

# **Studies in Computational Intelligence**

Volume 735

## **Series editor**

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland  
e-mail: [kacprzyk@ibspan.waw.pl](mailto:kacprzyk@ibspan.waw.pl)

### *About this Series*

The series “Studies in Computational Intelligence” (SCI) publishes new developments and advances in the various areas of computational intelligence—quickly and with a high quality. The intent is to cover the theory, applications, and design methods of computational intelligence, as embedded in the fields of engineering, computer science, physics and life sciences, as well as the methodologies behind them. The series contains monographs, lecture notes and edited volumes in computational intelligence spanning the areas of neural networks, connectionist systems, genetic algorithms, evolutionary computation, artificial intelligence, cellular automata, self-organizing systems, soft computing, fuzzy systems, and hybrid intelligent systems. Of particular value to both the contributors and the readership are the short publication timeframe and the worldwide distribution, which enable both wide and rapid dissemination of research output.

More information about this series at <http://www.springer.com/series/7092>

Krzysztof Dyczkowski

# Intelligent Medical Decision Support System Based on Imperfect Information

The Case of Ovarian Tumor Diagnosis

Krzysztof Dyczkowski  
Faculty of Mathematics and Computer  
Science, Department of Imprecise  
Information Processing Methods  
Adam Mickiewicz University  
Poznań  
Poland

ISSN 1860-949X ISSN 1860-9503 (electronic)  
Studies in Computational Intelligence  
ISBN 978-3-319-67004-1 ISBN 978-3-319-67005-8 (eBook)  
<https://doi.org/10.1007/978-3-319-67005-8>

Library of Congress Control Number: 2017951167

© Springer International Publishing AG 2018

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by Springer Nature  
The registered company is Springer International Publishing AG  
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

*I dedicate this book to my Family: my wife  
Anna, sons Kuba and Michał, and to my  
Parents*

# Preface

This monograph is the result of scientific research conducted between 2012 and 2016 by an interdisciplinary team of scientists from the Department of Imprecise Information Processing Methods at Faculty of Mathematics and Computer Science of Adam Mickiewicz University (AMU) in Poznań and The Division of Gynecologic Surgery of the Poznań University of Medical Sciences (PUMS). The AMU team includes prof. dr hab. Maciej Wygralak, dr inż. Anna Stachowiak, dr Patryk Żywica, and mgr Andrzej Wójtowicz. The PUMS team is composed of: prof. dr hab. n.med Dariusz Szpurek, dr hab. n.med. Rafał Moszyński, and dr n.med. Sebastian Szubert. The author of this monograph is the project leader.

Medical diagnosis and the proper differentiation of ovarian tumors are very important issues. Every year, about 2500 women die in Poland due to this type of cancer and over 14,000 in the USA. Poland is at the top of the list with the highest mortality due to this cancer. In addition, these tumors are difficult to diagnose because they give too few unambiguous symptoms in the early stage. Differentiating this form of tumor is in many cases difficult and requires a physician with extensive experience, advanced medical equipment, and repeated use of imprecise and incomplete data.

Physicians specializing in this type of disease have noted the need for a computerized system to support the diagnosis of such tumors, which would help the less-experienced physicians in the diagnosis and differentiation of tumors. There is also a need for a central database for standardized gathering of medical information on cases of patients treated in different centers in Poland, which would allow us to improve the developed diagnostic algorithms in the future.

The book outlines the theoretical foundations and the description of the *OvaExpert* system created by our team. The system supports the physician in the diagnosis of ovarian tumors using computational intelligence methods with particular focus on the problem of data incompleteness. The first part of the book introduces basic information on medical diagnosis and the diagnosis of ovarian tumors. The statistics concerning this disease have also been presented. In the next part of the book, we have introduced the necessary elements of fuzzy set theory and its extensions such as IVFSs and IFSs. The key part of the monograph is the

description of original algorithms based on the cardinalities of IVFSs. The algorithm descriptions are supplemented with the analysis of the efficacy of the presented algorithms, performed on the data of patients treated in the Poznań medical center. The last chapter describes the *OvaExpert* system, its key elements, and technologies and analyzes the prediction efficacy of algorithms used in the system against other existing methods. The book ends with a brief description of the ongoing and planned research and development works connected with the system.

I would like to thank the whole team involved in the project, without whom this book could not have been written. In particular, I would like to thank prof. dr hab. Maciej Wygralak for supporting me throughout the years, for pointing me in scientific directions and motivating to do further research. I extend my gratitude to colleagues from my Department: dr Anna Stachowiak, dr Andrzej Wójtowicz, and dr Patryk Żywica for their daily cooperation and hard work, commitment, as well as conceptual and technical support. I am grateful to the doctors from the Poznań University of Medical Sciences who cooperated with us, and especially to dr hab. Rafał Moszyński and dr n.med. Sebastian Szubert whom I would like to thank for introducing me to the medical realm of knowledge and for tremendous support during the creation of the system, and to prof. dr hab. Dariusz Szpurek, for his interest in the subject and creating favorable conditions for cooperation between our teams.

Finally, I would like to thank my beloved wife, Anna, for her understanding throughout the whole time which I devoted to my research and for her unwavering support in difficult times.

Poznań, Poland  
April 2017

Krzysztof Dyczkowski

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Medical Foundations</b>	<b>5</b>
2.1	Computer-Based Systems for Decision-Making Support in Medicine	5
2.1.1	Solutions in Medical Diagnosis Support in the World	8
2.1.2	Solutions in Cancer Diagnosis Support in the World	9
2.2	Ovarian Tumors	11
2.2.1	The Concept of a Tumor	11
2.2.2	The Scale of the Problem	12
2.2.3	Diagnosis and Differentiation of Tumors	15
2.2.4	Diagnostic Models	17
2.3	Methods of Classification Evaluation	23
2.3.1	Types of Classification Errors	23
2.3.2	Performance Measures	24
<b>3</b>	<b>Elements of Fuzzy Set Theory</b>	<b>29</b>
3.1	Fuzzy Sets	29
3.2	Triangular Operations and Negations	30
3.2.1	The Notion of a Triangular Norm	30
3.2.2	Basic Families of Triangular Operations	33
3.2.3	Negations	34
3.3	Basic Operations on Fuzzy Sets	35
3.4	Cardinality of a Fuzzy Set	37
3.4.1	Scalar Approach	38
3.4.2	Fuzzy Approach	42
3.4.3	Fuzzy Approach with the Use of Cardinality Patterns	46
3.5	Data Aggregation	50
3.5.1	Aggregation Operators	50



<b>4</b>	<b>Cardinalities of Interval-Valued Fuzzy Sets and Their Applications in Decision Making with Imperfect Information</b>	55
4.1	Decision Making with Imperfect Data	55
4.2	Interval-Valued Fuzzy Sets	56
4.2.1	Operations on Interval-Valued Fuzzy Sets	58
4.3	Atanassov's Intuitionistic Fuzzy Sets	58
4.4	Scalar Cardinality of IVFSs ( <i>Sigma F-Counts</i> )	60
4.5	Fuzzy Cardinality of IVFSs ( <i>f-FCounts</i> )	63
4.6	Decision Making Model Based on Cardinalities of IVFSs	71
4.6.1	Introduction to Group Decision Making (GDM)	71
4.6.2	Decision Algorithms	72
4.6.3	Algorithm Using Scalar Cardinality <i>Sigma F-Count</i>	74
4.6.4	Algorithm Using Scalar Cardinality <i>f-FEcount</i>	77
4.7	Evaluation of Algorithms Efficacy	79
4.7.1	Decision Intervals Modelling	81
4.7.2	Uncertaintification of Diagnostic Models	82
4.7.3	Decision Making Procedure	83
4.7.4	Evaluation	84
4.7.5	Results	88
<b>5</b>	<b>OvaExpert System</b>	95
5.1	Features of the OvaExpert System	95
5.1.1	Medical Data Gathering	96
5.1.2	Easy and Intuitive User Interface	96
5.1.3	Multiple Diagnostic Models	98
5.1.4	Bipolar Presentation of Diagnosis	98
5.1.5	Gathering of Expert Knowledge	99
5.1.6	System Application in the Diagnosis Process	99
5.1.7	System Architecture	99
5.2	Diagnostic Models	103
5.2.1	Decision Module Based on Aggregation of Diagnostic Models (OEA)	104
5.2.2	Decision Module Based on Interval-Valued Fuzzy Set Cardinality (FSC)	105
5.2.3	Decision Module Based on Prototypes and Uncertainty-Aware Similarity Measure (IVFC)	105
5.2.4	Prognostic Efficacy of Individual Decision Modules	106
<b>6</b>	<b>Summary</b>	109
6.1	Ongoing Research	109
6.1.1	Using Fuzzy Control (Rule-Based System) as a Prediction Method	110
6.1.2	Application of Deep Learning	110

Contents	xi
6.1.3 Using Stochastic Orderings for Classifier Evaluation . . . . .	110
6.2 Future Plans for Development . . . . .	110
<b>References</b> . . . . .	113
<b>Index</b> . . . . .	121

# Acronyms and Abbreviations

Acc	Accuracy
ADNEX	The Assesment of Different Neoplasia in the adneXa
Alc	Alcazar's model
AUC	Area under an ROC curve
BI-RADS	Breast Imaging Reporting Data System
BMI	Body Mass Index
CBR	Case-Based Reasoning
CEE	Central and Eastern Europe
COG	Center of Gravity
DBMS	Database management system
Dec	Decisiveness
EBM	Evidence-Based Medicine
FCS	Family of all Crisp Sets
FDA	Food and Drug Administration
FE	FECCount
FFS	Family of all Fuzzy Sets
FG	FGCount
FIGO	International Federation of Gynecology and Obstetrics
FL	FLCount
FN	False Negative
FP	False Positive
FSC	<i>Ovaexpert</i> decision method based on counting
GDM	Group Decision Making
GFE	Family of all generalized cardinal numbers
GI-RADS	Gynecologic Imaging Reporting and Data System
GP	General Practitioner
IFS	Atanassov's Intuitionistic Fuzzy Set
IOTA	International Ovarian Tumor Analysis Group
IVFC	OvaExpert decision method based on Interval-Valued Fuzzy Classifier
IVFS	Interval-Valued Fuzzy Set

LR	Logistic Regression
LR1	IOTA Logistic regression-based model 1
LR2	IOTA Logistic regression-based model 2
Max	Maximum
Min	Minimum
MSK	Memorial Sloan-Kettering Cancer Center
MVC	Model-View-Controller
NICE	National Institute for Health and Care Excellence
OEA	OvaExpert decision method based on aggregation
OECD	The Organization for Economic Cooperation and Development
ORM	Object-Relational Mapping
OWA	Ordered Weighted Averaging
PACS	Picture Archiving and Communication System
PBD	Percent Breast Density
PDI	Polytomous Discrimination Index
PPV	Positive Predictive Value
Prec	Precision
RIS	Radiology Information System
RMI	Risk of Malignancy Index
ROC	Receiver Operating Characteristic
ROMA	Risk of Ovarian Malignancy Algorithm
SaaS	Software as a Service
SD	Doppler Index
SDR	Standardized Death Rate
Sen	Sensitivity
SM	Sonomorphological Index
Sneg	Family of all strong negations
Spec	Specificity
SR	IOTA Simple Rules diagnostic model
Supp	Support of a fuzzy set
Tim	Timmerman's logistic regression model
TN	True Negative
TNR	True Negative Rate
TP	True Positive
TPR	True Positive Rate
USG	Ultrasonography

# List of Figures

Fig. 2.1	Ovarian cancer—number of new cases and deaths per 100,000 people . . . . .	15
Fig. 2.2	IOTA simple rules. . . . .	21
Fig. 2.3	Example of ROC curve for a classifier based on CA125 and HE4 biomarkers, for which $AUC = 0.927$ . . . . .	26
Fig. 3.1	Cardinality pattern $f_{1,t,p}$ for $t = 0.5$ and $p = 2$ . . . . .	39
Fig. 3.2	Cardinality pattern $f_{2,t,p}$ for $t = 0.5$ and $p = 2$ . . . . .	39
Fig. 3.3	Cardinality pattern $f_{3,t_1,t_2,p}$ for $t_1 = 0.3$ , $t_2 = 0.7$ and $p = 1$ . . . . .	40
Fig. 3.4	Cardinality pattern $f_{4,t_1,t_2}$ for $t_1 = 0.3$ , $t_2 = 0.7$ . . . . .	40
Fig. 3.5	Cardinality pattern $f_{5,t,p}$ for $t = 0.5$ , $p = 12$ . . . . .	41
Fig. 3.6	Generalized cardinal numbers of set $A$ (3.39): <b>a</b> $FG(A)$ , <b>b</b> $FL(A)$ , <b>c</b> $FE(A)$ . . . . .	48
Fig. 3.7	Generalized f-weighted cardinal numbers of set $A$ (3.53) and cardinality pattern $f_{1,t}$ (3.25) with parameters $t = 0.5$ , $p = 2$ : <b>a</b> $FG_f(A)$ , <b>b</b> $FL_f(A)$ , <b>c</b> $FE_f(A)$ . . . . .	48
Fig. 3.8	Generalized f-weighted cardinal numbers of set $A$ (3.53) and cardinality pattern $f_{2,t,p}$ (3.26) with parameters $t = 0.5$ and $p = 2$ : <b>a</b> $FG_f(A)$ , <b>b</b> $FL_f(A)$ , <b>c</b> $FE_f(A)$ . . . . .	48
Fig. 3.9	Generalized f-weighted cardinal numbers of set $A$ (3.53) and cardinality pattern $f_{3,t_1,t_2,p}$ (3.27) with parameters $t_1 = 0.3$ , $t_2 = 0.7$ , $p = 1$ : <b>a</b> $FG_f(A)$ , <b>b</b> $FL_f(A)$ , <b>c</b> $FE_f(A)$ . . . . .	49
Fig. 3.10	Generalized f-weighted cardinal numbers of set $A$ (3.53) and cardinality pattern $f_{4,t_1,t_2}$ (3.28) with parameters $t_1 = 0.3$ , $t_2 = 0.7$ : <b>a</b> $FG_f(A)$ , <b>b</b> $FL_f(A)$ , <b>c</b> $FE_f(A)$ . . . . .	49
Fig. 3.11	Generalized f-weighted cardinal numbers of set $A$ (3.53) and cardinality pattern $f_{5,t,p}$ (3.29) with parameters $t = 0.5$ , $p = 12$ : <b>a</b> $FG_f(A)$ , <b>b</b> $FL_f(A)$ , <b>c</b> $FE_f(A)$ . . . . .	50
Fig. 4.1	Interval-valued fuzzy set $\tilde{A}$ (4.5). . . . .	58
Fig. 4.2	Atanassov's intuitionistic fuzzy set $\mathcal{A}$ (4.13). . . . .	59

Fig. 4.3	Graphical interpretation of IVFS $\tilde{A}$ (4.23).....	62
Fig. 4.4	Graphical interpretation of IVFS $\tilde{A}$ (4.30).....	64
Fig. 4.5	Fuzzy cardinalities of IVFS $\tilde{A}$ (3.53) for cardinality pattern $f = f_{id}$ : <b>a</b> $FG_f(\tilde{A})$ , <b>b</b> $FL_f(\tilde{A})$ , <b>c</b> $FE_f(\tilde{A})$ .....	64
Fig. 4.6	Graphical interpretation of IVFS $\tilde{A}$ (4.32).....	65
Fig. 4.7	Fuzzy cardinalities of IVFS $\tilde{A}$ (3.53) with cardinality pattern $f = f_{id}$ : <b>a</b> $FG_f(\tilde{A})$ , <b>b</b> $FL_f(\tilde{A})$ , <b>c</b> $FE_f(\tilde{A})$ .....	65
Fig. 4.8	Graphical interpretation of IVFS $\tilde{A}$ (4.34).....	66
Fig. 4.9	Fuzzy cardinalities of IVFS $A$ (3.53) with cardinality pattern $f = f_{id}$ : <b>a</b> $FG_f(\tilde{A})$ , <b>b</b> $FL_f(\tilde{A})$ , <b>c</b> $FE_f(\tilde{A})$ .....	66
Fig. 4.10	Graphical interpretation of IVFS $\tilde{A}$ (4.36).....	67
Fig. 4.11	Fuzzy cardinalities of IVFS $\tilde{A}$ (3.53) with cardinality pattern $f = f_{id}$ : <b>a</b> $FG_f(\tilde{A})$ , <b>b</b> $FL_f(\tilde{A})$ , <b>c</b> $FE_f(\tilde{A})$ .....	67
Fig. 4.12	Graphical interpretation of IVFS $\tilde{A}$ (4.38).....	68
Fig. 4.13	Fuzzy cardinalities of IVFS $\tilde{A}$ (4.38) with cardinality pattern $f = f_{id}$ : <b>a</b> $FG_f(\tilde{A})$ , <b>b</b> $FL_f(\tilde{A})$ , <b>c</b> $FE_f(\tilde{A})$ .....	68
Fig. 4.14	Fuzzy cardinalities of IVFS $\tilde{A}$ (4.38) with cardinality pattern $f = f_1$ (3.25): <b>a</b> $FG_f(\tilde{A})$ , <b>b</b> $FL_f(\tilde{A})$ , <b>c</b> $FE_f(\tilde{A})$ .....	69
Fig. 4.15	Fuzzy cardinalities of IVFS $\tilde{A}$ (4.38) with cardinality pattern $f = f_2$ (3.26): <b>a</b> $FG_f(\tilde{A})$ , <b>b</b> $FL_f(\tilde{A})$ , <b>c</b> $FE_f(\tilde{A})$ .....	69
Fig. 4.16	Fuzzy cardinalities of IVFS $\tilde{A}$ (4.38) with cardinality pattern $f = f_3$ (3.27): <b>a</b> $FG_f(\tilde{A})$ , <b>b</b> $FL_f(\tilde{A})$ , <b>c</b> $FE_f(\tilde{A})$ .....	70
Fig. 4.17	Fuzzy cardinalities of IVFS $\tilde{A}$ (4.38) with cardinality pattern $f = f_4$ (3.28): <b>a</b> $FG_f(\tilde{A})$ , <b>b</b> $FL_f(\tilde{A})$ , <b>c</b> $FE_f(\tilde{A})$ .....	70
Fig. 4.18	Fuzzy cardinalities of IVFS $\tilde{A}$ (4.38) with cardinality pattern $f = f_5$ (3.29): <b>a</b> $FG_f(\tilde{A})$ , <b>b</b> $FL_f(\tilde{A})$ , <b>c</b> $FE_f(\tilde{A})$ .....	71
Fig. 4.19	Example of six information sources as IVFS $\tilde{D}$ .....	73
Fig. 4.20	Graphical interpretation of IVFS $\tilde{A}$ (4.46).....	76
Fig. 4.21	Cardinalities of IVFSs <b>a</b> $FE_f(\tilde{P})$ , <b>b</b> $FE_f(\tilde{C})$ .....	78
Fig. 4.22	Cardinalities of IVFSs <b>a</b> $FE_f(\tilde{P})$ , <b>b</b> $FE_f(\tilde{C})$ .....	79
Fig. 4.23	Distribution of patients with respect to percentage-wise gaps in diagnostic data .....	80
Fig. 4.24	Graphical summary of the current and the proposed approach. <i>Rectangles</i> represent diagnostic models on various stages. <i>Vertical arrows</i> designate transformations of diagnostic models (uncertaintification and counting). The third stage (decision) has been represented as an ellipsis. <i>Horizontal arrows</i> designate flow of patient data and diagnosis .....	83

Fig. 4.25	Division of data sets. Data of patients with more than 50% of missing data have not been taken into account in the experiment. . . . .	84
Fig. 4.26	Distribution of malignity classes in training and test sets . . . . .	85
Fig. 4.27	Visualization of the training phase. Data flow is represented by <i>arrows</i> and <i>rectangles</i> represent operations on data . . . . .	86
Fig. 4.28	Results achieved by original diagnostic models (with respect to total costs) with various missing data levels . . . . .	87
Fig. 4.29	Results achieved by uncertaintified diagnostic models (with respect to total costs) with various missing data levels . . . . .	89
Fig. 4.30	Results achieved by algorithms based on cardinality <i>sigma f-Count</i> . The graphs presents one algorithms from each group . . . . .	89
Fig. 4.31	Results achieved by algorithms based on <i>f-FEcount</i> . The graphs presents one algorithms from each group . . . . .	90
Fig. 4.32	Comparison of quality measure results for original diagnostic models on test set. . . . .	90
Fig. 4.33	Comparison of quality measure results for uncertaintified diagnostic models on test set. . . . .	90
Fig. 4.34	Results of prediction efficecy indicators achieved by individual algorithms on test data. . . . .	91
Fig. 4.35	Comparison of decision power of methods <i>FSC</i> and <i>OEA</i> on training data depending on amount of missing data. Efficacy area of original models has been marked in <i>grey</i> . . . . .	92
Fig. 4.36	Efficacy of decision methods <i>FSC</i> and <i>OEA</i> in comparison to efficacy of original diagnostic models on test data . . . . .	92
Fig. 5.1	An example of the OvaExpert user interface. . . . .	97
Fig. 5.2	A screenshot of the OvaExpert system presenting the recommended diagnosis in a bipolar form. . . . .	100
Fig. 5.3	A diagram of interaction between the physician and the OvaExpert system. . . . .	101
Fig. 5.4	UML diagram showing key features of the OvaExpert system (Żywica et al. [164]) . . . . .	102
Fig. 5.5	Method of medical data anonymization in the OvaExpert system . . . . .	103

# List of Tables

Table 2.1	FIGO classification for ovarian tumors (after Society of Gynecologic Oncology [112]) . . . . .	12
Table 2.2	Countries with the highest incidence of ovarian cancer in 2012 . . . . .	14
Table 2.3	Types of classification errors . . . . .	23
Table 2.4	Types of classification errors taking into account lack of decision . . . . .	27
Table 4.1	Diagnostic attributes used for selected diagnostic models . . . . .	80
Table 4.2	Cost matrix used for models assessment . . . . .	87
Table 4.3	Decision efficacy results for algorithms based on scalar cardinality of type <i>sigma f-Count</i> for various decision algorithms and cardinality patters with bootstrap percentile 95% confidence intervals, achieved on test data. Decisiveness, sensitivity, specificity are expressed in percentages . . . . .	91
Table 4.4	Decision efficacy results for algorithms based on fuzzy cardinality of type <i>f-FECount</i> for various decision algorithms and cardinality patters with bootstrap percentile 95% confidence intervals, achieved on test data. Decisiveness, sensitivity, specificity are expressed in percentages . . . . .	93
Table 5.1	Quality measures of the original diagnostic models compared with the new models of the OvaExpert system. Results obtained on test set . . . . .	106
Table 5.2	The McNemar test with Benjamini-Hochberg correction between the original diagnostic models and the decision modules ( $\alpha = 0.05$ ). Results obtained on test set . . . . .	107



# Abstract

This book deals with a computer-supported medical diagnosis with particular focus on ovarian tumor diagnosis. The book presents theoretical foundations (both medical and mathematical) of the intelligent *OvaExpert* system which supports the decision-making process in tumor diagnosis. The main purpose for creating *OvaExpert* was to support a gynecologist in the process of predicting the malignancy of ovarian tumors by applying the existing diagnostic models and using modern methods of computational intelligence which allow for imprecision and imperfection of the medical data, both of which are common features of the everyday medical practice. We had good reasons to focus on this particular cancer. Ovarian cancer is difficult to diagnose and has high mortality, especially in the Central and Eastern Europe. The book presents novel methods based on interval-valued fuzzy sets and the theory of their cardinalities. The algorithms applied have been verified by analyzing their efficacy on real data.