Cawangan Kedah
Kampus Sungai Petani

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

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

Maz Jamilah Masnan

Faculty of Applied & Human Sciences, Institute of Engineering Mathematics, Universiti Malaysia Perlis,
Mathematical Sciences & Data Analytics, Centre of Excellence for Social Innovation & Sustainability, Universiti Malaysia Perlis,
Centre of Testing, Measurements & Appraisal (CeTMA), Universiti Utara Malaysia,
Institute of Sustainable Agrotechnology, Universiti Malaysia Perlis,
Centre of Excellence for Advanced Sensor Technology, Universiti Malaysia Perlis,
7Faculty of Electrical & Electronics Engineering, Universiti Malaysia Pahang

4 - 5 AUGUST 2021

CONTENT

1 INTRODUCTION
2 OBJECTIVES
3 LITERATURE REVIEW
4 METHODOLOGY
5 RESULTS AND DISCUSSIONS
6 CONCLUSIONS

Introduction

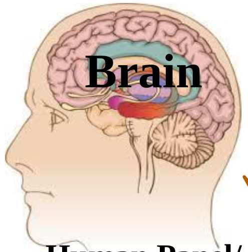
Conventional testing method in food & beverage manufacturing

Artificial sensors to mimic trained human panel

Unidentified Sample

+

=



Human Panel/ Trained Expert

Identified Sample

E-Tongue

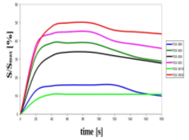
+

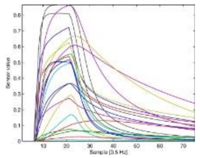
E-Nose

Electronic Nose

Electronic Tongue

PROBLEM
Multi Sensor Data Fusion - which model is appropriate to be applied?



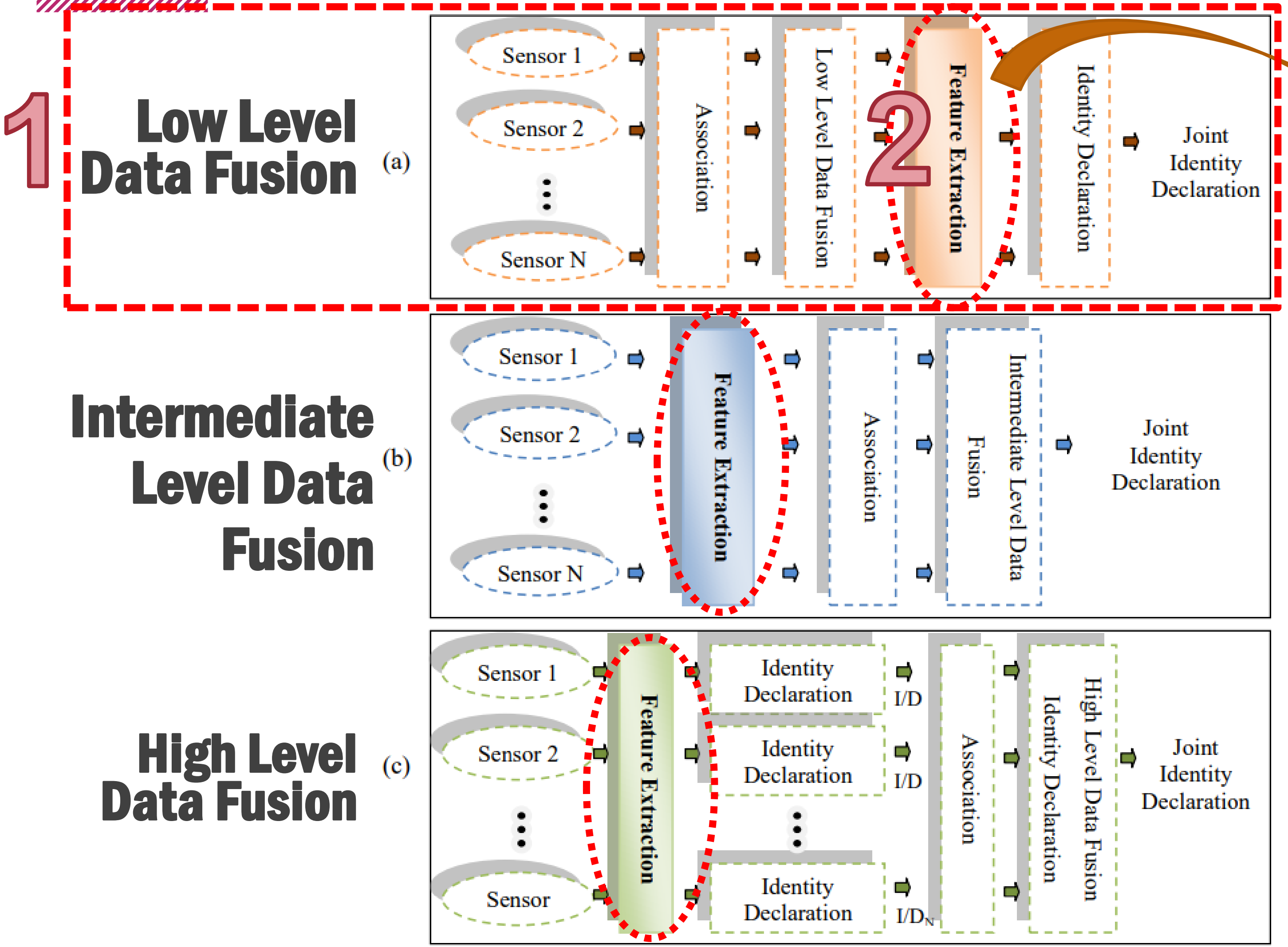


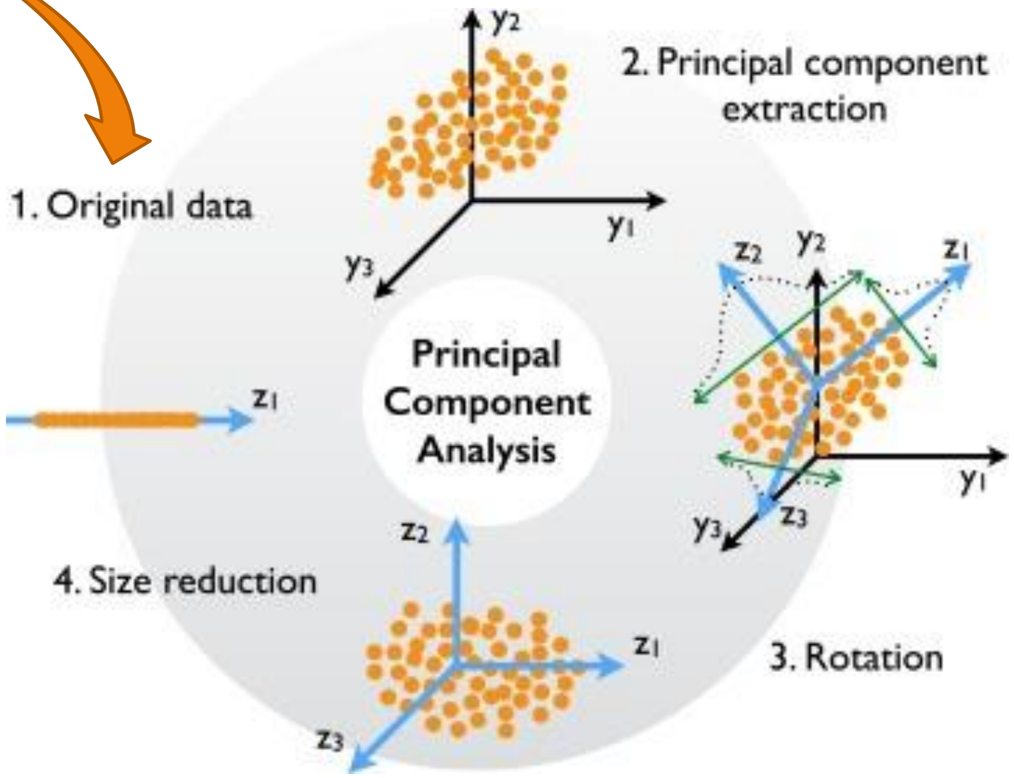
Data Analysis System

PROBLEM
Conventional Feature Extraction

Drawback:-
- **Discrepancy due to human fatigue/stress**
- **Time consuming**
- **Expensive & impossible for online monitoring**

Introduction (Joint Directors of Laboratories (JDL) Data Fusion Framework – Hall, 1992)



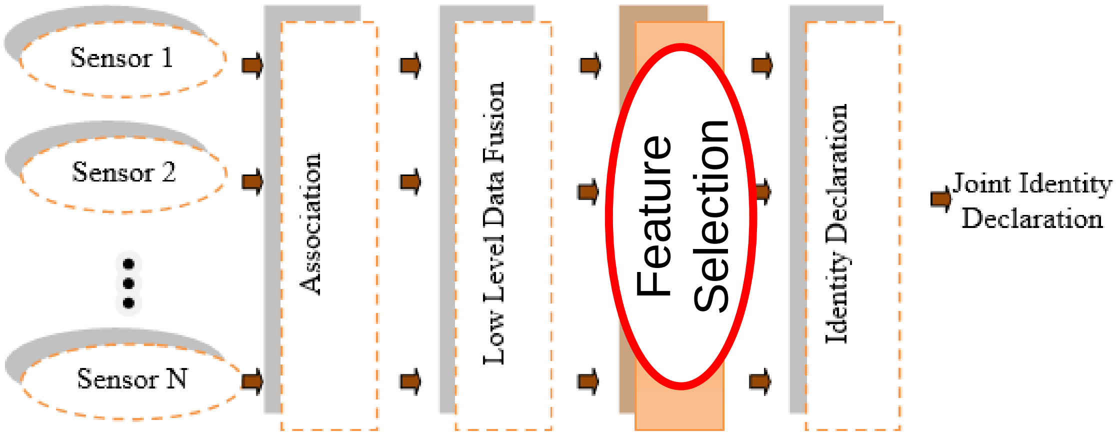


The resulting linear combinations of features are difficult to interpret especially when specific features are of interest.

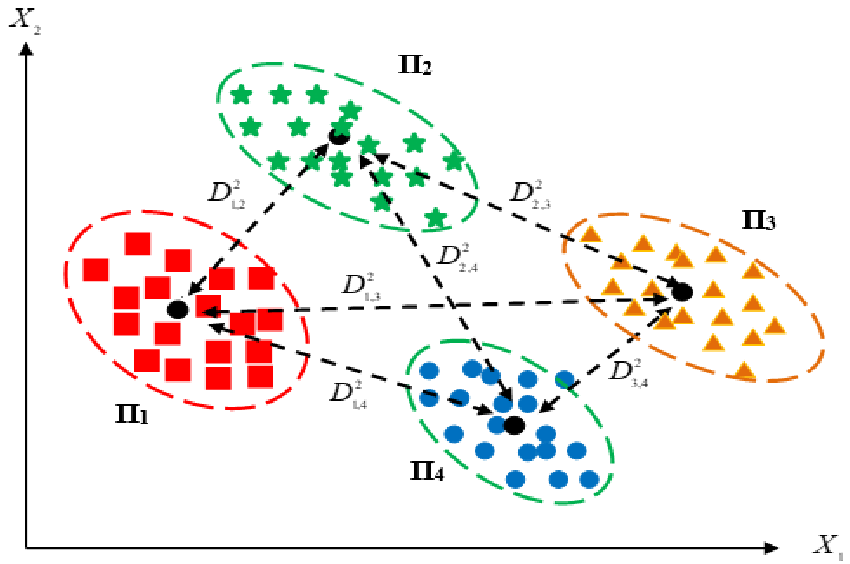
Objectives

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**.



Low Level Data Fusion



Literature Review

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.

$$\Delta^2 = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \quad (1)$$

Calculate the pair-wise distances among multi-class for a feature

$$D_{ij}^2 = \left(\overline{X}_i - \overline{X}_j\right)^T S_{ij}^{-1} \left(\overline{X}_i - \overline{X}_j\right) \quad (2)$$

Finding the average distance for a feature

$$D_{U_{x(ip)}}^2 = \frac{1}{{}^{\pi}C_2} \sum_{i=1}^{\pi} \sum_{j=i+1}^{\pi-1} D_{ij}^2 \quad (3) \rightarrow D_{U_{x(ip)}}^2 \in [0, \infty)$$

Methodology

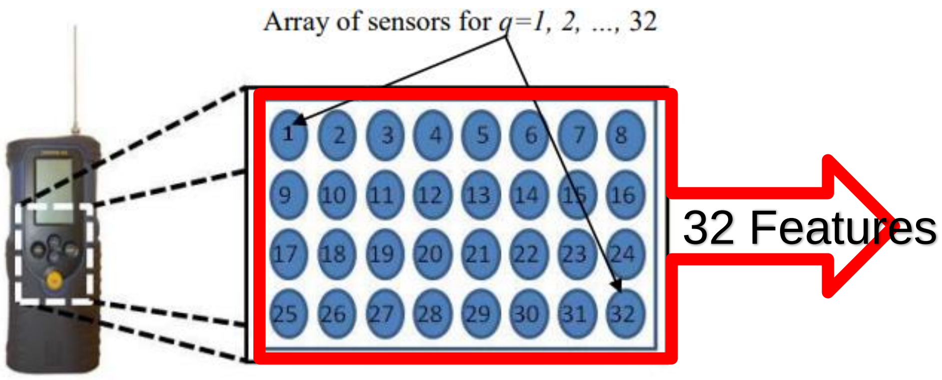


Figure 1.2: Illustration for Array of Sensors Attached in an e-nose (32-array)

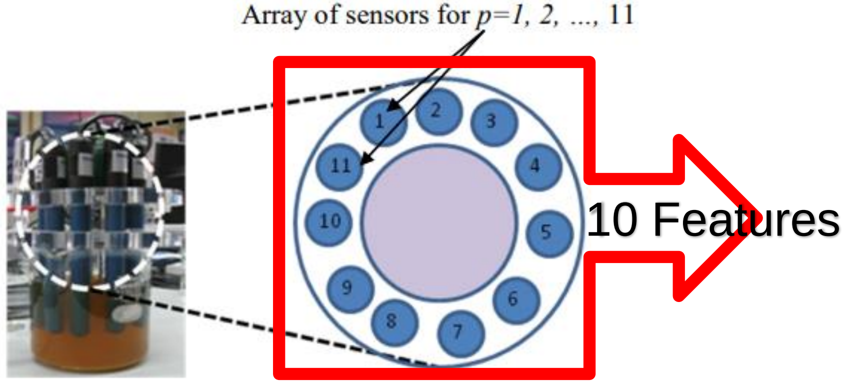


Figure 1.3: Illustration for Array of Sensors Attached in an e-tongue (11-array)

Table 3.1: Illustration of Single Sensor Data and Fused Data

Group (i)	Observation (k)	Fused Features (l)							
		X_{ik1}	X_{ik2}	...	X_{ikp}	Y_{ik1}	Y_{ik2}	...	Y_{ikq}
		e-tongue Data				e-nose Data			
1	1								
1	2								
⋮	⋮								
1	n_1								
2	1								
2	2								
	⋮								
2	n_2								
⋮	⋮								
g	1								
g	2								
⋮	⋮								
g	n_g								

Electronic Tongue
No of features = 10
No of classes = 11

Electronic Nose
No of features = 32
No of classes = 11

Low Level Data Fusion
No of features = 42
No of classes = 11
Data : 4 datasets honey (Agromas, As-Syifa, Syair Timur and Tayyibah)
Classes : pure honey (PH), cane sugar (CS), beet sugar (BS), 20BS:80PH, 40BS:60PH, 60BS:40PH, 80BS:20PH, 20CS:80PH, 40CS:60PH, 60CS:40PH, 80CS:20PH

Methodology

Input : $\mathbf{X}_{ip} = (\mathbf{X}_{ip_1} + \mathbf{X}_{ip_2})$ - matrix of observation for fused feature set

Output : Ranked fused features $\mathscr{F} = \left[\mathbf{D}^2_{U_1} > \mathbf{D}^2_{U_2} > \dots > \mathbf{D}^2_{U_p}\right]$

Step 1 : **Initialize** $\mathbf{X}_{ip} = \left[X_{i1}, \dots, X_{ip_1}, \dots, X_{ip2}, \dots, X_{ip}\right]$

Step 2 : **For** each feature $\left(X_{ip_1}, X_{ip_2}\right) \in \mathbb{Z}^p$

i. Calculate the unbounded Mahalanobis distance for ${}^{\pi}C_2$ pair-wise classes using criterion (3) for each fused feature in $\mathbb{Z}^p$

$$D^2_{ij} = \frac{\left(\overline{X}_i - \overline{X}_j\right)^2}{S_{ij}}, \left(i = 1, 2, \dots, \pi; j = i+1, \dots, \pi - 1; i \neq j\right)$$

ii. Calculate the average distance for ${}^{\pi}C_2$ pair-wise classes using criterion (4) for each fused feature in $X_{ip} \in \mathbb{Z}^p$

$$D^2_{U_{x(ip)}} = \frac{1}{{}^{\pi}C_2} \sum_{i=1}^{\pi} \sum_{j=i+1}^{\pi-1} D^2_{ij}$$

$[0, \infty)$ *unbounded Mahalanobis distance.*

iii. Store $D^2_{U_{x(ip)}}$ into $\mathscr{F} = [\ ]$

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

Step 4 : **Rank** $D^2_{U_{x(ip)}}$ (from largest to smallest values) where $\mathscr{F} = \left[\mathbf{D}^2_{U_1} > \mathbf{D}^2_{U_2} > \dots > \mathbf{D}^2_{U_p}\right]$. **Filter** $\mathscr{F}$ for best feature subset evaluation.

Single feature calculation

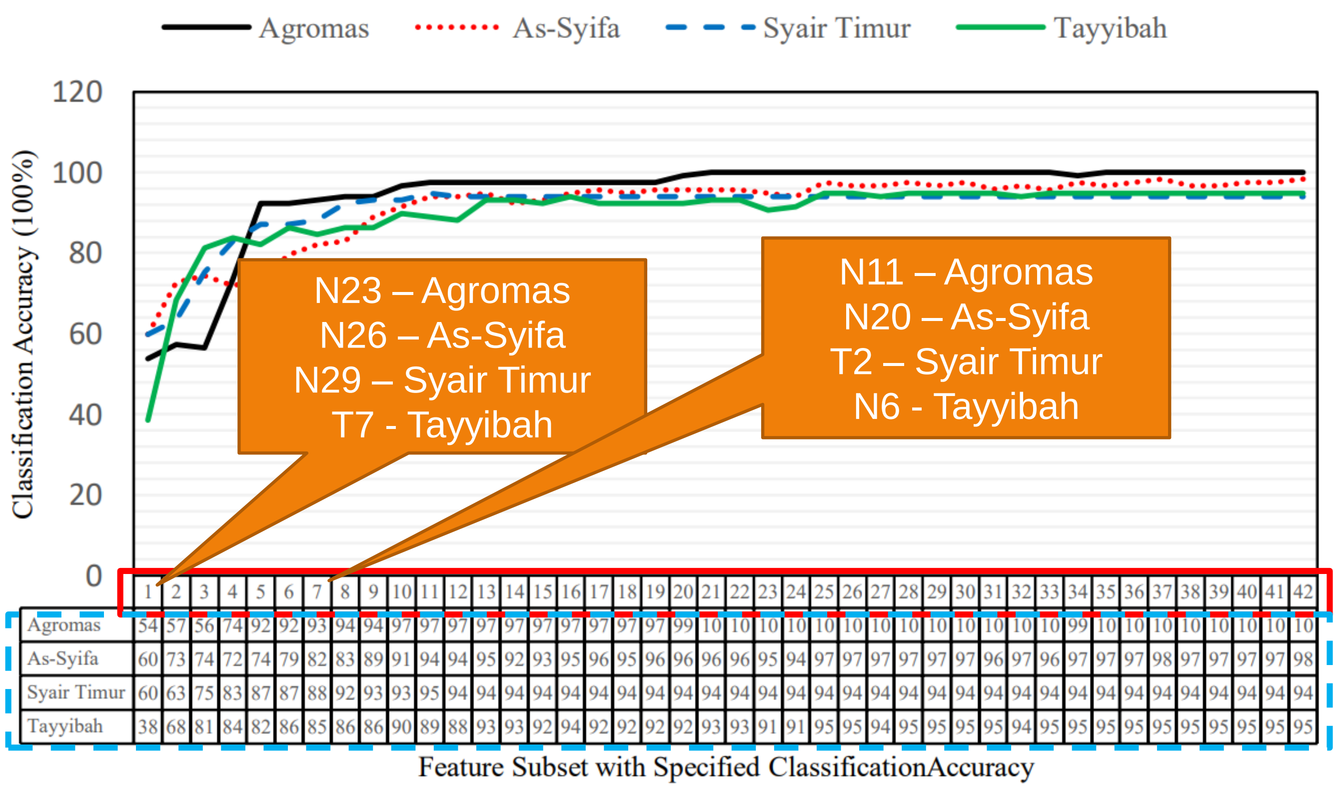
Distances for g groups produce ${}^{g}C_2$ group pairs $(g = 9)$

	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

Result & Discussion

Table 2: Ranked features for four types of honeys.

Agromas		As-Syifa		Syair Timur		Tayyibah	
Ranked Feature	Unbounded Distance	Ranked Feature	Unbounded Distance	Ranked Feature	Unbounded Distance	Ranked Feature	Unbounded Distance
N23	11,949.06	*N26*	1937.62	*N29*	12593.81	*T7*	84016
N5	11,343.20	*N5*	1872.11	*N5*	6375.09	*T2*	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	*T2*	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	*T1*	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



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42
Agromas	54	57	56	74	92	92	93	94	94	97	97	97	97	97	97	97	97	97	97	99	10	10	10	10	10	10	10	10	10	10	10	10	99	10	10	10	10	10	10	10	10	10
As-Syifa	60	73	74	72	74	79	82	83	89	91	94	94	95	92	93	95	96	95	96	96	96	96	95	94	97	97	97	97	97	97	96	97	96	97	97	97	98	97	97	97	97	98
Syair Timur	60	63	75	83	87	87	88	92	93	93	95	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94
Tayyibah	38	68	81	84	82	86	85	86	86	90	89	88	93	93	92	94	92	92	92	92	93	93	91	91	95	95	94	95	95	95	95	94	95	95	95	95	95	95	95	95	95	95

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

Conclusion

Feature selection (unbounded Mahalanobis distance) can be used to replace the conventional feature extraction for low level data fusion involving different sensor devices.

The proposed method has shown potential result of classifying pure honey from adulterated honey concentrations.

Unlike feature extraction, distance-based 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.

THANK YOU