

Contents

Declaration of Authorship	i
Abstract	iii
Acknowledgements	v
Contents	vi
List of Figures	ix
List of Tables	x
Abbreviations	xi
1 Introduction	1
1.1 Introduction	1
1.2 Problem Statement	2
1.3 Research Motivation	5
1.3.1 Why Feature Selection is important	5
1.3.2 Challenges of Feature Selection	5
1.3.3 Why Binary Cuckoo Search	6
1.3.4 Why Rough Set Theory	7
1.3.5 Limitations of Existing Work	7
1.4 Research Goals	8
1.5 Research Methodology	9
1.6 Organization of the Thesis	10
2 Background	11
2.1 Classification	11
2.1.1 Data Representation	11
2.1.2 Learning and Evaluation	12
2.1.3 Classification Performance	13

2.1.4	Classification Algorithms	14
2.1.4.1	Decision Tree	14
2.1.4.2	Naïve Bayes	15
2.2	Feature Selection	15
2.2.1	General Feature Selection Steps	15
2.2.2	Filter and Wrapper approaches	17
2.3	Summary	17
3	Literature Review	19
3.1	Introduction	19
3.2	Rough Set Theory	21
3.3	Ant Colony Optimization	24
3.4	Particle Swarm Optimization	25
3.5	Artificial Bee Colony	26
3.6	Cuckoo Search	27
3.7	Comparison	31
3.8	Summary	31
4	Proposed Algorithm	33
4.1	Introduction	33
4.2	New Objective Function	35
4.2.1	Frequent values	36
4.2.2	Dependency Degree	37
4.2.3	Balancing between the Percentage of Distinct Values and Dependency Degree	37
4.2.4	Balancing Between the Quality and Number of Selected Features	38
4.3	New Initialization Mechanisms	38
4.4	New Updating Mechanisms	41
4.4.1	New Global Search	42
4.4.2	New Local Search	42
4.4.3	Global versus local Ratio (New Switching Mechanism)	43
4.5	New Stopping Criterion	43
4.6	Summary	45
5	Evaluation and Results	46
5.1	Evaluation Methodology	46
5.1.1	Datasets selection	46
5.1.2	Evaluation method	49
5.1.3	Benchmarking and Experiment Design	50
5.1.3.1	Training, Learning and Test Sets	51
5.1.3.2	Feature Reduction	52
5.2	Results and Discussion	52
5.2.1	Comparisons between MBCSFS and BCSFS	53
5.2.2	Analysis of Computational Time	57

5.2.3	Analysis of convergence	58
5.2.4	Analysis of New Objective Function "3OG"	60
5.2.5	Classification Performance Before and After MBCSFS	63
5.2.6	Comparisons between MBCSFS, PSO with CFS, and Genetic with CFS	63
5.3	Summary	66
6	Conclusion	68
6.1	Introduction	68
6.2	Contributions	69
6.2.1	Objective function	69
6.2.2	Modified Binary Cuckoo Search	69
6.3	Results	70
6.4	Limitations and Assumptions	71
6.5	Future Work	72
A	Rough Set Theory	73
B	Lévy Flights	75
C	New Stopping Criterion	77
D	Classifications of Dimensionality Reduction	79
	Bibliography	80