





e-PROCEEDINGS

of The 5th International Conference on Computing, Mathematics and Statistics (iCMS2021)

4-5 August 2021 Driving Research Towards Excellence





e-Proceedings of the 5th International Conference on Computing, Mathematics and Statistics (iCMS 2021)

Driving Research Towards Excellence

Editor-in-Chief: Norin Rahayu Shamsuddin

Editorial team:

Dr. Afida Ahamad Dr. Norliana Mohd Najib Dr. Nor Athirah Mohd Zin Dr. Siti Nur Alwani Salleh Kartini Kasim Dr. Ida Normaya Mohd Nasir Kamarul Ariffin Mansor

e-ISBN: 978-967-2948-12-4 DOI

Library of Congress Control Number:

Copyright © 2021 Universiti Teknologi MARA Kedah Branch

All right reserved, except for educational purposes with no commercial interests. No part of this publication may be reproduced, copied, stored in any retrieval system or transmitted in any form or any means, electronic or mechanical including photocopying, recording or otherwise, without prior permission from the Rector, Universiti Teknologi MARA Kedah Branch, Merbok Campus. 08400 Merbok, Kedah, Malaysia.

The views and opinions and technical recommendations expressed by the contributors are entirely their own and do not necessarily reflect the views of the editors, the Faculty or the University.

Publication by Department of Mathematical Sciences Faculty of Computer & Mathematical Sciences UiTM Kedah

TABLE OF CONTENT

PART 1: MATHEMATICS

| | Page |
|--|------|
| STATISTICAL ANALYSIS ON THE EFFECTIVENESS OF SHORT-TERM PROGRAMS DURING COVID-19 PANDEMIC: IN THE CASE OF PROGRAM BIJAK SIFIR 2020 Nazihah Safie, Syerrina Zakaria, Siti Madhihah Abdul Malik, Nur Baini Ismail, Azwani Alias Ruwaidiah | 1 |
| Idris | |
| RADIATIVE CASSON FLUID OVER A SLIPPERY VERTICAL RIGA PLATE WITH VISCOUS DISSIPATION AND BUOYANCY EFFECTS Siti Khuzaimah Soid, Khadijah Abdul Hamid, Ma Nuramalina Nasero, NurNajah Nabila Abdul Aziz | 10 |
| GAUSSIAN INTEGER SOLUTIONS OF THE DIOPHANTINE EQUATION $x^4 + y^4 = z^3$ FOR $x \neq y$ <i>Shahrina Ismail, Kamel Ariffin Mohd Atan and Diego Sejas Viscarra</i> | 19 |
| A SEMI ANALYTICAL ITERATIVE METHOD FOR SOLVING THE EMDEN- FOWLER EQUATIONS Mat Salim Selamat, Mohd Najir Tokachil, Noor Aqila Burhanddin, Ika Suzieana Murad and Nur Farhana Razali | 28 |
| ROTATING FLOW OF A NANOFLUID PAST A NONLINEARLY SHRINKING SURFACE WITH FLUID SUCTION <i>Siti Nur Alwani Salleh, Norfifah Bachok and Nor Athirah Mohd Zin</i> | 36 |
| MODELING THE EFFECTIVENESS OF TEACHING BASIC NUMBERS THROUGH MINI TENNIS TRAINING USING MARKOV CHAIN Rahela Abdul Rahim, Rahizam Abdul Rahim and Syahrul Ridhwan Morazuk | 46 |
| PERFORMANCE OF MORTALITY RATES USING DEEP LEARNING APPROACH Mohamad Hasif Azim and Saiful Izzuan Hussain | 53 |
| UNSTEADY MHD CASSON FLUID FLOW IN A VERTICAL CYLINDER WITH POROSITY AND SLIP VELOCITY EFFECTS Wan Faezah Wan Azmi, Ahmad Qushairi Mohamad, Lim Yeou Jiann and Sharidan Shafie | 60 |
| DISJUNCTIVE PROGRAMMING - TABU SEARCH FOR JOB SHOP SCHEDULING PROBLEM S. Z. Nordin, K.L. Wong, H.S. Pheng, H. F. S. Saipol and N.A.A. Husain | 68 |
| FUZZY AHP AND ITS APPLICATION TO SUSTAINABLE ENERGY PLANNING DECISION PROBLEM <i>Liana Najib and Lazim Abdullah</i> | 78 |
| A CONSISTENCY TEST OF FUZZY ANALYTIC HIERARCHY PROCESS Liana Najib and Lazim Abdullah | 89 |
| FREE CONVECTION FLOW OF BRINKMAN TYPE FLUID THROUGH AN COSINE OSCILLATING PLATE | 98 |

Siti Noramirah Ibrahim, Ahmad Qushairi Mohamad, Lim Yeou Jiann, Sharidan Shafie and Muhammad Najib Zakaria

RADIATION EFFECT ON MHD FERROFLUID FLOW WITH RAMPED WALL106TEMPERATURE AND ARBITRARY WALL SHEAR STRESS106

Nor Athirah Mohd Zin, Aaiza Gul, Siti Nur Alwani Salleh, Imran Ullah, Sharena Mohamad Isa, Lim Yeou Jiann and Sharidan Shafie

PART 2: STATISTICS

| A REVIEW ON INDIVIDUAL RESERVING FOR NON-LIFE INSURANCE Kelly Chuah Khai Shin and Ang Siew Ling | 117 |
|---|-----|
| STATISTICAL LEARNING OF AIR PASSENGER TRAFFIC AT THE MURTALA MUHAMMED INTERNATIONAL AIRPORT, NIGERIA <i>Christopher Godwin Udomboso and Gabriel Olugbenga Ojo</i> | 123 |
| ANALYSIS ON SMOKING CESSATION RATE AMONG PATIENTS IN HOSPITAL SULTAN ISMAIL, JOHOR Siti Mariam Norrulashikin, Ruzaini Zulhusni Puslan, Nur Arina Bazilah Kamisan and Siti Rohani Mohd Nor | 137 |
| EFFECT OF PARAMETERS ON THE COST OF MEMORY TYPE CHART Sakthiseswari Ganasan, You Huay Woon and Zainol Mustafa | 146 |
| EVALUATION OF PREDICTORS FOR THE DEVELOPMENT AND PROGRESSION OF DIABETIC RETINOPATHY AMONG DIABETES MELLITUS TYPE 2 PATIENTS <i>Syafawati Ab Saad, Maz Jamilah Masnan, Karniza Khalid and Safwati Ibrahim</i> | 152 |
| REGIONAL FREQUENCY ANALYSIS OF EXTREME PRECIPITATION IN PENINSULAR MALAYSIA <i>Iszuanie Syafidza Che Ilias, Wan Zawiah Wan Zin and Abdul Aziz Jemain</i> | 160 |
| EXPONENTIAL MODEL FOR SIMULATION DATA VIA MULTIPLE IMPUTATION IN THE PRESENT OF PARTLY INTERVAL-CENSORED DATA <i>Salman Umer and Faiz Elfaki</i> | 173 |
| THE FUTURE OF MALAYSIA'S AGRICULTURE SECTOR BY 2030 Thanusha Palmira Thangarajah and Suzilah Ismail | 181 |
| MODELLING MALAYSIAN GOLD PRICES USING BOX-JENKINS APPROACH Isnewati Ab Malek, Dewi Nur Farhani Radin Nor Azam, Dinie Syazwani Badrul Aidi and Nur Syafiqah Sharim | 186 |
| WATER DEMAND PREDICTION USING MACHINE LEARNING: A REVIEW Norashikin Nasaruddin, Shahida Farhan Zakaria, Afida Ahmad, Ahmad Zia Ul-Saufie and Norazian Mohamaed Noor | 192 |
| DETECTION OF DIFFERENTIAL ITEM FUNCTIONING FOR THE NINE- QUESTIONS DEPRESSION RATING SCALE FOR THAI NORTH DIALECT | 201 |

Suttipong Kawilapat, Benchlak Maneeton, Narong Maneeton, Sukon Prasitwattanaseree, Thoranin Kongsuk, Suwanna Arunpongpaisal, Jintana Leejongpermpool, Supattra Sukhawaha and Patrinee Traisathit

ACCELERATED FAILURE TIME (AFT) MODEL FOR SIMULATION PARTLY 210 INTERVAL-CENSORED DATA

Ibrahim El Feky and Faiz Elfaki

MODELING OF INFLUENCE FACTORS PERCENTAGE OF GOVERNMENTS' RICE 217 RECIPIENT FAMILIES BASED ON THE BEST FOURIER SERIES ESTIMATOR 217

Chaerobby Fakhri Fauzaan Purwoko, Ayuning Dwis Cahyasari, Netha Aliffia and M. Fariz Fadillah Mardianto

CLUSTERING OF DISTRICTS AND CITIES IN INDONESIA BASED ON POVERTY 225 INDICATORS USING THE K-MEANS METHOD 225

Khoirun Niswatin, Christopher Andreas, Putri Fardha Asa OktaviaHans and M. Fariz Fadilah Mardianto

ANALYSIS OF THE EFFECT OF HOAX NEWS DEVELOPMENT IN INDONESIA 233 USING STRUCTURAL EQUATION MODELING-PARTIAL LEAST SQUARE

Christopher Andreas, Sakinah Priandi, Antonio Nikolas Manuel Bonar Simamora and M. Fariz Fadillah Mardianto

A COMPARATIVE STUDY OF MOVING AVERAGE AND ARIMA MODEL IN 241 FORECASTING GOLD PRICE

Arif Luqman Bin Khairil Annuar, Hang See Pheng, Siti Rohani Binti Mohd Nor and Thoo Ai Chin

CONFIDENCE INTERVAL ESTIMATION USING BOOTSTRAPPING METHODS 249 AND MAXIMUM LIKELIHOOD ESTIMATE

Siti Fairus Mokhtar, Zahayu Md Yusof and Hasimah Sapiri

DISTANCE-BASED FEATURE SELECTION FOR LOW-LEVEL DATA FUSION OF 256 SENSOR DATA

M. J. Masnan, N. I. Maha3, A. Y. M. Shakaf, A. Zakaria, N. A. Rahim and N. Subari

BANKRUPTCY MODEL OF UK PUBLIC SALES AND MAINTENANCE MOTOR 264 VEHICLES FIRMS

Asmahani Nayan, Amirah Hazwani Abd Rahim, Siti Shuhada Ishak, Mohd Rijal Ilias and Abd Razak Ahmad

INVESTIGATING THE EFFECT OF DIFFERENT SAMPLING METHODS ON 271 IMBALANCED DATASETS USING BANKRUPTCY PREDICTION MODEL

Amirah Hazwani Abdul Rahim, Nurazlina Abdul Rashid, Abd-Razak Ahmad and Norin Rahayu Shamsuddin

INVESTMENT IN MALAYSIA: FORECASTING STOCK MARKET USING TIME 278 SERIES ANALYSIS

Nuzlinda Abdul Rahman, Chen Yi Kit, Kevin Pang, Fauhatuz Zahroh Shaik Abdullah and Nur Sofiah Izani

PART 3: COMPUTER SCIENCE & INFORMATION TECHNOLOGY

ANALYSIS OF THE PASSENGERS' LOYALTY AND SATISFACTION OF AIRASIA 291 PASSENGERS USING CLASSIFICATION 291

Ee Jian Pei, Chong Pui Lin and Nabilah Filzah Mohd Radzuan

HARMONY SEARCH HYPER-HEURISTIC WITH DIFFERENT PITCH 299 ADJUSTMENT OPERATOR FOR SCHEDULING PROBLEMS

Khairul Anwar, Mohammed A.Awadallah and Mohammed Azmi Al-Betar

A 1D EYE TISSUE MODEL TO MIMIC RETINAL BLOOD PERFUSION DURING 307 RETINAL IMAGING PHOTOPLETHYSMOGRAPHY (IPPG) ASSESSMENT: A DIFFUSION APPROXIMATION – FINITE ELEMENT METHOD (FEM) APPROACH Harnani Hassan, Sukreen Hana Herman, Zulfakri Mohamad, Sijung Hu and Vincent M. Dwyer

INFORMATION SECURITY CULTURE: A QUALITATIVE APPROACH ON 325 MANAGEMENT SUPPORT

Qamarul Nazrin Harun, Mohamad Noorman Masrek, Muhamad Ismail Pahmi and Mohamad Mustaqim Junoh

APPLY MACHINE LEARNING TO PREDICT CARDIOVASCULAR RISK IN RURAL 335 CLINICS FROM MEXICO

Misael Zambrano-de la Torre, Maximiliano Guzmán-Fernández, Claudia Sifuentes-Gallardo, Hamurabi Gamboa-Rosales, Huizilopoztli Luna-García, Ernesto Sandoval-García, Ramiro Esquivel-Felix and Héctor Durán-Muñoz

ASSESSING THE RELATIONSHIP BETWEEN STUDENTS' LEARNING STYLES 343 AND MATHEMATICS CRITICAL THINKING ABILITY IN A 'CLUSTER SCHOOL' Salimah Ahmad, Asyura Abd Nassir, Nor Habibah Tarmuji, Khairul Firhan Yusob and Nor Azizah Yacob

STUDENTS' LEISURE WEEKEND ACTIVITIES DURING MOVEMENT CONTROL 351 ORDER: UiTM PAHANG SHARING EXPERIENCE

Syafiza Saila Samsudin, Noor Izyan Mohamad Adnan, Nik Muhammad Farhan Hakim Nik Badrul Alam, Siti Rosiah Mohamed and Nazihah Ismail

DYNAMICS SIMULATION APPROACH IN MODEL DEVELOPMENT OF UNSOLD 363 NEW RESIDENTIAL HOUSING IN JOHOR

Lok Lee Wen and Hasimah Sapiri

WORD PROBLEM SOLVING SKILLS AS DETERMINANT OF MATHEMATICS 371 PERFORMANCE FOR NON-MATH MAJOR STUDENTS 371

Shahida Farhan Zakaria, Norashikin Nasaruddin, Mas Aida Abd Rahim, Fazillah Bosli and Kor Liew Kee

ANALYSIS REVIEW ON CHALLENGES AND SOLUTIONS TO COMPUTER 378 PROGRAMMING TEACHING AND LEARNING

Noor Hasnita Abdul Talib and Jasmin Ilyani Ahmad

PART 4: OTHERS

ANALYSIS OF CLAIM RATIO, RISK-BASED CAPITAL AND VALUE-ADDED 387 INTELLECTUAL CAPITAL: A COMPARISON BETWEEN FAMILY AND GENERAL TAKAFUL OPERATORS IN MALAYSIA Nur Amalina Syafiga Kamaruddin, Norizarina Ishak, Siti Raihana Hamzah, Nurfadhlina Abdul Halim and Ahmad Fadhly Nurullah Rasade THE IMPACT OF GEOMAGNETIC STORMS ON THE OCCURRENCES OF 396 EARTHOUAKES FROM 1994 TO 2017 USING THE GENERALIZED LINEAR MIXED MODELS N. A. Mohamed, N. H. Ismail, N. S. Majid and N. Ahmad **BIBLIOMETRIC ANALYSIS ON BITCOIN 2015-2020** 405 Nurazlina Abdul Rashid, Fazillah Bosli, Amirah Hazwani Abdul Rahim, Kartini Kasim and Fathiyah Ahmad@Ahmad Jali GENDER DIFFERENCE IN EATING AND DIETARY HABITS AMONG UNIVERSITY 413 **STUDENTS** Fazillah Bosli, Siti Fairus Mokhtar, Noor Hafizah Zainal Aznam, Juaini Jamaludin and Wan Siti Esah Che Hussain MATHEMATICS ANXIETY: A BIBLIOMETRIX ANALYSIS 420 Kartini Kasim, Hamidah Muhd Irpan, Noorazilah Ibrahim, Nurazlina Abdul Rashid and Anis Mardiana Ahmad

PREDICTION OF BIOCHEMICAL OXYGEN DEMAND IN MEXICAN SURFACE 428 WATERS USING MACHINE LEARNING 428

Maximiliano Guzmán-Fernández, Misael Zambrano-de la Torre, Claudia Sifuentes-Gallardo, Oscar Cruz-Dominguez, Carlos Bautista-Capetillo, Juan Badillo-de Loera, Efrén González Ramírez and Héctor Durán-Muñoz

DISTANCE-BASED FEATURE SELECTION FOR LOW-LEVEL DATA FUSION OF SENSOR DATA

M. J. Masnan^{1,2}, N. I. Mahat³, A. Y. M. Shakaf⁴, A. Zakaria⁵, N. A. Rahim⁶ and N. Subari⁷
¹ Faculty of Applied & Human Sciences, Institute of Engineering Mathematics, Universiti Malaysia Perlis, ² Mathematical Sciences & Data Analytics, Centre of Excellence for Social Innovation & Sustainability, Universiti Malaysia Perlis, ³ Centre of Testing, Measurements & Appraisal (CeTMA), Universiti Utara Malaysia, ⁴ Institute of Sustainable Agrotechnology, Universiti Malaysia Perlis, ⁷ Faculty of Electrical & Electronics Engineering, Universiti Malaysia Pahang (¹ mazjamilah@unimap.edu.my, ³ noridayu@uum.edu.my, ⁴ aliyeon@unimap.edu.my, ⁵ ammarzakaria@unimap.edu.my, ⁶ norasmadi@unimap.edu.my, ⁷ aziansubari@ump.edu.my)

Low level data fusion offers a mechanism for raw data from different sensor devices to be fused in an attempt to improve classification performance. This scenario creates a challenge for engineers to deal with large number of features over the smaller number of observations, or also known as high dimensional problem. Traditionally, engineers prefer to apply feature extraction in choosing important features for classification task. Unfortunately, the traditional method is bound to certain limitations. Thus, the objective of this study is to propose a feature selection based on unbounded Mahalanobis distance $[0, \infty)$ to replace the feature extraction phase in the low-level data fusion. The average pair-wise distances for the fused features were calculated and filtered from largest to smallest values, and features with the larger distance value are considered important. Classification results using the ranked features in a feature subset selection manner have shown that the proposed distance provides an effective and easy approach in choosing the important features that lead to good classification performance.

Keywords: Feature Selection, Mahalanobis Distance, Classification, Data Fusion

1. Introduction

Nowadays, the application of sensor devices in food and beverage manufacturing industries has become a common setting. Sensor devices are used to replace the trained human panel which may have some drawbacks such as discrepancy due to human fatigue or stress, time consuming, expensive and impossible for on line monitoring, and finally the evaluation perceptions depend on panelists' training. Therefore, complimentary yet reliable artificial sensors that mimic the trained human panel's smell and taste system were proposed to overcome the drawbacks. The application of electronic nose and electronic tongue were introduced to substitute these systems which has begun in the early 1980s and 1990s, respectively. Unfortunately, these sensors work independently. In order to combine the information from the two sensors to increase the classification or identification accuracy, the multi sensor data fusion method was introduced.

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) which consists of low-level data fusion, intermediate-level data fusion and high-level data fusion. A custom phase that involves in all level of data fusion is the feature extraction process where the 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). Due to some weaknesses of feature extraction method specifically in choosing specified features, this study intends to introduce a simple yet effective method to replace the conventional method. Therefore, the objective of this study is to propose a new approach to replace the feature extraction phase with feature selection method for the low-level data fusion of two sensor devices i.e. electronic nose and electronic tongue. A distance-based feature

selection using unbounded Mahalanobis distance criterion is applied for classification of multi-class problem.

2. Review of Literature

Survey of literature was conducted regarding the low-level data fusion, feature extraction and feature selection related to the application of sensor devices in food and beverage industries that may involve the application of electronic nose, electronic tongue and other sensor devices.

2.1 Low Level Data Fusion

The simplest model for multi sensor data fusion is the low-level data fusion as shown in Figure 1. It implies a concatenation of raw data from different sensors from similar or dissimilar sensors (Rudnitskaya et al., 2006). The resulting data matrix as a result from the low-level data fusion contains n rows of observations from different classes with p columns of features that are obtained from the fused sensors. Basically, this framework combines signals from similar or different sensors to produce new data that is predicted to be more informative for classification.



Figure 1: Framework for low level data fusion (Hall, 1997)

Several research that have applied electronic nose and electronic tongue, and/or other sensors include Vera et al. (2011), Zakaria et al. (2011), Zakaria et al. (2010), and Rodriguez-Mendez et al. (2004). The low-level data fusion is preferred due to its ability to fuse raw data i.e. signals from different or similar sensors before feature extraction is imposed towards the fused data matrix. Some researchers presumed this framework as the most efficient compared to the intermediate-level and high-level data fusion because it avoids loss of information since all available raw data from each sensor are employed in the next phase.

2.2 Feature Extraction

Traditionally, feature extraction is employed in the framework of low, intermediate and high-level data fusion. Most previous researchers who applied multi sensor data fusion based on JDL process model were only focused on the use of principal component analysis as the feature extraction approach (Prieto et al., 2011, Vera et al., 2011, and Zakaria et al., 2010). Feature extraction is favorable due to its advantageous to deal with high multicollinearity among the p fused features from the applied sensor devices. It also allows visualization most of the information contained in raw data matrix in p principal components which are orthogonal to each other. Unfortunately, the resulting linear combinations of the p principal components are difficult to interpret especially when specific or subset of features is/are of interest. Furthermore, decision to include certain number of principal components for the next phase in the fusion framework are either based on the average eigenvalue, percentage of variance explained or simply using the scree graph. In this context, the inclusion of all

the fused features without considering the relevancy or informative of a single feature is ignored. Thus, studies on feature selection as an alternative to feature extraction i.e. principal component analysis tool in multi sensor data fusion is addressed in this study.

2.3 Feature Selection

Dealing with multi sensor data fusion means more features are being considered for the construction of classifiers especially for low level data fusion. This is due to the nature of this framework that will fuse all the resulting features from the applied sensors which usually lead to the problem of curse of dimensionality. Classification within this condition may be inaccurate and the inclusion of too many features may harm the performance of a sample classifier (Foithong et al., 2012 and McLachlan, 1992). In addition, unlike feature extraction method where all features are included in the analysis, feature selection is able to identify and include only relevant features for the next phase in data fusion framework. Feature selection is a process of identifying the most useful features in describing differences among the possible class (McLachlan, 1992). According to Dernoncourt et al. (2014), feature selection is the process of removing irrelevant features in reducing the dimensionality of data to processed, decreasing the execution time and improving the predictive accuracy of classifier (Chandrashekar and Shahin, 2014).

There are two common evaluation functions for feature selection that are allocation and separation criteria. The general aim of an evaluation function is to measure the discrimination ability of a feature or subset of features to distinguish different class labels. Allocation criterion measures the effectiveness of fused feature in predicting class membership using the overall error rate of the optimal rule formed by the p fused features. Whereas, separation criterion assesses the effectiveness of p fused feature which tends to maximally distinguish or separate the observations to its classes (Masnan et al., 2017). Thus, allocation criterion is appropriate if the objective is to identify feature useful to form a classifier with high accuracy for allocating future observations. Whilst, separation criterion is relevant if the objective is to identify features that are useful to describe separation between classes. Established by the later approach, a method that can be applied is the distance-based feature selection that aims to identify useful features by their discriminating power based on highest distance for separation among classes from the available p fused features. Ray and Turner (1992) have introduced unbounded and bounded Mahalanobis distance-based evaluation criteria for multi class classification problem both in the distribution free and Gaussion distribution cases for the recognition of hand-printed numeric characters. This study applies the unbounded Mahalanobis distance criterion to replace the feature extraction phase in the low-level data fusion framework for multi-class classification problem.

3. Methodology

The proposed methodology in this study is an extension work from Ray and Turner (1992) with regards to the unbounded Mahalanobis distance for feature evaluation, but with some modification so that it can be applied in multi sensor data fusion particularly the low-level data fusion. The calculation of unbounded Mahalanobis distance within the range $[0, \infty)$ in this study is obtained based on the pair-wise distances between multi-class ($\pi > 2$) classification. This criterion is practical in selecting relevant features since it can provide a measure of similarity between multivariate populations and uses covariance information between features to weight the contribution of all features to the unbounded distance. Furthermore, the unbounded distance gives less weight to features that have high variance and high correlation (Masnan et al., 2017).

Figure 2 illustrates the graphical presentation of the pair-wise distance for multi class classification problem. Suppose the preprocessed data matrix for investigation are \mathbf{X}_{ip_1} with p_1 features are from electronic nose, and \mathbf{X}_{ip_2} with p_2 features from electronic tongue, and the fused features denotes as $\mathbf{X}_{ip} = (\mathbf{X}_{ip_1} + \mathbf{X}_{ip_2})$. All the \mathbf{X}_{ip} observations are partitioned to multi-class (π)

where $i = 1, 2, ..., \pi$, and k = 1, 2, ..., n, with a total of $p = (p_1 + p_2)$ fused feature from the electronic nose and electronic tongue, respectively. For this study, even though different sensor devices were applied, the low-level data fusion is successfully implemented by controlling similar number of observations within similar number of classes for each data collected using the electronic nose and electronic tongue.



Figure 2: Graphical representation of pair-wise unbounded Mahalanobis distance for multi-class

In order to evaluate the fused features that are useful to describe separation between classes, the univariate unbounded Mahalanobis distances $\left(\Delta_{Uij}^2\right)$ which is estimated by D_{Uij}^2 are calculated. These calculations were developed using algorithms that were executed using RStudio which simplifies the calculation of pair-wise distance for every single fused features. Equations (1) to (4) are the calculation for the estimated pair-wise-class (i, j) distances of D_{Uij}^2 required for the feature evaluation where $(i=1, 2, ..., \pi; j=i+1, ..., \pi-1; i \neq j)$ and n is the sample size of the respective class. Equation (3) is the simplified calculation of the univariate Mahalanobis distance where the inverse of S_{ij} in (2) is the reciprocal of the respective common variance. The summation of each of the resulting ${}^{\pi}C_2$ pair-wise distances in (3) divided by ${}^{\pi}C_2$ gives the average unbounded Mahalanobis distance of each of the evaluated feature where $D_{U_{x(ip)}}^2 \in [0, \infty)$. The process is repeated

for each of the *p* fused feature independently.

$$D_{ij}^{2} = \left(\overline{X}_{i} - \overline{X}_{j}\right)^{T} S_{ij}^{-1} \left(\overline{X}_{i} - \overline{X}_{j}\right)$$
(1)

$$S_{ij} = \frac{\sum_{k=1}^{n_1} (X_{ik} - \overline{X}_i) (X_{ik} - \overline{X}_i)^T + \sum_{k=1}^{n_2} (X_{jk} - \overline{X}_j) (X_{jk} - \overline{X}_j)^T}{n_1 + n_2 - 2}$$
(2)

$$D_{ij}^2 = \frac{\left(\overline{X}_i - \overline{X}_j\right)^2}{S_{ij}} \tag{3}$$

$$D_{U_{x(ip)}}^{2} = \frac{1}{\pi C_{2}} \sum_{i=1}^{\pi} \sum_{j=i+1}^{\pi-1} D_{ij}^{2}$$
(4)

The average pair-wise distances for each p fused features will be obtained using (4), and the generated values are ranked from largest to smallest. The largest average pair-wise distance values based on the ranked features is considered as the indicator of good discriminant features to describe separation between classes. The objective of performing univariate unbounded Mahalanobis distance is to select the best subset of features that has been ranked and further selected for best performance

of classification. For simplicity of showing the ${}^{\pi}C_2$ pairwise-class distances per feature, Table 1 represents the combinations of the resulting distances for a feature x, $D_{U_1}^2$.

| | π_1 | π_1 | π_1 | π_1 | | π |
|------------|---------------|---------------|---------------|---------------|---|--------------|
| π_1 | 0 | $D_{1,2}^{2}$ | $D_{1,3}^2$ | $D_{1,4}^2$ | | $D_{l\pi}^2$ |
| π_2 | $D_{2,1}^2$ | 0 | $D_{2,3}^2$ | $D_{2,4}^2$ | | $D_{2\pi}^2$ |
| π 3 | $D_{3,1}^2$ | $D_{3,2}^2$ | 0 | $D_{3,4}^2$ | | $D^2_{3\pi}$ |
| π_4 | $D_{4,1}^2$ | $D_{3,2}^2$ | $D_{4,3}^2$ | 0 | | $D^2_{4\pi}$ |
| : | ÷ . | : | : | ÷ • | 0 | ÷ |
| π | $D^2_{\pi l}$ | $D^2_{\pi 2}$ | $D^2_{\pi 3}$ | $D^2_{\pi 4}$ | | 0 |

Table 1: The generated ${}^{\pi}C_2$ pair-wise distances among classes for univariate feature

The diagonal elements are the distances of the same centroids while either of the pair-wise distances in the lower or upper diagonal is useful for the required distance measure. In this study, the lower diagonal pair-wise class's distances are used. The algorithm for the unbounded pair-wise Mahalanobis distance (4) for low-level data fusion of electronic nose and electronic tongue is described as follows. Basically, this algorithm can be applied to replace the traditional feature extraction i.e. principal component analysis.

_

| Input : | $X_{ip} = (X_{ip} + X_{ip})$ - matrix of observation for fused feature set |
|------------------------|---|
| Output : | Ranked fused features $\mathcal{F} = \begin{bmatrix} D_{U_1}^2 > D_{U_2}^2 > \dots > D_{U_p}^2 \end{bmatrix}$ |
| Step 1 : | Initialize X _{<i>ip</i>} = $\begin{bmatrix} X_{i1},, X_{ip_1},, X_{ip2},, X_{ip} \end{bmatrix}$ |
| Step 2 : | For each feature $\begin{pmatrix} X_{ip_1}, X_{ip_2} \end{pmatrix} \in \mathbb{Z}^p$ |
| | i. Calculate the unbounded Mahalanobis distance for ${}^{\pi}C_{2}$ pair-wise |
| | classes using criterion (3) for each fused feature in \mathbb{Z}^p $D_{ij}^2 = \frac{\left(\overline{X}_i - \overline{X}_j\right)^2}{S_{ij}}, (i = 1, 2,, \pi; j = i + 1,, \pi - 1; i \neq j)$ |
| | ii. Calculate the average distance for ${}^{\pi}C_{2}$ pair-wise classes using |
| | criterion (4) for each fused feature in $X_{ip} \in \mathbb{Z}^p$ |
| | $D_{U_{x(ip)}}^{2} = \frac{1}{\pi C_{2}} \sum_{i=1}^{\pi} \sum_{j=i+1}^{\pi-1} D_{ij}^{2}$ |
| | iii. Store $D^2_{U_{x(ip)}}$ into $\mathcal{F} = [$ |
| Step 3 : | End of for loop |
| Step 4 : | Rank $D^2_{U_{r(in)}}$ (from largest to smallest values) where |
| | $\mathcal{F} = \left[D_{U_1}^2 \stackrel{\wedge (\psi)}{>} D_{U_2}^2 > \dots > D_{U_p}^2 \right].$ |
| Step 4 : Filter | \mathcal{F} for best feature subset evaluation. |
| | |

The evaluation of the algorithm will be illustrated using four datasets of different honeys for the classification of pure and adulterated honey. The secondary dataset was obtained from the Center of

Excellence for Advanced Sensor Technology (CEASTech), Universiti Malaysia Perlis. In order to evaluate whether the fused features with the highest average pair-wise unbounded Mahalanobis distance is able to separate multi-class problem, the ranked features will be used to perform classification of pure honey from its adulterated concentrations. Linear discriminant analysis classifier will be used to evaluate the performance of the selected subset of features for the four honey datasets namely Agromas, As-Syifa, Syair Timur and Tayyibah. The apparent error rate of the classification will be implemented using the leave-one-out approach which is believed can improved the classification process.

4. Result and Discussion

Table 2 presents the ranking of features based on the output of the algorithm for the unbounded pairwise Mahalanobis distance for low-level data fusion. There are 32 and 10 array of sensors for electronic nose (N) and electronic tongue (T), respectively. These array of sensors became the features used for feature evaluation. Features from both devices were numbered based on their total array of sensors attached in both sensors. From the results, obviously each different honey recorded different features' performance which means no specific similar features dominate the highest ranking.

Table 2: Ranked fused features for four types of honeys.

| Aş | Agromas As-Syifa | | Sya | ir Timur | Tayyibah | | |
|---------|------------------|---------|-----------|-----------|-----------|------------|-----------|
| Ranked | Unbounded | Ranked | Unbounded | Ranked | Unbounded | Ranked | Unbounded |
| Feature | Distance | Feature | Distance | Feature | Distance | Feature | Distance |
| N23 | 11,949.06 | N26 | 1937.62 | N29 | 12593.81 | <i>T</i> 7 | 84016 |
| N5 | 11,343.20 | N5 | 1872.11 | N5 | 6375.09 | <i>T2</i> | 17337.5 |
| N29 | 10,027.52 | N29 | 1815.44 | N23 | 6014.23 | N29 | 4120.59 |
| N31 | 4,040.16 | N31 | 1116.44 | N31 | 5832.96 | N23 | 3493.41 |
| N9 | 3,680.18 | N15 | 1086.9 | N26 | 5555.79 | N31 | 2661.3 |
| N26 | 3,482.48 | N9 | 1058.89 | N9 | 4647.56 | N5 | 2235.94 |
| N11 | 3,074.39 | N20 | 994.57 | <i>T2</i> | 4354.62 | N6 | 2188.87 |
| N6 | 2,418.34 | N16 | 928.61 | N11 | 3749.95 | N26 | 2043.21 |
| N20 | 2,300.65 | N23 | 917.45 | N6 | 3650.01 | N9 | 1883.41 |
| N10 | 1,793.69 | T11 | 888.82 | N20 | 3259.16 | N20 | 1156.73 |
| N17 | 1,778.24 | N17 | 886.27 | N17 | 2369.79 | N10 | 1097.63 |
| N15 | 1,407.55 | N13 | 832.86 | N28 | 2264.82 | N17 | 1052.48 |
| N16 | 1,161.90 | N8 | 800.05 | N10* | 1997.47 | T11 | 1037.43 |
| N28 | 1,129.29 | N21 | 779.23 | N1 | 1778.03 | N22 | 1029.37 |
| N22 | 1,124.58 | N11 | 755.48 | N8 | 1742.91 | N8 | 1020.89 |
| N8 | 1,094.92 | N18 | 727.66 | N18 | 1742.85 | N28 | 1019.52 |
| N18 | 1,074.86 | N28 | 712.29 | N15 | 1695.26 | N18 | 900.7 |
| N13 | 869.54 | N7 | 626.57 | N16 | 1346.31 | N15 | 870.4 |
| N12 | 770.64 | N12 | 616.34 | N22 | 1207.47 | N16 | 802.34 |
| T11 | 762.64 | N10 | 582.97 | N30 | 1074.31 | T9 | 660.37 |
| N30 | 741.57 | N1 | 580.04 | N3 | 913.47 | N11 | 657.82 |
| N4 | 669.04 | N3 | 545.42 | T9 | 889.88 | N12 | 554.35 |
| N7 | 646.60 | N4 | 516.91 | N12 | 868.62 | N13 | 544 |
| N14 | 574.10 | N14 | 502.47 | N13 | 842.9 | N27 | 537.6 |
| N21 | 546.70 | T2 | 498.87 | N19 | 826.82 | <i>T1</i> | 513.37 |
| N19 | 536.86 | N2 | 436.33 | N4 | 818.7 | N19 | 475.41 |
| N2 | 532.30 | N22 | 423.87 | N27 | 761.42 | N7 | 474.97 |
| N1 | 500.41 | N25 | 405.44 | N2 | 700.4 | N1 | 446.23 |
| N27 | 459.53 | N19 | 385.8 | N7 | 687.86 | N21 | 440.02 |
| N3 | 446.28 | N27 | 347.33 | N21 | 682.05 | N14 | 434.49 |
| N24 | 318.36 | N6 | 344.93 | N25 | 587.72 | N30 | 416.91 |
| N25 | 317.15 | N24 | 337.84 | N24 | 569.58 | N25 | 321.53 |

| Ag | Agromas | | As-Syifa | | Syair Timur | | Tayyibah | |
|------------|-----------|------------|-----------|------------|-------------|-----------|-----------|--|
| Ranked | Unbounded | Ranked | Unbounded | Ranked | Unbounde | Ranked | Unbounded | |
| Feature | Distance | Feature | Distance | Feature | d Distance | Feature | Distance | |
| <i>T2</i> | 236.18 | T9 | 318.68 | T11 | 542.8 | N3 | 318.65 | |
| <i>T3</i> | 193.98 | <i>T</i> 7 | 281.94 | N14 | 510.64 | N24 | 291.55 | |
| Τ7 | 180.85 | N30 | 261.39 | T1 | 341.75 | N2 | 274.43 | |
| T1 | 133.79 | <i>T3</i> | 258.33 | <i>T</i> 8 | 215.54 | N4 | 263.23 | |
| N32 | 119.50 | T8 | 148.62 | <i>T5</i> | 177.28 | T8 | 185.14 | |
| T5 | 96.45 | T1 | 132.06 | N32 | 142.51 | N32 | 143.86 | |
| T9 | 69.64 | T4 | 96.43 | T4 | 92.6 | Т3 | 114.05 | |
| <i>T</i> 8 | 55.04 | T5 | 70.97 | <i>T3</i> | 62.53 | T4 | 98.99 | |
| T4 | 42.19 | N32 | 67.79 | <i>T</i> 7 | 60.66 | <i>T5</i> | 90.02 | |
| T10 | 41 37 | T10 | 43 47 | T10 | 11.09 | T10 | 76.1 | |

| Table 2: Ranked fused | features fo | r four types (| of honevs (| (continued) |
|-----------------------|-------------|----------------|-------------|-------------|
| rubie 2. Runkeu rubeu | reatures ro | i ioui types | or noneys, | (continueu) |

Based on the results in Table 2, the top-ranked features can be considered random i.e. without any specific pattern of domination from a specific features for all dataset. However, from the observation of the top-ten-ranked features, several similar features appeared in all the dataset without any specific order e.g. N5, N9, N20, N23, N26, N29, and N31. The evaluation of the ranked features was made using the generated feature subset based on the inclusion of one feature at a time which begin with the top-ranked to the bottom-ranked. For example, the classification of Agromas honey from several adulterated concentrations, the generated feature subset begun with (subset 1=N23), (subset 2=N23, N5), (subset 3=N23, N5, N29), etc. until the last feature subset that contain all 42 fused features (subset 42=N23, N5, N29, ..., T10). The Agromas honey achieved 100% correct classification by using the 21 top-ranked features, and the result remain unchanged even though more features were included, except for the inclusion of feature 34 (T3) which affect the accuracy to drop to 99%. Figure 2 describes the behavior of the classification accuracy for each of the generated feature subset for all the datasets involved.



Feature Subset with Specified ClassificationAccuracy

Figure 2: Classification performance of ranked feature subset (inclusion of one ranked-feature)

5. Conclusion

Feature selection based on the evaluation of unbounded Mahalanobis distance can be used to replace the conventional feature extraction for the low-level data fusion involving different sensor devices. The proposed method has shown potential result of classifying pure honey from adulterated honey concentrations by selecting several specific top-ranked features. Unlike feature extraction, distancebased feature selection offers a simple and objective method to identify features that are useful to describe separation between classes which lead to good classification performance. However, the low-level data fusion can be a disadvantage if the features from one of the sensor devices were ranked at the bottom. This can be observed from the electronic tongue where most of

References

- Chandrashekar, G. and Sahin, F. (2014). A survey on feature selection methods. *Computers and Electrical Engineering*, 40, 16-28.
- Dernoncourt, D., Hanczar, B. and Zucker, J-D. (2014). Analysis of feature selection stability on high dimension and small sample. *Computational Statistics and Data Analysis*, 71, 681-693.
- Foithong, S., Pinngern, O. and Attachoo, B. (2012). Feature subset selection wrapper based on mutual information and rough sets. *Expert System with Applications*, 39, 574-584.
- Hall, D. L. and Llinas, J. (1997). An introduction to multisensor data fusion. Proceedings of the IEEE. 85, 6 23.
- Hall, D. L. (1992). Mathematical techniques in multi sensor data fusion. Boston: Artech House Inc.
- Masnan, M. (2017). Discriminant analysis of multi sensor data fusion based on percentile forward feature selection (Unpublished Doctoral thesis), Universiti Utara Malaysia, Sintok, Malaysia.
- McLachlan, G. J. (1992). *Discriminant analysis and statistical pattern recognition*. USA: John Wiley & Sons Inc.
- Prieto, N., Gay, M., Vidal, S., Aagaard, O., de Saja, J. A. and 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.
- Ray, S., and Turner, L. F. (1992). Mahalanobis distance-based two new feature evaluation criteria. *Information Sciences*, 60, 217-245.
- Rudnitskaya, A., Kirsanov, D., Legin, A., Beullens, K., Lammertyn, J., Nicolai, B. M., and Irudayaraj, J. (2006). Analysis of apples varieties – comparison of electronic tongue with different analytical techniques. *Sensors and Actuators B*, 116, 23-28.
- Vera, L., Aceña, L., Guash, J., Boque, R., Mestres, M., and Busto, O. (2011). Discrimination and sensory description of beers through data fusion. *Talanta*, 87, 136-142.
- Zakaria, A., Shakaff, A. Y. M., Adom, A. H., Ahmad, M. N., Masnan, M. J., Aziz, A. H. A., ..., and Kamarudin, L. M. (2010). Improved classification of *orthosiphon stamineus* by data fusion of electronic nose and tongue sensors. *Sensors*, 10, 8782-8796.





