TABLE OF CONTENTS

LIST OF FIGURES ............................................................. vii

LIST OF TABLES ............................................................... xv

Chapter 1 INTRODUCTION AND SCOPE OF THE THESIS ............ 1

  1.1 Introduction ............................................................ 3
  1.2 Use of Soft Computing in Bioinformatics ............................. 5
  1.3 Molecular Biology ...................................................... 7
      1.3.1 The central dogma of molecular Biology ....................... 7
      1.3.2 Gene expression ................................................. 10
      1.3.3 Microarray technology: measurement of gene expression .... 11
  1.4 Gene Expression Data Analysis ....................................... 13
  1.5 Relevance of Fuzzy Set Theory in Bioinformatics ................. 15
  1.6 Relevance of Artificial Neural Networks in Bioinformatics ...... 17
  1.7 Clustering and Cluster Validity Indices ............................. 18
  1.8 Methodology for Identifying Some Possible Genes Mediating a Disease 19
  1.9 Description of Data Sets ............................................. 20
      1.9.1 Human lung expression data .................................... 20
      1.9.2 Human colon expression data .................................... 21
      1.9.3 Human breast cell expression data ............................... 21
      1.9.4 Human soft tissue sarcoma expression data ................... 21
1.9.5 Human lymphocytes and plasma cell expression data .......................... 22

1.10 Scope of the Thesis ................................................................. 22
   1.10.1 Comparative analysis of cluster validity indices in identifying some possible genes mediating certain cancers ................................. 22
   1.10.2 Gaussian Fuzzy Index (GFI) for cluster validation: Identification of high quality biologically enriched clusters of genes and selection of some possible genes mediating certain cancers .......................... 23
   1.10.3 Interval based fuzzy systems for identification of important genes from microarray gene expression data: Application to carcinogenic development ......................................................... 23
   1.10.4 Selection of genes mediating certain cancers using a Neuro-Fuzzy Approach ................................................................. 24
   1.10.5 Development of a fuzzy entropy based method for detecting altered gene-gene interactions in carcinogenic state ......................... 25
   1.10.6 Fuzzy Correlated Association Mining (FCAM) from microarray gene expression data: Application to various cancers ........................... 26
   1.10.7 Conclusions and scope for further research .............................. 27

Chapter 2 COMPARATIVE ANALYSIS OF CLUSTER VALIDITY INDICES IN IDENTIFYING SOME POSSIBLE GENES MEDIATING CERTAIN CANCERS ................................. 29

   2.1 Introduction ........................................................................... 31
   2.2 Method for comparing cluster validity indices ......................... 32
      2.2.1 Algorithm ...................................................................... 32
   2.3 Results and Discussion ............................................................ 37
      2.3.1 Comparison using pathway database ................................. 37
      2.3.2 Using functional enrichment ......................................... 41
2.3.3 Comparative results using $z$-score ............................................. 43
2.3.4 Comparative analysis ............................................................ 44
2.3.5 Selection of some possible genes mediating certain cancers ....... 45
2.4 Conclusion ................................................................................. 46

Chapter 3 GAUSSIAN FUZZY INDEX ($GFI$) FOR CLUSTER VALIDATION: IDENTIFICATION OF HIGH QUALITY BIOLOGICALLY ENRICHED CLUSTERS OF GENES AND SELECTION OF SOME POSSIBLE GENES MEDIATING CERTAIN CANCERS 49

3.1 Introduction ................................................................................. 51
3.2 Methodology .............................................................................. 53
  3.2.1 Gaussian Fuzzy Index ($GFI$) for cluster validation ................. 53
3.3 Results ....................................................................................... 55
  3.3.1 Description of the datasets ...................................................... 56
  3.3.2 Results on synthetic and Iris data ............................................ 56
  3.3.3 Comparative results using pathway database ......................... 57
  3.3.4 Comparative results using functional enrichment ................. 58
  3.3.5 Comparative results using $z$-score ......................................... 62
  3.3.6 Comparative results using pathway database, functional enrich-
      ment and $z$-score altogether .................................................... 63
  3.3.7 Selection of some possible genes mediating certain cancers .... 63
3.4 Conclusions ................................................................................. 66

Chapter 4 INTERVAL BASED FUZZY SYSTEMS FOR IDENTIFICATION OF IMPORTANT GENES FROM MICROARRAY GENE EXPRESSION DATA: APPLICATION TO CARCINOGENIC DEVELOPMENT ........................................... 69
4.1 Introduction ......................................................... 71
4.2 Methodology ......................................................... 73
  4.2.1 Linguistic Fuzzy Rule Generation and Grouping (LFRGG) .... 73
  4.2.2 Expression Interval based Rule Generation and Grouping (EIRGG) 78
4.3 Results and Discussion ........................................... 80
  4.3.1 Analysis of the results using LFRGG and EIRGG ............... 81
  4.3.2 Comparative analysis of LFRGG and EIRGG with other existing
       methods ............................................................. 82
  4.3.3 Validation of the results obtained by both LFRGG and EIRGG ... 87
4.4 Conclusions ......................................................... 89

Chapter 5  SELECTION OF GENES MEDIATING CERTAIN CANCERS,
USING A NEURO-FUZZY APPROACH  ................................ 91

5.1 Introduction ......................................................... 93
5.2 Some Existing Methods ........................................... 95
5.3 Proposed Methodology ............................................ 98
  5.3.1 Grouping of genes ............................................. 99
  5.3.2 Generation of data ............................................ 102
  5.3.3 Neuro-Fuzzy Supervised and Unsupervised models [146, 147, 148] 103
  5.3.4 Algorithm-NFS .............................................. 105
  5.3.5 Algorithm-NFU .............................................. 110
  5.3.6 Selection of the most important group .......................... 111
  5.3.7 Selection of important genes from the most important group ... 112
5.4 Results and Discussion ........................................... 112
  5.4.1 Comparison and validation of results ........................ 114
5.5 Conclusions ......................................................... 134
Chapter 6 DEVELOPMENT OF A FUZZY ENTROPY BASED METHOD FOR DETECTING ALTERED GENE-GENE INTERACTIONS IN CARCINOGENIC STATE

6.1 Introduction

6.2 Methodology

6.2.1 Step 1 – Measuring information content (entropy) of a gene

6.2.2 Step 2 – Measuring conditional entropy of a gene on another

6.2.3 Step 3 – Measuring information gain and building gene dependency matrix

6.2.4 Step 4 – Quantizing the elements of gene dependency matrix

6.2.5 Step 5 – Building of gene dependency network (GDN) and altered gene dependency network (AGDN)

6.2.6 Step 6 – Selection of Influential genes from altered gene dependency network (AGDN)

6.3 Results

6.3.1 Analysis of the results

6.3.2 Validation of results

6.4 Conclusion

Chapter 7 FUZZY CORRELATED ASSOCIATION MINING (FCAM) FROM MICROARRAY GENE EXPRESSION DATA: APPLICATION TO VARIOUS CANCERS

7.1 Introduction

7.2 Materials and Methods

7.2.1 Mathematical Preliminaries

7.2.2 Algorithm

7.2.3 Altered set of associations from normal to diseased state