



**Bio-Inspired Sensor Data Fusion for Herbal Tea
Flavour Assessment**

by

**Nur Zawatil Isqi Binti Zakaria
(1130610576)**

A thesis submitted in fulfilment of the requirements for the degree of
Master of Science (Mechatronic Engineering)

**School of Mechatronic Engineering
UNIVERSITI MALAYSIA PERLIS**

2017

ACKNOWLEDGEMENT

First of all, *Alhamdulillah*, all praises to Almighty God. Finally, I managed to complete my master thesis after all the hardship and tribulations during this adventurous journey with Allah blessing and guidance. In addition, I would like to take this opportunity to acknowledge and appreciate the efforts of the people in the table below that helped me during my research and the documentation of this thesis. May Allah grant them the best rank in this world and hereafter.

No.		Names	Remarks
1.	Supervisor and Co-supervisor	Prof. Dr. Ali Yeon Bin. Md. Shakaff and Dr. Ammar Bin Zakaria	Guidance, motivations and patience
2.	Department	Ministry of Education, CEASTech, School of Mechatronics Engineering, Agrotech UniMAP and Centre of graduate study	Finance, instruments, samples, guidance and knowledges sharing
3.	Parents	Zakaria Ismail and Zauyah Md Zain	Support, love, patience and finance assisted
4.	Siblings and their family	Nur Zatul-Iffah, Zulkhairi and Mohamad Zulfadzli	Supports and finance assisted
5.	Sifu	Dr. Maz Jamilah Masnan, Assoc. Prof Dr. Abu Hasan Abdullah, Abdul Halis, Rohani Farook, Azian Subari, and Syahida Sulaiman	Knowledge, patience and lunch treats
6.	Inspirations	Prof. Dr. Mohd. Noor Ahmad and Pengkalan Asam team (Wani, Siti, Mubaraq, Farhanah, Kak Aza, Kak Dayah, Dayah.and Zul)	Provide wonderful environment for my thesis writing and knowledge sharing.
7.	Special people	Dr. Latifah Munirah, Nurul Maisairah, Syahida, Hamizah, Nor Hanani, Syamimi Wahida, Nurul Aini and Nurlisa	Visiting me in hospital due to car accident and assist in my research work.
8.	Al-Fateh Educational Institute	Pak Teh Zaki Muin (Principal), Tc.Rabiha and all teachers.	Give me special leave from teaching in order to do thesis corrections.

TABLE OF CONTENTS

	PAGE
THESIS DECLARATION	i
ACKNOWLEDGEMENT	ii
TABLE OF CONTENTS	iii
LIST OF TABLES	x
LIST OF FIGURES	xvi
LIST OF ABBREVIATIONS	xix
LIST OF SYMBOLS	xxii
ABSTRAK	xxiii
ABSTRACT	xxiv
CHAPTER 1 INTRODUCTION	
1.1 Background of Flavour	1
1.2 Bio-inspired Sensor	2
1.3 Problem Statement	3
1.4 Research Objectives	4
1.5 Scope of Research	4
1.6 Contribution of Research	5
1.7 Thesis Outline	6
CHAPTER 2 LITERATURE REVIEW	
2.1 Ready to Drink Tea	7
2.1.1 Real Tea	10

2.1.2	Herbal	11
2.2	Sensory Quality	13
2.2.1	Flavour	13
2.3	Flavour assessment	15
2.3.1	Conventional Approaches	17
2.3.2	Bio-Inspired Sensor	19
2.3.2.1	Bio-Inspired Data Fusion	22
2.4	Data Processing	24
2.4.1	Data Distribution Analysis	24
2.4.1.1	Normality Test	25
2.4.2	Pre-Processing	26
2.4.2.1	Data Cleaning	27
2.4.2.2	Data Transformation	27
2.4.2.3	Data Integration	28
2.4.2.4	Dimension Reduction	28
2.4.2.4.1	Features Extraction	30
2.4.3	Classification	32
2.4.3.1	Parametric	33
2.4.3.1.1	Linear Discriminant Analysis (LDA)	33
2.4.3.2	Non-parametric	34
2.4.3.2.1	Fisher Discriminant Analysis (FDA)	35
2.4.3.2.2	K-Nearest Neighbour (KNN)	36
2.4.3.2.3	Artificial Neural Network (ANN)	37
2.4.3.2.4	Support Vector Machine (SVM)	43
2.5	Automated Quality Grading	45
2.6	Summary	47

CHAPTER 3 METHODOLOGY

3.1	Overview	48
3.2	Practical Issues	49
3.2.1	Standard Sample Preparation Procedure	49
3.2.2	Flavour preservation	51
3.2.2.1	Herbal Infusion	51
3.2.2.2	Commercialize flavour and masking agent	52
3.2.3	E-nose and E-tongue cleansing	53
3.2.4	Magnetic stirring	54
3.2.5	E-tongue storage	54
3.2.6	Sample container	54
3.3	Design of Experiment (D.O.E) and Data Collection	54
3.3.1	Design of Experiment for Electronic Nose and Electronic Tongue	55
3.4	Data Collection	57
3.4.1	Portable Electronic Nose (PEN 3)	57
3.4.1.1	Testing Condition	58
3.4.2	Electronic Tongue	59
3.4.2.1	Testing Condition	60
3.4.3	GC	61
3.4.3.1	SPME-GC Setting	61
3.4.3.2	GC/MS Setting	62
3.4.3.3	Testing Condition	62
3.5	Data Analysis	63
3.5.1	Distribution analyses	64
3.5.1.1	Normality test	64
3.5.2	Pre-Processing Technique	64
3.5.2.1	E-nose and E-tongue	65

3.5.2.1.1	Feature Extraction and Dimension Reduction	65
3.5.2.1.2	Data Fusion	66
3.5.2.1.2.1	Low Level Data Fusion	66
3.5.2.1.2.2	Intermediate level data fusion	67
3.5.2.2	GC/MS	68
3.5.3	Data division	71
3.5.4	Applied Classification Technique	74
3.5.4.1	FDA	76
3.5.4.2	KNN	77
3.5.4.2.1	Parameter optimization	77
3.5.4.3	SVM	78
3.5.4.3.1	Parameter optimization	80
3.5.4.4	PNN	81
3.5.4.4.1	Parameter optimization	81
3.6	Proposed Automatic Bio-Inspired Flavour Assessment and Grading System	82
3.7	Summary	84

CHAPTER 4 RESULT AND DISCUSSION

4.1	Background	86
4.2	Different Commercial Flavour	87
4.2.1	Normality Test	87
4.2.1.1	E-nose	87
4.2.1.2	E-tongue	88
4.2.2	Features Extraction	89
4.2.2.1	E-nose	89
4.2.2.2	E-tongue	90
4.2.3	Parameters Optimisation	91

4.2.3.1	K-nearest neighbour (KNN)	92
4.2.3.2	Support vector machine (SVM)	92
4.2.3.3	Probabilistic neural network (PNN)	93
4.2.4	Classifier Performance Evaluation	94
4.2.4.1	Low Level Data Fusion (LLDF)	94
4.2.4.2	Intermediate Level Data Fusion (ILDF)	97
4.3	Different Type of Tea and Manufacturer for <i>O.stamineus</i>	100
4.3.1	Normality Test	100
4.3.1.1	E-Nose	100
4.3.1.2	E-Tongue	101
4.3.2	Features Extraction	102
4.3.2.1	E-Nose	102
4.3.2.2	E-Tongue	103
4.3.3	Parameter Optimisation	104
4.3.3.1	K-nearest neighbour (KNN)	104
4.3.3.2	Support vector machine (SVM)	105
4.3.3.3	Probabilistic neural network (PNN)	107
4.3.4	Classifier performance evaluation	107
4.3.4.1	Low Level Data Fusion (LLDF)	107
4.3.4.2	Intermediate Level Data Fusion (ILDF)	112
4.3.4.3	Gas Chromatography Mass Spectrometry (GC/MS)	116
4.4	Different Concentration	121
4.4.1	Normality Test	121
4.4.1.1	E-nose	121
4.4.1.2	E-tongue	122
4.4.2	Features Extraction	123
4.4.2.1	E-nose	123

4.4.2.2	E-tongue	124
4.4.3	Parameter Optimisation	125
4.4.3.1	K-nearest neighbour (KNN)	125
4.4.3.2	Support vector machine (SVM)	126
4.4.3.3	Probabilistic neural network (PNN)	128
4.4.4	Classifier performance evaluation	128
4.4.4.1	Low Level Data Fusion (LLDF)	128
4.4.4.2	Intermediate Level Data Fusion (ILDF)	133
4.4.4.3	Gas Chromatography Mass Spectrometry (GC/MS)	136
4.5	Different Type Bitter Masking Agent	140
4.5.1	Normality Test	141
4.5.1.1	E-nose	141
4.5.1.2	E-tongue	141
4.5.2	Features Extraction	142
4.5.2.1	E-nose	142
4.5.2.2	E-tongue	143
4.5.3	Parameter Optimisation	145
4.5.3.1	K-nearest neighbour (KNN)	145
4.5.3.2	Support vector machine (SVM)	146
4.5.3.3	Probabilistic neural network (PNN)	147
4.5.4	Classifier performance evaluation	147
4.5.4.1	Low Level Data Fusion (LLDF)	147
4.5.4.2	Intermediate Level Data Fusion (ILDF)	151
4.6	Proposed Automatic Bio-Inspired Flavour Assessment and Grading System	156
4.7	Discussion	161

CHAPTER 5 CONCLUSION

5.1 Conclusion 162

5.2 Research Achievements 164

5.3 Future Development 165

REFERENCES 166

APPENDIX-A 183

APPENDIX-B 186

APPENDIX-C 205

LIST OF AWARDS AND PUBLICATIONS 214

©This item is protected by original copyright

LIST OF TABLES

NO.		PAGE
3.1	List of tea sample	55
3.2	List of tested PEN3 method testing for D.O.E	56
3.3	List of sensors description in E-nose.	58
3.4	List of sensors and description in E-tongue.	60
3.5	Total observations used for training and testing stage.	73
3.6	List of samples for each dataset and amounts.	73
3.7	List of parameter setting for FDA.	77
3.8	List of parameters and tested value	78
3.9	List of kernel applied in SVM	80
3.10	List of parameters and tested value	81
4.1	Standardized canonical discriminant function coefficients	90
4.2	Standardized canonical discriminant function coefficients	91
4.3	Optimum parameter for KNN	92
4.4	Results for SVM optimisation parameter for different commercial flavour	93
4.5	Confusion matrix and performance percentage for linear discriminant analysis (FDA) technique	95
4.6	Confusion matrix and performance percentage for k-nearest neighbour (KNN) technique	95
4.7	Confusion matrix and performance percentage for support vector machine (SVM) technique	96
4.8	Confusion matrix and performance percentage for probabilistic neural network (PNN) technique	96
4.9	Summary for percentage accuracy and misclassification error for all methods	97
4.10	Confusion matrix and performance percentage for linear discriminant analysis (FDA) technique	97

4.11	Confusion matrix and performance percentage for k-nearest neighbour (KNN) technique	98
4.12	Confusion matrix and performance percentage for support vector machine (SVM) technique	98
4.13	Confusion matrix and performance percentage for probabilistic neural network (PNN) technique	99
4.14	Summary for percentage accuracy and misclassification error for all methods	99
4.15	Standardized canonical discriminant function coefficients	103
4.16	Standardized canonical discriminant function coefficients	104
4.17	Optimum parameter for KNN	105
4.18	SVM parameter optimisation result for LLDF and ILDF	106
4.19	SVM parameter optimisation result for GC/MS TIC	106
4.20	Confusion matrix and performance percentage for linear discriminant analysis (FDA) technique	108
4.21	Confusion matrix and performance percentage for k-nearest neighbour (KNN) technique	108
4.22	Confusion matrix and performance percentage for support vector machine (SVM) technique	109
4.23	Confusion matrix and performance percentage for probabilistic neural network (PNN) technique	110
4.24	Comparison of different classifiers' sensitivity and specificity for different types and brands dataset in LLDF framework (in percentage)	111
4.25	Summary for percentage accuracy and misclassification error for all methods	112
4.26	Confusion matrix and performance percentage for linear discriminant analysis (FDA) technique	113
4.27	Confusion matrix and performance percentage for k-nearest neighbour (KNN) technique	113
4.28	Confusion matrix and performance percentage for support vector machine (SVM) technique	114

4.29	Confusion matrix and performance percentage for probabilistic neural network (PNN) technique	114
4.30	Comparison of different type of classifiers' sensitivity and specificity for each class of different types and brands dataset in ILDF framework (in percentage)	115
4.31	Summary for percentage accuracy and misclassification error for all methods	116
4.32	Confusion matrix and performance percentage for linear discriminant analysis (FDA) technique	117
4.33	Confusion matrix and performance percentage for k-nearest neighbour (KNN) technique	118
4.34	Confusion matrix and performance percentage for support vector machine (SVM) technique	118
4.35	Confusion matrix and performance percentage for probabilistic neural network (PNN) technique	119
4.36	Comparison of different type of classifiers' sensitivity and specificity for each class of different types and brands dataset using GC/MS (in percentage)	120
4.37	Summary for percentage accuracy and misclassification error for all methods	120
4.38	Standardized canonical discriminant function coefficients	124
4.39	Standardized canonical discriminant function coefficients	125
4.40	Optimum parameter for KNN	126
4.41	SVM parameter optimisation result for LLDF and ILDF	127
4.42	SVM parameter optimisation result for GC/MS TIC	127
4.43	Confusion matrix and performance percentage for linear discriminant analysis (FDA) technique	129
4.44	Confusion matrix and performance percentage for k-nearest neighbour (KNN) technique	130
4.45	Confusion matrix and performance percentage for support vector machine (SVM) technique	130

4.46	Confusion matrix and performance percentage for probabilistic neural network (PNN) technique	131
4.47	Comparison of different type of classifiers' sensitivity and specificity for different concentrations (Agro) dataset in LLDF framework (in percentage)	132
4.48	Summary for percentage accuracy and misclassification error for all methods	132
4.49	Confusion matrix and performance percentage for linear discriminant analysis (FDA) technique	133
4.50	Confusion matrix and performance percentage for k-nearest neighbour (KNN) technique	134
4.51	Confusion matrix and performance percentage for support vector machine (SVM) technique	134
4.52	Confusion matrix and performance percentage for probabilistic neural network (PNN) technique	135
4.53	Comparison of different type of classifiers' sensitivity and specificity for each class of different types and brands dataset in ILDF framework (in percentage)	135
4.54	Summary for percentage accuracy and misclassification error for all methods	136
4.55	Confusion matrix and performance percentage for linear discriminant analysis (FDA) technique	137
4.56	Confusion matrix and performance percentage for k-nearest neighbour (KNN) technique	138
4.57	Confusion matrix and performance percentage for support vector machine (SVM) technique	138
4.58	Confusion matrix and performance percentage for probability neural network (PNN) technique	139
4.59	Comparison of different type of classifiers' sensitivity and specificity for each class of different concentrations dataset using GC/MS (in percentage)	139
4.60	Summary for percentage accuracy and misclassification error for all methods	140
4.61	Standardized canonical discriminant function coefficients	143
4.62	Standardized canonical discriminant function coefficients	145

4.63	Optimum parameter for KNN	145
4.64	SVM parameter optimisation result for LLDF and ILDF	146
4.65	Confusion matrix and performance percentage for linear discriminant analysis (FDA) technique	149
4.66	Confusion matrix and performance percentage for k-nearest neighbour (KNN) technique	149
4.67	Confusion matrix and performance percentage for support vector machine (SVM) technique	150
4.68	Confusion matrix and performance percentage for probabilistic neural network (PNN) technique	150
4.69	Comparison of different type of classifiers' sensitivity and specificity for each class of different types of masking agent dataset in LLDF framework (in percentage)	151
4.70	Summary for percentage accuracy and misclassification error for all methods	151
4.71	Confusion matrix and performance percentage for linear discriminant analysis (FDA) technique	153
4.72	Confusion matrix and performance percentage for k-nearest neighbour (KNN) technique	153
4.73	Confusion matrix and performance percentage for support vector machine (SVM) technique	154
4.74	Confusion matrix and performance percentage for probabilistic neural network (PNN) technique	154
4.75	Comparison of different type of classifiers' sensitivity and specificity for each class of different types of masking agent dataset in ILDF framework (in percentage)	155
4.76	Summary for percentage accuracy and misclassification error for all methods	156
A.1	List of PEN3 parameters	183
B.1	Tests of Normality for different commercial brand E-nose dataset	186
B.2	Descriptive Statistics for different commercial brand E-nose dataset	187

B.3	Tests of Normality for different commercial brand E-tongue dataset	188
B.4	Descriptive Statistics for different commercial brand E-tongue dataset	190
B.5	Tests of Normality for different type of tea and brand E-nose dataset	190
B.6	Descriptive Statistics for different type of tea and brand E-nose dataset	192
B.7	Tests of Normality for different type of tea and brand E-tongue dataset	193
B.8	Descriptive Statistics for different type of tea and brand E-tongue dataset	195
B.9	Tests of Normality for different concentration E-nose dataset	195
B.10	Descriptive Statistics for E-nose dataset	197
B.11	Tests of Normality for different concentration E-tongue dataset	197
B.12	Descriptive Statistics for different concentration E-tongue dataset	199
B.13	Tests of Normality for different masking agent E-nose dataset	200
B.14	Descriptive Statistics for different masking agent E-nose dataset	202
B.15	Tests of Normality for different masking agent E-tongue dataset (Shapiro-Wilk)	202
B.16	Descriptive Statistics for different masking agent E-tongue dataset	204

LIST OF FIGURES

NO.		PAGE
1.1	Comparison between human olfactory system and electronic nose.	2
2.1	Beverages division (Fellows & Hampton, 1992)	8
2.2	Tea manufacture — major steps and corresponding type of tea (K. Wang et al., 2011)	10
2.3	Flavour detection mechanism -Illustration by Lydia V. Kibiuk, Baltimore, MD; Devon Stuart, Harrisburg, PA (Neuroscienc, 2012)	15
2.4	Division in classification or pattern recognition	32
2.5	Architecture for (a) SOM, (b) MLP, (c) RBF and (c) PNN. (Perceptron & Networks, 2015)	40
3.1	Flow chart for research methodology	48
3.2	E-nose setup for volatile compound evaluation	59
3.3	E-tongue setup for non-volatile compound evaluation	61
3.4	Flow for data analysis	63
3.5	Illustration of fusing data in low level data fusion	67
3.6	Illustration of fusing data in intermediate level data fusion	68
3.7	Pre-processing flow for GC/MS	69
3.8	Plot for (a) original and (b) processed relative intensity GC/MS-TIC data	71
3.9	E-nose characteristic response curve of 10 array sensor values during tea infusion sample measurement	72
3.10	E-tongue characteristic response curve of 11 array sensor values during tea infusion sample measurement	72
3.11	Process flow for parameter optimization for each classifier.	75
3.12	Bio-inspired flavour assessment and grading system via data fusion.	83
3.13	Grading algorithm	84
4.1	Scatter plot for different types and brands	158

4.2	Scatter plot for different concentrations.	159
4.3	Scatter plot for different masking agents	160
A.1	LDA plot for different measurement time; (a) 30s, (b) 40s and (c) 60s	185
B.2	Normal probability plot for different commercial brand E-nose dataset	187
B.3	Normal probability plot for different commercial brand E-tongue dataset	190
B.4	Normal probability plot for different type of tea and brand E-nose dataset	192
B.5	Normal probability plot for different type of tea and brand E-tongue dataset	194
B.6	Normal probability plot for different concentration E-nose dataset	197
B.7	Normal probability plot for different concentration E-tongue dataset	199
B.8	Normal probability plot for different masking agent E-nose dataset	201
B.9	Normal probability plot for different masking agent E-tongue dataset	204
C.10	Results of PNN parameter optimization for different commercial flavour dataset.	205
C.11	Results of PNN parameter optimization for different types of tea and manufacturer dataset.	206
C.12	Results of PNN parameter optimization for different concentration dataset.	206
C.13	Results of PNN parameter optimization for different type bitter masking agent dataset.	206
C.14	Results of PNN parameter optimization for different commercial flavour dataset.	207
C.15	Results of PNN parameter optimization for different types of tea and manufacturer dataset.	207
C.16	Results of PNN parameter optimization for different concentration dataset.	208
C.17	Results of PNN parameter optimization for different type bitter masking agent dataset.	208

C.18	Results of KNN parameter optimization for different commercial flavour dataset.	209
C.19	Results of KNN parameter optimization for different types of tea and manufacturer dataset.	209
C.20	Results of KNN parameter optimization for different concentration dataset.	210
C.21	Results of KNN parameter optimization for different type bitter masking agent dataset.	210
C.22	Results of KNN parameter optimization for different commercial flavour dataset.	211
C.23	Results of KNN parameter optimization for different types of tea and manufacturer dataset.	211
C.24	Results of KNN parameter optimization for different concentration dataset.	212
C.25	Results of KNN parameter optimization for different type bitter masking agent dataset.	212
C.26	Results of KNN parameter optimization for different types of tea and manufacturer dataset.	213
C.27	Results of KNN parameter optimization for different concentration dataset.	213

LIST OF ABBREVIATIONS

AD	Anderson-Darling
BP	Back propagation
CDA	Canonical discriminant analysis
CDF	Computable document format
D.O.E	Design of experiment
E.I	Electron-ionization
EEG	Electroencephologram
E-nose	Electronic nose
E-tongue	Electronic tongue
FE	Features extraction
FID	Flame ionisation detector
FDA	Linear Discriminant Analysis with fisher criterion
FS	Features selection
FTIR	Fourier transform IR spectroscopy
GC	Gas Chromatography
GC-O	GC-Olfactometry
GDA	Generalized Discriminant Analysis
GRNN	General regression neural network
HLDF	High Level Data Fusion
HPLC	High performance liquid chromatography
HS	Headspace
ILDF	Intermediate Level Data Fusion
KNN	K-nearest neighbour

KS	Kolmogrov-Smirnov
LDA	Linear discriminant analysis
LF	Lilliefors
LLDF	Low Level Data Fusion
LS	Least Squares
MLP	Multi-Layer Perceptron
MOS	Metal oxide semiconductor
MS	Mass Spectrometry
MSDF	Multi sensor data fusion
MSE	Mean squared error
MVA	Multivariate analysis
NMR	Nuclear magnetic resonance
NN	Neural network
OAA	One against all
OA0	One against one
PCA	Principal component analysis
PEN3	Portable electronic nose
PH	Pureherb
PLS	Partial least square
PNN	Probabilistic neural network
POL	Polens
QP	Quadratic programming
RBF	Radial basis fuction
RH	Rainhill
RTD	Ready To Drink

SIM	Selective ion monitoring
SMO	Sequential Minimal Optimization
SOM	Self-organizing map
SPME	Solid phase microextraction
SVM	Support vector machine
SW	Shapiro-Wilk
TIC	Total Ion Chromatography
V.F.C	Volatile flavour compound

©This item is protected by original copyright

LIST OF SYMBOLS

$\mu_1\mu_2$	Mean vector
C	Box constraint
C_1C_2	Covariance matrices
G/G_0	Ratio of conductance
M/Z	Mass/charge
V	Voltage
α_i	Inequalities
β	Linear model coefficient
ξ_i	Slack variable in SVM

©This item is protected by original copyright

Bio-Inspirasi Gabungan Data Sensor Untuk Penilaian Perasa Teh Herba

ABSTRAK

Produk-produk berasaskan herba menjadi amalan pengeluaran meluas di kalangan pengeluar untuk pasaran antarabangsa dan tempatan. Memandangkan bertambahnya pengeluaran bagi memenuhi permintaan pasaran, adalah sangat penting bagi pengeluar supaya memastikan produk mereka telah memenuhi kriteria dan kualiti tertentu yang telah ditetapkan oleh pengawal kualiti. Salah satu produk berasaskan herba yang terkenal ialah teh herba. Tesis ini mengkaji penilaian-penilaian rasa berdasarkan inspirasi bio dalam konteks gabungan data melibatkan e-hidung dan e-telinga. Objektif kajian ini adalah untuk mendapatkan pengelasan yang tepat bagi pelbagai jenis dan jenama teh herba, pengelasan beberapa agen 'masking' rasa dan yang terakhir pengelasan berdasarkan perbezaan kepekatan teh herba. Dalam penyelidikan ini, dua tahap gabungan data dimanfaatkan iaitu gabungan data tahap rendah (LLDF) dan gabungan data tahap pertengahan (ILDF). Empat teknik pengelasan; 'Fisher Linear Data Analysis (FDA)', 'Support Vector Machine (SVM)', 'k-Nearest Neighbour (KNN)' dan 'Probability Neural Network (PNN)' telah diuji dalam mencari pengelas terbaik bagi mencapai objektif penyelidikan. Dalam menilai prestasi pengelas, penganggar ralat berdasarkan pengesahan silang 'k-fold' dan 'leave-one-out' (LOO) telah digunakan. Pengelasan berdasarkan data GC/MS TIC turut disertakan sebagai satu perbandingan kepada prestasi klasifikasi menggunakan pendekatan-pendekatan gabungan. Secara umumnya, melalui gabungan data tahap rendah dan gabungan data tahap pertengahan, KNN mengatasi teknik pengelasan yang lain untuk tiga penilaian rasa. Bagaimanapun, keputusan-keputusan pengelasan berdasarkan data GC/MS TIC adalah berubah-ubah bagi aplikasi yang berbeza. Memandangkan KNN dapat memberikan keupayaan pengelasan yang tinggi, sistem automatik pengredan dibina berdasarkan algoritma teknik tersebut.

Bio-Inspired Sensor Data Fusion for Herbal Tea Flavour Assessment

ABSTRACT

Herbal-based products are becoming a widespread production trend among manufacturers for the domestic and international markets. As the production increases to meet the market demand, it is very crucial for the manufacturer to ensure that their products have met specific criteria and fulfil the intended quality determined by the quality controller. One of famous herbal-based product is herbal tea. This thesis investigates bio-inspired flavour assessments in a data fusion framework involving an E-nose and E-tongue. The objectives are to attain good classification of different types and brands of herbal tea, classification of different flavour masking effects and finally classification of different concentrations of herbal tea. Two data fusion levels were employed in this research, low level data fusion (LLDF) and intermediate level data fusion (ILDF). Four classification approaches; Fisher Linear Data Analysis (FDA), Support Vector Machine (SVM), k-Nearest Neighbour (KNN) and Probability Neural Network (PNN) were examined in search of the best classifier in order to achieve the research objectives. In order to evaluate the classifiers' performance, an error estimator based on k-fold cross validation and leave-one-out were applied. Classification based on GC/MS TIC data was also included as a comparison to the classification performance using fusion approaches. Generally, KNN outperformed the other classification techniques for the three flavour assessments in the low level data fusion and intermediate level data fusion. However, the classification results based on GC/MS TIC data varies in different application. Since KNN provide the highest classification performance, automatic grading system was developed based on this technique.