ARTIFICIAL INTELLIGENCE TECHNIQUES IN BREAST CANCER DIAGNOSIS AND PROGNOSIS

SERIES IN MACHINE PERCEPTION AND ARTIFICIAL INTELLIGENCE*

Editors: **H. Bunke** (Univ. Bern, Switzerland)
P. S. P. Wang (Northeastern Univ., USA)

- Vol. 23: Machine Learning and Perception
(Eds. *G. Tasini, F. Esposito, V. Roberto and P. Zingaretti*)
- Vol. 24: Spatial Computing: Issues in Vision, Multimedia and Visualization Technologies
(Eds. *T. Caelli, Peng Lam and H. Bunke*)
- Vol. 25: Studies in Pattern Recognition Memorial Volume in Honor of K S Fu
(Eds. *H. Freeman*)
- Vol. 26: Neural Network Training Using Genetic Algorithms
(Eds. *L. C. Jain, R. P. Johnson and A. F. J. van Rooij*)
- Vol. 27: Intelligent Robots — Sensing, Modeling & Planning
(Eds. *B. Bolles, H. Bunke and H. Noltemeier*)
- Vol. 28: Automatic Bankcheck Processing
(Eds. *S. Impedovo, P. S. P. Wang and H. Bunke*)
- Vol. 29: Document Analysis II
(Eds. *J. J. Hull and S. Taylor*)
- Vol. 30: Compensatory Genetic Fuzzy Neural Networks and Their Applications
(*Y.-Q. Zhang and A. Kandel*)
- Vol. 31: Parallel Image Analysis: Tools and Models
(Eds. *S. Miquet, A. Montanvert and P. S. P. Wang*)
- Vol. 33: Advances in Oriental Document Analysis and Recognition Techniques
(Eds. *S.-W. Lee, Y. Y. Tang and P. S. P. Wang*)
- Vol. 34: Advances in Handwriting Recognition
(Ed. *S.-W. Lee*)
- Vol. 35: Vision Interface — Real World Applications of Computer Vision
(Eds. *M. Cheriet and Y.-H. Yang*)
- Vol. 36: Wavelet Theory and Its Application to Pattern Recognition
(*Y. Y. Tang, L. H. Yang, J. Liu and H. Ma*)
- Vol. 37: Image Processing for the Food Industry
(*E. R. Davies*)
- Vol. 38: New Approaches to Fuzzy Modeling and Control — Design and Analysis
(*M. Margaliot and G. Langholz*)

*For the complete list of titles in this series, please write to the Publisher.

ARTIFICIAL INTELLIGENCE TECHNIQUES IN BREAST CANCER DIAGNOSIS AND PROGNOSIS

Editors

Ashlesha Jain

Ajita Jain

Sandhya Jain

Lakhmi Jain

World Scientific

Singapore • New Jersey • London • Hong Kong

Published by

World Scientific Publishing Co. Pte. Ltd.

P O Box 128, Farrer Road, Singapore 912805

USA office: Suite 1B, 1060 Main Street, River Edge, NJ 07661

UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

**ARTIFICIAL INTELLIGENCE TECHNIQUES IN BREAST CANCER
DIAGNOSIS AND PROGNOSIS**

Series in Machine Perception and Artificial Intelligence — Vol. 39

Copyright © 2000 by World Scientific Publishing Co. Pte. Ltd.

All rights reserved. This book, or parts thereof, may not be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system now known or to be invented, without written permission from the Publisher.

For photocopying of material in this volume, please pay a copying fee through the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA. In this case permission to photocopy is not required from the publisher.

ISBN 981-02-4374-X

This book is printed on acid-free paper.

Printed in Singapore by World Scientific Printers

**This book is dedicated to
the victims of cancer**

The proceeds of royalty will be donated to the Anti-Cancer Foundation

This page is intentionally left blank

Contributors

Chapter 1

R. Kates, N. Harbeck, and M. Schmitt

Technische Universität München

Munich

Germany

A. Jain

Private Practitioner

Adelaide

Australia

A. Jain

University of Adelaide

Adelaide

Australia

Chapter 2

M. Bottema, G.N. Lee, and S. Lu

Flinders University of South Australia

Adelaide

Australia

Chapter 3

R. Kates, N. Harbeck, K. Ulm, H. Graeff, and M. Schmitt

Technische Universität München

Munich

Germany

F. Jaenicke

Universitätsfrauenklinik Eppendorf

Hamburg

Germany

Chapter 4

L.M. Roberts

University of Wisconsin-Milwaukee

Milwaukee, WI

U.S.A.

Chapter 5

M.G. Daidone and D. Coradini

Istituto Nazionale per lo Studio e la Cura dei Tumori

Milan

Italy

Chapter 6

H.-P. Chan, N. Petrick, and B. Sahiner

University of Michigan

Ann Arbor, MI

U.S.A.

Chapter 7

S.S. Cross, J. Downs, P. Drezet, Z. Ma, and R.F. Harrison

University of Sheffield

Sheffield

U.K.

Chapter 8

W.N. Street

University of Iowa

Iowa City, IA

U.S.A.

PREFACE

The main aim of this book is to present a sample of recent research in the application of novel artificial intelligence paradigms to the diagnosis and prognosis of breast cancer. These paradigms include neural networks, fuzzy logic and evolutionary computing. Artificial intelligence techniques offer advantages, such as adaptation, fault tolerance, learning and human like behavior, over conventional computing techniques.

The idea is to combine the pathological, intelligent and statistical approaches to enable simple and accurate diagnosis and prediction of prognosis.

This book contains 8 chapters. Chapter 1 provides a brief background on breast cancer including epidemiology, diagnosis, prognosis and management of the disease. It also highlights the benefits of artificial intelligence paradigms.

Bottema, Lee, and Lu present the techniques for automatic image feature extraction for diagnosis and prognosis of breast cancer in Chapter 2. Artificial intelligence methods process information in order to arrive at decisions. Image analysis methods are described in the context of extracting information from images relevant to the detection, classification, and prognosis of breast cancer.

Chapter 3, by Kates, Harbeck, Ulm, Jaenicke, Graeff, and Schmitt, discusses the clinical requirements for decision support strategies in breast cancer. It is demonstrated that a statistical mode of neural network operation emphasizing the avoidance of unnecessary complexity can recognize and represent the complicated dependence of the disease on factors, and distinguish this dependence from the noise.

Roberts presents MammoNet, a Bayesian network aiming to perform accurate, reliable and consistent diagnosis of breast cancer in Chapter 4. The network creates a differential diagnosis by specifying the observed symptoms and computing the posterior probability of the various diagnoses using standard probability formulas.

Daidone and Coradini focus on predicting prognosis and treatment response in breast cancer patients in Chapter 5. Integration of pathological, clinical and statistical approaches is shown to contribute significantly to the prediction of prognosis.

Chapter 6 by Chan, Petrick, and Sahiner is on computer-aided breast cancer diagnosis. It discusses computer vision techniques for detection and characterization of masses and microcalcifications on mammograms as well as the effects of computer-aided diagnosis algorithms on radiologists' diagnostic accuracy.

Cross, Downs, Drezet, Ma, and Harrison examine in Chapter 7 four different decision support technologies: logistic regression, data-derived decision trees, multilayer perceptrons, and adaptive resonance theory mapping neural networks.

The final chapter, by Street, discusses Xcyt, a system for remote cytological diagnosis and prognosis of breast cancer. Xcyt is a software system for providing expert diagnosis and prognosis of breast cancer based on fine needle aspirates. This chapter addresses important topics in imaging, diagnosis and prognosis. The Xcyt system is being made available via the World Wide Web.

It is our objective to develop an intelligent, cost-effective, reliable and easily accessible system for the diagnosis and prognosis of breast cancer. It is our belief that this book will stimulate the imagination of researchers in the application of artificial intelligence paradigms in this important area of medicine.

This book will be useful to researchers, physicians, and students who are interested in the application of artificial intelligence paradigms in breast cancer diagnosis and prognosis.

We are grateful to the authors for their contributions and to all the reviewers for their expertise in the various fields. We acknowledge the assistance of Berend Jan and Irene van der Zwaag and Shaheed Mehta in the preparation of the manuscript. We thank Professor H. Bunke for the opportunity to publish this book, and the World Scientific Publishing Company for their excellent editorial assistance.

CONTENTS

Chapter 1.

	An Introduction to Breast Cancer Diagnosis, Prognosis, and Artificial Intelligence	1
	<i>N. Harbeck, R.E. Kates, M. Schmitt, A. Jain, and A. Jain</i>	
1	Breast Cancer Characteristics	2
1.1	Epidemiology	2
1.2	Biology	3
2	Breast Cancer Diagnosis and Management	4
2.1	Diagnosis	4
2.2	Disease Management	5
3	Breast Cancer Prognosis and Risk Assessment	7
4	Artificial Intelligence	9
5	Summary	12
	References	13

Chapter 2.

	Automatic Image Feature Extraction for Diagnosis and Prognosis of Breast Cancer	17
	<i>M.J. Bottema, G.N. Lee, and S. Lu</i>	
1	Introduction	17
2	Image Modalities	18
2.1	Mammography	18
2.1.1	Screening Mammography	18
2.1.2	Clinical Mammography	19
2.1.3	Detection and Classification of Masses	20
2.1.4	Detection of Stellate Lesions	21
2.1.5	Detection and Classification of Microcalcifications	22
2.2	Ultrasound	25
2.3	Fine Needle Aspiration Slides	26
3	Image Features and Analysis	26
3.1	Geometric Features	27
3.1.1	The Boundary	28
3.1.2	Area, Volume, Radius, Height, Contrast	30
3.1.3	The Curse of $\text{Perimeter}^2/4\pi$ Area	31
3.1.4	Shape	33

3.2 Texture Features	38
3.2.1 Run-Length Matrices	38
3.2.2 Co-occurrence Matrices	38
3.2.3 Markovian Statistics	39
3.2.4 Example of Texture Arrays	39
3.2.5 Fractal Dimension	40
3.3 Morphology	40
3.4 Snakes	42
3.5 Wavelets	44
4 Summary	46
Acknowledgments	47
References	47

Chapter 3.

Decision Support in Breast Cancer: Recent Advances in Prognostic and Predictive Techniques

R. Kates, N. Harbeck, K. Ulm, F. Jaenicke, H. Graeff, and M. Schmitt

1 Clinical Requirements for Decision Support Strategies in Breast Cancer	55
1.1 Risk-Group Assessment	55
1.2 Tumor-Biological Factors: The Plasminogen Activation System	56
1.3 The Role of Statistical Models in Clinical Decision Support for Breast Cancer	58
2 Intelligent Systems and Statistical Techniques for Prognosis and Prediction	58
2.1 Advanced Statistical Models as Intelligent Systems	58
2.2 The Cox Proportional Hazards Model and Generalizations	59
2.3 Classification and Regression Trees	62
2.4 Intelligent Systems as Statistical Models	64
3 Breast Cancer as a Complex Disease	66
3.1 A Representative Study	66
3.2 Node-Negative Breast Cancer: A Multivariate Proportional Hazards (Cox) Prognostic Model	67
3.3 Univariate Risk Assessment in Subgroups	68
3.4 Analysis of Time-Varying Effects	71
3.5 Classification and Regression Trees (CART)	73
4 Neural Survival Models	75
4.1 Neural Nets and Prognosis in a Clinical Setting	75

4.2	Previous Approaches to Survival Modeling Using Neural Nets	75
4.3	Neural Topology Used for Survival Modeling	76
4.3.1	Transformations and Input Neurons	78
4.3.2	Hidden Neurons	78
4.3.3	Output Nodes	79
4.4	A Survival Model Using Neural Nets to Define the Time-Dependent Risk Structure	79
4.5	Net Optimization	81
4.5.1	Initialization	81
4.5.2	Training	81
4.5.3	Complexity Reduction	83
5	Simulated Follow-Up Data	84
5.1	Simulation Technique	84
5.2	Simulated "Nonlinear Diseases"	84
5.3	Statistical Tools	85
5.4	Performance	86
6	Conclusions and Outlook	91
	Acknowledgments	94
	References	95

Chapter 4.

	MammoNet: A Bayesian Network Diagnosing Breast Cancer	101
	<i>L.M. Roberts</i>	
1	Introduction	101
2	Bayesian Networks	102
2.1	Probability Notation	104
2.1.1	Conditional Probability	105
2.1.2	Joint Probability	106
2.2	Benefits of Bayesian Networks	106
3	Breast Cancer Problem Overview	108
3.1	Risk Factors and Physical Symptoms	108
3.2	Mammographic Findings	109
3.2.1	Direct Mammographic Findings: Mass Analysis	109
3.2.2	Direct Mammographic Findings: Calcification Analysis	110
3.2.3	Indirect Mammographic Findings	110
4	<i>MammoNet</i> : Bayesian Network Diagnostic Tool for Breast Cancer	111
4.1	<i>MammoNet</i> Network Topology	112

4.1.1 *MammoNet* Network Terminology 114

4.1.2 *MammoNet* Data Acquisition..... 114

4.2 *MammoNet* Demographic Features 115

4.2.1 Demographic Feature: Age 115

4.2.2 Demographic Feature: Age of Menarche 116

4.2.3 Demographic Feature: Child Bearing Age 116

4.2.4 Demographic Feature: Number of First Degree Relatives
with History of Breast Cancer 116

4.2.5 Demographic Feature: Previous Biopsy at Same Site 116

4.3 *MammoNet* Hypothesis Node 117

4.4 *MammoNet* Physical Examination Symptoms 118

4.5 *MammoNet* Mammographic Indirect Findings 119

4.6 *MammoNet* Mammographic Direct Indications 120

4.6.1 Mammographic Direct Indication: Mass 120

4.6.2 Mammographic Direct Indications: Calcifications 122

5 Decision Support System 123

5.1 System Architecture 124

5.1.1 User Interface 125

5.1.2 Knowledge Representation Module 126

5.1.3 Multimodal Discourse Module 128

5.1.4 Bayesian Network Inference Engine 133

6 Results 141

7 Conclusion 144

Acknowledgments 146

References 146

Chapter 5.

**Predicting Prognosis and Treatment Response in
Breast Cancer Patients 149**

M.G. Daidone and D. Coradini

1 Introduction 150

2 Patients and Methods 151

2.1 Case Series 152

2.2 Laboratory Determinations 152

2.3 Clinico-Biological Database 153

2.4 Statistical Approaches 153

3 Continuous versus Categorical Biological Variables 156

4 Biomarkers and Prediction of Prognosis 158

4.1	Proliferation-Related Variables	159
4.2	Hormone Receptors	161
4.3	Integration of Biological and Clinicopathologic Information	164
5	Biomarkers and Prediction of Treatment Response	166
5.1	Chemotherapy	167
5.2	Hormonal Therapy	170
6	Discussion and Conclusions	173
	Acknowledgments	174
	References	174

Chapter 6.

	Computer-Aided Breast Cancer Diagnosis	179
	<i>H.-P. Chan, N. Petrick, and B. Sahiner</i>	
1	Introduction	179
2	Computerized Detection of Microcalcifications	180
2.1	Methods	181
2.1.1	Convolution Neural Network Classifier	183
2.1.2	Convolution Neural Network Training — An Example	185
2.2	Analysis of Detection Accuracy	187
2.2.1	FROC Analysis	187
2.2.2	Performance of Computer Detection Program in Unknown Cases	188
2.3	Effects of Computer-Aided Detection on Radiologists' Performance	190
3	Computerized Classification of Microcalcifications	191
3.1	Methods	193
3.1.1	Feature Extraction	194
3.1.2	Feature Selection	197
3.1.3	Feature Classification	201
3.2	Analysis of Classification Accuracy	202
3.2.1	Comparison of Feature Selection Methods	202
3.2.2	Computer-Aided Classification of Microcalcifications	204
4	Computerized Detection of Masses	207
4.1	Methods	209
4.1.1	Density-Weight Contrast Enhancement Segmentation	210
4.1.2	Local Growing	212
4.1.3	Morphological Feature-Based FP Reduction	214
4.1.4	Texture Feature-Based FP Reduction	214

4.2 Analysis of Detection Accuracy 215

 4.2.1 Database of Digitized Mammograms 216

 4.2.2 Definition of a True Positive 216

 4.2.3 Training and Testing 217

 4.2.4 Performance of Mass Detection 218

 4.2.5 Performance of Computer Detection Program with
 Unknown Cases 221

5 Computerized Characterization of Breast Masses 222

 5.1 Methods 225

 5.1.1 First Stage Segmentation 226

 5.1.2 Second Stage Segmentation 227

 5.1.3 Feature Extraction 235

 5.1.4 Classifier Design 237

 5.2 Analysis of Classification Accuracy 238

 5.3 Radiologists' Performance for Breast Mass Characterization
 Using CAD 240

6 Classifier Design 242

7 Summary 246

Acknowledgments 246

References 247

Chapter 7.

**Which Decision Support Technologies are Appropriate
for the Cytodiagnosis of Breast Cancer? 265**

S.S. Cross, J. Downs, P. Drezet, Z. Ma, and R.F. Harrison

1 Introduction 265

2 The Problem Domain 266

3 Previous Work in This Area 268

4 The Dataset Used in Our Studies 269

 4.1 Study Population 269

 4.2 The Defined Observations that were Used as Input Variables 270

 4.3 Partitioning of the Data 271

5 Human Performance 272

 5.1 Logistic Regression 273

6 Data Derived Decision Tree 276

7 Multilayer Perceptron Neural Networks 277

8 Adaptive Resonance Theory Mapping (ARTMAP) Neural Networks 280

 8.1 Potential Advantages of ARTMAP 280

8.2	ARTMAP Architecture and Methodolgy	281
8.3	Results from the Cascaded System	285
8.4	Symbolic Rule Extraction	285
9	Which Technique Produces the Most Suitable Decision Support System?	289
	Acknowledgments	290
	References	291

Chapter 8.

	Xcyt: A System for Remote Cytological Diagnosis and Prognosis of Breast Cancer	297
	<i>W.N. Street</i>	
1	Introduction	297
2	Imaging	299
2.1	Sample Preparation	300
2.2	Automatic Detection of Nuclei	301
2.3	Representation of Nuclear Boundaries	304
2.4	Algorithm Improvements	305
2.5	Nuclear Morphometric Features	306
3	Diagnosis	307
3.1	MSM-T: Machine Learning via Linear Programming	308
3.2	Predictive Results and Clinical Usage	310
4	Prognosis	311
4.1	Median-Based Separation	313
4.2	A Neural Network Approach	315
5	Remote Execution	320
6	Conclusions	321
	Acknowledgments	321
	References	322

Index