

# Contents

<i>Preface</i>	v
1. Biomedical Databases and Data Mining	1
1.1 Databases and Knowledge Discovery in Biomedicine . . .	2
1.2 Outline of this Book . . . . .	5
2. DYNASTAT: A Methodology for Modeling of Multi-agent Systems	11
2.1 Introduction . . . . .	11
2.2 Literature Review . . . . .	13
2.3 DYNASTAT Methodology . . . . .	16
2.4 Use of UML 2.2 in the Framework of DYNASTAT Methodology . . . . .	18
2.5 UML to Model Medical Multi-agent Systems . . . . .	23
2.6 Possible Applications . . . . .	30
2.7 Conclusion . . . . .	33
3. SciPort: An Extensible Data Management Platform for Biomedical Research	35
3.1 Introduction . . . . .	35
3.2 Related Work . . . . .	38
3.3 Unified Scientific Data Modeling . . . . .	39
3.4 Document Authoring and Searching . . . . .	47
3.5 Sharing Distributed Biomedical Data . . . . .	56
3.6 The Architecture of SciPort . . . . .	59
3.7 Conclusion . . . . .	62

4.	An Integrative Framework for Anonymizing Clinical and Genomic Data	65
4.1	Introduction . . . . .	65
4.2	Related Work . . . . .	70
4.3	The DIANOVA Framework . . . . .	77
4.4	Algorithms for Realizing DIANOVA . . . . .	83
4.5	Extensions of DIANOVA . . . . .	88
4.6	Conclusion . . . . .	88
5.	Data Integration Challenges: A Systems Biology Perspective	91
5.1	Introduction . . . . .	91
5.2	Modeling Biological Systems . . . . .	92
5.3	Biological and Mathematical Data Sources . . . . .	94
5.4	Various Data Exchange Formats in Systems Biology . . . . .	96
5.5	Building an Integrative Framework to Combine Modeling and Biological Data Sources . . . . .	98
5.6	Analysis of the Developed Integrative Environment . . . . .	111
5.7	Conclusion . . . . .	113
6.	Ontology-based Data Integration: A Case Study in Clinical Trials	115
6.1	Introduction . . . . .	115
6.2	System Architecture and Overview . . . . .	117
6.3	The CTDM Ontology . . . . .	118
6.4	Ontology-based Data Integration of Study Relevant Information . . . . .	121
6.5	Assembly of ETL Processes Based on Ontology Mappings	130
6.6	Case Study and Evaluation . . . . .	132
6.7	Related Work . . . . .	135
6.8	Conclusion . . . . .	136
7.	A Data Warehouse for <i>Ca. Glomeribacter Gigasporarum</i> Bacterium	139
7.1	Introduction . . . . .	139
7.2	State-of-the-art of Metagenomics for Genomic Comparison	141
7.3	BIOBITS System Architecture . . . . .	143
7.4	Software Modules to Support Researchers' Activities . . . . .	149
7.5	Conclusion . . . . .	154

8. Quality of Medical Data: A Case Study	157
8.1 Introduction . . . . .	158
8.2 Case Study . . . . .	165
8.3 Generation of a Summary Table . . . . .	172
8.4 Conclusion . . . . .	173
9. Efficient EMD-based Similarity Search in Medical Image Databases	175
9.1 Introduction . . . . .	175
9.2 Dimensionality Reduction for the EMD . . . . .	182
9.3 Query Processing Algorithm . . . . .	192
9.4 Evaluation on Medical Data Sets . . . . .	193
9.5 Conclusion . . . . .	200
10. Fast Multimedia Querying for Medical Applications	203
10.1 Introduction . . . . .	203
10.2 Related Work . . . . .	206
10.3 Subspace Tree . . . . .	208
10.4 Experiments . . . . .	215
10.5 Conclusion . . . . .	218
11. Ensemble Feature Selection in Biomedical Applications	221
11.1 Introduction . . . . .	222
11.2 Evaluation of Feature Selection Approaches . . . . .	225
11.3 Ensemble Feature Selection . . . . .	228
11.4 Biomedical Example . . . . .	230
11.5 Computational Approach . . . . .	232
11.6 Results . . . . .	233
11.7 Discussion . . . . .	235
11.8 Conclusion . . . . .	238
12. Analysis of Breast Cancer Genomic Data by Fuzzy Association Rule Mining	241
12.1 Introduction . . . . .	242
12.2 Microarrays . . . . .	247
12.3 Association Rule Mining . . . . .	247
12.4 Fuzzy Association Rules . . . . .	251
12.5 Dataset . . . . .	254

12.6	Extracting the Fuzzy Association Rules . . . . .	257
12.7	Results . . . . .	268
12.8	Conclusion . . . . .	277
13.	Graph Mining on Brain Co-activation Networks	279
13.1	Introduction . . . . .	281
13.2	Related Work . . . . .	282
13.3	Method . . . . .	284
13.4	Experiments . . . . .	288
13.5	Conclusion . . . . .	293
14.	Automatic Identification of Surgery Indicators	295
14.1	Introduction . . . . .	295
14.2	Background . . . . .	297
14.3	Approach . . . . .	302
14.4	Experiment . . . . .	308
14.5	Conclusion . . . . .	318
15.	Incremental Learning of Medical Data for Multi-step Patient Health Classification	321
15.1	Introduction . . . . .	321
15.2	The Bayes Tree . . . . .	325
15.3	Experimental Evaluation . . . . .	330
15.4	Incorporating Medical Knowledge Data Bases . . . . .	337
15.5	Application of Anytime Classification . . . . .	340
15.6	Conclusion . . . . .	343
	<i>Bibliography</i>	345