulated to produce sensor-specific opinions about the human-like sensing modalities. Cole, Covington and Gardner (2011) have successfully confirmed that flavor can be assessed by combining these two artificial sensors. Thus, even though e-nose and e-tongue are not integrated since each device works independently, data fusion techniques can be applied for further data analysis (Zakaria, Masnan, Zakaria and Shakaff, 2010). Presumably, e-nose and e-tongue is functioning successfully when good classification result is attained. Perhaps, to accomplish the purpose of e-nose and e-tongue to mimic the human panel‟s smell and taste is by obtaining good classification accuracy as the main goal. However, one of the challenges to achieve such purpose is to deal with variability of sensor arrays from both sensors. In real practice, sensor arrays from e-nose are highly correlated while sensor arrays from e-tongue are less correlated among each other. Such scenarios are caused by the nature of fully selective and partially selective of the e-nose and e-tongue, respectively. Since these systems are not yet integrated like the way human smell and taste system behave, different levels of multi sensor data fusion approaches can be employed for the mimicking purposes.

One of the multi sensor data fusion frameworks employed in food industries is the Joint Directors of Laboratories (JDL) Data Fusion Framework or sometimes known as JDL process model (Hall, 1992). Figure 1.4 illustrates the whole JDL data fusion framework. The framework enumerates in detail three different approaches for fusing sensor data namely *low level data fusion* (LLDF), *intermediate level data fusion*

(ILDF) and *high level data fusion* (HLDF). Obviously, a common technique involves in all level of data fusion is called the *feature extraction*. In feature extraction phase, raw data are transformed into a new form of reduced data set that is still represents the original information which is useful for further classification process.

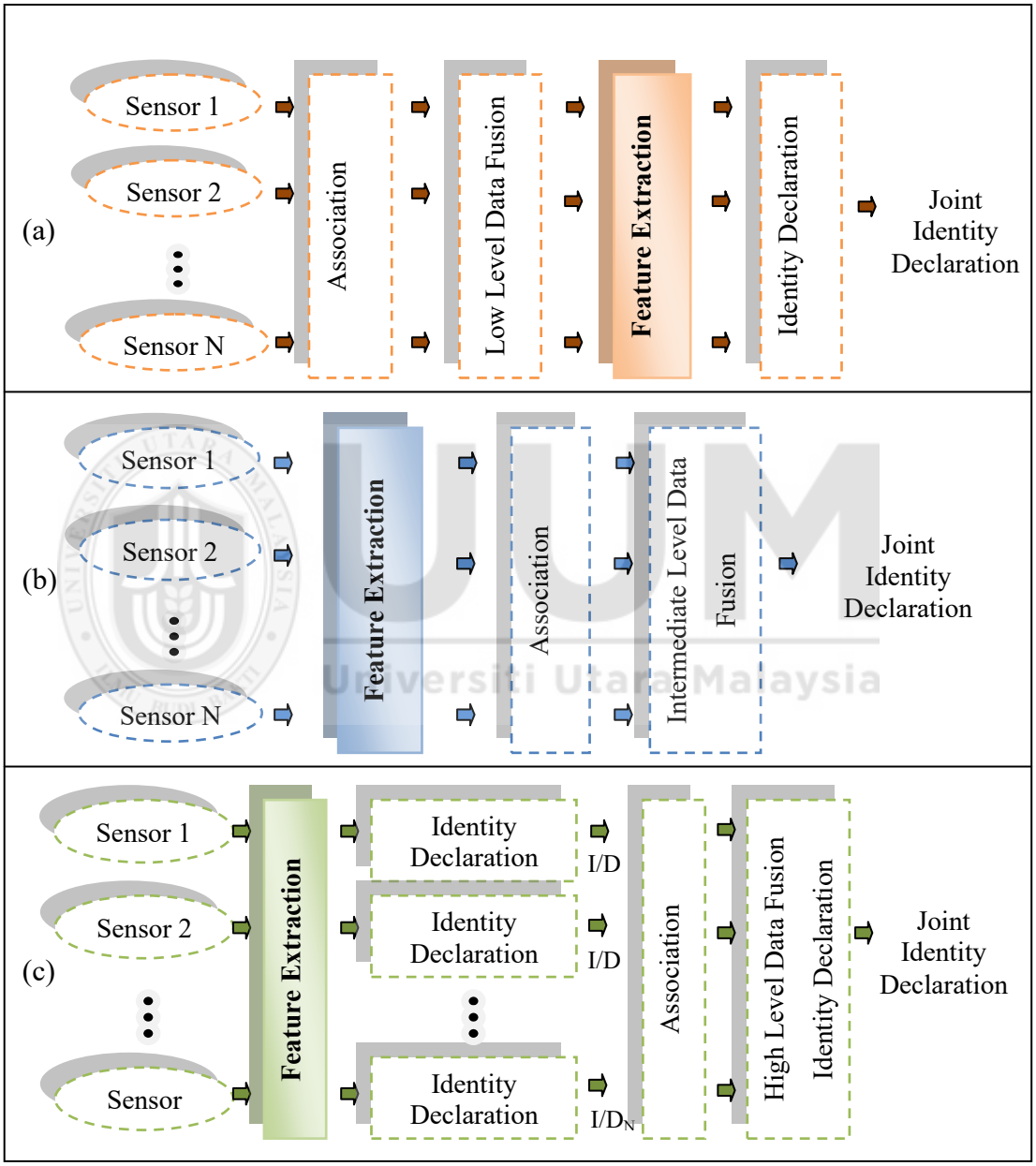

Figure 1.4. Diagrams for the JDL Data Fusion Frameworks (a) LLDF model (b) ILDF Model, and (c) HLDF Model. (Hall, 1992)

Many studies in the food related industries have been devoted to the specific feature extraction i.e. Principal Component Analysis technique where various approaches in

decision making to choose important features were suggested, evaluated and implemented. Despite of the popularity of this technique, some pitfalls of this approach remain ambiguous. The method is highly dependence on a transformation approach for selecting useful features which has several limitations. The multi sensor data fusion model with feature extraction has successfully been applied in food-based problems, but the method of feature selection has received little attention in this application, example works include (Masnan et al., 2012; Banerjee, Tudu, Shaw, Jana, Bhattacharyya, & Bandyophadhyay, 2012; and Zakaria, Masnan, Zakaria, & Shakaff, 2014).

The obvious weakness in the transformation approach is that it manipulates all the features in the analysis although some of the features may contain noise or irrelevant for the classification. Such extraneous features would only ruin the classification accuracy and in turn, reduce the potential of identifying useful features from the dataset. In addition, by applying feature extraction approach, new features defined by several functions containing all original features are formulated which then conceal the significant features. If in the case where identification of important features is of interest, the transformation approach may not be a suitable choice because it is lack of interpretability. Furthermore, the issue of indistinctness of retaining the appropriate number of principal components is another challenge.

In order to address an effective alternative to the discussed problems, this study attempts to explore the advantageous of *feature selection* method. Feature selection is a study of algorithms to reduce dimensionality of data which aims to improve classification performance. For a dataset of size n, and P is the number of observed features from e-nose, and/or Q is the number of observed features from e-tongue, the

aim of feature selection is to reduce the dimension P to p where $p \leq P$ or, to reduce the dimension Q to q where $q \leq Q$. This technique is commonly used when useful features are needed to be retained while *irrelevant* and/or *redundant* features are to be removed. In this study, relevant features are defined as features that could explain the separation between groups. Therefore, they are identified using the distance-based criterion that measures the separation between groups. Many distance-based criteria are possible to be applied such as Euclidean distance, Bhattacharyya distance, Kullback-Leibler divergence etc., but this study opts to use Mahalanobis distance Δ^2 that was first introduced by Prasanta Chandra Mahalanobis in 1936. Further details of this criterion are elaborated in the next section 1.2.

1.2 Motivation and Problem Statement

The motivation and problem for this study begins with the importance of fusion of different sensor devices, which later leads to greater number of features to deal with. Most researchers in related area practiced on implementing features extraction (Masnan et al., 2012; Prieto et al., 2011; Zakaria et al., 2011; Vera, Aceña, Guash, Boque, Mestres, & Busto, 2011; and Apetrei et al., 2010), but the capability of such strategy is arguable as most features produced by sensor devices are highly correlated (Zhang & Yan, 2015; and Wang, Tyo, & Hayat, 2007; and Ciosek, Brzózka & Wróblewski, 2004). Alternatively, the idea of substituting feature extraction phase in the multi sensor data fusion model with feature selection is possible. However, feature selection needs one to determine a selection criterion which could lead to the best possible set of features for classification purposes. As such purpose, a selection criteria based on maximizes separation among groups sounds promising (Ray & Turner, 1992; Achariyapaopan & Childers, 1985; and Jain

& Waller, 1978). Unfortunately, estimated pair-wise distances among multi-group of features that represent the groups‟ separation headed to another problem of unbounded $[0,\infty)$ separation values. Such unbounded separation values leave researchers with an argument on "how large is large for the estimated pair-wise distances shall be defined for maximum separation?" To overcome this problem, the proposed work on bounded $[0,1]$ Mahalanobis distance Δ_A^2 by Ray and Turner (1992) was considered, and it became the main focus of this study. Among other issue raised from the application of unbounded and bounded Mahalanobis distance $(\Delta^2$ and $\Delta_A^2)$, repectively, is the difference in the classification performance based on subset of features generated by the Mahalanobis distances $(\Delta^2$ and $\Delta_A^2)$. These trails are elaborated sequentially along with the trend of multi sensor sensor fusion in the following paragraphs.

Previous studies have revealed that simultaneous utilization of e-nose and e-tongue sensors is important to increase the amount of information extracted from a specific sample (Di Natale et al., 2000; Prieto et al., 2011). Some other investigations were demonstrated in Buratti, Benedetti, Scampicchio, and Pangerod (2004); Cosio, Ballabio, Benedetti, and Gigliotti (2007); Zakaria et al., (2010); Cole et al., (2011); Baldwin, Bai, Plotto, and Dea (2011); and Zakaria et al., (2011). Generally, the advantages of fusing e-nose and e-tongue in the food research have created a significant impact towards the food research domain, in which further improvement were recorded in the classification results. The significance of fusing only the e-nose and e-tongue as the sensory evaluation is important for the food safety, quality assessment, authenticity estimation, and freshness evaluation.

These systems are perceived as the human-like inspired sensor technologies to produce outputs well correlated with the human sensory panels; with which the taste sensor own the intelligent capability to break down the information included in the chemical substances to the basic information of taste quality (Ghasemi-Varnamkhasti, Mohtasebi, & Siadat, 2010; Baldwin et al., 2011). In general, these sensors are capable in measuring the chemical compounds (Apetrei, et al., 2010; Cole et al., 2011) in the form of liquid and odor from sample which is critical for such research.

In addition, recent trend in multi sensor data fusion research has shifted to the use of more sensors, for instance the use of e-nose and e-tongue with ultra violet spectrometers (as electronic eye or e-eye), Fourier Transformed Infra-Red (FTIR) spectroscopy, gas chromatography–mass spectrometry (GC-MS) and/or other kinds of sensors devices. This trend is maneuvered by the compeling needs to meet the complexity of food production research especially that directed to fulfill customer perception and acceptance. However, more sensors employed in a research does not necessarily implies the better the research is from the perspective of sensors-integrated function. The research goals within the context of food industry appear to be the determinant to the preference sensors for application. Some examples of research which involved e-nose, e-tongue and other sensors were presented by Apetrei et al., (2010); Vera et al., (2011); and Prieto et al., (2011); and Prieto et al., (2011).

As far as this research is concerned, several studies were performed using the applications of e-nose and e-tongue within the context of JDL process model which offer means to fuse at different level either at low, intermediate or high levels.

Previous studies have shown that fusing at different level has its own advantages and disadvantages, which differ from one study to another. For example, fusing at low level should be considered as the most efficient approach, but this comes with limitation such as rarely identical or commensurate sensors are applied in an analysis (Steinmetz, Sévila, & Bellon-Maurel, 1999). Steinmetz et al. (1999) further argued that fusing at intermediate level and high level use less information with respect to the raw signals provided by each sensor which include errors in the fusion process. However, fusing at the later levels are well adapted to practical cases involving sensor devices of different modalities.

In addition, fusing data from e-nose and e-tongue leads to high dimensional data problem, or rather easily expressed by $P+Q \rangle n$ problem where $P+Q$ denotes the number of fused features, and n is the number of observations. One obvious problem with data spaces of dimensionality higher than three is the difficulty to visualize data belonging to group. Moreover, high dimensional data has some unexpected mathematical properties, and as the number of dimensions increases, distance measure between groups become less meaningful. Besides, high dimensional data may create singular covariance matrices in which obtaining an inverse covariance matrix is impossible, hence classical rules cannot be constructed. The singularity problem is mainly caused by the correlated features of the applied sensors especially the e-nose. In addition, as the dimension of data increases and exceeds the number of sample, the fused data increases its complexity as well, but the classification performance is better than the single sensor system (Boilot, Hines, Gongora, & Folland, 2003). Therefore, an approach for selecting a subset of features in order to

optimize the performance of the fusion system is desired (Wide et al., 1998; Boilot et al., 2003).

As there is no simple rule for selecting a proper fusion technique, a wide range of techniques has potential applicability. Earlier, Steinmetz et al. (1999) admitted that the selection of sensor fusion method is a difficult task. Huang, Cai and Xu (2007) also brought up the same issue of determining the level of data fusion of different problems that remain a dilemma. Di Natale et al. (2000) further pointed out that there is an issue of choosing the most convenient data fusion level for maximum information from the measurements to be achieved. From the foregoing discussion, it is clear that there is no common one-fits-all architecture to fuse data from multiple sensors particularly e-nose and e-tongue. In general, implementation of data fusion approach is still unclear for a specific application and suffers from few problems as early as in the process of fusing data, selecting the appropriate fusion level, and singularity matrices which leads to dimension reduction. It is therefore, necessary to conduct an in-depth study by exploiting discriminant analysis and feature selection technique to produce a good classification rule ideal for fusion of e-nose and e-tongue.

Obviously, comprehensive research that demonstrates in-depth studies in the process of data fusion framework is still limited. This is mainly concerned with the feature extraction phase available in each data fusion level. Most previous researches that apply JDL process model were only focused on the use of Principal Component Analysis (PCA) as the feature extraction approach (Masnan, Mahat, Shakaff, Adom & Saad, 2012; Prieto et al., 2011; Vera et al., 2011; Zakaria et al., 2011; Zakaria et al., 2010; Cosio et al., 2007; Buratti et al., 2004; & Rodriguez-Mendez et al., 2004).

One of the key distinctions between feature extraction and feature selection reveals in their outcomes i.e. the features to be selected or extracted. Say six features $P_1, P_2, \cdots, P_6$ are considered for selection or extraction and only three relevant and useful features are required. If both approaches result in three features, the three *selected features* are a subset of six original features (say P_1, P_3, P_6), but the three *extracted features* (PC_is) are some combination of six original features (say $PC_1 = \sum_{i=1}^{6} a_i P_i$, $PC_2 = \sum_{i=1}^{6} b_i P_i$, and $PC_3 = \sum_{i=1}^{6} c_i P_i$) where a_i, b_i and c_i are some constants. The difference in the outcome of both approaches clearly provides a more objective choice to selection.

Even though the greatest advantage of PCA is the ability to disentangle the redundancy effect inhibited in the sensor data particularly data from e-nose and e-tongue, the selection of relevant features for further process is of interest. That is why recently feature selection techniques receive special attention from researchers in this area of study as an alternative to the feature extraction phase especially in a single sensor domain. Not to mentioned its potential and advantages in the multi sensor data fusion discipline. McLachlan (1992) stated that the application of feature selection in multiple linear regression and discriminant analysis not only leads to simpler models, but frequently improves prediction or classification accuracy. Therefore, the manipulation of feature selections in the data fusion of JDL process model with the objectives to select significant features to describe the groups" separation is worth to discover. And its suitability as well as its applicability for a specific data fusion levels is rather significance for the fusion of e-nose and e-tongue to closely mimic the human senses.

For that reason, this study aims to learn the best feature selection criteria and strategy to find the optimal subset of features meant for good classification performance. The discovery to figure out the best feature subsets is mainly based on the idea of group separation criteria using the unbounded $[0, \infty)$ Mahalanobis distance Δ^2 and bounded $[0,1]$ Mahalanobis distance Δ_A^2. There are few reasons why the Mahalanobis distance criteria is chosen compared to other distance functions. For features that have larger variance, it will receive relatively less weight when Mahalanobis distance is applied. Similarly, two highly correlated features do not contribute as much as two features that are less correlated (Rencher, 2002, pg. 76). Thus, the application of Mahalanobis distance which contains the inverse of covariance matrix is actually standardizing all features to the same variance and eliminating the correlations effect. These make the Mahalanobis distance suitable for feature selection of multi-group problem with highly correlated features.

Despite the different selection of feature subset generated by the unbounded $[0, \infty)$ Mahalanobis distance Δ^2 and bounded $[0,1]$ Mahalanobis distance Δ_A^2, a convention that leads to the use of earlier distance to the later distance are discussed. The problem of unbounded Mahalanobis distance is that Δ^2 values may increase to ∞, and it may pose difficulty in the comparison of different feature sets of its total pair-wise distance ($^gC_2$) in multi-group case. In *g*-group problem, the single large value of Δ^2 in the set of $^gC_2$ would lead to a high value of the average distance which would than fail to represent the average separability of the *g* groups. Thus, to overcome the drawback is to transform the Δ^2 values before averaging process in

such a way that the transformed measure lies within $[0,1]$ range using the Mahalanobis distance Δ_A^2.

1.3 Research Objectives

This thesis aims to investigate the potential application of distance-based feature selection in replacing the traditional feature extraction approach in the LLDF and ILDF models of the JDL model. The following objectives are expected to be attained in order to fulfill the research study:

i. to develop *univariate feature selection algorithms* using the *unbounded Mahalanobis distance* (Δ^2) and *bounded Mahalanobis distance* (Δ_A^2) for the LLDF and ILDF models,

ii. to develop *multivariate feature selection algorithms* using the *unbounded Mahalanobis distance* (Δ^2) and *bounded Mahalanobis distance* (Δ_A^2) for the LLDF and ILDF models,

iii. to construct the *parametric classification rules* based on the *percentile forward feature selection* for each of the developed algorithms in objective (i) and objective (ii), and

iv. to *evaluate the performance* of the constructed parametric classification rules.

1.4 Significance of Study

Studies on a single model of multi sensor data fusion have been done by several researchers. The LLDF was applied in the studies done by Sundic, Marco, Samitier, and Wide (2000), Di Natale et al., (2000), Boilot et al., (2003), Rodríguez-Méndez et al., (2004), Zakaria et al., (2010), and Zakaria et al., (2011). For the applications of ILDF model, studies were carried out by Rong, Ping, and Wenlei (2000), and Guru, Suraj, and Manjunath (2010), while Xiaobo, and Jiewen, (2005), Tao, and Veldhuis (2009), and Doeswijk et al., (2011) manipulated the HLDF model in their research.

There were also studies that use two different multi sensor data fusion models for different classification purposes. As far as this research is concerned, only one research applied the LLDF and ILDF models (Vera et al., 2011), as well as the ILDF and HLDF (Steinmetz et al., 1999). Similarly goes to the simultaneous manipulation of the three multi sensor data fusion models (i.e. LLDF, ILDF and HLDF) where very few studies were done using these models. Such study can be found in Huang et al., (2007) that describes conceptually the overall picture of the multi sensor data fusion. Other case studies with the application of the three models can be referred to Roussel, Bellon-Maurel, Roger, and Grenier (2003) and Rudnitskaya et al., (2006) where both studies manipulated electronic tongue and Fourier transform infrared spectroscopy (FTIR); and aroma sensors, FTIR and ultraviolet spectrometer, respectively. For that reason, this thesis aims to exploit all the LLDF, ILDF and HLDF models in the fusion of e-nose and e-tongue.

A thorough study of all the multi sensor data fusion models applied in the fusion of e-nose and e-tongue is very important in the food research community, especially when the application of fusion has proven to be an advantage. This is due to the

indistinctive results from the previous studies especially when more than one multi sensor data fusion models were applied in a research. Furthermore, there is no such consensus that can be found from the literature about the preferable multi sensor data fusion models to be used. Some researchers declared that LLDF model is the best (Rudnitskaya et al., 2006), while others found ILDF is better (Vera et al., 2011; and Steinmetz et al., 1999). On the other hand, HLDF model is claimed to be the best (Doeswijk et al., 2011; and Roussel et al., 2003). These show that different level of data fusion models (i.e. LLDF, ILDF and HLDF) adapt dissimilar information for classification due to different experiment settings, sensor devices applied and sample for test. Hence, the attempt to evaluate the usefulness and significant of all the multi sensor data fusion models involving the fusion of e-nose and e-tongue is expected to produce important findings of the best model suitable for discriminant analysis of respective fused sensors data. The findings are hoped to explicate the confusion and endless debate among the researchers of which models or mechanism worth for the fusion of e-nose and e-tongue.

Another aspect that shall be highlighted is the potential use from the fusion of e-nose and e-tongue to replace or to complement the existing sensory panels in the food industries. Baldwin et al. (2011) discussed this issue in detail. Since sensory panels may not always be available and/or quality control personnel may not be consistent in evaluating samples, hence they urge the need of such technology for faster analysis. Furthermore, the fusion mechanism has the potential as the promising tools to mimic the human sensory system (Ghasemi-Varnamkhasti et al., 2010). Even though with technological advances and promising results, the fusion of e-nose and e-tongue are still unable to mimic the biological human sensory systems (Di Rosa, Leone, Cheli, & Chiofalo, 2017), the success of fusing these sensors data and further

classify them will be a remarkable findings to the development of replacement methods for sensory panels for objective measurement of food products in a consistent and cost effective manner. Above all, the success of feature selection technique to replace the conventional feature extraction in every data fusion level namely the LLDF, ILDF and HLDF provide significant findings to the issue of fusion of e-nose and e-tongue data.

From the perspective of feature selection, this is an attempt to propose an approach to replace the feature extraction phase in the JDL fusion model specifically for the low LLDF and ILDF. Despite applying the conventional Mahalanobis distance that gives values in the range $[0 \to \infty]$ to represent the separability between group means, a more practical measure based on the bounded Mahalanobis distance that give finite range $[0,1]$ distance value is examined. The resulting finite distance value to represent the average separability has overcome the problem of m multi-group feature selection problem. These criteria is then used to create the parametric classification rule where the performance of both search is measured using the leave-one-out technique. Figure 1.5 illustrates the proposed methodological changes for the conventional multi sensor data fusion model with the feature selection approach.

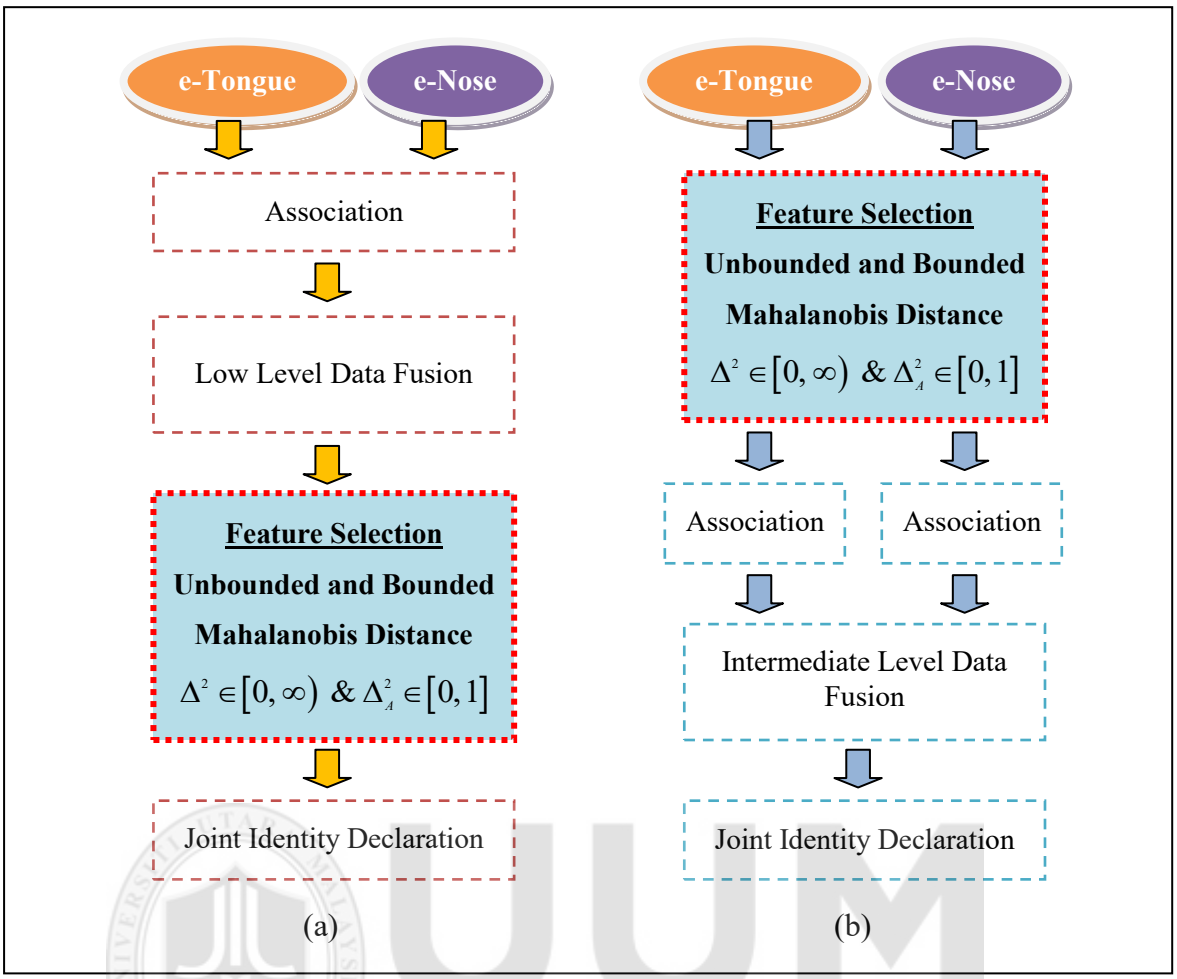

Figure 1.5. Proposed Methodological Changes for Multi Sensor Data Fusion (a) LLDF Model, and (b) ILDF Model using Feature Selection of Unbounded and Bounded Mahalanobis Distances

1.5 Scope of Study and Assumptions

This thesis addresses the problem of *feature selection* involves in multi sensor data fusion framework (LLDF and ILDF) of JDL process model specific for the exploitation of e-nose and e-tongue in the food related research. It provides an alternative to the conventional approach of extracting features included in all level of data fusion framework. Concentrations are given to determine the optimal feature subsets based on the distance-based criteria i.e. the unbounded $[0,\infty)$ Mahalanobis distance Δ^2 and bounded $[0,1]$ Mahalanobis distance Δ_A^2. The major concern in LLDF model is to select important features from the combination features of e-nose

and e-tongue. In this case, the nature of highly correlated features among e-nose sensors and moderately correlated features among e-tongue sensors would be a real challenge. Commonly, features from e-nose would dominate the classification performance as compared to the e-tongue. However, selection of discriminative features is totally depends on the performance of features that can provide greater distance among groups by minimizing the influence of the numerical values.

While for the ILDF, important features from both e-nose and e-tongue are selected independently. Once discriminant features were selected from both sensors, then only these features would be fused before they are transferred for classification puposes. However, cautious is given to the resulted fused features. There is a tendency that the selected features from both sensors may still be irrelevant once they were fused. In order to accomplish the search of the relevant features based on the selected criterion, percentile forward feature selection is further applied. The percentile forward feature selection is a bottom up search procedure that adds new features to a feature subset one at a time based on certain percentiles of the ranked features until the final feature subset is obtained.

The search strategy is suitable with the exploited filtering approach in identifying the optimal feature subset based on the separation criterion. Thus, optimal feature subset is limited to only the top highest percentiles $\left(\mathcal{P}_H\right)$ of the ranked features that produce highest univariate distance values. In accordance to the selected search, another concern for discriminant analysis of this research is the accurate estimation of the misclassification rates. Despite many available error rate estimators, the leave-one-out approach is employed. Several conditions of data in hand which include unequal

yet small sample sizes among quite a large number of groups are the main reasons for the estimator to be used.

Finally, to draw the conclusions of which data fusion model that can best fit the fusion models of e-nose and e-tongue, the non-parametric classification techniques will be tested using the predefined features in every level. All the intended algorithms for the fusion of both sensors were created using the statistical computing platform called **`RStudio`** (version 1.0.136) with a 64-bit capacity. Data manipulated in this thesis are basically secondary datasets from e-nose and e-tongue available in the Centre of Excellence for Advanced Sensor Technology (CEASTech), Universiti Malaysia Perlis. However, the data obtained may vary from one dataset to another since the applied e-nose and e-tongue are different in each experiment, which reflect different feature dimension, unequal sample size as well as different subject of sample. Since datasets used for this thesis is secondary data, the following assumptions are made throughout the research.

i. Data is presumed to comply with the appropriate data collection methods recommended for each sensor, as it is collected by the experts from the field. Therefore, the validity of sensor data is not argued in this thesis.

ii. The experiment for obtaining data from e-nose and e-tongue are performed separately. Therefore, such data sets are considered independent of each other.

iii. Signals based on the base line (offset of the calibration curve) are manipulated for the preprocessing of e-nose and e-tongue data, and the preprocessing of each sensor differed.

All related reviews of the above discussions are included in Chapter 2. The reviews are presented within three main subtopics; the electronic sensors, the need and some multi sensor data fusion frameworks, as well as the ongoing debates of the preference for multi sensor data fusion model; the next concerns are the feature selection issues in discriminant analysis where different selection criterion, approaches, strategies and some stopping rules are elaborated; and finally some existing techniques of parametric and nonparametric classifications are described.

Chapter 3 covers the research methodology applied in this study. The intended algorithms for every level of multi sensor data fusion are also included. The results and findings of the study are reported in Chapter 4. And finally Chapter 5 illustrates the conclusions and future works for improvement.

CHAPTER TWO
MULTI SENSOR DATA FUSION, FEATURE SELECTION AND CLASSIFICATION TECHNIQUES

2.1 The Electronic Sensors

The evolution of *e-nose* as an artificial olfactory is verified by the reported research on *odor* detection using an array of eight different electrochemical cells by Hartman and colleagues in 1964 (Phaisanggittisagul, 2007). However, only 20 years later, the development of an electronic instrument which is capable to detect and to recognize complex odors called e-nose is rationalized. The term "electronic nose" was first appeared in the literature around late 1980s (Gardner, 1988 in Gardner & Bartlett, 1999). Definition of e-nose by Gardner and Bartlett (1994) in Gardner et al., (1999) has generally been accepted as "an electronic nose is an instrument which comprises an array of electronic chemical sensors with partial specificity and an appropriate pattern recognition system, capable of recognizing simple or complex odors."

E-nose is an instrument which mimics the sense of smell (Peris & Escuder-Gilabert, 2009). In order to understand e-nose, it is useful to understand the basic components of smell process. Smell that constituted by an odor is stimulated in the human olfactory system that consists of three essential elements. Craven, Gardner and Bartlett (1996) described the system includes an array of olfactory receptor cells situated in the roof of the nasal cavity, the olfaction bulb which is situated above the nasal cavity, and the brain. Similarly, e-nose system is composed of three elements such as electronic sensor array, signal pre-processor and pattern recognition system. Figure 2.1 illustrates the basic components of the human olfactory system and an e-nose system.

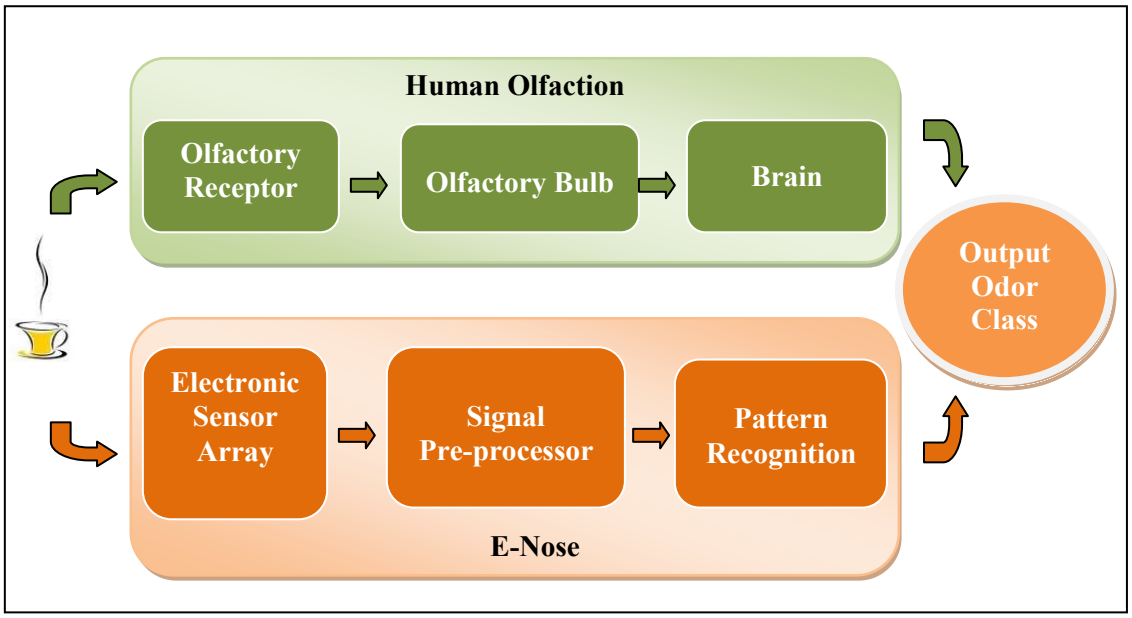

Figure 2.1. Typical Block Diagram of Human Olfaction and E-Nose

Ideally, an e-nose may attempts to mimic the original human nose, but its performance is still far behind compared to the capability of human nose (Schaller, Bosset & Escher, 1998); the sensors still present a number of weak points (Di Rosa et al., 2017). The first element of an e-nose is electronic sensor array as shown in Fig. 2.1, which is also known as the sampling unit. The array of electronic chemical sensors with partial specificity (Hine, Llobet & Gardner, 1999) is responsible to provide dynamic responses (i.e. electrical signals in the form of resistance change) resulting from the interactions between an odor sample (i.e. odorant molecules in the form of volatile compounds) and the sensing materials (Gardner et al., 1999; Peris et al., 2009; Phaisangittisagul, Nagle & Areekul, 2010). Since volatile compounds are responsible for the aroma of foodstuffs (García-González & Aparicio, 2002), appropriate sampling technique is important for the sensor array to generate good signal response that lead to better odor classification (Peris et al., 2009).

The success of e-nose to analyze gases has led to the development of an array of sensors that work in liquid surroundings (Rodríguez-Méndez, Apetrei & De Saja,

2010). It all started in 1982 from the work of Persaud and Dodd on e-nose that can make fine discriminations between complex odorant mixtures containing various ratios of odorants without the necessity of highly specialized peripheral receptors (Ciosek et al., 2004). Then, a system for *liquid* analysis was first described by Otto and Thomas in 1985 (Ciosek et al., 2004; Ciosek, Brzózka, Wróblewski, Martinelli, Di Natale & D‟Amico, 2005). However, the term *e-tongue* or artificial gustation became widely recognized in the late 1980s (Ciosek et al., 2004). The first multi-sensor system for liquid analysis was based on a poor selectivity approach introduced by Toko and coworkers from Kyushu University in 1990, which referred as the taste sensor (Winquist, Krantz-Rülcker & Lundström, 2003). The term taste sensor is rather specific in nature with the ability to respond to the basic tastes of human tongue such as sourness, sweetness, bitterness, saltiness and umami (Toko, 2000). Oliveri, Casolina and Forina (2010) suggested a straightforward description of e-tongue and taste sensor; the former term has a wider meaning, embracing all possible applications while the later is exclusively refer to sensory like evaluations.

According to Gutiérrez et al. (2011) a clear definition of e-tongue did not appear until 2004. They defined e-tongue as an analytical instrument comprising an array of nonspecific, poorly selective, chemical sensors with partial specificity (cross sensitivity) to different compounds in a solution and an appropriate pattern recognition or multivariate calibration tool, capable of recognizing the quantitative and qualitative composition of simple and complex solutions (Legin, Rudnitskaya, Lvova, Vlasov, Di Natale & D‟Amico, 2003). Unlike the e-nose systems, e-tongue is not only employed for recognition and classification, but also for quantitative determination of multiple component concentrations (Legin et al., 2003; Rudnitskaya et al., 2006). Major differences compared to e-nose lie in the applied sensors or

sensor systems and their tasks for measurement (Hauptman, Borngraeber, Schroeder & Auge, 2000). How does e-tongue works? E-tongue is an artificial taste system that imitates human taste system (Toko, 1996; Kovács, Sipos, Szöllösi, Kókai, Székely & Fekete, 2011). Therefore, they are developed based on the mechanisms of biological systems. In a gustatory system, substances producing taste are received by the biological membrane of gustatory cells in non-specific taste buds on the tongue. Information on taste substances is transduced into electric signals which are then transmitted along the nerve fiber to the brain where the taste is perceived (Escuder-Gilabert & Peris, 2010). Figure 2.2 describes basic components of e-tongue and the human taste system.

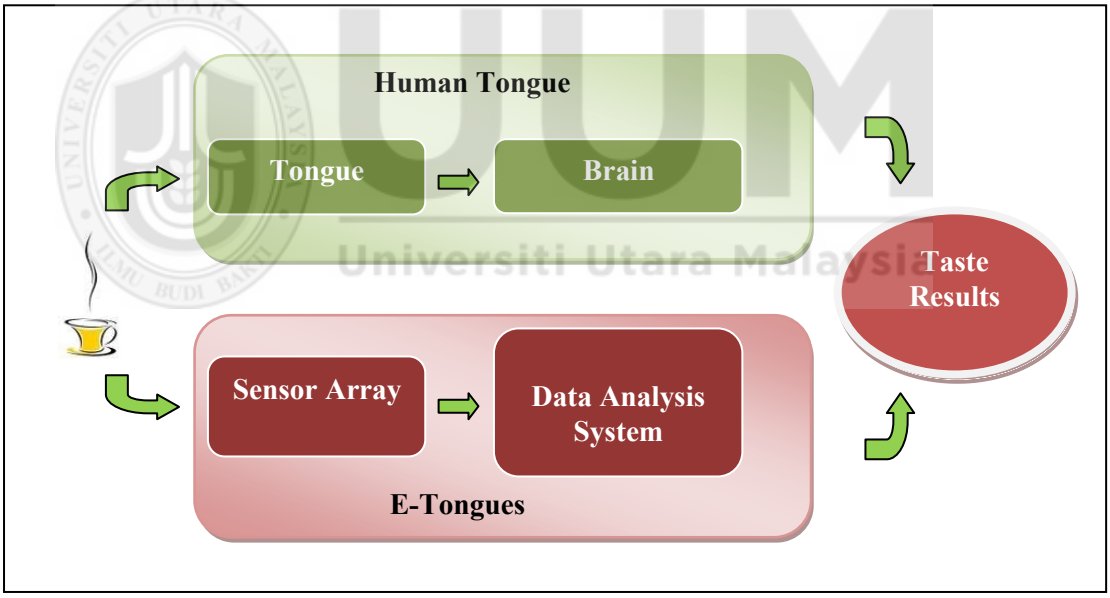

Figure 2.2. Typical Block Diagram of Human Tongue and E-Tongue

E-tongue systems compose of a sensor array and a data analysis system i.e. pattern recognition block (Ciosek & Wróblewski, 2011; Rodríguez-Méndez et al., 2010; Wei, Wang & Liao, 2009; Legin et al., 2003), or hardware and software components as defined by Jamal, Khan and Imam, 2009. Other researchers describe the components of e-tongues are of three elements including the sensors, sensor interface

and a processing unit (Abdul Aziz, Shakaff, Farook, Adom, Ahmad & Mahat, 2011; Ciosek & Wróblewski, 2006); or four elements consisting of automatic sampler, array of chemical sensors with different selectivity, instrumentation to obtain the signal, and software with appropriate algorithm to process the signal and get the required results (Escuder-Gilabert & Peris, 2010). However, for simplicity of explanation, the first description is preferable for this research.

Several successful applications of e-nose are proven in the food industry mainly related to the process monitoring, shelf-life investigation, freshness evaluation, authenticity assessment, and some other food quality control studies (Peris et al., 2009). Some specific applications include rice infestation (Zhou & Wang, 2011), food disease diagnosis and spoilage detection (Casalinuovo, Di Pierro, Coletta & Di Francesco, 2006), egg freshness determination (Dutta, Hines, Gardner, Udrea & Boilot, 2003), red meat spoilage classification (El Barbri, Llobet, Bari, Correig & Bouchikhi, 2008), classification of agarwood oil (Hidayat, Md. Shakaff, Ahmad & Adom, 2010), and classification of different brands of Espresso coffee (Pardo et al., 2000).

Jamal et al. (2009) exhibited that the hardware is used for the capacitance measurements of sensorial units, while the software controls the data acquisition process, perform calculations and analyze the electrical signals. An example of hardware components of e-tongues further described by Jamal et al. (2009) are the signal generator, signal amplifier, multiplexer, data acquisition board, and a computer, as well as the software component that deals with electrical signals and provide capacitance values for further analysis. The electrical signals produced are not necessarily specific for any particular species in the liquid, instead a signal

pattern is generated which can be related to certain features or qualities of the sample using appropriate pattern recognition software (Rodríguez-Méndez et al., 2010).

2.1.1 The Need for Multi Sensor Data Fusion

Multi sensor data fusion is an evolving technique related to problem of how to combine data from one or multiple (and possibly diverse) sensors in order to make inferences about a physical event, activity or situation. Mitchell (2007) defined multi sensor data fusion as the theory, techniques, and tools which are used for combining sensor data, or data derived from sensory into a common representational format. Also, the definition includes multiple measurements produced at different time instants, by a single sensor as described by Smith and Erickson (1991).

The applications of this concept span a wide domain including the military and nonmilitary applications. Though the military services of the United State of America pioneered the data fusion applications in the late 1970s, but only in the late 1980s a small numbers of military data fusion systems were in operation. Among the applications include ocean surveillance, air-to-air and surface-to-air defense, battlefield intelligence, surveillance and target acquisition, and strategic warning and defense. Nonmilitary applications include law enforcement, remote sensing, automated monitoring of equipment, medical diagnosis, and robotics (Hall, 1992).

Some other specific applications of multi sensor data fusion are in the area of multimodal biometric systems using face and palm-print (Raghavendra, Dorizzi, Rao & Kumar, 2011), renewable energy system (Li, Luo & Jin, 2010), color texture analysis (Wu, Li & Liao, 2007), face and voice outdoor multi-biometric system (Vajaria, Islam, Mohanty, Sarkar, Sarkar & Kasturi, 2007), personal authentication

(Duc, Bigun, Bigun, Maitre & Fischer, 1997; Kumar, Wong, Shen & Jain, 2006), image recognition (Sun, Zeng, Liu, Heng & Xia, 2005), medical decision making (Harper, 2005), and road traffic accidents (Sohn & Lee, 2003).

Further extensive applications of multi sensor data fusion are in the field of pattern recognition and classification of food science research. Food related industries have become more challenging ever as the demand increases each year. It is a common phenomenon occurs in every part of the globe where the food supply chain exists either in the form of fresh food or processed food products. As the demand arises especially for processed food products, certainly it imposes manufacturers to the so called efficient and immediate methods of monitoring products produced. Not only from the perspective of natural and human resources, applied technology, and business strategy, but beyond that is the product quality. In the case of processed food products, usually customers would perceive good quality product based on flavor, color, aroma, texture, nutrition, and microbial content (Korel & Balaban, 2009).

For that reasons, one of the biggest challenges to manufacturer is to preserve the intended quality of food products as it should be. In order to do that, a smart way to assess and maintain the quality parameters is required. Traditionally for the taste and aroma (or easily perceive as flavor), human panels or trained experts are responsible for the evaluation of quality parameters. However, this approach suffers from some drawbacks, for example; (i) discrepancy due to human fatigue or stress; (ii) time consuming; (iii) expensive (García-González et al., 2002); (iv) impossible for on line monitoring (Cole et al., 2011; Ghasemi-Varnamkhasti et al., 2010; Buratti et al., 2004); and (v) the evaluation perceptions depends on panelists‟ training (García-

González et al., 2002). Thus, alternative method to the sensory properties of a product involving taste and smell organoleptic panels for the quality parameter assessment is highly desirable (Cole et al., 2011; Ghasemi-Varnamkhasti et al., 2010; Zhang and Jia, 2007).

Flavor is generally understood to be the overall experience from the combination of oral and nasal stimulation and is principally derived from a combination of the human senses of taste and smell (Cole et al., 2011). The intention to overcome the drawback of human panel leads to the evolution of flavor sensor that combined the artificial olfactory (e-nose) and artificial gustation (e-tongue) (Craven et al., 1996; Toko, 1996). The application of e-nose and e-tongue begin in the early 1980s and 1990s respectively, where several commercial instruments have become available (Schaller et al., 1998) mainly to reduce the need for flavor panels (Cole et al., 2011) and to increase the performance of the measurement (Winquist, Lundström & Wide, 1999).

An increasing demand for more accurate estimation of the subject being studied through measurement and detection has directed the emerging technology to fuse data from multiple sensors. Since then, more extended research were carried out, and the application widespread in various applications (Raghavendra et al., 2011; Li et al., 2010; Wu et al., 2007; Vajaria et al., 2007; Kumar et al., 2006; Sun et al., 2005; Harper, 2005; Sohn et al., 2003; Duc et al., 1997), especially in the food industry (Cole, et al., 2011; Bruwer, MacGregor & Bourg Jr., 2007; Cosio et al., 2007; Buratti et al., 2004; Olafsdottir et al., 2004; Boilot, et al., 2003; Cimander, Carlsson & Mandenius, 2002; Rong et al., 2000; Steinmetz, et al., 1999).

Certainly combination of these sensors has the potential to mimic the human flavor panels since measurement data from the sensors are manipulated to produce sensor-specific opinions about the human-like sensing modalities (Wide, et al., 1998). This has been previously suggested by Hall (1992) as he described multi sensor data fusion as the technique consists of algorithms, methods and procedures to assemble data of different origin and nature in order to increase the information about an event or a phenomenon. Di Natale et al. (2000) also agreed that the utilization of such system is expected to enhance the amount of information extracted from a sample.

Several applications of multi sensor data fusion applied in the food related issues include flavor sensing system (Cole et al., 2011), storage condition of olive oil (Cosio et al., 2007), variation of snack food textural measurement (Bruwer et al., 2007), quality control of yoghurt fermentation (Cimander et al., 2002), wine analysis (Buratti et al., 2004; Rong et al., 2000), determination of fish quality (Olafsdottir et al., 2004), discrimination of standard fruit solutions (Boilot et al., 2003) and fruit quality assessment (Steinmetz et al., 1999).

2.1.2 Multi Sensor Data Fusion Model

There are several multi sensor data fusion frameworks that have been developed for different kinds of applications. Among them are JDL process model, Thomopoulus architecture, multi sensor integration fusion model, behavioral knowledge-based data fusion model, waterfall model, distributed blackboard data fusion architecture, as well as Omnibus data fusion model (further reading please refer to Esteban, Starr, Willetts, Hannah & Bryanston-Cross, 2005). One of the earliest and mostly applied data fusion frameworks is the JDL process model, developed by the U.S. Joint Directors of Laboratory (JDL) Data Fusion Working Group (DFS) (Hall & Llinas,

1997). The process model is mainly being researched by the U.S. military services beginning the late 1970s to support the developments of military applications (Huang, 2009; Esteban et al., 2005). The JDL-DFS group was established in 1986 to unify the terminology and procedure in data fusion, where eventually the JDL process model was developed (Huang, 2009) and has gain attention from multidisciplinary researchers.

The JDL process model is a conceptual model which identifies the process, functions, categories of techniques, and specific techniques applicable to data fusion through different function levels. In 2004, the process model was revised. The data fusion function levels is partitioned into five data fusion function levels (Huang, 2009); level 0 is the signal/feature assessment, level 1 is the entity assessment, level 2 is the situation assessment, level 3 is the impact assessment, and finally level 4 is the performance assessment (Huang, 2009; Hall et al., 1997). Hall (1992) has discussed the above functions and introduced mathematical techniques for fusing data using data fusion algorithms which serve as the reference for both military and nonmilitary applications. Extensive discussions of the data fusion techniques are refereed to Hall (1992).

A typical application of multi sensor data fusion is called multi sensor identity fusion. Generally, in identity fusion, the interest is to combine identity declaration data from multiple sensors to obtain joint estimate of identity which is more specific and accurate, than any of the declarations from an individual sensor. In order to implement the data fusion system, one of the key issues is to decide where in the data flow does the fusion takes place. Usually, fusion could occur either at the raw data level (prior to feature extraction), at the feature vector level (prior to identity

declaration), or at the decision level (after each sensor has made an independent declaration of identity) (Hall, 1992). The initial concept is commonly known as low level data fusion or data level fusion, followed by intermediate level data fusion or feature level fusion, and the final is high level data fusion or decision level fusion.

Before further discussions on different levels of data fusion, it is helpful to understand some common processes involve in each level such as association, feature extraction, as well as identity declaration. Association is performed towards the raw data to ensure data being fused relates to the same object. Feature extraction transforms the data output from sensor into a well-organized representation and simplified data format for much complex analyses. Details on the application of each identity fusion level are described in the following sections 2.1.2.1 to 2.1.2.3. Further reading is referred to Hall (1992). Another complete yet current dimension of multi sensor data fusion and can be found in Mitchell (2007). The later fusion model is not much covered in this thesis since the application of the model is less employed in the food research.

2.1.2.1 Low Level Data Fusion

The first model of multi sensor data fusion is the low level data fusion (LLDF), which implies concatenation of raw data (features) from similar or different sensors (Rudnitskaya et al., 2006; Di Natale et al., 2000; and Winquist et al., 1999). The resulting data matrix, $\left(\mathbf{X}_{N\times P}\right)$ has *n* rows representing the number samples from different groups and *p* column signifying the total number of features from all the sensors. It combines the signals provided by the respective sensors to produce new data that is predicted to be more informative before any further processing begun.

In order to perform low level data fusion, the sensors must be identical, in such a way that the signals must be commensurate for the data can be combined together. After data are fused, the data will undergo an extraction process usually by means of principal component analysis (PCA). PCA allows visualization most of the information contained in a raw data matrix in few dimensions called principal components (PCs), which are orthogonal to each other (Vera et al., 2011). The PCA calculation on $\mathbf{X}_{N\times P}$ is achieved via an eigenvector decomposition of the corresponding covariance matrix, $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ (Bruwer et al., 2007). Here, only the eigenvectors corresponding to the dominant eigenvalues are kept for classification. The whole process of LLDF is outlined in Figure 2.3.

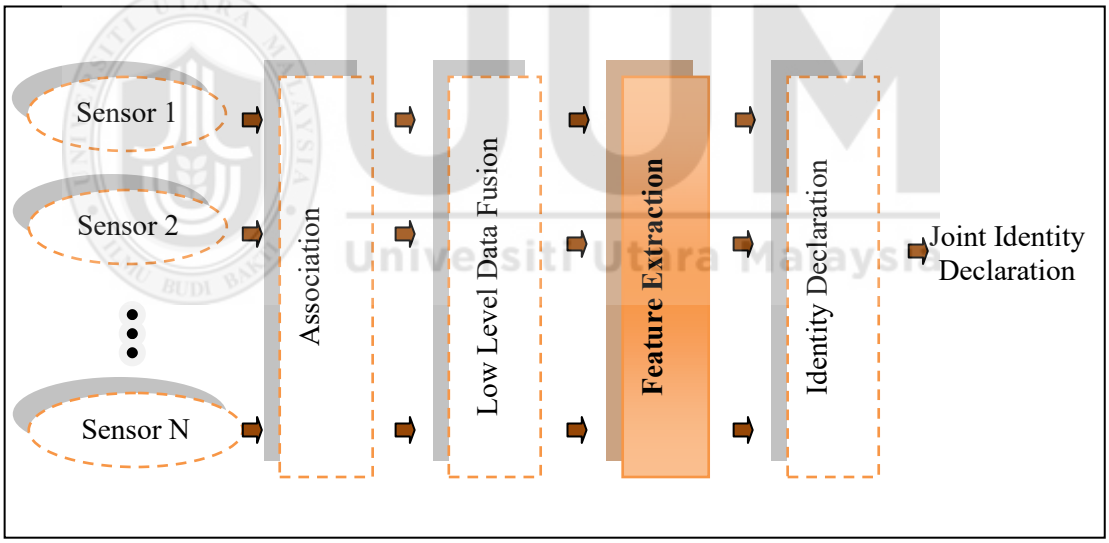

Figure 2.3. Framework of Low Level Data Fusion (Adapted from Hall, 1997)

Some researches that employed this model include classification of pure and/or adulterated honey (Masnan et al., 2012), discrimination and sensory description of beers (Vera et al., to, 2011), classification of different honey samples (Zakaria et al., 2011), classification of orthosiphon stamineus (Zakaria et al., 2010), multimodal characterization of red wines using three sensory modalities (Rodríguez-Méndez et

al., 2004), discrimination of white grapes varieties (Roussel et al., 2003) and some others. Table 2.1 summarizes the application of e-nose and e-tongue and/or other sensors as well as its general remarks. The application of other sensors is also described in Table 2.2.

Table 2.1

Summary of Studies for Fusion of E-Nose and E-Tongue and/or Other Sensors Using LLDF

Authors	**Sample**	**Classification Methods**	**General Remarks LLDF**
Masnan et al., (2012)	Honey	PCA, LDA	Classification of the intermediate level is better than low level.
Prieto et al., (2011)	Wine	PCA, PLS-DA	Discrimination result is significantly improved when signals are fused
Gil-Sánchez et al., (2011)	Wine	PCA, Cluster	Sensor fusion able to monitor evolution of wines as a function of day
Zakaria et al., (2011)	Honey	PCA, LDA, PNN	Classification of fusion method is better than single sensor
Apetrei et al., (2010)	Virgin olive oil	PCA, PLS-DA	Discrimination and prediction is improved by fusing three sensors
Zakaria et al., (2010)	Tea	PCA (for EDA), LDA	Classification of fusion is better than single sensor
Cosio et al., (2007)	Extra virgin olive oil	PCA (for EDA), LDA	Classification performance of fusion system is the same as classification using e-nose
Buratti et al., (2004)	Wine	PCA, LDA, CART	Better classification obtained using e-nose and e-tongue. Chemical and color analysis are not helpful.
Di Natale et al., (2000)	Urine, milk	PCA	Cooperation of e-nose and e-tongue improves classification performances
*Sundic et al., (2000)	Potato chips, potato cream	Fuzzy logics, ANN	1. Considerable improvement of classification performance for fusion 2. Classification performance for fusion is highly correlated to the increase of e-nose performance

Table 2.1 Continued

*Winquist et al., (1999)	Different juices	PCA, Proj. latent structure	Slight improvement when fuse the sensors
*Wide et al., (1998)	Different juices	PCA, ANN	Fusion technique improves the system performance

Table 2.2

Summary of Studies for Fusion of Other Sensors Using LLDF

Authors	**Applied Sensors**	**Sample**	**Classification Methods**	**General Remarks LLDF**
Aranda-Sanchez et al., (2009)	Non-destructive acoustic impact technique & colorimeter sensors	Tomato	Bayesian classifier	Data fusion helps to develop an optimum system for more accurate sorting decisions.
Boilot et al., (2003)	4 different e-noses	Juices-apple, pear, peach	PCA, GA, PNN	Classification of fusion method is better than single modality.
Cimander et al., (2002)	E-nose, NIRS, Bioreactor probes	Yoghurt fermentation	ANN, PCA	Sensor fusion improves neural network stability and reliability.

2.1.2.2 Intermediate Level Data Fusion

Intermediate level data fusion (ILDF) is simply the fusion of preferred extracted features from different sensors. In this fusion level, extracted features from several data sources or from a multiple representation of single data source of different time periods are chosen (Rudnitskaya et al., 2006). The extracted feature matrices are concatenated to form a single extracted feature matrix as an input for classification purposes. The resulting merged data matrix is three-dimensional; the first dimension equal to the number of samples from different groups, the second and third dimensions are the number of extracted features from the first and second sensors.

The extracted features describe above comes from the calculation of covariance matrix $\mathbf{X}^T\mathbf{X}$ that is performed to each sensor data. The resulting eigenvectors corresponding to the dominant eigenvalues from both sensors are fused before further classification. Figure 2.4 displays the process of ILDF. Meanwhile, the most common feature extraction technique that is usually associated with the dimension reduction process is PCA.

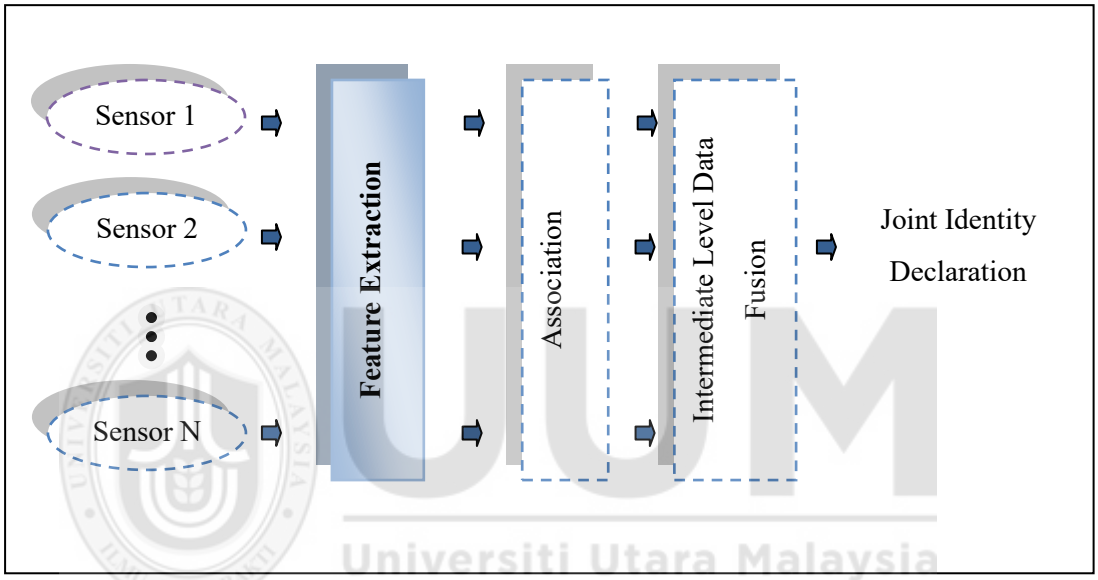

Figure 2.4. Framework of Intermediate Level Data Fusion (Adapted from Hall, 1997)

Several researchers who applied this level include classification of pure and/or adulterated honey (Masnan et al., 2012), Vera et al., (2011) in the discrimination and sensory description of beers, Guru et al., (2010) in an attempt of fusing covariance matrices of PCA and Fisher linear discriminant (FLD), and Steinmetz et al., (1999) in the fusion methodology for fruit quality assessment and others. Different remarks were given for the fusion of e-nose and e-tongue using ILDF and/or LLDF as illustrated in Table 2.3 and in Table 2.4 for fusion of other sensors using ILDF and/or HLDF.

Table 2.3

Summary of Studies for Fusion of E-Nose and E-Tongue Using LLDF and/or ILDF

Authors	Sample	Classification Methods	General Remarks ILDF
Masnan et al., (2012)	Honey	PCA, LDA	Classification of the intermediate level is better than low level.
Vera et al., (2011)	Beer	LDA	1. ILDF is better than LLDF 2. Classification using fusion method is better than single sensor
Rodríguez-Méndez et al., (2004)	Wine	PCA	Discrimination of fusion system is significantly improve than single sensor
Rong et al., (2000)	Wine	Fuzzy NN	Classification using fusion approach is better

Table 2.4

Summary of Studies for Fusion of Other Sensors Using ILDF and/or HLDF

Authors	Applied Sensors	Sample	Classification Methods	General Remarks ILDF
Guru et al., (2010)	Secondary data	Video shoots, iris, zoo and wine dataset	PCA, FLDA	Performance of fusion system improved by fusing classifiers during feature extraction phase.
Steinmetz et al., (1999)	Destructive and nondestructive sensors	Melon, peach	PCA	Intermediate level is better than high level. And the choice of sensor fusion method is a difficult task.

2.1.2.3 High Level Data Fusion

In high level data fusion (HLDF), each sensor performs a transformation to obtain an independent declaration of identity. More or less, the extraction process in the ILDF is continued for separate classification analysis. The identity declarations from each

sensor are subsequently fused. Identity declaration can be of two forms, identity result (hard fusion) and scores (soft fusion) (Huang, 2009). Either one of the forms is used as inputs for the decision fusion. Among the techniques for fusing declaration of identity include voting methods, Bayesian inference, Dempster-Shafer"s method, generalized evidence processing theory (Hall, 1992) and knowledge based expert system (Huang, 2009). Figure 2.5 shows the framework of HLDF.

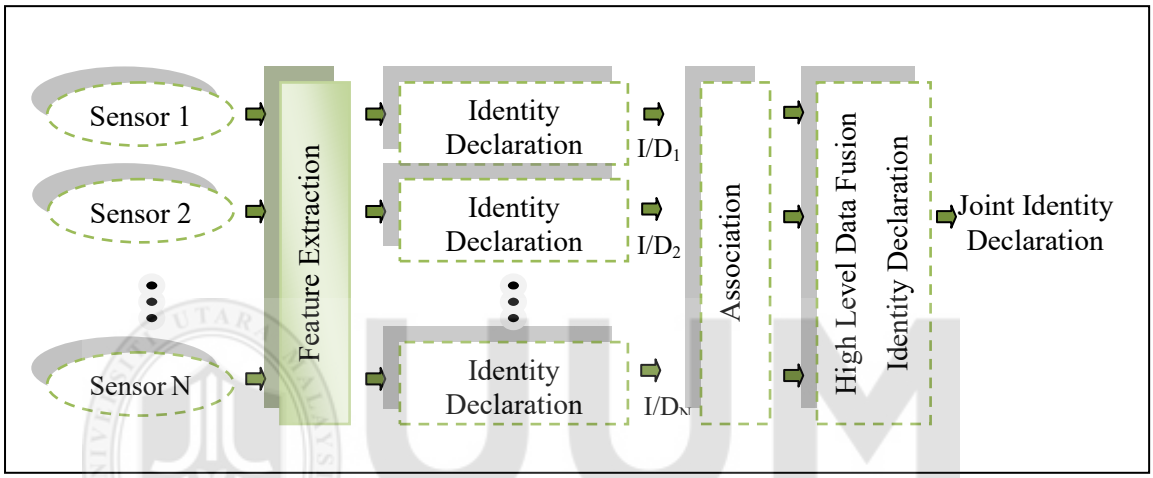

Figure 2.5. Framework of High Level Data Fusion (Adapted from Hall, 1997)

Related researches applying HLDF include an attempt in increasing predictive performance of high level data fusion (Doeswijk et al., 2011), threshold-optimized decision level fusion applied in biometrics (Tao, & Veldhuis, 2009), assessing apple quality using three sensors (Xiaobo & Jiewen, 2005), discrimination of white grapes varieties (Roussel et al., 2003) and fusion methodology for fruit quality assessment (Steinmetz et al., 1999).

It can be concluded from Table 2.5 that most researchers agreed with the fusion of e-nose and e-tongue, whom some are with additional sensors, would provide better classification results compared to the utilization of single sensor. However, in certain cases, additional sensors other than e-nose and e-tongue may not be helpful such as

in the case of Buratti et al., (2004). Other than that, we may also observe that the performance of e-nose sensor is dominant than e-tongue, in contrast to the fusion method. This scenario is revealed from the work by Cosio et al., (2007), and in the case of Sundic et al., (2000) where they admit that the improved performance of fusion method is correlated to the increase performance of e-nose. Nonetheless, none of the previous works that involves e-nose and e-tongue has venture the HLDF.

Table 2.5

Summary of Studies for Fusion of Other Sensors Using HLDF, ILDF and/or LLDF

Authors	**Applied Sensors**	**Sample**	**Classification Methods**	**General Remarks**
Doeswijk et al., (2011)	Not mentioned	Tomato	PLSDA	High level data fusion should always be considered for predictive purposes.
Rudnitskaya et al., (2006)	E-tongue, FTIR	Apple	PCA, PLS DISCRIM, PLS regression	Best results for discrimination of apple varieties and determination of organic acid content were obtained using LLDF
Xiaoboe et al., (2005)	E-nose, machine vision, NIR (spectrophometer)	Apple	ANN	High level fusion is able to provide more information for quality assessment.
Roussel et al., (2003)	Aroma sensors, FTIR, UV spectrometer	White grapes varieties	Bayesian classifier	High level fusion improves the white grape classifications than the low level.
Steinmetz et al., (1999)	Destructive and nondestructive sensors	Melon, peach	PCA	Intermediate level is better than high level.

2.1.3 Discussions of LLDF, ILDF and HLDF

So far we have discussed three types of data fusion model that can be applied when dealing with different sensors data. Each model offers different architecture of fusing different sensors data. Further investigations on other sensor mechanisms than e-nose and e-tongue that engage different data fusion models in food research domain were also observed. Generally, we noticed more researchers have broadened their exploration to low, intermediate and/or high level fusion. However, the verdict for the best data fusion model especially involving e-nose and e-tongue is still inconclusive. For example, Rudnitskaya et al., (2006) claimed LLDF is better, but Roussel et al., (2003) to the contrary favored the HLDF than the other fusion levels. In contrast, Masnan et al. (2012) discovered ILDF is better than LLDF. Recognition was also given to the HLDF for the ability to provide more information (Xiaobo and Jiewen, 2005) and its predictive reason (Doeswijk et al., 2011). The disagreement among researchers regarding the most suitable fusion model to be applied for food discriminant using sensor mechanisms, and specifically in the context of e-nose and e-tongue will be discovered by this research.

Earlier, Steinmetz et al., (1999) have forwarded their concerns on the difficulties of choosing the right fusion model. Even though they admit LLDF as the most efficient model, this model can only be used when identical or commensurate sensors are applied. They further commented that ILDF and HLDF methods are well adapted to practical cases with different sensors. Unfortunately, these models allow loss of information with respect to raw signal provided by each sensor where errors are transmitted in the fusion process. According to Gigli, Bossé and Lampropoulos, (2007) HLDF is preferred for its feasibility, most tolerant to individual errors in a data stream subsystem, has lower computational complexity than ILDF, and due to

its low coupling of information, it is more robust to the removal or addition of individual data sources. However, the main disadvantage of HLDF is that information lost from a lower level fusion cannot be recovered at a higher level. Besides, the HLDF model is much concern on the process of voting some classification rules which more suitable in machine learning rather than statistics.

Another concern that is of importance with regards to the data fusion levels is the phase of feature extraction in each model. Hall (1992) did not discuss much about feature extraction stage from the statistical point of view except in section 5.2 (pp. 138-144). However his clarifications regarding extracted features for identification process from different dimensions and applications are significant for deeper understanding. As described in the previous studies as well as reviews in sections 2.1.2.1, 2.1.2.2 and 2.1.2.3 of this thesis regarding LLDF, ILDF and HLDF, the most commonly applied feature extraction method in every data fusion level involving is obviously PCA. Other research such as Zakaria et al., (2014) used LDA as a tool to extract features. This practice can be observed from the summary of studies illustrated in Tables 2.1 and 2.2 for LLDF, Tables 2.3 and 2.4 for ILDF and Table 2.5 for HLDF. Most studies related to sensor data fusion employed PCA as the dimension reduction technique. To name a few, it includes Byrne, O‟‟sullivan, Bredie, Anderson and Martens, (2003), Faber, Mojet and Poelman, (2003), Gimeno, Ansorena, Astiasarán and Bello, (2000), Hansen, Petersen and Byrne, (2005), Thybo, Kühn and Martens, (2003), as well as Zamora and Guirao (2004).

There are some issues to be highlighted besides the preferable application of feature extraction using PCA in the JDL data fusion model. PCA is the most recognized dimension reduction tool (Mallet, Coomans & de Vel, 1996) for data that exhibit

multi-collinerity (Masnan et al., 2012; Bruwer et al., 2007; Steinmetz et al., 1999). Despite its popularity, PCA approach is confusing and sometimes is indefinite in nature. PCA can be calculated either using the covariance or correlation matrix (Masnan et al,. 2012, pp. 5-6), where the covariance matrix is suitable for features that have merely equal variance, while correlation matrix is suitable for features with unequal scales of measures. Jolliffe (2002, pp. 22-26) elaborated these concepts in detail. According to him, a major argument for using correlation rather than covariance matrices to define PCs is that the results of analyses for different sets of random variables are more directly comparable than for analyses based on covariance matrices. Misconception may occur if these distinctions are misunderstood. Most reviewed studies in LLDF, ILDF and HLDF concealed these issues for discussions except Bruwer et al. (2007) whom apply covariance matrix as well as Zakaria et al. (2010 & 2011) and Masnan et al. (2012) whom exploited correlation matrix.

However, the most haziness issue in using PCA as dimension reduction method is in determining the number of principal components to retain. There are commonly three indicators to decide; average eigenvalue, elbow at a scree graph and/or proportion of total variance explained. Joliffe (2002) claimed these ad hoc rules-of-thumbs are intuitively plausible in justification. The average eigenvalue rule or sometimes called Kaiser‟s rule is constructed specific for use with correlation matrices but can be adapted for some types of covariance matrices (Jolliffe, 2002). Further explanation for such situation is referred to Jolliffe (2002, p. 115). According to Rencher (2002), by using this rule, retain associated PCs whose eigenvalues (λ s) are greater than the

average of the eigenvalues, $\sum_{i=1}^{p} \lambda_i / p$, where for a correlation matrix the average is 1.

Although this indicator is widely applied and the default in many software packages, Rencher (2002) criticized that this indicator keeps those components that account for more variance than the average variance of the variables, which caused a wide gap between two eigenvalues that fall on both sides of the average. This rule was rarely utilized in the reviewed case of LLDF, ILDF, and HLDF except in Masnan et al. (2012). Less attention was also given to the scree graph approach in the discussion of multi sensor data fusion. Rencher (2002) recommended retaining those eigenvalues in the steep curve (elbow) before the first one on the straight line. Alas, in practice, the turning point between the steep curve and the straight line may be indistinct or there may be more than one discernible bend. Probably this approach is the least applied among all and may only be used to confirm the number of principal components selected as reported in Masnan et al. (2012).

The third indicator, the proportion of variance explained, is the most frequently employed approach in extracting features in LLDF, ILDF and HLDF studies. Here, the number of principal components that are kept is depending on the cumulative percentage of total variation which one desires where the selected principal components contribute say 80% or 90%. Even though the challenge of this method lies in selecting an appropriate threshold percentage (Rencher, 2002), Jolliffe (2002) notified a sensible cutoff is very often in the range of 70% to 90%, or it can sometimes be higher or lower depending on the practical details of a dataset. But, Rencher (2002) further warned that too high percentage would run the risk of retaining components that are either sample specific (a component may not

generalize the population or sample) or variable specific (a component is dominated by a single variable and does not represent a composite summary of several variables). The inconsistencies of appropriate cutoff percentages to retain associated principal components can be observed in the previous studies. Table 2.6 signifies the discrepancies. Some of the reviews may involved few cases in their studies such as in Vera et al., (2011), Apetrei et al., (2010) and Rudnitskaya et al., (2006) and some others do not even reveal their threshold percentages for instance Rudnitskaya et al., (2006) and Cimander et al., (1999).

Table 2.6

Varieties of Selected Proportion of Total Variance Explained and Number of Retained Principal Components Used by Different Researchers

Authors	**Fusion Level**	**% Total Variance Explained**	**Number of Retained Principal Components**
Vera et al., (2011)	LLDF	Case 1 : 65%	Case 1 : 7
		Case 2 : 73%	Case 2 : 10
		Case 3 : 82%	Case 3 : 15
	ILDF	Case 1 : 97%	Case 1 : 2
		Case 2 : 96%	Case 2 : 4
		Case 3 : 62%	Case 3 : 1
Prieto et al., (2011)	LLDF	60%	3
Gil-Sánchez et al., (2011)	LLDF	69%	2
Zakaria et al., (2011)	LLDF	Greater than 80%	3
Apetrei et al., (2010)	LLDF	Case 1 : 63%	Case 1 : 3
		Case 2 : 59%	Case 2 : 3
Zakaria et al., (2010)	LLDF	Greater than 80%	2
Cosio et al., (2007)	LLDF	61%	2
Rudnitskaya et al., (2006)	LLDF	LLDF : unmentioned	LLDF : unmentioned
	ILDF	ILDF : unmentioned	ILDF : 5 & 7
	HLDF	HLDF : unmentioned	HLDF : 8
Rodríguez-Méndez et al., (2004)	ILDF	90%	2

The indistinctness issues in retaining the appropriate number of principal components for dimension reduction applied in multi sensor data fusion particularly involving e-nose and e-tongue has become a real challenge. Even though there is some flexibility upon the techniques for deciding the correct number of principal components for further classification, the indefinite nature of the accurate threshold to be used limiting the prospect of PCA as a feature extraction tool.

Additionally, comments and critiques pointed out by Héberger and Andrade (2004) as well as Wood et al. (2005) in their articles that the significance of features selected by PCA is not known and the greater percentage of total variation does not automatically exemplify that the selected PCs should provide good features for discrimination. These discoveries are some evidence of PCA's disadvantages. Furthermore, when it comes to interpretation of certain features, PCA is lack of interpretability (Fraiman, Justel & Svarc, 2008; and Jolliffe, 2002). Fraiman, et al. (2008) further added that the resulting linear combinations of features in PCA are difficult to interpret unless most of the coefficients of the linear combination are not significant. Thus, studies on feature selection as an alternative to dimension reduction tool in multi sensor data fusion is the core of this research study and are extended in the following section 2.2.

2.2 Feature Selection

Dealing with sensor fusion means more features are being considered for the construction of classifier especially for the LLDF model which results to considerably small sample size against large number of features. The rise of feature dimensionality spaces usually leads to a problem known as curse of dimensionality which was introduced by Bellman (1961). When this occurred, an enormous number

of samples are required to perform accurate predictions for problems with high dimensionality. Classification within these conditions may lead to inaccurate parameter estimation, and the inclusion of too many features may harm the performance of a sample classifier (Foithong, Pinngern & Attachoo, 2012; McLachlan, 1992; and Achariyapapaopan & Childers, 1985). Chandrashekar and Sahin (2014) added if a system uses irrelevant features, it will use this information for new data leasing to poor generalization. In such situations, consideration might be given to the extraction of some important features using feature extraction which turn out to be unfavorable as illustrated in section 2.1.3, or to employ subset of available features using feature selection.

Feature selection refers to the procedure of selecting p relevant features based on certain criteria from a set of k original sensor features, $P_{\mathcal{P}} = \{p_1, p_2, \ldots, p_k\}$, for p where $p < k$. It is a process of identifying the most useful features in describing differences among the possible groups (McLachlan, 1992). According to Wankhande et al. (2013) and Dernoncourt, Hanczar and Zucker (2014), feature selection is the process of removing irrelevant features in reducing the dimensionality of data to be processed, decreasing the execution time and improving the predictive accuracy of the classifier (Chandrashekar & Sahin, 2014; Marra & Wood, 2011; Nakariyakul & Casasent, 2009; and Kanal & Chandrasekaran, 1971). Generally, the goal of feature selection is to determine the right (Wankhande et al., 2013) and optimal subset of features (Achariyapapaopan et al., 1985, Zhang & Sun, 2002; and Schulerud & Albregtsen, 2004) that maximizes the information contents or predictive accuracy. Perhaps, the intentions of performing feature selection illustrated by McLachlan (1992) are reasonable and straightforward; if the objective of a study is to form a

discriminant rule for the prediction of unclassified subjects, the error rate is preferred for feature subset evaluation. Alternatively, if the aim of a study is to identify the useful features in discriminating different groups, then the appropriate measure for feature selection is the separation among groups provided by the feature subset.

More researchers are convinced with the benefits offered by the concept of feature selection especially as an alternative to the feature extraction technique. For example, Youn (2004) outlines several advantages for the feature selection to be the preference in the classification of microarray dataset. With a generalization to multi sensor, the benefit includes (i) feature selection identifies relevant features which imply the subset of discriminating array sensors, and (ii) feature selection gives a better generalization error since only relevant features are included for classification. The selection of the most informative feature set leads to an improvement in the classification accuracy (Pechenizkiy, 2005), faster and more cost-effective classification performance, and better understanding of the underlying process of the observed dataset (Dash & Liu, 1997; Kabir, Islam & Murase 2010; Rueda, Oommen & Henriquez, 2010; Vergara & Llobet, 2011; and Li, Wu, Li & Ding, 2013).

The subject of feature selection has been applied to various conditions such as supervised, semi-supervised or unsupervised cases (Yen, Chen & Lin, 2010; Dy, 2008; Liu & Motoda, 2008), causal feature selection (Guyon, Aliferis & Elisseeff, 2008), as well as weighting and local methods (Kononenko & Šikonja, 2008; Huang, Xu, Ng & Ye, 2008; Domeniconi & Gunopulos, 2008; and Sun, 2008). Supervised feature selection is preferable when predictors are specifically selected for the purpose of increasing accuracy or to find a subset of predictors to reduce the complexity of the model. On the contrary, when the outcome is ignored during

elimination of predictors, the technique is called unsupervised feature selection. These details are available in Kuhn and Johnson (2013). Among the latest expansion of feature selection is the group feature selection that is employed in the area of data stream. Data stream is a sequence of digitally encoded coherent signals (packets of data) used to transmit or receive information that is in the process of being transmitted.

Broadly speaking, feature selection aims to choose those observed features that are most discriminative for classifying subjects to their correct groups. These subset of features or sometimes known as discriminators refers to a subset of the original features that is able to distinguish the structure of the existing groups and is used to construct a classification rule (Mahat, 2006). Further discussions are available in Li et al., (2013) and Wankhande et al., (2013). Perhaps a straightforward feature selection method contains the following steps (Dash & Liu, 1997); (i) generation procedure for choosing appropriate discriminators, (ii) evaluation function for choosing the best possible discriminators, (iii) stopping criterion to stop the selection process, and finally (iv) validation procedure to finalize the final set of discriminators. The last step, however, is not considered as part of the feature selection process. Hence, literatures and discussions will be focusing to the first three steps.

2.2.1 Feature Subset Generation Procedure

The generation procedure is a searching procedure to generate subsets of features that meet the goal of searching (Dash & Liu, 1997). Commonly feature selection methods are divided into two; *deterministic* and *stochastic* feature selections. *Deterministic* methods include branch and bound which also known as exponential

search, and sequential method. Even though branch and bound method gives optimal feature subsets, the application is impractical since the complexity grows exponentially with the number of features, even for moderate number of features (Vergara et al., 2011). Mallet et al. (1996) and Dy (2008) commented that this method is inefficient if the number of features is 30 or more. In addition, this procedure exploits much time and only work with a monotonic criterion function. Further explanation of such method is referred to Fukunaga (1990, pp. 491-503) and Web (2002, pp. 312-313).

Meanwhile, *stochastic* feature selection or known as randomized search algorithms attempts to overcome the computational costs of exponential methods. Stracuzzi (2008) claimed that this method would be useful when the space of possible feature subsets is prohibitively large and when the deterministic feature selection algorithms are prone to get trapped in local optima. The techniques include *genetic algorithms* (GA), *simulated annealing* (SA) and the *greedy search* procedures including forward, backward and stepwise searches. *Genetic algorithms* (GA) are modeled on the principle of evolution, developed by Siedlecki and Sklansky (1989), can also be used to find subset of features (Chandrashekar & Sahin, 2014). In GA approach, a given feature subset is represented as binary string (a "chromosome") of length *P*, with a zero or one in position *i*, $i = 1, 2, \ldots, P$ denoting the absence or presence of feature *i* in the set (Jain & Zongker, 2002). GA adopts the crossover and mutation as search mechanisms to randomly search for a good solution (Nakariyakul & Casasent, 2009).

Caution is given to the selection of parameter for the algorithm to perform well, and, Nakariyakul and Casasent (2009) reported previous comparative study showed that

the performance of GA degraded as the number of original features *P* increases. Jain and Zongker (2002) criticized this method does not attempt to find best subset of a specified size and it is hard to find the overall best subset since the chromosome score is heavily influenced by the subset size. According to Vergara et al. (2011), GA have been shown to be able to solve optimization problems by exploring all regions of the potential solution space and exponentially searching promising areas through mutation, crossover and selection operations applied to individual (chromosomes) in a population (set of possible solutions). Although GA is useful for selecting features, Jouan-Rimbaud et al. (1996) in Vergara et at. (2011) have shown that the solution found by GAs should be investigated carefully because the algorithm does not prevent meaningless features from being selected.

SA is a stochastic technique that randomizes the search procedure and applying sophisticated control strategies. It applies larger step-widths at the beginning of the minimum search procedure to prevent the minimum search from being trapped in smaller local minima (Schürmann, 1996, p. 204). Unfortunately, Vergara et al. (2011) remarked these techniques are well suited to find a global optimum of feature selection problem, but at the cost of lengthy computation.

The sequential methods significantly reduce the number of trials to be performed during the search by applying local search. This method is based on the previously known forward and backward feature selection (Liu & Motoda, 2008), where in forward selection one starts with an empty feature subset and adds relevant features into the subset according to a procedure; while for backward selection begins with a full set of features and removes an unimportant feature procedurally. Since both

searches might be too costly for large number of features, hence more efficient algorithms are developed (Liu & Motoda, 2008) called sequential strategies.

This method is considered as suboptimal procedure (Schulerud & Albregtsen, 2004). The technique generally uses greedy techniques where it assesses a set of potential feature combination which avoids examining every feature combination. This makes sequential methods much simpler and faster than the deterministic and stochastic searching methods. One consequence of this approach is the risk of getting trapped in local minima of the search space (Vergara et al., 2011; Dy, 2008). Among the most popular methods include forward selection, backward selection and stepwise selection. There are also different kinds of suboptimal selection techniques, and readers are suggested to refer Web (2002, pp. 314-317).

2.2.1.1 Forward Selection

Forward selection is a bottom-up searching procedure that adds a new feature to a feature set one at a time (Jain & Zongker, 2002), until the final features set is reached (Web, 2002). All features are first examined in turn to find the one that optimizes the chosen criteria (Hastie et al., 2009). After the single best feature (from *k* features) is chosen (called as the discriminator), then it will be paired with each of the remaining (*k* - 1) features so that the pair which give the optimized determined criterion could be identified. Next, the chosen paired of discriminators will be paired with each of the remaining (*k* - 2) features in attempt to find a set of triple features that optimize the criterion. Such process continues until *p* discriminators have been selected which involve $k+(k-1)+\ldots+(k-p+1)=\frac{1}{2}p(2k-p+1)$ criterion function evaluations (Krzanowski, 2000). This method is computationally attractive, but it suffers from

the nesting effect i.e. the subset of the four best features chosen must contain the subset of the three best features and so on (Schulerud & Albregtsen, 2004; Nakariyakul & Casasent, 2009). Guyon and Elisseeff (2003) agreed that this search is computationally more efficient than backward search, but they commented weaker subsets are obtained using this search because the importance of features is not assessed in the context of other variables not included yet.

2.2.1.2 Backward Selection

Backward selection or rather known as backward elimination is a top-down analogy of the forward selection (Jain & Zongker, 2002). From the complete set of original features, each feature which has least impact is omitted in turn (Hastie et al., 2009). The process of omitting features repeated from subset of (k - 1), (k - 2), … based on the intended optimized criterion. The process continues until p discriminators are left, and this involves $\frac{1}{2}p(2k-p+1)$ criterion function evaluations (Krzanowski, 2000). In forward selection, once a discriminator is selected in the retained set, it will remain in the set. Similarly, the backward elimination would not consider adding the discriminator that has been removed. In order to overcome such rigid selection of features, the stepwise selection combines the forward selection and backward elimination on the current set of discriminators is possible. However, this selection approach suffers the nesting effect as in the forward selection.

Jain and Zongker (2002) commented that forward selection is faster than backward. They further commented that the different performance of forward and backward is expected since the forward selection starts with small subsets and enlarges the subsets, while backward starts with large subsets and shrink them. And it is

computationally expensive to determine the criterion value for large subsets than for small subsets. Guyon and Elisseeff (2003) remarked this search may outsmart forward selection by eliminating at the first step the variable that by itself provides the best separation to retain the two features that together perform best. Schulerud & Albregtsen (2004) added this search also suffer from the nesting problem. However, Kuhn and Johnson (2013) suggested improvement for this search using non-inferential criteria such as Akaike Information Criterion (AIC) statistic, to add or remove features from the model.

2.2.1.3 Stepwise Selection

Stepwise selection is combination of the forward and backward approaches (Rencher, 2002), and thus overcomes the nesting problem (Schulerud & Albregtsen, 2004). At each step of the process, features that are not included in the present subset are examined in turn to determine which feature is the best for inclusion. Then, all the features that are retained in the present subset are examined in turn to determine if any of the features can be excluded without loss. Feedbacks with regards to the application of this search are varied. The use of stepwise procedures should be confined to problems with large number of features that the computation time required to examine all cases is prohibitive (McCabe, 1975). Pfeiffer (1985) preferred this search because it is an important tool to reduce the number of features without substantial loss of the discriminatory power. Krzanowski (2000) claimed that stepwise procedures remain the universally most popular tools for feature selection with at least a satisfactory subset even though the globally optimum one is obtained. However, Fraiman et al. (2008) highlighted that this search is not feasible when many features are considered. Even though stepwise search procedure makes the

search less greedy, it increases the repetition of hypothesis testing (Kuhn & Johnson, 2013; Marra & Wood, 2011).

2.2.1.4 Other Feature Search

Rough set theory has shown to be successful for selecting relevant and non-redundant features from a given data set, but this approach is slow in terms of computational speed (Maji & Garai, 2013), and the joint dependency of the features for high dimensional real life data sets cannot be estimated correctly (Peng et al., 2005). In this approach, the real valued features are divided into several discrete partitions and the dependency of approximation of a feature is calculated (Maji & Garai, 2013). The major concern in the computation of the dependency of real valued features is the inherent error exists in the discretization process. Liu et al. (2009) claimed this is a popular technique to obtain feature subset based on consistency.

Canonical variate analysis provides more possibilities for feature selection (Wood et al., 2005). This technique seeks for linear combination of the P features known as canonical variates within g groups which project the data onto an r-dimensional subspace $\left(\text{where } r \le \min\left(P, g-1\right)\right)$ in order to maximize the between groups to within groups variation, subject to the canonical variates being uncorrelated within and between groups.

The *plus-l-minus-r search* applies the forward selection to add l features and uses backward selection to remove r features from the resultant subset in each selection cycle (Schulerud & Albregtsen, 2004). The selection process is repeated until the required number of features p is obtained. This approach avoids the nesting problem.

Sequential Forward Floating Selection (SFFS) and *Sequential Backward Floating Selection* (SBFS) are an extension of forward and backward selection proposed by Pudil et al. (1994). These approaches efficiently remove the nesting problem by dynamically backtracking after each sequential step to locate a better subset (Nakariyakul & Casasent, 2009). SFFS uses the basis of forward selection to add one feature at a time to the selected feature subset, and every time a new feature is added to the present feature set, the algorithm backtracks using the backward selection to remove one feature at a time to find better subset. In reverse, SBFS method begins with all features *k*, removes one feature at a time and conditionally adds a feature to the resultant subset only if a better subset can be located. Reported studies in Nakariyakul and Casasent (2009) claimed that these floating search approaches performed better that the conventional forward and backward selections. Jain and Zongker (2002) stated the floating methods show results comparable to the optimal algorithm branch and bound, despite being for the most part, faster than the branch and bound algorithm.

Somol et al. (1999) proposed *adaptive versions* of the floating search methods. This method adds or removes more than one feature in each sequential step to obtain a better subset. Once the desired number of features *p* is almost attained, the adaptive methods increase the number of features to add or remove from the current subset in each step primarily for thorough searches. Findings from two data sets reported by Somol et al. (1999) showed that the *adaptive sequential forward floating selection* (ASFFS) algorithm presented slightly better results than SFFS, but with longer searching times. Ferri et al. (1994) in Jain and Zongker (2002) compared the performance of GA, SFFS, and forward selection methods on data set with maximum

360 dimensions and concluded that SFFS gives good performance on every high dimensional problem.

In common machine learning practices, the processes of feature selection and classification are set up in sequential procedures known as filters, wrappers, and embedded procedures (Blum & Langley, 1997). Filter procedure, ranks each feature based on some evaluation function of the individual predictive power and eventually chooses the best first *p* features. Primary assumption of this procedure is that the features are independent of each other and the final selection criterion can be obtained as a sum of or product of criteria evaluated for each feature independently (Cios, Swiniarski, Pedrycz & Kurgan, 2007). Vergara and Llobet (2011) explained the need to trim the number of features by assessing the features individually is purposely to understand their relative influence towards the classification performance. A common objective of ranking or filtering each feature is to select relevant (Ray & Turner, 1992; Vergara & Llobet, 2011) features or feature subset that may provide substantial predictive power (Cios et al., 2007). Wrappers treat classifiers as black boxes and aim at finding a feature subset that has minimum cross-validation error on the training data. Approaches like sequential forward selection, genetic algorithms and simulated annealing are examples of wrapper. Embedded methods such as decision tree intrinsically selects a subset of features in the training of classifiers, or optionally estimate the importance of features from the coefficients in the classifiers like support vector machine (Yan & Zhang, 2015).

2.2.2 Evaluation Function for Selecting Features

At this stage, the interest is to obtain an optimal features or subset of features using an evaluation function that can give the best classification performance. An evaluation function attempts to measure the discrimination ability of a feature or a subset of features to distinguish different group labels (Dash & Liu, 1997). Different evaluation criteria may lead to different optimal feature subsets (Pfeiffer, 1985; Dash & Liu, 1997). For instance, McLachlan (1992, pp. 91-93) stated the preference of selection criterion can be either *allocatory* or *separatory*, which depends on the objective of the discriminant analysis. Allocatory criterion measures effectiveness of feature vector in predicting group membership using the overall error rate of the optimal rule formed from the *p* features. Whereas, separatory criterion assesses the effectiveness of feature vector which tends to maximally distinguish or separate the population. Allocatory optimality is basically definable only when stringent assumptions are met while in vague situations a separatory function may sometimes usefully serve as an allocator (Geisser, 1976). Dash and Liu (1997) specifically divided evaluation for selecting feature into five different criteria which are distance, information, dependence, consistency, and classifier error rate.

The choice of criterion for selecting features has to be fitted with the goal of classification. Habbema and Hermans (1977) stressed that the criterion for the features selection has to be linked as closely as possible to the practical aim of discriminant analysis. They further described the two aims as (i) description of the differences between the groups on the basis of the sample data as descriptive discriminant analysis, and (ii) allocation of future elements whose origin is not known with certainty through creating allocation rule is called the predictive

discriminant analysis. Mahat (2006) further simplified the connection between the selection criterion and the aim of discriminant analysis properly. Ideally, allocation criterion i.e. classifier performance is appropriate if the objective is to identify features useful to form a classification rule with high accuracy for allocating future objects. Whilst, separation criterion is relevant if the objective is to identify features that are useful to describe separation between groups. Reviews upon the evaluation function for feature selection are based on these two criteria.

2.2.2.1 Allocation Criterion

Perhaps allocation criterion is the most widely applied feature selection criterion in classification problem. A large amount of literatures on this criterion are available from the earliest year of publication to the latest one. Despite the varieties of approaches, basically we may group the approaches according to the criteria suggested by Dash and Liu (1997), except the distance criterion. Among the earlier and simple feature selection was discussed by Weiner and Dunn (1966) where features were selected based on the largest Studentized (*t*-Statistics) differences between two sample means as defined in (2.1). Let $\bar{X}_1$ be the mean of a feature in Π_1 and $\bar{X}_2$ is the mean in Π_2, with N_1 and N_2 are the size of Π_1 and Π_2 respectively. Then, the *t*-statistics is

$$t = \frac{(\bar{X}_1 - \bar{X}_2)}{s\sqrt{\frac{1}{N_1} + \frac{1}{N_2}}} \tag{2.1}$$

McCabe (1975) proposed a test statistic U that measures the ratio of the estimated generalized variance within, to the estimated generalized variance total. Small values of U indicate good discrimination. Murray (1977) highlighted a method to eliminate features by discarding a pair of highly correlated features in a multiple comparison test. Young and Odell (1984) discussed two linear feature selection techniques; the principal component method, and the singular value decomposition method, which is based on M method of feature selection and is estimated using $\hat{M}$ defined by (2.2) where m denotes the distinct classes with known *a priori* probability. Unfortunately, this approach is inappropriate if the sample size from each group is greater or equal to the number of features.

$$\hat{M} = \left[\hat{\mu}_2 - \hat{\mu}_1 \middle| \ldots \middle| \hat{\mu}_m - \hat{\mu}_1 \middle| \hat{\Sigma}_2 - \hat{\Sigma}_1 \middle| \ldots \middle| \hat{\Sigma}_m - \hat{\Sigma}_1 \right] \tag{2.2}$$

Pfeiffer (1985) studied different criterion of Q_i which consider the ratio of probability density functions at point x_{ij} for a given group for nonparametric classification case. Hsu (1989) argued the application of partial F ratios, partial Wilks‟s Lambdas and p-values for evaluation of the unique contribution of a predictor to group discriminability. Three caveats concerning the these statistics; they may well lack of generalizability across different sets of predictors, elimination of two or more predictors because of low partial Fs or high p-values could result in a large decrease in the separability of groups, and finally despite their meaningful and interpretable in discriminant analysis, they are difficult to interpret in the stepwise approach.

Several evaluations of feature selection approaches that depend on data distributions include the Gini Index, information gain and information gain ratio. In this context, when a feature set is tested, the probability distribution of different classes is calculated and employed to measure the relevance with respect to the class concept (Lin, 2013). Thus, for mutli-group data, more diverse results are obtained than those in two-group problems, makes it hard to implement. Some investigations of such selection include Jin-Jie et al. (2008) who proposed information-theoretic criteria based on quadratic mutual information of feature selection and constructive criterion that overcome the earlier mutual information (common and Uniform distributions). Lin (2013) formulated the enhanced entropy denoted as *EH* as a new feature evaluation criteria defined by (2.3) with the success of approving a relevant feature relies on low entropy and low data variance. In this design, $EH(x,T)$ is compared with the initial status before splitting, and Aggregation Gain, $AG(x,T)$ is used to regard an improvement. In other word, the higher the $AG(x,T)$, the more relevant feature *x* is to the class.

$$EH(x,T)=\sum_{i-1}^{n}\frac{|T_i|}{|T|}\times H(x,T_i)\times\sigma^2(T_i) \tag{2.3}$$

Another criterion that usually applied in feature ranking mode (also known as filter method) includes the correlation criteria and mutual information which helps us understands the relevance of a feature (Chandrashekar & Sahin, 2014). The simplest correlation criteria is the Pearson correlation coefficient defined in (2.4) where x_i is the i_{th} feature, *Y* is the output (group), $cov(x_i,Y)$ is the covariance, $var(x_i)$ and $var(Y)$ are the variances of *x* and *Y*, respectively. However, this ranking can only detect

linear dependencies between feature and target. Yin et al. (2013) implied it can be overcome by using simple nonlinear preprocessing techniques on the feature before the correlation coefficients is computed.

$$R(i) = \frac{cov(x_i, Y)}{\sqrt{var(x_i) \cdot var(Y)}} \tag{2.4}$$

Mutual information or rather known as information theoretic ranking criteria uses the measure of dependency between two variables which begin with the Shannons definition for entropy. Suppose discrete feature X is observed, then the conditional entropy for continuous features is given by (2.5) which implies that by observing a feature X, the uncertainty in the output Y is reduced. The decrease in uncertainty is given by another function $I(Y,X) = H(Y) - H(Y|X)$. If X and Y are independent, then the mutual information is zero and greater than zero if they are dependent. For continuous features, we may need to find the probability densities of x_i and y as well as the joint density of $p(x_i, y)$. Obviously, for the continuous problem, the application of this criterion is burdensome since we often do not know the densities and are hard to estimate from data. Suggestions for further readings include Yin et al. (2013), Foithong et al. (2012) and Liu et al. (2009).

$$H(Y|X) = -\sum_{x_i} \sum_{y} p(x_i, y) \log(p(y|x)) \tag{2.5}$$

An extension for mutual information used as a filter in order to obtain maximum classification or prediction performance with a minimal subset of features by reducing the redundancies among the selected features to a minimum is the minimum

redundancy-maximum relevance criteria. The entropy-based on this criterion indicates that the higher it is for a feature, the more the feature is needed which can be obtained by subtracting the redundancy from relevance. Detail of such criterion is referred to Sakar et al., (2012) and Foithong et al., (2012). Similar to minimum redundancy-maximum relevance criteria, but with slightly different strategy is the maximum relevance and significance criterion (Maji & Paul, 2011), maximum weight and minimum redundancy (Wang et al., 2013).

Rayleigh coefficient can also be used as a criterion in reducing the number of selected features. Louw and Steel (2006) applied this criterion as they believe it optimize each step of feature identification for deletion. Considering α as *n*-vector of elements, while $\mathrm{M}^{(i)}$ and $\mathrm{N}^{(i)}$ as the Gram matrix, the criterion in (2.6) is calculated at each step to determine which variable *i* that should be eliminated.

$$R^{(i)}(\alpha) = \frac{\alpha' \mathrm{M}^{(i)} \alpha}{\alpha' \mathrm{N}^{(i)} \alpha} \tag{2.6}$$

Generally, allocation criterion is more concerned in finding the best subsets of features in the formation of good discriminant rule with good prediction power. This approach is actually somewhat similar with the regularization method proposed by Friedman (1989) in McLachlan (1992, pg. 152). He argued the influential subset of features has to be surprisingly small for the subset-selection techniques to be competitive with other regularization method or even without regularization at all. The feature subset selection within this perspective focuses on single feature relevance, relevance in the context of certain particular selection, or feature subset relevance towards predictive improvement or to gain in prediction accuracy. A

journal written by John et al. (1994) described definitions of relevance based on assumption that all features and the group are Boolean without noise. Chandrashekar and Sahin (2014) said the basic property of relevant features is those features which provide a measurement of feature‟s usefulness in discriminating the different classes. The later definition may also relate to the following criterion.

2.2.2.2 Separation Criterion

Another criterion that receives great interest is the separation criterion or sometimes known as the distance-based criterion. Han et al. (2013) referred this criterion as a measure that characterizes the quality of a feature based on its ability to discriminate instances of a group with instances from other groups. A large value of distance indicates that it is easier to discriminate between the groups (Afifi et al., 2004, p. 259). Another description by Maji and Garai (2013) that reflects the separation criterion is the search of best possible set of features that have optimal saliencies for which the inter-group (respectively, intra-group) distances are maximized (respectively, minimize). In other word, instances from different groups (i.e. between-group) should have feature values that are more distinctive than values from the same group (i.e. within-group). McLachlan (1992) discussed this issue in detail in his book and regards this feature selection as most useful in describing differences among possible groups. Mahat (2006) also mentioned that this criterion is sometimes preferred as an alternative to classifier performance because it avoids heavy computation.

Feature selection of this type is mostly performed in probabilistic approaches. Normally, probabilistic feature selection involves the use of certain bounds on the

probability of error which reflect the intuitive justification for the measure of separability between the groups. In particular, Kittler (1975) believed that the greater the distance, the smaller the probability of error. Kittler (1975) and Chittineni (1980) elucidated several types of a general two-group probabilistic distance measures that follow the Gaussian density distribution functions which include divergence (2.7), Bhattacharyya (2.8), Jeffreys-Matusita (2.9), and Kullback-Leibler (2.10), where $p\left(X\middle|\Pi_i\right)$ is the probability density functions of the patterns in groups Π_i, $\int(...)dX$ is the multivariate integral, and these are defined as follows.

$$J=\int\left[p\left(X\middle|\Pi_1\right)-p\left(X\middle|\Pi_2\right)\right]\log\left[\frac{p\left(X\middle|\Pi_1\right)}{p\left(X\middle|\Pi_2\right)}\right]dX \tag{2.7}$$

$$B=-\ln\int\left[p\left(X\middle|\Pi_1\right)p\left(X\middle|\Pi_2\right)\right]^{1/2}dX \tag{2.8}$$

$$JM=\left[\int\left(p\left(X\middle|\Pi_1\right)-p\left(X\middle|\Pi_2\right)\right)^2dX\right]^{1/2} \tag{2.9}$$

$$KL=\int\ln\left[\frac{p\left(X\middle|\Pi_1\right)}{p\left(X\middle|\Pi_2\right)}\right]p\left(X\middle|\Pi_1\right)dX \tag{2.10}$$

Han et al. (2012) proposed a new criterion function known as separability index matrix that apply functions (2.8) and (2.11), which they claimed is capable of providing discriminant features for classification with a very low computational cost. Unfortunately, the criterion is dependent with the classifiability of a feature that works best for two-group but not multi-group problems. Another multi-group studies

that applied criteria (2.7), (2.8), and (2.9) were performed by Gunal and Adizkan (2008). Surprisingly, their results for transformed (2.7) i.e. taking the standardization form of the divergence within a certain range of values and (2.9) gave similar outputs, and the same results recorded from criterion (2.7) and (2.8). In a nut shell, Gunal and Adizkan (2008) generalized that criteria (2.7), (2.8) and (2.9) were unable to beat their proposed novel separability measure found on subspace analysis, the common subspace and Fisher subspace.

Probabilistic approach of feature selection for multivariate data with multiple groups is complicated. Kittler (1975), Chittineni (1980), Davijver and Kittler (1982), and Ray and Turner (1992) whom remarked the most challenging part to deal with this criterion is to have the knowledge of the conditional probability densities of data that belongs to a particular family of probability density functions. Furthermore, in order to obtain effective probabilistic measure, it is necessary to have information about the probabilistic structure and the computation of the measures which usually involves the integration of multivariate density functions which have to be estimated from the training or sample data (Jain & Waller, 1978; Ray & Turner, 1992). Furthermore, for multi group problem, it seems hard to be implemented.

Mahalanobis distance Δ^2 (Mahalanobis, 1936) is known as an important feature selection criterion based on distance. Web (2002, pp. 426-427), Chittineni (1980) and Kittler (1975) claimed this distance function as a probabilistic distance measure when Bhattacharyya function in (2.8) has equal covariance matrices $\Sigma_1 = \Sigma_2 = \Sigma$.

However, Ray and Turner (1992) preferred to specify Mahalanobis distance Δ^2 as simple non-probabilistic measure which only takes into account the effect of

correlations between features, and makes it a potential powerful feature evaluation criterion. Given two Gaussian distributions with means $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ between two groups Π_1 and Π_2, and a common dispersion matrix $\boldsymbol{\Sigma}$, Mahalanobis distance Δ^2 is another useful feature selection. It is defined by equation (2.11).

$$\Delta^2 = \left(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\right)^T \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\right) \tag{2.11}$$

Afifi et al. (2004) strictly discussed the theoretical background of Δ^2 applied for features that follow multivariate normal distribution. The parameters $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$, and $\boldsymbol{\Sigma}$ are usually unknown and often, the training set of n_i instances from each group are used to estimate these parameters. Let $\bar{\mathbf{X}}_1$ and $\bar{\mathbf{X}}_2$ be the corresponding sample means vectors for two groups Π_1 and Π_2, then the estimated Mahalanobis distance D^2 between the two groups is defined by the equation (2.12), where **S** is the pooled variance-covariance matrix, and, n_1 and n_2 (such that $n_1 + n_2 = n$) are the sample size of data from groups Π_1 and Π_2, respectively.

$$D^2 = \left(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2\right)^T \mathbf{S}^{-1} \left(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2\right) \tag{2.12}$$

$$\mathbf{S} = \frac{\sum_{i=1}^{n_1} \left(X_{1i} - \bar{X}_1\right)\left(X_{1i} - \bar{X}_1\right)^T + \sum_{i=1}^{n_2} \left(X_{2i} - \bar{X}_2\right)\left(X_{2i} - \bar{X}_2\right)^T}{n_1 + n_2 - 2} \tag{2.13}$$

Jain and Waller (1978) have applied this concept in their investigation of peaking phenomenon based on two ranking types *i.e.* worst to best ranking and best to worst ranking. Achariyapaopan and Childers (1985) have also examined recursive suboptimal feature selection using Mahalanobis distance based on filter approach

where for a set of one, any feature regardless of its power can be selected, but for best two features is the combination of the first and the last ranked features, while for the best of three features is the combination of the first, middle and the last ranked features. Studies by Jain and Waller (1978) and Achariyapaopan and Childers (1985) where performed in the context of two equiprobable multivariate Gaussian densities with a common covariance matrix.

However, difficulties occurred in determining the probability of error since the conditional distribution of the chosen statistics for the decision rule should initially be determined. This technical flaw has long been tackled by previous researchers including Anderson (1951), John (1960), John (1961), Sitgreaves (1961) in Jain and Waller (1978), and Ray and Turner (1992). Besides, Roberts and Hanka (1982) have also explored the effect of correlation between features on the discriminatory potential of a feature subset. If two features are highly correlated positively, then the pairwise Mahalanobis distance depends critically on the individual distances, whereas if they are highly correlated negatively the pairwise performance will be a substantial improvement on the single feature.

Ray and Turner (1992) have introduced two new [0, 1] bounded Mahalanobis distance-based evaluation criteria known as Δ_A^2 and Δ_B^2 for multi-group classification problem both in the distribution free and Gaussian distribution cases. Both criteria are illustrated in equations (2.14) and (2.15).

$$\Delta_A^2 = \frac{\pi_1 \pi_2 \Delta^2}{1 + \pi_1 \pi_2 \Delta^2} \tag{2.14}$$

$$\Delta_B^2 = 1 - \exp\left(-\frac{\Delta^2}{8}\right) \tag{2.15}$$

The proposed criteria were tested for the recognition problem of hand-printed numeric characters, and proven to be more powerful over the direct use of the unbounded Mahalanobis distance Δ^2 criterion (Ray & Turner, 1992). Further details of the Mahalanobis distance-based criteria Δ_A^2 and Δ_B^2 according to the distribution free and Gaussian distribution is available in Ray and Turner (1992). Few studies for the application of [0, 1] bounded Mahalanobis distance-based criteria of Δ_A^2 are disclosed. Masnan et al. (2015) implemented [0, 1] bounded Mahalanobis distance-based criteria Δ_A^2 to filter the best subset of discriminant features for the classification of multi-group of honey.

The criterion has been discovered sensible to perform sensor closeness test and significantly improve the evaluation based on [0, 1] distance value. In the case of sensor closeness test, the [0, 1] bounded Mahalanobis distance-based criteria Δ_A^2 has overcome the exceptionally high value of average distance which lead to the failure to represent the average separability of groups. However, in evaluation purposes, difficulties in making decision based on unbounded magnitude of Δ^2 can be resolved using [0, 1] bounded Mahalanobis distance Δ_A^2.

An optimum measure of the effectiveness of a set of features is usually reflected by the Bayesian error probability (P_e). Devijver and Kittler (1982) have proven the relationships between Δ^2 and P_e in the distribution free case where there cannot be

any exact relationship as in equation (2.16) where π_1 and π_2 are the *a priori* probabilities of two groups.

$$P_e \leq \frac{2\pi_1\pi_2}{1+\pi_1\pi_2\Delta^2} \tag{2.16}$$

Ray and Turner (1992) further shown the derivation of Δ_A^2 based on equation (2.16) where they believed it was the only distribution free upper bound of P_e available in terms of Δ^2, denoted by equation (2.17).

$$P_e^U(A) \leq \frac{2\pi_1\pi_2}{1+\pi_1\pi_2\Delta^2} \tag{2.17}$$

After some imposed operations to equation (2.17), the criterion Δ_A^2 in equation (2.14) is obtained where $\Delta_A^2 \in [0,1]$ which corresponds to a normalizing transformation on $\Delta^2 \in [0,\infty)$. Among other properties of equation (2.14) are (i) Δ_A^2 is monotonically an increasing function of Δ^2, (ii) Δ_A^2 is symmetric with respect to π_1 with the property equation $\Delta_A^2(\pi_1 = r) = \Delta_A^2(\pi_1 = 1-r)$ since ($\pi_2 = 1-\pi_1$), (iii) $\Delta_A^2(\pi_1 = r) < \Delta_A^2(\pi_1 = s)$ iff min(r, 1–r) < min(s, 1–s). Details of each proof can be referred to Ray and Turner (1992).

In summary, since Δ_A^2 is a function of the Mahalanobis distance Δ^2, and that Δ^2 can be solved in terms of Δ_A^2, the distribution free upper bound (2.16) which is

expressed in terms of Δ^2 can be expressed by Δ_A^2. This implies that the error bound associated with Δ_A^2 is the same as that associated with Δ^2. Δ^2 increases unboundedly, whereas Δ_A^2 is upper bounded by 1. Because of the boundedness, Δ_A^2 is expected to perform better than Δ^2 in a multi group classification problem, as proved by Ray and Turner (1992). Another simple non probabilistic separation criterion is the Fisher ratio that does not assume the populations are normally distributed (Rencher, 2002, p. 300, Afifi et al., 2004, p. 259, and Johnson & Wichern, 2007, p. 590). The Fisher ratio provides a good measure of group separability because the distance increases as the between group difference increases, and the within group spread decreases (Han et al., 2013). The Fisher ratio is defined by equation (2.18) as the ratio of the between group difference to the within group spread as follows:

$$\lambda_{ijl} = \frac{\left(\bar{x}_{il} - \bar{x}_{jl}\right)^2}{\left(\sigma_{il}^2 + \sigma_{jl}^2\right)} \tag{2.18}$$

where $\bar{x}_{il}$, $\bar{x}_{jl}$, $\sigma^2{}_{il}$, and $\sigma^2{}_{jl}$ are the means and the variances of instances of the *i*th group and *j*th group in the direction of the *l*th feature, while λ_{ijl} indicates the group separation between the *i*th group and *j*th group in the direction of the *l*th feature.

Thus, for multi-group problem, Han et al. (2013) defined the generalized Fisher ratio by equation (2.19) where λ_l is the average group separability measures in the direction of the *l*th feature and *g* is the total number of groups, while *i* and *j* indicate the group indices such that $l \le i,$ and $j \le g$, also ω_i and ω_j are the mixing weights

for the *i*th and *j*th group, respectively. Yin et al. (2013) commented that a negative aspect of using Fisher criterion to obtain feature ranking is that it does not reveal the mutual information among features.

$$\lambda_l = \frac{1}{g(g-1)} \frac{\sum_{i=1}^{g}\sum_{j=1}^{g} \omega_i \omega_j \lambda_{ijl}}{\sum_{i=1}^{g}\sum_{j=1}^{g} \omega_i \omega_j}, \; i \neq j \tag{2.19}$$

Interclass distance is another simple group separability which can be used to assess discriminatory potential of pattern representations in a given space (Devijver & Kittler, 1982). This criterion is not defined explicitly via class conditional probability functions. Thus, its estimate can be computed based on the training set without prior determination of the probabilistic structure of the groups. Good separability among groups is displayed by the greater average pairwise distance between objects of different groups. Despite its simplicity, the only disadvantage is that this concept cannot serve as true indicators of mutual group overlap (Devijver & Kittler, 1982).

2.2.3 Stopping Criterion

A fundamental interest associated with the designing of pattern classifiers is to determine the optimal number of selected features for a given sample size. This can be set by deciding when the contribution of an additional feature is meaningful or meaningless. A stopping criterion determines when the feature selection process should stop. Several types of stopping criteria are: (a) the search completes; (b) some given bound is reached, where a bound can be a specified number (minimum number of features or maximum number of iterations); (c) subsequent addition (or deletion)

of any feature does not produce a better subset; and (d) a sufficiently good subset is selected (e.g., a subset may be sufficiently good if its classification error rate is less than the allowable error rate for a given task).

Jain and Waller (1978) have studied two optimal number approach based on assumptions that each feature is equally good, and the features do not have equal discrimination power for classification. An approximate criterion used to determine whether or not a subsequent additional feature would be helpful for a classifier is given by equation (2.20) where $\delta\Delta_P^2$ denotes the increase in the Mahalanobis distance of each additional feature so that the average of error rate remains constant.

$$\Delta_{P+1}^2 - \Delta_P^2 = \delta\Delta_P^2 = \frac{\Delta^2{}_p}{2N-3-p} \tag{2.20}$$

Applying equation (2.20) as a threshold for finding optimum number of ranked features according to equation (2.11) from best to worst, feature that does not increase $\delta\Delta_P^2$ will be excluded. One can see as the sample size *N* increases, the value of $\delta\Delta_P^2$ decreases. This indicates that if the sample size is approaching infinity, the selected *p* features contribute nothing to the separation.

On the other hand, Achariyapaopan and Childers (1985) proposed a recursive algorithm for selecting near optimal features set by introducing a weighting factor $W = \sum^{-1}\left(\mu_1 - \mu_2\right)$. Their argument was significant feature x_i should be weighted more than insignificant feature. Selection begins by calculating the weight of each feature. The best feature based on the largest *W* is then selected and the Γ_1 as in

equation (2.21) is then calculated. Next, the best two feature combinations are determined by calculating the Γ_2. If $\Gamma_2 > \Gamma_1$, the second feature is included in the optimal features. Otherwise, the feature is excluded. The calculation of Γ in equation (2.21) was based on the approximation of the probability of error derived by Lachenbruch (1968).

$$\Gamma = \left(\frac{(N-2-f)(N-5-f)}{(N-3)(N-3-f)} \right)^{1/2} \times \frac{\Delta^2}{\left(\Delta^2 + (4f/N)\right)^{1/2}} \tag{2.21}$$

Yongli et al. (2013) have applied a criterion function $\min\left(f(c,m)\right)$ defined by equation (2.22), where c is the current classification rate, and m is the current misclassification rate. The evaluation criterion $f(c,m)$ which is a quadratic function with the minimum value would give the optimal feature subset needed. The criterion is used to obtain optimal feature subsets for their improved feature selection algorithm based on Mahalanobis distance.

$$f(c,m) = \sqrt{\left((1-c)100.m.100\right)} \tag{2.22}$$

Appropriate statistical test can be applied for the stopping criterion. Let $\tilde{d}$ be a sample based distance between group, Mahat (2006, p. 115) illustrates simple explanation of different stopping rule for the selection strategies. In forward selection strategy, feature that produces the highest value of $\tilde{d}$ after being chosen in the set of discriminators will be selected in the strategy; thus, stop the selection when including a new discriminator does not significantly increase the value of $\tilde{d}$. While in

backward elimination strategy, feature that produces the highest value of $\tilde{d}$ after being eliminated in the set of discriminators will be deleted; thus, stop the selection when deleting further discriminators significantly decreases the value of $\tilde{d}$.

Based on the above mentioned setting, whenever a set of discriminators have been identified, the following stopping rules may applied (Mahat, 2006):

i. Testing for no additional information (forward selection).

A new discriminator is chosen when its contribution together with the previous selected discriminators gives adequate information for separating the two groups in the forward selection. The null hypothesis ($H_o : \Delta_j = \Delta_p$) is performed for each $j = 1, 2, \ldots, p$. This test checks at every step j the adequacy of the information in the j discriminators as compared to the information in all P discriminators. The following statistic is used to test the whole set of selected discriminators when each new potential discriminator is chosen

$$F_{calc} = \frac{N - p + 1}{p - j} \times \frac{n_1 n_2 \left(D_p^2 - D_j^2 \right)}{n_1 n_2 D_j^2 + N\left(n_1 + n_2 \right)} \tag{2.23}$$

Where $N = n_1 + n_2 - 2$, D_p^2 is the sample-based Mahalanobis distance using all the p continuous features and D_j^2 is the sample based Mahalanobis distance using the current j selected discriminators at step j. For a fixed value

of type I error, α, these steps will be continued until the first j discriminators give $F_{calc} \le F_{(p-j,N-p+1,1-\alpha)}$.

ii. F to enter (forward selection).

The null hypothesis $H_o : \Delta_j^2 = \Delta_{j-1}^2$ where $j = 2, 3, \ldots, p,$ is to test for adequacy of information given by a new potential discriminator, selected at step j of the forward selection, compared to the previous selected discriminators. The test statistics is

$$F_F = (N - j + 1) \times \frac{n_1 n_2 \left(D_j^2 - D_{j-1}^2\right)}{n_1 n_2 D_{j-1}^2 + N(n_1 + n_2)} \tag{2.24}$$

The selection of potential discriminator is continuously performed until the first j discriminators gives $F_F \le F_{(1,N-j+1,1-\alpha)}$.

iii. Testing for no additional information (backward elimination).

In backward elimination, the associated tests in (i) and (ii) are reversed. The null hypothesis for no additional information gain is $H_o : \Delta_p = \Delta_{(p-j)}$, for $j = 1, 2, \ldots, p$. And the test statistics is

$$F_R^* = \frac{N - p + 1}{j} \times \frac{n_1 n_2 \left(D_p^2 - D_{p-j}^2\right)}{n_1 n_2 D_{p-j}^2 + N(n_1 + n_2)} \tag{2.25}$$

And the deletion continues until $F_R^* \le F_{(j,N-p+1,1-\alpha)}$.

iv. F to delete (backward elimination).

The null hypothesis is $H_o : \Delta_j = \Delta_{j-1}$ where $j = 2, 3, \ldots, p$, and the associated test statistic is

$$F_F^* = (N - j + 1) \times \frac{n_1 n_2 \left(D_{j-1}^2 - D_j^2\right)}{n_1 n_2 D_{j-1}^2 + N(n_1 + n_2)} \tag{2.26}$$

The deletion continues until $F_F^* \le F_{(1,N-j+1,1-\alpha)}$.

2.3 Classification Rules

Schaller et al. (1998) illustrated that supervised or trained learning methods classify an odor by developing a mathematical model based on training data, i.e. samples with known properties to a set of given descriptor. Test samples are then evaluated against a knowledge base and predicted class membership is deduced. These methods enable the system to reduce parameters such as volatile, temperature and humidity, and train a system to concentrate at particular combinations of sensors to measure a given odor. Supervised methods include artificial neural network (ANN) such as multi-layer perceptron (MLP), probabilistic neural network (PNN), and radial basis function network (RBF) (García-González et al., 2002; Zhang, Balaban & Principe, 2003; Dutta, Das, Stocks & Morgan, 2006); regression analysis that include partial least square (PLS), multiple linear regression (MLR), principal

component regression (PCR) and ridge regression (RR) (García-González et al., 2002); discriminant function analysis (DFA) which include linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA) (Shaffer et al., 1999; Rodriguez, et al., 2010); learning vector quantization (LVQ) (Shaffer et al., 1999); and support vector machine (SVM) (Di Natale, Davide, and Di Amico, 1995).

The abovementioned methods can also be categorized as parametric approach or nonparametric approach. *Parametric* approaches assume that data can be described by a probability density function such as multi-normal distribution (Gardner et al., 1999), while *nonparametric* approaches do not make any underlying assumptions about the probability distribution of data (Hines et al., 1999).

2.3.1 Parametric versus Nonparametric Classification Approaches

Fisher"s Linear Discriminant Analysis (FLDA) was derived by Fisher (1936) that originally deals for two-group problem. Rao (1948) generalized Fisher"s Linear Discriminant Analysis (FLDA) from two-group problem to multiple-group problem which he called it as linear discriminant analysis (LDA) or rather known as multiple linear discriminant analysis. To understand the multiple-group problem, let us look at the two-group problem first. Let $\mathbf{x}$ in groups Π_1 and Π_2 with means $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ and covariances $\sum_1$ and $\sum_2$ the linear combination of $\mathbf{x}$ be written as $Y = a_1x_1 + a_2x_2 + \ldots + a_nx_n = a^Tx$ where $\underline{a}$ is a set of parameters of the model. Then $\boldsymbol{Y}$ will have means $\boldsymbol{a}^{\boldsymbol{T}}\boldsymbol{\mu}_i$ and variances $\boldsymbol{a}^{\boldsymbol{T}}\sum_i \boldsymbol{a}$ for $i = 1, 2$.

Since the parameters $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ and $\sum$ are usually unknown, estimation for each parameter is often based on the training sets of $n(\Pi_j)$. Assuming both groups have equal covariance $\sum$ and the variance of $\boldsymbol{Y}$ is $\boldsymbol{a}^{\boldsymbol{T}} \sum \boldsymbol{a}$ in both groups, substitute μ_i by $\bar{x}_i$ (the estimated mean vector in Π_i), and $\sum$ with S (the pooled estimate of variance), thus Fisher"s approach tries to maximize the ratio of the between groups variance to within groups variance. The optimal criterion of FLDA is the linear combination $y = \hat{a}^T x = (\bar{x}_1 - \bar{x}_2)^T S^{-1} x$ that maximize the ratio:

$$r(\hat{a}) = \frac{\left(\hat{a}^T \bar{x}_1 - \hat{a}^T \bar{x}_2\right)^2}{\hat{a}^T S \hat{a}} \tag{2.27}$$

Therefore, for $r(\hat{a})$ to be maximize is to find the best value of $\hat{a}$ given by $S^{-1}(\bar{x}_1 - \bar{x}_2)$ which gives the maximum ratio as $(\bar{x}_1 - \bar{x}_2)^T S^{-1}(\bar{x}_1 - \bar{x}_2)$. Thus, allocation rule based on Fisher linear discriminant function (FLDF) is given by:

Allocate x_0 to Π_1 if $y_0 - \hat{m} \geq 0$, or allocate x_0 to Π_2 if $y_0 - \hat{m} < 0$, where

$$y_0 = (\bar{x}_1 - \bar{x}_2)^T S^{-1} x_0 \geq \hat{m} = \frac{1}{2}\left(\bar{x}_1 - \bar{x}_2\right)^T S^{-1}(\bar{x}_1 + \bar{x}_2) \tag{2.28}$$

For $g > 2$ populations problem, the motivation behind FLDA is to find a reasonable representation of the populations that involves only a few linear combinations of observations, such as $Y = a_1^T x, a_2^T x, \ldots a_n^T x$. The main purpose of FLDA is to separate population, and can also be used to classify. The g populations are not necessarily assumed to be multivariate normal, but the population covariance

matrices are assumed equal, $\sum_1 = \sum_2 = \ldots = \sum_g = \sum$. Based on the estimation of each parameter, the ratio for multiple group problems is defined by:

$$\frac{\hat{a}^T B\hat{a}}{\hat{a}^T W\hat{a}} = \frac{\hat{a}^T \left(\sum_{j=1}^{g} (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})^T \right) \hat{a}}{\hat{a}^T \left[\sum_{j=1}^{g} \sum_{i=1}^{n(\Pi_j)} (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})^T \right] \hat{a}} \tag{2.29}$$

where B is the sample between groups and W is equal to S, the estimate of $\sum$ based on the sample within groups matrix, and the optimization of $\hat{a}$ is represented in the form of eigenvectors $\hat{e}$ such that $SB\hat{e} = \hat{\lambda}\left(n(\Pi_1) + n(\Pi_2) + \ldots + n(\Pi_j) - g\right)\hat{e}$. Let $\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_s > 0$ denote the $s \leq \min(g - 1, p)$ nonzero eigenvalues of $W^{-1}B\hat{e} = \hat{\lambda}\hat{e}$, and $\hat{e}_1, \hat{e}_2, \ldots, \hat{e}_s$ be the corresponding eigenvectors, then the vector coefficient $\hat{a}$ that maximize the ratio (2.29) is given by $\hat{a}_1 = \hat{e}_1$. The linear combination $\hat{a}_1^T x$ is called the sample first discriminant, when $\hat{a}_2 = \hat{e}_2$ then $\hat{a}_2^T x$ is called the sample second discriminant and continue until $\hat{a}_k = \hat{e}_k$ the sample jth discriminant, $j \leq s$. Thus, allocation rule based on Fisher linear disriminant function (FLDF) is given by:

Allocate x to Π_j if

$$\sum_{j=1}^{r} \left(y_j - \bar{y}_{kj} \right)^2 = \sum_{j=1}^{r} \hat{a}_j^T \left[\left(x - \bar{x}_k \right) \right]^2 \leq \sum_{j=1}^{r} \left[\hat{a}_j^T \left(x - \bar{x}_i \right) \right]^2 \quad \text{for all } i \neq k \tag{2.30}$$

where $\hat{a}_j$ is defined in (2.30) , $\bar{y}_{kj} = \hat{a}_j^T \bar{x}_k$ and $r \leq s$. Further details of these concepts can be found in (Hastie, Tibshirani, and Friedman, 2009, pp. 106-111), (Rencher, 2002, pp. 304-305), and (Dillon, and Goldstein, 1984, pp. 400-403).

Although FLDA approach does not strictly assume the groups are normally distributed, it does implicitly assume the covariance matrices of the groups are equal since a pooled covariance matrix is used. When the groups follow normal distribution and have equal covariance matrix $\left(\sum_1 = \sum_2 = \cdots = \sum_g\right)$, Fisher"s discriminant rule is equivalent to LDA with minimum equal costs of misclassification (ECM) rule with equal prior probability and equal cost of misclassification. Further details are referred to allocation rules of ECM in Johnson et al., (2007, p. 608). If $\sum_1 = \sum_2 = \ldots = \sum_{\boldsymbol{g}}$ does not hold, the classification rules can easily be altered to preserve optimality of classification rates. By substituting S_i, the sample covariance matrix for the *i*th group in the distance function of

$$D_i^2(y) = \left(y - \bar{y}_i\right)^T S_i^{-1} \left(y - \bar{y}_i\right), \qquad \text{for } i = 1, 2, ..., g \tag{2.31}$$

then (2.31) cannot be reduced to a linear function of *y*, but remains as a quadratic function (Rencher, 2002, p. 306). Thus, discrimination rules based on S_i are called quadratic classification rule or known as QDA which was proposed by Smith (1947).

However, Fatti, Hawkins and Raath (in Hawkins, 1982) have noted that if the distribution of the populations is not normal, the optimality of the linear discriminant function can still be improved. Throughout their simulation studies of applying the LDF and QDF to continuous data, they made several generalizations. First, for

distribution of data with lighter tail than the normal, the LDF or QDF should perform very adequately; second, LDF and QDF will perform poorly if the distributions are heavy tailed and skewed; and finally, if the distribution are heavy tailed but symmetric, QDF may perform reasonably well in terms of overall error rate when the training samples are very large. However, if the training samples are small, then heavy tailed would introduce excessively large sampling errors into parameter estimates, which lead to an unreliable discriminant function.

2.3.2 Other Nonparametric Approaches

Methods we have discussed so far can be considered as parametric and nonparametric methods. Even though LDA could be either parametric or nonparametric (known as FLDA), depending on the assumption of the group conditional densities of the samples being studied, LDA and QDA mark as the most influential methods under the multivariate normality condition. Among other methods that grab the attention of researcher in the multi sensor data fusion problem is the nonparametric techniques. When these techniques applied, the group conditional densities are not known (McLachlan, 1992, p.283). Such techniques include *kernel*, *nearest neighbor*, classification and regression trees (*CART*) and artificial neural networks (*ANN*) and of course the FLDA. These techniques are better alternative to parametric procedures, but, they usually need large amount of training set.

Sewell (2009) described in detail the history and evolution of *kernel* method. Kernel discriminant analysis estimates the distribution of variables in each group using one of a variety of complex functions known as kernel density estimates. It is commonly

used in a nonlinear problem where one can map the problem from the input space to a new (higher dimension) space called the feature space by doing a nonlinear transformation using suitable chosen basis functions and then use a linear model in the feature space. The linear model in the feature space corresponds to a nonlinear model in the input space which makes it useful in the classification and regression problems. However, the choice of kernel function is crucial for the success of all kernel algorithms because the kernel constitute prior knowledge that is available about a task. Further readings of this technique can be found in a monograph on the theory of kernel written by Berg, Christensen and Ressel (1984).

The *nearest neighbor* (1-NN) rule classify based on the "nearness" and then attempt to find groups of subjects that are as near as possible to each other. This is the most fundamental and simplest supervised classification techniques. However it tends to be computationally intensive and very robust to noise. The nearest neighbor of (1-NN) and (*k*-NN) are among the preferred rule. Using the *k*-NN classifier, a test pattern is assigned to the class that is the most frequent among k-nearest neighbors in the training set. Usually *k* is selected to be odd in order to avoid ties. The *k*-NN rule becomes optimal when *k* tends to infinity. Such approach has been discussed by (Ciosek et al., 2006) in the application of e-nose.

CART was developed by Breiman, Friedman, and Olshen (1984), is a tree building method in which data is split repeatedly in groups according to different parent nodes, child nodes and terminal nodes. Exhaustive search of all possible splits is performed until the optimal univariate splits is found. However, this method is biased toward selecting predictor variables having more levels and it is very computer intensive (Johnson et al., 2007). As far as this study is concerned, this method is

rarely applied in the multi sensor data fusion domain. Sources of discussion are available in (Worth, and Cronin, 2003; Buratti et al., 2004; Gigli et al., 2007).

ANN are defined as structures comprised of densely interconnected adaptive simple processing elements called nodes that are able to perform massive parallel computations for data processing and knowledge representation. This technique is meant for data with nonlinear nature, and usually being applied in the fusion of e-nose and e-tongue (Gutierrez-Osuna, 2002). This technique is favored in fusion domain may be because it resemble the structure and workings of the human brain through mathematical models. However, the potential disadvantage of this approach is that no statistical information can be created to express the model and the number of training samples as well as the training time required to analyze are largely dependent on the number of adjustable parameter for NN (more adjustable parameter increase the training sample and training time) (Ghasemi-Varnamkhasti et al., 2010). Some extensions of this technique were proposed by: (Rong et al., 2000 called fuzzy NN) and (Boilot et al., 2003 known probabilistic NN). Further discussions from the multi sensor data fusion perspective are included in (Wide et a., 1998; Cimander et al., 2002; Xiabo et al., 2005; Ghasemi-Varnamkhasti et al., 2010).

Additional readings regarding other reported techniques applied in the multi sensor data fusion domain is accessible in: Bayesian classifier (Rousell et al., 2003), genetic algorithm (Rousell et al., 2003; Llobet, Brezmas, Vilanova and Correig, 2006; Gigli et al., 2007), support vector machine (SVM) classifier (Pardo et al., 2008; Dixon and Brereton, 2009; Wang et al., 2009; Tao et al., 2009), partial least square (Apetrei et al., 2010; Daeswijk et al., 2011), cluster analysis (Huang, Qiu and Guo, 2009), and fuzzy C-means (Sundic et al., 2000).

2.3.3 Evaluation of Constructed Classifier

Once we have obtained the classifier that can allocate observations to its specific group, it is important to evaluate the performance of the classifier. One way to judge the ability of classification procedures to predict group membership is to find the probability of misclassification known as *error rate*. For example, Johnson et al. (2007) suggested that Fisher‟s classifier can be optimal from a minimum expected cost of misclassification (ECM) or minimum total probability of misclassification (TPM) only when multivariate normality holds, such that $f(x_N|\pi_j)$ and $f(x_T|\pi_j)$ for $j = 1, 2, ..., g$, are multivariate normal distributions with means $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots, \boldsymbol{\mu}_g$ with covariance matrix $\sum_1, \sum_2, \ldots, \sum_g$.

However, in real case, this may not be the case. Rencher (2002, p. 314) described that classification rules developed using Fisher‟s concept can be considered as parametric or nonparametric for certain condition in the allocation rule. In the absence of multivariate normality, Fisher‟s classifier can be viewed as providing an approximation to the total sample information. Johnson et al. (2007) suggested the values of the first few discriminants themselves can be checked for normality and rule (2.28) employed. Since Fisher‟s linear discriminant method provides a discrimination rule minimizing the expected misclassification error (Hastie et al., 2009), this method is considered non-cost-sensitive. Thus, Johnson and Wichern (2007, p. 619) recommended a computation of estimated error rate for allocation rule (2.28) using Lachenbruch‟s estimate, i.e. the expected actual error rate (AER) given by

$$\hat{E}(AER) = \frac{\sum_{j=1}^{g} n_{j_M}^{(H)}}{\sum_{j=1}^{g} n_i} \tag{2.32}$$

where $n_{j_M}^{(H)}$ is the number of misclassified holdout observations in the *j*th group, $j = 1, 2, ..., g$. The holdout concept will be described further in the next section.

A much simpler and straight forward measure of classification performance is the apparent error rate (AER). This approach does not depend on the form of the parent populations and it can be applied for any classification procedure (Johnson and Wichern, 2007). It can be defined as the fraction of observations in the training sample that are misclassified by the sample classification function. AER can be easily calculated from the confusion matrix, which shows the actual versus predicted group membership. For n_1 objects from Π_1 and n_2 objects from Π_2, the confusion matrix is as follows

Table 2.7

Confusion Matrix Table for Two Groups (Π_1, Π_2)

		Predicted Membership		
		Π_1	Π_2	
Actual Membership	Π_1	n_{1C}	$n_{1W} = n_1 - n_{1C}$	n_1
	Π_2	$n_{2W} = n_2 - n_{2C}$	n_{2C}	n_2

Where

n_{1C} = number of objects from Π_1 correctly classified as Π_1

n_{1W} = number of objects in Π_1 misclassified as Π_2

n_{2C} = number of objects from Π_2 correctly classified as Π_2

n_{2W} = number of objects in Π_2 misclassified as Π_1

Therefore, the proportion of objects in the training set that is misclassified known as the apparent error rate is calculated as

$$AER = \left(\frac{n_{1W} + n_{2W}}{n_1 + n_2} \right) \times 100 \tag{2.33}$$

Unfortunately, Efron (1986) criticized this approach exhibits biases especially when the size of training set is small. This drawback can be further improved using a resampling method called leave-one-out.

Once we have developed the suitable classifier, the next issue need to be considered is the classifier performance. Our main interest is to design a discriminant rule that can classify future object well, which eventually gives low *misclassification rate* or the *error rate*. Generally, it is difficult to obtain an analytic expression for the error rate and therefore it must be estimated from the available data. Web (2002) expressed that the error rate measure suffers from the disadvantage that it is only a single measure of performance, treating all correct classifications equally and all misclassifications with equal weight. Among the common measure of error rate are apparent error rate, true error rate, expected error rate and Bayes error rate. Further readings on these measures are referred to Web (2002).

There are few techniques that can be applied to measure the error rate. Among them are the *re-substitution* method, *hold out* method, *U-method* or *cross validation* method, *Jackknife* method and *bootstrap* method. The re-substitution method (use the design set to estimate the error) yields what we call the apparent error rate. Error of this type is consistent, but can be severely optimistically bias (Dillon et al., 1984). The hold-out method splits the data into two mutually exclusive sets, the training and test set. The classifier is designed using training set and performance evaluated on the independent test set. Devijver and Kittler (1982) in Web (2002) criticized that this method makes inefficient use of the data and gives pessimistically biased error estimate. Dillon et al. (1984) added that this method requires large sample, and in their absence, either the function or its estimate of misclassification is likely to suffer, but the estimation of this method is consistent and unbiased.

One of the cross-validation branches is the leave-one-out method. The method is among the mostly applied method especially in the case of small sample size. It is an improved version of the sample-splitting or partitioning procedure (Rencher, 2002). The method makes use of all the available data without serious bias in the estimating error rate such that calculate the error by using *n-1* samples in the design set, and testing on the remaining sample, and this is repeated for all n subsets of size *n-1* (Web, 2002). The disadvantage of this method exist when the sample is large, result in computationally expensive to design *n* classifiers. This method yield almost unbiased estimates of the misclassification probabilities (Dillon et al., 1984). Despite the fact that this method is unbiased, several researchers criticized the method with regard to the large variance and mean square error.

The Jackknife method is a procedure for reducing the bias of the apparent error rate (Web, 2002). Dillon et al. (1984) reported that this method employ two stages: use the cross validation to obtain good estimates of classification error rate, and then use jackknife analysis to consider coefficient stability. These processes simultaneously require large number of successive runs and become increasingly time consuming and expensive.

Bootstrap refers to a class procedure that samples the observed distribution, with replacement, to generate sets of observations that may be used to correct for bias (Web, 2002). This method provides nonparametric estimates of the bias and variance of an estimator which proven to be superior to many other techniques. Even though it is computationally intensive, many researchers find it to be an attractive technique which has recorded many developments from Efron (1979) who introduced the technique.

CHAPTER THREE
RESEARCH METHODOLOGY

3.1 Introduction

In the previous chapter, we have highlighted in specific the traditional feature extraction method that was widely applied for the dimension reduction of the multi sensor data fusion model involving e-nose and e-tongue sensors. Some complications in applying feature extraction in the respective areas have also been discussed in detail. This study proposes alternative idea which implements feature selection in the classification process of sensors data as planned in Figure 3.1.

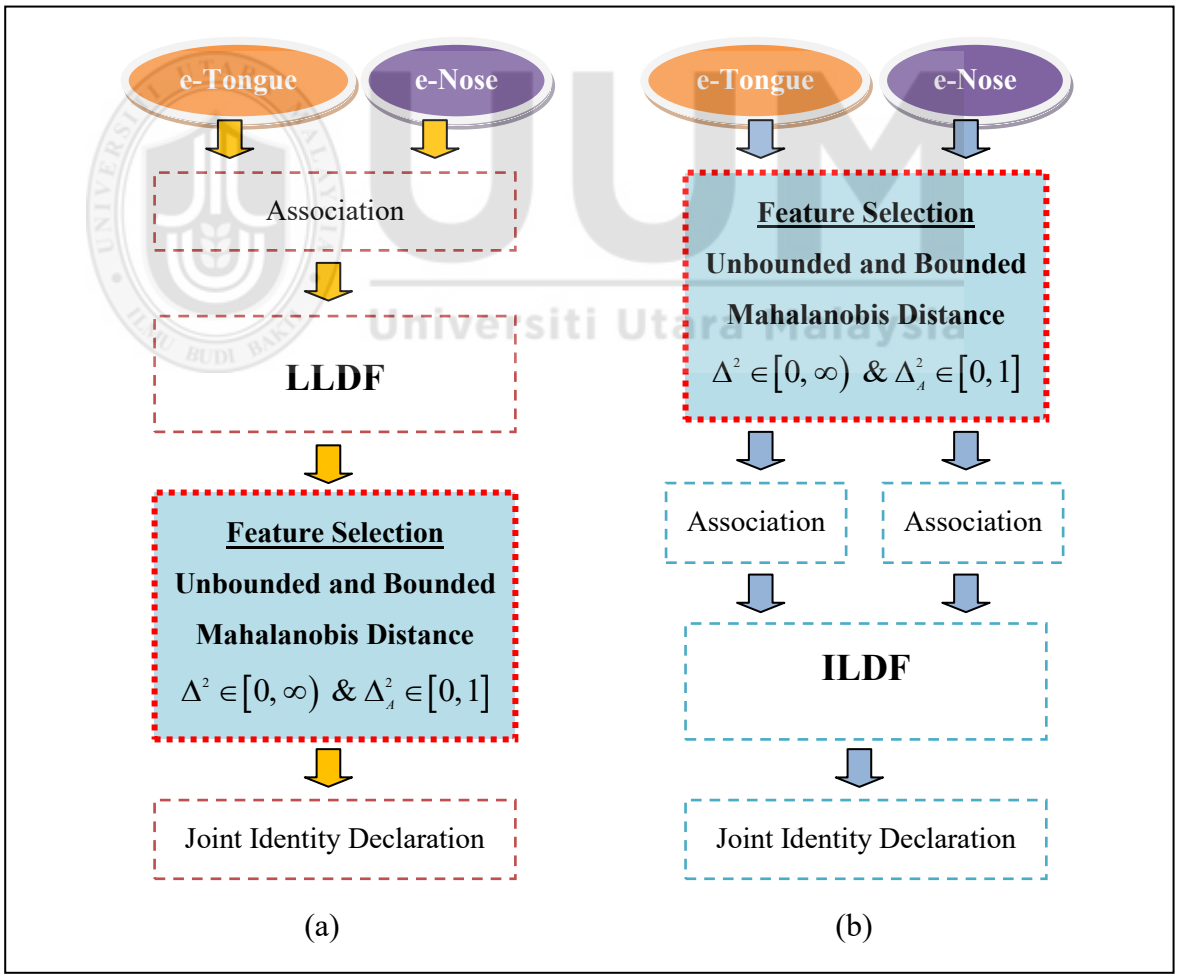

Figure 3.1. Proposed Methodological Changes for Multi Sensor Data Fusion (a) LLDF Model, and (b) ILDF Model Using Feature Selection of Unbounded and Bounded Mahalanobis Distances

In relation to the discussions of some drawbacks of PCA in section 2.1.3, Figure 3.2 illustrates the basis of principal components (PC) applied for extraction of features based on the highest variance counted in the first few selected PCs. As highlighted in the literature, this approach lacks of interpretability and unable to show obvious and well-defined feature"s contribution.

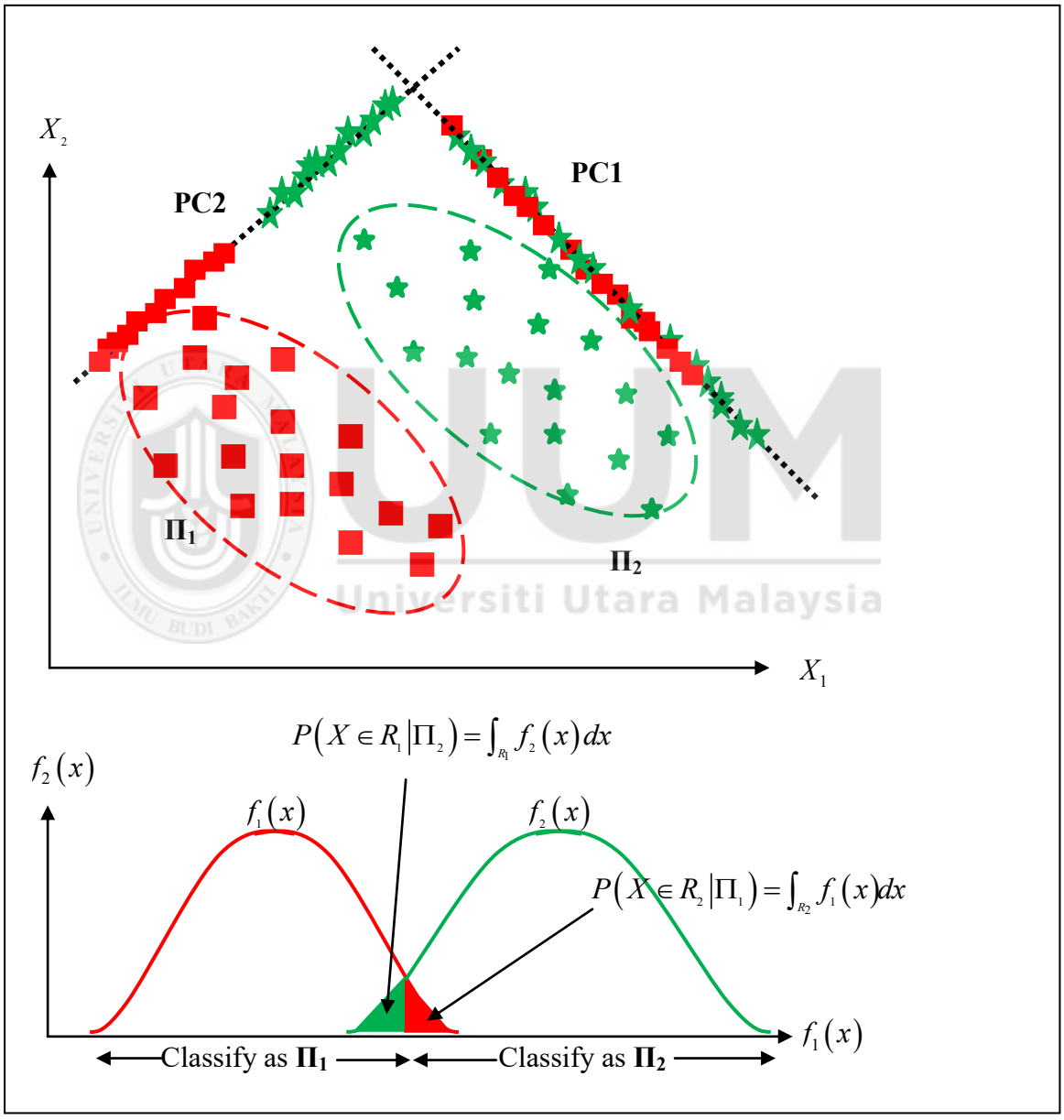

Figure 3.2. Illustration of the Application of PCA and Probability Distribution Function in Dimension Reduction and Classification

Another well-known approach to perform feature selection is based on the probability distribution of data. The most challenging part to perform this criterion is to have the knowledge of the conditional probability densities of data that belongs to a particular family of probability density functions. This can only be done by performing some integration procedures. If only two groups are involved, the process may not be burdensome, but for multi-group problem, integration of multivariate density functions which have to be estimated from sample data would be complicated and time consuming due to more features involved for calculation. Again, Figure 3.2 shows the application of probability distribution approach for classification of two group problem which involve integration of probability distribution functions of groups. Due to this difficulty, feature selection as well as classification using the knowledge of data distribution is avoided.

Therefore, for this study, emphasis was given to the feature selection issues based on distance approach which disregards the above hassles. Main focus was dedicated to the role of two Mahalanobis distance functions in selecting relevant features elaborated in section 2.2.2.2. The performance of the original Mahalanobis distance Δ^2 proposed by P. C. Mahalanobis (1936) and its extension Δ_A^2 suggested by Ray and Turner (1992) will be further explored, compared and evaluated. The two functions Δ^2 and Δ_A^2 both in equations (2.11) and (2.14) differ in the calculated distance value and are estimated using D^2 and D_A^2. The first function produces $[0, +\infty)$ unbounded distance, while the later gives $[0, 1]$ bounded distance.

The effect between the two distances (Δ^2 and Δ_A^2) is not noticeable unless a thorough assessment is performed. We believe each function leads to dissimilar

evaluation of significant features which eventually affects the classification performance. Figure 3.3 shows graphically how the pair-wise distances between multi-groups were calculated using criterion D^2 which later was used to calculate the average distance Δ_A^2. This study theoretically employs the Mahalanobis distance concept to find the maximum average distance values from both D^2 and D_A^2 as an indicator of good discriminant features.

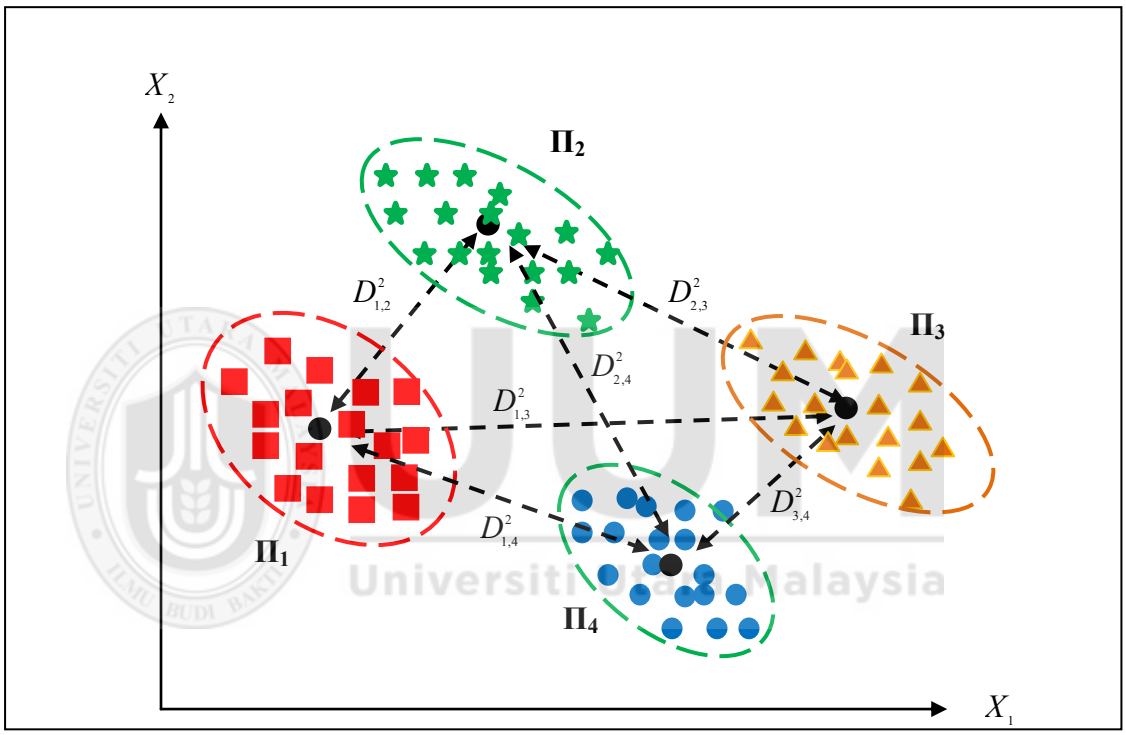

Figure 3.3. Graphical Representation of Pair-Wise Mahalanobis Distance Δ^2 / Δ_A^2 Between Multi-Group Means

The concepts of calculating pair-wise distances using Δ^2 and Δ_A^2 were found to be very useful and practical for this study compared to other distance measures. It provides a measure of similarity between multivariate populations and uses covariance information between features to weight the contributions of all features to the distance. Furthermore, the distance gives less weight to those features that have high variance and to those features that have high correlation. Euclidean distance was found to be inappropriate since this distance gives excess weight to features that were

highly correlated and gives additional weight to features that have high variance and similar information (Jalan, 2009). Taking the advantages of the distance provided by Δ^2 and Δ_A^2, the study adopted the distance to measure the features‟ separability.

In consequence, to accomplish the objectives of the study, the designed research methodology in this chapter has been technically evaluated to accommodate certain characteristics of single and fusion of e-nose and e-tongue data. Among the characteristics include:

i. multivariate data under study are all continuous in nature,

ii. data belong to multi-groups case,

iii. small number of groups‟ sample size where each $n(\Pi_j) < 30$ for $j = 1, 2, \ldots, g$,,

iv. sample size is smaller than the number of features (p and q) , $(n < p)$, $(n < q)$, $(n < p + q)$ for p is the feature dimension of e-tongue, q is the feature dimension of e-nose, and $p+q$ is the feature dimension for the fused features,

v. data belong to multi-groups that follow normal and non-normal distributions, commonly groups‟ distributions are nearly symmetric with lighter tail, few groups‟ distributions are skewed but overall, the range for all groups are very small, and

vi. some differences in the nature of fused data where e-nose features are generally highly correlated positively, while e-tongue features are moderately and highly correlated positively and negatively.

In order to generalize (within the scope of the research) the capability of the LLDF and ILDF functions, we proposed a research methodology that applies a new feature selection criteria and procedure with common classification rule, stopping criterion and performance evaluation. Since data were mix in distribution, the parametric classification rule was employed in the second strategy. The constructed rule was evaluated using leave-one-out approach. In addition, investigations were implemented using two data fusion models; LLDF and ILDF models. The HLDF is not considered in this research since the procedure of identity declaration is more of the machine learning approach.

3.2 Percentile Forward Feature Selection and Algorithms for Data Fusion

Generally, the new feature selection criteria and procedure is suitable for any high dimensional continuous dataset. This approach is suitable to reduce the dimension of data by identifying and selecting the most useful features that describe greatest separability among all possible groups. The distance criteria are basically suitable for data that are highly correlated among the features which is the case for the datasets applied in this research. Thus, this approach offers an advantage for multi-group data where information within features is redundant of each others. In addition, for the univariate feature ranking where the investigation of finding the most discriminative features to the least discriminative one, again, the selected criteria are able to assign less weight to features that have high variance. Further discussions of these criteria can be found in sections 3.2 and 3.3.

As for the selection procedure of highly discriminative features that is important for higher classification accuracy, this research proposed a new feature selection approach suitable for the aforementioned datasets. The new approach is called

Percentile Forward Feature Selection (PFFS). The input features for this procedure is the decreasing-order-ranked features according to the separability among groups. This separability is identified based on the average distance value calculated for every feature in the dataset. Highest average distance values are believed to be influential in giving higher classification accuracy. Therefore, the features must be ranked in decreasing order (from the highest to lowest average distance) for ease of feature selection. This ordering is purposely implemented to ensure features that provide larger separability among the existing groups are given the priority to be selected first. In this case, the common well known feature selection approaches such as the backward or stepwise feature selections are no longer applicable for such data ranking. However, if the features are ranked in increasing order, then, the backward feature selection can be applied accodingly.

To be more specific and for ease of feature selection implementation, the rank-and-percentile form is imposed on the decreasing-order-ranked features. The ranked features are ranked again from 1 (the top highest distance in the decreasing-order-ranked) to the *n*th (the bottom lowest distance in the decreasing-order-ranked) observation, accordingly, with the calculated percentile $\left[\left(\frac{f-i}{f}\right)\times 100\right]$, for $i=0,1,2,...,n-1$, where f is the number of fused features, respectively, for each new ranked feature. Say, for n=42 features i.e. for the LLDF model, the percentiles for the top highest distance, top second highest distance, top third highest distance, exectra, are represented by the 100.00% $\left(\mathcal{P}_{100}\right)$, 97.50% $\left(\mathcal{P}_{97.5}\right)$, 95.10% $\left(\mathcal{P}_{91.5}\right)$, respectively. The lowest distance is represented by 0.00% $\left(\mathcal{P}_{0}\right)$. For the implementation of ILDF model, the same processes are repeated accordingly. However, the percentile is

calculated using the number of features from single sensor, $(\tilde{t} = 1, 2, \ldots, p)$ and $(\tilde{n} = 1, 2, \ldots, q)$, for e-tongue and for e-nose, respectively.

Thus, in order to select a subset of features p from a set of k original features $P_{\mathscr{p}} = \{p_1, p_2, \ldots, p_k\}$, where $p < k$, the feature selection is performed by adding two features simultaneously in each feature subset. These can be described in the following order.

Feature Subset Generation	Number of Feature Included	Rank[Percentile]	Percentile Notation
Subset 1	2 features	1,2 [100.0-97.5]	$(\mathscr{P}_{100.0\rightarrow97.5})$
Subset 2	4 features	1,2,3,4 [100.0-92.6]	$(\mathscr{P}_{100.0\rightarrow92.6})$
Subset 3	6 features	1,2,3,4,5,6 [100.0-87.8]	$(\mathscr{P}_{100.0\rightarrow87.8})$
:	:	:	:

For the LLDF model, the first subset include two features of percentile $(\mathscr{P}_{100.0\rightarrow97.5})$, the second subset include other additional two features of percentile $(\mathscr{P}_{100.0\rightarrow92.6})$ and others. However, for the ILDF model, the inclusion involve only single feature from e-tongue and e-nose into the feature subset. Here, the feature selection guarantees the inclusion of features from both sensors start from the first feature subset. Unfortunately, this is not the case for feature selection in the LLDF model where feature with higher separability measure regardless of from which sensor will be selected and the inclusion of features equally from both sensors is uncontrollable.

In general, PFFS provides a simple and straight forward procedure in selecting the potential features. This is because the selection focuses on features with the highest separability that are only located in the top few percentiles of the overall ranked features. PFFS is also flexible where the recommended percentile to include

influential features is up to the researcher‟s preference or intuitive judgement. This makes PFFS a heuristic feature selection approach where practical method is employed. but it does not guarantee to be optimal or perfect, but sufficient for the classification goal. The heuristic method is, however, can be employed to speed up the process of finding a satisfactory solution. Moreover, the approach gurantee only discriminative features are selected, and the least discriminative features that are considered negligible will be discarded without affecting the classification performance.

Eventhough the decision of how many features are supposed to be in a feature subset is based on researcher‟s judgement, few conditions can be considered for the stopping criteria of the feature search. These conditions are primarily important for comparing the performance of the selection criteria either the unbounded or bounded Mahalanobis distance. Such conditions include

i. Univariate and multivariate distances for feature ranking and selected feature subset. This condition is useful in identifying which criterion would give the best feature ranking and able to recognize discriminant features distinctively.

ii. Discriminative power of the first selected feature subset. This condition evaluates the performance of the selected features in the first subset in achieving the higher classification accuracy. The higher the classification accuracy, the better the discriminative power of the selected features.

iii. Highest classification accuracy for the first feature subset. Basically, if the features selected from the first feature subset produce higher

classification accuracy compared to the other criterion, then, the criterion is better than the other.

iv. Maximum convergence of correct classification before deterioration based on feature subset. This condition looks for specific feature subset that could achieve highest classification accuracy before the classification performance deteriorates. The feature subset with less number of features but gives higher classification accuracy is considered as the best criterion.

Taking into considerations the selection procedure and criteria for feature selection, the main framework for algorithms of LLDF and ILDF can be illustrated below. Determination of the selected percentiles for the LLDF and ILDF models for this research were based on the following settings after taking into accout the predetermined conditions.

Model	Feature Source	Selected Percentile & Number of Feature (D^2)	Selected Percentile & Number of Feature (D_A^2)
LLDF	Fusion	$(\mathcal{P}_{100.0\rightarrow 68.2})$ 14 features	$(\mathcal{P}_{100.0\rightarrow 68.2})$ 14 features
ILDF	e-tongue	$(\mathcal{P}_{100.0\rightarrow 66.6})$ 4 features to $(\mathcal{P}_{100.0\rightarrow 44.4})$ 6 features	$(\mathcal{P}_{100.0\rightarrow 66.6})$ 4 features to $(\mathcal{P}_{100.0\rightarrow 44.4})$ 6 features
	e-nose	$(\mathcal{P}_{100.0\rightarrow 90.3})$ 4 features to $(\mathcal{P}_{100.0\rightarrow 83.8})$ 6 features	$(\mathcal{P}_{100.0\rightarrow 90.3})$ 4 features to $(\mathcal{P}_{100.0\rightarrow 83.8})$ 6 features

Figure 3.4. Proposed Percentiles for the Forward Feature Selection of the LLDF and ILDF Models using the Unbounded and Bounded Mahalanobis Distances.

A. Algorithm of Low Level Data Fusion (LLDF) with Feature Selection

[**Step 1**]

i. Fuse original data from e-nose and e-tongue.

ii. Compute the univariate unbounded Mahalanobis distance D^2 and the univariate bounded Mahalanobis distance D_A^2 of the fused features which then produce distance values $\left(\mathcal{F}_{D^2}\right)$ and $\left(\mathcal{F}_{D_A^2}\right)$, respectively. Rank $\left(\mathcal{F}_{D^2}\right)$ and $\left(\mathcal{F}_{D_A^2}\right)$ from the largest to smallest distance values. Filter the ranked fused features $\left(\mathcal{F}_{D^2}\right)$ and $\left(\mathcal{F}_{D_A^2}\right)$, and select the top 68.2 percentile $\left(\mathcal{P}_{100.0\rightarrow 68.2}\right)$ of the highest average distance from $\left(\mathcal{F}_{D^2}\right)$ and $\left(\mathcal{F}_{D_A^2}\right)$ as the subset of potential discriminators for the next step. The selection of 68.2 percentile is believed to fulfill the suggested four conditions.

[**Step 2**]

i. Utilize the subset of discriminators identified in step 1 to find the combination of features using $\left(\mathcal{P}_{100.0\rightarrow 68.2}\right)$ percentiles forward feature selection.

ii. Calculate the multivariate unbounded Mahalanobis distance D^2 and the multivariate bounded Mahalanobis distance D_A^2 for each selected subset.

iii. Construct a parametric classification rule using selected discriminators and training objects.

iv. Measure classification performance using leave-one-out error rate.

B. Algorithm of Intermediate Level Data Fusion (ILDF) with Feature Selection

[**Step 1**]

i. Compute the unbounded Mahalanobis distance D^2 and the bounded Mahalanobis distance D_A^2 for each single sensor data of e-tongue (*T*) and e-nose (*N*), respectively, to generate $\left(T_{D^2}\right)$, $\left(T_{D_A^2}\right)$, $\left(N_{D^2}\right)$, and $\left(N_{D_A^2}\right)$. Rank each $T_{D^2}, T_{D_A^2}, N_{D^2}$, and $N_{D_A^2}$ from the largest to smallest distance values.

ii. Fuse the ranked single features of e-tongue and e-nose $T_{D^2}, T_{D_A^2}$, N_{D^2}, and $N_{D_A^2}$ to get the ranked fused features $\left(T_{D^2} \text{ and } N_{D^2}\right)$ and $\left(T_{D_A^2} \text{ and } N_{D_A^2}\right)$. Filter the ranked fused features $\left(T_{D^2} \text{ and } N_{D^2}\right)$ and $\left(T_{D_A^2} \text{ and } N_{D_A^2}\right)$ and select the top 66.6 percentile $\left(\mathcal{P}_{100.0\rightarrow 66.6}\right)$ or up to 44.4 percentile $\left(\mathcal{P}_{100.0\rightarrow 44.4}\right)$ of the highest average distance from $\left(T_{D^2} \text{ and } T_{D_A^2}\right)$ and select the top 90.3 percentile $\left(\mathcal{P}_{100.0\rightarrow 90.3}\right)$ or up to 83.8 percentile $\left(\mathcal{P}_{100.0\rightarrow 83.8}\right)$ of the highest average distance from $\left(N_{D^2} \text{ and } N_{D_A^2}\right)$ as the subset of potential discriminators for the next step. The selected percentiles either $\left(\mathcal{P}_{100.0\rightarrow 90.3}\right)$ or $\left(\mathcal{P}_{100.0\rightarrow 83.8}\right)$ are considered to satisfy the illustrated conditions.

[**Step 2**]

i. Utilize the subset of discriminators identified in step 1 to find the combination of features using the $\left(\mathcal{P}_{100.0\rightarrow 90.3}\right)$ or $\left(\mathcal{P}_{100.0\rightarrow 83.8}\right)$ percentiles forward feature selection.

ii. Calculate the multivariate unbounded Mahalanobis distance D^2 and the multivariate bounded Mahalanobis distance D_A^2 for each selected subset.

iii. Construct a parametric classification rule using selected discriminators and training objects.

iv. Measure classification performance using leave-one-out error rate.

The overall algorithms can be further referred in Figure 3.1. The determination of percentile values in selecting the number of features in the LLDF and ILDF models were set based on the performance of each of the selected feature subsets.

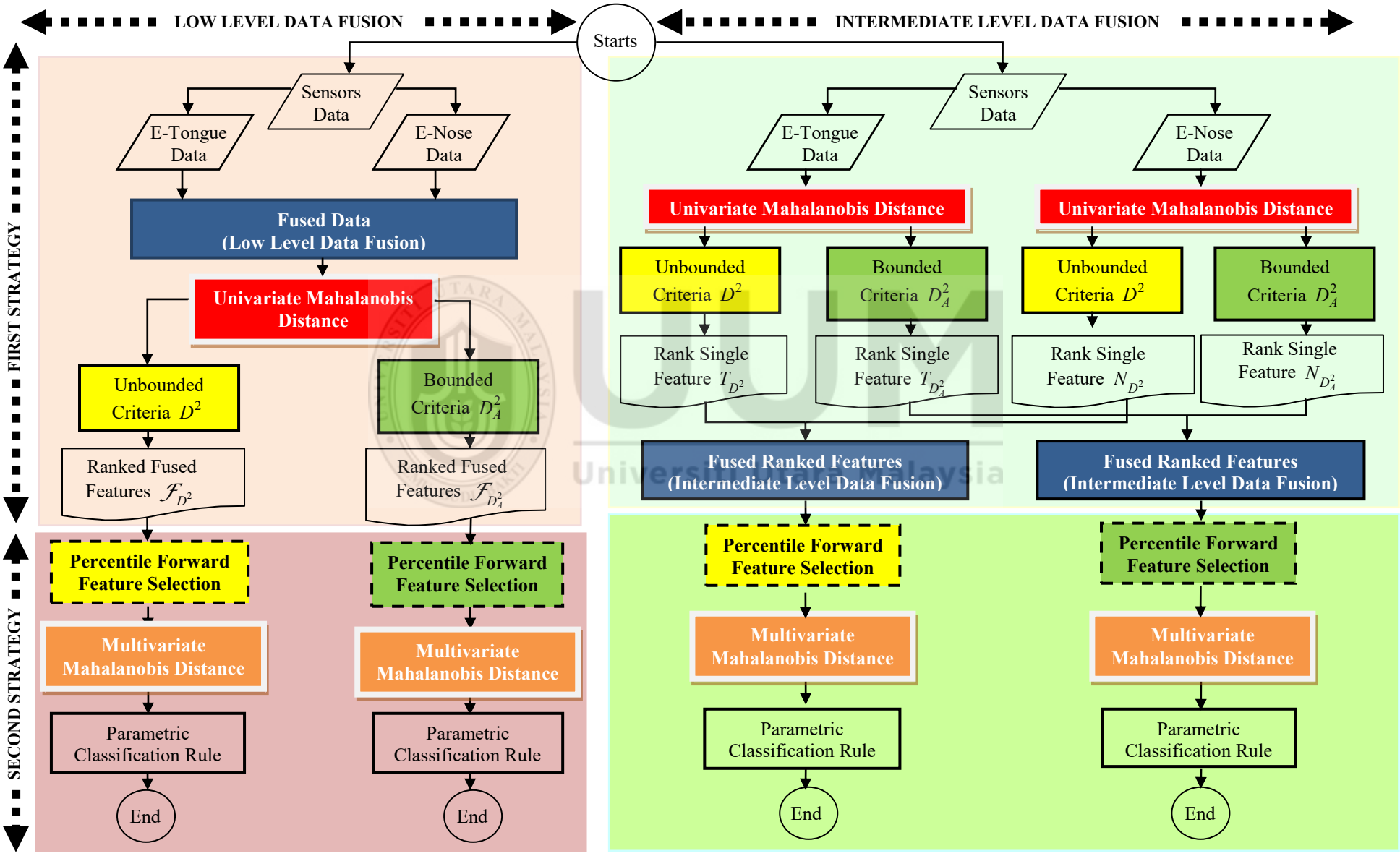

Figure 3.5. Proposed Feature Selection Strategies using the Unbounded $\left(D^2\right)$ and Bounded $\left(D_A^2\right)$ Mahalanobis Distances for LLDF and ILDF

3.3 Univariate Mahalanobis Distance

Suppose a random matrix for investigation $\mathbf{X}$ and $\mathbf{Y}$ are both partitioned into multi-group $(\mathcal{G} > 2)$, Π_i $(i = 1,2,\ldots,g)$, the ith of which comprises n_i rows and constitutes a random sample from p and q-variate populations. The matrices denote by $\mathbf{x}_{il}$ and $\mathbf{y}_{il}$ are signals data observed on the kth individual of the ith group $(i = 1,2,\ldots,g; k = 1,2,\ldots,n_i)$ from sensors e-tongue with dimension features $(\tilde{t} = 1,2,\ldots,p)$ and e-nose $(\tilde{n} = 1,2,\ldots,q)$. Hence, all objects in the multi-groups $\mathcal{G} = (\Pi_i,\ i = 1,2,\ldots,g)$ can be presented as $Z^T = (x^T, y^T)$ and the fused sensor data as $\mathcal{F}^T = (x+y)^T$.

Then $x_{i11}, x_{i22},\ldots, x_{in_ip}$ and $y_{i11}, y_{i22},\ldots, y_{in_iq}$ are assumed to be random samples from multivariate populations whose mean vectors $\boldsymbol{\mu}_{il} = (\bar{\mathbf{x}}_{i1}, \bar{\mathbf{x}}_{i2},\ldots, \bar{\mathbf{x}}_{ip})$, $\boldsymbol{\mu}_{il} = (\bar{\mathbf{y}}_{i1}, \bar{\mathbf{y}}_{i2},\ldots, \bar{\mathbf{y}}_{iq})$ and dispersion matrix $\boldsymbol{\Sigma}_{il}$. Let $\bar{\mathbf{x}}_{il} = \frac{1}{n_i}\sum_{i=1}^{g}\sum_{l=1}^{p} x_{il}$,

$\bar{\mathbf{y}}_{il} = \frac{1}{n_i}\sum_{i=1}^{g}\sum_{l=1}^{q} y_{il}$, $s_{il}(x) = \frac{\sum_{i=1}^{n_i}(x_{il}-\bar{x}_l)(x_{il}-\bar{x}_l)'}{n_i-1}$ and $s_{il}(y) = \frac{\sum_{i=1}^{n_i}(y_{il}-\bar{y}_l)(y_{il}-\bar{y}_l)'}{n_i-1}$,

where $\bar{\mathbf{x}}_{il}$ and $\bar{\mathbf{y}}_{il}$ are the maximum likelihood estimators of μ_{il} and $s_{il}(x)$ and $s_{il}(y)$ are the unbiased estimators of $\boldsymbol{\Sigma}_i$ for $(i = 1,2,\ldots,g)$, $(\tilde{t} = 1,2,\ldots,p)$ and $(\tilde{n} = 1,2,\ldots,q)$, respectively.

The features‟ structure for single and fused sensors data are described as in Table 3.1. Let begins with univariate Mahalanobis distance Δ^2 and Δ_A^2, being estimated

using D^2 and D_A^2. Selection of features begins by applying the univariate distance functions using D^2 and D_A^2 to all available features which contains several groups.

Table 3.1

Illustration of Single Sensor Data and Fused Data

Group (*i*)	**Observation (*k*)**	**Fused Features (*I*)**							
		X_{ik1}	X_{ik2}	…	X_{ikp}	Y_{ik1}	Y_{ik2}	…	$Y_{ik(p+q)}$
		e-tongue Data				**e-nose Data**			
		X_{ik1}	X_{ik2}	…	X_{ikp}	Y_{ik1}	Y_{ik2}	…	Y_{ikq}
1	1	x_{111}	x_{112}	…	x_{11p}	y_{111}	y_{112}	…	y_{11q}
1	2	x_{121}	x_{122}	…	x_{12p}	y_{121}	y_{122}	…	y_{12q}
⋮	⋮	⋮	⋮	…	⋮	⋮	⋮	…	⋮
1	n_1	x_{1n_11}	x_{1n_12}	…	x_{1n_1p}	y_{1n_11}	y_{1n_12}	…	y_{1n_1q}
2	1	x_{211}	x_{212}	…	x_{21p}	y_{211}	y_{212}	…	y_{21q}
2	2	x_{221}	x_{222}	…	x_{22p}	y_{221}	y_{222}	…	y_{22q}
⋮	⋮	⋮	⋮	…	⋮	⋮	⋮	…	⋮
2	n_2	x_{2n_21}	x_{2n_22}	…	x_{2n_2p}	y_{2n_21}	y_{2n_22}	…	y_{2n_2q}
⋮	⋮	⋮	⋮	…	⋮	⋮	⋮	…	⋮
g	1	x_{g11}	x_{g12}	…	x_{g1p}	y_{g11}	y_{g12}	…	y_{g1q}
g	2	x_{g21}	x_{g22}	…	x_{g2p}	y_{g21}	y_{g22}	…	y_{g2q}
⋮	⋮	⋮	⋮	…	⋮	⋮	⋮	…	⋮
g	n_g	x_{gn_g1}	x_{gn_g2}	…	x_{gn_gp}	y_{gn_g1}	y_{gn_g2}	…	y_{gn_gq}

For simplicity of showing the ${}^gC_2$ pairwise-groups distances per feature, Table 3.2 represents the combinations of the resulting distances for D^2 and D_A^2. The diagonal elements are the distances of the same centroids while either of the pair-wise distances on the lower or upper diagonal is useful for the required distance measure. In our case, we used the lower diagonal pair-wise groups‟ distances.

Table 3.2

The ${}^{g}C_{2}$ Pairwise Mahalanobis Distance for Univariate Feature

Distances for g groups produce ${}^{g}C_{2}$ group pairs $(g=9)$									
	G_1	G_2	G_3	G_4	G_5	G_6	G_7	G_8	G_9
G_1	0	$D^2_{1,2}$	$D^2_{1,3}$	$D^2_{1,4}$	$D^2_{1,5}$	$D^2_{1,6}$	$D^2_{1,7}$	$D^2_{1,8}$	$D^2_{1,9}$
G_2	$D^2_{2,1}$	0	$D^2_{2,3}$	$D^2_{2,4}$	$D^2_{2,5}$	$D^2_{2,6}$	$D^2_{2,7}$	$D^2_{2,8}$	$D^2_{2,9}$
G_3	$D^2_{3,1}$	$D^2_{3,2}$	0	$D^2_{3,4}$	$D^2_{3,5}$	$D^2_{3,6}$	$D^2_{3,7}$	$D^2_{3,8}$	$D^2_{3,9}$
G_4	$D^2_{4,1}$	$D^2_{3,2}$	$D^2_{4,3}$	0	$D^2_{4,5}$	$D^2_{4,6}$	$D^2_{4,7}$	$D^2_{4,8}$	$D^2_{4,9}$
G_5	$D^2_{5,1}$	$D^2_{5,2}$	$D^2_{5,3}$	$D^2_{5,4}$	0	$D^2_{5,6}$	$D^2_{5,7}$	$D^2_{5,8}$	$D^2_{5,9}$
G_6	$D^2_{6,1}$	$D^2_{6,2}$	$D^2_{6,3}$	$D^2_{6,4}$	$D^2_{6,5}$	0	$D^2_{6,7}$	$D^2_{6,8}$	$D^2_{6,9}$
G_7	$D^2_{7,1}$	$D^2_{7,2}$	$D^2_{7,3}$	$D^2_{7,3}$	$D^2_{7,5}$	$D^2_{7,6}$	0	$D^2_{7,8}$	$D^2_{7,9}$
G_8	$D^2_{8,1}$	$D^2_{8,2}$	$D^2_{8,3}$	$D^2_{8,4}$	$D^2_{8,5}$	$D^2_{8,6}$	$D^2_{8,7}$	0	$D^2_{8,9}$
G_9	$D^2_{9,1}$	$D^2_{9,2}$	$D^2_{9,3}$	$D^2_{9,4}$	$D^2_{9,5}$	$D^2_{9,6}$	$D^2_{9,7}$	$D^2_{9,8}$	0

Equations (3.1) to (3.4) are applied to obtain the estimated pairwise-groups (i, j) distances of D^2 and D^2_A, respectively, where $(i=1,2,\ldots,g;\ j=i+1,\ldots,g-1)$ and $i \neq j$. The calculation of univariate Mahalanobis distance was simplified by equation (3.3) where the inverse of s_{ij} in equation (3.2) is basically the reciprocal of the respective common variance.

$$\mathbf{D}^2_{ij} = \left(\bar{\mathbf{X}}_i - \bar{\mathbf{X}}_j\right)^T \mathbf{S}^{-1}_{ij} \left(\bar{\mathbf{X}}_i - \bar{\mathbf{X}}_j\right) \tag{3.1}$$

$$\mathbf{S}_{ij} = \frac{\sum_{k=1}^{n_1}\left(X_{ik} - \bar{X}_i\right)\left(X_{ik} - \bar{X}_i\right)^T + \sum_{k=1}^{n_2}\left(X_{jk} - \bar{X}_j\right)\left(X_{jk} - \bar{X}_j\right)^T}{n_1 + n_2 - 2} \tag{3.2}$$

$$\mathbf{D}^2_{ij} = \frac{\left(\bar{\mathbf{X}}_i - \bar{\mathbf{X}}_j\right)^2}{S_{ij}} \tag{3.3}$$

$$\mathbf{D}^2_{A(ij)} = \frac{D^2_{ij}}{4 + D^2_{ij}} \tag{3.4}$$

The summation of each of the resulting ${}^{g}C_2$ pair-wise distances using equation (3.1) divided by ${}^{g}C_2$ gave the average Mahalanobis distance of each evaluated feature. It can be simply obtained using equation (3.5) for features $\tilde{t} = 1, 2, \ldots, p$, $\tilde{n} = 1, 2, \ldots, q$, and $f = 1, 2, \ldots, p, \ldots p+q$, respectively. The values of D^2 basically gives the $[0, \infty)$ *unbounded Mahalanobis distance*. The process was repeated for the entire features of interest, independently.

$$\mathbf{D}^2_{l} = \frac{1}{{}^{g}C_2} \sum_{i=1}^{g} \sum_{j=i+1}^{g-1} D^2_{ij} \tag{3.5}$$

$$\mathbf{D}^2_{A} = \frac{1}{{}^{g}C_2} \sum_{i=1}^{g} \sum_{j=i+1}^{g-1} D^2_{A(ij)} \tag{3.6}$$

Once the ${}^{g}C_2$ pair-wise distances of D^2_{ij} using equation (3.1) were available, each element of D^2_{ij} was computed using equation (3.4) which was originally derived by equation (2.16), where equal *a priori* probabilities $(\pi_1 = \pi_2 = 0.5)$ were applied. Then only the average distance for the second criterion that is the $[0, 1]$ *bounded Mahalanobis distance value* is obtained. The respective calculation for the average bounded Mahalanobis distance is as in equation (3.6). Such process was repeated for the entire features of interest. The maximum average distance values of D^2 and D^2_A for each feature are then ranked from the largest to smallest. The objective of performing univariate Mahalanobis distance is to select the best feature set ranked in the percentiles of $(\mathcal{P}_{100.0 \to 90.3})$, $(\mathcal{P}_{100.0 \to 83.8})$ or up to $(\mathcal{P}_{100.0 \to 68.2})$ of the largest ranked

distance of the LLDF. While the percentiles of $\left(\mathcal{P}_{100.0\to 90.3}\right)$ up $\left(\mathcal{P}_{100.0\to 83.8}\right)$ for e-nose features and $\left(\mathcal{P}_{100.0\to 66.6}\right)$ up to $\left(\mathcal{P}_{100.0\to 44.4}\right)$ for e-tongue features for the ILDF model as the input for the next strategy.

3.4 Multivariate Mahalanobis Distance

The next search for maximum average distance for subset of features of $\tilde{t} < p$, $\tilde{n} < q$ or $f < p+q$ selected from the specified percentiles of fused and single ranked features in the previous step is described using the multivariate Mahalanobis distance. The quadratic form between the centroids of those features is given by equation (3.7) where $\mu_{il} = \left(\bar{x}_{i1}, \bar{x}_{i2}, \ldots, \bar{x}_{il}\right)$ and $\mu_{jl} = \left(\bar{x}_{j1}, \bar{x}_{j2}, \ldots, \bar{x}_{jl}\right)$ represent the vector mean values for the multi-groups of $\left(i = 1, 2, \ldots, g;\ j = i+1, \ldots, g-1\right)$, $i \neq j$, $l = \tilde{t}$ or $\tilde{n}$ or f, and $\mathbf{S}_{ijl}^{-1}$ is the inverse of the pooled covariance matrix for the features $\tilde{t} = 1, 2, \ldots, p$, $\tilde{n} = 1, 2, \ldots, q$, and $f = 1, 2, \ldots, p, \ldots p+q$, respectively for e-tongue, e-nose and fusion.

$$\mathbf{D}_{ijl}^{2} = \left(\bar{\mathbf{X}}_{il} - \bar{\mathbf{X}}_{jl}\right)^{T} \mathbf{S}_{ijl}^{-1} \left(\bar{\mathbf{X}}_{il} - \bar{\mathbf{X}}_{jl}\right) \tag{3.7}$$

Unlike in the first stage where features were assumed to be independent, features were not assumed to be independent for the second stage. This was to enable comparison between D^2 and D_A^2 in their ranking of discriminant features. In the second stage, the output of the unbounded Mahalanobis distance D_{ijl}^2 was our interest and the ${}^{g}C_2$ pair-wise distances for the aforementioned percentiles of the specified subset of features such as in Table 3.2 were obtained. Next, the bounded Mahalanobis distance D_A^2 were obtained using equation (3.4) where we recalculated

each element of D_{ijl}^2 with equal *a priori* probabilities. Then the average distance for the criteria of the unbounded and bounded Mahalanobis distance values were obtained. The findings of the best feature subset within the selected percentiles were embedded with the classification process where the performance of selected feature subsets was compared based on the perspective of bounded and unbounded criteria.

For the calculation of the proposed Mahalanobis distances, this study modified the original R-code of the pair-wise Mahalanobis distances for multi-grouped data available in package HDMD version 1.2 (26-02-2013) of Statistical Analysis Tools for High Dimension Molecular Data by McFerrin and McFerrin (2013). The modified code is available in Appendix A. Several reasons to modify the original R-code are as follows.

i. The original R-code was developed to fulfill the high dimensional molecular data. In contrast, this deals with sensory data. Several attempts were executed with the HDMD using datasets in this study, but failed due to the input covariance matrix used. The input data for the original R-code is either the covariance or correlation matrix. However, this study uses the inverse covariance matrix.

ii. The original HDMD only calculates the unbounded Mahalanobis distance. However, this study requires both calculations of the unbounded and bounded Mahalanobis distances. Thus, to fulfill the calculations requirement of the proposed unbounded and bounded Mahalanobis distances, the original code needs to be modified accordingly.

iii. This study has earned the permission from the authors (McFerrin & McFerrin, 2013) to modify the original R-code so that the computation of the pair-wise unbounded and bounded Mahalanobis distances can be implemented.

iv. Ultimately, the main focus of this study was to perform selection of discriminative features subsets that produces high classification accuracy. In specific, this phase was involved in the implementation of the univariate Mahalanobis distance which entails many processes. However, the calculation of the multivariate Mahalanobis distance is to measure the average distance among multi-group data which consider more than one feature. Thus, the average distance obtained from the multivariate Mahalanobis distance is not as significant as the average distance of the univariate distance.

Finally, the function of the univariate and multivariate Mahalanobis distance can be highlighted by emphasizing the output produced. By performing the univariate Mahalanobis distance, a thorough evaluation of all features was made based on the measure of separability. Feature with higher average distance demonstrates large separability. As a result, all available features in the dataset were ranked according to the separability scored. The ranked features are the input for the classification process. Unlike the univariate case, multivariate Mahalanobis distance was applied to measure the average distance for more than one feature. The calculation involved in the feature subsets based on the selected percentiles. Here, the interest was to show the separability of the selected features in the feature subsets.

3.5 Bounded and Unbounded Mahalanobis Distances as Criteria for Discriminant Features

Total distance for the ${}^{g}C_{2}$ pair-wise groups of the first criterion that was the *unbounded Mahalanobis* D_{ij}^{2} describes in Table 3.2, can be obtained by adding all the multi-group distance combinations of $(i=1,2,\ldots,g;\ j=i+1,\ldots,g-1)$ where $i \neq j$. In our case, lower diagonal element of distance of Table 3.2 was summed for the calculation purposes. Generally, the summation of the lower diagonal divided by the value of ${}^{g}C_{2}$ gives the average of Mahalanobis distance of each evaluated feature or feature subset. In this process, the *expected value approach* was used to obtain the average distance, and the indication of good feature or subset of features was the *maximum average value*. Basically it takes value in the range $[0,\infty)$ with higher values represents greater separability of group centroids.

In the case of multi-groups classification problems, when the expected value approach was applied and the pair-wise groups distance was adopted, serious difficulty occurred. For example, in a set of *g*-groups problem where $g > 2$, if one of the lower diagonal elements of D_{ij}^{2} is exceptionally large, it would lead to a high value of average distance. This would then leads to failure of representing the average separability of the *g* groups. Thus, by transforming the boundless distance values produced by Δ^{2} using Δ_{A}^{2} before the averaging process would overcome the problem. The transformed value lies within a finite range 0 to 1, known as *bounded Mahalanobis distance*. The bounded Mahalanobis distance $\Delta_{A}^{2} \in [0,1]$ corresponds to a normalizing transformation on $\Delta^{2} \in [0,\infty)$. The normalization process was derived

from the distribution free of the relationship between Δ^2 and the upper bounded Bayesian error probability, P_e, as outlined in equation (2.16) in reported review in section 2.2.2.2.

Since Δ_A^2 is a function Δ^2, and Δ^2 can be solved in terms of Δ_A^2, the distribution free upper bound (2.18) which was expressed in terms of Δ^2 can be expressed by Δ_A^2. This implies that the error bound associated with Δ_A^2 was the same as that associated with Δ^2. As Δ^2 increases unboundedly, Δ_A^2 increases upper bounded by 1. Therefore, because of the boundedness nature, Δ_A^2 is expected to perform better than Δ^2 in a multi-groups classification problem (Ray & Turner, 1992).

3.6 Proposed Discriminant Analysis for Low Level Data Fusion

The constructions of the full sequence of PFFS were completed by joining all the pieces of the proposed strategies begin with the LLDF model. For the LLDF model, fused features were the input. Once feature and/or subset of features with greater separability between groups‟ centroids are obtained, next step to perform classification is continued. Feature and/or subset of features with the largest average distance among multi-groups would lead to accurate group recognition during classification process. It is expected that the higher the average distance, the better the accuracy of classification rule. In this study, selection of features was implemented using percentile forward selection approach. The process begins with no discriminator in a classification rule. Then, the best potential discriminator with the maximum criterion value approach (largest average distance) from the percentile ranking of the fused features $\mathcal{F}_{D^2} = \left[f_{\mathcal{P}_{100.0}}, f_{\mathcal{P}_{97.5}}, f_{\mathcal{P}_{91.5}}, \ldots, f_{\mathcal{P}_{0.0}} \right]$ and

$\mathcal{F}_{D_A^2} = \left[f_{\mathcal{P}_{100.0}}, f_{\mathcal{P}_{97.5}}, f_{\mathcal{P}_{95.1}}, \ldots, f_{\mathcal{P}_{0.0}} \right]$ is selected for both criterion Δ^2 and Δ_A^2, respectively.

This process is described in Algorithms 3.1 and 3.2. Since $f_{\mathcal{P}_{100.0}}$ was considered the best feature for its greatest separability between all the pair-wise multi-groups centroids, the first step was compulsory to begin with $f_{\mathcal{P}_{100.0}}$. The next selection of discriminator was based on the remaining $\left(\mathcal{F}_{D^2} - f_{\mathcal{P}_{100.0}}\right)$. The second discriminator was paired with $f_{\mathcal{P}_{100.0}}$ that satisfy the condition i.e. the next largest average distance from the paired discriminators. The processes continue until all respective features in $\mathcal{F}_{D^2} = \left[f_{\mathcal{P}_{100.0}}, f_{\mathcal{P}_{97.5}}, f_{\mathcal{P}_{91.5}}, \ldots, f_{\mathcal{P}_{0.0}} \right]$ and $\mathcal{F}_{D_A^2} = \left[f_{\mathcal{P}_{100.0}}, f_{\mathcal{P}_{97.5}}, f_{\mathcal{P}_{95.1}}, \ldots, f_{\mathcal{P}_{0.0}} \right]$, but limited to cumulative $\left(\mathcal{P}_{100.0 \to 68.2}\right)$ were selected as subset of discriminators for the development of classification rule and error assessment.

The implementation of PFFS using separability criterion in this fashion was expected to overcome the nesting problem since focus was given to the top highest percentiles of features with the largest average distance for both unbounded and bounded Mahalanobis distances, Δ^2 and Δ_A^2, respectively. In addition, since only efficient features were included, this approach was computationally effective because inefficient features were not assessed in the search by $\left(\mathcal{F}_{D^2} - f_{\mathcal{P}_{100.0}} - f_{\mathcal{P}_{97.5}} - f_{\mathcal{P}_{95.1}} \cdots - f_{\mathcal{P}_{68.2}}\right)$, and $\left(\mathcal{F}_{D_A^2} - f_{\mathcal{P}_{100.0}} - f_{\mathcal{P}_{97.5}} - f_{\mathcal{P}_{95.1}} \cdots - f_{\mathcal{P}_{68.2}}\right)$. Furthermore, the search does not require any stopping rule to be applied which simplified the implementation process.

A simple and straight forward approach to measure accuracy of the selected rule is the apparent error rate (AER). Known to be suitable for any type of classification and does not require any knowledge of the parent population. It measures how effective the rule was in allocating objects to its correct groups. The proportion of objects in the training sample that was misclassified by the applied rule was obtained from the confusion matrix table. As criticized for being bias, the apparent error rate can be improved through a resampling procedure such as leave-one-out. Leave-one-out attempts to remove the bias of the apparent error rate and mostly applied for small sample size. It uses an object as a test set and the remaining $(n-1)$ to build a rule. The advantage of this approach is that the size of training set is almost as large as the entire data set in each repetition and suitable for small data set. The process of these steps is describes in the following steps:

i. omit an object k from sample in turn, for $k = 1, 2, \ldots, n_i$ where $i = 1, 2, \ldots, g$ and $n = n_1 + n_2 + \ldots + n_g$,
ii. based on the remaining objects, $(n-1)$, use the forward selection to select l from the original $\mathcal{F}^T = (x+y)^T$ fused features,
iii. construct a classification rule using those selected l features (discriminators) and then $(n-1)$ training objects,
iv. classify the omitted object k using the corresponding l discriminators, and
v. once all objects have been taken out in turn, compute the proportion of misclassified object.

The proposed strategy to perform discriminant analysis was based on the linear discriminant analysis (LDA) for the classification purposes. The cross validation

using the leave-one-out approach was implemented in the classification rule using the option `CV=TRUE` in the following function `fit <- lda(Group ~ feature1 + feature2 + … + featureN,data=data.file, na.action= "na.omit", CV=TRUE)`. The algorithms for fused feature ranking using the unbounded and bounded Mahalanobis distances are outlined in Algorithms 3.1 and 3.2, respectively. And the overall steps of the discriminant analysis of the LLDF model based on the unbounded and bounded Mahalanobis distances are illustrated in Figures 3.6 and 3.7, respectively.

Algorithm 3.1

Fused Feature Ranking for Unbounded $\left[0,+\infty\right)$ Mahalanobis Distance [D^2] for Low Level Data Fusion ~ Strategy 1

Input : $Z=\left[x_1,x_2,\ldots,x_p,y_1,y_2,\ldots,y_{p+q}\right]$ - original fused feature set

Output : Ranking criterion function $\left(\mathcal{P}_{100.0\rightarrow 68.2}\right)$ and ranked fused features $\mathcal{F}_{D^2}=\left[f_{\mathcal{P}_{100.0}},f_{\mathcal{P}_{97.5}},f_{\mathcal{P}_{95.1}},\ldots,f_{\mathcal{P}_{0.0}}\right]$

Step 1 : **Initialize** $\mathcal{F}_{D^2}=\left[f_1,f_2,\ldots,f_{p+q}\right]$

Step 2 : **For** each feature $\left(X,Y\right)\in Z^{p+q}$

i. Calculate the Mahalanobis distance for ${}^{g}C_2$ pair-wise groups using criterion (3.3) for each fused features in Z^{p+q}

$$\mathbf{D}_{ij}^2=\frac{\left(\bar{\mathbf{X}}_i-\bar{\mathbf{X}}_j\right)^2}{S_{ij}},\quad i=1,2,\ldots,g\,;\,j=i+1,\ldots,g-1$$

ii. Calculate the average distance for ${}^{g}C_2$ pair-wise groups using criterion (3.5) for each fused features in Z^{p+q}

$$\mathbf{D}_{ij}^2=\frac{1}{{}^{g}C_2}\sum_{i=1}^{g}\sum_{j=i+1}^{g-1}D_{ij}^2$$

iii. Calculate the average distance for each fused features in Z^{p+q} by obtaining $f_l=\sum_{l=1}^{p+q}D_l^2$ where $l=1,2,\ldots,p+q$

iv. Store f_l into $\mathcal{F}_{D^2}=[\ ]$

Step 3 : **End** of **for** loop

Step 4 : **Rank** feature $\left(X,Y\right)\in Z^{p+q}$ (from largest to smallest values) based on $\mathcal{F}_{D^2}=\left[f_1,f_2,\ldots,f_{p+q}\right]$

Step 5 : **Filter** $\mathcal{F}_{D^2}=\left[f_1,f_2,\ldots,f_{p+q}\right]$ to keep only first $\left(\mathcal{P}_{100.0\rightarrow 68.2}\right)$ of $\mathcal{F}_{D^2}=\left[f_{\mathcal{P}_{100.0}},f_{\mathcal{P}_{97.5}},f_{\mathcal{P}_{95.1}},\ldots,f_{\mathcal{P}_{0.0}}\right]$ for the next feature subset selection

---

Algorithm 3.2

Fused Feature Ranking for Bounded $[0,1]$ Mahalanobis Distance [D_A^2] for Low Level Data Fusion ~ Strategy 1

Input : $Z = \left[x_1, x_2, \ldots, x_p, y_1, y_2, \ldots, y_{p+q} \right]$ - original fused feature set

Output : Ranking criterion function $\left(\mathcal{P}_{100.0 \to 68.2} \right)$ and ranked fused features $\mathcal{F}_{D_A^2} = \left[f_{\mathcal{P}_{100.0}}, f_{\mathcal{P}_{97.5}}, f_{\mathcal{P}_{95.1}}, \ldots, f_{\mathcal{P}_{0.0}} \right]$

Step 1 : **Initialize** $\mathcal{F}_{D^2} = \left[f_1, f_2, \ldots, f_{p+q} \right]$

Step 2 : **For** each feature $\left(X, Y \right) \in Z^{p+q}$

i. Calculate the Mahalanobis distance for ${}^{g}C_2$ pair-wise groups using criterion (3.4) with equal *a priori* probability for each fused features in Z^{p+q}

$$\mathbf{D}^2_{A(ij)} = \frac{D_{ij}^2}{4 + D_{ij}^2}, \quad i = 1, 2, \ldots, g \; ; \; j = i+1, \ldots, g-1$$

ii. Calculate the average distance of ${}^{g}C_2$ pair-wise groups using criterion (3.6) for each fused features in Z^{p+q}

$$D^2{}_A = \frac{1}{{}^{g}C_2} \sum_{i=1}^{g} \sum_{j=i+1}^{g-1} D_{Aij}^2$$

iii. Calculate the average distance for each fused features in Z^{p+q} by obtaining $f_l = \sum_{l=1}^{p+q} D_{Al}^2$ where $l = 1, 2, \ldots, p+q$

iv. Store f_l into $\mathcal{F}_{D_A^2} = [\ ]$

Step 3 : **End** of **for** loop

Step 4 : **Rank** feature $\left(X, Y \right) \in Z^{p+q}$ (from largest to smallest values) based on $\mathcal{F}_{D_A^2} = \left[f_1, f_2, \ldots, f_{p+q} \right]$

Step 5 : **Filter** $\mathcal{F}_{D_A^2} = \left[f_1, f_2, \ldots, f_{p+q} \right]$ to keep only first $\left(\mathcal{P}_{100.0 \to 68.2} \right)$ of $\mathcal{F}_{D_A^2} = \left[f_{\mathcal{P}_{100.0}}, f_{\mathcal{P}_{97.5}}, f_{\mathcal{P}_{95.1}}, \ldots, f_{\mathcal{P}_{0.0}} \right]$ for the next feature subset selection

---

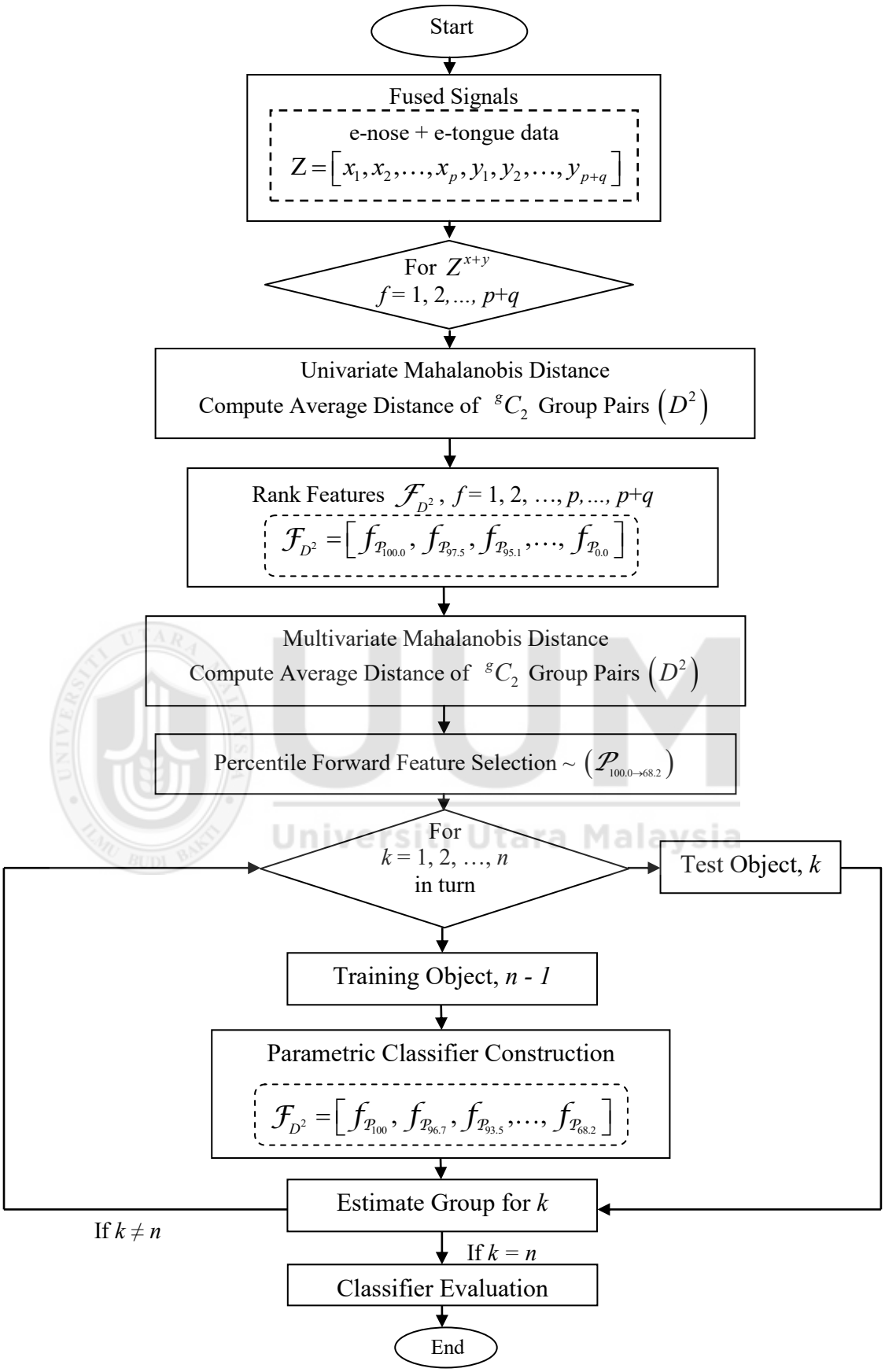

Figure 3.6. Flow Chart of Discriminant Analysis for the LLDF Model (Criterion D^2)

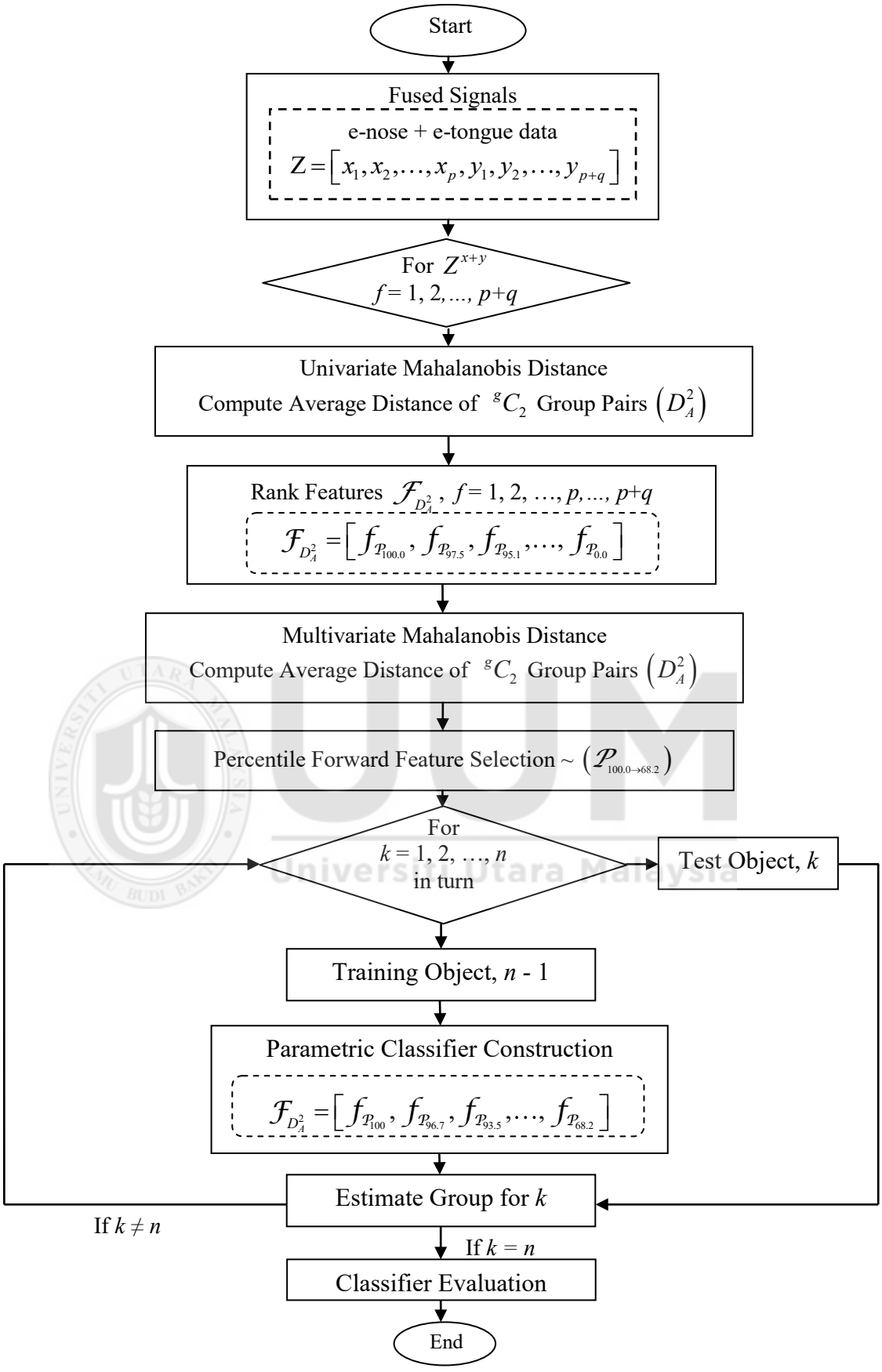

Figure 3.7. Flow Chart of Discriminant Analysis for the LLDF Model (Criterion D_A^2)

3.7 Proposed Discriminant Analysis for Intermediate Level Data Fusion

Next, the construction of the whole sequence of PFFS and other proposed strategies for the ILDF model. The strategies of this model were similar to the LLDF model, but the difference is the way the features being fused. Unlike the previous model where fusion of features begun first, for the intermediate level, fusion takes place after the ranking and filtering procedures. The procedure is described in step 1 for the second model in previous section 3.1, as well as Algorithms 3.3 and 3.4. Evaluation and calculation of the criteria Δ^2 and Δ_A^2, were implemented for each features from e-tongue, $X=\left[x_1,x_2,\ldots,x_p\right]$ and e-nose, $Y=\left[y_1,y_2,\ldots,y_q\right]$, separately. Once the ranking was done, the evaluation of subset features was performed by the inclusion of top best $\left(\mathcal{P}_{100.0\rightarrow 90.3}\right)$ up to $\left(\mathcal{P}_{100.0\rightarrow 44.4}\right)$ of each sensor based on the maximum expected value of average distance.

The top best features of e-tongue, $\left(\mathcal{P}_{100.0\rightarrow 66.6}\right)$ to $\left(\mathcal{P}_{100.0\rightarrow 44.4}\right)\in T_{D^2}$ and e-nose, $\left(\mathcal{P}_{100.0\rightarrow 90.3}\right)$ to $\left(\mathcal{P}_{100.0\rightarrow 83.8}\right)\in N_{D^2}$ were fused to form fusion of filtered ranking features $\mathcal{F}_{D^2}=\left[(\mathcal{P}_{100.0\rightarrow 44.4})\in T_{D^2},(\mathcal{P}_{100.0\rightarrow 83.8})\in N_{D^2}\right]$ for criterion Δ^2. While best features of e-tongue, $\left(\mathcal{P}_{100.0\rightarrow 66.6}\right)$ to $\left(\mathcal{P}_{100.0\rightarrow 44.4}\right)\in T_{D_A^2}$ and e-nose, $\left(\mathcal{P}_{100.0\rightarrow 90.3}\right)$ or $\left(\mathcal{P}_{100.0\rightarrow 83.8}\right)\in N_{D_A^2}$ are then fused for criterion Δ_A^2 resulting fusion of filtered ranking features $\mathcal{F}_{D_A^2}=\left\lfloor(\mathcal{P}_{100.0\rightarrow 44.4})\in T_{D_A^2},(\mathcal{P}_{100.0\rightarrow 83.8})\in N_{D_A^2}\right\rfloor$. Using this model, features from either of the e-tongue or e-nose have the equal chances to be chosen for the first and the following search.

The best potential discriminator $f_1 \in \mathcal{F}_{D^2}$ and $f_1 \in \mathcal{F}_{D_A^2}$ with the maximum criteria of Δ^2 and Δ_A^2 from $\mathcal{F}_{D^2} = \left[(\mathcal{P}_{100.0 \to 44.4}) \in T_{D^2}, (\mathcal{P}_{100.0 \to 83.8}) \in N_{D^2} \right]$ and $\mathcal{F}_{D_A^2} = \left\lfloor (\mathcal{P}_{100.0-44.4}) \in T_{D_A^2}, (\mathcal{P}_{100.0-83.8}) \in N_{D_A^2} \right\rfloor$, respectively, were selected. This process is previously outlined in strategy 1 of the intermediate level model as well as Algorithms 3.3 and 3.4. The next selection of discriminator was based on the remaining $\left(\mathcal{F}_{D^2} - \left(f_{100.0} \in T_{D^2} \text{ or } f_{100.0} \in N_{D^2} \right) \right)$. The second discriminator was paired with the next selected $f_l \in \mathcal{F}_{D^2}$ that maximize the criterion i.e. the next largest average distance from the paired discriminators. The same procedure was repeated for the second criterion Δ_A^2.

The processes continued until all features in $\mathcal{F}_{D^2} = \left[(\mathcal{P}_{100.0 \to 44.4}) \in T_{D^2}, (\mathcal{P}_{100.0 \to 83.8}) \in N_{D^2} \right]$ and $\mathcal{F}_{D_A^2} = \left\lfloor (\mathcal{P}_{100.0 \to 44.4}) \in T_{D_A^2}, (\mathcal{P}_{100.0 \to 83.8}) \in N_{D_A^2} \right\rfloor$ were selected as subset of discriminators for the development of classification rule and error assessment. The implementation of percentile forward selection using separability criterion in this fashion was expected to overcome the nesting problem for the intermediate level data fusion since focus was given to the best features with the largest average distance for both unbounded and bounded Mahalanobis distances, D^2 and D_A^2, respectively. In addition, since only efficient features were included, this approach is computationally effective because inefficient features were not assessed in the search by $\left(\mathcal{F}_{D^2} - \mathcal{F}_{D^2}(\mathcal{P}_{100.0}) \cdots - \mathcal{F}_{D^2}(\mathcal{P}_{44.4}) \right)$ and $\left(\mathcal{F}_{D_A^2} - \mathcal{F}_{D_A^2}(\mathcal{P}_{100.0}) \cdots - \mathcal{F}_{D_A^2}(\mathcal{P}_{83.8}) \right)$, for e-tongue and e-nose, respectively. Furthermore, the search does not require any stopping rule to be applied taking into

consideration of predetermined conditions, which simplified the implementation process.

The rest of the procedures that involve the leave-one-out error estimation and the implementation of parametric rule based on LDA similar to the low level data fusion model were repeated for the intermediate level model. The algorithms for single feature ranking using the unbounded and bounded Mahalanobis distances are outline in Algorithms 3.3 and 3.4, respectively. While the overall steps of the discriminant analysis of the ILDF model based on the unbounded and bounded Mahalanobis distances are illustrated in Figures 3.8 and 3.9, respectively.

Algorithm 3.3

Single Feature Ranking for Unbounded $\left[0,+\infty\right)$ Mahalanobis Distance [D^2] for Intermediate Level Data Fusion ~ Strategy 1

Input : $X=\left[x_1,x_2,\ldots,x_p\right]$ & $Y=\left[y_1,y_2,\ldots,y_q\right]$ - original single sensor feature set

Output : Ranking criterion function, $\left(\mathcal{P}_{100.0\rightarrow 44.4}\right.$ or $\left.\mathcal{P}_{100.0\rightarrow 83.8}\right)$; fused single feature ranking from e-tongue and e-nose of $\mathcal{F}_{D^2}=\left[(\mathcal{P}_{100.0\rightarrow 44.4})\in T_{D^2},(\mathcal{P}_{100.0\rightarrow 83.8})\in N_{D^2}\right]$

Step 1 : **Initialize** $T_{D^2}=[\ ]$ & $N_{D^2}=[\ ]$

Step 2 : **For** each feature $x_k\in X$ & $y_k\in Y$

i. Calculate the Mahalanobis distance for ${}^{g}C_2$ pair-wise groups using criterion (3.3) for each features $X=\left[x_1,x_2,\ldots,x_p\right]$ and $Y=\left[y_1,y_2,\ldots,y_q\right]$

$$\mathbf{D}_{ij}^2=\frac{\left(\bar{\mathbf{X}}_i\text{-}\bar{\mathbf{X}}_j\right)^2}{S_{ij}},\quad i=1,2,\ldots,g\ ;\ j=i+1,\ldots,g-1$$

$$\mathbf{D}_{ij}^2=\frac{\left(\bar{\mathbf{Y}}_i\text{-}\bar{\mathbf{Y}}_j\right)^2}{S_{ij}},\quad i=1,2,\ldots,g\ ;\ j=i+1,\ldots,g-1$$

ii. Calculate the average distance for ${}^{g}C_2$ pair-wise groups using criterion (3.5) for each features X and Y

$$\mathbf{D}_{ij}^2 = \frac{1}{{}^{g}C_2}\sum_{i=1}^{g}\sum_{j=i+1}^{g-1} D_{ij}^2$$

iii. Calculate the average distance $D(X)_l = \sum_{l=1}^{p} D_l^2$ and $D(Y)_l = \sum_{l=1}^{q} D_l^2$, for e-tongue and e-nose sensors, respectively.

iv. Store $D(X)_l$ and $D(Y)_l$ into $T_{D^2} = [\ ]$ & $N_{D^2} = [\ ]$

Step 3 : **End** of **for** loop

Step 4 : **Rank** feature $T_{D^2} = [\ ]$ & $N_{D^2} = [\ ]$ (from largest to smallest)

Step 5 : **Filter** $T_{D^2} = [\ ]$ & $N_{D^2} = [\ ]$ to keep only first $(\mathcal{P}_{100.0\rightarrow 44.4}) \in T_{D^2}$ and $(\mathcal{P}_{100.0\rightarrow 83.8}) \in N_{D^2}$, respectively

Step 6 : **Fused** $(\mathcal{P}_{100.0\rightarrow 44.4}) \in T_{D^2}$ and $(\mathcal{P}_{100.0\rightarrow 83.8}) \in N_{D^2}$ to produce $\mathcal{F}_{D^2} = \left[(\mathcal{P}_{100.0\rightarrow 44.4}) \in T_{D^2}, (\mathcal{P}_{100.0\rightarrow 83.8}) \in N_{D^2}\right]$

Algorithm 3.4

Single Feature Ranking for Bounded $[0,1]$ Mahalanobis Distance $[D_A^2]$ for Intermediate Level Data Fusion ~ Strategy 1

Input : $X = \left[x_1, x_2, \ldots, x_p\right]$ & $Y = \left[y_1, y_2, \ldots, y_q\right]$ - original single sensor feature set

Output : Ranking criterions function, $(\mathcal{P}_{100.0\rightarrow 44.4}$ or $\mathcal{P}_{100.0\rightarrow 83.8})$; fused single feature ranking from e-tongue and e-nose of $\mathcal{F}_{D_A^2} = \left[(\mathcal{P}_{100.0\rightarrow 44.4}) \in T_{D_A^2}, (\mathcal{P}_{100.0\rightarrow 83.8}) \in N_{D_A^2}\right]$

Step 1 : **Initialize** $T_{D^2} = [\ ]$ & $N_{D^2} = [\ ]$

Step 2 : **For** each feature $x_k \in X$ & $y_k \in Y$

i. Calculate the Mahalanobis distance for ${}^{g}C_2$ pair-wise groups using criterion (3.4) with equal *a priori*

probability for each features $X=\left[x_1,x_2,\ldots,x_p\right]$ and $Y=\left[y_1,y_2,\ldots,y_q\right]$

$$\mathbf{D}^2_{A(ij)}(X)=\frac{D^2_{ij}}{4+D^2_{ij}},\quad i=1,2,\ldots,g\,;\,j=i+1,\ldots,g-1$$

$$\mathbf{D}^2_{A(ij)}(Y)=\frac{D^2_{ij}}{4+D^2_{ij}},\quad i=1,2,\ldots,g\,;\,j=i+1,\ldots,g-1$$

ii. Calculate the average distance of ${}^gC_2$ pair-wise groups using criterion (3.6) for each features $X=\left[x_1,x_2,\ldots,x_p\right]$ and $Y=\left[y_1,y_2,\ldots,y_q\right]$

$$D^2{}_A(X)=\frac{1}{{}^gC_2}\sum_{i=1}^{g-1}\sum_{j=i+1}^{g}D^2_{Aij}$$

$$D^2{}_A(Y)=\frac{1}{{}^gC_2}\sum_{i=1}^{g-1}\sum_{j=i+1}^{g}D^2_{Aij}$$

iii. Calculate the average distance $D(X)_l=\sum_{l=1}^{p}D^2_{Al}$ and $D(Y)_l=\sum_{l=1}^{q}D^2_{Al}$, for e-tongue and e-nose sensors, respectively.

iv. Store $D(X)_l$ and $D(Y)_l$ into $T_{D^2_A}=[\ ]$ & $N_{D^2_A}=[\ ]$

Step 3 : **End** of **for** loop

Step 4 : **Rank** feature $T_{D^2_A}=[\ ]$ & $N_{D^2_A}=[\ ]$ (from largest to smallest)

Step 5 : **Filter** $T_{D^2_A}=[\ ]$ & $N_{D^2_A}=[\ ]$ and keep only first$\left(\mathcal{P}_{100.0\rightarrow 44.4}\right)\in T_{D^2_A}$ and $\left(\mathcal{P}_{100.0\rightarrow 83.8}\right)\in N_{D^2_A}$, respectively

Step 6 : **Fused** $\left(\mathcal{P}_{100.0\rightarrow 44.4}\right)\in T_{D^2_A}$ and $\left(\mathcal{P}_{100.0\rightarrow 83.8}\right)\in N_{D^2_A}$ to produce

$$\mathcal{F}_{D^2_A}=\left[(\mathcal{P}_{100.0\rightarrow 44.4})\ \in T_{D^2_A},(\mathcal{P}_{100.0\rightarrow 83.8})\in N_{D^2_A}\right]$$

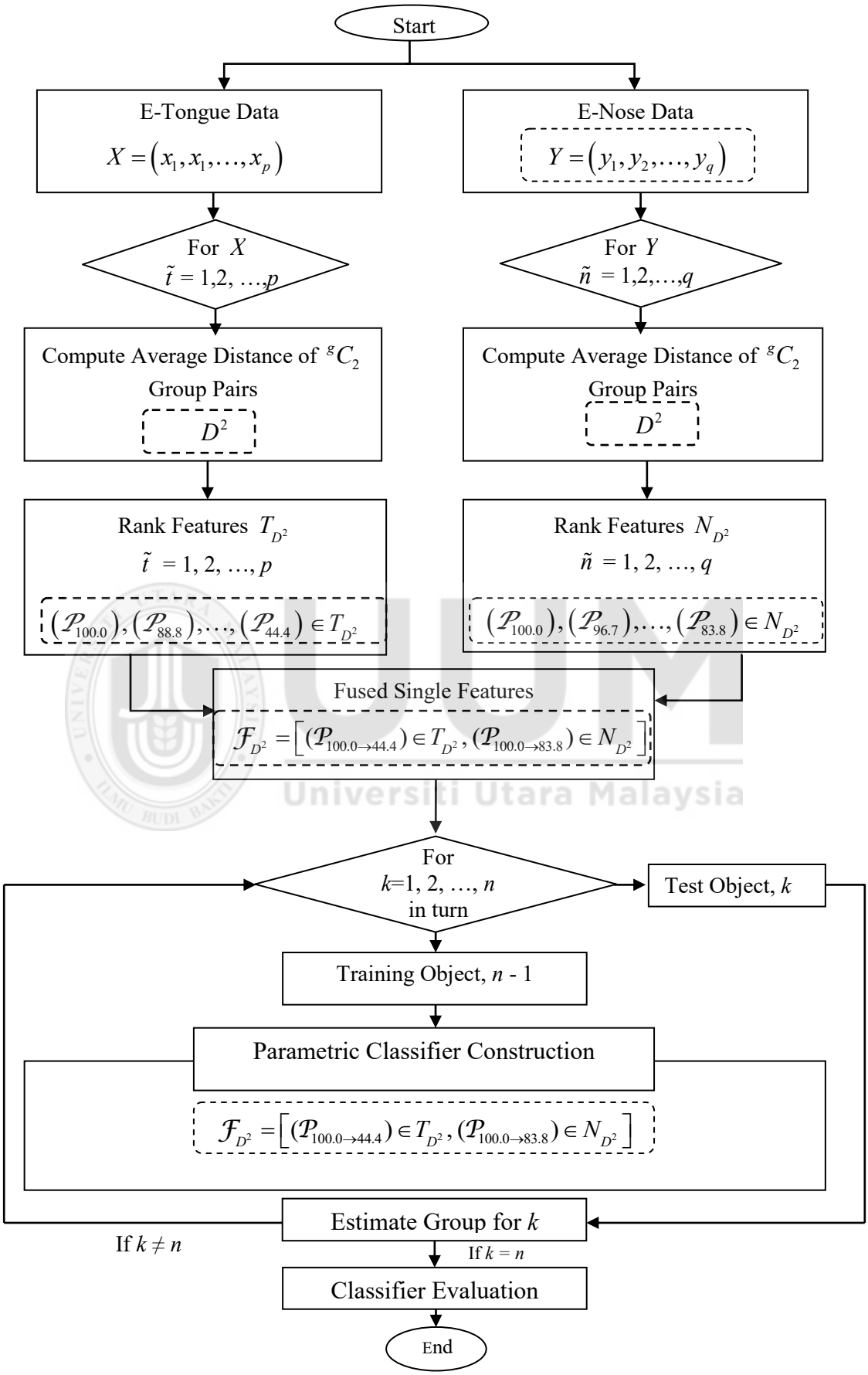

Figure 3.8. Flow Chart of Discriminant Analysis for the ILDF Model (Criterion D^2)

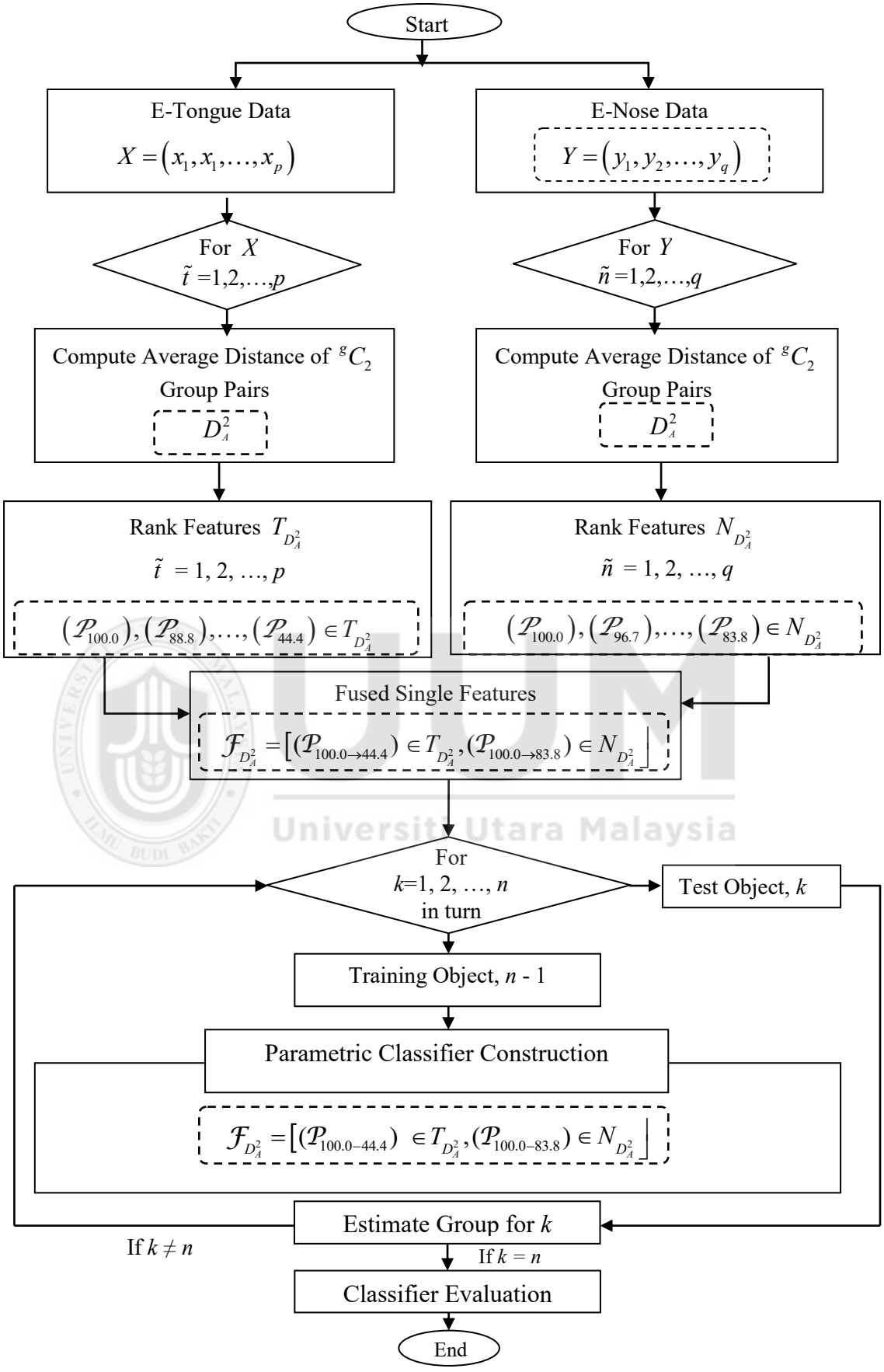

Figure 3.9. Flow Chart of Discriminant Analysis for the ILDF Model (Criterion D_A^2)

3.8 Applications to Real Data

The secondary data applied for this study was obtained from the Centre of Excellence for Advanced Sensor Technology (CEASTech), University Malaysia Perlis. For the purpose of consistency issue in comparing the performance of the feature selection algorithms and classification accuracies, ten datasets of different brands of pure Tualang honey were chosen. All the pure Tualang honey brands were purchased from the local market namely Agromas (AG), As-Syifa (AS), Syair Timur (ST), Tualang 3 (T3), Tayyibah (T), Tualang King (TK), Tualang TLH (TLH), Tualang Napis (TN), Wild Tualang (WT), and Yubalam Bahtera (YB). For each pure honey type, three bottles of three different batches of honey were purchased. And from each bottle, three samples of 5 ml honey were taken for experiments following the e-tongue and e-nose procedures, hence produced nine samples for each pure honey.

In addition, for each pure honey dataset, several adulterated honey samples which was mixed with different percentages (20%, 40%, 60% and 80%) of beetroot sugar (BS) and cane sugar (CS) were also prepared. The BS and CS were imported from Germany and United Kingdom, respectively. The adulterated samples went through similar e-tongue and e-nose procedures as the experiments of pure honey. Briefly, as a result of the experiments using the e-tongue and e-nose sensors, each dataset contains 27 observations of pure honey type, nine observations of pure beetroot, nine observations of pure cane sugar, nine observations of four different adulterated honeys concentrations (i.e. 20%, 40%, 60% and 80%) with BS and nine observations of four different adulterated honeys concentrations (i.e. 20%, 40%, 60% and 80%) with CS. These made up each honey dataset with 117 observations from 11 groups.

Hence, overall, we have ten datasets of pure Tualang honey and its adulterated solutions. Table 3.3 illustrates an example of a dataset for Agromas (AG) and its adulterated solution based on the abovementioned mixture ratios which made the composition of 11 groups. The rest of the pure Tualang honey brand datasets were organized similar to AG as in Table 3.3. The procedure of collecting data using e-tongue and e-nose devices were performed independently. The dimensions for e-tongue and e-nose have been described in detail in sections 3.2 and Table 3.1.

Table 3.3

Description of AG Tualang Honey Dataset with Adulterated Concentrations

Label (Pure Honey + Sugar Solution)	**Grouping (Ratio for Pure Honey : Sugar Solution)**	**No. of Observation**
AG	100% Pure honey	27
BS	100% BS	9
CS	100% CS	9
AGBS20	20% BS + 80% Pure honey	9
AGBS40	40% BS + 60% Pure honey	9
AGBS60	60% BS + 40% Pure honey	9
AGBS80	80% BS + 20% Pure honey	9
AGCS20	20% CS + 80% Pure honey	9
AGCS40	40% CS + 60% Pure honey	9
AGCS60	60% CS + 40% Pure honey	9
AGCS80	80% CS + 20% Pure honey	9
Total	11	117

3.8.1 Setup and Measurement for E-Tongue

The chalcogenide-based potentiometric e-tongue was made up of eight distinct ion-selective sensors from Sensor Systems (St. Petersburg, Russia). The potentiometric sensors used in this experiment were as follows.

Sensor Label	Description
Fe^{3+}	Ion-selective sensor for Iron ions
Cd^{2+}	Ion-selective sensor for Cadmium ions
Cu^{2+}	Ion-selective sensor for Copper ions
Hg^{2+}	Ion-selective sensor for Mercury ions
Ti^{+}	Ion-selective sensor for Titanium ions
S^{2-}	Ion-selective sensor for Sulfur ions
Cr^{4+}	Ion-selective sensor for Chromium ions
Ag^{+}	Ion-selective sensor for Silver ions

The e-tongue system was implemented by arranging an array of potentiometric sensors around the reference, pH and ORP probes. Each sensor output was connected to the analogue input of a data acquisition board (NI USB-6008) from National Instruments (Austin TX, USA). A 5% (w/v) solution of honey in distilled water was prepared and stirred for three minutes at 1,000 rpm before making any measurements. Each sample was replicated five times. For each measurement, the e-tongue was steeped simultaneously and left over for five minutes, and the potential readings were recorded for the whole duration. After each sampling, the e-tongue was dipped for one minute in 10% ethanol, stirred at 400 rpm and rinsed twice using distilled water (stirred at 400 rpm for two minutes) to remove any sticky residues from previous samples sticking on the sensor surface to avoid contaminating the next sample.

3.8.2 Setup and Measurement for E-Nose

The Cyranose320 (e-nose) from Smith Detection^{TM} which uses 32 non-selective sensors of different types of polymer matrix blended with carbon black was employed. The combination of these 32 sensors as an array allows the qualitative and capable of performing quantitative assessments of complex solutions. Each sample was drawn from a bottle using 10ml syringe and kept in a 13 x 100 mm test tube and seal with a silicone stopper. Each sample was replicated four 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 ± °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 30 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 on the honey samples.

	Cycle	Time (*s*)	Pump Speed
Sampling Setting	Baseline Purge	10	120 mL/min
	Sample Draw	30	120 mL/min
	Idle Time	3	-
	Air Intake Purge	80	160 mL/min

3.8.3 Data Pre-Processing

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 pre-process the data of both modalities. The maximum sensor response, S_t 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 $S_{frac} = [S_t - S_0]/S_0$. 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 any temperature, humidity and temporal drifts.

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 the four 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 five of the replicated samples.

Each S_{frac} data point from each e-nose and e-tongue sensor formed the S_{frac} matrix. This S_{frac} matrix was processed separately and scaled using z-score ($S_{frac,}1$) to zero mean and one standard deviation (taken from MATLAB statistical toolbox). This is to ensure that all sensor responses were commensurate and no particular sensor dominates the results.

3.8.4 Initial Multivariate Data Analysis

Some preliminary data analyses were performed on all the collected datasets before further discriminant analyses were done. The analyses were purposesly implemented to gain some information about the structure of the groups, and to determine whether the samples fulfill the requirements of the selected classification rule. It includes the multivariate normality test, homogeneity of the variance covariance matrices, and the multicolinearity among features. Since all the features are in continuous form, the testing processes were implemented straightforwardly. The only challenge emerged was the small sample size for each multi-group factors compared to the number of features which creates high dimensional data $(n < p)$ for both sensors.

It is important to know whether the investigated datasets is following multivariate normal for the optimal parametric rule is applied correctly. Otherwise, nonparametric rule may be more appropriate. Therefore, it will be worthwhile to perform some investigations on the selected datasets. For the multivariate normality assumption, data were tested univariately and multivariately. The univariate test for each group was performed separately using the Shapiro Wilk test procedure which is suitable for a sample less than 50 objects. The tested j continuous variables $x_{ij}\ (j = 1,\ldots, p+q; i = 1,\ldots, g)$ is said to have a normal distribution if the obtained p-value is greater than a predefined type I error, α *i.e.* $\alpha = 0.05$. The tested hypotheses are as follows:

H_0 : *The data is from a normally distributed population*
H_1 : *The data is not from a normally distributed population*

The test was carried out for all the available datasets with the following **R** routine:

```
variable <- colnames(DATA)[c(2:11)]
res <- vector("list", length(variable))
for (i in 1:length(variable))
{
res[[i]]   <-   lapply(split(DATA[,variable[i]]   ,DATA$Group),
shapiro.test)
}
```

The multivariate normality test based on the kurtosis and skewness, as well as the Q-Q plot can be performed using the well known Mardia multivariate normality test. Unfortunately, Mardia procedure failed because the requirement for the inversion of sample covariance matrices was not fulfilled. A problem with high dimensional data which was the case for both e-tongue and e-nose data was the nonsingular estimate for the covariance matrix. Since all the available multivariates normality tests required the inversion covariance matrices, a method of nonparametric inference for multivariate data was applied.

The nonparametric model underlying the **R** package **npmv** introduced by Ellis, Burchett, Harrar, & Bathke, (2017) simply stated that the multivariate observations vectors $\mathbf{X}_{ij}$ are independent and within the same groups, they follow the same *p*-variate distribution. Typical global statistical hypotheses were "Are the *g* samples from the same population (multivariate distribution)?"

```
# Load required packages 'npv'
library(npmv)
nonpartest(var1|var2|var3|…|varp~Group,DATA,permreps=1000)
```

Another test of homogeneity of the variance covariance matrices was also conducted to check whether the groups fulfilled the requirement of having equal covariance matrix Σ. The hypotheses to be tested are $H_0: \Sigma_1 = \ldots = \Sigma_g = \Sigma$ against $H_1: \Sigma_i \neq \Sigma_j$ *for at least one* $i \neq j$. The common test that can be used to test for

equality of covariance matrices is the Box"s M-test. Unfortunately, the test approximation works well only if each sample size exceeds 20 and if p and g do not exceed 5. Since the datasets contain unequal group sizes with $n = 9$ and $n = 27$, with number of groups of eleven, the test were carried out using the nonparametric approaches. The test called the Fligner-Killeen test was performed based on the median of features for the null hypothesis is that the variances in each of the groups (samples) are the same. The **R** routine is available in stats package as follows:

```
# Load required packages 'stats'
library(stats)
# Specify data
testdata <- as.data.frame([DATA,3:13])
# Load required packages 'car' and 'coin'
library(coin)
test.var1 <- fligner.test(var1 ~ Group, data = testdata)
:
:
Test.varp <- fligner.test(varp ~ Group, data = testdata)
```

Finally, multiple correlation *t*-tests from a correlation matrix were also performed using **R** package „biotools". The following results (for one sample data) were among the outputs and plots for all the tests performed on all ten datasets in hand.

3.9 Conclusion

This chapter describes the methodology to perform percentile forward feature selection for the LLDF and ILDF model. Four main algorithms were proposed for the models. The unbounded and bounded Mahalanobis distances to be applied have been discussed in details in terms of its computation to get the average pairwise distances for the criteria. Distinctions between the univariate Mahalanobis distance and multivariate Mahalanobis distance were also discussed. The proposed algorithms to implement the fused feature ranking and the single feature ranking for the

unbounded $\left(D^2\right)$ and bounded $\left(D_A^2\right)$ distances, respectively, for the LLDF and ILDF combining the proposed two strategies were also made. The overall processes of the discriminant analysis using the four algorithms are available in Figures 3.6, 3.7, 3.8 and 3.9. And finally, the ten datasets of the adulterated honey that were tested using each of the algorithms are described. The next chapter illustrates the findings of all the study objectives using the proposed algorithms and discriminant analyses.

CHAPTER FOUR
RESULT AND DISCUSSION

4.1 Introduction

This chapter presents the results of some investigations on the proposed strategies of percentiles forward feature selection for discriminant analysis of LLDF and ILDF fusion models. The search for discriminative features for both fusion models were justified using the bounded $[0,1]$ Mahalanobis distance $\left(\Delta_A^2\right)$ that is believed can overcome some weaknesses of the conventional method of unbounded $[0,+\infty)$ Mahalanobis distance $\left(\Delta^2\right)$, or it performes at least as good as the former one. To verify this claim, the explorations were performed following the seven research objectives which include:

i. to develop *univariate feature selection algorithms* using the *unbounded Mahalanobis distance* $\left(\Delta^2\right)$ and *bounded Mahalanobis distance* $\left(\Delta_A^2\right)$ for the LLDF and ILDF models,

ii. to develop *multivariate feature selection algorithms* using the *unbounded Mahalanobis distance* $\left(\Delta^2\right)$ and *bounded Mahalanobis distance* $\left(\Delta_A^2\right)$ for the LLDF and ILDF models,

iii. to construct the *parametric classification rules* based on the *percentile forward feature selection* for each of the developed algorithms in objective (i) and objective (ii), and

iv. to *evaluate the performance* of the constructed parametric classification rules.

In order to achieve these objectives, investigations were conducted using ten different datasets of honey that were described earlier in section 3.8. Even though the setting of each of the dataset was almost the same, the repetitive process of the proposed feature selection was hoped to rationalize the potential of the bounded Mahalanobis distance over the unbounded version especially in selecting the discriminative features that lead to better classification performance. Findings are elaborated based on the LLDF model (sections 4.2 and 4.3) followed by the ILDF model (sections 4.4 and 4.5).

4.2 Results for Low Level Data Fusion

The implementation of PFFS for the LLDF model was performed using Algorithm 3.1 for the unbounded $[0, \infty)$ Mahalanobis distance estimated by D^2, and Algorithm 3.2 for the bounded $[0,1]$ Mahalanobis distance estimated by D_A^2. The results of these procedures were the ten lists of ranked e-nose and e-tongue features for all the datasets. A sample of such results for AG honey type can be observed from Table 4.1. The table clearly specifies the feature ranking from both sensors. The unbounded criterion $\left(D^2\right)$ ranked feature from e-nose sensor (N23) with the distance score 11,949.06 as the most discriminative feature and ranked feature from e-tongue sensor (T10) with the distance score 41.37 as the least discriminative feature. However, the bounded criterion $\left(D_A^2\right)$ ranked feature N20 with the distance score 0.95 as the most discriminative feature and ranked feature T7 with the distance score 0.67 as the least discriminative feature.

Table 4.1

Results of Fused Feature Ranking for LLDF based on Bounded and Unbounded Mahalanobis Distance for AG Honey

D^2 **Criterion**				D_A^2 **Criterion**			
Feature	*Criterion Value*	*Rank*	*Percentile*	*Feature*	*Criterion Value*	*Rank*	*Percentile*
N23	11,949.06	1	100.00%	N20	0.947622	1	100.00%
N5	11,343.20	2	97.50%	N29	0.945641	2	97.50%
N29	10,027.52	3	95.10%	N5	0.939915	3	95.10%
N31	4,040.16	4	92.60%	N26	0.93304	4	92.60%
N9	3,680.18	5	90.20%	N10	0.932675	5	90.20%
N26	3,482.48	6	87.80%	N23	0.932147	6	87.80%
N11	3,074.39	7	85.30%	N28	0.926292	7	85.30%
N6	2,418.34	8	82.90%	N11	0.913003	8	82.90%
N20	2,300.65	9	80.40%	N22	0.909697	9	80.40%
N10	1,793.69	10	78.00%	N17	0.909596	10	78.00%
N17	1,778.24	11	75.60%	N31	0.903392	11	75.60%
N15	1,407.55	12	73.10%	N19	0.900304	12	73.10%
N16	1,161.90	13	70.70%	N9	0.897842	13	70.70%
N28	1,129.29	14	68.20%	N8	0.895816	14	68.20%
N22	1,124.58	15	65.80%	N18	0.89035	15	65.80%
N8	1,094.92	16	63.40%	N15	0.882092	16	63.40%
N18	1,074.86	17	60.90%	N30	0.878124	17	60.90%
N13	869.54	18	58.50%	N13	0.877008	18	58.50%
N12	770.64	19	56.00%	N12	0.875399	19	56.00%
T11	762.64	20	53.60%	N7	0.87438	20	53.60%
N30	741.57	21	51.20%	N6	0.871188	21	51.20%
N4	669.04	22	48.70%	N21	0.870555	22	48.70%
N7	646.60	23	46.30%	N16	0.869101	23	46.30%
N14	574.10	24	43.90%	T11	0.862588	24	43.90%
N21	546.70	25	41.40%	N27	0.861947	25	41.40%
N19	536.86	26	39.00%	N2	0.857348	26	39.00%
N2	532.30	27	36.50%	N24	0.855591	27	36.50%
N1	500.41	28	34.10%	N1	0.853093	28	34.10%
N27	459.53	29	31.70%	N3	0.851567	29	31.70%
N3	446.28	30	29.20%	N14	0.849068	30	29.20%

Table 4.1 Continued

N24	318.36	31	26.80%	T9	0.848139	31	26.80%
N25	317.15	32	24.30%	N4	0.845732	32	24.30%
T2	236.18	33	21.90%	T2	0.835134	33	21.90%
T3	193.98	34	19.50%	N25	0.81475	34	19.50%
T7	180.85	35	17.00%	T1	0.814565	35	17.00%
T1	133.79	36	14.60%	T5	0.807811	36	14.60%
N32	119.50	37	12.10%	N32	0.781032	37	12.10%
T5	96.45	38	9.70%	T8	0.779217	38	9.70%
T9	69.64	39	7.30%	T4	0.729904	39	7.30%
T8	55.04	40	4.80%	T3	0.678862	40	4.80%
T4	42.19	41	2.40%	T10	0.675177	41	2.40%
T10	41.37	42	0.00%	T7	0.66831	42	0.00%

From this LLDF ranking list, the first $\left(\mathscr{P}_{100.0\rightarrow 68.2}\right)$ percentiles features from both criteria were selected for the next multivariate distance calculation and the classification procedure (performed pair by pair). This selection is specified in the dashed-box in Table 4.1. The selected discriminative features for the unbounded criterion $\left(D^2\right)$ were N23, N5, N29, N31, N9, N26, N11, N6, N20, N10, N17, N15, N16, and N28. For the bounded criterion $\left(D_A^2\right)$, N20, N29, N5, N26, N10, N23, N28, N11, N22, N17, N31, N19, N9, and N8 were selected. Repeat the process for the rest of the datasets. For the selections of AG honey, none of the feature from e-tongue was selected due to smaller average distance score.

There were nine other similar tables as Table 4.1 to represent the results of feature ranking for the rest of honey types. However, the tables are available in Appendix B. Feature subsets appeared in the next Tables 4.2 to 4.11 were all the highest average

distance scores of the 68.2 percentiles $\left(\mathcal{P}_{100.0\rightarrow 68.2}\right)$ selected features resulted from the first procedure of fused feature ranking.

Next, the percentiles forwards feature selection were performed towards the selected features based on the unbounded criterion $\left(D^2\right)$ using subsets of [N23, N5, N29, N31, N9, N26, N11, N6, N20, N10, N17, N15, N16, and N28]; and bounded criterion $\left(D_A^2\right)$ using subsets [N20, N29, N5, N26, N10, N23, N28, N11, N22, N17, N31, N19, N9, and N8]. The percentiles forwards feature selection procedures for both criteria were performed for every cumulative 2.4 or 2.5 percentiles until the accumulated $\left(\mathcal{P}_{100.0\rightarrow 68.2}\right)$ percentiles were achieved.

In this second strategy, the processes were implemented as described in Figure 3.6 for further discriminant analysis based on the unbounded criterion $\left(D^2\right)$ and Figure 3.7 for the bounded criterion $\left(D_A^2\right)$. The investigations were implemented in turn. For example, the first subset [N23, N5] =$\left(\mathcal{P}_{100.0\rightarrow 97.5}\right)$ percentile, the next subset [N23, N5, N29, N31] = $\left(\mathcal{P}_{100.0\rightarrow 92.6}\right)$, followed by the rest of the feature subsets, until its finally done for $\left(\mathcal{P}_{100.0\rightarrow 68.2}\right)$ feature subset. Table 4.2 is the results of discriminant analysis for both criteria $\left(D^2\right)$ and $\left(D_A^2\right)$. The table illustrates the performance of parametric classification (LDA rule) with the leave-one-out error estimation as well as the multivariate unbounded $\left(D^2\right)$ and bounded $\left(D_A^2\right)$ distance according to every feature subset.

Table 4.2

Classification Performances for Subset of Ranked Fused Features and the Multivariate Mahalanobis Distance for AG Honey (LLDF)

Ranked Features	**Criterion D^2**		**Ranked Features**	**Criterion D_A^2**	
	Accuracy (100%)	**Distance**		**Accuracy (100%)**	**Distance**
N23, N5	57.26	6.4697×10^{-4}	N20, N29	66.67	6.7906×10^{-9}
N23, N5, N29, N31	73.50	6.8589×10^{-4}	N20, N29, N5, N26	93.16	5.5525×10^{-7}
N23, N5, N29, N31, N9, N26	92.31	6.9310×10^{-4}	**N20, N29, N5, N26, N10, N23**	**97.44**	1.2297×10^{-5}
N23, N5, N29, N31, N9, N26, N11, N6	94.02	7.7855×10^{-4}	N20, N29, N5, N26, N10, N23, N28, N11	96.58	1.2328×10^{-5}
N23, N5, N29, N31, N9, N26, N11, N6, N20, N10	96.58	7.8016×10^{-4}	N20, N29, N5, N26, N10, N23, N28, N11, N22, N17	96.58	1.2332×10^{-5}
N23, N5, N29, N31, N9, N26, N11, N6, N20, N10, N17, N15	**97.44**	7.8063×10^{-4}	N20, N29, N5, N26, N10, N23, N28, N11, N22, N17, N31, N19	97.44	1.2637×10^{-5}
N23, N5, N29, N31, N9, N26, N11, N6, N20, N10, N17, N15, N16, N28	97.44	7.8172×10^{-4}	N20, N29, N5, N26, N10, N23, N28, N11, N22, N17, N31, N19, N9, N8	94.87	1.2668×10^{-5}

The classification performance was given in percentages. The multivariate distances were also calculated and recorded for each of features subset. Tables 4.3 to 4.11 show the overall results for the classification performance and the multivariate Mahalanobis distance for honey types AG, AS, ST, T, T3, TK, TLH, TN, WT and

YB, respectively. Generally, the classification performances of the unbounded $\left(D^2\right)$ and bounded $\left(D_A^2\right)$ distances were varied for each honey dataset.

Table 4.2 shows the results of the unbounded distance criteria which begins with selection of feature subset using features N23 and N5 as the influential features, while the bounded distance criteria preferred features N20 and N29. However, the classification performance seemed to be better for the later criterion which begun with 66.67% of correct classification and converge to the highest correct classification in the third subset selection.

Table 4.3 portrays that discriminative features selected by the unbounded criterion were better than the first feature subset of the bounded criterion. The result was supported by the higher classification accuracy of the unbounded criterion. The highest correct classifications were only recorded by the classification of the sixth feature subsets from both criteria.

Table 4.3

Classification Performances for Subset of Ranked Fused Features and the Multivariate Mahalanobis Distance for AS Honey (LLDF)

Ranked Features	**Criterion D^2**		**Ranked Features**	**Criterion D_A^2**	
	Accuracy (100%)	**Distance**		**Accuracy (100%)**	**Distance**
N26, N5	72.65	$2.3511\text{x}10^{-5}$	N15, N5	68.38	$4.0388\text{x}10^{-7}$
N26, N5, N29, N31	71.79	$2.6341\text{x}10^{-5}$	N15, N5, N23, N11	74.36	$3.5921\text{x}10^{-6}$
N26, N5, N29, N31, N15, N9	79.49	$2.6851\text{x}10^{-5}$	N15, N5, N23, N11, N29, N26	85.47	$3.7580\text{x}10^{-6}$
N26, N5, N29, N31, N15, N9, N20, N16	82.91	$2.7296\text{x}10^{-5}$	N15, N5, N23, N11, N29, N26, N8, N18	85.47	$3.7580\text{x}10^{-6}$

Table 4.3 Continued

N26, N5, N29, N31, N15, N9, N20, N16, N23, T11	91.45	4.5724×10^{-4}	N15, N5, N23, N11, N29, N26, N8, N18, N20, N2	86.32	3.7709×10^{-6}
N26, N5, N29, N31, N15, N9, N20, N16, N23, T11, N17, N13	**94.02**	4.5732×10^{-4}	**N15, N5, N23, N11, N29, N26, N8, N18, N20, N2, N4, T11**	**93.16**	8.2308×10^{-6}
N26, N5, N29, N31, N15, N9, N20, N16, N23, T11, N17, N13, N8, N21	92.31	4.5737×10^{-5}	N15, N5, N23, N11, N29, N26, N8, N18, N20, N2, N4, T11, N1, N3	90.60	8.2333×10^{-6}

Table 4.4 depicts the results of bounded distance criteria begun the selection feature subset using features N6 and T2 as the influential features with higher accuracy compared to the unbounded criterion with lower correct classification rate. The accuracy rate for the unbounded criterion started to converge at the sixth feature subset selection. The unbounded criterion seemed less effective compared to the other criterion.

Table 4.4

Classification Performances for Subset of Ranked Fused Features and the Multivariate Mahalanobis Distance for ST Honey (LLDF)

Ranked Features	**Criterion D^2**		**Ranked Features**	**Criterion D_A^2**	
	Accuracy (100%)	**Distance**		**Accuracy (100%)**	**Distance**
N29, N5	63.25	2.3524×10^{-5}	N6, T2	74.36	2.3656×10^{-7}
N29, N5, N23, N31	76.92	2.0990×10^{-4}	N6, T2, N31, N29	71.79	2.6280×10^{-7}
N29, N5, N23, N31, N26, N9	87.18	2.1263×10^{-4}	N6, T2, N31, N29, N26, N5	80.34	7.3720×10^{-7}

Table 4.4 Continued

N29, N5, N23, N31, N26, N9, T2, N11	92.30	2.2483×10^{-4}	N6, T2, N31, N29, N26, N5, N17, T11	88.89	5.5281×10^{-6}
N29, N5, N23, N31, N26, N9, T2, N11, N6, N20	93.16	2.4703×10^{-4}	N6, T2, N31, N29, N26, N5, N17, T11, N23, *, N20 *(omit N22)	93.16	8.6750×10^{-6}
N29, N5, N23, N31, N26, N9, T2, N11, N6, N20, N17, N28	**94.02**	2.4743×10^{-4}	**N6, T2, N31, N29, N26, N5, N17, T11, N23, N20, T1, N18**	**96.58**	8.6756×10^{-6}
N29, N5, N23, N31, N26, N9, T2, N11, N6, N20, N17, N28, *, N1, N8 *(omit N10)	93.16	2.4751×10^{-4}	N6, T2, N31, N29, N26, N5, N17, T11, N23, N20, T1, N18, N9, N16	96.58	8.6866×10^{-6}

Similar patterns as achieved by ST honey type in Table 4.4 were observed from Table 4.5 for T honey. Except that the convergence of this dataset was achieved at the last feature subset. And the selection of features as well as highest classification accuracy was less effective for the unbounded criterion.

Table 4.5

Classification Performances for Subset of Ranked Fused Features and the Multivariate Mahalanobis Distance for T Honey (LLDF)

Ranked Features	**Criterion D^2**		**Ranked Features**	**Criterion D_A^2**	
	Accuracy (100%)	**Distance**		**Accuracy (100%)**	**Distance**
T7, T2	68.38	1.3081×10^{-4}	N23, N6,	74.36	1.2652×10^{-6}
T7, T2, N29, N23	83.76	1.9277×10^{-4}	N23, N6, T11, N8	83.76	1.4984×10^{-6}
T7, T2, N29, N23, N31, N5	86.32	2.1021×10^{-4}	N23, N6, T11, N8, N22, N10	92.31	1.4986×10^{-6}

Table 4.5 Continued

T7, T2, N29, N23, N31, N5, N6, N26	86.32	2.2342×10^{-4}	N23, N6, T11, N8, N22, N10, N5, N18	92.31	1.8625×10^{-6}
T7, T2, N29, N23, N31, N5, N6, N26, N9, N20	88.89	2.2358×10^{-4}	N23, N6, T11, N8, N22, N10, N5, N18, N9, N26	92.31	1.8704×10^{-6}
T7, T2, N29, N23, N31, N5, N6, N26, N9, N20, N10, N17	88.89	2.2359×10^{-4}	N23, N6, T11, N8, N22, N10, N5, N18, N9, N26, N20, N17	92.31	1.8725×10^{-6}
T7, T2, N29, N23, N31, N5, N6, N26, N9, N20, N10, N17, T11, N22	**93.16**	2.1074×10^{-4}	**N23, N6, T11, N8, N22, N10, N5, N18, N9, N26, N20, N17, N19, T2**	**94.02**	3.6080×10^{-6}

Again, similar patterns in the results of Table 4.6 were observed following the results of Table 4.4 and 4.5. Except that the convergence of this dataset was achieved at the last feature subset. And the selection of features as well as highest classification accuracy was less effective for the unbounded criterion.

Table 4.6

Classification Performances for Subset of Ranked Fused Features and the Multivariate Mahalanobis Distance for T3 Honey (LLDF)

Ranked Features	**Criterion D^2**		**Ranked Features**	**Criterion D_A^2**	
	Accuracy (100%)	**Distance**		**Accuracy (100%)**	**Distance**
N29, N5	76.92	1.6931×10^{-6}	N6, T2	84.62	2.4560×10^{-6}
N29, N5, N23, N31	82.91	7.7784×10^{-5}	N6, T2, N10, N31	86.32	2.4596×10^{-6}
N29, N5, N23, N31, N26, N9	88.89	7.8137×10^{-5}	N6, T2, N10, N31, N29, N26	90.60	2.4667×10^{-6}

Table 4.6 Continued

N29, N5, N23, N31, N26, N9, N10, N17	90.60	7.8147×10^{-5}	N6, T2, N10, N31, N29, N26, N5, N17	92.31	2.5392×10^{-6}
N29, N5, N23, N31, N26, N9, N10, N17, N20, N6	92.31	9.1028×10^{-5}	**N6, T2, N10, N31, N29, N26, N5, N17, T11, N23**	**95.73**	4.3066×10^{-6}
N29, N5, N23, N31, N26, N9, N10, N17, N20, N6, N8, T2	**94.87**	2.3705×10^{-4}	N6, T2, N10, N31, N29, N26, N5, N17, T11, N23, N22, N20	95.73	4.3082×10^{-6}
N29, N5, N23, N31, N26, N9, N10, N17, N20, N6, N8, T2, N22, N15	94.87	2.3707×10^{-4}	**N6, T2, N10, N31, N29, N26, N5, N17, T11, N23, N22, N20, T1, N18**	**96.58**	4.3082×10^{-6}

Table 4.7 shows the results of features selection of the bounded distance criteria, which confirmed the discriminative power of its selection is better than the unbounded criterion. In addition, the bounded criterion scored higher classification accuracy compared to the unbounded criterion.

Table 4.7

Classification Performances for Subset of Ranked Fused Features and the Multivariate Mahalanobis Distance for TK Honey (LLDF)

Ranked Features	**Criterion D^2**		**Ranked Features**	**Criterion D_A^2**	
	Accuracy (100%)	**Distance**		**Accuracy (100%)**	**Distance**
N23, N29	72.65	6.0223×10^{-5}	N5, N23	82.05	1.4067×10^{-6}
N23, N29, N5, N9	89.74	7.8790×10^{-5}	N5, N23, T11, N6	88.89	1.9562×10^{-6}
N23, N29, N5, N9, N26, N20	89.74	7.9155×10^{-5}	**N5, N23, T11, N6, N10, N11**	**92.31**	1.9562×10^{-6}

Table 4.7 Continued

N23, N29, N5, N9, N26, N20, N17, N31	89.74	8.0114×10^{-5}	**N5, N23, T11, N6, N10, N11, N17, N29**	**94.87**	1.9841×10^{-6}
N23, N29, N5, N9, N26, N20, N17, N31, N16, N10	**92.31**	8.0126×10^{-5}	N5, N23, T11, N6, N10, N11, N17, N29, N22, N18	94.02	1.9842×10^{-6}
N23, N29, N5, N9, N26, N20, N17, N31, N16, N10, N15, N13	92.31	8.0146×10^{-5}	N5, N23, T11, N6, N10, N11, N17, N29, N22, N18, T2, N15	93.16	2.9793×10^{-6}
N23, N29, N5, N9, N26, N20, N17, N31, N16, N10, N15, N13, N18, N6	88.89	9.2349×10^{-5}	N5, N23, T11, N6, N10, N11, N17, N29, N22, N18, T2, N15, N16, N19	92.31	2.9792×10^{-6}

The results of dataset in Table 4.8 were better performed by the unbounded criterion in terms of the discriminative power of its feature subset. However, in terms of the convergence of highest classification was achieved by the bounded criterion at the fifth feature subsets.

Table 4.8

Classification Performances for Subset of Ranked Fused Features and the Multivariate Mahalanobis Distance for TLH Honey (LLDF)

Ranked Features	**Criterion D^2**		**Ranked Features**	**Criterion D_A^2**	
	Accuracy (100%)	**Distance**		**Accuracy (100%)**	**Distance**
N29, N9	83.76	3.3638×10^{-8}	N6, N20	72.65	1.1681×10^{-8}
N29, N9, N26, N17	91.45	4.9322×10^{-8}	N6, N20, N9, N18	90.60	1.1825×10^{-8}
N29, N9, N26, N17, N10, N5	92.31	3.4975×10^{-6}	N6, N20, N9, N18, N22, N8	94.87	1.1836×10^{-8}

Table 4.8 Continued

N29, N9, N26, N17, N10, N5, N20, N31	94.87	$3.8907x10^{-6}$	N6, N20, N9, N18, N22, N8, N26, T11	97.44	$3.4863x10^{-7}$
N29, N9, N26, N17, N10, N5, N20, N31, N18, N15	97.44	$3.9036x10^{-6}$	**N6, N20, N9, N18, N22, N8, N26, T11, N28, *, N17 (* omit N7)**	**98.29**	$3.4866x10^{-7}$
N29, N9, N26, N17, N10, N5, N20, N31, N18, N15, N8, N22	**98.29**	$3.9094x10^{-6}$	N6, N20, N9, N18, N22, N8, N26, T11, N28, N17, N15, *, N31 *(omit N21)	97.44	$3.4878x10^{-7}$
N29, N9, N26, N17, N10, N5, N20, N31, N18, N15, N8, N22, N16, N23	96.58	$9.1033x10^{-5}$	N6, N20, N9, N18, N22, N8, N26, T11, N28, N17, N15, N31, N5, N30	97.44	$4.4377x10^{-7}$

Obvious difference can be spotted from the results in Table 4.9 above, where the bounded criterion outperformed the unbounded criterion in the selection of the first feature subset as well as the convergence of highest correct classification rate.

Table 4.9

Classification Performances for Subset of Ranked Fused Features and the Multivariate Mahalanobis Distance for TN Honey (LLDF)

Ranked Features	**Criterion D^2**		**Ranked Features**	**Criterion D_A^2**	
	Accuracy (100%)	**Distance**		**Accuracy (100%)**	**Distance**
T3, N23	71.79	2.0409×10^{-4}	N22, T11	92.31	2.3486×10^{-7}
T3, N23, N29, N9	85.47	2.1010×10^{-4}	**N22, T11, N29, N10**	**94.02**	3.7112×10^{-7}
T3, N23, N29, N9, N26, N31	86.32	2.1500×10^{-4}	N22, T11, N29, N10, N26, N17	92.31	3.7693×10^{-7}
T3, N23, N29, N9, N26, N31, N20, N10	83.76	2.1530×10^{-4}	N22, T11, N29, N10, N26, N17, N18, N19	88.89	3.7695×10^{-7}
T3, N23, N29, N9, N26, N31, N20, N10, N6, N8	88.03	2.3586×10^{-4}	**N22, T11, N29, N10, N26, N17, N18, N19, N9, T2**	**95.73**	1.2905×10^{-6}
T3, N23, N29, N9, N26, N31, N20, N10, N6, N8, T11, N22	**94.87**	2.5562×10^{-4}	N22, T11, N29, N10, N26, N17, N18, N19, N9, T2, N28, N6	94.87	1.3624×10^{-6}
T3, N23, N29, N9, N26, N31, N20, N10, N6, N8, T11, N22, N17, N18	94.87	2.5565×10^{-4}	N22, T11, N29, N10, N26, N17, N18, N19, N9, T2, N28, N6, N20, N31	96.58	1.3879×10^{-6}

For Table 4.10 above, quite similar performance were recorded for both the unbounded and bounded criteria. However, convergence of highest correct classification rate is shown to favor the bounded criterion.

Table 4.10

Classification Performances for Subset of Ranked Fused Features and the Multivariate Mahalanobis Distance WT Honey (LLDF)

Ranked Features	**Criterion D^2**		**Ranked Features**	**Criterion D_A^2**	
	Accuracy (100%)	**Distance**		**Accuracy (100%)**	**Distance**
N5, N29	74.36	9.3746×10^{-6}	N23, N28	74.36	3.2974×10^{-6}
N5, N29, N23, N26	85.47	2.6217×10^{-4}	N23, N28, N20, N31	83.76	3.3699×10^{-6}
N5, N29, N23, N26, N9, N20	88.89	2.6302×10^{-4}	N23, N28, N20, N31, N6, N26	87.18	3.8612×10^{-6}
N5, N29, N23, N26, N9, N20, N31, N6	88.03	2.9777×10^{-4}	N23, N28, N20, N31, N6, N26, N5, N10	89.74	5.2536×10^{-6}
N5, N29, N23, N26, N9, N20, N31, N6, N11, N17	90.60	2.9800×10^{-4}	**N23, N28, N20, N31, N6, N26, N5, N10, N29, N9**	**91.45**	5.4209×10^{-6}
N5, N29, N23, N26, N9, N20, N31, N6, N11, N17, N28, N10	**91.45**	2.9837×10^{-4}	N23, N28, N20, N31, N6, N26, N5, N10, N29, N9, N22, N8	91.45	5.4215×10^{-6}
N5, N29, N23, N26, N9, 20, N31, N6, N11, N17, N28, N10, N15, N8	91.45	2.9846×10^{-4}	N23, N28, N20, N31, N6, N26, N5, N10, N29, N9, N22, N8, N15, N16	93.16	5.4230×10^{-6}

Finally, the dataset in Table 4.11 describes that the bounded criterion once again outperformed the unbounded criterion in terms of discriminative power of feature subsets. The convergence of highest classification accuracy shown equally achieved in the last feature subsets from both bounded and unbounded criteria.

Table 4.11

Classification Performances for Subset of Ranked Fused Features and the Multivariate Mahalanobis Distance for YB Honey (LLDF)

Ranked Features	**Criterion D^2**		**Ranked Features**	**Criterion D_A^2**	
	Accuracy (100%)	**Distance**		**Accuracy (100%)**	**Distance**
N5, N23	58.97	2.7559×10^{-4}	N11, N28	60.68	4.2519×10^{-11}
N5, N23, N29, N26	79.49	2.9294×10^{-4}	N11, N28, N9, N26	66.67	1.0872×10^{-9}
N5, N23, N29, N26, N6, N9	84.62	3.2889×10^{-4}	N11, N28, N9, N26, N20, N6	76.92	3.08535×10^{-8}
N5, N23, N29, N26, N6, N9, N31, N20	84.62	3.4173×10^{-4}	N11, N28, N9, N26, N20, N6, N23, N17	91.45	3.6059×10^{-6}
N5, N23, N29, N26, N6, N9, N31, N20, N11, N17	93.16	3.4217×10^{-4}	N11, N28, N9, N26, N20, N6, N23, N17, N5, T11	96.58	5.7087×10^{-6}
N5, N23, N29, N26, N6, N9, N31, N20, N11, N17, N28, N8	91.45	3.4307×10^{-4}	N11, N28, N9, N26, N20, N6, N23, N17, N5, T11, N25, N10	97.44	5.7107×10^{-6}
N5, N23, N29, N26, N6, N9, N31, N20, N11, N17, N28, N8, N10, T11	**98.29**	3.3965×10^{-4}	**N11, N28, N9, N26, N20, N6, N23, N17, N5, T11, N25, N10, N8, N18**	**98.29**	5.7131×10^{-6}

4.3 Discussion for Feature Selection in Low Level Data Fusion

The discussions of the LLDF results are described based on the predetermined conditions that include the univariate and multivariate distance for feature ranking and selected feature subset, selections discriminative power of the selected feature subsets, classification accuracy (first feature subset and highest accuracy), and

maximum convergence of correct classification before deterioration based on feature subset.

The choice of features to be selected is based on the porposed Percentile Forward Feature Selection which earlier described in Section 3.2. Generally, the findings of the studies for LLDF were consistence with the previous study by Ray and Turner (1992) from the perspective of the performance of bounded distance $\left(D_A^2\right)$. Based on the results in Tables 4.2 to 4.11 and Figures 4.1 to 4.3, the following discussions were made.

Univariate and multivariate distances for feature ranking and selected feature subset. In the first process of fused feature ranking, the univariate Mahalanobis distance based on the unbounded and bounded criteria has shown its ability to recognize discriminant features uniquely. Distances given by the unbounded criterion represent the average distance for the particular feature across 11 groups. Higher distances were recorded among e-nose features for both unbounded and bounded criteria. Dissimilar to the previous criteria, bounded distance gave even distinctive features using average distance among 11 groups between distance values of 0 to 1. The multivariate Mahalanobis distance for the unbounded and bounded criteria were very small due to small range of the original datasets. From the finding in Tables 4.2 to 4.11, we can see that the multivariate Mahalanobis distance increases very slowly for both criteria.

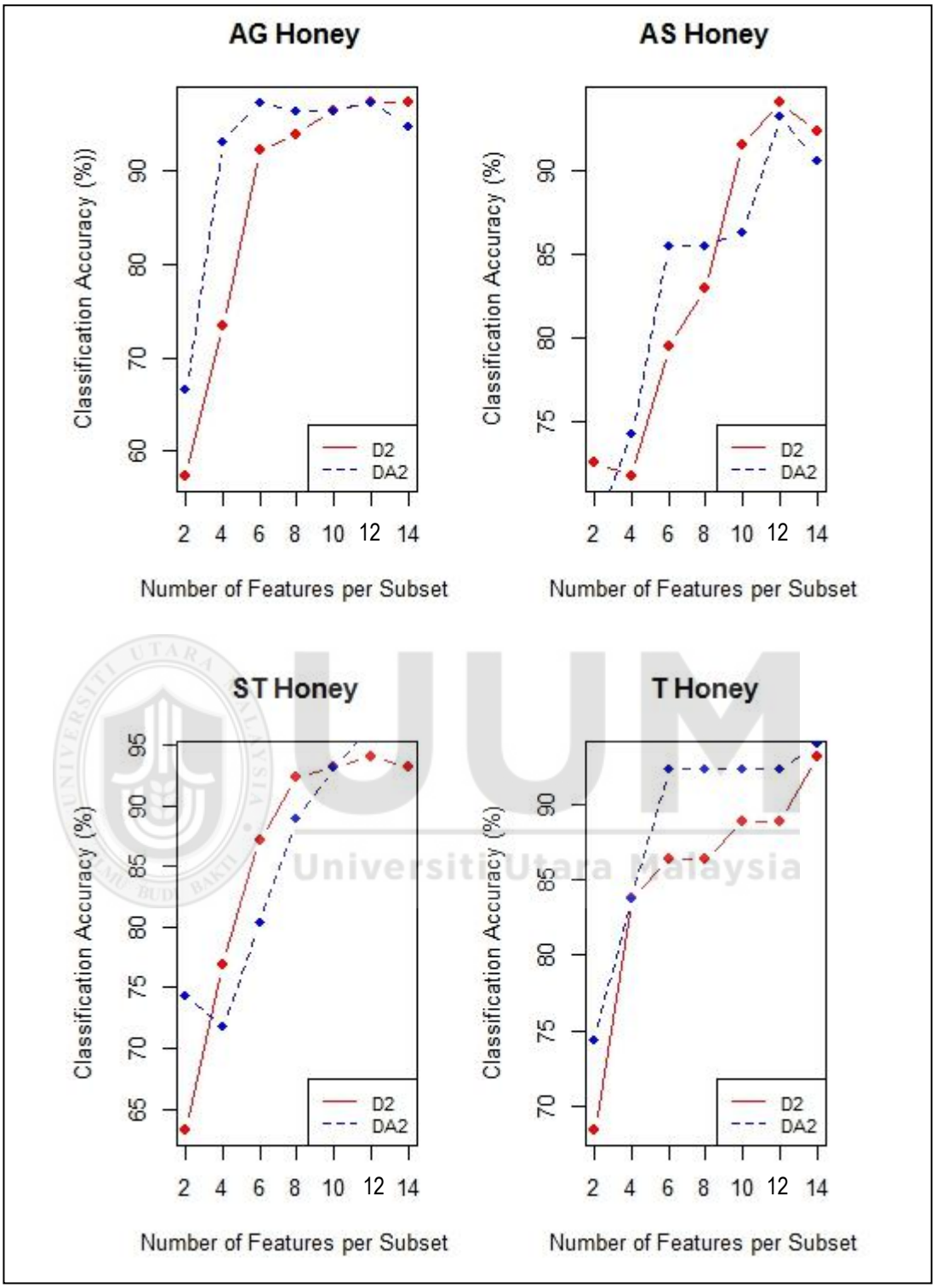

Figure 4.1. Comparison of Classification Accuracy based on D^2 and D_A^2 for Feature Subsets of AG, AS, ST and T Honey Types (LLDF)

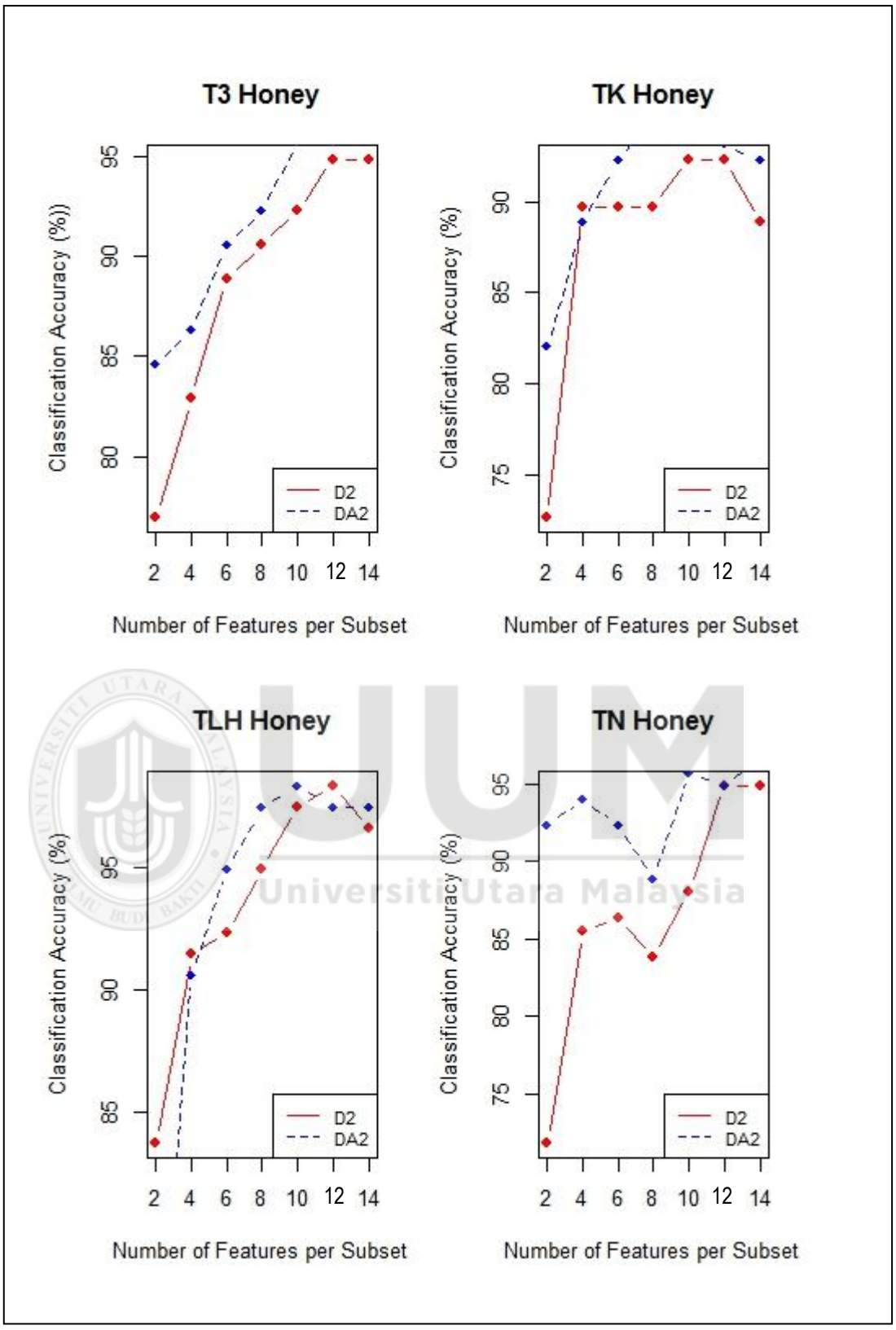

Figure 4.2. Comparison of Classification Accuracy based on D^2 and D_A^2 for Feature Subsets of T3, TK, TLH and TN Honey Types (LLDF)

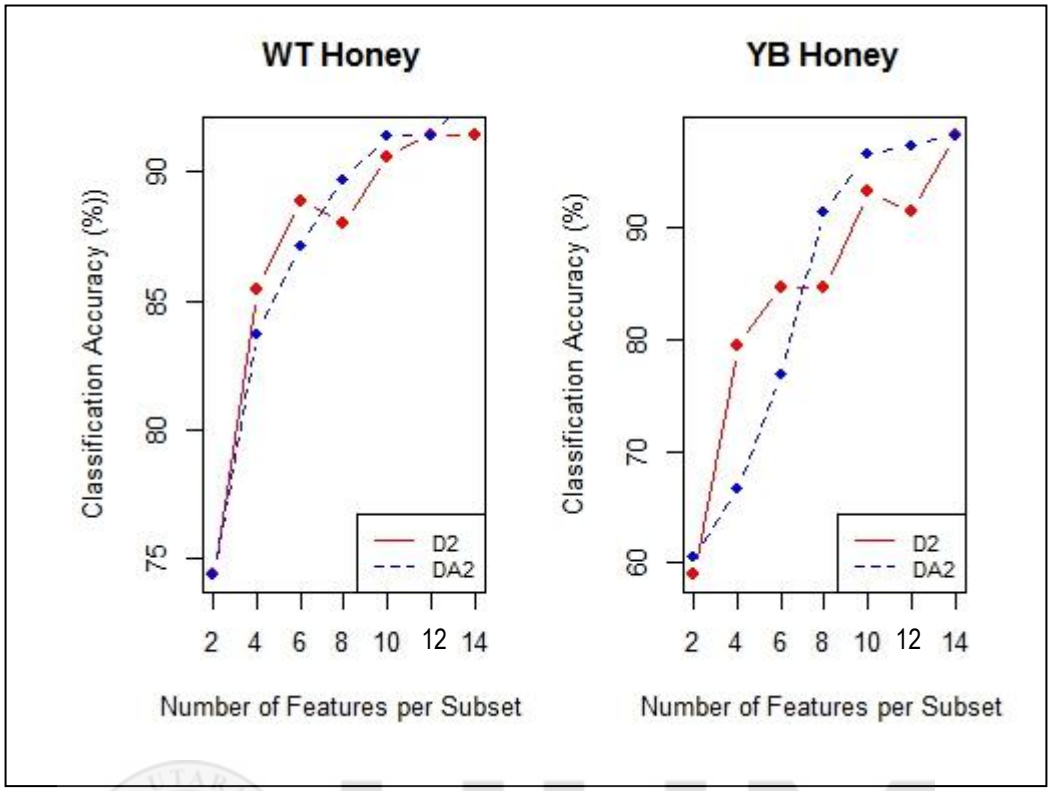

Figure 4.3. Comparison of Classification Accuracy based on D^2 and D_A^2 for WT and YB Honey Types (LLDF)

Discriminative power of the first selected feature subset. Commonly, features selected for both criteria $\left(D^2\right)$ and $\left(D_A^2\right)$ were almost the same, but the features were ranked differently. These can be observed from the summary of the selected features of AG and ST honey (below), the selected features for the unbounded criteria are highlighted in orange while the one for the bounded are highlighted in blue. The results indicate that both criteria were able to detect discriminative features, but of different discriminative power. From the results in Tables 4.2 to 4.11, obviously feature subsets selected by bounded criteria gave higher discriminative power.

Table 4.12

Illustration for the Comparison of Ranked Fused Features (LLDF Model) for AG and ST Honey Dataset.

LLDF (AG)	D^2	N5	N6	N8	N9	N 10	N 11	N 15	N 16	N 17	N 19	N 20	N 22	N 23	N 26	N 28	N 29	N 31
	Rank	2	8		5	10	7	12	13	11		9		1	6	28	3	4
	D_A^2	N5	N6	N8	N9	N 10	N 11	N 15	N 16	N 17	N 19	N 20	N 22	N 23	N 26	N 28	N 29	N 31
	Rank	3		14	13	5	8			10	12	1	9	6	4	7	2	11

LLDF (ST)	D^2	N1	N5	N6	N8	N9	N 11	N 16	N 17	N 18	N 20	N 23	N 26	N 28	N 29	N 31	T1	T2	T 11
	Rank	13	2	9	14	6	8		11		10	3	5	12	1	4		7	
	D_A^2	N1	N5	N6	N8	N9	N 11	N 16	N 17	N 18	N 20	N 23	N 26	N 28	N 29	N 31	T1	T2	T 11
	Rank		6	1		13		16	7	12	10	9	5		4	3	11	2	8

For example, Table 4.12 shows for AG honey, features N20 (ranked 1st) and N29 (ranked 2nd) selected using D_A^2 criterion gave 66.67% correct classification compared to features N23 (ranked 1st) and N5 (ranked 2nd) selected using D^2 criterion gave 57.26% correct classification. The same pattern occurred for ST honey where features N6 (ranked 1st) and T2 (ranked 2nd) selected using D_A^2 criterion gave 74.36% correct classification compared to features N29 (ranked 1st) and N5 (ranked 2nd) selected using D^2 criterion gave 63.25% correct classification. These can be observed clearly from Figures 4.1 and 4.2 below. Out of 10 datasets tested, the selected first pair of features by the bounded criterion recorded 7 times (AG, ST, T, T3, TK, TN and YB) of higher discriminative power, one (WT) of equal discriminative power, and two (AS and TLH) by the unbounded criterion.

Classification Accuracy (first feature subset and highest accuracy). Generally, the first selected feature subset based on the bounded criterion begun with higher classification accuracy compared to the unbounded criterion. For example, higher

correct classification 66.67% was recorded for the bounded criterion for AG honey, while the unbounded criterion gave only 57.26% correct classification. The same pattern occurred for ST honey where higher correct classification 74.36% was recorded for the bounded criterion, but the unbounded criterion gave only 63.25% correct classification. Further details referred to Figures 4.1 to 4.3 above. In terms of highest classification recorded for both criteria, out of 10 datasets tested, six highest classification accuracies were recorded by the bounded criterion (ST, T, T3, TK, TN, WT), two equal correct classification for both criteria (AG, TLH, YB), and one highest correct classification by the unbounded criterion (AS). However, varied highest accuracies were recorded at different feature subsets of each of the honey dataset of both criteria.

Maximum convergence of correct classification before deterioration based on feature subset. This refers to the highest classification accuracy achieved based on some particular subset of features from both criteria. Observation of this point can be found from Figures 4.1, 4.2 and 4.3 where the first highest peak occurred. In specific, the results of this point are recorded in the following Table 4.12. Generally, feature subsets generated by the bounded criterion converge to the highest classification accuracy faster within certain initial feature subset as highlighted in Table 4.12.

Table 4.13

Comparison of Performance for the Unbounded and Bounded Feature Selection based on Feature Subset Number and Correct Classification (ILDF)

DATASET	D^2		D_A^2	
	Subset Number	**Accuracy**	**Subset Number**	**Accuracy**
AG	6	97.43	4	96.58
AS	6	94.02	6	93.16
ST	6	94.02	6	96.58

Table 4.13 Continued

T	7	93.16	3	92.31
T3	6	94.87	5	95.73
TK	5	92.31	4	94.87
TLH	6	98.29	5	98.29
TN	6	94.87	2	94.02
WT	6	91.45	5	91.45
YB	7	98.29	7	98.29

Overall, the bounded distance criterion D_A^2 performed better than the unbounded criterion D^2 in all aspect of discriminative power of the first selected feature subsets, classification accuracy of the first feature subset and its highest accuracy, and maximum convergence of correct classification before deterioration based on feature subset. These findings proof that the bounded distance criterion D_A^2 is superior in selecting subset of features that is important for higher classification accuracy for LLDF model. One weakness of this model is that there is no guarantee for features from both sensors to be selected especially when their individual ranking scored low average distance. From the perspective of the adulterated honey classification, confusion to recognize an observation belongs to its original group was largely influenced by the adulterated honey mixture as describe in Table 3.3. Meticulous observations from the classification reports executed by LDA rule lead to similar pattern of wrongly classified the adulterated honey especially the mixture of pure honey with cane sugar and pure honey with beet sugar.

4.4 Results for Intermediate Level Data Fusion

The implementation of PFFS for the ILLDF model was performed using Algorithm 3.3 for the unbounded $[0, \infty)$ Mahalanobis distance estimated by D^2, and Algorithm

3.4 for the bounded $[0,1]$ Mahalanobis distance estimated by D_A^2. The results of these procedures were the ten separate lists of ranked e-nose and e-tongue features for all the datasets. An example of such results for AG honey type can be observed from Table 4.14 for e-nose data and Table 4.15 for e-tongue data. The table clearly itemized the feature ranking from both sensors when performed separately, in accordance to the ILDF model.

To mention some, the feature ranking of e-nose data for AG honey based on the unbounded distance were led by N23 (11,949.06), N5 (11,343.20), N29 (10,027.52), etc. and for the bounded distance, the superior features were N20 (0.9476), N29 (0.9456), N5 (0.9399) and follows. For the e-tongue, superior features for the unbounded criterion were ranked by T11 (762.64), T2 (236.18), T3 (193.98), etc. and for the bounded distance, features were led by T11 (0.8626), T9 (0.8471), T2 (0.8351) and as follows. The least important features for both criteria can be observed from Tables 4.14 and 4.15. These tables were obtained for all ten data set for both e-nose and e-tongue sensors. However, only for the AG honey results were included in this section, and the results of the other types of honey are attached in Appendix C.

From this ILDF ranking list, select the first $(\mathcal{P}_{100.0\to83.8})and(\mathcal{P}_{100.0\to44.4})$ percentiles of features from e-nose and e-tongue list, and fuse the features to form the 1st feature subset, and repeat for the next $(\mathcal{P}_{96.7})$and$(\mathcal{P}_{88.8})$, $(\mathcal{P}_{93.5})$and$(\mathcal{P}_{77.7})$ to form the 2nd and 3rd feature subsets from fusion of e-nose and e-tongue data. Then, performed classification for every selected feature subset to measure the classification accuracy. Unlike the LLDF model where we have specified to select the first $(\mathcal{P}_{68.2})$ features

from the ranked list, in the ILDF model, the feature subsets selection stop when the maximum accuracy achieved begun to deteriorate. Repeat the process for the rest of the ranked features.

Table 4.14

Results of Feature Ranking for ILDF based on Bounded and Unbounded Mahalanobis Distance for e-nose AG Honey

D^2				D_A^2			
Feature	**Criterion Value**	**Rank**	**Percent**	**Feature**	**Criterion Value**	**Rank**	**Percent**
N23	11,949.06	1	100.00%	N20	0.94762	1	100.00%
N5	11,343.20	2	96.70%	N29	0.94564	2	96.70%
N29	10,027.52	3	93.50%	N5	0.93991	3	93.50%
N31	4,040.16	4	90.30%	N26	0.93304	4	90.30%
N9	3,680.18	5	87.00%	N10	0.93268	5	87.00%
N26	3,482.48	6	83.80%	N23	0.93215	6	83.80%
N11	3,074.39	7	80.60%	N28	0.92629	7	80.60%
N6	2,418.34	8	77.40%	N11	0.913	8	77.40%
N20	2,300.65	9	74.10%	N22	0.9097	9	74.10%
N10	1,793.69	10	70.90%	N17	0.9096	10	70.90%
N17	1,778.24	11	67.70%	N31	0.90339	11	67.70%
N15	1,407.55	12	64.50%	N19	0.9003	12	64.50%
N16	1,161.90	13	61.20%	N9	0.89784	13	61.20%
N28	1,129.29	14	58.00%	N8	0.89582	14	58.00%
N22	1,124.58	15	54.80%	N18	0.89035	15	54.80%
N8	1,094.92	16	51.60%	N15	0.88209	16	51.60%
N18	1,074.86	17	48.30%	N30	0.87812	17	48.30%
N13	869.54	18	45.10%	N13	0.87701	18	45.10%
N12	770.64	19	41.90%	N12	0.8754	19	41.90%
N30	741.57	20	38.70%	N7	0.87438	20	38.70%
N4	669.04	21	35.40%	N6	0.87119	21	35.40%
N7	646.60	22	32.20%	N21	0.87055	22	32.20%
N14	574.10	23	29.00%	N16	0.8691	23	29.00%

Table 4.14 Continued

N21	546.70	24	25.80%	N27	0.86195	24	25.80%
N19	536.86	25	22.50%	N2	0.85735	25	22.50%
N2	532.30	26	19.30%	N24	0.85559	26	19.30%
N1	500.41	27	16.10%	N1	0.85309	27	16.10%
N27	459.53	28	12.90%	N3	0.85157	28	12.90%
N3	446.28	29	9.60%	N14	0.84907	29	9.60%
N24	318.36	30	6.40%	N4	0.84573	30	6.40%
N25	317.15	31	3.20%	N25	0.81475	31	3.20%
N32	119.50	32	0.00%	N32	0.78103	32	0.00%

Table 4.15

Results of Feature Ranking for ILDF based on Bounded and Unbounded Mahalanobis Distance for e-tongue AG Honey

D^2				D_A^2			
Feature	**Criterion Value**	**Rank**	**Percent**	**Feature**	**Criterion Value**	**Rank**	**Percent**
T11	762.64	1	100.00%	T11	0.86259	1	100.00%
T2	236.18	2	88.80%	T9	0.84814	2	88.80%
T3	193.98	3	77.70%	T2	0.83513	3	77.70%
T7	180.85	4	66.60%	T1	0.81457	4	66.60%
T1	133.79	5	55.50%	T5	0.80781	5	55.50%
T5	96.45	6	44.40%	T8	0.77922	6	44.40%
T9	69.64	7	33.30%	T4	0.7299	7	33.30%
T8	55.04	8	22.20%	T3	0.67886	8	22.20%
T4	42.19	9	11.10%	T10	0.67518	9	11.10%
T10	41.37	10	0.00%	T7	0.66831	10	0.00%

Table 4.16

Classification Performances for Subset of Ranked Features and the Multivariate Mahalanobis Distance for AG Honey (ILDF)

Ranked Features	Criterion D^2		Ranked Features	Criterion D_A^2	
	Accuracy (100%)	Distance		Accuracy (100%)	Distance
N23, T11	77.78	4.6587×10^{-4}	N20, T11	73.50	1.3393×10^{-6}
N23, T11, N5, T2	93.16	7.0840×10^{-4}	N20, T11, N29, T9	98.29	2.6791×10^{-5}
N23, T11, N5, T2, N29, T3	100	3.3208×10^{-4}	N20, T11, N29, T9, N5, T2	98.29	3.5279×10^{-5}
N23, T11, N5, T2, N29, T3, N31, T7	99.15	1.0450×10^{-2}	N20, T11, N29, T9, N5, T2, N26, T1	100	3.5306×10^{-5}

Findings from AG honey data shows that the unbounded criterion is better in the selection of discriminant feature, and converge faster to 100% correct classification. The feature subsets formed by the unbounded criterion also proved that it‟s the correct features sub setting.

Table 4.17

Classification Performances for Subset of Ranked Features and the Multivariate Mahalanobis Distance for AS Honey (ILDF)

Ranked Features	Criterion D^2		Ranked Features	Criterion D_A^2	
	Accuracy (100%)	Distance		Accuracy (100%)	Distance
N26, T11	83.76	2.8362×10^{-4}	N15, T11	80.34	5.1572×10^{-6}
N26, T11, N5, T2	94.02	3.0537×10^{-4}	N15, T11, N5, T2	88.03	5.5300×10^{-6}
N26, T11, N5, T2, N29, T9	93.16	3.7995×10^{-4}	N15, T11, N5, T2, N23, T1	94.02	8.1263×10^{-6}
N26, T11, N5, T2, N29, T9, N31, T7	96.58	6.8728×10^{-4}	N15, T11, N5, T2, N23, T1, N11, T3	96.58	1.2886×10^{-5}
				96.58	1.2968×10^{-5}

Table 4.17 Continued

N26, T11, N5, T2, N29, T9, N31, T7, N15, T3	95.73	$1.4796x10^{-3}$	N15, T11, N5, T2, N23, T1, N11, T3, N29, T8

Similar pattern appeared for the second dataset where the unbounded criterion outperformed the bounded criterion. But none of the feature subsets recorded perfect classification.

Table 4.18

Classification Performances for Subset of Ranked Features and the Multivariate Mahalanobis Distance for ST Honey (ILDF)

Ranked Features	**Criterion D^2**		**Ranked Features**	**Criterion D_A^2**	
	Accuracy (100%)	**Distance**		**Accuracy (100%)**	**Distance**
N29, T2	70.09	$3.9728x10^{-5}$	N6, T2	74.36	$8.1560x10^{-7}$
N29, T2, N5, T9	78.63	$2.1503x10^{-4}$	N6, T2, N31, T11 * (omit N10)	91.45	$1.3990x10^{-5}$
N29, T2, N5, T9, N23, T11	89.74	$1.0995x10^{-3}$	N6, T2, N31, T11, N29, T1	94.02	$1.4067x10^{-6}$
N29, T2, N5, T9, N23, T11, N31, T1	93.16	$1.1023x10^{-3}$	N6, T2, N31, T11, N29, T1, N26, T9	93.16	$4.2582x10^{-6}$
N29, T2, N5, T9, N23, T11, N31, T1, N26, T8	92.31	$1.1040x10^{-3}$	N6, T2, N31, T11, N29, T1, N26, T9, N5, T8	91.45	$4.7475x10^{-6}$

For this dataset, the bounded criterion seems outperformed the unbounded criterion with faster convergence to maximum classification in the 3rd feature subset. In this dataset, feature N10 was omitted though was ranked with higher distance since the data vector creates singularity problem.

Table 4.19

Classification Performances for Subset of Ranked Features and the Multivariate Mahalanobis Distance for T Honey (ILDF)

Ranked Features	**Criterion D^2**		**Ranked Features**	**Criterion D_A^2**	
	Accuracy (100%)	**Distance**		**Accuracy (100%)**	**Distance**
N29, T7	92.39	1.0988×10^{-5}	N23, T11	87.18	1.3072×10^{-6}
N29, T7, N23, T2	83.76	1.9212×10^{-4}	N23, T11, N6, T2	87.18	3.2572×10^{-6}
N29, T7, N23, T2, N31, T11	94.87	1.7947×10^{-4}	N23, T11, N6, T2, N8, T8	97.43	3.2573×10^{-6}
N29, T7, N23, T2, N31, T11, N5, T9	91.45	2.9707×10^{-4}	N23, T11, N6, T2, N8, T8, N22, T1	94.02	3.2574×10^{-6}

Unlike other findings, this data demonstrates that the first feature subset with high discriminative power was given by the unbounded criterion, but faster convergence was obtained by the bounded criterion i.e. by the 3rd subset of features.

Table 4.20

Classification Performances for Subset of Ranked Features and the Multivariate Mahalanobis Distance for T3 Honey (ILDF)

Ranked Features	**Criterion D^2**		**Ranked Features**	**Criterion D_A^2**	
	Accuracy (100%)	**Distance**		**Accuracy (100%)**	**Distance**
N9, T2	91.45	1.3293×10^{-4}	N6, T2	84.62	2.4560×10^{-6}
N9, T2, N5, T11	95.73	9.8515×10^{-5}	N6, T2, N10, T11	93.16	1.7758×10^{-6}
N9, T2, N5, T11, N23, T1	97.44	1.7818×10^{-4}	N6, T2, N10, T11, N31, T1	94.02	1.7763×10^{-6}
N9, T2, N5, T11, N23, T1, N31, T9	96.58	3.366×10^{-4}	N6, T2, N10, T11, N31, T1, N29, T9	96.58	4.5752×10^{-6}
N9, T2, N5, T11, N23, T1, N31, T9, N26, T3	100	9.1028×10^{-5}	N6, T2, N10, T11, N31, T1, N29, T9, N26, T8	94.87	4.5744×10^{-6}

Again for this dataset, unbounded criterion outperformed the bounded criterion with higher classification accuracy by the first feature subset with 91.45% correct classification. By having the fifth subsets for the unbounded criterion, the correct classification converges to 100%.

Table 4.21

Classification Performances for Subset of Ranked Features and the Multivariate Mahalanobis Distance for TK Honey (ILDF)

Ranked Features	**Criterion D^2**		**Ranked Features**	**Criterion D_A^2**	
	Accuracy (100%)	**Distance**		**Accuracy (100%)**	**Distance**
N23, T11	81.20	7.6756×10^{-5}	N5, T11	85.47	4.0099×10^{-7}
N23, T11, N29, T2	85.47	1.3180×10^{-4}	N5, T11, N23, T2	88.89	2.7277×10^{-6}
N23, T11, N29, T2, N5, T9	88.89	3.0581×10^{-4}	N5, T11, N23, T2, N6, T1	93.16	2.9514×10^{-6}
N23, T11, N29, T2, N5, T9, N9, T1	93.16	3.0599×10^{-4}	N5, T11, N23, T2, N6, T1, N10, T8	94.87	2.9516×10^{-6}
N23, T11, N29, T2, N5, T9, N9, T1, N26, T8	96.58	3.0634×10^{-4}	N5, T11, N23, T2, N6, T1, N10, T8, N11, T9	94.87	5.7790×10^{-6}
N23, T11, N29, T2, N5, T9, N9, T1, N26, T8, N20, T7	95.72	3.4862×10^{-4}	N5, T11, N23, T2, N6, T1, N10, T8, N11, T9, N17, T5	94.87	5.7813×10^{-6}

The bounded criterion appears to perform better for TK honey dataset when the first subset of features gave higher classification accuracy compared to the unbounded criterion. However, in terms of convergences to highest accuracy, the unbounded criterion seems to be better.

Table 4.22

Classification Performances for Subset of Ranked Features and the Multivariate Mahalanobis Distance for TLH Honey (ILDF)

Ranked Features	Criterion D^2		Ranked Features	Criterion D_A^2	
	Accuracy (100%)	Distance		Accuracy (100%)	Distance
N29, T11	86.32	1.8835×10^{-5}	N6, T11	79.49	3.4784×10^{-7}
N29, T11, N9, T2	94.02	7.4032×10^{-5}	N6, T11, N20, T2	92.31	1.3632×10^{-6}
N29, T11, N9, T2, N26, T10	93.16	7.4041×10^{-5}	N6, T11, N20, T2, N9, T1	96.58	1.3632×10^{-6}
N29, T11, N9, T2, N26, T10, N17, T5	96.66	7.4068×10^{-5}	N6, T11, N20, T2, N9, T1, N18, T8	96.58	1.3632×10^{-6}

Again for this dataset, higher discriminative power belongs to the unbounded criterion. But the bounded criteria converge to maximum accuracy in the third feature subset faster than the unbounded criteria.

Table 4.23

Classification Performances for Subset of Ranked Features and the Multivariate Mahalanobis Distance for TN Honey (ILDF)

Ranked Features	Criterion D^2		Ranked Features	Criterion D_A^2	
	Accuracy (100%)	Distance		Accuracy (100%)	Distance
N23, T3	71.79	2.0409×10^{-4}	N22, T11	92.31	2.3486×10^{-7}
N23, T3, N29, T11	92.31	2.3035×10^{-4}	N22, T11, N29, T2	96.58	1.1472×10^{-6}
N23, T3, N29, T11, N5, T2	95.73	3.3661×10^{-4}	N22, T11, N29, T2, N10, T1	97.44	1.1473×10^{-6}
N23, T3, N29, T11, N5, T2, N9, T1	100	3.3696×10^{-4}	N22, T11, N29, T2, N10, T1, N26, T7	97.44	1.9685×10^{-6}

Table 4.23 Continued

N23, T3, N29, T11, N5, T2, N9, T1, N26, T7	98.29	4.2680×10^{-4}	N22, T11, N29, T2, N10, T1, N26, T7, N5, T8	97.44	2.1097×10^{-6}

Obvious results were identified from this dataset where the bounded criterion successfully found the first two fusion feature with higher discriminative power. However, the bounded criterion converges to 100% correct classification early than the bounded distance.

Table 4.24

Classification Performances for Subset of Ranked Features and the Multivariate Mahalanobis Distance for WT Honey (ILDF)

Ranked Features	**Criterion D^2**		**Ranked Features**	**Criterion D_A^2**	
	Accuracy (100%)	**Distance**		**Accuracy (100%)**	**Distance**
N5, T11	93.16	2.0357×10^{-5}	N23, T11	92.31	3.5512×10^{-6}
N5, T11, N29, T2	96.16	5.9270×10^{-5}	N23, T11, N28, T2	96.58	4.2118×10^{-6}
N5, T11, N29, T2, N23, T10	94.02	3.1295×10^{-4}	N23, T11, N28, T2, N20, T10	94.87	4.2183×10^{-6}
N5, T11, N29, T2, N23, T10, N26, T7	91.45	3.4956×10^{-4}	N23, T11, N28, T2, N20, T10, N31, T1	99.15	4.2901×10^{-6}

Performance for this dataset is almost similar with small difference. In terms of discriminative power, the first feature subset for unbounded criterion leads the list but with small difference to the bounded criterion. However, the bounded criterion converges to the maximum correct classification with 0.85% error.

Table 4.25

Classification Performances for Subset of Ranked Features and the Multivariate Mahalanobis Distance for YB Honey (ILDF)

Ranked Features	Criterion D^2		Ranked Features	Criterion D_A^2	
	Accuracy (100%)	Distance		Accuracy (100%)	Distance
N5, T11	82.05	2.6609×10^{-5}	N11, T11	84.62	2.0198×10^{-7}
N5, T11, N23, T2	92.31	3.2583×10^{-4}	N11, T11, N28, T1	92.31	2.0196×10^{-7}
N5, T11, N23, T2, N29, T1	97.44	3.3887×10^{-4}	N11, T11, N28, T1, N9, T2	91.45	1.1308×10^{-6}
N5, T11, N23, T2, N29, T1, N26, T10	94.87	3.4256×10^{-4}	N11, T11, N28, T1, N9, T2, N26, T9	98.29	4.1374×10^{-6}

Finally, for the YB honey, apparently the bounded criterion has selected the first feature subset with higher discriminative power compared to the unbounded criterion. And, the bounded criterion also managed to converge to the highest classification accuracy in the fourth feature subset.

4.5 Discussion for Feature Selection in Intermediate Level Data Fusion

The discussions of the ILDF findings are expressed based on the univariate and multivariate distance for feature ranking and selected feature subset, selections discriminative power of the selected first feature subsets, classification accuracy (first feature subset and highest accuracy), and maximum convergence of correct classification before deterioration based on feature subset. The choice of features to be selected for the ILDF model is based on the porposed PFFS which earlier described in section 3.2.

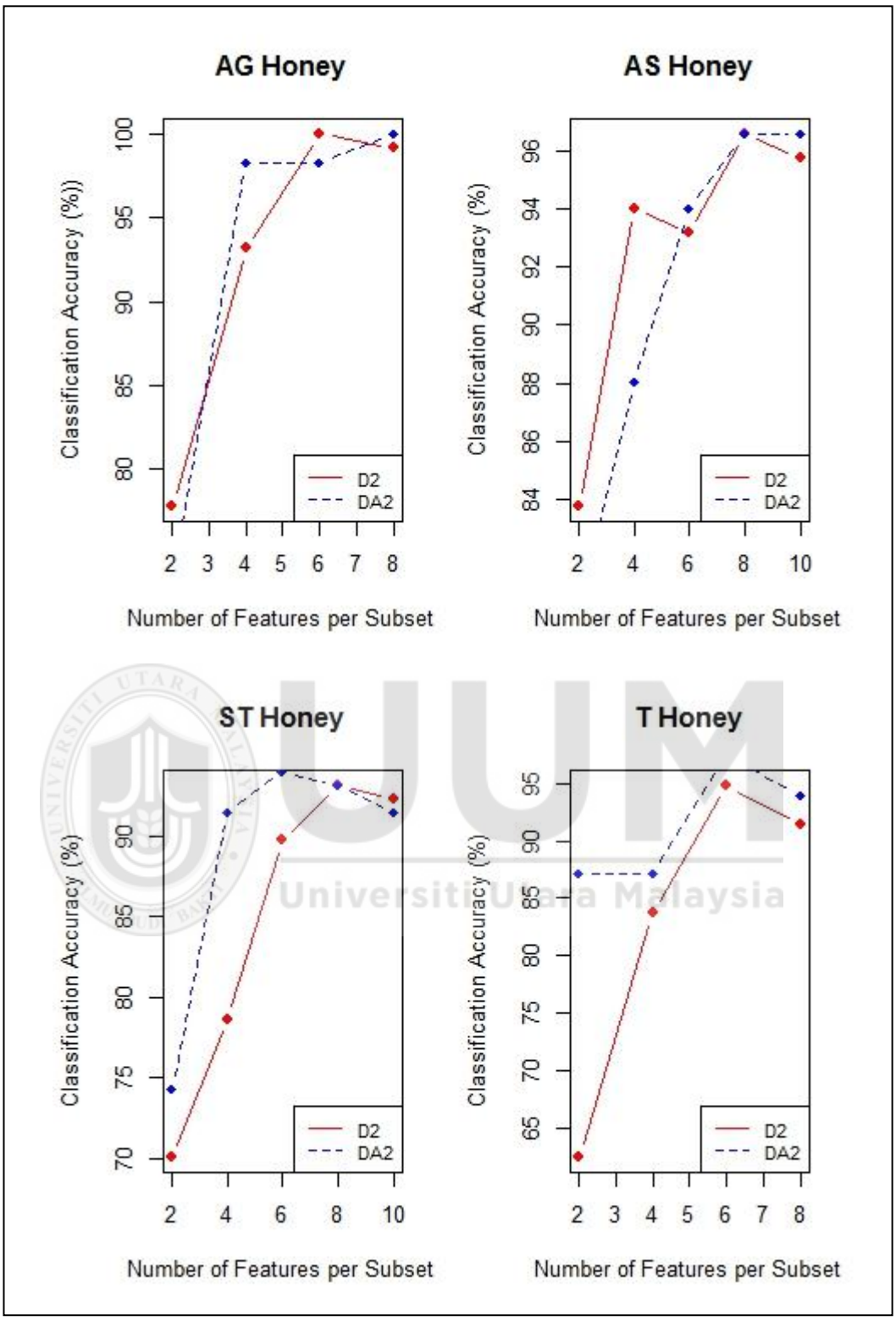

Figure 4.4. Comparison of Classification Accuracy based on D^2 and D_A^2 for Feature Subsets of AG, AS, ST and T Honey Types (ILDF)

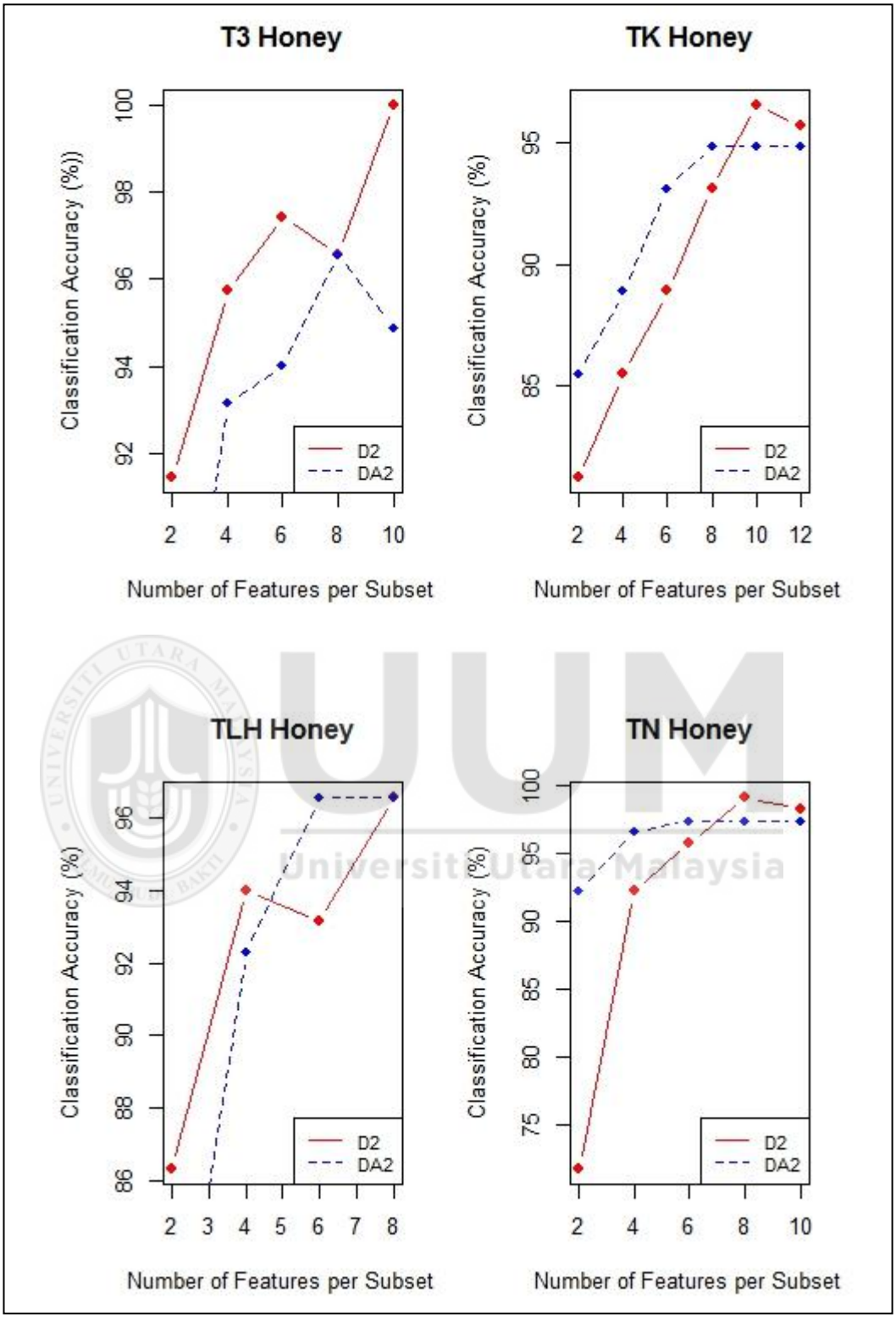

Figure 4.5. Comparison of the Classification Accuracy based on D^2 and D_A^2 for Feature Subsets of T3, TK, TLH and TN Honey Types (ILDF)

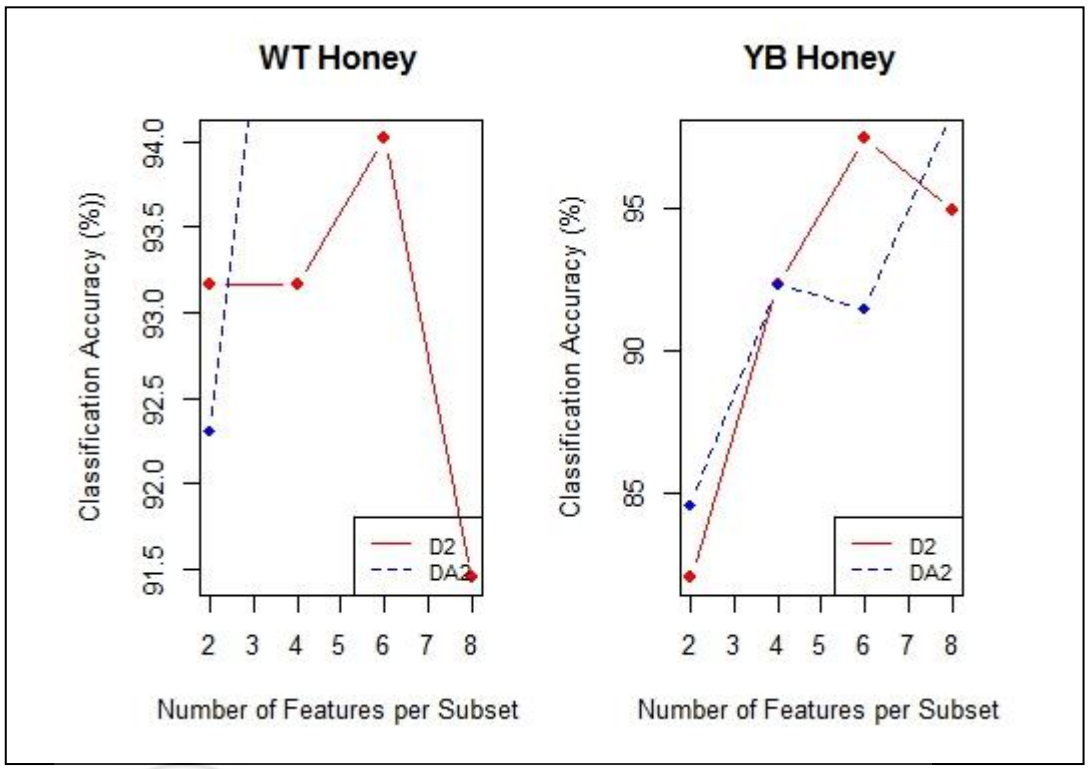

Figure 4.6. Comparison of the Classification Accuracy based on D^2 and D_A^2 for Feature Subsets of WT and YB Honey Types (ILDF)

Based on the results in Tables 4.16 to 4.25 and Figures 4.4 to 4.6, the following discussions were made.

Univariate and multivariate distances for feature ranking and selected feature subset. In the first process of feature ranking, the univariate Mahalanobis distance based on the unbounded and bounded criteria has shown its ability to recognize discriminant features distinctively. Distances given by the unbounded criterion represent the average distance for the particular feature across 11 groups. The highest average distance were among the e-nose features with as high as 12,593.81 (ILDF - dataset ST - feature N29), but not for the e-tongue features which scored much lower average distance. Unlike the previous criteria, bounded distance gave even noticeable features using average distance among 11 groups between the values 0 to 1.

Comparatively, by using these criteria, e-nose features recorded higher bounded distance compared to e-tongue features.

Discriminative power of the first selected feature subset. Regularly, features selected by both criteria were almost the same, but the features were ranked differently. Consider the selected features for AG and ST (below), the selected features for the unbounded criteria are highlighted in green while the bounded criterion is highlighted in grey. The selection results indicate that both criteria were able to detect discriminative features, but of different discriminative power.

Table 4.26

Illustration for the Comparison of Ranked Fused Features (ILDF Model) for AG and ST Honey Dataset.

ILDF (AG)	D^2	**N5**	**N20**	**N23**	**N26**	**N29**	**T1**	**T2**	**T3**	**T9**	**T11**
	Rank	3		1		5		4	6		2
	D_A^2	**N5**	**N20**	**N23**	**N26**	**N29**	**T1**	**T2**	**T3**	**T9**	**T11**
	Rank	5	1		7	3	8	6		4	2

ILDF (ST)	D^2	**N5**	**N6**	**N23**	**N26**	**N29**	**N31**	**T1**	**T2**	**T8**	**T9**	**T11**
	Rank	3		5	9	1	7	8	2	10	4	6
	D_A^2	**N5**	**N6**	**N23**	**N26**	**N29**	**N31**	**T1**	**T2**	**T8**	**T9**	**T11**
	Rank	9	1		7	5	3	6	2	10	8	4

Based on the above summary, for dataset AG honey, generally the unbounded criteria have different agreement on the selection of features from e-nose (D^2=>N23, D_A^2=>N20) but complete agreement for features from e-tongue (D^2=>T11, D_A^2=>T11). The same occurred to dataset ST where the selection of features by D^2 was feature N29, and D_A^2 was feature N6, but both criteria agreed to

choose T11 as their discriminative feature. Out of 10 datasets tested, this pattern also appeared in dataset AS (T11), T3 (T2), TK (T11), TLH (T11), WT (T11), and YB (T11).

Classification Accuracy (first feature subset and highest accuracy). For the ILDF model, performance of both criteria in terms of achieving correct classification by using the first selected feature subset is even. Out of ten datasets, five datasets (AG, AS, T3, TLH, and WT) obtained higher accuracy rate based on their first subset feature using the unbounded criteria. Whereas, the other five datasets (ST, T, TK, TN and YB) recorded higher accuracy using the bounded criteria based on the first feature subset. Details of these can be referred to Tables 4.16 to 4.25.

Maximum convergence of correct classification before deterioration based on feature subset. This refers to the highest classification accuracy achieved based on some particular subset of features from both criteria. It has been summarized in Table 4.27 below and can be observed in Figures 4.4, 4.5 and 4.6 where the first highest peak occurred from both criteria. The unbounded distance obtained maximum convergence sixth times with one even compared to the bounded distance. However, Table 4.27 shows that maximum convergence for the bounded criteria were attained as early as in the second feature subset to fourth feature subset. And for the unbounded criteria, maximum convergence were between the third to fifth feature subset

Table 4. 27

Comparison of Performance for the Unbounded and Bounded Feature Selection based on Feature Subset Number and Correct Classification (ILDF)

DATASET	D^2		D_A^2	
	Subset Number	**Accuracy**	**Subset Number**	**Accuracy**
AG	3	100.00%	4	100.00%
AS	4	96.58%	4	96.58%
ST	4	93.16%	3	94.02%
T	3	94.87%	3	97.43%
T3	5	100.00%	5	94.87%
TK	5	96.58%	4	94.87%
TLH	4	96.58%	3	96.58%
TN	4	99.15%	3	97.44%
WT	3	94.02%	2	96.58%
YB	3	97.44%	4	98.29%

Overall, the bounded distance criterion D_A^2 performed better than the unbounded criterion D^2 in all aspect of discriminative power of the first selected feature subsets, classification accuracy of the first feature subset and its highest accuracy, and maximum convergence of correct classification before deterioration based on feature subset. These findings proof that the bounded distance criterion D_A^2 is superior in selecting subset of features that is important for higher classification accuracy for LLDF model.

4.6 Conclusion

Generally, the feature selection for fusion of features was successfully performed for LLDF where the discriminant features were identified based on the univariate fused feature ranking. The confirmation of the selected discriminant feature subset was obtained by measuring the accuracy of classification using the chosen feature

subsets. These processes were repeated over different selected feature subsets using percentile forward feature selection $(\mathcal{P}_H)$ for all the ten datasets. Then, the multivariate Mahalanobis distances were also computed to show the average distance among the included feature subsets and groups. For the LLDF model, features from e-nose dominate the highest fused feature ranking for both unbounded and bounded criteria, left out the features from e-tongue from being included among discriminant features. Only small number of e-tongue features was procured in the selected discriminant feature subsets for the LLDF.

From the perspective of the ILDF model, basically the implementation of feature selection to replace feature extraction was also effective. Even though from the comparison of the classification accuracy based on the overall selected feature subsets between the LLDF and ILDF shows a difference performance, the ILDF model has shown more flexibility in selecting features from the two sensors. In fact, the number of discriminant features selected for the ILDF was lower than the number of discriminant features for the LLDF model. The findings could be because the way features were fused after individual feature ranking was performed separately. The way the ranked features from both sensors were fused was based on $(\mathcal{P}_{100.0\rightarrow44.4})$ and $(\mathcal{P}_{100.0\rightarrow83.8})$ percentile forward feature selection for e-tongue and e-nose, respectively. The implementation of these selections seems effective for ILDF model.

From the perspective of the adulterated honey classification, confusion to recognize an observation belongs to its original group was largely influenced by the adulterated honey mixture as describe in Table 3.3. Meticulous observations from the classification reports executed by LDA rule lead to similar pattern of wrongly

classified of the adulterated honey especially the mixture of pure honey with cane sugar and pure honey with beet sugar.

From the results, the study concludes that the bounded Mahalanobis distance $\left(D_A^2\right)$ is comparatively better than the unbounded Mahalanobis distance $\left(D^2\right)$ in performing feature subsets selection to trace the discriminant features. These findings were applicable for the multi sensor data fusion of e-tongue and e-nose specifically for the LLDF and ILDF models.

In summary, this chapter has provided answers to all the objectives of this study. The following chapter will summarize the overall conclusions of this study, the key contributions of the thesis, as well as some possible limitations and the direction for future work.

CHAPTER FIVE
CONCLUSION AND FUTURE WORK

5.1 Conclusion of Study

This study aims to study the potential discriminant analysis using feature selection to replace the traditional discriminant analysis (PCA) via feature extraction in multi sensor data fusion. The study revealed that limited works were devoted to study the weaknesses of the traditional approach. One big weakness of the approach is the failure to identify specific discriminant features that contribute to highest classification accuracy. The discriminant analysis with feature extraction approach is highly dependent only on few selected combination of features trigged by percent of variance explained or the largest eigenvalues. Thus, the feature extraction approach is less objective in highlighting the significance of each specific feature and their contribution towards the classification accuracy.

Two models of multi sensor data fusion applied for this study are the low and intermediate level data fusions which are commonly preferred by engineers in multi sensor data fusion. Since insufficient researches are available to improvise the traditional approach, this study tries to introduce a simple yet effective method to perform classification in the low and intermediate data fusion problem. The sensor fusion for this study comprised the e-nose (which mimics the human smell sense) and e-tongue (which mimics the human taste sense). Features from e-nose were formed by the 32 array of sensors, while features from e-tongue were created by 11 array of sensor which works based on gaseous and liquid, respectively.

The fusion of features from the sensors created a new dimension of data fusion which befitted the curse of dimension within multi-group $(\mathcal{G}=11)$. The investigation of the appropriate criteria which suited the previous aforementioned challenges preceded the potential of distance based feature selection which measure the separability among the multi-group centroids (group means). The maximum separability among the multi-group centroids given by the average distance of a feature became the focal solution to be proposed. The concept was adopted from a study done 25 years ago.

Ray and Turner (1992) have proposed the bounded Mahalanobis distance-based for two new feature evaluation criteria $\left(\Delta_A^2 \text{ and } \Delta_B^2\right)$ in a pattern recognition problem of isolated hand printed numeric characters. As far as this study is concerned, this study proved that one of the proposed criterions i.e. $\left(\Delta_A^2\right)$ has great potential to be applied in the problem of identifying discriminant features in low and intermediate level data fusion. In order to show the flexibility and potential of the bounded Mahalanobis distance $\left(\Delta_A^2\right)$, the unbounded Mahalanobis distance $\left(\Delta^2\right)$ was also included in the analyses.

The research methodology applied in this study mainly devoted to the unbounded and bounded Mahalanobis distance criteria to identify the most discriminative features that give the highest classification accuracy. The performances of both criteria will be compared and contrast according to some evaluation points. In order to achieve these, four study objectives were formed. These objectives were implemented based on Figure 3.5, the proposed feature selection strategies based on the unbounded $\left(D^2\right)$ and bounded $\left(D_A^2\right)$ mahalanobis distance for LLDF and ILDF.

Specific implementations of the proposed feature selection were referred to Algorithm 3.1 for fused feature ranking for unbounded distance for LLDF, Algorithm 3.2 for fused feature ranking for bounded distance for LLDF, Algorithm 3.3 for single feature ranking for unbounded distance for ILDF, and Algorithm 3.4 for single feature ranking for bounded distance for ILDF. The classification accuracy using LDA with the leave-one-out error estimation were implemented as described in Figures 3.6 and 3.7 for the unbounded and bounded criteria for LLDF, and Figures 3.8 and 3.9 for the unbounded and bounded criteria for ILDF.

For this study, no stopping rule is needed since the most potential features were selected based on the average distance (which measures the average separability among multi-group) obtained from the ranking process. A heuristic selection of the number of features to be included for the feature subset selection was determined for the LLDF and ILDF models, called the percentile forward feature selection. For the LLDF model, the percentile forward feature selection is based on the top highest percentiles $\left(\mathcal{P}_{100.0\rightarrow 68.2}\right)$ of the ranked features.

Model	Feature Source	Selected Percentile & Number of Feature (D^2)	Selected Percentile & Number of Feature (D_A^2)
LLDF	Fusion	$\left(\mathcal{P}_{100.0\rightarrow 68.2}\right)$ 14 features	$\left(\mathcal{P}_{100.0\rightarrow 68.2}\right)$ 14 features

It means, only the highest 14 ranked features were included for the discriminant search. The ranked features in LLDF which came from the e-tongue and e-nose sensors were fused at the data level, then the univariate unbounded and bounded distance criteria were computed, and finally they were ranked from highest to lowest. By doing these, the study is being simplified in the sense that no global search is needed, and concentration was given only to potential feature.

The similar concept was applied for the ILDF model. To simplify the search of discriminant feature subset, a bit different percentile forward feature selection was applied in the second model. Since features from the e-tongue and e-nose were fused at the intermediate level, the fusion occurred when the most discriminant features from both sensors were identified from the individual ranked list. The way the percentile forward feature selection was performed was based on the following percentiles from each ranked features, separately.

Model	Feature Source	Selected Percentile & Number of Feature (D^2)	Selected Percentile & Number of Feature (D_A^2)
ILDF	e-tongue	$(\mathcal{P}_{100.0\rightarrow66.6})$ 4 features to $(\mathcal{P}_{100.0\rightarrow44.4})$ 6 features	$(\mathcal{P}_{100.0\rightarrow66.6})$ 4 features to $(\mathcal{P}_{100.0\rightarrow44.4})$ 6 features
	e-nose	$(\mathcal{P}_{100.0\rightarrow90.3})$ 4 features to $(\mathcal{P}_{100.0\rightarrow83.8})$ 6 features	$(\mathcal{P}_{100.0\rightarrow90.3})$ 4 features to $(\mathcal{P}_{100.0\rightarrow83.8})$ 6 features

With this approach, potential features from both sensors have the equal chances to be included in the feature selection, regardless of their overall ranking. By doing this, individual performance is reflected knowing that e-tongue and e-nose performed the object detection uniquely according to their designed functions. Again, by implementing percentile forward feature selection, no global search is required, and concentration was given only to promising feature.

One advantage of the second model is that, the selection process of the discriminant features gave credit to both sensors equally where domination of single sensor is avoided. For this study, other feature search approaches such as backward and stepwise feature selections were not preferable because the approaches were

impractical for the proposed procedures. For example, adapting the stepwise or backward search would cause the inclusion process will be dominated by features from e-nose due to higher average distance, while deletion would prefer the features from e-tongue due to small distance.

Through the implementation of the algorithms and classification rules, all the research objectives were attained, and the appropriate discriminant analysis based on feature selection to replace the discriminant analysis based on feature extraction was demonstrated and confirmed. In order to prove which of the two criteria either the unbounded or bounded performed better, several comparisons in terms of the univariate/multivariate distance for feature ranking, discriminative power of the first selected feature subset, classification accuracy (first feature subset and highest accuracy) and maximum convergence of correct classification before deteriorate were observed and analyzed. Through the findings, the unbounded Mahalanobis distance $\left(\Delta_A^2\right)$ estimated by $\left(D_A^2\right)$ was found to be more influential criteria to select discriminant feature subset based on average separability for both LLDF and ILDF models. In addition, the bounded criterion was able to select fewer numbers of features in its feature subsets compared to the unbounded distance criterion.

In summary, this study has introduced alternative strategies for the traditional discriminant analysis based on feature extraction in the classification problem of multi sensor data fusion namely low and intermediate level data fusion. The applications of the unbounded and bounded Mahalanobis distance for the proposed discriminant analysis based on feature selection have shown some promising results. Among the two criteria, the bounded criterion revealed better achievements in terms of less number of features in the selected feature subsets, fast convergence of

accurate classification and the classification accuracy of the first selected pair of features was higher. And these findings can be observed when applied in the ILDF model.

5.2 Contribution of Study

The findings of this study have contributed to three different aspects including the knowledge of feature selection, methodology, algorithms and application. Each of these is explained in the following details:

1. Knowledge of feature selection. This study has developed a new feature selection approach called percentile forward feature selection that suitable for fused data. The implementation of the proposed feature subset selection suits the low and intermediate level of data fusion models where discriminant features were successfully identified. Apart from that, the concept of unbounded and bounded Mahalanobis distances appeared to be the essence of value added to the research. These two criteria could be an important consideration when performing feature selection based on separability among multi-group centroids.

2. Discriminant Analysis Methodology for the LLDF and ILDF. For years, the LLDF and ILDF have been implemented using the traditional discriminant analysis based on feature extraction. Dependency on the first few combinations of features given the highest variance explained or highest eigenvalues restricted the possibility to identify the most discriminant features that contribute to the utmost classification accuracy. The criteria of unbounded and bounded Mahalanobis distance have created a new dimension

of study in the LLDF and ILDF methodology. Discriminant analysis based on feature selection using the unbounded and bounded Mahalanobis distance gave some flexibility and practicality to the implementation of feature selection and classification of LLDF and ILDF models.

3. Algorithms Development. All the previous contributions were accomplished via development of some associated algorithms. In specific, this study has designed new algorithms to compute univariate unbounded and bounded Mahalanobis distances and perform fused feature ranking for the discriminant analysis in LLDF and ILDF. Next, additional algorithms were created by modifying a built-in **R** function from `HDMD package` to compute the multivariate unbounded and bounded Mahalanobis distances for the final evaluation of feature subset selection in the LLDF and ILDF models. These algorithms have simplified the execution of every procedures of discriminant analysis in the LLDF and ILDF models.

4. Applications of Discriminant Analysis based on Feature Selection. In this study, the proposed criteria and procedures were tested on ten datasets of honey collected from two different sensors (e-tongue and e-nose). Knowing that the sensors functioned uniquely according to the designed procedure, when their data are to be fused and analyzed, an appropriate approach should be followed so that the sensors contributions are equally assessed and valued. Therefore, a more reliable and accurate classification result could be found.

5.3 Direction for Future Work

This work has not yet answered all problems related to replacing the traditional discriminant analysis based on feature extraction with the discriminant analysis based on feature selection approach. Future work is necessary to fill some flaws of this study, so that better classification results of this type can be produced. Directions for future research are summarized as follows.

1. Chapter 3 of this study proposed percentile forward feature selection which heuristically determines what percentiles of ranked features to be included for the feature subset selection in the LLDF and ILDF models. Different percentiles of higher values may lead to new findings of which discriminative features influential for correct classification. The study also has not attempted to pair the features with the highest univariate Mahalanobis distance with the features of the lowest distance to be in subset of feature, and the classification results remain unresolved.

2. Classification rule for this study is based on the parametric rule with classification accuracy as the performance measure. This is due to the nature of some of the variables that follow the normal distribution and some that are not. Thus, it may be worthwhile to execute the proposed procedures and algorithms using the nonparametric classification rules and compare the findings.

3. The study only tested dataset of adulterated honey. Even though the dataset are quite many, different datasets other than adulterated honey could produce different distance and classification results. However, classification of pure

honey from the adulterated honey with cane sugar or beet sugar is a challenging task. What more when the adulterated honeys were purposely mixed to imitate the real honey, and the dataset were formed into multi-group of many different adulterated honeys.

4. As previously mentioned, the study only includes two sensors such as e-tongue and e-nose. These fused sensors produced about 42 features which became the input for the percentile forward feature selection. Similar and /or somewhat different findings were obtained when the LLDF or the ILDF models were applied using the two distance criteria. Therefore, the proposed procedures and algorithms using the nonparametric classification rules could be tested for different other sensors, where a deeper look at the relationships among selected discriminant features could be learnt

5. Finally, application of the proposed procedures and algorithms should be implemented in the higher level data fusion. Less study is devoted for this model, but the execution of this model applying the proposed approaches would lead to interesting findings.

REFERENCES

Abdul Aziz, A. H., Md. Shakaff, A. Y., Farook, R., Adom, A. H., Ahmad, M. N., & Mahat, N. I. (2011). Simple implementation of an electronic tongue for taste assessments of food and beverages products. *Sensors and Transducers Journal*, *132* (9), 136-150.

Achariyapapaopan, T., & Childers, D. G. (1985). Optimum and near optimum feature selection for multivariate data. *Signal Processing*, *8*, 121-129.

Afifi, A., May, S., & Clark, V. A. (2004). *Computer-aided multivariate analysis*. CRC Press.

An, A. (2003). Learning classification rules from data. *Computers and Mathematics with Applications*, *45*, 737-748.

Anderson, T. W. (1951). Classification by multivariate analysis, Psychometrika 16, 31-50.

Apetrei, C., Apetrei, I. M., Villanueva, S., de Saja, J. A., Gutiererez-Rosales, F., & Rodriguez-Mendez, M. L. (2010). Combination of an e-nose, an e-tongue, and e-eye for the characterization of olive oils with different degrees of bitterness. *Analytica Chimica Acta*, *663*, 91-97. doi:10.1016/j.aca.2010.01.034

Aranda-Sanchez, J. I., Baltazar, A., & González-Aguilar, G. (2009). Implementation of a Bayesian classifier using repeated measurements for discrimination of tomato fruit ripening stages. *Biosystems Engineering*, *102*, 274-284. doi:10.1016/j.biosystemseng.2008.12.005

Baldwin, E. A., Bai, J., Plotto, A., & Dea, S. (2011). Electronic noses and tongues: applications for the food and pharmaceutical industries. *Sensors*, *11*, 4744-4766. doi:10.3390/s110504744

Banerjee, R., Tudu, B., Shaw, L., Jana, A., Bhattacharyya, N., & Bandyopadhyay, R. (2012). Instrumental testing of tea by combining the responses of electronic nose and tongue. *Journal of Food Engineering*, *110*, 356-363. doi:10.1016/j.foodeng.2011.12.037

Berrueta, L. A., Alonso-Salces, R. M., & Héberger, K. (2007). Supervised pattern recognition in food analysis. *Journal of Chromatography A*, *1158*, 196-214. doi:10.1016/j.chroma.2007.05.024

Blum, A. L., & Langley, P. (1997). Selection of relevant features and examples in machine learning. *Artificial Intelligence*, *97*, 245-271. doi:10.1016/S0004-3702(97)00063-5

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 B*, *88*, 80-88.

Breiman, L., Freidman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and Regression Trees*. Wadsworth & Brooks/Cole Advanced Books & Software, Monterey, CA.

Bruwer, M., MacGregor, J. F., & Bourg Jr., W. M. (2007). Fusion of sensory and mechanical testing data to define measures of snack food texture. *Food Quality and Preference*, *18*, 890-900. doi:10.1016/j.foodqual.2007.03.001

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*, *525*, 133-139. doi:10.1016/j.aca.2004.07.062

Byrne, D. V., O‟sullivan, M. G., Bredie, W. L. P., Anderson, H. J., & Martens, M. (2003). Descriptive sensory profiling and physical/chemical analyses of warmed-over flavour in pork patties from carriers and non-carriers of RN allele. *Meat Science*, *63*, 211-224.

Casalinuovo, I. A., Di Pierro, D., Coletta, M, & Di Francesco, P. (2006). Application of electronic noses for disease diagnosis and food spoilage detection. *Sensors*, *6*, 1428-1439.

Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. *Computers and Electrical Engineering*, *40*, 16-28. doi:10.1016/j.compeleceng.2013.1.024

Chittineni, C. B. (1980). Efficient feature-subset selection with probabilistic distance criteria. *Information Sciences*, *22*, 19-35.

Cimander, C., Carlsson, M., & Mandenius, C. (2002). Sensor fusion for on-line monitoring of yoghurt fermentation. *Journal of Biotechnology*, *99*, 237-248.

Cios, K. J., Swiniarski, R. W., Pedrycz, W., & Kurgan, L.A. (2007). Feature extraction and selection methods. In *Data Mining*, Springer US, 133-233. doi:// 10.1007/978-0-387-36795-8_7 (chapter in book).

Ciosek, P., & Wróblewski, W. (2011). Potentiometric electronic tongues for foodstuff and biosample recognition-an overview. *Sensors*, *11*, 4688-4701. doi: 10.3390/s110504688

Ciosek, P., Brzózka, Z., & Wróblewski, W. (2004). Classification of beverages using a reduced sensor array. *Sensors and Actuators B*, *103*, 76-83. doi:10.1016/j.snb.2004.04.038

Ciosek, P., Brzózka, Z., Wróblewski, W., Martinelli, E., Di Natale, C., & D‟Amico, A. (2005). Direct and two-stage data analysis procedures based on

PCA, PLS-DA and ANN for ISE-based electronic tongue-effect of supervised feature extraction. *Talanta*, *67*, 590-596. doi: 10.1016/j.talanta.2005.03.006

Ciosek, P., & Wróblewski, W. (2006). The analysis of sensor array data with various pattern recognition techniques. *Sensors and Actuators B*, *114*, 85-93. doi: 10.1016/j.snb.2005.04.008

Cole, M., Covington, J. A., & Gardner, J. W. (2011). Combined electronic nose and electronic tongue for a flavor sensing system. *Sensors and Actuators B*, *156*, 2, 832-839. doi:10.1016/j.snb.2011.02.049

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*, *101*, 485-491.

Craven, M. A., Gardner, J. W., & Bartlett, P. N. (1996). Electronic noses – development and future prospects. *Trends in Analytical Chemistry*, *15* (9), 486-493.

Dash, M., & Liu, H. (1997). Feature Selection for Classification. *Intelligent Data Analysis*, *1*, 131-156.

Dernoncourt, D., Hanczar, B., & Zucker, J-D. (2014). Analysis of feature selection stability on high dimension and small sample. *Computational Statistics and Data Analysis*, *71*, 681-693. http://dx.doi.org/10.1016/j.csda.2013.07.012

Devijver, P. A. & Kittler, J. (1982). *Pattern recognition: a statistical approach*. London: Prentice-Hall.

Dillon, W. R., & Goldstein, M. (1984). *Multivariate analysis, methods and applications*. USA: John Wiley & Sons.

Di Natale, C., Davide, F., & Di Amico, A. (1995). Pattern recognition in gas sensing: well-stated techniques and advances. *Sensors and Actuators B*, *23*, 111-118.

Di Natale, C., Paolesse, R., Macagnano, A., Mantini, A., D'Amico, A., Legin, A., ... & Vlasov, Y. (2000). Electronic nose and electronic tongue integration for improved classification of clinical and food samples. *Sensors and Actuators B: Chemical*, *64*(1), 15-21.

Di Rosa, A. R., Leone, F., Cheli, F., & Chiofalo, V. (2017). Fusion of electronic nose, electronic tongue and computer vision for animal source food authentication and quality assessment – a review. *Journal of Food Engineering*, *210*, 62-75.

Dixon, S. J., & Brereton, R. G. (2009). Comparison of performance of five common classifiers represented as boundry methods: Euclidean distance to

centroids, linear discriminant analysis, quadratics discriminant analysis, learning vector quantization and support vector machines, as dependent on data structure. *Chemometrics and Intelligent Laboratory Systems*, *95*, 1-17. doi:10.1016/j.chemolab.2008.07.010

Doeswijk, T. G., Smilde, A. K., Hageman, J. A., Mesterhuis, J. A., & van Eeuwijk, F. A. (2011). On the increase of predictive performance with high-level data fusion. *Analytica Chimica Acta*, *705*, 41-47. doi:10.1016/j.aca.2011.03.025

Doeswijk, T. G., Hageman, J. A., Westerhuis, J. A., Tikunov, Y., Bovy, A., & van Eeuwijk, F. A. (2011). *Chemometrics and Intelligent Laboratory System*, *107*, 371-376. doi:10.1016/j.chemolab.2011.05.010

Domeniconi, C., & Gunopulos, D. (2008). Local feature selection for classification. In H. Liu & H. Motoda (Eds.), *Computational methods of feature selection* (pp. 211-232). Boca Raton, FL: Chapman & Hall

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, 18*, 835-843.

Dutta, R., Hines, E. L., Gardner, J. W., Udrea, D., & Boilot, P. (2003). Non-destructive egg freshness determination: an electronic nose based approach. *Measurement Science and Technology*, *14*, 190-198.

Dutta, R., Das, A., Stocks, N. G., & Morgan, D. (2006). Stochastic resonance-based electronic nose: a novel way to classify bacteria. *Sensors and Actuators B*, *115*, 17-27. doi:10.1016/j.snb.2005.08.033

Dy, J. G. (2008). Unsupervised feature selection. In H. Liu & H. Motoda (Eds.), *Computational methods of feature selection* (pp. 19-40). Boca Raton, FL: Chapman & Hall.

El Barbri, N., Llobet, E., El Bari, N., Correig, X., & Bouchikhi, B. (2008). Electronic nose based on metal oxide semiconductor sensors as an alternative technique for the spoilage classification of red meat. *Sensors*, *8*, 142-156.

Escuder-Gilabert, L. & Peris, M. (2010). Review: highlights in recent applications of electronic tongues in food analysis. *Analytica Chimica Acta*, *665*, 15-25. doi:10.1016/j.aca.2010.03.017

Esteban, J., Starr, A., Willetts, R., Hannah, P., & Bryanston-Cross, P. (2005). A review of data fusion models and architectures: towards engineering guidelines. *Neural Comput. & Applic.*, *14*, 273-281.

Faber, N. M., Mojet, J., & Poelman, A. A. M. (2003). Simple improvement of consumer fit in external preference mapping. *Food Quality and Preference*, *14*, 455-461.

Farber, O., & Kadmon, R. (2003). Assessment of alternative approaches for bioclimatic modeling with special emphasis on the Mahalanobis distance. *Ecological Modelling*, *160*, 115-130.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, *7*, 179-188.

Foithong, S., Pinngern, O., & Attachoo, B. (2012). Feature subset selection wrapper based on mutual information and rough sets. *Expert System with Applications*, *39*, 574-584. doi:10.1016/j.eswa.2011.07.048

Fraiman, R., Justel, A., & Svarc, M. (2008). Selection of variables for cluster analysis and classification rules. *Journal of the American Statistical Association*, *103(483)*, 1294-1303.

Fukunaga, K. (1990). *Introduction to Statistical Pattern Recognition*. West Lafayette, Indiana: Academic Press Inc.

García-González, D. L., & Aparicio, R. (2002). Sensors: From Biosensors to the Electronic Nose. *Grasas y Aceites*, *53, Fasc. 1*, 96-114.

Gardner, J. W., & Bartlett, P. N. (1993). Brief history of electronic nose. *Sensors and Actuator B*, *18*, 211-217.

Gardner, J. W., & Bartlett, P. N. (1999). *Electronic noses: principles and application*. Oxford University Press, Oxford.

Geisser, S. (1976). Discrimination, allocatory and separatory, linear aspects. *Classifcation and Clustering*, 301-330.

Ghasemi-Varnamkhasti, M., Mohtasebi, S. S., & Siadat, M. (2010). Biomimetic-based odor and taste sensing systems to food quality and safety characterization: an overview on basic principles and recent achievements. *Journal of Food Engineering*, *100*, 377-387. doi:10.1016/j.jfoodeng.2010.04.032

Gigli, G., Bossé, É., & Lampropoulos, G. A. (2007). An optimized architecture for classification combining data fusion and data-mining. *Information Fusion, 8*, 366-378. doi:10.1016/j.inffus.2006.02.002

Gil-Sánchez, L., Soto, J., Martínez-Máñez, R., Garcia-Breijo, J. I., & Llobet, E. (2011). A novel humid electronic nose combined with an electronic tongue for assessing deterioration of wine. *Sensors and Actuators A*, *171*, 152-158. doi:10.1016/j.sna.2011.08.006

Gimeno, O., Ansorena, D., Astiasarán, I., & Bello, J. (2000). Characterization of chorizo de pamplona: instrumental of colour and texture. *Food Chemistry*, *69*, 195-200.

Gualdrón, O., Llobet, E., Brezmes, J., Vilanova, X., & Correig, X. (2006). Coupling fast variable selection methods to neural network-based classifiers: Application to multisensor systems. *Sensors and Actuators B: Chemical*, *114*(1), 522-529.

Gunal, S., & Edizkan, R. (2008). Subspace based feature selection for pattern recognition. *Information Sciences*, *178*, 3716-3726. doi:10.1016/j.ins.2008.06.001

Gutiérrez, M., Domingo, C., Vila-Planas, J., Ipatov, A., Capdevila, F., Demming, S., …, Jiménez-Jorquera, C. (2011). Hybrid electronic tongue for the characterization and quantification of grape variety in red wines. *Sensors and Actuators B*, *156*, 695-702. doi: 10.1016/j.snb.2011.02.020

Gutierrez-Osuna, R. (2002). Pattern analysis for machine olfaction: a review. *IEEE Sensors Journal*, *2*, 189-202.

Guru, D. S., Suraj, M. G., & Manjunath, S. (2010). Fusion of covariance matrices of PCA and FLD. *Pattern Recognition Letters*, *32*(3), 432-440. doi:10.1016/j.patrec.2010.10.006

Guyon, I., Aliferis, C., & Elisseeff, A. (2008). Causal feature selection. In H. Liu & H. Motoda (Eds.), *Computational methods of feature selection* (pp. 63-85). Boca Raton, FL: Chapman & Hall

Habbema, J. D. F., & Hermans, J. (1977). Selection of variables in discriminant analysis by F-statistic and error rate. *Technometrics*, *19*(*4*), 487-493.

Hall, D. L. (1992). *Mathematical techniques in multi sensor data fusion*. Boston: Artec House Inc.

Hall., D. L., & Llinas, J. (1997). An introduction to multi sensor data fusion. *Proceedings of the IEEE*, *58*, 6-22.

Han, J., Lee, S. W., & Bien, Z. (2013). Feature subset selection using separability index matrix. *Information Sciences*, *223*, 102-118. http://dx.doi.org/10.1016/j.ins.2012.09.042

Hansen, T., Petersen, M. A., & Byrne, D. V. (2005). Sensory based quality control utilizing an electronic nose and GC_MS analyses to predict end-product quality from raw materials. *Meat Science*, *69*, 621-634.

Harper, P. R. (2005). A review and comparison of classification algorithms for medical decision making. *Health Policy*, *71*, 315-331. doi: 10.1016/j.healthpol.2004.05.002

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The element of statistical learning; data mining, inference and prediction*. New York: Springer Series in Statistics.

Hauptman, P., Borngraeber, R., Schroeder, J., & Auge, J. (2000). Artificial electronic tongue in comparison to the electronic nose – state of the art and trends. *IEEE/EIA International Frequency Control Symposium and Exhibition*, 22-29.

Hawkins, D. W. (Ed.). (1982). *Topics in applied multivariate analysis*. USA: Cambridge University Press. 17–18.

Héberger, K., & Andrade, J. M. (2004). Procrustes rotation and pair-wise correlation: a parametric and a non-parametric method for variable selection. *Croatia Chemica Acta*, *77*(1-2), 117-125.

Hidayat, W., Md. Shakaff, A. Y., Ahmad, M. N., & Adom, A. H. (2010). Classification of Agarwood Oil Using an Electronic Nose. *Sensors*, *10*, 4675-4685. doi:10.3390/s100504675

Hines, E. L., Llobet, E., & Gardner, J.W. (1999). Electronic noses: areview of signal processing techniques. *IEEE Proc-Circuits Devices Syst*. *146*(6), 297-310.

Hsu, L. M. (1989). Discriminant analysis: a comment. *Journal of Counseling Psychology*, *36*(2), 244-247.

Huang, J., Cai, Y., & Xu, X. (2007). A hybrid genetic algorithm for feature selection wrapper based on mutual information. *Pattern Recognition Letters*, *28*, 1825-1844. doi:10.1016/j.patrec.2007.05.011

Huang, C. (2009). *Data fusion in scientific data mining* (Unpublished doctoral dissertation). Rensselaer Polytechnic Institute, Troy, NY.

Huang, J. Z., Xu, J., Ng, M., & Ye, Y. (2008). Weighting method for feature selection in k-means. In H. Liu & H. Motoda (Eds.), *Computational methods of feature selection* (pp. 193-210). Boca Raton, FL: Chapman & Hall

Jain, A. K., & Waller, W. G. (1978). On the optimal number of features in the classification of multivariate Gaussian data. *Pattern recognition*, *10*(5-6), 365-374.

Jain, A, & Zongker, D. (2002). Feature selection: evaluation, application and small sample performance. IEEE Transaction on Pattern Analysis and Machine Learning, 19(2), 153-158. doi:10.1109/34.574797

Jalan, J. (2009). Feature selection, statistical modeling and its applications to universal JPEG steganalyzer (Unpublished master dissertation). Iowa State University, Ames, Iowa, USA.

Jamal, M., Khan, M. R., & Imam, S. A. (2009). Electronic tongue and their analytical application using artificial neural network approach: a review. *MASAUM Journal of Reviews and Surveys*, *1*(1), 130-137.

Jin-Jie, H., Ning, L., Shuang-Quan, L., & Yun-Ze, C. (2008). Feature selection for classificatory analysis based on information-theoretic criteria. *Acta Automatica Sinica*, *34*(3), 383-388. doi: 10.3724/SP.J.1004.2008.00383

John, S. (1960). The distribution of Walds classification statistic when the dispersion matrix is known. *Shankyā 21*, 371-376.

John, S. (1961). Errors in discrimination. *The Annals of Mathematical Statistics*, *32*, 1125-1144.

John, G. H., Khovi, R., & Pfleger, K. (1994). Irrelevant featuresand the subset selection problem. In William W. Cohen, & Haym Hirsh, (Eds.), *Machine Learning: Proceedings of the Eleventh International Conference*, 121-129.

Johnson, R. A., & Wichern, D. W. (2007). *Applied multivariate statistical analysis*. USA: Pearson Prentice Hall.

Jolliffe, I. T. (2002). *Principal Component Analysis*. 2nd Edition, New York, Inc., Springer-Verlag.

Kabir, M. M., Islam, M. M., & Murase, K. (2010). A new wrapper feature selection approach using neural network. *Neurocomputing*, *73*, 3273-3283. doi:10.1016/j.neucom.2010.04.003

Kanal, L., & Chandrasekaran, B. (1971). On dimensionality and sample size in statistical pattern classification. *Pattern Recognition*, *3*, 225-234.

Khaleghi, B., Khamis, A., Karray, F. O., & Razavi, S. N. (2012). Multisensor data fusion: a review of the state–of-the-art. *Information fusion*. doi:10.1016/j.inffus.2011.08.001

Kononenko, I., & Šikonja, M. R. (2008). Non-Myopic feature quality evaluation with (R) Relief. In H. Liu & H. Motoda (Eds.), *Computational methods of feature selection* (pp. 169-192). Boca Raton, FL: Chapman & Hall

Korel, F., & Ö. Balaban, M. (2009). Electronic nose technology in food analysis. In Ötles, Semih (Ed.) *Handbook of food analysis instruments* (pp. 365 – 374), Florida: CRC press.

Kovács, Z., Sipos, L., Szöllösi, D., Kókai, Z., Székely, G., & Fekete, A. (2011). Electronic tongue and sensory evaluation for sensing apple juice taste attributes. *Sensor Letters*, *9*, 1273-1281. doi:10.1166/sl.2011.1687

Krzanowski, W. J. (2000). *Principles of multivariate analysis: A user's perspective*. New York: Oxford University Press.

Kuhn, M., & Johnson, K. (2013). Applied predictive modeling. *Springer Science and Business Media*, New York, 487-519. doi:10.1007/978_1_4614_6849_3_19

Kumar, A., Wong, D. C. M., Shen, H. C., & Jain, A. K. (2006). Personal authentication using hand images. *Pattern Recognition Letters*, *27*, 1478-1486. doi:10.1016/j.patrec.2006.02.021

Legin, A., Rudnitskaya, A., Lvova, L., Vlasov, Y., Di Natale, C., & D"Amico, A. (2003). Evaluation of Italian wine by the electronic tongue: recognition, quantitative analysis, and correlation with human sensory perception. *Analytica Chimica Acta*, *484*, 33-44. doi:10.1016/S0003-2670(03).0031-5

Li, H., Wu, X., Li, Z., & Ding, W. (2013). Group feature selection with streaming features. *2013 IEEE 13th International Conference on Data Mining,* pp. 1109-1114. doi: 1109/ICDM.2013.137

Li, J., Luo, S., & Jin, J. S. (2010). Sensor data fusion for accurate cloud presence prediction using Dempster-Shafer evidence theory. *Sensors*, *10*, 9384-9396. doi:10.3390/s101009384

Lin, H. (2013). Feature selection based on cluster and variability analyses for ordinal multi-class classification problems. *Knowledge-Based Systems*, *37*, 94-104. http://dx.doi. Org/10.1016/j.knosys.2012.07.018

Liu, H., Motoda, H., & Yu, L. (2004). A selective sampling approach to active feature selection. *Artificial Intelligence*, *159*, 49-74. doi:10.1016/j.srtint.2004.05.009

Liu, H., & Motoda, H. (2008). Less is more. In H. Liu & H. Motoda (Eds.), *Computational methods of feature selection* (pp. 3-17). Boca Raton, FL: Chapman & Hall.

Liu, H., Sun., J., Liu, L., & Zhang, H. (2009). Feature selection with dynamic mutual information. *Pattern Recognition*, *42*, 1330-1339. doi: 10.1016/j.patcog.2008.10.028

Louw, N., & Steel, S. J. (2006). Input variable selection in Kernel Fisher discriminant analysis. In Spiliopoulou, M., Kruse, R., Borgelt, C., Nürnberg, A., Gaul, W., (Eds), *From Data and Information Analysis to Knowledge Engineering*, Springer Berlin Heidelberg, 126-133. doi: 10.1007/3-540-31314-1_14

Maji, P., & Garai, P. (2013). On fuzzy-rough attribute selection: criteria of max-dependancy, max-relevance, min-redundancy, and max-significance. *Applied Soft Computing*, *13*, 3968-3980. http://dx.doi.org/10.1016/j.asoc.2012.09.006

Mahalanobis, P. C. (1936). On the generalized distance in statistics. Proc. Natl. Inst. Sci. (India), 12, 49-55.

Mahat, N. I. (2006). *Some investigations in discriminant analysis with mixed variables* (Unpublished doctoral dissertation). University of Exeter, Devon, UK.

Maji, P., & Paul, S. (2011). Rough set based maximum relevance-maximum significance criterion and gene selection from microarray data. International Journal of Approximate Reasoning, 52, 408-426. doi.org/10.1016/j.ijar.2010.09.006

Mallet, Y., Coomans, D., & de Vel, O. (1996). Recent development in discriminant analysis on high dimensional spectral data. *Chemometrics and Intelligent Laboratory Systems*, *35*, 157-173.

Marra, G., & Wood, S. N. (2011). Practical variable selection for generalized additive models, *Computational Statistics and Data Analysis*, *55*, 2373-2387. doi:10.1016/j.csda.2011, 02004

Masnan, M. J., Mahat, N. I., Shakaff, A. Y. M., Abdullah, A. H., Zakaria, N. Z. I., Yusuf, N., ... & Aziz, A. H. A. (2015, May). Understanding Mahalanobis distance criterion for feature selection. In M. F. Ramli, & A. K. Junoh (Eds.), *AIP Conference Proceedings* (Vol. 1660, No. 1, p. 050075). AIP Publishing. doi: 10.1063/1.4915708

Masnan, M. J., Mahat, N. I., Shakaff, A. Y. M., & Abdullah, A. H. (2015, December). Sensors closeness test based on an improved [0, 1] bounded Mahalanobis distance Δ2. In *AIP Conference Proceedings* (Vol. 1691, No. 1, p. 050017). AIP Publishing.

Masnan, M. J., Zakaria, A., Shakaff, A. Y. M, Mahat, N. I., Hamid, H., Subari, N., & Saleh, J. M. (2012). Principal Component Analysis – A Realization of Classification Success in Multi Sensor Data Fusion. In Sanguansat, P., *Principal Component Analysis – Engineering Application*, Croatia; Rijeka, 1-24.

Masnan, M., Mahat, N. I ., Zakaria, A., Shakaff, A. Y. M., Adom, A. H., & Sa"ad, F. S. A. (2012). Enhancing classification performance of multisensory data through extraction and selection of features. *Procedia Chemistry*, *6*, 132-140.

McCabe, G. P. (1975). Computations for variable selection in discriminant analysis. *Technometrics*, *17*(*1*), 103-109.

McFerrin, L., & McFerrin, M. L. (2013). Package „HDMD". *Stazeno z http:// cran.r-project.org/web/packages/HDMD/HDMD. pdf (14.6. 2013)*.

McLachlan, G. J. (1992). *Discriminant analysis and statistical pattern recognition*. USA: John Wiley & Sons Inc.

Mitchell, H.B. (2007). *Multi-sensor data fusion, an introduction*. Berlin, Heidelberg: Springer.

Murray, G. D. (1977). A cautionary note on selection of variables in discriminant analysis. *Appl. Statist*., *26*(*3*), 246-250.

Nakariyakul, S., & Casasent, D. P. (2009). An improvement on floating search algorithm for feature subset selection. Pattern Recognition, 42, 1932-1940. doi:10.1016/j.patcog.2008.11.018

Olafsdottir, G., Nesvadba, P., Di Natale, C., Careche, M., Oehlenschläger, J., Tryggvadóttir, S. V., … & Jørgensen, B. M. (2004). Multisensor for fish quality determination. *Trends in Food Science & Technology*, *15*, 86-93. doi:10.1016/j.tifs.2003.08.006

Oliveri, P., Casolino, M. C., & Forina, M. (2010). Chemometric brains for artificial tongues. *Advances in Food and Nutrition Research*, *61*, 57-116. doi:10.1016/S1043-4526(10)61002-9

Pardo, M., Niederjaufner, G., Benussi, G., Comini, E., Faglia, G., Sberveglieri, … Lundstrom, I. (2000). Data preprocessing enhances the classification of different brands of Espresso coffee with an electronic nose. *Sensors and Actuators B*, *69*, 397-403.

Pardo, M., & Sberveglieri, G. (2008). Random forests and nearest shrunken centroids for the classification of sensor array data. *Sensors and Actuators B: Chemical*, *131*(1), 93-99.

Pechenizkiy, M. (2005). *Feature extraction for supervised learning in knowledge discovery systems* (Unpublished doctoral dissertation). University of Jyväskylä, Findland.

Peng, H., Long, F., & Ding, C. (2005). Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *27*(*8*), 1226-1238.

Peris, M., & Escuder-Gilabert, L. (2009). A 21st century technique for food control: electronic noses. *Analytica Chimica Acta*, *638*, 1-15. doi:10.1016/j.aca.2009.02.009

Pfeiffer, K. P. (1985). Stepwise variable selection and maximum likelihood estimation of smoothing factors of kernel functions for nonparametric discriminant functions evaluated by different criteria. *Computers and Biomedical Research*, *18*, 46-61.

Phaisanggittisagul, E. (2007). *Signal processing using wavelets for enhancing electronic nose performance* (Unpublished doctoral dissertation). North Carolina State University, Raleigh, NC.

Phaisanggittisagul, E., Nagle, H. T., & Areekul, V. (2010). Intelligent method for sensor subset selection for machine olfaction. *Sensors and Actuators B*, *145*, 507-515. doi:10.1016/j.snb.2009.12.063

Prieto, N., Gay, M., Vidal, S., Aagaard, O., de Saja, J. A., & Rodriguez-Mendez, M. L. (2011). Analysis of the influence of the type of closure in the

organoleptic characteristics of a red wine by using an electronc panel. *Food Chemistry*, *129*, 589-594. doi:10.1016/j.foodchem.2011.04.071

Pudil, P., Novovicova, J., & Kittler, J. (1994). Floating search methods in feature selection. *Pattern Recognition Letter*, *15*, 1119-1125.

Raghavendra, R, Dorizzi, B., Rao, A., & Kumar, G. H. (2011). Designing efficient fusion schemes for multimodal biometric systems using face and palmprint. *Pattern Recognition*, *44*, 1076-1088. doi:10.1016/j.patcog.2011.11.008

Rao, C. R. (1948), The utilization of multiple measurements in problems of biological classification. *J. Roy. Statist. Soc. B*, *10*, 159 – 203.

Ray, S., & Turner, L. F. (1992). Mahalanobis distance-based two new feature evaluation criteria. *Information Sciences*, *60*, 217-245.

Rencher, A. C. (2002). *Methods of multivariate analysis* (online). John Wiley & Sons, Inc. http://www3.interscience.wiley.com/cgi-bin/summary/104086842/SUMMARY?CRETRY=1&SRET...

Roberts, S. J., & Hanka, R. (1982). An interpretation of mahalanobis distance in the dual space. *Pattern Recognition*, *15(4)*, 325-333.

Rodríguez-Méndez, M. L., Apetrei, C., & De Saja, J. A. (2010). Electronic tongues purposely designed for organoleptic characterization of olive oils. *Olives and Olive Oil in Health and Disease Prevention*, *Natural Components section*, 525-532.

Rodríguez-Méndez,M. L., Arrieta, A. A., Parra, V., Bernal, A., Vegas, A., Villanueva, S., Gutiérreze-Osuna, R., & De Saja, J. A. (2004). Fusion of three sensory modalities for the multimodal characterization of red wines. *IEEE Sensors Journal*, *4*, 348-354.

Rong, L., Ping, W., & Wenlei, H. (2000). A novel method for wine analysis based on sensor fusion technique. *Sensors and Actuators B*, *66*, 246-250.

Roussel, S., Bellon-Maurel, V., Roger, J., & Grenier, P. (2003). Fusion aroma, FT-IR and UV sensor data based on the Bayesian inference. Application to the discrimination of white grape varieties. *Chemometrics and Intelligent Laboratory Systems*, *65*, 209-219.

Rudnitskaya, A., Kirsanov, D., Legin, A., Beullens, K., Lammertyn, J., Nicolai, B. M., & Irudayaraj, J. (2006). Analysis of apples varieties – comparison of electronic tongue with different analytical techniques. *Sensors and Actuators B*, *116*, 23-28. doi:10.1016/j.snb.2005.11.069

Rueda, L., Oommen, B. J., & Henriquez, C. (2010). Multi-class pairwise linear dimensionality reduction using heteroscedastic schemes. *Pattern Recognition*, *43*, 2456-2465.

Sakar, C. O., Kursun, O., & Gurgen, F. (2012). A feature selection method based on kernel canonical correlation analysis and the minimum redundancy-maximum relevance filter method. *Expert System with Applications*, *39*, 3432-3437. doi:10.1016/j.eswa.2011.09.031

Schaller, E., Bosset, J. O., & Escher, F. (1998). "Electronic noses" and their application to food. *Lebensm.-Wiss.U.-Techno., Review Article, 31*, 305-316.

Schulerud, H., & Albregtsen, F. (2004). Many are called, but few are chosen. Feature selection and error estimation in high dimensional spaces. *Computer Methods and Programs in Biomedicine*, *73*, 91-99. doi:10.1016/S0169-2607(03)00018-X

Schürmann, J. (1996). *Pattern classification: a unified view of statistical and neural approaches*. New York: Wiley.

Sewell, M. (2009). *Kernel methods*. www.svm.org/kernels/kernel-methods.pdf

Shaffer, R. E., Rose-Pehrsson, S. L., & McGill, R. A. (1999). A comparison study of chemical sensor array pattern recognition algorithms. *Analytica Chimica Acta*, *384*, 305-317.

Siedlecki, W., & Sklansky, J. (1989). A note on genetic algorithms for large-scale feature selection. Pattern Recognition Letters, 10, 335-347.

Smith, C. R., & Erickson, G. J. (1991). Multisensor data fusion: concepts and principles. *IEEE Pacific Rim Conference on Communications, Computers and Signal Processing*, 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*, *41*, 1-14.

Somol, P., Pudil, P., Novocicova, J., Paclik, P. (1999). Adaptive floating search methods in feature selection. *Pattern Recognition Letters*, *20*, 1157-1163.

Steinmetz, V., Sévila, F., & Bellon-Maurel, V. (1999). A methodology for sensor fusion design: application to fruit quality assessment. *J. agri. Engng Res., 74,* 21-31.

Stracuzzi, D. J. (2008). Randomized feature selection. In H. Liu & H. Motoda (Eds.), *Computational methods of feature selection* (pp. 3-17). Boca Raton, FL: Chapman & Hall.

Sun, Y. (2008). Feature weighting through local learning. In H. Liu & H. Motoda (Eds.), *Computational methods of feature selection* (pp. 233-254). Boca Raton, FL: Chapman & Hall

Sun, Q. S, Zeng, S. G., Liu, Y., Heng, P. A., & Xia, D. S. (2005). *Pattern Recognition, 38*, 2437-2448. doi:10.1016/j.patcog.2004.12.013

Sundic, T., Marco, S., Samitier, J., & Wide, P. (2000). Electronic tongue and electronic nose data fusion in classification with neural network and fuzzy logic based models. *Instrumentation and Measurement Technology Conference, Proceedings of the 17th IEEE, 3*, 1474-1479. doi:10.1109/IMTC.2000.848719

Tao, Q., & Veldhuis, R. (2009). Threshold-optimized decision level fusion and its application to biometrics. *Pattern Recognition, 42*, 823-836. doi:10.1016/j.patcog.2008.09.036

Thybo, A. K., Kühn, B. F., & Martens, H. (2003). Explaining Danish children‟s preferences for apples using instrumental, sensory and demographic/behavioural data. *Food Quality and Preference*, *15*, 53-63.

Toko, K. (1996). Taste sensor with global selectivity. *Materials Science and Engineering C4*, 69-82.

Toko, K. (2000). Taste sensor. *Sensor and Actuators B, 64*, 205-215.

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*, *28*, 1572-1580. doi:10.1016/j.patrec.2007.03.019

Vera, L., Aceña, L., Guash, J., Boque, R., Mestres, M., & Busto, O. (2011). Discrimination and sensory description of beers through data fusion. *Talanta*, *87,* 136-142. doi://10.1016/j.talanta.2011.09.052

Vergara, A., & Llobet, E. (2011). Feature selection and sensor array optimization in machine olfaction. In Hines, E. L., & Leeson, M. S. (Eds.), *Intelligent Systems for Machine Olfaction: Tools and Methodologies* (pp. 1-61). doi:10.4018/978-1-61520-915-6

Wang, Z., Tyo, J. S., & Hayat, M. M. (2007). Data interpretation for spectral sensors with correlated bands. *Journal of Optical Society of America*, *24*(9), 2864-2870.

Wang, S. J., Mathew, A., Chen, Y., Xi, L. F., Ma, L., & Lee, J. (2009). Empirical analysis of support vector machine ensemble classifiers. *Expert Systems with applications*, *36*(3), 6466-6476.

Wang, J., Wu, L., Kong, J., Li, Y., & Zhang, B. (2013). Maximum weight and minimum redundancy: a novel framework for feature subset selection. *Pattern Recognition*, *46*, 1616-1627. http://dx.doi.org/10.1016/j.patcog.2012.11.025

Wankhande, K., Rane, D., & Thool, R. (2013). A new feature selection algorithm for stream data classification. *IEEE 2013 International Conference on*

Advances in Computing, Communications and Informatics (ICACCI), pp. 1843-1848.

Web, A. R. (2002). *Statistical Pattern Recognition*, West Sussex, England: John Wiley and Sons Ltd.

Wei, Z., Wang, J., & Liao, W. (2009). Technique potential for classification of honey by electronic tongue. *Journal of Food Engineering*, *94*, 260-266. doi://10.1016/j.foodeng.2009.03.016

Weiner, J. M., & Dunn, O. J. (1996). Elimination of variates in linear discrimination problems. *Biometrics*, *22*(2), 268-275.

Wide, P., Winquist, F., Bergsten, P., & Petriu, E. M. (1998). The human-based multisensory fusion method for artificial nose and tongue sensor data. *IEEE Transactions on Instrumentation and Measurement*, *47*(5), 1072-1077.

Winquist, F., Lundström, I., & Wide, P. (1999). The combination of an electronic tongue and electronic nose. *Sensors and Actuators B*, *58*, 512-517.

Winquist, F., Krantz-Rülcker, C., & Lundström, I. (2003). Electronic tongues and combinations of artificial senses. In Baltes, H., Fedder, G. K., & Korvink. J. G. (Eds.), *Sensors Update* (pp. 279-306). Weinheim, Germany: Wiley VCH.

Woods, M. P. (1998). Symposium on „Taste, flavour and palatability"Taste and flavour perception. *Proceedings of the Nutrition Society*, *57*(04), 603-607.

Wood, M., Jolliffe, I. T., & Horgan, G. W. (2005). Variable selection for discriminant analysis of fish sound using matrix correlations. *Journal of Agricultural, Biological and Envronmental Statistics*, *10*(3), 321-336.

Worth, A. P., & Cronin, M. T. D. (2003). The use of discriminant analysis, logistics regression and classification tree analysis in the development of classification models for human health effects. *Journal of Molecular Structure (Theochem), 622*, 97-111.

Wu, Y., Li, M., & Liao, G. (2007). Multiple features data fusion method in color texture analysis. *Applied Mathematics and Computation*, *185*, 784-797. doi:10.1016/j.amc.2006.06.116

Xiaobo, Z., & Jiewen, Z. (2005). Apple quality assessment by fusion three sensors. *IEEE*, 389-392.

Yan, K., & Zhang, D. (2015). Feature selection and analysis on correlated gas sensor data with recursive feature elimination. *Sensors and Actuators B: Chemical*, *212*, 353-363.

Yen, C., Chen, L., & Lin, S. (2010). Unsupervised feature selection: minimize information redundancy of features. *International Conference on*

Technologies and Applications of Artificial Intelligence, 247-254. doi:10.1109/TAAI.2010.49

Yin, L., Ge, Y., Xiao, K., Wang, X., & Quan, X. (2013). Feature selection for high dimensional imbalance data. *Neurocomputing*, *105*, 3-11. http://dx.doi.org/10.1016/j.neucom.2012.04.039

Yongli, Z., Weiming, T., Yungui, Z., Hongzhi, C. (2013). An improved feature selection algorithm based on Mahalanobis distance for network intrusion detection. International Conference on Sensor Network Security Technology and Privacy Communication System (SNS & PCS), 69-73. doi:10.1109/SNS-PCS.2013.6553837

Youn, E. S. (2004). *Feature selection and discriminant analysis in data mining* (Unpublished doctoral dissertation). University of Florida, Gainesville, FL.

Young, D. M., & Odell, P. L. (1984). A formulation and comparison of two linear feature selection techniques applicable to statistical classification. *Pattern Recognition*, *17*(3), 331-337.

Zakaria, A., Shakaff, A. Y. M., Adom, A. H., Ahmad, M. N., Masnan, M. J., Aziz, A. H. A., …, & Kamarudin, L. M. (2010). Improved classification of *orthosiphon stamineus* by data fusion of electronic nose and tongue sensors. *Sensors*, *10*, 8782-8796. doi:10.3390/s101008782

Zakaria, A., Shakaff, A. Y. M., Masnan, M. J., Ahmad, M. N., Adom, A. H., Jaafar, M. N., ... & Subari, N. (2011). A biomimetic sensor for the classification of honeys of different floral origin and the detection of adulteration. *Sensors*, *11*(8), 7799-7822.

Zakaria, N. Z. I., Masnan, M. J., Zakaria, A., & Shakaff, A. Y. M. (2014). A Bio-Inspired Herbal Tea Flavour Assessment Technique. *Sensors*, *14*(*7*), 12233-12255; doi:10.3390/s140712233

Zakaria, N. Z. I., Masnan, M. J., Zakaria, A., & Shakaff, A. Y. M. (2014). A Bio-Inspired Herbal Tea Flavour Assessment Technique. *Sensors*, *14*, 12233-12255. doi:10.3390/s140712233

Zamora, M. C., & Guirao, M. (2004). Performance comparison between trained assessors and wine experts using specific sensory attributes. *Journal of Sensory Studies*, *19*, 530-545.

Zhang, D., & Yan, K. (2015). Feature selection and analysis on correlated gas sensor data with recursive feature elimination. *Sensors and Actuators B: Chemical*, *212*, 353-363.

Zhang, H., Balaban, M. Ö., & Principe, J.C. (2003). Improving pattern recognition of electronic nose data with time-delay neural network. *Sensors and Actuators B*, *96*, 385-389. doi:10.1016/S0925-4005(03)00574-4

Zhang, X., & Jia, Y. (2007). A linear discriminant analysis framework based on random subspace for face recognition. *Pattern Recognition*, *40*, 2585-2591.

Zhang, H., & Sun, G. (2002). Feature selection using tabu search method. *Pattern Recognition*, *35*, 701-711.

Zhou, B., & Wang, J. (2011). Use of electronic nose technology for identifying rice infestation by Nilaparvata lugens. *Sensors and Actuators B*, *160*, 15-21. doi:10.1016/j.snb.2011.07.002

Appendix A
DEVELOPED R ALGORITHMS FOR THE UNIVARIATE AND MULTIVARIATE MAHALANOBIS DISTANCES

A. Algorithms for fused feature ranking based on univariate unbounded Mahalanobis distance $\left(D^2\right)$

```
univariate.mahalanobisU <- function(variable, grouping)
{
  n <- nrow(variable)
  g <- as.factor(grouping)
  lev <- lev1 <- levels(g)
  counts <- as.vector(table(g))
  ng = length(lev1)
  group.mean <- aggregate(variable, by = list(groupFUN =
  "mean")
  xbargroup <- as.vector(group.mean)
  colnames(xbargroup) <- c("Group", "GroupMean")

  group.var <- aggregate(variable, by = list(grouping), FUN =
  "var") #group.var = data.frame
  vargroup <- as.vector(group.var)
  colnames(vargroup) <- c("Group", "GroupVariance")

  str(xbargroup)
  str(vargroup)

  Distance = matrix(nrow = ng, ncol = ng)
  dimnames(Distance) <- list(rownames(Distance, do.NULL =
  FALSE, prefix = "g"), colnames(Distance, do.NULL = FALSE,
  prefix = "g"))

  Means = round(xbargroup$GroupMean, digits=10)
  Variance = round(vargroup$GroupVariance digits=10)
  Distance = round(Distance, digits=3)

  for (i in 1:ng) {
    for (j in 1:ng) {
      if (i > j)

     Distance[i, j] <- ((Means[i]- Means[j])^2)*((counts[i]
     +counts[j])2) /(Variance[i]+Variance[j])
    }
  }
  return(Distance)
}
```

B. Algorithms for fused feature ranking based on univariate bounded Mahalanobis distance $\left(D_A^2\right)$

```
univariate.mahalanobisU <- function(variable, grouping)
{
  n <- nrow(variable)
  g <- as.factor(grouping)
  lev <- lev1 <- levels(g)
  counts <- as.vector(table(g))
  ng = length(lev1)
  group.mean <- aggregate(variable, by = list(groupFUN =
  "mean")
  xbargroup <- as.vector(group.mean)
  colnames(xbargroup) <- c("Group", "GroupMean")

  group.var <- aggregate(variable, by = list(grouping), FUN =
  "var") #group.var = data.frame
  vargroup <- as.vector(group.var)
  colnames(vargroup) <- c("Group", "GroupVariance")

  str(xbargroup)
  str(vargroup)

  Distance = matrix(nrow = ng, ncol = ng)
  dimnames(Distance) <- list(rownames(Distance, do.NULL =
  FALSE, prefix = "g"), colnames(Distance, do.NULL = FALSE,
  prefix = "g"))

  Means = round(xbargroup$GroupMean, digits=10)
  Variance = round(vargroup$GroupVariance digits=10)
  Distance = round(Distance, digits=3)

  for (i in 1:ng) {
    for (j in 1:ng) {
      if (i > j)

     Distance[i, j] <- ((Means[i]-Means[j])^2)*
     ((counts[i]+counts[j])-2)/(Variance[i]+ Variance[j])
     Distance[i, j] <- Distance[i, j]/(4+Distance[i, j])
     }
  }
  return(Distance)
}
```

C. Algorithms for multivariate unbounded Mahalanobis distance $\left(D^2\right)$

```
library(HDMD)
pairwise.MVmahal <- function (x, grouping, cov, inverted =
FALSE, digits = 3, ...)
{
  x <- if (is.vector(x))
    matrix(x, ncol = length(x))
  else as.matrix(x)
  if (!is.matrix(x))
    stop("x could not be forced into a matrix")
  if (length(grouping) == 0) {
    grouping = t(x[1])
    x = x[2:dim(x)[2]]
    cat("assigning grouping\n")
    print(grouping)
  }
  n <- nrow(x)
  p <- ncol(x)
  if (n != length(grouping)) {
    cat(paste("n: ", n, "and groups: ", length(grouping),
              "\n"))
    stop("nrow(x) and length(grouping) are different")
  }
  g <- as.factor(grouping)
  g
  lev <- lev1 <- levels(g)
  counts <- as.vector(table(g))
  if (any(counts == 0)) {
    empty <- lev[counts == 0]
    warning(sprintf(ngettext(length(empty), "group %s is
    empty", "groups %s are empty"), paste(empty, collapse = "
    ")), domain = NA)
    lev1 <- lev[counts > 0]
    g <- factor(g, levels = lev1)
    counts <- as.vector(table(g))
  }
  ng = length(lev1)
  group.means <- tapply(x, list(rep(g, p), col(x)), mean)
  #if (missing(cov)) {
  #if (is.null(poolcov)) {
  #inverted = FALSE
  #cov = cor(x)
  #  cov = poolcov(x)
  #}
  #else {
  #  if (dim(cov) != c(p, p))
  #    stop("cov matrix not of dim = (p,p)\n")
  #}
  Distance = matrix(nrow = ng, ncol = ng)
  dimnames(Distance) = list(names(group.means),
  names(group.means))
  Means = round(group.means, digits)
  Cov = round(cov, digits)
  Distance = round(Distance, digits)
```

```
  for (i in 1:ng) {
    Distance[i, ] = mahalanobis(group.means, group.means[i, ],
    cov, inverted)
  }
  result <- list(means = group.means, cov = cov, distance =
Distance)
  result
}
```

D. Algorithms for multivariate bounded Mahalanobis distance $\left(D_A^2\right)$

```
library(HDMD)
pairwise.MVmahal <- function (x, grouping, cov, inverted =
FALSE, digits = 3, ...)
{
  x <- if (is.vector(x))
    matrix(x, ncol = length(x))
  else as.matrix(x)
  if (!is.matrix(x))
    stop("x could not be forced into a matrix")
  if (length(grouping) == 0) {
    grouping = t(x[1])
    x = x[2:dim(x)[2]]
    cat("assigning grouping\n")
    print(grouping)
  }
  n <- nrow(x)
  p <- ncol(x)
  if (n != length(grouping)) {
    cat(paste("n: ", n, "and groups: ", length(grouping),
              "\n"))
    stop("nrow(x) and length(grouping) are different")
  }
  g <- as.factor(grouping)
  g
  lev <- lev1 <- levels(g)
  counts <- as.vector(table(g))
  if (any(counts == 0)) {
    empty <- lev[counts == 0]
    warning(sprintf(ngettext(length(empty), "group %s is
empty",
                             "groups %s are empty"),
paste(empty, collapse = " ")),
            domain = NA)
    lev1 <- lev[counts > 0]
    g <- factor(g, levels = lev1)
    counts <- as.vector(table(g))
  }
  ng = length(lev1)
  group.means <- tapply(x, list(rep(g, p), col(x)), mean)
  #if (missing(cov)) {
  #if (is.null(poolcov)) {
  #inverted = FALSE
  #cov = cor(x)
```

```
  #  cov = poolcov(x)
  #}
  #else {
  #  if (dim(cov) != c(p, p))
  #    stop("cov matrix not of dim = (p,p)\n")
  #}
  Distance = matrix(nrow = ng, ncol = ng)
  dimnames(Distance) = list(names(group.means),
names(group.means))
  Means = round(group.means, digits)
  Cov = round(cov, digits)
  Distance = round(Distance, digits)
  for (i in 1:ng) {
    Distance[i, ] = mahalanobis(group.means, group.means[i,
    ], cov, inverted)
  }
  result <- list(means = group.means, cov = cov, distance =
Distance)
  result
}
```

Appendix B
Results of Fused Feature Ranking for LLDF based on Bounded and Unbounded Mahalanobis Distances

Results of Fused Feature Ranking for LLDF based on Bounded and Unbounded Mahalanobis Distance for AS Honey

D^2				D_A^2			
Feature	**Criterion Value**	**Rank**	**Percent**	**Feature**	**Criterion Value**	**Rank**	**Percent**
N26	1937.62	1	100.00%	N15	0.9326	1	100.00%
N5	1872.11	2	97.50%	N5	0.9285	2	97.50%
N29	1815.44	3	95.10%	N23	0.9214	3	95.10%
N31	1116.44	4	92.60%	N11	0.9181	4	92.60%
N15	1086.9	5	90.20%	N29	0.914	5	90.20%
N9	1058.89	6	87.80%	N26	0.91	6	87.80%
N20	994.57	7	85.30%	N8	0.9099	7	85.30%
N16	928.61	8	82.90%	N18	0.9077	8	82.90%
N23	917.45	9	80.40%	N20	0.8984	9	80.40%
T11	888.82	10	78.00%	N2	0.8969	10	78.00%
N17	886.27	11	75.60%	N4	0.8965	11	75.60%
N13	832.86	12	73.10%	T11	0.8963	12	73.10%
N8	800.05	13	70.70%	N1	0.8916	13	70.70%
N21	779.23	14	68.20%	N3	0.8907	14	68.20%
N11	755.48	15	65.80%	N16	0.8883	15	65.80%
N18	727.66	16	63.40%	N31	0.8874	16	63.40%
N28	712.29	17	60.90%	N21	0.8834	17	60.90%
N7	626.57	18	58.50%	N9	0.882	18	58.50%
N12	616.34	19	56.00%	T2	0.8751	19	56.00%
N10	582.97	20	53.60%	N19	0.8743	20	53.60%
N1	580.04	21	51.20%	N13	0.8726	21	48.70%
N3	545.42	22	48.70%	N12	0.8726	22	48.70%
N4	516.91	23	46.30%	N14	0.8722	23	46.30%
N14	502.47	24	43.90%	N28	0.8555	24	43.90%
T2	498.87	25	41.40%	N25	0.8487	25	41.40%
N2	436.33	26	39.00%	N7	0.848	26	39.00%
N22	423.87	27	36.50%	T1	0.8386	27	36.50%
N25	405.44	28	34.10%	T3	0.8383	28	34.10%
N19	385.8	29	31.70%	T8	0.8379	29	31.70%
N27	347.33	30	29.20%	N30	0.8329	30	29.20%
N6	344.93	31	26.80%	N6	0.8129	31	26.80%
N24	337.84	32	24.30%	N27	0.8104	32	21.90%
T9	318.68	33	21.90%	N17	0.8104	33	21.90%
T7	281.94	34	19.50%	N22	0.805	34	19.50%
N30	261.39	35	17.00%	T9	0.7969	35	17.00%
T3	258.33	36	14.60%	T7	0.788	36	14.60%
T8	148.62	37	12.10%	T5	0.7791	37	12.10%
T1	132.06	38	9.70%	N24	0.7716	38	9.70%
T4	96.43	39	7.30%	N10	0.7405	39	7.30%
T5	70.97	40	4.80%	N32	0.704	40	4.80%
N32	67.79	41	2.40%	T10	0.5924	41	2.40%
T10	43.47	42	0.00%	T4	0.5177	42	0.00%

Results of Fused Feature Ranking for LLDF based on Bounded and Unbounded Mahalanobis Distance for ST Honey

D^2				D_A^2			
Feature	Criterion Value	Rank	Percent	Feature	Criterion Value	Rank	Percent
N29	12593.81	1	100.00%	N6	0.9642	1	100.00%
N5	6375.09	2	97.50%	T2	9226	2	97.50%
N23	6014.23	3	95.10%	N10*	0.9219	3	95.10%
N31	5832.96	4	92.60%	N31	0.9203	4	92.60%
N26	5555.79	5	90.20%	N29	0.9194	5	90.20%
N9	4647.56	6	87.80%	N26	0.919	6	87.80%
T2	4354.62	7	85.30%	N5	0.9106	7	85.30%
N11	3749.95	8	82.90%	N17	0.909	8	82.90%
N6	3650.01	9	80.40%	T11	0.9063	9	80.40%
N20	3259.16	10	78.00%	N23	0.9016	10	78.00%
N17	2369.79	11	75.60%	N22*	0.895	11	75.60%
N28	2264.82	12	73.10%	N20	0.89	12	73.10%
N10*	1997.47	13	70.70%	T1	0.8868	13	70.70%
N1	1778.03	14	68.20%	N18	0.8836	14	68.20%
N8	1742.91	15	65.80%	N9	0.881	15	65.80%
N18	1742.85	16	63.40%	N16	0.8804	16	63.40%
N15	1695.26	17	60.90%	N11	0.8777	17	60.90%
N16	1346.31	18	58.50%	N8	0.877	18	58.50%
N22	1207.47	19	56.00%	N15	0.8731	19	56.00%
N30	1074.31	20	53.60%	N28	0.8696	20	53.60%
N3	913.47	21	51.20%	N30	0.8599	21	51.20%
T9	889.88	22	48.70%	N19	0.8373	22	48.70%
N12	868.62	23	46.30%	N24	0.8372	23	46.30%
N13	842.9	24	43.90%	N13	0.8326	24	43.90%
N19	826.82	25	41.40%	N7	0.8323	25	41.40%
N4	818.7	26	39.00%	N21	0.8225	26	39.00%
N27	761.42	27	36.50%	N12	0.8156	27	36.50%
N2	700.4	28	34.10%	T9	0.8091	28	34.10%
N7	687.86	29	31.70%	N14	0.8053	29	31.70%
N21	682.05	30	29.20%	N25	0.8041	30	29.20%
N25	587.72	31	26.80%	N4	0.8032	31	26.80%
N24	569.58	32	24.30%	N2	0.8001	32	24.30%
T11	542.8	33	21.90%	N27	0.7924	33	21.90%
N14	510.64	34	19.50%	N3	0.7886	34	19.50%
T1	341.75	35	17.00%	T8	0.7881	35	17.00%
T8	215.54	36	14.60%	N1	0.7825	36	14.60%
T5	177.28	37	12.10%	T3	0.7425	37	12.10%
N32	142.51	38	9.70%	T5	0.742	38	9.70%
T4	92.6	39	7.30%	T10	0.7155	39	7.30%
T3	62.53	40	4.80%	T4	0.7119	40	4.80%
T7	60.66	41	2.40%	N32	0.7013	41	2.40%
T10	11.09	42	0.00%	T7	0.6736	42	0.00%

Results of Fused Feature Ranking for LLDF based on Bounded and Unbounded Mahalanobis Distance for T Honey

D^2				D_A^2			
Feature	Criterion Value	Rank	Percent	Feature	Criterion Value	Rank	Percent
T7	84016	1	100.00%	N23	0.9517	1	100.00%
T2	17337.5	2	97.50%	N6	0.9514	2	97.50%
N29	4120.59	3	95.10%	T11	0.9276	3	95.10%
N23	3493.41	4	92.60%	N8	0.9152	4	92.60%
N31	2661.3	5	90.20%	N22	0.9106	5	90.20%
N5	2235.94	6	87.80%	N10	0.9105	6	87.80%
N6	2188.87	7	85.30%	N5	0.9105	7	85.30%
N26	2043.21	8	82.90%	N18	0.9082	8	82.90%
N9	1883.41	9	80.40%	N9	0.9073	9	80.40%
N20	1156.73	10	78.00%	N26	0.9	10	78.00%
N10	1097.63	11	75.60%	N20	0.899	11	75.60%
N17	1052.48	12	73.10%	N17	0.8975	12	73.10%
T11	1037.43	13	70.70%	N19	0.8973	13	70.70%
N22	1029.37	14	68.20%	T2	0.8937	14	68.20%
N8	1020.89	15	65.80%	N28	0.8861	15	65.80%
N28	1019.52	16	63.40%	N29	0.8825	16	63.40%
N18	900.7	17	60.90%	N15	0.879	17	60.90%
N15	870.4	18	58.50%	T8	0.8762	18	58.50%
N16	802.34	19	56.00%	N16	0.8741	19	56.00%
T9	660.37	20	53.60%	N1	0.8677	20	53.60%
N11	657.82	21	51.20%	N13	0.8668	21	51.20%
N12	554.35	22	48.70%	T1	0.8653	22	48.70%
N13	544	23	46.30%	N7	0.8635	23	46.30%
N27	537.6	24	43.90%	N11	0.8607	24	43.90%
T1	513.37	25	41.40%	N21	0.8592	25	41.40%
N19	475.41	26	39.00%	N2	0.8588	26	39.00%
N7	474.97	27	36.50%	N4	0.8581	27	36.50%
N1	446.23	28	34.10%	N3	0.8504	28	34.10%
N21	440.02	29	31.70%	T9	0.8484	29	31.70%
N14	434.49	30	29.20%	N12	0.8464	30	29.20%
N30	416.91	31	26.80%	N31	0.8442	31	26.80%
N25	321.53	32	24.30%	N14	0.8426	32	24.30%
N3	318.65	33	21.90%	N27	0.8302	33	21.90%
N24	291.55	34	19.50%	N30	0.8077	34	19.50%
N2	274.43	35	17.00%	N25	0.8066	35	17.00%
N4	263.23	36	14.60%	N24	0.8002	36	14.60%
T8	185.14	37	12.10%	T4	0.7996	37	12.10%
N32	143.86	38	9.70%	T5	0.7527	38	9.70%
T3	114.05	39	7.30%	T10	0.7168	39	7.30%
T4	98.99	40	4.80%	N32	0.6727	40	4.80%
T5	90.02	41	2.40%	T3	0.5971	41	2.40%
T10	76.1	42	0.00%	T7	0.5829	42	0.00%

Results of Fused Feature Ranking for LLDF based on Bounded and Unbounded Mahalanobis Distance for T3 Honey

D^2				D_A^2			
Feature	Criterion Value	Rank	Percent	Feature	Criterion Value	Rank	Percent
N29	3688.43	1	100.00%	N6	0.9642	1	100.00%
N5	3619.54	2	97.50%	T2	0.9226	2	97.50%
N23	2725.01	3	95.10%	N10	0.9219	3	95.10%
N31	2314.98	4	92.60%	N31	0.9203	4	92.60%
N26	2284.21	5	90.20%	N29	0.9194	5	90.20%
N9	1939.9	6	87.80%	N26	0.919	6	87.80%
N10	1787.67	7	85.30%	N5	0.9106	7	85.30%
N17	1356.67	8	82.90%	N17	0.909	8	82.90%
N20	1260.21	9	80.40%	T11	0.9063	9	80.40%
N6	1076.32	10	78.00%	N23	0.9016	10	78.00%
N8	1056.54	11	75.60%	N22	0.895	11	75.60%
T2	1002.97	12	73.10%	N20	0.89	12	73.10%
N22	986.39	13	70.70%	T1	0.8868	13	70.70%
N15	959.63	14	68.20%	N18	0.8836	14	68.20%
N18	930.71	15	65.80%	N9	0.881	15	65.80%
N16	925.82	16	63.40%	N16	0.8804	16	63.40%
N28	890.07	17	60.90%	N11	0.8777	17	60.90%
N11	716.85	18	58.50%	N8	0.877	18	58.50%
N13	692.65	19	56.00%	N15	0.8731	19	56.00%
T11	685.45	20	53.60%	N28	0.8696	20	53.60%
N21	628.14	21	51.20%	N30	0.8599	21	51.20%
N12	626.89	22	48.70%	N19	0.8373	22	48.70%
N1	573.16	23	46.30%	N24	0.8372	23	46.30%
N7	512.38	24	43.90%	N13	0.8326	24	43.90%
N14	477.42	25	41.40%	N7	0.8323	25	41.40%
N19	411.58	26	39.00%	N21	0.8225	26	39.00%
N4	377.42	27	36.50%	N12	0.8156	27	36.50%
N27	371.13	28	34.10%	T9	0.8091	28	34.10%
N2	350.85	29	31.70%	N14	0.8053	29	31.70%
N25	337.76	30	29.20%	N25	0.8041	30	29.20%
N30	320.51	31	26.80%	N4	0.8032	31	26.80%
N3	300.53	32	24.30%	N2	0.8001	32	24.30%
T1	298.75	33	21.90%	N27	0.7924	33	21.90%
T9	288.86	34	19.50%	N3	0.7886	34	19.50%
N24	210.66	35	17.00%	T8	0.7881	35	17.00%
T3	179.84	36	14.60%	N1	0.7825	36	14.60%
T8	128.25	37	12.10%	T3	0.7425	37	12.10%
N32	97.12	38	9.70%	T5	0.742	38	9.70%
T7	74.44	39	7.30%	T10	0.7155	39	7.30%
T4	58.29	40	4.80%	T4	0.7119	40	4.80%
T5	54.34	41	2.40%	N32	0.7013	41	2.40%
T10	52.63	42	0.00%	T7	0.6736	42	0.00%

Results of Fused Feature Ranking for LLDF based on Bounded and Unbounded Mahalanobis Distance for TK Honey

D^2				D_A^2			
Feature	Criterion Value	Rank	Percent	Feature	Criterion Value	Rank	Percent
N23	2663.26	1	100.00%	N5	0.9366	1	100.00%
N29	2538.28	2	97.50%	N23	0.9269	2	97.50%
N5	2069.87	3	95.10%	T11	0.9262	3	95.10%
N9	2066.03	4	92.60%	N6	0.9207	4	92.60%
N26	2050.25	5	90.20%	N10	0.9166	5	90.20%
N20	1580.12	6	87.80%	N11	0.895	6	87.80%
N17	1568.36	7	85.30%	N17	0.8876	7	85.30%
N31	1553.67	8	82.90%	N29	0.8851	8	82.90%
N16	1230.39	9	80.40%	N22	0.8844	9	80.40%
N10	1189.51	10	78.00%	N18	0.8678	10	78.00%
N15	1095.07	11	75.60%	T2	0.8665	11	75.60%
N13	1062.34	12	73.10%	N15	0.8644	12	73.10%
N18	1005.69	13	70.70%	N16	0.8618	13	70.70%
N6	977.35	14	68.20%	N19	0.8608	14	68.20%
N8	916.76	15	65.80%	N12	0.8595	15	65.80%
N22	865.9	16	63.40%	T1	0.8587	16	63.40%
T11	862.7	17	60.90%	N8	0.8579	17	60.90%
N21	860.6	18	58.50%	N26	0.8507	18	58.50%
N11	840.33	19	56.00%	T8	0.8479	19	56.00%
N28	786.16	20	53.60%	N20	0.8475	20	53.60%
N12	775.97	21	51.20%	N29	0.8462	21	51.20%
T2	749.26	22	48.70%	N13	0.8439	22	48.70%
N7	685.13	23	46.30%	N25	0.8296	23	46.30%
N14	540.21	24	43.90%	N28	0.8207	24	43.90%
N1	526.24	25	41.40%	N14	0.816	25	41.40%
N25	434.22	26	39.00%	T9	0.8112	26	39.00%
N19	424.72	27	36.50%	N1	0.8102	27	36.50%
N3	396.13	28	34.10%	N7	0.81	28	34.10%
N4	360.71	29	31.70%	N21	0.8082	29	31.70%
N27	343.16	30	29.20%	N4	0.8041	30	29.20%
N2	325.63	31	26.80%	N31	0.8022	31	26.80%
T9	316.08	32	24.30%	N2	0.7969	32	24.30%
N30	266.31	33	21.90%	N27	0.7815	33	21.90%
T1	205.23	34	19.50%	N31	0.7576	34	19.50%
N24	175.27	35	17.00%	N24	0.7345	35	17.00%
T8	159.03	36	14.60%	T5	0.7164	36	14.60%
T7	121.11	37	12.10%	T10	0.7095	37	12.10%
N32	106.62	38	9.70%	T4	0.6858	38	9.70%
T10	99.81	39	7.30%	T7	0.6042	39	7.30%
T5	65.49	40	4.80%	T3	0.5951	40	4.80%
T4	45.26	41	2.40%	N30	0.5621	41	2.40%
T3	26.77	42	0.00%	N32	0.5095	42	0.00%

Results of Fused Feature Ranking for LLDF based on Bounded and Unbounded Mahalanobis Distance for TLH Honey

D^2				D_A^2			
Feature	Criterion Value	Rank	Percent	Feature	Criterion Value	Rank	Percent
N29	2865.03	1	100.00%	N6	0.9391	1	100.00%
N9	2833.27	2	97.50%	N20	0.9362	2	97.50%
N26	2526.77	3	95.10%	N9	0.9353	3	95.10%
N17	2326.74	4	92.60%	N18	0.9259	4	92.60%
N10	2151.57	5	90.20%	N22	0.9239	5	90.20%
N5	1997.04	6	87.80%	N8	0.9195	6	87.80%
N20	1986.03	7	85.30%	N26	0.9189	7	85.30%
N31	1898.18	8	82.90%	T11	0.9161	8	82.90%
N18	1889.85	9	80.40%	N28	0.9108	9	80.40%
N15	1711.96	10	78.00%	N7*	0.9087	10	78.00%
N8	1539.02	11	75.60%	N17	0.9079	11	75.60%
N22	1457.59	12	73.10%	N15*	0.905	12	73.10%
N16	1416.93	13	70.70%	N21	0.9036	13	70.70%
N23	1328.33	14	68.20%	N31	0.9032	14	68.20%
N12	1292.17	15	65.80%	N5	0.8974	15	65.80%
N13	1139.75	16	63.40%	N30	0.8969	16	63.40%
N28	1069.02	17	60.90%	N13	0.896	17	60.90%
N21	992.47	18	58.50%	N23	0.8954	18	58.50%
N27	959.76	19	56.00%	N27	0.8952	19	56.00%
N7	926.02	20	53.60%	N10	0.8944	20	53.60%
N11	900.11	21	51.20%	N12	0.8926	21	51.20%
T11	868.3	22	48.70%	N1	0.8916	22	48.70%
N1	851.84	23	46.30%	N29	0.8852	23	46.30%
N19	772.37	24	43.90%	N3	0.8846	24	43.90%
N3	751.01	25	41.40%	N19	0.8827	25	41.40%
N6	698.65	26	39.00%	N11	0.8798	26	39.00%
N25	664.92	27	36.50%	N2	0.8792	27	36.50%
N14	632.77	28	34.10%	N25	0.8678	28	34.10%
N2	578.12	29	31.70%	N16	0.8644	29	31.70%
T2	428.15	30	29.20%	N14	0.8631	30	29.20%
N4	422.11	31	26.80%	N4	0.8435	31	26.80%
T10	419.94	32	24.30%	T2	0.8404	32	24.30%
N24	378.11	33	21.90%	N24	0.8121	33	21.90%
N30	345.88	34	19.50%	T1	0.7814	34	19.50%
T5	201.75	35	17.00%	T8	0.7734	35	17.00%
T9	148.68	36	14.60%	T10	0.7712	36	14.60%
N32	98.14	37	12.10%	T5	0.7628	37	12.10%
T1	94.16	38	9.70%	N32	0.7611	38	9.70%
T8	79.59	39	7.30%	T7	0.6885	39	7.30%
T7	77.88	40	4.80%	T9	0.6707	40	4.80%
T4	27.32	41	2.40%	T3	0.6143	41	2.40%
T3	24.61	42	0.00%	T4	0.6131	42	0.00%

Results of Fused Feature Ranking for LLDF based on Bounded and Unbounded Mahalanobis Distance for TN Honey

D^2				D_A^2			
Feature	Criterion Value	Rank	Percent	Feature	Criterion Value	Rank	Percent
T3	5306	1	100.00%	N22	0.9255	1	100.00%
N23	3747.14	2	97.50%	T11	0.9229	2	97.50%
N29	2640.31	3	95.10%	N29	0.9216	3	95.10%
N5	2182.42	4	92.60%	N10	0.9156	4	92.60%
N9	2025	5	90.20%	N26	0.9137	5	90.20%
N26	1770.22	6	87.80%	N5	0.9082	6	87.80%
N31	1473.64	7	85.30%	N17	0.9033	7	85.30%
N20	1411.19	8	82.90%	N18	0.8995	8	82.90%
N10	1326	9	80.40%	N19	0.8956	9	80.40%
N6	1209.44	10	78.00%	N9	0.8902	10	78.00%
N8	1206.69	11	75.60%	T2	0.887	11	75.60%
T11	1153.05	12	73.10%	N28	0.8814	12	73.10%
N22	1134.53	13	70.70%	N6	0.8779	13	70.70%
N17	1110.75	14	68.20%	N20	0.8706	14	68.20%
N18	1110.75	15	65.80%	N31	0.8698	15	65.80%
N11	1108.25	16	63.40%	N21	0.8694	16	63.40%
N15	1016.42	17	60.90%	N16	0.8664	17	60.90%
N1	913.06	18	58.50%	N12	0.8648	18	58.50%
N13	901.48	19	56.00%	N7	0.8647	19	56.00%
N16	896.82	20	53.60%	N23	0.8644	20	53.60%
N12	800.97	21	51.20%	N27	0.863	21	51.20%
N28	790.66	22	48.70%	N15	0.8628	22	48.70%
N3	782.48	23	46.30%	N8	0.8603	23	46.30%
N21	692.11	24	43.90%	N2	0.8553	24	43.90%
N7	646.58	25	41.40%	N3	0.8544	25	41.40%
N27	628.73	26	39.00%	N1	0.8528	26	39.00%
N25	609.36	27	36.50%	N30	0.8527	27	36.50%
N2	595.51	28	34.10%	N25	0.8475	28	34.10%
N19	573.71	29	31.70%	N14	0.8446	29	31.70%
N4	544.58	30	29.20%	N13	0.8438	30	29.20%
N14	530.46	31	26.80%	N4	0.8262	31	26.80%
T2	492.3	32	24.30%	T1	0.8181	32	24.30%
N24	331.69	33	21.90%	N11	0.8163	33	21.90%
N30	257.81	34	19.50%	N24	0.7868	34	19.50%
T1	142.21	35	17.00%	N32	0.7464	35	17.00%
T7	115.22	36	14.60%	T7	0.7394	36	14.60%
N32	107.78	37	12.10%	T8	0.7383	37	12.10%
T8	100.75	38	9.70%	T9	0.7261	38	9.70%
T10	89.38	39	7.30%	T3	0.7189	39	7.30%
T5	62.51	40	4.80%	T5	0.706	40	4.80%
T4	46.21	41	2.40%	T10	0.6816	41	2.40%
T9	42.27	42	0.00%	T4	0.6441	42	0.00%

Results of Fused Feature Ranking for LLDF based on Bounded and Unbounded Mahalanobis Distance for WT Honey

D^2				D_A^2			
Feature	Criterion Value	Rank	Percent	Feature	Criterion Value	Rank	Percent
N5	4127.81	1	100.00%	N23	0.9408	1	100.00%
N29	4037.63	2	97.50%	N28	0.9369	2	97.50%
N23	3405.52	3	95.10%	N20	0.9356	3	95.10%
N26	3347.08	4	92.60%	N31	0.9352	4	92.60%
N9	2942.97	5	90.20%	N6	0.9344	5	90.20%
N20	2461.75	6	87.80%	N26	0.9315	6	87.80%
N31	2299.67	7	85.30%	N5	0.9306	7	85.30%
N6	2016.16	8	82.90%	N10	0.9272	8	82.90%
N11	1739.11	9	80.40%	N29	0.9247	9	80.40%
N17	1666.27	10	78.00%	N9	0.9199	10	78.00%
N28	1488.67	11	75.60%	N22	0.914	11	75.60%
N10	1444.13	12	73.10%	N8	0.9127	12	73.10%
N15	1443.47	13	70.70%	N15	0.9104	13	70.70%
N8	1416.52	14	68.20%	N16	0.9085	14	68.20%
N16	1315.33	15	65.80%	N11	0.9052	15	65.80%
N18	1259.19	16	63.40%	N27	0.9038	16	63.40%
T11	1070.54	17	60.90%	N25	0.9016	17	60.90%
T2	1016.2	18	58.50%	N12	0.9006	18	58.50%
N12	1012.62	19	56.00%	N31	0.8956	19	56.00%
N1	899.3	20	53.60%	N18	0.8947	20	53.60%
N22	881.08	21	51.20%	T11	0.8947	21	51.20%
N3	834.49	22	48.70%	N17	0.8915	22	48.70%
N13	831.93	23	46.30%	N1	0.8898	23	46.30%
N7	809.65	24	43.90%	N30	0.8803	24	43.90%
N27	732.23	25	41.40%	N14	0.8796	25	41.40%
N30	723.78	26	39.00%	N4	0.8778	26	39.00%
N21	719.4	27	36.50%	N7	0.8732	27	36.50%
N4	668.77	28	34.10%	N2	0.8725	28	34.10%
N19	631.36	29	31.70%	N13	0.8683	29	31.70%
N14	588.74	30	29.20%	N19	0.8616	30	29.20%
N2	542.75	31	26.80%	N21	0.8595	31	26.80%
N25	471.72	32	24.30%	T2	0.8198	32	24.30%
T10	374.6	33	21.90%	N24	0.8061	33	21.90%
N24	296.97	34	19.50%	T10	0.7894	34	19.50%
T7	263	35	17.00%	T1	0.778	35	17.00%
N32	154.08	36	14.60%	T8	0.777	36	14.60%
T1	132.66	37	12.10%	T9	0.763	37	12.10%
T3	104.01	38	9.70%	N32	0.7011	38	9.70%
T5	70.1	39	7.30%	T5	0.6943	39	7.30%
T8	63.53	40	4.80%	T3	0.6848	40	4.80%
T4	43.16	41	2.40%	T4	0.6747	41	2.40%
T9	36.63	42	0.00%	T7	0.6743	42	0.00%

Results of Fused Feature Ranking for LLDF based on Bounded and Unbounded Mahalanobis Distance for YB Honey

D^2				D_A^2			
Feature	Criterion Value	Rank	Percent	Feature	Criterion Value	Rank	Percent
N5	6474	1	100.00%	N11	0.9467	1	100.00%
N23	5287.73	2	97.50%	N28	0.9459	2	97.50%
N29	4512.24	3	95.10%	N9	0.9406	3	95.10%
N26	3137.66	4	92.60%	N26	0.9354	4	92.60%
N6	3085.16	5	90.20%	N20	0.9296	5	90.20%
N9	2511.35	6	87.80%	N6	0.9218	6	87.80%
N31	2324.08	7	85.30%	N23	0.9216	7	85.30%
N20	1784.21	8	82.90%	N17	0.9216	8	82.90%
N11	1685.89	9	80.40%	N5	0.921	9	80.40%
N17	1329.32	10	78.00%	T11	0.9134	10	78.00%
N28	1278.17	11	75.60%	N25	0.911	11	75.60%
N8	1072.09	12	73.10%	N10	0.9088	12	73.10%
N10	1015.58	13	70.70%	N8	0.9075	13	70.70%
T11	960.45	14	68.20%	N18	0.9061	14	68.20%
N16	941.06	15	65.80%	N15	0.904	15	65.80%
N22	915.22	16	63.40%	N27	0.8891	16	63.40%
N15	889.7	17	60.90%	N12	0.8866	17	60.90%
N18	880.81	18	58.50%	N30	0.8852	18	58.50%
N30	856.89	19	56.00%	N16	0.8807	19	56.00%
N21	689.68	20	53.60%	N13	0.8799	20	53.60%
N13	679.78	21	51.20%	N7	0.8794	21	51.20%
N12	639.55	22	48.70%	N22	0.8742	22	48.70%
N7	517.44	23	46.30%	N14	0.8728	23	46.30%
N4	493.31	24	43.90%	N19	0.8703	24	43.90%
N1	484.69	25	41.40%	N21	0.8695	25	41.40%
N14	461.52	26	39.00%	N1	0.8686	26	39.00%
N25	447.94	27	36.50%	N31	0.8682	27	36.50%
N27	399.56	28	34.10%	N29	0.8596	28	34.10%
N3	369.81	29	31.70%	N4	0.8577	29	31.70%
N19	339.41	30	29.20%	N3	0.848	30	29.20%
N2	313.8	31	26.80%	N2	0.8372	31	26.80%
T2	218.66	32	24.30%	T1	0.8312	32	24.30%
N24	209.73	33	21.90%	N24	0.8243	33	21.90%
T1	202.2	34	19.50%	T2	0.8204	34	19.50%
T10	164.84	35	17.00%	N32	0.7648	35	17.00%
N32	108.32	36	14.60%	T9	0.7566	36	14.60%
T7	85.19	37	12.10%	T10	0.7477	37	12.10%
T3	57.36	38	9.70%	T8	0.7058	38	9.70%
T8	46.07	39	7.30%	T7	0.6164	39	7.30%
T9	37.04	40	4.80%	T3	0.6067	40	4.80%
T5	32.67	41	2.40%	T4	0.6042	41	2.40%
T4	22.2	42	0.00%	T5	0.5806	42	0.00%

Appendix C
Results of Single Feature Ranking for ILDF based on Bounded and Unbounded Mahalanobis Distances

Results of Single Feature Ranking for ILDF based on Bounded and Unbounded Mahalanobis Distance for e-nose (AS honey)

D^2				D_A^2			
Feature	**Criterion Value**	**Rank**	**Percent**	**Feature**	**Criterion Value**	**Rank**	**Percent**
N26	1,937.62	1	100.00%	N15	0.9326	1	100.00%
N5	1,872.11	2	96.70%	N5	0.9285	2	96.70%
N29	1,815.44	3	93.50%	N23	0.9214	3	93.50%
N31	1,116.44	4	90.30%	N11	0.9181	4	90.30%
N15	1,086.90	5	87.00%	N29	0.914	5	87.00%
N9	1,058.89	6	83.80%	N26	0.91	6	83.80%
N20	994.57	7	80.60%	N8	0.9099	7	80.60%
N16	928.61	8	77.40%	N18	0.9077	8	77.40%
N23	917.45	9	74.10%	N20	0.8984	9	74.10%
N17	886.27	10	70.90%	N2	0.8969	10	70.90%
N13	832.86	11	67.70%	N4	0.8965	11	67.70%
N8	800.05	12	64.50%	N1	0.8916	12	64.50%
N21	779.23	13	61.20%	N3	0.8907	13	61.20%
N11	755.48	14	58.00%	N16	0.8883	14	58.00%
N18	727.66	15	54.80%	N31	0.8874	15	54.80%
N28	712.29	16	51.60%	N21	0.8834	16	51.60%
N7	626.57	17	48.30%	N9	0.882	17	48.30%
N12	616.34	18	45.10%	N19	0.8743	18	45.10%
N10	582.97	19	41.90%	N12	0.8726	19	41.90%
N1	580.04	20	38.70%	N13	0.8726	20	38.70%
N3	545.42	21	35.40%	N14	0.8722	21	35.40%
N4	516.91	22	32.20%	N28	0.8555	22	32.20%
N14	502.47	23	29.00%	N25	0.8487	23	29.00%
N2	436.33	24	25.80%	N7	0.848	24	25.80%
N22	423.87	25	22.50%	N30	0.8329	25	22.50%
N25	405.44	26	19.30%	N6	0.8129	26	19.30%
N19	385.80	27	16.10%	N17	0.8104	27	16.10%
N27	347.33	28	12.90%	N27	0.8104	28	12.90%
N6	344.93	29	9.60%	N22	0.805	29	9.60%
N24	337.84	30	6.40%	N24	0.7716	30	6.40%
N30	261.39	31	3.20%	N10	0.7405	31	3.20%
N32	67.79	32	0.00%	N32	0.704	32	0.00%

Results of Feature Ranking for ILDF based on Bounded and Unbounded Mahalanobis Distance for e-tongue (AS honey)

D^2				D_A^2			
Feature	**Criterion Value**	**Rank**	**Percent**	**Feature**	**Criterion Value**	**Rank**	**Percent**
T11	888.82	1	100.00%	T11	0.8963	1	100.00%
T2	498.87	2	88.80%	T2	0.8751	2	88.80%
T9	318.68	3	77.70%	T1	0.8386	3	77.70%
T7	281.94	4	66.60%	T3	0.8383	4	66.60%
T3	258.33	5	55.50%	T8	0.8379	5	55.50%
T8	148.62	6	44.40%	T9	0.7969	6	44.40%
T1	132.06	7	33.30%	T7	0.788	7	33.30%
T4	96.43	8	22.20%	T5	0.7791	8	22.20%
T5	70.97	9	11.10%	T10	0.5924	9	11.10%
T10	43.47	10	0.00%	T4	0.5177	10	0.00%

Results of Feature Ranking for ILDF based on Bounded and Unbounded Mahalanobis Distance for e-nose (ST honey)

D^2				D_A^2			
Feature	**Criterion Value**	**Rank**	**Percent**	**Feature**	**Criterion Value**	**Rank**	**Percent**
N29	12,593.81	1	100.00%	N6	0.9642	1	100.00%
N5	6,375.09	2	96.70%	N10*	0.9219	2	96.70%
N23	6,014.23	3	93.50%	N31	0.9203	3	93.50%
N31	5,832.96	4	90.30%	N29	0.9194	4	90.30%
N26	5,555.79	5	87.00%	N26	0.919	5	87.00%
N9	4,647.56	6	83.80%	N5	0.9106	6	83.80%
N11	3,749.95	7	80.60%	N17	0.909	7	80.60%
N6	3,650.01	8	77.40%	N23	0.9016	8	77.40%
N20	3,259.16	9	74.10%	N22	0.895	9	74.10%
N17	2,369.79	10	70.90%	N20	0.89	10	70.90%
N28	2,264.82	11	67.70%	N18	0.8836	11	67.70%
N10	1,997.47	12	64.50%	N9	0.881	12	64.50%
N1	1,778.03	13	61.20%	N16	0.8804	13	61.20%
N8	1,742.91	14	58.00%	N11	0.8777	14	58.00%
N18	1,742.85	15	54.80%	N8	0.877	15	54.80%
N15	1,695.26	16	51.60%	N15	0.8731	16	51.60%
N16	1,346.31	17	48.30%	N28	0.8696	17	48.30%
N22	1,207.47	18	45.10%	N30	0.8599	18	45.10%
N30	1,074.31	19	41.90%	N19	0.8373	19	41.90%
N3	913.47	20	38.70%	N24	0.8372	20	38.70%
N12	868.62	21	35.40%	N13	0.8326	21	35.40%
N13	842.90	22	32.20%	N7	0.8323	22	32.20%
N19	826.82	23	29.00%	N21	0.8225	23	29.00%
N4	818.70	24	25.80%	N12	0.8156	24	25.80%
N27	761.42	25	22.50%	N14	0.8053	25	22.50%
N2	700.40	26	19.30%	N25	0.8041	26	19.30%
N7	687.86	27	16.10%	N4	0.8032	27	16.10%
N21	682.05	28	12.90%	N2	0.8001	28	12.90%
N25	587.72	29	9.60%	N27	0.7924	29	9.60%
N24	569.58	30	6.40%	N3	0.7886	30	6.40%
N14	510.64	31	3.20%	N1	0.7825	31	3.20%
N32	142.51	32	0.00%	N32	0.7013	32	0.00%

Results of Feature Ranking for ILDF based on Bounded and Unbounded Mahalanobis Distance for e-tongue (ST honey)

D^2				D_A^2			
Feature	**Criterion Value**	**Rank**	**Percent**	**Feature**	**Criterion Value**	**Rank**	**Percent**
T2	4,354.62	1	100.00%	T2	9226	1	100.00%
T9	889.88	2	88.80%	T11	0.9063	2	88.80%
T11	542.80	3	77.70%	T1	0.8868	3	77.70%
T1	341.75	4	66.60%	T9	0.8091	4	66.60%
T8	215.54	5	55.50%	T8	0.7881	5	55.50%
T5	177.28	6	44.40%	T3	0.7425	6	44.40%
T4	92.60	7	33.30%	T5	0.742	7	33.30%
T3	62.53	8	22.20%	T10	0.7155	8	22.20%
T7	60.66	9	11.10%	T4	0.7119	9	11.10%
T10	11.09	10	0.00%	T7	0.6736	10	0.00%

Results of Feature Ranking for ILDF based on Bounded and Unbounded Mahalanobis Distance for e-nose (T honey)

D^2				D_A^2			
Feature	**Criterion Value**	**Rank**	**Percent**	**Feature**	**Criterion Value**	**Rank**	**Percent**
N29	4,120.59	1	100.00%	N23	0.9517	1	100.00%
N23	3,493.41	2	96.70%	N6	0.9514	2	96.70%
N31	2,661.30	3	93.50%	N8	0.9152	3	93.50%
N5	2,235.94	4	90.30%	N22	0.9106	4	90.30%
N6	2,188.87	5	87.00%	N10	0.9105	5	87.00%
N26	2,043.21	6	83.80%	N5	0.9105	6	83.80%
N9	1,883.41	7	80.60%	N18	0.9082	7	80.60%
N20	1,156.73	8	77.40%	N9	0.9073	8	77.40%
N10	1,097.63	9	74.10%	N26	0.9	9	74.10%
N17	1,052.48	10	70.90%	N20	0.899	10	70.90%
N22	1,029.37	11	67.70%	N17	0.8975	11	67.70%
N8	1,020.89	12	64.50%	N19	0.8973	12	64.50%
N28	1,019.52	13	61.20%	N28	0.8861	13	61.20%
N18	900.70	14	58.00%	N29	0.8825	14	58.00%
N15	870.40	15	54.80%	N15	0.879	15	54.80%
N16	802.34	16	51.60%	N16	0.8741	16	51.60%
N11	657.82	17	48.30%	N1	0.8677	17	48.30%
N12	554.35	18	45.10%	N13	0.8668	18	45.10%
N13	544.00	19	41.90%	N7	0.8635	19	41.90%
N27	537.60	20	38.70%	N11	0.8607	20	38.70%
N19	475.41	21	35.40%	N21	0.8592	21	35.40%
N7	474.97	22	32.20%	N2	0.8588	22	32.20%
N1	446.23	23	29.00%	N4	0.8581	23	29.00%
N21	440.02	24	25.80%	N3	0.8504	24	25.80%
N14	434.49	25	22.50%	N12	0.8464	25	22.50%
N30	416.91	26	19.30%	N31	0.8442	26	19.30%
N25	321.53	27	16.10%	N14	0.8426	27	16.10%
N3	318.65	28	12.90%	N27	0.8302	28	12.90%
N24	291.55	29	9.60%	N30	0.8077	29	9.60%
N2	274.43	30	6.40%	N25	0.8066	30	6.40%
N4	263.23	31	3.20%	N24	0.8002	31	3.20%
N32	143.86	32	0.00%	N32	0.6727	32	0.00%

Results of Feature Ranking for ILDF based on Bounded and Unbounded Mahalanobis Distance for e-tongue (T honey)

D^2				D_A^2			
Feature	**Criterion Value**	**Rank**	**Percent**	**Feature**	**Criterion Value**	**Rank**	**Percent**
T7	84,016.00	1	100.00%	T11	0.9276	1	100.00%
T2	17,337.50	2	88.80%	T2	0.8937	2	88.80%
T11	1,037.43	3	77.70%	T8	0.8762	3	77.70%
T9	660.37	4	66.60%	T1	0.8653	4	66.60%
T1	513.37	5	55.50%	T9	0.8484	5	55.50%
T8	185.14	6	44.40%	T4	0.7996	6	44.40%
T3	114.05	7	33.30%	T5	0.7527	7	33.30%
T4	98.99	8	22.20%	T10	0.7168	8	22.20%
T5	90.02	9	11.10%	T3	0.5971	9	11.10%
T10	76.10	10	0.00%	T7	0.5829	10	0.00%

Results of Feature Ranking for ILDF based on Bounded and Unbounded Mahalanobis Distance for e-nose (T3 honey)

D^2				D_A^2			
Feature	**Criterion Value**	**Rank**	**Percent**	**Feature**	**Criterion Value**	**Rank**	**Percent**
N29	3,688.43	1	100.00%	N6	0.9642	1	100.00%
N5	3,619.54	2	96.70%	N10	0.9219	2	96.70%
N23	2,725.01	3	93.50%	N31	0.9203	3	93.50%
N31	2,314.98	4	90.30%	N29	0.9194	4	90.30%
N26	2,284.21	5	87.00%	N26	0.919	5	87.00%
N9	1,939.90	6	83.80%	N5	0.9106	6	83.80%
N10	1,787.67	7	80.60%	N17	0.909	7	80.60%
N17	1,356.67	8	77.40%	N23	0.9016	8	77.40%
N20	1,260.21	9	74.10%	N22	0.895	9	74.10%
N6	1,076.32	10	70.90%	N20	0.89	10	70.90%
N8	1,056.54	11	67.70%	N18	0.8836	11	67.70%
N22	986.39	12	64.50%	N9	0.881	12	64.50%
N15	959.63	13	61.20%	N16	0.8804	13	61.20%
N18	930.71	14	58.00%	N11	0.8777	14	58.00%
N16	925.82	15	54.80%	N8	0.877	15	54.80%
N28	890.07	16	51.60%	N15	0.8731	16	51.60%
N11	716.85	17	48.30%	N28	0.8696	17	48.30%
N13	692.65	18	45.10%	N30	0.8599	18	45.10%
N21	628.14	19	41.90%	N19	0.8373	19	41.90%
N12	626.89	20	38.70%	N24	0.8372	20	38.70%
N1	573.16	21	35.40%	N13	0.8326	21	35.40%
N7	512.38	22	32.20%	N7	0.8323	22	32.20%
N14	477.42	23	29.00%	N21	0.8225	23	29.00%
N19	411.58	24	25.80%	N12	0.8156	24	25.80%
N4	377.42	25	22.50%	N14	0.8053	25	22.50%
N27	371.13	26	19.30%	N25	0.8041	26	19.30%
N2	350.85	27	16.10%	N4	0.8032	27	16.10%
N25	337.76	28	12.90%	N2	0.8001	28	12.90%
N30	320.51	29	9.60%	N27	0.7924	29	9.60%
N3	300.53	30	6.40%	N3	0.7886	30	6.40%
N24	210.66	31	3.20%	N1	0.7825	31	3.20%
N32	97.12	32	0.00%	N32	0.7013	32	0.00%

Results of Feature Ranking for ILDF based on Bounded and Unbounded Mahalanobis Distance for e-tongue (T3 honey)

D^2				D_A^2			
Feature	**Criterion Value**	**Rank**	**Percent**	**Feature**	**Criterion Value**	**Rank**	**Percent**
T2	1,002.97	1	100.00%	T2	0.9226	1	100.00%
T11	685.45	2	88.80%	T11	0.9063	2	88.80%
T1	298.75	3	77.70%	T1	0.8868	3	77.70%
T9	288.86	4	66.60%	T9	0.8091	4	66.60%
T3	179.84	5	55.50%	T8	0.7881	5	55.50%
T8	128.25	6	44.40%	T3	0.7425	6	44.40%
T7	74.44	7	33.30%	T5	0.742	7	33.30%
T4	58.29	8	22.20%	T10	0.7155	8	22.20%
T5	54.34	9	11.10%	T4	0.7119	9	11.10%
T10	52.63	10	0.00%	T7	0.6736	10	0.00%

Results of Feature Ranking for ILDF based on Bounded and Unbounded Mahalanobis Distance for e-nose (TK honey)

D^2				D_A^2			
Feature	**Criterion Value**	**Rank**	**Percent**	**Feature**	**Criterion Value**	**Rank**	**Percent**
N23	2,663.26	1	100.00%	N5	0.9366	1	100.00%
N29	2,538.28	2	96.70%	N23	0.9269	2	96.70%
N5	2,069.87	3	93.50%	N6	0.9207	3	93.50%
N9	2,066.03	4	90.30%	N10	0.9166	4	90.30%
N26	2,050.25	5	87.00%	N11	0.895	5	87.00%
N20	1,580.12	6	83.80%	N17	0.8876	6	83.80%
N17	1,568.36	7	80.60%	N29	0.8851	7	80.60%
N31	1,553.67	8	77.40%	N22	0.8844	8	77.40%
N16	1,230.39	9	74.10%	N18	0.8678	9	74.10%
N10	1,189.51	10	70.90%	N15	0.8644	10	70.90%
N15	1,095.07	11	67.70%	N16	0.8618	11	67.70%
N13	1,062.34	12	64.50%	N19	0.8608	12	64.50%
N18	1,005.69	13	61.20%	N12	0.8595	13	61.20%
N6	977.35	14	58.00%	N8	0.8579	14	58.00%
N8	916.76	15	54.80%	N26	0.8507	15	54.80%
N22	865.90	16	51.60%	N20	0.8475	16	51.60%
N21	860.60	17	48.30%	N29	0.8462	17	48.30%
N11	840.33	18	45.10%	N13	0.8439	18	45.10%
N28	786.16	19	41.90%	N25	0.8296	19	41.90%
N12	775.97	20	38.70%	N28	0.8207	20	38.70%
N7	685.13	21	35.40%	N14	0.816	21	35.40%
N14	540.21	22	32.20%	N1	0.8102	22	32.20%
N1	526.24	23	29.00%	N7	0.81	23	29.00%
N25	434.22	24	25.80%	N21	0.8082	24	25.80%
N19	424.72	25	22.50%	N4	0.8041	25	22.50%
N3	396.13	26	19.30%	N31	0.8022	26	19.30%
N4	360.71	27	16.10%	N2	0.7969	27	16.10%
N27	343.16	28	12.90%	N27	0.7815	28	12.90%
N2	325.63	29	9.60%	N31	0.7576	29	9.60%
N30	266.31	30	6.40%	N24	0.7345	30	6.40%
N24	175.27	31	3.20%	N30	0.5621	31	3.20%
N32	106.62	32	0.00%	N32	0.5095	32	0.00%

Results of Feature Ranking for ILDF based on Bounded and Unbounded Mahalanobis Distance for e-tongue (TK honey)

D^2				D_A^2			
Feature	**Criterion Value**	**Rank**	**Percent**	**Feature**	**Criterion Value**	**Rank**	**Percent**
T11	862.70	1	100.00%	T11	0.9262	1	100.00%
T2	749.26	2	88.80%	T2	0.8665	2	88.80%
T9	316.08	3	77.70%	T1	0.8587	3	77.70%
T1	205.23	4	66.60%	T8	0.8479	4	66.60%
T8	159.03	5	55.50%	T9	0.8112	5	55.50%
T7	121.11	6	44.40%	T5	0.7164	6	44.40%
T10	99.81	7	33.30%	T10	0.7095	7	33.30%
T5	65.49	8	22.20%	T4	0.6858	8	22.20%
T4	45.26	9	11.10%	T7	0.6042	9	11.10%
T3	26.77	10	0.00%	T3	0.5951	10	0.00%

Results of Feature Ranking for ILDF based on Bounded and Unbounded Mahalanobis Distance for e-nose (TLH honey)

D^2				D_A^2			
Feature	**Criterion Value**	**Rank**	**Percent**	**Feature**	**Criterion Value**	**Rank**	**Percent**
N29	2,865.03	1	100.00%	N6	0.9391	1	100.00%
N9	2,833.27	2	96.70%	N20	0.9362	2	96.70%
N26	2,526.77	3	93.50%	N9	0.9353	3	93.50%
N17	2,326.74	4	90.30%	N18	0.9259	4	90.30%
N10	2,151.57	5	87.00%	N22	0.9239	5	87.00%
N5	1,997.04	6	83.80%	N8	0.9195	6	83.80%
N20	1,986.03	7	80.60%	N26	0.9189	7	80.60%
N31	1,898.18	8	77.40%	N28	0.9108	8	77.40%
N18	1,889.85	9	74.10%	N7	0.9087	9	74.10%
N15	1,711.96	10	70.90%	N17	0.9079	10	70.90%
N8	1,539.02	11	67.70%	N15	0.905	11	67.70%
N22	1,457.59	12	64.50%	N21	0.9036	12	64.50%
N16	1,416.93	13	61.20%	N31	0.9032	13	61.20%
N23	1,328.33	14	58.00%	N5	0.8974	14	58.00%
N12	1,292.17	15	54.80%	N30	0.8969	15	54.80%
N13	1,139.75	16	51.60%	N13	0.896	16	51.60%
N28	1,069.02	17	48.30%	N23	0.8954	17	48.30%
N21	992.47	18	45.10%	N27	0.8952	18	45.10%
N27	959.76	19	41.90%	N10	0.8944	19	41.90%
N7	926.02	20	38.70%	N12	0.8926	20	38.70%
N11	900.11	21	35.40%	N1	0.8916	21	35.40%
N1	851.84	22	32.20%	N29	0.8852	22	32.20%
N19	772.37	23	29.00%	N3	0.8846	23	29.00%
N3	751.01	24	25.80%	N19	0.8827	24	25.80%
N6	698.65	25	22.50%	N11	0.8798	25	22.50%
N25	664.92	26	19.30%	N2	0.8792	26	19.30%
N14	632.77	27	16.10%	N25	0.8678	27	16.10%
N2	578.12	28	12.90%	N16	0.8644	28	12.90%
N4	422.11	29	9.60%	N14	0.8631	29	9.60%
N24	378.11	30	6.40%	N4	0.8435	30	6.40%
N30	345.88	31	3.20%	N24	0.8121	31	3.20%
N32	98.14	32	0.00%	N32	0.7611	32	0.00%

Results of Feature Ranking for ILDF based on Bounded and Unbounded Mahalanobis Distance for e-tongue (TLH honey)

D^2				D_A^2			
Feature	**Criterion Value**	**Rank**	**Percent**	**Feature**	**Criterion Value**	**Rank**	**Percent**
T11	868.30	1	100.00%	T11	0.9161	1	100.00%
T2	428.15	2	88.80%	T2	0.8404	2	88.80%
T10	419.94	3	77.70%	T1	0.7814	3	77.70%
T5	201.75	4	66.60%	T8	0.7734	4	66.60%
T9	148.68	5	55.50%	T10	0.7712	5	55.50%
T1	94.16	6	44.40%	T5	0.7628	6	44.40%
T8	79.59	7	33.30%	T7	0.6885	7	33.30%
T7	77.88	8	22.20%	T9	0.6707	8	22.20%
T4	27.32	9	11.10%	T3	0.6143	9	11.10%
T3	24.61	10	0.00%	T4	0.6131	10	0.00%

Results of Feature Ranking for ILDF based on Bounded and Unbounded Mahalanobis Distance for e-nose (TN honey)

D^2				D_A^2			
Feature	**Criterion Value**	**Rank**	**Percent**	**Feature**	**Criterion Value**	**Rank**	**Percent**
N23	3,747.14	1	100.00%	N22	0.9255	1	100.00%
N29	2,640.31	2	96.70%	N29	0.9216	2	96.70%
N5	2,182.42	3	93.50%	N10	0.9156	3	93.50%
N9	2,025.00	4	90.30%	N26	0.9137	4	90.30%
N26	1,770.22	5	87.00%	N5	0.9082	5	87.00%
N31	1,473.64	6	83.80%	N17	0.9033	6	83.80%
N20	1,411.19	7	80.60%	N18	0.8995	7	80.60%
N10	1,326.00	8	77.40%	N19	0.8956	8	77.40%
N6	1,209.44	9	74.10%	N9	0.8902	9	74.10%
N8	1,206.69	10	70.90%	N28	0.8814	10	70.90%
N22	1,134.53	11	67.70%	N6	0.8779	11	67.70%
N17	1,110.75	12	64.50%	N20	0.8706	12	64.50%
N18	1,110.75	13	61.20%	N31	0.8698	13	61.20%
N11	1,108.25	14	58.00%	N21	0.8694	14	58.00%
N15	1,016.42	15	54.80%	N16	0.8664	15	54.80%
N1	913.06	16	51.60%	N12	0.8648	16	51.60%
N13	901.48	17	48.30%	N7	0.8647	17	48.30%
N16	896.82	18	45.10%	N23	0.8644	18	45.10%
N12	800.97	19	41.90%	N27	0.863	19	41.90%
N28	790.66	20	38.70%	N15	0.8628	20	38.70%
N3	782.48	21	35.40%	N8	0.8603	21	35.40%
N21	692.11	22	32.20%	N2	0.8553	22	32.20%
N7	646.58	23	29.00%	N3	0.8544	23	29.00%
N27	628.73	24	25.80%	N1	0.8528	24	25.80%
N25	609.36	25	22.50%	N30	0.8527	25	22.50%
N2	595.51	26	19.30%	N25	0.8475	26	19.30%
N19	573.71	27	16.10%	N14	0.8446	27	16.10%
N4	544.58	28	12.90%	N13	0.8438	28	12.90%
N14	530.46	29	9.60%	N4	0.8262	29	9.60%
N24	331.69	30	6.40%	N11	0.8163	30	6.40%
N30	257.81	31	3.20%	N24	0.7868	31	3.20%
N32	107.78	32	0.00%	N32	0.7464	32	0.00%

Results of Feature Ranking for ILDF based on Bounded and Unbounded Mahalanobis Distance for e-tongue (TN honey)

D^2				D_A^2			
Feature	**Criterion Value**	**Rank**	**Percent**	**Feature**	**Criterion Value**	**Rank**	**Percent**
T3	5,306.00	1	100.00%	T11	0.9229	1	100.00%
T11	1,153.05	2	88.80%	T2	0.887	2	88.80%
T2	492.30	3	77.70%	T1	0.8181	3	77.70%
T1	142.21	4	66.60%	T7	0.7394	4	66.60%
T7	115.22	5	55.50%	T8	0.7383	5	55.50%
T8	100.75	6	44.40%	T9	0.7261	6	44.40%
T10	89.38	7	33.30%	T3	0.7189	7	33.30%
T5	62.51	8	22.20%	T5	0.706	8	22.20%
T4	46.21	9	11.10%	T10	0.6816	9	11.10%
T9	42.27	10	0.00%	T4	0.6441	10	0.00%

Results of Feature Ranking for ILDF based on Bounded and Unbounded Mahalanobis Distance for e-nose (WT honey)

D^2				D_A^2			
Feature	**Criterion Value**	**Rank**	**Percent**	**Feature**	**Criterion Value**	**Rank**	**Percent**
N5	4,127.81	1	100.00%	N23	0.9408	1	100.00%
N29	4,037.63	2	96.70%	N28	0.9369	2	96.70%
N23	3,405.52	3	93.50%	N20	0.9356	3	93.50%
N26	3,347.08	4	90.30%	N31	0.9352	4	90.30%
N9	2,942.97	5	87.00%	N6	0.9344	5	87.00%
N20	2,461.75	6	83.80%	N26	0.9315	6	83.80%
N31	2,299.67	7	80.60%	N5	0.9306	7	80.60%
N6	2,016.16	8	77.40%	N10	0.9272	8	77.40%
N11	1,739.11	9	74.10%	N29	0.9247	9	74.10%
N17	1,666.27	10	70.90%	N9	0.9199	10	70.90%
N28	1,488.67	11	67.70%	N22	0.914	11	67.70%
N10	1,444.13	12	64.50%	N8	0.9127	12	64.50%
N15	1,443.47	13	61.20%	N15	0.9104	13	61.20%
N8	1,416.52	14	58.00%	N16	0.9085	14	58.00%
N16	1,315.33	15	54.80%	N11	0.9052	15	54.80%
N18	1,259.19	16	51.60%	N27	0.9038	16	51.60%
N12	1,012.62	17	48.30%	N25	0.9016	17	48.30%
N1	899.30	18	45.10%	N12	0.9006	18	45.10%
N22	881.08	19	41.90%	N31	0.8956	19	41.90%
N3	834.49	20	38.70%	N18	0.8947	20	38.70%
N13	831.93	21	35.40%	N17	0.8915	21	35.40%
N7	809.65	22	32.20%	N1	0.8898	22	32.20%
N27	732.23	23	29.00%	N30	0.8803	23	29.00%
N30	723.78	24	25.80%	N14	0.8796	24	25.80%
N21	719.40	25	22.50%	N4	0.8778	25	22.50%
N4	668.77	26	19.30%	N7	0.8732	26	19.30%
N19	631.36	27	16.10%	N2	0.8725	27	16.10%
N14	588.74	28	12.90%	N13	0.8683	28	12.90%
N2	542.75	29	9.60%	N19	0.8616	29	9.60%
N25	471.72	30	6.40%	N21	0.8595	30	6.40%
N24	296.97	31	3.20%	N24	0.8061	31	3.20%
N32	154.08	32	0.00%	N32	0.7011	32	0.00%

Results of Feature Ranking for ILDF based on Bounded and Unbounded Mahalanobis Distance for e-tongue (WT honey)

D^2				D_A^2			
Feature	**Criterion Value**	**Rank**	**Percent**	**Feature**	**Criterion Value**	**Rank**	**Percent**
T11	1,070.54	1	100.00%	T11	0.8947	1	100.00%
T2	1,016.20	2	88.80%	T2	0.8198	2	88.80%
T10	374.60	3	77.70%	T10	0.7894	3	77.70%
T7	263.00	4	66.60%	T1	0.778	4	66.60%
T1	132.66	5	55.50%	T8	0.777	5	55.50%
T3	104.01	6	44.40%	T9	0.763	6	44.40%
T5	70.10	7	33.30%	T5	0.6943	7	33.30%
T8	63.53	8	22.20%	T3	0.6848	8	22.20%
T4	43.16	9	11.10%	T4	0.6747	9	11.10%
T9	36.63	10	0.00%	T7	0.6743	10	0.00%

Results of Feature Ranking for ILDF based on Bounded and Unbounded Mahalanobis Distance for e-nose (YB honey)

D^2				D_A^2			
Feature	**Criterion Value**	**Rank**	**Percent**	**Feature**	**Criterion Value**	**Rank**	**Percent**
N5	6,474.00	1	100.00%	N11	0.9467	1	100.00%
N23	5,287.73	2	96.70%	N28	0.9459	2	96.70%
N29	4,512.24	3	93.50%	N9	0.9406	3	93.50%
N26	3,137.66	4	90.30%	N26	0.9354	4	90.30%
N6	3,085.16	5	87.00%	N20	0.9296	5	87.00%
N9	2,511.35	6	83.80%	N6	0.9218	6	83.80%
N31	2,324.08	7	80.60%	N17	0.9216	7	80.60%
N20	1,784.21	8	77.40%	N23	0.9216	8	77.40%
N11	1,685.89	9	74.10%	N5	0.921	9	74.10%
N17	1,329.32	10	70.90%	N25	0.911	10	70.90%
N28	1,278.17	11	67.70%	N10	0.9088	11	67.70%
N8	1,072.09	12	64.50%	N8	0.9075	12	64.50%
N10	1,015.58	13	61.20%	N18	0.9061	13	61.20%
N16	941.06	14	58.00%	N15	0.904	14	58.00%
N22	915.22	15	54.80%	N27	0.8891	15	54.80%
N15	889.70	16	51.60%	N12	0.8866	16	51.60%
N18	880.81	17	48.30%	N30	0.8852	17	48.30%
N30	856.89	18	45.10%	N16	0.8807	18	45.10%
N21	689.68	19	41.90%	N13	0.8799	19	41.90%
N13	679.78	20	38.70%	N7	0.8794	20	38.70%
N12	639.55	21	35.40%	N22	0.8742	21	35.40%
N7	517.44	22	32.20%	N14	0.8728	22	32.20%
N4	493.31	23	29.00%	N19	0.8703	23	29.00%
N1	484.69	24	25.80%	N21	0.8695	24	25.80%
N14	461.52	25	22.50%	N1	0.8686	25	22.50%
N25	447.94	26	19.30%	N31	0.8682	26	19.30%
N27	399.56	27	16.10%	N29	0.8596	27	16.10%
N3	369.81	28	12.90%	N4	0.8577	28	12.90%
N19	339.41	29	9.60%	N3	0.848	29	9.60%
N2	313.80	30	6.40%	N2	0.8372	30	6.40%
N24	209.73	31	3.20%	N24	0.8243	31	3.20%
N32	108.32	32	0.00%	N32	0.7648	32	0.00%

Results of Feature Ranking for ILDF based on Bounded and Unbounded Mahalanobis Distance for e-tongue (YB honey)

D^2				D_A^2			
Feature	**Criterion Value**	**Rank**	**Percent**	**Feature**	**Criterion Value**	**Rank**	**Percent**
T11	960.45	1	100.00%	T11	0.9134	1	100.00%
T2	218.66	2	88.80%	T1	0.8312	2	88.80%
T1	202.20	3	77.70%	T2	0.8204	3	77.70%
T10	164.84	4	66.60%	T9	0.7566	4	66.60%
T7	85.19	5	55.50%	T10	0.7477	5	55.50%
T3	57.36	6	44.40%	T8	0.7058	6	44.40%
T8	46.07	7	33.30%	T7	0.6164	7	33.30%
T9	37.04	8	22.20%	T3	0.6067	8	22.20%
T5	32.67	9	11.10%	T4	0.6042	9	11.10%
T4	22.20	10	0.00%	T5	0.5806	10	0.00%