



UNIVERSITI TEKNOLOGI MARA

**AGARWOOD CLASSIFICATION
BASED ON ODOR PROFILE
USING INTELLIGENT SIGNAL
PROCESSING TECHNIQUE**

MUHAMMAD SHARFI BIN NAJIB

Thesis submitted in fulfillment
of the requirements for the degree of
Doctor of Philosophy

Faculty of Electrical Engineering

June 2012

ABSTRACT

This thesis presents the classification of Agarwood from Malaysia and Indonesia regions based on signal processing technique. Signal processing for the Agarwood classification is a new area and has yet been actively implemented. In this thesis, the Agarwood has been pre-identified by experts using 32 sensor arrays to measure the Agarwood odor profile. General Agarwood pattern has been plot in 2D diagram. The odor profile from different samples have been normalized and pre-processed and visualized in 3D and 2D plot to find unique patterns. The variation of patterns that has been visualized has been marked as different group samples. From 32 data sensor arrays, several significant data sensor array have been pre-processed using principal component analysis (PCA) as data reduction process. The selected data from PCA are applied as input to compute sensor centroid for k-NN and ANN model design. To test the robustness of the classification techniques, the data sets are randomized for both k-NN classifier and ANN model. The classification results of the k-NN classifier and the ANN model utilizing significant sensor centroid new features for Agarwood grades and regions. It was found that the k-NN classifier and the ANN model is able to classify 100% of Agarwood grade and region.

TABLE OF CONTENTS

	Page
CONFIRMATION BY PANEL OF EXAMINERS	ii
AUTHOR'S DECLARATION	iii
ABSTRACT	iv
ACKNOWLEDGEMENT	v
TABLE OF CONTENTS	vi
LIST OF TABLES	xi
LIST OF FIGURES	xiv
LIST OF ABBREVIATION	xx
CHAPTER ONE: INTRODUCTION	1
1.1 BACKGROUND	1
1.2 PROBLEM STATEMENT	1
1.3 RESEARCH OBJECTIVE	2
1.4 RESEARCH SCOPE	3
1.5 THESIS LAYOUT	3
CHAPTER TWO : LITERATURE REVIEW	5
2.1 INTRODUCTION	5
2.2 AGARWOOD	5
2.2.1 Agarwood Grades and Regions	6
2.2.2 Sensing and Detection Methods	7
2.2.3 Physical-based detection	11
2.2.4 Chemical-based detection	11
2.3 SIGNAL PROCESSING	13
2.3.1 E-Nose	14
2.3.2 PCA	17
2.3.3 Centroid	17
2.3.4 Intelligent Classification Technique	18

2.3.5	Classification Analysis from E-Nose Data	21
2.3.6	Data Splitting and randomization	23
2.3.7	K-Nearest Neighbor (k-NN)	25
2.3.8	Artificial Neural Network	25
2.3.9	Performance Measures of k-NN and ANN classification	28
2.3.10	Summary	28
CHAPTER THREE: THEORITICAL BACKGROUND		30
3.1	INTRODUCTION	30
3.2	DATA ANALYSIS	30
3.2.1	Principal Component Analysis	30
3.2.2	Centroid	32
3.2.3	Proximity Matrix	32
3.3	K-NEAREST NEIGHBOR	34
3.3.1	Distance Metric and Rules	35
3.4	ANN	36
3.4.1	ANN Training Algorithm	39
3.5	VALIDATION	43
3.5.1	R-Squared Error Analysis	43
3.5.1	Residual	43
3.5.2	Sum Square Error	43
3.5.3	Classification Accuracy	44
3.6	SUMMARY	44
CHAPTER FOUR: METHODOLOGY		45
4.1	INTRODUCTION	45
4.2	SAMPLE PREPARATION	45
4.2.1	Sample Background	46
4.2.2	Heating of Grinding Fresh Sample	47
4.2.3	Data Measurement	50
4.3	FEATURE EXTRACTION	50
4.3.1	Data Normalization	54
4.3.2	Data Arrangement	55

4.3.3	Data Pre-analysis	55
4.3.4	Sensor Centroid	56
4.4	CLASSIFICATION	56
4.4.1	k-NN Model	56
4.4.1.1	Data Preparation for k-NN	58
4.4.1.2	Data Randomization	60
4.4.1.3	Data Splitting	60
4.4.1.4	Parameter Optimization	60
4.4.1.5	k-NN Training	61
4.4.1.6	k-NN Testing	62
4.4.1.7	k-NN Classification Performance Measure	62
4.4.2	ANN Model	62
4.4.2.1	Data Preparation for ANN	64
4.4.2.2	Data Randomization	65
4.4.2.3	Data Splitting	65
4.4.2.4	Parameter Optimization	65
4.4.2.5	ANN Training	65
4.4.2.6	ANN Testing	66
4.4.2.7	ANN Classification Performance Measures	66
4.5	SUMMARY	67
CHAPTER FIVE: RESULTS AND DISCUSSION		68
5.1	INTRODUCTION	68
5.2	SAMPLE PREPARATION	68
5.2.1	Identification of Agarwood Sample	69
5.2.2	Grinding Fresh Sample	70
5.3	DATA COLLECTION	71
5.3.1	Agarwood Grade	71
5.3.2	Agarwood Region	72
5.4	FEATURE EXTRACTION	77
5.4.1	Data Normalization and Visualization	77
5.4.2	Data Selection	95
5.4.3	Pattern Analysis for Agarwood Grade	95

5.4.4	Pattern Analysis for Agarwood Region	97
5.4.5	Sensor Centroid for Agarwood Grades Pattern	101
5.4.6	Sensor Centroid for Agarwood Region Pattern	102
5.4.7	Summary	104
5.5	AGARWOOD GRADES CLASSIFICATIONS	104
5.5.1	k-NN model	105
5.5.2	Agarwood Grade Classification using k-NN Model	105
5.5.2.1	Agarwood Grade Classifications using k-NN Model based on 32 sensor arrays	106
5.5.2.2	Agarwood Grade Classifications using k-NN Model based on 16 sensor arrays	114
5.5.3	ANN Model for Agarwood Grade Classification	121
5.5.4	Agarwood Grade Classifications using ANN Model	122
5.5.4.1	Agarwood Grade Classifications using ANN Model based on 32 sensor arrays	123
5.5.4.2	Agarwood Grade Classifications using ANN Model based on 16 sensor arrays	133
5.5.5	Summary	142
5.6	AGARWOOD REGIONS CLASSIFICATIONS	144
5.6.1	k-NN model for Agarwood regions	144
5.6.2	Agarwood Region Classifications using k-NN Model	144
5.6.2.1	Agarwood Region Classifications using k-NN model based on 32 sensor array	145
5.6.2.2	Agarwood Region Classifications using k-NN model based on 16 sensor array	153
5.6.3	ANN Model for Agarwood Regions	161
5.6.4	Agarwood Region Classifications using ANN Model	161
5.6.4.1	Agarwood Region Classifications using ANN Model based on 32 sensor array	162
5.6.4.2	Agarwood Region Classifications using ANN Model based on 16 sensor array	172
5.6.5	Summary	182

CHAPTER SIX: CONCLUSION AND SUGGESTION FOR FUTURE WORK	
WORK	183
6.1 CONCLUSION	183
6.2 SUGGESTION FOR FUTURE WORK	184
REFERENCES	185
AUTHOR'S PROFILE	2 0

LIST OF TABLES

Tables	Title	Page
Table 2.1	Accuracy result of various classification method using different type of E-Nose	15
Table 4.1	Parameter setting of Training Method	49
Table 4.2	Data arrangement for Agarwood Grade Sensor Centroid Computation	59
Table 4.3	Data arrangement for Agarwood Region Sensor Centroid Computation	59
Table 4.4	Methods to find optimal k-NN model	61
Table 4.5	A sample of data arrangement for a region	65
Table 5.1	Component Matrix	76
Table 5.2	Sensor Resistance Response	78
Table 5.3	Sensor Resistance Response	81
Table 5.4	Agarwood feature variations form all samples (Melaka, Super Pagoh, Ulu Tembeling, Pagoh and Indonesia)	90
Table 5.5	Component Matrix of Normalised Agarwood Data	91
Table 5.6	Proximity Matrix (MELAKA sample)	93
Table 5.7	Proximity Matrix (SUPER PAGO sample)	93
Table 5.8	Proximity Matrix (ULU TEMBELING sample)	94
Table 5.9	Proximity Matrix (PAGO sample)	94
Table 5.10	Data arrangement for Sensor Centroid Computation for Normalised Agarwood Grade Sample	101
Table 5.11	Data arrangement for Sensor Centroid Computation of Agarwood Normalised Region Sample	102
Table 5.12	The k-NN Classification Matrix Training Data for Agarwood Grade using 32-sensor array	111
Table 5.13	The k-NN Classification Matrix Testing Data for Agarwood Grade using 32-sensor array	111
Table 5.14	Summary of Performance Measures of the k-NN Classification for Agarwood grade randomised data using	113

	32-sensor array	
Table 5.15	The k-NN Classification Matrix Training data of the k-NN Classification for Agarwood Grade using 16-sensor array	119
Table 5.16	The k-NN Classification Matrix Testing data of the k-NN Classification for Agarwood Grade using 16-sensor array	119
Table 5.17	Summary of Performance Measures of the k-NN Classification of Agarwood grade randomised data using 16-sensor array	121
Table 5.18	The ANN Classification Matrix for Agarwood Grade Training Data using 32-sensor array	130
Table 5.19	The ANN Classification Matrix for Agarwood Grade Testing Data using 32-sensor array	130
Table 5.20	Summary of Performance Measures of the ANN Classification for Agarwood grade randomised data using 32-sensor array	132
Table 5.21	The ANN Classification Matrix for Agarwood Grade Training Data using 16-sensor array	140
Table 5.22	The ANN Classification Matrix for Agarwood Grade Testing Data using 16-sensor array	140
Table 5.23	Summary of Performance Measures of the ANN Classification for Agarwood Grade randomised data using 16-sensor array	142
Table 5.24	The k-NN Classification Matrix for Agarwood Region Training Data using 32-sensor array	150
Table 5.25	The k-NN Classification Matrix for Agarwood Region Testing Data using 32-sensor array	150
Table 5.26	Summary of Performance Measures of k-NN Classification for Agarwood Region randomised data using 32-sensor array	152
Table 5.27	The k-NN Classification Matrix for Agarwood Region Testing Data using 16-sensor array	158
Table 5.28	The k-NN Classification Matrix for Agarwood Region	158

	Testing Data using 16-sensor array	
Table 5.29	Summary of Performance Measures of the k-NN Classification for Agarwood region randomised data using 16-sensor array	160
Table 5.30	The ANN Classification Matrix for Agarwood Region Training Data using 32-sensor array	169
Table 5.31	The ANN Classification Matrix for Agarwood Region Testing Data using 32-sensor array	169
Table 5.32	Summary of Performance Measures of the ANN Classification for Agarwood Region randomised data using 32-sensor array	171
Table 5.33	The ANN Classification Matrix for Agarwood Region Training Data using 16-sensor array	179
Table 5.34	The ANN Classification Matrix for Agarwood Region Testing Data using 16-sensor array	179
Table 5.35	Summary of Performance Measures of the ANN Classification for Agarwood Region randomized data using 16-sensor array	181

LIST OF FIGURES

Figures	Title	Page
Figure 2.1	Schematic Diagram of Hydrodistillation	9
Figure 2.2	Isolation methods of essential oils from	10
Figure 2.3	Pattern Recognition Systems Process Flow [177]	22
Figure 2.4	Artificial Neural Network Architecture Back propagation (BPNN)	27
Figure 3.1	Illustration of k-NN Model [261]	34
Figure 3.2	Multilayer feed-forward network structure for single hidden layer	36
Figure 4.1	Overall Research Methodology	45
Figure 4.2	Sample Preparation	46
Figure 4.3	Instrumentation Setup	48
Figure 4.4	The Instrumentation Setup of E-Nose hardware, PC-Nose Software and the heating system	48
Figure 4.5	Feature Extraction for Sensor Centroid Computation (D) and classification of ANN input features	53
Figure 4.6	Feature Extraction using Sensor Centroid for classification of k-NN input features	54
Figure 4.7	k-NN classification	58
Figure 4.8	ANN classification	64
Figure 5.1	Samples of Agarwood	69
Figure 5.2	A sample of Agarwood from the region Ulu Tembeling and identified as A grade.	69
Figure 5.3	A sample of grinded and identified Agarwood from Ulu Tembeling (Grade A)	70
Figure 5.4	Samples of Agarwood resin	70
Figure 5.5	3D Pattern Variations of unnormalised Agarwood Samples, (a) High Grade Sample, (b) Low Grade Sample (50 data measured x 32 number of features) in overall dataset arrangement	71

Figure 5.6	Agarwood Sample from High Grade Samples, (b) Low Grade Samples	72
Figure 5.7	3D Pattern Variations of Agarwood Sample from (a) Melaka, (b) Super Pagoh, (c) Ulu Tembeling, (d) Pagoh and (e) Indonesia (50 data measured x 32 number of features) in overall dataset arrangement	73
Figure 5.8	3D Pattern Variations of Agarwood Sample from (a) Melaka, (b) Super Pagoh, (c) Ulu Tembeling, (d) Pagoh and (e) Indonesia	74
Figure 5.9	2D Pattern Variations of Agarwood Sample from (a) Melaka, (b) Super Pagoh, (c) Ulu Tembeling, (d) Pagoh and (e) Indonesia	75
Figure 5.10	A normalised sample measurements of the Agarwood from Melaka	80
Figure 5.11	Normalized sample measurements from all Agarwood samples (Melaka, Ulu Tembeling, Super Pagoh, Pagoh, Indonesia)	84
Figure 5.12	3D Pattern Variations of Agarwood Sample from (a) Melaka, (b) Super Pagoh, (c) Ulu Tembeling, (d) Pagoh and (e)Indonesia (50 data measured x 16 number of features) in overall dataset arrangement	85
Figure 5.13	3D Normalised Pattern Variations of Agarwood Sample from (a) Melaka, (b) Super Pagoh, (c) Ulu Tembeling, (d) Pagoh and (e) Indonesia	86
Figure 5.14	2D Normalised Pattern Variations of Agarwood Sample from (a) Melaka, (b) Super Pagoh, (c) Ulu Tembeling, (d) Pagoh and (e) Indonesia	87
Figure 5.15	Boxplot analysis of Normalised Agarwood Sample from Melaka	88
Figure 5.16	Boxplot analysis of Normalised Agarwood Sample from Super Pagoh	88
Figure 5.17	Boxplot analysis of Normalised Agarwood Sample from Ulu Tembeling	89

Figure 5.18	Boxplot analysis of Normalises Agarwood Sample from Pagoh	89
Figure 5.19	Boxplot analysis of Normalised Agarwood Sample from Indonesia	90
Figure 5.20	Combination of Agarwood Normalised Sample from (a) High Grade, (b) Low Grade after pre-analysis (PCA and boxplot)	96
Figure 5.21	Agarwood Normalised Sample from (a) High Grade, (b) Low Grade with Boxplot Analysis	97
Figure 5.22	3D Agarwood Normalised Sample from (a) Melaka, (b) Super Pagoh, (c) Ulu	98
Figure 5.23	Boxplot analysis of Agarwood Normalised Sample from (a) Melaka, (b) Super Pagoh, (c) Ulu Tembeling, (d) Pagoh and (e) Indonesia	100
Figure 5.24	Sensor Centroid Agarwood Normalised Sample from (a) High Grade , (b) Low Grade	101
Figure 5.25	Boxplot analysis of Agarwood Normalised Sample from (a) Melaka, (b) Super Pagoh, (c) Ulu Tembeling, (d) Pagoh and (e) Indonesia	103
Figure 5.26	The k-NN training errors with varying k versus percentage of accuracy	106
Figure 5.27	The k-NN training errors with varying distance versus percentage of accuracy	107
Figure 5.28	The k-NN training errors with varying rules versus percentage of accuracy	108
Figure 5.29	Regression Plot of Agarwood Grades Classification	109
Figure 5.30	Classification from testing data	110
Figure 5.31	Training data error percentage	111
Figure 5.32	Testing data error percentage	112
Figure 5.33	The k-NN training errors with varying k 34versus percentage of accuracy	114
Figure 5.34	The k-NN training errors with varying distance versus percentage of accuracy	115

Figure 5.35	The k-NN training errors with varying rules versus percentage of accuracy	116
Figure 5.36	Regression Plot of Agarwood Grades Classification	117
Figure 5.37	Classification from testing data	118
Figure 5.38	Training data error percentage	119
Figure 5.39	Testing data error percentage	120
Figure 5.40	The ANN final training errors with varying hidden layer size versus percentage of accuracy	123
Figure 5.41	The ANN final training errors with varying learning rate versus percentage of accuracy	124
Figure 5.42	The ANN final training errors with varying momentum rate versus percentage of accuracy	125
Figure 5.43	The ANN final training errors with varying epoch versus percentage of accuracy	126
Figure 5.44	Training mean of square error (SSE) plot over 7 epochs for networks with (a) 80; (b) 30; (c) 70; (d) 20; (e) 50; (f) 100; (g) 90; (h) 60; (i) 10; (j) 40 epoch	127
Figure 5.45	Regression Plot of Agarwood Grades Classification	128
Figure 5.46	Classification from testing data	129
Figure 5.47	Training data error percentage	130
Figure 5.48	Testing data error percentage	131
Figure 5.49	The ANN final training errors with varying hidden layer size versus percentage of accuracy	133
Figure 5.50	The ANN final training errors with 51 varying learning rate versus percentage of accuracy	134
Figure 5.51	The ANN final training errors with varying momentum rate versus percentage of accuracy	135
Figure 5.52	The ANN final training errors with varying epoch versus percentage of accuracy	136
Figure 5.53	Training mean of square error (SSE) plot over 7 epochs for networks with (a) 60; (b) 10; (c) 80; (d) 70; (e) 30; (f) 40; (g) 100; (h) 90; (i) 50; (j) 20 epoch	137
Figure 5.54	Regression Plot of Agarwood Grades Classification	138

Figure 5.55	Classification from testing data	139
Figure 5.56	Training data error percentage	140
Figure 5.57	Testing data error percentage	141
Figure 5.58	The k-NN training errors with varying k versus percentage of accuracy	145
Figure 5.59	The k-NN training errors with varying distance versus percentage of accuracy	146
Figure 5.60	The k-NN training errors with varying rules versus percentage of accuracy	147
Figure 5.61	Regression Plot of Agarwood Region Classification	148
Figure 5.62	Classification produced from testing data	149
Figure 5.63	Training data error percentage	151
Figure 5.64	Testing data error percentage	151
Figure 5.65	The k-NN training errors with varying k versus percentage of accuracy	153
Figure 5.66	The k-NN training errors with varying distance versus percentage of accuracy	154
Figure 5.67	The k-NN training errors with varying rules versus percentage of accuracy	155
Figure 5.68	Regression Plot of Agarwood Region Classification	156
Figure 5.69	Classification produced from testing data	157
Figure 5.70	Training data error percentage	159
Figure 5.71	Testing data error percentage	159
Figure 5.72	The ANN final training errors with varying hidden layer size versus percentage of accuracy	162
Figure 5.73	The ANN final training errors with varying learning rate versus percentage of accuracy	163
Figure 5.74	The ANN final training errors with varying momentum rate versus percentage of accuracy	164
Figure 5.75	The ANN final training errors with varying epoch versus percentage	165
Figure 5.76	Training mean of square error (SSE) plot over 14 epochs for networks with (a) 30; (b) 70; (c) 10; (d) 60; (e) 80; (f)	166

	100; (g) 50; (h) 90; (i) 20; (j) 40 epoch	
Figure 5.77	Regression Plot of Agarwood Region Classification	167
Figure 5.78	Classification produced by training data	168
Figure 5.79	Training data error percentage	170
Figure 5.80	Testing data error percentage	170
Figure 5.81	The ANN final training errors with varying hidden layer size versus percentage of accuracy	172
Figure 5.82	The ANN final training errors with varying learning rate versus percentage of accuracy	173
Figure 5.83	The ANN final training errors with varying momentum rate versus percentage of accuracy	174
Figure 5.84	The ANN final training errors with varying epoch versus percentage	175
Figure 5.85	Training mean of square error (SSE) plot over 15 epochs for networks with (a) 70; (b) 10; (c) 50; (d) 20; (e) 80; (f) 90; (g) 100; (h) 60; (i) 30; (j) 40 epoch	176
Figure 5.86	Regression Plot of Agarwood Region Classification	177
Figure 5.87	Classification produced by training data	178
Figure 5.88	Training data error percentage	180
Figure 5.89	Testing data error percentage	180

LIST OF ABBREVIATION

Abbreviation

2D	-	2-Dimensional
3D	-	3-Dimensional
AI	-	Artificial Intelligence
ANN	-	Artificial Neural Network
ASP	-	Advance Signal Processing
BPNN	-	Backpropagation Neural Network
CPN	-	Cohonen Propagation Neural Network
C_n	-	No. of Centroid
DM	-	Data Measured
EO	-	Essential Oil
FE	-	Feature Extraction
FFT	-	Fast Fourier Transform
F_n	-	No. of Feature
FRIM	-	Forest Research Institute Malaysia
GC-MS	-	Gas Chromatography/Mass Spectrometry
GC	-	Gas Chromatography
GC-O	-	Gas Chromatography-olfactomaetry
GC-MS-FID	-	Gas Chromatography Flame Ionization Detector
k	-	k-NN variable
k-NN	-	k-nerarest Neighbour
LDA	-	Linear Discriminant Analysis
MS	-	Mass Spectrometer
P1	-	Process 1
P2	-	Process 2
P3	-	Process 3
P4	-	Process 4
P5	-	Process 5
PAHs	-	Polycyclic Aromatics Hydrocarbons
PCA	-	Principal Component Analysis
PNN	-	Probablistic Neural Network
R^2	-	Correlation Coefficient

SC	-	Sensor Centroid
SFE	-	Super Fluid Extraction
SOP	-	Standard Operation Procedure
SSE	-	Sum Squared Error
Std _{min}	-	Minimum Standard Deviation
Std _{max}	-	Maximum Standard Deviation
<i>Str</i>	-	Strategy
SVM	-	Support Vector Machine
SVD	-	Singular Value Decomposition
TH	-	True High
TL	-	True Low
TR1	-	True Region 1
TR2	-	True Region 2
TR3	-	True Region 3
TR4	-	True Region 4
TR5	-	True Region 5
Ω	-	Ohm
Ω_{min}	-	Minimum Ohm
Ω_{max}	-	Maximum Ohm
Peak _{min}	-	Minimum Peak Value
Peak _{max}	-	Maximum Peak Value

CHAPTER ONE

INTRODUCTION

1.1 BACKGROUND

Agarwood is a type of resins which comes from Karas tree mostly found in Asia region. The tree can be obtained directly from forest or from plantations. The Karas tree that is found from forest is actually a controlled plant. The Agarwood source from the controlled plant is actually limited to a certain extent as compared to the tree that is inoculated. In order to increase the Agarwood massively, the Karas tree is planted in well-structured plantations. The harvest of the Karas tree can only be processed after several years until the Agarwood is formed in the core of the tree. The Agarwood is formed whenever the part of the Karas tree is injured. The injury process at the Karas tree occurs via two ways. The first one is through natural process and the other one through inoculation process. The rate of the Agarwood formation through the first process is slower as compared to the second one [1]. Nevertheless, the quality of the Agarwood grading depends on its unique scent, formation and characteristics of the physical of the Agarwood itself [2]. Countries that involves in the Agarwood are clustered into two categories, consumer and producer country. The consumer countries are mainly from Middle East while the producer countries are mostly from South East Asia. It is mainly used for aroma therapy and incents. As the demand on the Agarwood is getting higher, an apparent classification standard for the Agarwood in the Malaysian market will be very useful to consumers and traders in order to control its quality.

1.2 PROBLEM STATEMENT

A lot of works have been done in identifying the Agarwood. The famous method of identification of the Agarwood is through essential oil composition [2-3]. Currently, there are many research which focuses on the quality of the Agarwood

essential oil. In order to achieve the quality of the essential oil, the Agarwood wood must be identified first before being extracted to essential oil. Presently, the grading of the Agarwood quality and oils is classified based on expertise of human sense, color, chemical-based properties and density of the agarwood resin with no scientific accuracy measurement done [3]. However, odor-based approach that relies on experts is still questionable especially in terms of its data measurability and consistency. Based on that issue, this research will use scientific approach using E-Nose which will complement the existing odor-based approach (human experts). However, E-Nose may not possible to be used directly; it needs extra signal processing to process raw data statistically. The data signal processing will increase the possibility of extracting significant features from Agarwood sensor array profile. In this thesis, the intelligent signal processing model will be employed based on E-nose signal in classifying the Agarwood grades and regions. Artificial neural network (ANN) and k-nearest neighbor (k-NN) will be used as the intelligent classifier. Both classifiers have shown great performance and high rate of classification accuracy.

1.3 RESEARCH OBJECTIVE

The main aim of this thesis is to develop an intelligent classification model for the detection of the Agarwood regions and grades using signal processing technique. The proposed classification model must reflect significant achievement in classification performance in terms of classification accuracy and residual errors. The objectives of this thesis are the followings:

1. To pre-process normalized Agarwood measured data using a Principal Component Analysis (PCA) as a pattern analysis technique and a new technique based on the sensor centroid statistical technique to extract Agarwood unique features.
2. To develop an intelligent classification model using k-NN and ANN classifier based on several parameters and extracted Agarwood unique features.
3. To perform comparative study of different k-NN classifiers and ANN models with several strategies and approach in classifying Agarwood.

4. To propose a new data processing technique with a novel sensor centroid features for k-NN classifier.
5. To cross-validate the k-NN classifier and ANN model using several classification performance measures.

1.4 RESEARCH SCOPE

The scope of the thesis include data collection at Forest Research Institute Malaysia (FRIM) research laboratory under controlled condition, based on E-Nose C320 measurements, data analysis using principal component analysis (PCA) and sensor centroid feature extraction technique, classification of Agarwood grades (high and low grade) and regions using selected classifiers. The samples are limited to Malaysian (Melaka, Pagoh, Super Pagoh and Ulu Tembeling) and Indonesian samples. Each data set per sample is ten (10) from each region. The experiment is follows the registered Standard Operation Procedure (SOP) provided by FRIM [4]. The total of overall datasets consists of 10 datasets from each region multiplied by 5 sample regions multiplied by 32 sensor arrays which total of 1600 overall data. The software tool that is used in this work is MATLAB software, and Microsoft excel.

1.5 THESIS LAYOUT

This thesis is organized into six chapters. The content of each chapter is briefed as follows:

Chapter 1 introduces the background of this thesis which includes the Agarwood, its origin and clusters, the type of Agarwood classification factor, the problem being addressed and the objectives and scope of this research.

Chapter 2 reviews the literature related to the work in this thesis. It elaborates the research history in Agarwood and Signal Processing techniques.

Chapter 3 explains the fundamental aspects of data analysis, feature extraction techniques, classification techniques, error analysis, and classification accuracy in this thesis.

Chapter 4 presents the methodology applied in achieving the objectives of this thesis. It covers sample preparation, feature extraction, classification of the k-NN classifier design and the ANN model.

Chapter 5 discusses the data measurements and pre-analysis which include sample preparation, instrumentation setup, data collection, data analysis and feature extraction. The results are represented in various graphical representations. This chapter also discusses the results of Agarwood grades and regions classifications using significant features extracted using k-NN classification techniques and ANN model design. The results are also shown using several significant plots. The classification results of k-NN and ANN classifications techniques are evaluated using several performance measures.

Chapter 6 presents and discusses the conclusion for thesis with some recommended potential future improvements of the current work.

CHAPTER TWO

LITERATURE REVIEW

2.1 INTRODUCTION

This chapter consists of two main sections; literature reviews on Agarwood (Section 2.2) and the signal processing (Section 2.3). Section 2.2 is the detail reviews on the history of Agarwood grades and regions (Section 2.2.1), sensing and detection (Section 2.2.2), physical-based detection (Section 2.2.3) and chemical-based detection (Section 2.2.4). In Section 2.3, reviews on the chemical-based detection specifically on E-nose (Section 2.3.1), data processing and feature extraction using PCA (Section 2.3.2) and centroid (Section 2.3.3) are presented. Next, general classification technique, Intelligent Classification Technique (Section 2.3.4), classifications analysis from E-nose data (Section 2.3.5) is reviewed respectively. For data preparation, data splitting and randomization (Section 2.3.6) is briefly described. Specific intelligent classification technique is reviewed in K-Nearest Neighbor (k-NN) (Section 2.3.7) and Artificial Neural Network (Section 2.3.8) is reviewed on several applications based on E-nose. The last section of this chapter is reviews on Performance measure that had been done using k-NN and ANN technique (Section 2.3.9). This chapter ends with summary of the overall literature reviews.

2.2 AGARWOOD

Agarwood is a forest product where it is reported as one of the highly valuable commodity [5-6]. Agarwood has different names in different countries. Agarwood is known as Gaharu in Malaysia and Indonesia [2, 7], Jin-koh in Japan [8-9], Ch'en Hsiang in China [8-9] and Oudh in United Arab Emirates [10]. The main markets for

these products are in South and East Asia and the Middle East [11] and it also traded internationally [12].

There are many applications that involve Agarwood in various forms such as in raw Agarwood, processed Agarwood and most commonly burnt into incense especially in cultural activities, religious functions and medicine purpose [8-9]. Normally raw Agarwood is processed into sculpture, beads and essential oil or liquors [8-9], powder and timber pieces [10].

Benefits of Agarwood in medicine which has been found recently are a pharmacological effect on the central nervous system achieved by oral administration or abdominal injection [13-14] and human health [15].

2.2.1 Agarwood Grades and Regions

Agarwood may be classified into various grades: Grade A and B (high grade), C and D (low grade) [16]. While in India they are graded into; Grade I: Black Agar (high grade), Grade II: Bankang and Grade III: Pjutas, Kalaguchi and Grade IV: Dhum (Low Grade) [4]. Grade C of Agarwood is often distilled to obtain Agarwood Oils [16]. Its cost ranges from RM 30,000.00 to RM 40,000.00 per kilogram. Good quality of Agarwood which is widely used as incense [13] and usually requested from a company to produce high quality of Agarwood oil [10].

Another grade of Agarwood is low quality where the Agarwood is touched up with small layer of Agarwood powder and mixed with wax where it is then heated [10]. It costs ranges from RM 10,000.00 to RM 15,000.00 per kilogram while the low quality ones are used for essential oil extraction [2, 10].

Agarwood are graded based on several quality indicators and they are often graded according to physical properties, Agarwood formation, and its unique scent [10, 16], chemical properties [17] and country of origin [10]. Physical properties are often indicated by color, thickness and density [10].

In chemical properties, it has been known by extracting Agarwood into essential oil. Essential oils are volatile compound produced by plants for other purpose besides nutrition. It is categorized under natural product and Agarwood is one of the natural products class. There are several major classes of natural product