ACKNOWLEDGEMENT

It is my foremost responsibility to express my profound gratitude and sincere respects to my Research Guide Dr. Neelu Khare for her continuous support throughout my work. Her talents, dedication and enthusiasm for research will be a continuing source of inspiration.

I am very grateful to Honourable Chancellor Dr. G. Viswanathan, for his support and encouragement without that this work would certainly not have been possible.

Very special thanks to Vice Presidents Mr. Sankar Viswanathan, Mr. Sekar Viswanathan and Mr. G. V. Selvam, for helping me to keep my sanity over the entire period of research.

I wish to express my heart-felt gratitude to the Doctoral Committee members Dr. N.P Gopalan & Dr. Manikandan.

I gratefully acknowledge the Vice Chancellor Dr. Anand Samuel, for providing me unflinching encouragement and support in various ways.

Many thanks to the Pro-Vice Chancellor Dr. S. Narayanan.

I express my sincere thanks to the Dr. G. Buvaneswari, Dean, Academics Research for facilitating all the process towards completion of my Ph. D program.

I express my sincere thanks to Dr. Aswani Kumar Cherukuri, Dean, School of Information Technology and Engineering, for his invaluable suggestions to enhance my research work.

I wish to express my heartfelt thanks to Dr. Sree Hari Reddy, General Physician, Sree Diabetes Center, Kurnool, Andhra Pradesh for providing me the data and helping me out in understanding the attributes of diabetes disease.

I am deeply indebted to my family for having given me the much required motivation and inspiration for hard-work, determination and self-confidence. I wish to thank my friends for their timely help.

Finally, and most importantly, I would like to thank the Almighty God, for it is under his grace that I live, learn and flourish.

Place : Vellore

Date : 21.12.2017

G Thippa Reddy
# CONTENTS

Chapter 1 Introduction

1.1 DATA MINING ................................. 1

1.2 DATA MINING ON MEDICAL DATA .......... 2

1.3 ACQUISITION OF MEDICAL DATA .......... 3

1.4 NEED OF MEDICAL DATA MINING .......... 3

1.5 CATEGORIZATION OF MEDICAL DATA MINING ... 5

   1.5.1 PREDICTIVE ANALYSIS .................. 5

   1.5.2 DESCRIPTIVE ANALYSIS .................. 7

1.6 OPTIMIZATION AND METAHEURISTIC ALGORITHMS IN DATA MINING 8

1.7 MOTIVATION ................................ 9

Chapter 2 Literature Review

2.1 GENERAL REVIEW OF DATA MINING ON MEDICAL DATA 12

   2.1.1 INFORMATION MINING FROM MEDICAL DATA 12

   2.1.2 CLASSIFICATION OF MEDICAL DATA .......... 14

   2.1.3 DIAGNOSIS OF HEART DISEASE USING PREDICTIVE MINING METHODS 17

   2.1.4 DIAGNOSIS OF DIABETES USING PREDICTIVE MINING METHODS 20

   2.1.5 HEART DISEASE DIAGNOSIS BY DATA MINING .......... 21

   2.1.6 DIABETES DIAGNOSIS BY DATA MINING ......... 21

2.2 MEDICAL DATA MINING BASED ON INTELLIGENT METHODS 23

2.3 DATA MINING APPROACHES FOR HEART DISEASE ....... 25

2.4 DATA MINING APPROACHES FOR DIABETES .......... 28

2.5 DIABETES DIAGNOSIS USING DATA MINING ......... 31

2.6 TRENDS AND FUTURE DEVELOPMENTS .......... 33

2.7 CONCLUSIONS ................................ 35

Chapter 3 Conclusion

3.1 SUMMARY ................................... 37

3.2 FUTURE RESEARCH DIRECTIONS ............... 39

3.3 ACKNOWLEDGEMENTS ......................... 41

3.4 REFERENCES ................................ 42

Appendix A Sample Data Set

Appendix B Sample Analysis

Appendix C Sample Graphs

Appendix D Sample Algorithms
Rule Based Algorithm

4.1 INTRODUCTION

4.2 HEART DISEASE PREDICTION USING RULE BASED FFBAT FUZZY CLASSIFIER

4.3 RESULTS AND DISCUSSION

4.3.1 DATASET DISCUSSION

4.3.2 PERFORMANCE EVALUATION

4.3.3 COMPARISON OF ACCURACY WITH OTHER REPORTED RESULTS

4.5 SUMMARY

Chapter 5 An Efficient System for Heart Disease Prediction using Hybrid OFBAT with Rule-Based Fuzzy Logic Model

5.1 INTRODUCTION

5.2 OFBAT-FUZZY CLASSIFIER BASED HEART DISEASE PREDICTION

5.2.1 FEATURE REDUCTION USING LPP

5.2.2 HEART DISEASE PREDICTION USING RULE BASED FUZZY CLASSIFIER

5.3 RESULTS EVALUATION AND DISCUSSION

5.3.1 PERFORMANCE EVALUATION

5.3.2 PERFORMANCE ANALYSIS

5.3.3 COMPARISON OF THE ACCURACY OF OTHER REPORTED RESULTS

5.4 SUMMARY
Chapter 6 Cuckoo Search Optimized Reduction and Fuzzy Logic Classifier for Heart Disease and Diabetes Prediction

6.1 INTRODUCTION

6.2 ROUGH SET THEORY

6.3 AN APPROACH FOR PROPOSED HEART AND DIABETES PREDICTION

6.4 PERFORMANCE EVALUATION AND DISCUSSION

6.4.1 EVALUATION METRICS

6.4.2 PERFORMANCE EVALUATION OF PROPOSED APPROACH

6.4.3 COMPARISON OF THE OTHER APPROACHES

6.5 SUMMARY

Chapter 7 Conclusion

References

List of Publications
LIST OF FIGURES

1.1 Acquisition Model of Medical Data 4
1.2 Organization of Data Mining Methods 6
1.3 Graphical Representation of Chronological Statistics of Heart Disease in India 10
1.4 Graphical Representation of Chronological Statistics of Diabetes in India 11
2.1 Pie Chart Representation of the Overall Contribution of Data Mining and Classification on Medical Data 22
2.2 Taxonomy of Medical Data Mining Based on Intelligent Methods 24
2.3 Chronological Review on Medical Data Mining Based on Intelligent Methods 25
2.4 Taxonomy of Heart Disease Prediction Methods 26
2.5 Chronological Review on Heart Disease Prediction 27
2.6 Taxonomy of Diabetes Prediction Methods 28
2.7 Chronological Review of Diabetes Prediction 29
2.8 Contribution of Meta-Heuristic Algorithms on Medical Data in Chronological Order 30
2.9 Categorization of Meta-Heuristic Algorithm on Medical Data Based on Inspiration 34
3.1 Process Flow Diagram of FFBAT-RBFL 50
3.2 Schematic Representation of RBFL Algorithm 52
3.3 Solution Encoding 56
3.4 Accuracy Plot for Population Size 10 58
3.5 Accuracy Plot for Population Size 15 58
5.2 Performance Evaluation of Accuracy Plot for Population Size 10 76
5.3 Performance Evaluation of Accuracy Plot for Population Size 15 77
5.4 Performance Evaluation of Accuracy Plot for Population Size 20 77
5.5 Comparative Analysis of Proposed and Other Approaches for Population Size 10 78
5.6 Comparative Analysis of Proposed and Other Approaches for Population Size 15 79
5.7 Comparative Analysis of Proposed and Other Approaches for Population Size 20 79
5.8 Comparative Analysis of Proposed and Other Approaches 80
5.9 Accuracy Plot between Proposed and Existing Approach 80
6.1 The Overall Process of Proposed Disease Prediction System 91
6.2 Comparative Analysis of Different Dimension Reduction Approach 91
6.3 Fitness Plot of Proposed CS+RS 92
6.4 Fitness Plot of LPP + RBFL 92
6.5 Fitness Plot of RS + FL 92
6.6 Performance of Proposed Approach based on Accuracy 93
6.7 Performance of Proposed Approach based on Sensitivity 94
6.8 Performance of Proposed Approach Based on Specificity 93
6.9 Time Taken for Training Against the Number of Data
for (a) Cleaveland Dataset (b) Hungarian Dataset
(c) Switzerland Dataset (d) Real-Time Data Set

LIST OF TABLES

2.1 Descriptive Mining for Clinical Diagnosis 32
2.2 Predictive Mining for Clinical Diagnosis 35
2.3 Diagnosis of Heart Disease Using Predictive Mining Methods 37
2.4 Diagnosis of Diabetes Using Predictive Mining Methods 40
2.5 Heart Disease Diagnosis by Descriptive Mining Methods 42
2.6 Diabetes Diagnosis by Descriptive Mining Methods 40
2.7 Optimization Problems 44
3.1 Dataset Description 62
3.2 Confusion matrix 63
6.1 Comparative Analysis of Proposed Approach with other Approaches based on Accuracy, Sensitivity and Specificity 90
<table>
<thead>
<tr>
<th>ACRONYMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS   : Cuckoo Search</td>
</tr>
<tr>
<td>PSO  : Particle Swarm Optimization</td>
</tr>
<tr>
<td>FF   : Firefly</td>
</tr>
<tr>
<td>FFBAT : Firefly-BAT</td>
</tr>
<tr>
<td>FN   : False Negative</td>
</tr>
<tr>
<td>FP   : False Positive</td>
</tr>
<tr>
<td>LPP  : Locality Preserving Projection</td>
</tr>
<tr>
<td>OFBAT : Oppositional Based Learning FFBAT</td>
</tr>
<tr>
<td>PSO  : Particle Swarm Optimization</td>
</tr>
<tr>
<td>RBFL : Rule Based Fuzzy Logic Classifier</td>
</tr>
<tr>
<td>RS   : Rough Sets</td>
</tr>
<tr>
<td>TN   : True Negative</td>
</tr>
<tr>
<td>TP   : True Positive</td>
</tr>
<tr>
<td>UCI  : University of California</td>
</tr>
</tbody>
</table>