

## TABLE OF CONTENTS

CHAPTER NO.	TITLE	PAGE NO.
	<b>ABSTRACT</b>	<b>iii</b>
	<b>LIST OF TABLES</b>	<b>xii</b>
	<b>LIST OF FIGURES</b>	<b>xiv</b>
	<b>LIST OF SYMBOLS AND ABBREVIATIONS</b>	<b>xix</b>
<b>1</b>	<b>INTRODUCTION</b>	<b>1</b>
1.1	DATA MINING IN LARGE DATABASE ANALYSIS	1
1.2	KNOWLEDGE DISCOVERY IN LARGE DATABASE	3
1.2.1	Challenges in Data Analysis	5
1.3	GENERAL OVERVIEW OF PREPROCESSING	7
1.3.1	Types of Pre-processing Techniques	8
1.3.2	Imputation Methods	13
1.3.3	Importance of Pre-processing Techniques	13
1.4	GENERAL OVERVIEW OF FEATURE SELECTION	14
1.4.1	Types of Feature Selection	15
1.4.2	Issues in Feature Selection	17
1.4.3	Ensemble Feature Selection Approach	18
1.4.4	Importance of Feature Selection Techniques	18
1.5	GENERAL OVERVIEW OF CLASSIFICATION	19
1.5.1	Ensemble Classification	20
1.6	MOTIVATION OF ENSEMBLE CLASSIFIER	21

<b>CHAPTER NO.</b>	<b>TITLE</b>	<b>PAGE NO.</b>
	1.6.1 Homogeneous Ensembles	22
	1.6.2 Heterogeneous Ensembles	22
1.7	MOTIVATION OF RESEARCH	22
1.8	OBJECTIVE OF RESEARCH	23
1.9	PROBLEM SPECIFICATION	23
	1.9.1 Issues and Challenges to Handle Large Dataset	23
1.10	RESEARCH CONTRIBUTION AND METHODOLOGY	24
1.11	ORGANIZATION OF THE THESIS	25
<b>2</b>	<b>LITERATURE SURVEY</b>	<b>27</b>
2.1	INTRODUCTION	27
2.2	VARIOUS APPROACHES BASED ON FEATURE SELECTION	27
2.3	VARIOUS APPROACHES BASED ON CLASSIFICATION	35
	2.3.1 Ensemble Classification	41
2.4	SUMMARY	51
<b>3</b>	<b>HYBRID AND ENSEMBLE FEATURE SELECTION MODELS</b>	<b>52</b>
3.1	INTRODUCTION	52
3.2	HYBRID FEATURE SELECTION MODEL	54
	3.2.1 Kernel principal Component Analysis	54
	3.2.2 KPCA SVM Model	56

<b>CHAPTER NO.</b>	<b>TITLE</b>	<b>PAGE NO.</b>
	3.2.3 Proposed Enhanced GA for Parameter Selection of KPCA SVM Model	59
3.3	ENSEMBLE FEATURE SELECTION MODEL	60
	3.3.1 Ensemble Feature Selection Construction	61
3.4	PROPOSED ENSEMBLE FEATURE SELECTION BASED ON HGPSO-HABC	64
	3.4.1 Particle Swarm Optimization (PSO)	65
	3.4.1.1 Genetic algorithm with Euclidean data distance	65
	3.4.2 Hybrid Artificial Bee Colony Algorithm (HABCA)	68
3.5	SIMULATION RESULTS	72
	3.5.1 Hybrid Feature Selection Model	73
	3.5.2 Ensemble Feature Selection Model	75
3.6	SUMMARY	79
<b>4</b>	<b>FUZZY BASED ENSEMBLE CLASSIFICATION MODEL ON LARGE DATASETS</b>	<b>80</b>
	4.1 INTRODUCTION	80
	4.2 FUZZY BASED CLASSIFICATION MODEL	81
	4.2.1 Adaptive Neuro Fuzzy Interference System (ANFIS)	81
	4.2.2 Fuzzy rule based Classifier	84
4.3	PROPOSED ENSEMBLE BASED CLASSIFICATION APPROACH	85
	4.3.1 Proposed Homogeneous ensemble classifier (Fuzzy k-Nearest Neighbor Algorithm (FKNN))	86

<b>CHAPTER NO.</b>	<b>TITLE</b>	<b>PAGE NO.</b>
4.3.2	Proposed Homogeneous Ensemble Classifier (Fuzzy Rough Positive Region based Nearest Neighbour (FRPRNN))	88
4.3.3	Proposed Heterogeneous Ensemble classification Model	91
4.3.3.1	Adaptive Neuro Fuzzy Inference System(ANFIS) Classifier	93
4.3.3.1.1	Architecture of ANFIS	93
4.3.3.1.2	Learning algorithm of ANFIS	95
4.3.3.2	Fuzzy Rule based classifier	101
4.3.3.3	Improved Fuzzy Rule based classifier	103
4.4	EXPERIMENTAL RESULTS	107
4.4.1	Performance on Classification Process	107
4.4.1.1	For homogenous Classification Process	108
4.4.1.2	For heterogeneous classification process	109
4.5	SUMMARY	114
<b>5</b>	<b>INTELLIGENT ENSEMBLE CLASSIFICATION MODEL FOR MEDICAL DIAGNOSIS</b>	<b>115</b>
5.1	INTRODUCTION	115
5.2	PROPOSED METHODOLOGY	118
5.2.1	Proposed Preprocessing Technique	118
5.2.2	Feature Selection Ensemble	129
5.2.2.1	Ensemble Construction	132

<b>CHAPTER NO.</b>	<b>TITLE</b>	<b>PAGE NO.</b>
	5.2.2.2 Feature selection approaches	133
	5.2.2.2.1 Particle swarm optimization (PSO)	133
	5.2.2.2.2 Hybrid artificial bee colony algorithm (HABCA)	138
5.2.3	Homogenous Ensemble Classifier	144
5.2.4	Heterogeneous Ensemble Classifier	147
	5.2.4.1 Fuzzy K nearest neighbour	148
	5.2.4.2 Adaptive Neuro Fuzzy Inference System (ANFIS) Classifier	148
	5.2.4.3 Fuzzy rule based classifier	152
5.3	SUMMARY	154
<b>6</b>	<b>EXPERIMENTAL RESULTS</b>	<b>155</b>
6.1	INTRODUCTION	155
6.2	PERFORMANCE EVALUATION OF FEATURE SELECTION PROCESS	163
	6.2.1 Running Time (Seconds) for Feature Selection Algorithm on UCI Data	165
	6.2.2 Feature Selection Accuracy	166
6.3	PERFORMANCE EVALUATION ON CLASSIFICATION PROCESS	168
	6.3.1 Average Classification Results of Homogenous and Heterogeneous Results	170
	6.3.1.1 Average mean square error	178
	6.3.1.2 Time convergence results	182
6.4	SUMMARY	186

<b>CHAPTER NO.</b>	<b>TITLE</b>	<b>PAGE NO.</b>
<b>7</b>	<b>CONCLUSION AND FUTURE WORK</b>	<b>187</b>
7.1	CONCLUSION	187
7.2	FUTURE WORK	188
	<b>REFERENCES</b>	<b>190</b>
	<b>LIST OF PUBLICATIONS</b>	