Contents

Certificate ii

Declaration by the Candidate iii

Copyright Transfer Certificate iv

Acknowledgements v

Contents vi

List of Figures xi

List of Tables xiii

Preface xvii

1 Introduction 1
- 1.1 Motivation . . . 1
- 1.2 Feature Selection . . . 3
  - 1.2.1 Benefits and Application of Feature Selection . . . 8
- 1.3 Thesis Structure . . . 9

2 Background 13
- 2.1 Rough Set Theory . . . 13
- 2.2 Fuzzy Set Theory . . . 15
- 2.3 Fuzzy Rough Set Theory . . . 16
- 2.4 Fuzzy Rough Set based Feature Selection . . . 17
  - 2.4.1 Related Work . . . 19
  - 2.4.2 Limitation of Fuzzy Set . . . 20

2.5 Intuitionistic Fuzzy Set 20
2.5.1 Intuitionistic Fuzzy Rough Set Theory 21
2.5.2 Related Work 22
2.6 Summary 23

3 Feature Selection Models and its Application 25
3.1 Divergence based Intuitionistic Fuzzy Rough Set Model 26
3.1.1 Feature selection using IF rough set model based on divergence measure 35
3.1.2 Experimentation 38
3.1.3 Application to tuberculosis treatment 39
3.1.3.1 Result 42
3.2 k-mean based Intuitionistic Fuzzy Rough Set Model 49
3.2.1 Feature Selection based on Intuitionistic Fuzzy Rough Set Model based on k-means 50
3.2.2 Application for Enhancing Prediction of Aptamer-protein Interacting Pairs 51
3.2.2.1 Result 52
3.3 Summary 55

4 A Fitting Model for Feature Selection 57
4.1 Fitting Model based on Intuitionistic Fuzzy Rough Set 58
4.2 Feature Selection based on Fitting Model 61
4.3 Experimentation 64
4.4 Summary 68

5 Feature Selection Model for Incomplete Data 69
5.1 Fuzzy Rough Model for Feature Selection and Missing Value Imputation 70
5.1.1 Feature Grouping 71
5.1.2 Missing Value Imputation 71
5.1.3 Search Heuristic for Finding Reduct 75
5.2 Experimentation 80
5.2.1 Results 81
5.2.1.1 By employing Parameter variation 82
By varying the level of induced missing Values 82
By varying the value of k (number of nearest neighbour for missing value imputation) 84
By varying the percentage of correlated features employed for obtaining reduced dataset using proposed approach 86
By varying the percentage of missing entries required in an instance to ignore the instance 88

5.2.1.2 Comparison with Existing Missing Value Imputation method . . . 89
5.2.1.3 Comparison with Other Feature Selection Algorithms 92
5.3 Summary . . . 93

6 Bireduct Model and its Application 95
6.0.1 Bireduct formulation . . . 96
6.1 Intuitionistic Fuzzy Bireducts for Data Reduction . . . 97
6.1.1 Intuitionistic Fuzzy Feature Selection . . . 98
6.1.2 Intuitionistic Fuzzy Instance Selection . . . 99
6.1.2.1 Method I . . . 99
6.1.2.2 Method II . . . 100
6.1.3 Simultaneous Intuitionistic Fuzzy Instance and Feature Selection 102
6.1.4 Heuristic Search Strategy for IF Bireducts . . . 105
6.2 Experimentation . . . 109
6.2.1 Results . . . 110
6.2.1.1 Using Parameter Variation . . . 111
6.2.1.2 Using Variants of Instance Selection . . . 116
6.2.1.3 Comparisons with Other FS Algorithms . . . 118
6.2.1.4 Comparison with Instance Selection and Feature Selection + Instance Selection Approaches . . . 121
6.2.1.5 Comparison with existing Bireduct approach . . . 124
6.3 Application to Cancer Treatment . . . 125
6.3.0.1 Results . . . 129
6.3.0.2 Comparison with Unreduced Dataset . . . 131
6.3.0.3 Comparison with Existing Approaches . . . 131
6.4 Summary . . . 132

7 Unsupervised Feature Selection 135
7.1 Feature Selection based on Fuzzy Rough Set in Unsupervised Domain 137
7.1.1 Feature subset quality evaluation . . . 138
7.1.1.1 Dependency measure . . . 138
7.1.2 Search strategy . . . 141
7.1.2.1 Reproduction 1 . . . 141
7.1.2.2 Reproduction 2 . . . 142
7.2 Experimentation . . . 144
7.2.1 Results . . . 145
7.2.1.1 Using variants of proposed approach . . . 146
7.2.1.2 Comparison with state of art dependency based approach . . . 147
7.2.1.3 Comparison with state of art non dependency based approach . . . 150

7.2.1.4 Comparison with supervised approach 152
7.3 Summary 153

8 Conclusion **159**
8.1 Feature Selection 160
8.2 Future Work 161

A Data Validation Techniques **163**

B Performance Evaluation Metrics **165**

C Statistical Testing **167**

Bibliography **169**