

CONTENTS

Preface.....	v
Chapter 1 Survey of Early Warning Systems for Environmental and Public Health Applications	1
1. Introduction.....	1
2. Disease Surveillance.....	3
3. Reference Architecture for Model Extraction.....	5
4. Problem Domain	9
5. Data Sources	10
6. Detection Methods.....	12
7. Summary and Conclusion.....	13
References	14
Chapter 2 Time-Lapse Cell Cycle Quantitative Data Analysis Using Gaussian Mixture Models	17
1. Introduction	18
2. Material and Feature Extraction	20
2.1. Material and cell feature extraction	20
2.2. Model the time-lapse data using AR model	23
3. Problem Statement and Formulation	24
4. Classification Methods	26
4.1. Gaussian mixture models and the EM algorithm	26
4.2. K-Nearest Neighbor (KNN) classifier.....	28
4.3. Neural networks.....	28
4.4. Decision tree.....	29
4.5. Fisher clustering	30
5. Experimental Results	30
5.1. Trace identification.....	31
5.2. Cell morphologic similarity analysis	33
5.3. Phase identification	35
5.4. Cluster analysis of time-lapse data	37

6. Conclusion	40
Appendix A	41
Appendix B.....	42
References	43
Chapter 3 Diversity and Accuracy of Data Mining Ensemble.....	47
1. Introduction.....	47
2. Ensemble and Diversity	49
2.1. Why needs diversity?	49
2.2. Diversity measures	51
3. Probability Analysis.....	52
4. Coincident Failure Diversity.....	52
5. Ensemble Accuracy	55
5.1. Relationship between random guess and accuracy of lower bound single models.....	55
5.2. Relationship between accuracy A and the number of models N	56
5.3. When model's accuracy $< 50\%$	57
6. Construction of Effective Ensembles	58
6.1. Strategies for increasing diversity	59
6.2. Ensembles of neural networks	60
6.3. Ensembles of decision trees	61
6.4. Hybrid ensembles	62
7. An Application: Osteoporosis Classification Problem	62
7.1. Osteoporosis problem.....	63
7.2. Results from the ensembles of neural nets	63
7.3. Results from ensembles of the decision trees.....	66
7.4. Results of hybrid ensembles.....	67
8. Discussion and Conclusions	68
References	70
Chapter 4 Integrated Clustering for Microarray Data	73
1. Introduction.....	73
2. Related Work	77
3. Data Preprocessing	81

4. Integrated Clustering.....	83
4.1. Clustering algorithms	83
4.2. Integration methodology	88
5. Experimental Evaluation.....	89
5.1. Evaluation methodology.....	89
5.2. Results	91
5.3. Discussion	93
6. Conclusions.....	94
References	94
Chapter 5 Complexity and Synchronization of EEG with Parametric Modeling	99
1. Introduction.....	100
1.1. Brief review of EEG recording analysis.....	100
1.2. AR modeling based EEG analysis.....	101
2. TVAR Modeling	104
3. Complexity Measure.....	105
4. Synchronization Measure	109
5. Conclusions.....	113
References	114
Chapter 6 Bayesian Fusion of Syndromic Surveillance with Sensor Data for Disease Outbreak Classification.....	119
1. Introduction.....	120
2. Approach.....	122
2.1. Bayesian belief networks.....	122
2.2. Syndromic data.....	126
2.3. Environmental data.....	128
2.4. Test scenarios	130
2.5. Evaluation metrics	130
3. Results.....	131
3.1. Scenario 1	131
3.2. Scenario 2	134
3.3. Promptness	135
4. Summary and Conclusions	136
References	137

Chapter 7	An Evaluation of Over-the-Counter Medication Sales for Syndromic Surveillance.....	143
1.	Introduction.....	143
2.	Background and Related Work.....	144
3.	Data.....	144
4.	Approaches	145
4.1.	Lead-lag correlation analysis	145
4.2.	Regression test of predictive ability.....	146
4.3.	Detection-based approaches	148
4.4.	Supervised algorithm for outbreak detection in OTC data	148
4.5.	Modified Holt-Winters forecaster	150
4.6.	Forecasting based on multi-channel regression.....	151
5.	Experiments	153
5.1.	Lead-lag correlation analysis of OTC data.....	153
5.2.	Regression test of the predicative value of OTC	154
5.3.	Results from detection-based approaches.....	156
6.	Conclusions and Future Work	158
	References	159
Chapter 8	Collaborative Health Sentinel.....	163
1.	Introduction.....	163
2.	Infectious Disease and Existing Health Surveillance Programs	166
3.	Elements of the Collaborative Health Sentinel (CHS) System..	170
3.1.	Sampling.....	170
3.2.	Creating a national health map	177
3.3.	Detection	177
3.4.	Reaction.....	183
3.5.	Cost considerations.....	184
4.	Interaction with the Health Information Technology (HCIT) World	185
5.	Conclusion	188
	References	189
	Appendix A - HL7.....	192

Chapter 9	A Multi-Modal System Approach for Drug Abuse Research and Treatment Evaluation: Information Systems Needs and Challenges	195
1.	Introduction.....	195
2.	Context.....	198
2.1.	Data sources	198
2.2.	Examples of relevant questions	199
3.	Possible System Structure.....	201
4.	Challenges in System Development and Implementation	204
4.1.	Ontology development	204
4.2.	Data source control, proprietary issues.....	205
4.3.	Privacy, security issues.....	205
4.4.	Costs to implement/maintain system.....	206
4.5.	Historical hypothesis-testing paradigm	206
4.6.	Utility, usability, credibility of such a system.....	206
4.7.	Funding of system development.....	207
5.	Summary	207
	References	208
Chapter 10	Knowledge Representation for Versatile Hybrid Intelligent Processing Applied in Predictive Toxicology ..	213
1.	Introduction.....	214
2.	Hybrid Intelligent Techniques for Predictive Toxicology Knowledge Representation	217
3.	XML Schemas for Knowledge Representation and Processing in AI and Predictive Toxicology	218
4.	Towards a Standard for Chemical Data Representation in Predictive Toxicology.....	220
5.	Hybrid Intelligent Systems for Knowledge Representation in Predictive Toxicology.....	225
5.1.	A formal description of implicit and explicit knowledge-based intelligent systems	226
5.2.	An XML schema for hybrid intelligent systems	228
6.	A Case Study	231
6.1.	Materials and methods.....	232
6.2.	Results	233

7. Conclusions.....	235
References	236
Chapter 11 Ensemble Classification System Implementation for Biomedical Microarray Data	239
1. Introduction.....	240
2. Background.....	241
2.1. Reasons for ensemble	241
2.2. Diversity and ensemble	241
2.3. Relationship between measures of diversity and combination method	243
2.4. Measures of diversity	243
2.5. Microarray data	244
3. Ensemble Classification System (ECS) Design.....	245
3.1. ECS overview	245
3.2. Feature subset selection.....	247
3.3. Base classifiers	248
3.4. Combination strategy.....	249
4. Experiments	250
4.1. Experimental datasets.....	250
4.2. Experimental results	252
5. Conclusion and Further Work.....	254
References	255
Chapter 12 An Automated Method for Cell Phase Identification in High Throughput Time-Lapse Screens	257
1. Introduction.....	258
2. Nuclei Segmentation and Tracking.....	259
3. Cell Phase Identification.....	260
3.1. Feature calculation.....	260
3.2. Identifying cell phase	262
3.3. Correcting cell phase identification errors.....	265
4. Experimental Results	266
5. Conclusion	272
References	272

Chapter 13 Inference of Transcriptional Regulatory Networks Based on Cancer Microarray Data	275
1. Introduction.....	275
2. Subnetworks and Transcriptional Regulatory Networks Inference	277
2.1. Inferring subnetworks using z-score.....	277
2.2. Inferring subnetworks based on graph theory	278
2.3. Inferring subnetworks based on Bayesian networks	279
2.4. Inferring transcriptional regulatory networks based on integrated expression and sequence data.....	283
3. Multinomial Probit Regression with Bayesian Gene Selection...	284
3.1. Problem formulation.....	284
3.2. Bayesian variable selection	286
3.3. Bayesian estimation using the strongest genes.....	288
3.4. Experimental results	289
4. Network Construction Based on Clustering and Predictor Design	293
4.1. Predictor construction using reversible jump MCMC annealing	293
4.2. CoD for predictors.....	295
4.3. Experimental results on a Myeloid line.....	296
5. Concluding Remarks	298
References	299
 Chapter 14 Data Mining in Biomedicine	 305
1. Introduction.....	305
2. Predictive Model Construction	306
2.1. Derivation of unsupervised models	307
2.2. Derivation of supervised models.....	311
3. Validation	316
4. Impact Analysis	318
5. Summary	319
References	319

Chapter 15 Mining Multilevel Association Rules from Gene Ontology and Microarray Data.....	321
1. Introduction.....	321
2. Proposed Methods	323
2.1. Preprocessing.....	323
2.2. Hierarchy-information encoding	324
3. The MAGO Algorithm	326
3.1. MAGO algorithm	327
3.2. CMAGO (Constrained Multilevel Association rules with Gene Ontology)	329
4. Experimental Results	330
4.1. The characteristic of the dataset	331
4.2. Experimental results	331
4.3. Interpretation	334
5. Concluding Remarks	335
References	336
 Chapter 16 A Proposed Sensor-Configuration and Sensitivity Analysis of Parameters with Applications to Biosensors..	339
1. Introduction.....	340
2. Sensor-System Configuration	342
3. Optical Biosensors	346
3.1. Relationship between parameters	347
3.2. Modelling of parameters.....	351
4. Discussion.....	356
5. Conclusion	358
References	359
 Epilogue	361
References	364
 Index	365