Table of Contents

Declara on II
Cer cate III
Abstract IV
Preface VII
Acknowledgement IX
List of Tables XVII
List of Figures XIX
List of Appendices XXI

Chapter 1

INTRODUCTION and OBJECTIVES

1.1. Introduction..... 1
1.2. Decision Support System in Disease Diagnosis..... 5
1.2.1. Needs..... 5
1.2.2 Usefulness of DSS..... 6
1.3. Objectives of the Study..... 7
1.3.1. Objectives..... 7
1.3.2. Area of the Study..... 8
1.4. Overview of Breast Cancer Disease..... 9
1.5. Overview of Cardiac Arrhythmia..... 10
1.6. Overview of Hepatitis Disease..... 13
1.7. Bene t Expected from the Proposed Study..... 14
1.8 Summary of the Research Work..... 15
References..... 18

Chapter 2

MEDICAL SCENARIO ON BREAST CANCER, CARDIAC ARRHYTHMIA AND HEPATITIS DISEASE... 19

2.1. Breast Cancer Domain..... 19
2.1.1. Introduction..... 19
2.1.2. Mortality Rate..... 19
2.1.3. Mortality Rate in India and Global Comparison..... 21

2.1.4. Reasons of Breast Cancer..... 23
2.1.5 Issues Related to Breast Cancer Diagnosis and Strategies to Overcome.. 24
2.2. Cardiac Arrhythmia Domain..... 27
2.2.1. Introduction..... 27
2.2.2. Mortality Rate..... 29
2.2.3. Mortality Rate in India..... 31
2.2.4. Reasons of Cardiac Arrhythmia..... 31
2.2.5. Issues Related to Cardiac Arrhythmia Diagnosis..... 32
2.3. Hepatitis Disease Domain..... 33
2.3.1. Introduction..... 33
2.3.2. Mortality Rate..... 33
2.3.3. Mortality Rate in India..... 36
2.3.4. Reasons of Hepa s..... 37
2.3.5. Issues Related to Hepatitis Disease Diagnosis..... 38
2.4. Issues Related to Healthcare Service in India..... 39
2.5. Needs of Alternative Health Care Management..... 41
2.6. Conclusion..... 41
Reference..... 42

Chapter 3

So Compu ng Techniques in Medical Diagnosis..... 44

3.1. Introduction..... 44
3.2. So Computing..... 44
3.2.1. Fuzzy Expert System..... 45
3.2.2. Genetic Algorithm..... 49
3.2.3. Artificial Neural Network..... 54
3.2.3.1. Ac va on Func on..... 57
3.2.3.2. Types of Architecture of Neural Network 60
3.2.3.3. Training or Learning Algorithms..... 61
3.2.3.4. Incremental Backpropagation Learning Network... 62
3.2.3.5. Levenberg-Marquardt Algorithm..... 63
3.2.3.6. Implementation of ANN..... 64
3.3. Conclusion..... 67
Reference..... 68

Chapter 4

PREPROCESSING OF DATA AND HANDLING MISSING VALUES 72

4.1. Introduction..... 72
4.2. Data Preprocessing..... 73
4.2.1. Data Cleaning..... 73
4.2.2. Data Integration and Transformation..... 74
4.2.3. Data Reduction..... 75
4.2.3.1. Numerosity Reduction..... 75
4.2.3.2. Dimensionality Reduction..... 77
4.2.3.2.A. Correla on-based Feature Subset Selection..... 77
4.2.3.2.B. Rough Set Theory..... 78
4.2.3.2.B.1. Feature Selection Using Rough Set..... 79
4.2.3.2.B.2. Reduct..... 79
4.2.3.2.B.3. Core..... 80
4.2.3.2.C. Principal Component Analysis..... 80
4.2.3.2.D. Genetic Algorithm..... 81
4.2.4. Missing Value Handling..... 82
4.2.4.A. Introduction..... 82
4.2.4.B. Missing Data Mechanisms..... 83
4.2.4.C. Techniques to Handle Missing Values..... 83
4.2.4.C.1. List-Wise Deletion..... 83
4.2.4.C.2. Pair-Wise Deletion..... 84
4.2.4.C.3. Single Imputation..... 84
4.2.4.C.4. Multiple Imputations..... 85
4.2.4.D. Principle of Analysis..... 86
4.2.4.E. Results..... 88
4.3. Discussion..... 90
4.4 Conclusion..... 91
Reference..... 92

Chapter 5

PERFORMANCE PREDICTION PARAMETERS..... 95

5.1. Introduction..... 95

5.2. Confusion Matrix..... 95
5.3. Area Under ROC Curve..... 98
5.3.1. Limita ons of ROC Curve..... 100
5.4. Kappa Sta s c..... 101
5.4.1. Limitations of Kappa Statistic..... 103
5.5. Akaike Information Criteria and Bayesian Information Criterion..... 104
5.5.1. Limita ons of AIC and BIC..... 105
5.6. Lift..... 106
5.6.1. Limitations of Lift Chart..... 107
5.7. Cumulative Gain..... 107
5.7.1. Limita ons of Cumulative Gain..... 109
5.8. Probability Threshold..... 109
5.8.1. Limitations of Probability Threshold..... 109
5.9. Conclusion..... 111
Reference..... 112

Chapter 6

BREAST CANCER DISEASE DIAGNOSIS..... 114

6.1. Introduction..... 114
6.2. Wisconsin Breast Cancer Database..... 117
6.3. Application..... 118
6.4. Data Preprocessing..... 119
6.4.1. Feature Extraction or Reduction..... 119
6.4.1.A. Correla on Based Feature Subset Selection..... 120
6.4.1.B. Feature Selection using Rough Set..... 121
6.4.1.C. Feature Selection using PCA..... 122
6.5. Artificial Neural Network Architecture..... 124
6.5.1. Incremental Backpropaga on Learning Network..... 126
6.5.2. Levenberg-Marquardt Algorithm..... 127
6.6. Modeling Results..... 127
6.7. Performance Measure..... 129
6.8. Experimental Results..... 129
6.9. Conclusion..... 131
Reference..... 132

Chapter 7

HEPATITIS DISEASE DIAGNOSIS..... 136

7.1. Introduction..... 136
7.2. UCI Arrhythmia Data Set..... 137
7.3. Application..... 138
7.4. Data Preprocessing..... 139
7.4.1. Feature Extraction or Reduction..... 139
7.4.2. Genetic Algorithm..... 140
7.4.3. Correla on-based Feature Subset Selection..... 141
7.5. Artificial Neural Network Architecture..... 143
7.5.1. Incremental Backpropagation Learning Network..... 147
7.5.2. Levenberg-Marquardt Algorithm..... 148
7.6. Modeling Results..... 148
7.7. Performance Evaluation Method..... 149
7.8. Experimental Results..... 150
7.9. Conclusion..... 152
Reference..... 153

Chapter 8

HEPATITIS DISEASE DIAGNOSIS..... 157

8.1. Introduction..... 157
8.2. UCI Hepatitis Data Set..... 158
8.3. Multiple Imputations..... 160
8.4. Feature Selection Techniques..... 161
8.4.A. Correla on-based Feature Subset Selection..... 162
8.4.B. Rough Set Theory..... 163
8.5. Applications..... 164
8.6. Data Preprocessing..... 165
8.7. Ar cial Neural Network..... 166
8.8. Artificial Neural Network Architecture..... 167
8.9. Training or Learning the Network..... 169
8.10. Modeling Results..... 169
8.11. Performance Prediction Parameters..... 170
8.12. Experimental Results..... 170
8.13. Conclusion..... 174
Reference..... 176

Chapter 9

ACCOMPLISHMENTS OF THE OBJECTIVES AND FUTURE SCOPE..... 179

9.1. Introduction……179

9.2. Accomplishments of the Objectives……. 179

9.3. Future Scope and Further Works…… 185

BIBLIOGRAPHY…… 187

INDEX ……201