

Radim Bris  
Jaroslav Majerník  
Krzysztof Pancerz  
Elena Zaitseva *Editors*

# Applications of Computational Intelligence in Biomedical Technology

# **Studies in Computational Intelligence**

Volume 606

## **Series editor**

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland  
e-mail: 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>

Radim Bris · Jaroslav Majerník  
Krzysztof Pancerz · Elena Zaitseva  
Editors

# Applications of Computational Intelligence in Biomedical Technology



Springer

*Editors*

Radim Bris

Katedra Aplikované Matematiky  
VŠB—Technical University of Ostrava  
Ostrava-Poruba  
Czech Republic

Jaroslav Majerník

Pavol Jozef Safárik University  
Kosice  
Slovakia

Krzysztof Pancerz

University of Information Technology  
and Management in Rzeszow  
Rzeszow  
Poland

Elena Zaitseva

Faculty of Management Science  
and Informatics  
University of Zilina  
Zilina  
Slovakia

ISSN 1860-949X

ISSN 1860-9503 (electronic)

Studies in Computational Intelligence

ISBN 978-3-319-19146-1

ISBN 978-3-319-19147-8 (eBook)

DOI 10.1007/978-3-319-19147-8

Library of Congress Control Number: 2015941160

Springer Cham Heidelberg New York Dordrecht London

© Springer International Publishing Switzerland 2016

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.

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media  
([www.springer.com](http://www.springer.com))

# Preface

Biomedical Technology represents one of the most dynamically developed scientific area that supports research activities aimed to discover, create and develop innovative applications of technological solutions to handle different biomedical problems. Technology development often requires multidisciplinary and team-oriented approaches and may lead to the realization of new and/or improved instruments that have substantial influence on quality and reliability of medical and biomedical progress and further research. In general, the Biomedical Technologies developed by technicians are transferred into practice by specialists in medicine. Here, the physicians utilize such technologies mainly to help their patients solve their health problems as best as possible and return them back to normal life as soon as possible. On the other hand, the managers of various institutions included in the healthcare system can create proper conditions to support physicians in their innovative and advanced work by accepting of new methods and by personal or even financial interventions. Computational Intelligence and Informatics play a dominant role in Biomedical Technologies as it is an actual.

There are various scientific teams all over the world attempting to develop modern and innovative methods for this field of research. Most of them usually consist of members that are specialists in some areas of Biomedical Technologies, in development of special tools, devices, equipment, methodologies, etc. All such teams have their unique experiences and knowledge based on different backgrounds, mathematical approaches, and techniques applied in investigation activities to reach their goals. Exchange of scientific results and combination of experiences obtained by these scientific teams can be very useful to generate international cooperation based on professionals in their areas as well as for the next investigation and development in Biomedical Technologies. Therefore, the papers included in this book represent a good opportunity to present and to find the latest results in selected parts of Biomedical Technologies. Most of them deal with problems of Biomedical and Medical Informatics, ranging from theoretical considerations to practical applications.

The authors of these book chapters discuss a lot of specific aspects of development methods and algorithms in Biomedical and Medical Informatics.

Algorithms for medical image processing, modelling methods and tool, medical decision-making support, estimation of risks of treatments, reliability of medical systems, problems of practical clinical applications and many other topics are covered by individual contributions. Decisions of different problems in development of medical decision-making support system are presented. For example, decision-making process connected with the choice of right operation technique plays a big role as a risk decreasing factor taking into account the risk of post-operative mortality and morbidity as well. Having a good mathematical (statistical) model a decision-maker can assess and predict results of the expensive Phadiatop medical test. Cox analysis can be used as a useful tool for the incidence of cancer, etc. Interesting results are discussed for estimation of risk in treatment and examination of reliability of medical system.

The book is composed by selected contributions presented and discussed at the International Workshop on Biomedical Technologies that was organized under the 10th International Conference on Digital Technologies 2014 (DT 2014) and held in July 2014 in Zilina, Slovakia. The organization of this Workshop was mainly possible thanks support of the International Visegrad Fund by Standard Grant No. 21320401 "International Workshop on Biomedical Technologies".

This book is intended for scientists interested in problems of Biomedical Technologies, for researchers and academic staff, for all dealing with Biomedical and Medical Informatics, for Ph.D. students, etc. Useful information is offered also to IT companies, developers of equipment and/or software for medicine and medical professionals as well.

We thank Jozef Kostolny for his help in the book pre-publishing preparation and technical support.

Radim Bris  
Jaroslav Majerník  
Krzysztof Pancerz  
Elena Zaitseva

# Contents

<b>Mesh-Based Modeling of Individual Cells and Their Dynamics in Biological Fluids . . . . .</b>	<b>1</b>
Ivan Cimrák, Iveta Jančigová, Renáta Tóthová and Markus Gusenbauer	
<b>Usage of Industrial Computed Tomography for Evaluation of Custom-Made Implants . . . . .</b>	<b>29</b>
Radovan Hudák, Jozef Živčák, Teodor Tóth, Jaroslav Majerník and Martin Lisý	
<b>Assessment of Mortality Risk for Patients Undergoing Colorectal Surgery Using Regression Modeling . . . . .</b>	<b>47</b>
Kateřina Janurová and Lubomír Martínek	
<b>Profiling of Software Requirements for the Pharmaceutical Enterprise Manufacturing Execution System . . . . .</b>	<b>67</b>
Vyacheslav Kharchenko, Oleksandr Gordieiev and Alina Fedoseeva	
<b>Application of Survival Analysis on Analysing the Association Between Chromosomal Aberrations and Carcinoma . . . . .</b>	<b>93</b>
Veronika Kubíčková	
<b>Modelling the Results of the Phadiatop Test Using the Logistic and Ordinal Regression . . . . .</b>	<b>103</b>
Pavlína Kuráňová	
<b>Complex Data Management in MRI Results Processing . . . . .</b>	<b>119</b>
Michal Kvet, Monika Vajsová and Karol Matiaško	
<b>Biometrics for Biomedical Applications . . . . .</b>	<b>143</b>
Kenneth Lai, Steven Samoil, Svetlana N. Yanushkevich and Gaetan Collaud	

<b>Multimedia Support of Medical Education Utilizing On-Line Archive of Video Lectures . . . . .</b>	<b>163</b>
Jaroslav Majerník and Pavol Jarčuška	
<b>Evaluation of Postural Stability Using Motion Analysis Techniques . . . . .</b>	<b>175</b>
Jaroslav Majerník and Jozef Živčák	
<b>Some Issues on an Object-Oriented Programming Language for Physarum Machines . . . . .</b>	<b>185</b>
Krzysztof Pancerz and Andrew Schumann	
<b>A Tool for Computer-Aided Diagnosis of Psychological Disorders Based on the MMPI Test: An Overview . . . . .</b>	<b>201</b>
Krzysztof Pancerz, Olga Mich, Andrzej Burda and Jerzy Gomuła	
<b>The Instance and Feature Selection for Neural Network Based Diagnosis of Chronic Obstructive Bronchitis . . . . .</b>	<b>215</b>
Sergey Subbotin	
<b>New Methods for the Reliability Analysis of Healthcare System Based on Application of Multi-State System . . . . .</b>	<b>229</b>
Elena Zaitseva, Vitaly Levashenko, Jozef Kostolný and Miroslav Kvassay	

# Mesh-Based Modeling of Individual Cells and Their Dynamics in Biological Fluids

Ivan Cimrák, Iveta Jančigová, Renáta Tóthová  
and Markus Gusenbauer

**Abstract** This text is aimed at providing both basic and advanced knowledge on the individual cell modeling in a flow. Besides the overview of various existing approaches, it is focused on mesh-based model and on its capabilities to cover complex mechano-elastic properties combined with adhesion and magnetic phenomena. We also describe validation procedures, offer an example of use of the model for better understanding of cell behavior and a short overview of future research directions.

## 1 Introduction

Computational models are able to handle multiple dynamically interacting variables and parameters, such as different elastic properties of cells, different environmental factors, or various geometries in microfluidic channels. This ability enables the investigation of how individual components of a system contribute to its overall performance via systematic manipulation of these variables and parameters.

In computer simulations, every parameter ranges over a set of values and multiple parameters can be varied simultaneously. This is in most cases impossible to achieve experimentally in an accurate and controlled fashion. Computational modeling gives us a tool to follow the behavior of the whole system as well as of individual cells.

To solve clinically relevant blood flow problems and test hypotheses regarding hemodynamic factors in vascular adaptation and disease, blood is often considered as a homogeneous viscous fluid. The underlying mathematical model for this flow is often based on Navier-Stokes equations or lattice-Boltzmann method [50].

---

I. Cimrák (✉) · I. Jančigová · R. Tóthová

Faculty of Management Science and Informatics, University of Žilina, Žilina, Slovakia  
e-mail: ivan.cimrak@fri.uniza.sk  
url: <http://cell-in-fluid.fri.uniza.sk>

M. Gusenbauer

Center for Integrated Sensor Systems, Danube University Krems,  
Krems an der Donau, Austria

In studies of sick (e.g. malaria-infected) cells, circulating tumor cells, or any other studies concerning special cells, one needs to capture the behavior of individual cells. In these cases, we need a model, in which the individual cells are separate entities with their own elastic properties. We have developed such model, with two principal components: fluid and elastic objects that are immersed in the fluid.

For the fluid, we use the lattice-Boltzmann method [6]. This method is based on fictive particles and gives the velocity, density and other properties of the fluid in each time step on a fixed grid. In this method, the governing equation is the lattice-Boltzmann equation describing the time evolution of the density of fictive particles.

The second component of the model governs elastic objects such as red blood cells (RBCs), platelets, any other type of biological cells or rigid objects, for example magnetic micro-beads. Our model is based on mechano-elastic properties of the object and covers the elasticity of the cell's membrane, represents solid structure of a rigid micro-bead, or it can include the influence of cell's internal structure on the elastic behavior of the cell. Additionally, the model is able to capture not only mechanical but also biological processes during the cell adhesion.

Both components, the fluid and the elastic objects, interact. The coupling is implemented through the drag force between the fluid and the object's mesh points. As a consequence, additional terms enter the lattice-Boltzmann equation representing external forces resulting from the deformation of the objects. Analogously, fluid forces are transferred to the object's mesh points causing not only the deformation but also the movement of the object.

We are primarily interested in blood flow in microfluidic devices, so our computational domains are rigid box-type channels with or without obstacles. In the following section we describe the model in detail. In Sect. 2.1 we introduce spring-network models that govern elastic behavior of the objects. The linking of the mesh with adhesion mechanism is described in Sect. 2.2. Here we show how the biological ligand-receptor model of adhesion can be modeled using the forces exerted on individual mesh points. In Sect. 2.4 we describe the implementation of the model into a computational framework Object-in-fluid.

In next Sect. 3 we present other models used for elastic objects in a flow. We focus on spring-network models and their different approaches to modeling the elasticity. Further, we discuss the ways how to employ the fluid-structure interactions.

To ensure whether the model corresponds to reality we present some validation procedures in Sects. 4.1, 4.2. We show how elastic parameters are determined to model correct stretching of a single red blood cell. Further we present a simplification of the adhesion model. The computational capabilities of the model are discussed in Sect. 4.3.

Besides the development of the model, we are interested in the applications of this model. In Sect. 5 we show first attempt to use our model for better understanding of cell behavior. We present analysis of local stresses of the cell's membrane during the motion in shear flow. Such experiment has not been performed experimentally.

In the last section we list several interesting directions for further research.

## 2 Model

The model consists of two parts. Lattice-Boltzmann (LB) method is used to govern the time-dependent evolution of the fluid. For the objects immersed in the fluid we utilize the immersed boundary (IB) method.

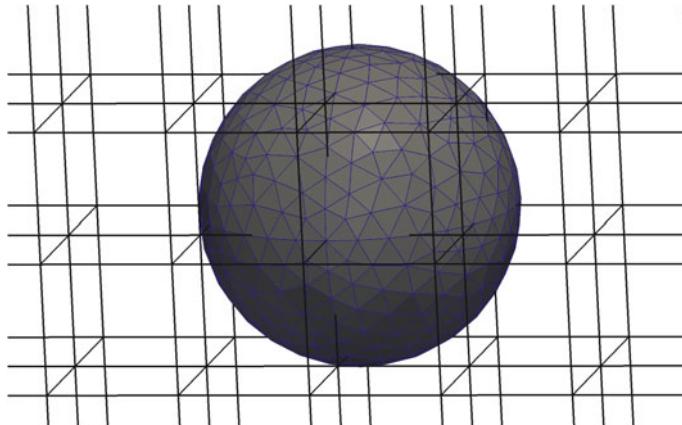
**Lattice Boltzmann method** is based on fictive particles. These particles perform consecutive propagations and collisions over a three-dimensional discrete lattice consisting of identical cubic cells, see Fig. 1. This lattice creates an Eulerian grid which is fixed during the entire simulation. The variable of interest in the LB method is  $n_i(x, t)$  which is the particle density function for the lattice point  $x$ , discrete velocity vector  $e_i$ , and time  $t$ . We use the D3Q19 version of the LB method (three dimensions with 19 discrete directions  $e_i$ , so  $i = 0, \dots, 18$ ). The governing equations in the presence of external forces, are

$$\underbrace{n_i(x + e_i \delta_t, t + \delta_t)}_{\text{propagation}} = n_i(x, t) - \underbrace{\Delta_i(\mathbf{n}(x, t))}_{\text{collision}} + \underbrace{f_i(x, t)}_{\text{external forces}} \quad (1)$$

where  $\delta_t$  is the time step and  $\Delta_i$  denotes the collision operator that accounts for the difference between pre- and post-collision states and satisfies the constraints of mass and momentum conservation.  $f_i$  is the external force exerted on the fluid. The macroscopic quantities such as velocity  $u$  and density  $\rho$  are evaluated from

$$\rho(x, t) = \sum_i n_i(x, t) \quad \text{and} \quad \rho(x, t)u = \sum_i n_i(x, t)e_i.$$

We refer to [1] and [17] for details on this implementation of the LB method.



**Fig. 1** Two different meshes in the model. The regular Eulerian cubic grid is used for lattice-Boltzmann method and unstructured Lagrangian triangular mesh is used for immersed boundary method

**Immersed Boundary method** is based on the discretization of object's boundary. The boundary is covered by a set of IB points linked by triangular mesh, which is called Lagrangian mesh, see Fig. 1. The positions of IB points are not restricted to any lattice. To take the mechano-elastic properties of the immersed objects into account, geometrical entities in this mesh (edges, faces, angles between two faces, ...) are used to model stretching, bending, stiffness, and other properties of the boundary. They define forces according to the current shape of the immersed object that are exerted on IB points. These forces cause motion of the IB points according to the Newton's equation of motion

$$m \frac{d^2 X_j}{dt^2} = F_j \quad (2)$$

where  $m$  is the mass of the IB point,  $X_j$  is the position and  $F_j$  is the force exerted on the particular IB point  $j$ . The sources of  $F_j$  are the above mentioned elasto-mechanical properties of the immersed object and the fluid-structure interaction.

Equations (1) and (2) describe two different model components on two different meshes: the motion of the fluid and the motion of the immersed objects. For coupling, we use an approach from [1] with a drag force that is exerted on a sphere moving in the fluid. Analogous to the Stokes formula for the sphere in a viscous fluid, we assume the force exerted by the fluid on one IB point  $j$  to be proportional to the difference of the velocity  $v$  of the IB point and the fluid velocity  $u$  at the same position,

$$F_j = \xi(v - u). \quad (3)$$

Here  $\xi$  is a proportionality coefficient which we will refer to as the friction coefficient. In the previous expression, the velocities  $v$  and  $u$  are computed at the same spatial location, whereas we posses  $u$  in fixed Eulerian grid points and  $v$  in moving Lagrangian IB points. Therefore, for computation of  $u$  at the IB point, we use linear interpolation of the  $u$  values from nearby fixed grid points.

There is also an opposite effect: not only fluid acts on the IB point, but also the IB point acts on the fluid. Therefore, the opposite force  $-F_j$  needs to be transferred back to the fluid.  $F_j$  from the location of the IB point is distributed to the nearest grid points. Distribution is inversely proportional to the cuboidal volumes with opposite corners being the IB point and the grid point [1].

From now on, we will not further elaborate on the lattice-Boltzmann equation. We are going to focus on other forces entering Eq. 2. In the following Sects. 2.1–2.3, we describe the forces resulting from object's elasticity, from its adhesion and magnetic phenomena as they relate to the objects.

## 2.1 Elasticity

The forces entering Eq. 2 originate either from the fluid-structure interaction, or from the elasticity of the object itself. To explain the idea behind the modeling of elasticity, mesh points may be considered to be connected with springs. Once

such spring is stretched, it exerts forces on the end-points. There are however more advanced mechanisms than springs acting between each two mesh points. Therefore, we prefer to call these mechanisms bonds rather than springs.

There are five elastic moduli describing elasticity of a red blood cell. We took the main ideas for force computations corresponding to different bonds from [18], however in [32, 33, 52] we concluded that different relations are more suitable.

Stretching force is identical to a force of a harmonic spring. It is responsible for tendency to preserve relaxed length of each edge in the triangular mesh. The actual expressions for force calculation are

$$\mathbf{F}_s(AB) = k_s \kappa(\lambda) \Delta L \mathbf{n}_{AB} \quad (4)$$

where  $k_s$  is the stretching coefficient and  $\kappa(\lambda)$  allows for nonlinearity of the stretching force (by making  $\lambda$  dependent on  $L$ ).  $L$  is the relaxed length of the edge  $AB$ ,  $\Delta L$  is the prolongation of this edge and  $\mathbf{n}_{AB}$  is the unit vector pointing from  $A$  to  $B$ .

Bending force is responsible for tendency to preserve the angle between two neighboring triangles. Expressions read as

$$\mathbf{F}_b(ABC) = k_b \frac{\Delta\theta}{\theta} \mathbf{n}_{ABC} \quad (5)$$

where  $k_b$  is the bending coefficient,  $\theta$  is the resting angle between two triangles that have common edge  $AB$ ,  $\Delta\theta$  is the deviation from this angle and  $\mathbf{n}_{ABC}$  is the unit normal vector to the triangle  $ABC$ .

Local area force keeps the area of individual triangles in the mesh fairly constant:

$$\mathbf{F}_a(A) = -k_a \frac{\Delta S_{ABC}}{\sqrt{S_{ABC}}} \mathbf{w}_A \quad (6)$$

where  $k_a$  is the local area coefficient,  $S_{ABC}$  is the relaxed area of triangle  $ABC$ ,  $\Delta S_{ABC}$  is the deviation from this resting state and  $\mathbf{w}_A$  is the unit vector pointing from the centroid of the triangle  $ABC$  to the vertex  $A$ . (Similar forces are assigned to vertices  $B$  and  $C$ ).

Global area force is similar to local area force, however, the difference is in keeping the overall object's surface constant:

$$\mathbf{F}_A(A) = -k_A \frac{\Delta S}{S} \mathbf{w}_A \quad (7)$$

where  $k_A$  is the global area coefficient,  $S$  is the relaxed area of the whole object,  $\Delta S$  is the deviation from this area and  $\mathbf{w}_A$  is again the unit vector pointing from the centroid of the triangle  $ABC$  to the vertex  $A$ .

Volume force is responsible for keeping the volume of the object constant:

$$\mathbf{F}_V(ABC) = -k_V \frac{\Delta V}{V} S_{ABC} \mathbf{n}_{ABC} \quad (8)$$

where  $k_V$  is the volume coefficient,  $S_{ABC}$  is the area of triangle  $ABC$ ,  $V$  is the volume of the whole object,  $\Delta V$  is the deviation from this relaxed volume and  $\mathbf{n}_{ABC}$  is the unit normal vector to the plane  $ABC$ .

The properties of the model are significantly influenced by elastic parameters  $k_s, k_b, k_a, k_A, k_V$ . To capture the proper behavior of real cells, one needs to calibrate these parameters according to data available from biological experiments. We address this topic in Sect. 4.1.

## 2.2 Adhesion

For applications concerning capture of circulating tumor cells, we need to implement the adhesion mechanisms. The biological concepts of adhesion rely on creating receptor-ligand complexes between the receptor patches located on the surface of the cell and ligands distributed on the functionalised surfaces. When a receptor patch (one of thousands on a single cell) approaches the ligand location, a bond between receptor and ligand (both being special types of proteins) is created with a given probability. This probability defines the rate of bond creation. Once the receptor-ligand complex is created, forces start to act attracting the cell to the surface. When the length of the receptor-ligand complex becomes larger (e.g. when the cell moves away from the surface), the bond breaks, and this happens again with a given probability.

Our model can be readily extended to include any other phenomenon based on forces exerted on the boundary mesh points of the elastic object. The adhesion mechanism introduced above is based on bonds between the receptor site on a cell and the ligand site on a wall. Once the stiffness  $\kappa$  of such bond is known, one can model this bond by exerting repulsive or attractive force on the corresponding mesh points.

The computation of the forces acting on the mesh points corresponding to a bond is based on the harmonic spring. If a bond exists and the distance between the receptor position and the ligand position is  $l$ , then the adhesion force is computed from

$$F_{adh} = \kappa l,$$

where  $\kappa$  is the stiffness of the bond. There are other possibilities, for example one can assume the resting length of the bond to be  $l_0$  and the force can then be proportional to the positive prolongation of the bond and zero otherwise

$$F_{adh} = \begin{cases} \kappa(l - l_0) & \text{for } l > l_0 \\ 0 & \text{otherwise.} \end{cases}$$

### 2.2.1 Dissociation and Association Dynamics

During the time the receptor and ligand are close enough, the bond is created with the on-rate  $k_{on}$ . This is represented in our model by the probability  $P_{on}$  of bond formation during time step  $\delta t$ . On the other hand, any bond can rupture with the off-rate  $k_{off}$  with dissociation probability  $P_{off}$  over time  $\delta t$ . Association and dissociation probabilities are expressed as

$$P_{on} = 1 - e^{-k_{on}\delta t}, \quad P_{off} = 1 - e^{-k_{off}\delta t}.$$

## 2.3 Magnetism

Besides the membrane elasticity and adhesion mechanism described in Sects. 2.1, 2.2, there is another phenomenon with extremely useful applications: magnetism. In biomedical applications the interaction of biological fluids with magnetic fields offers great opportunities. Magnetic properties can be used to separate certain cells, either directly by cells magnetization or by attaching magnetic beads to cells. Bioseparation is performed not only on the cell level but also on macro-molecules, antigens, DNA etc. Magnetic resonance imaging uses magnetic particles as a contrast agent for easier and much more accurate detection of disturbance regions. Destruction of tumor tissue is another promising technology. High-frequency magnetic fields heat magnetic particles that are inserted directly in the tumor and destroy nearby cells. Many other applications use and investigate the influence of magnetic fields with or without additional magnetic particles in the biological system.

Magnetism can be readily incorporated into our model. To understand the way how magnetic forces act on particles immersed in a fluid we start with description of magnetic bead structure. Magnetic beads for biomedical applications, which are available commercially, are often based on magnetic nanoparticles embedded in a polymer matrix [44]. These nanoparticles are generally assumed to have only a single domain because of their size (dipoles). Hence such polymer particles, which can be produced from nano- to several micrometers in size, have no remanent magnetization without an applied field.

The behavior of individual magnetic particles can be analyzed using molecular dynamics simulations. Classical equations of motion are numerically solved for each individual particle in the system. In our preliminary work we investigated particle dynamics using a simple Stokes drag [27]. The Object-in-fluid framework allows capturing magnetodynamic interactions in complex fluids. The simulation framework is able to cover a broad range of microfluidic applications with different flowing objects, e.g. interacting magnetic particles, blood cells and/or the combination of both. The developed simulation tool can calculate magnetic particle dynamics in homogeneous and inhomogeneous external fields incorporating also accurate flow characteristics.

### 2.3.1 Magnetic Particle Dynamics

A soft-magnetic particle experiences highly dynamic forces under the influence of a magnetic field  $B_{\text{ext}}$  and induced fields  $B_i$  of other particles (Fig. 2). A magnetic moment  $\mathbf{m}$  is created in every particle according to the magnetic susceptibility  $\chi$  and the particle volume  $V$ .

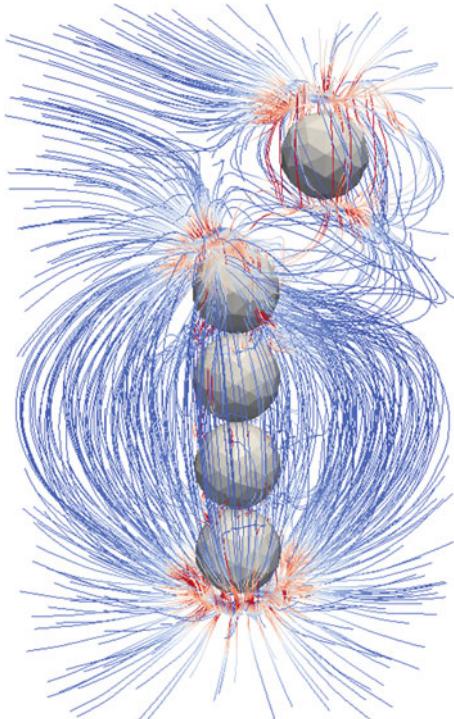
$$\mathbf{m} = V\chi \frac{\mathbf{B}}{\mu_0} = V\chi \frac{\mathbf{B}_{\text{ext}} + \mathbf{B}_i}{\mu_0} \quad (9)$$

The gradient force  $\mathbf{F}_g$  on a magnetic material (e.g. magnetic bead) is given by the negative gradient of the energy of the magnetic dipole moment  $\mathbf{m}$  in the field  $\mathbf{B}$  (Eq. 10). A magnetic dipole in an external field has the lowest energy with  $\mathbf{m}$  parallel to the field

$$\mathbf{F}_g = -\nabla(E) = \nabla(\mathbf{m} \cdot \mathbf{B}). \quad (10)$$

Solving Eq. 10 with the assumption, that a magnetic bead has only a single domain (superparamagnetic beads) with homogeneous magnetic moment  $\mathbf{m}$  ( $\frac{\partial \mathbf{m}}{\partial \mathbf{r}} = \partial_{\mathbf{r}} \mathbf{m} = 0$ ) reduces  $\mathbf{F}_g$  to

**Fig. 2** Complex field lines of interacting magnetic particles (dipoles). Chaining is preferred due to higher interaction forces at the poles. Red color of the field lines indicate higher magnetic field strengths



$$\mathbf{F}_g = \begin{pmatrix} F_x \\ F_y \\ F_z \end{pmatrix} = \begin{pmatrix} m_x \partial_x B_x + m_y \partial_x B_y + m_z \partial_x B_z \\ m_x \partial_y B_x + m_y \partial_y B_y + m_z \partial_y B_z \\ m_x \partial_z B_x + m_y \partial_z B_y + m_z \partial_z B_z \end{pmatrix}. \quad (11)$$

Other than in our preliminary work, where we have assumed to have partial homogeneity of the external field [27], a full description of magnetic particle movement means looking for solutions for all  $\partial_i B_j$  of the gradient force  $\mathbf{F}_g$ .

Cuboidal or cylindrical NdFeB permanent magnets are a possible source for a complex gradient field. Akoun and Yonnet [2] showed the analytic calculation of the magnetic field created by a cuboidal permanent magnet. Derby and Olbert [14] did the same for cylindrical permanent magnets or ideal solenoids. In case of homogeneous field sources (e.g. from a horse shoe magnet)  $\mathbf{F}_g$  reduces to the gradient field of other surrounding particles. Analytical equations to calculate forces between magnetized dipoles and how to avoid unphysical overlap of the particles are described in detail in [34, 37]. A magnetic particle interaction force  $\mathbf{F}_{\text{int}}$ , which depends on the particles radii, magnetization and the distance between them, and a corresponding excluded volume force  $\mathbf{F}_{\text{ex}}$  which cancels the magnetic interaction forces in direct contact of two particles are used.

### 2.3.2 Hydrodynamics

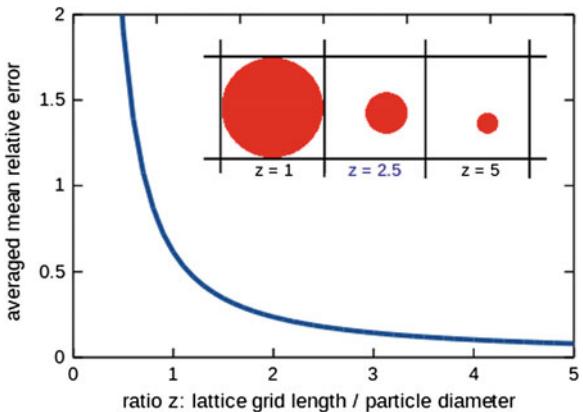
In the previous section magnetic particles are treated as point dipoles. A volume information is given by an excluded volume force  $\mathbf{F}_{\text{ex}}$  (collision detection) using the particles radii. Magnetic interactions are calculated in the center of mass of the particles. Treating fluid with a simple Stokes drag enables also the calculation of fluid interactions at the center of mass. Again, the volumetric representation is given by the particle radius  $a$  (Eq. 12). During laminar flow, the Stokes Law (Eq. 12) is used to calculate the force  $F_d$  on an object, with density  $\rho$  and dynamic viscosity  $\mu$  of the fluid and the relative velocity  $\mathbf{v}$  of the particle with respect to the fluid.

$$F_d = 6\pi\mu a \mathbf{v}. \quad (12)$$

The lattice-Boltzmann method allows a much more complex and accurate computation of fluids acting on the particles. In Sect. 2, fluid-structure interactions are described in detail. In case of magnetic particles, it is sufficient to use a single immersed boundary particle. A full boundary mesh is not necessary in general.

Mass points work well with radii much smaller than the lattice grid length of the fluid. Calibration of the correct flow behavior is done the same way as in previous work [10]. The particles with certain mass and radius get pushed and the velocity decrease is compared to an analytical expression (using the Newton's equation of motion). Using this method, the proper mass of the points can be computed. Figure 3 shows an average of different particles, their size ratio to the lattice grid length and the mean relative errors. With increase of particle size, the mean relative error gets dominant and needs to be treated with special care. The error limit is different for every application and must be set carefully.

**Fig. 3** The mean relative error of several mass-points (different mass and radii) is averaged. An increase of the error can be observed with larger particles compared to the lattice grid length



If the magnetic particles are larger than the set error limit, the volumetric representation of the particles has to change. One possible solution is the immersed boundary method. Here the stiffness constants are set in such a way, that the object's shape cannot change. At the center of mass, magnetic gradient forces are computed and then distributed to the boundary points. And again the Newton's equation of motion, as well as the lattice-Boltzmann equations are solved as described in Sect. 2. The model also allows the elasticity of the magnetized object, which in case of polymer beads, could be possible. In order to simulate attached magnetic nanoparticles on e.g. blood cells a similar approach can be used. Magnetic gradient forces are not computed in the center of mass, but at every single boundary point. The correct magnetization of the boundary points needs to be estimated from the amount of attached nanoparticles and their magnetic properties.

## 2.4 Computer Implementation

We implemented our model as an Object-in-fluid framework [9] forming an extension of a wide-spread software package ESPResSo [4]. This package enables simulation and analysis of molecular dynamics in many-particle coarse-grained bead-spring models in soft matter research in physics, chemistry and molecular biology. However, until recently, ESPResSo has not been equipped to handle the elasticity of flowing closed objects. Object-in-fluid framework enables ESPResSo to simulate the flow of elastic objects immersed in a fluid. The result of a simulation run is a complete data set of fluidic states and immersed object(s) information. The tool provides velocity and pressure profiles of the fluid given in each time step of the simulation. For immersed objects, e.g. biological cells or solid particles, Object-in-fluid delivers data on their three dimensional shapes represented by the objects' boundaries. Moreover, numerous quantities can be tracked during the simulations, such as local tension of the cells' membranes or fluid pressure on the surface of the cells.

Performing in-silico experiments using Object-in-fluid allows for generation of experimentally testable predictions and hypotheses. There are various analytical and statistical tools implemented in Object-in-fluid that let the user analyze the multicellular system and its sensitivity to parameter perturbations. More detailed description of the Object-in-fluid features are available either in official ESPResSo documentation [29] or at our research group’s webpage [7].

### 3 Other Modeling Approaches

There are other approaches to modeling elastic cells in flow. In this section we describe some of them that are qualitatively comparable to our model—i.e. we do not cover here continuous methods, such as modeling objects using finite elements or volume tracking methods, but focus on those that use spring network model for immersed objects.

#### 3.1 Elastic Objects

When modeling elastic behavior of cells by discretizing their surface, it is a natural idea to use springs for capturing the elastic behavior—not only the stretching, but also spring-like bending and area and volume conservation. However, even in this area, there are various approaches one can take besides the one described in Sect. 2.1.

One basic distinction is according to the relaxed state of the elastic object. In one approach, which we have also adopted, the relaxed state is a biconcave discoid shape (RBC shape) specified by the initial triangulation. An alternative is to use a sphere with the same surface area as a RBC as a relaxed state and energy-minimizing principles to obtain the RBC shape, e.g. in [53].

Another distinction is the actual modeling of elasticity. One can specify the forces directly, e.g. [18], or focus on elastic energies [41] and compute the forces by differentiating the elastic energies  $W$  with respect to positions  $\mathbf{r}_i$ , [53]:

$$\begin{aligned} E_s &= \frac{k_s}{2} \sum_{l=1}^{N_l} \left( \frac{\Delta L}{L_0} \right)^2, & E_b &= \frac{k_b}{2} \sum_{l=1}^{N_l} L_l \tan^2 \left( \frac{\Delta \theta}{2} \right), & E_A &= \frac{k_A}{2} \left( \frac{\Delta A}{A_0} \right)^2 A_0, \\ E_a &= \frac{k_a}{2} \sum_{e=1}^{N_e} \left( \frac{\Delta A_e}{A_e} \right)^2 A_e, & E_V &= \frac{k_V}{2} \left( \frac{\Delta V}{V_0} \right)^2 V_0 \\ W &= E_s + E_b + E_A + E_a + E_V, & \mathbf{F}_i &= -\frac{\partial W}{\partial \mathbf{r}_i} \end{aligned}$$

As evident when comparing the above expressions to forces described in Sect. 2.1, the formulas for the individual elastic forces or the corresponding in-plane energy terms also vary—some models include normalization by corresponding quantities in relaxed state (above), some consider difference in angles directly (our model), some apply trigonometric functions to the difference (above), etc.

Some of the models, e.g. [20], also take into account the thermal fluctuations and viscoelasticity of the membrane—strain rate dependence on time and molecular rearrangement when a stress is applied. These kinds of approaches sometimes also look at spectrin level modeling [13, 20] and offer large amount of details, but are limited by high computational cost. A possible solution to this problem is to reduce the number of degrees of freedom through mapping of a spectrin-based model onto coarse-grained structures, such as the spring-network models mentioned above.

On the other hand, spring network models can be coupled to other fluid solvers than lattice-Boltzmann, e.g. smoothed particle hydrodynamics method in [30]. In this case, the work was limited to 2D, but there is ambition to ultimately use the discrete inner particles to represent intracellular microstructures, and probe their evolution in response to chemical and mechanical stimuli. Other 2D spring network models have already include some detail—nucleus for circulating tumor cells or model the membrane as a bilayer, e.g. [48], however the penalty is that this models had to sacrifice the 3D dynamics.

Ultimately, the selection of the spring-network model depends on the level of detail and physical properties one wants to capture.

### 3.2 Fluid-Object Interactions

The interactions between fluid and immersed objects need to be treated carefully. The approach we took, using the drag force, described in Sect. 2, has the advantage of relatively straightforward implementation. The drawback though is that the no-slip condition, i.e. local velocity of the cell membrane being equal to the local velocity of the fluid at the same location, is not satisfied. As a consequence, our membrane is penetrable (technically, also a biological membrane is, but the exchange of material should be happening on the molecular level) and does not strictly conserve volume.

A remedy for this is to implement a bounce-back condition for the immersed boundary method. In [43], one can find criteria an IB method interpolation stencil should meet in order to produce physical results and a discussion of their mathematical significance. The main idea is to treat the moving elastic boundary similarly as one would treat a stationary rigid boundary and whenever the fictitious particles of a fluid population approach a node during propagation, they are put back to the node they came from with opposite velocity adjusted for the velocity of elastic object. The tricky part is to efficiently recompute the position of the moving boundary during this process. Also, in cases when high efficiency is needed, one might need to sacrifice the accuracy of the interpolation stencil and tolerate a certain level of numerical artifacts.

One way to reduce the computational cost in this case is to consider only one-way coupling of the fluid and the moving objects, i.e. objects receive information from the fluid but do not influence it. This is undesirable, because for example in a situation when the objects blocks a narrow opening, no fluid should pass through, but in one-way coupling model it would.

An alternative to straight tracking of particle populations in bounce-back scheme is the direct-forcing method [21]. In it a force is computed from Navier-Stokes equations, which when locally applied, equalizes the fluid velocity in the next step with current boundary point velocity, i.e. it is an explicit coupling scheme which is efficient in terms of computing time. We are planning to implement this approach to our model in the future.

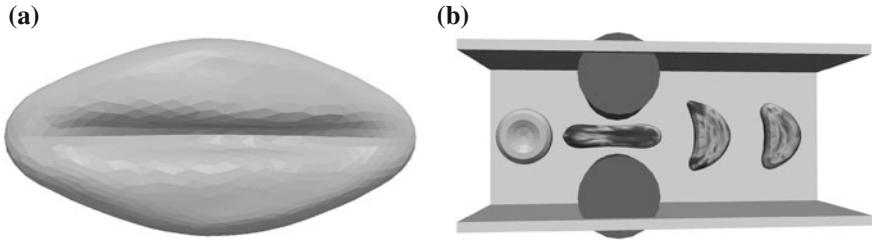
## 4 Validation of the Model

A natural next step after one selects a model, is to validate it, ideally using a comparison to biological experiments. It turns out that for some physical properties, it is not easy to obtain sufficiently detailed biological data, e.g. one can only get an overall behavior of the cell and not information on local parts of the membrane, or one has to compare with experimentally observed behavior rather than precise measurements. In this section, we cover some validation approaches as they relate to individual parts of the model.

### 4.1 Elasticity of Red Blood Cell

First, we describe several possibilities how to validate the elasticity of our cells since it is the core feature of our model. The biological experiment that is most widely used for this purpose is the optical tweezer experiment [39]. In this procedure, two silicone beads are attached to two opposite points on a red blood cell membrane and then pulled apart using optical tweezers. Cell prolongation and diameter decrease were measured as a response and simulations of the same procedure, Fig. 4a, can be used to calibrate the elastic parameters of the membrane [10].

Essentially, what one tries to accomplish using comparison like the one above, is to match a biophysical property—Young's modulus of elastic membrane—with a stiffness coefficient of springs in the spring network model. However, both in biological experiments and in simulations, it is difficult to separate the individual types of elastic behavior, e.g. stretching from bending, and thus it is useful to consider these properties also theoretically. We have investigated this relationship and found out that there is a functional dependence of the physical in-plane shear modulus of the cell membrane and the stretching coefficient. The dependence involves the mesh density and can be found in [11].



**Fig. 4** Two validation experiments: **a** A simulation of a red blood cell stretched by optical tweezers. **b** A red blood cell flowing through a channel with obstacles displayed in 4 different time steps. The grey scale corresponds to magnitude of stretching forces acting on the surface, with dark grey being the largest. Given sufficient time and no further disturbances, the cell returns to its relaxed state

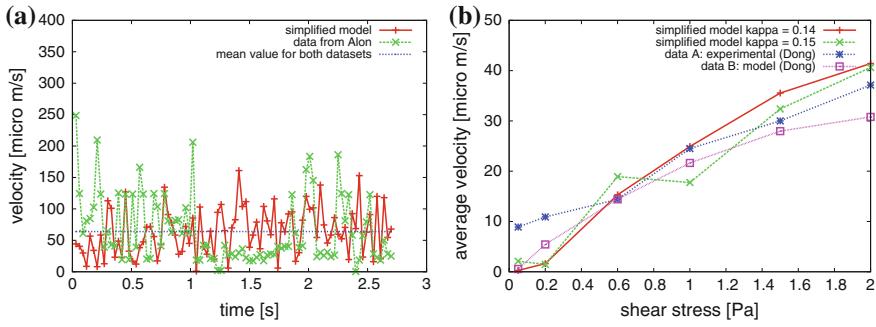
We have also found that a useful approach to monitoring how much elastic energy is “stored” in the given shape of the object is to use an energy-like characteristic that only involves the acting elastic forces [52]. This then allows us to separate the individual elastic moduli both when getting the global information about whole object—which type of elasticity is acting the least/most in a given situation. It can also be used to get local information about which parts of the surface are strained the most, Fig. 4b. This is useful when investigating the destruction thresholds of RBCs discussed in Sect. 5.

Other experiments that are frequently used for comparison of simulations and biology are parachute shapes of red blood cells in Poiseuille flow in micro vessels or micro glass capillaries [49] and tank treading motion in shear flow [5, 22]. Our model performs reasonably well in these.

## 4.2 Simplified Adhesion Model

The adhesion model introduced in Sect. 2.2 includes the dependence of  $k_{\text{off}}$  on the magnitude of the force exerted on the bond. This assumption is natural, however, the dependence of  $k_{\text{off}}$  on the bond prolongation is quite complicated, includes parameters such as bond detachment force and reactive compliance, that are quite difficult to measure. In [8] we consider a simplified model where  $k_{\text{off}}$  is a fixed constant and we illustrate that this simplified model captures important mesoscopic behavior caused by cell adhesion.

Cell rolling velocity is an important feature that must be correctly reproduced when modeling cell adhesion. This velocity is however not constant. During the rolling, the instant cell velocity varies randomly in a given range. The model should therefore capture at least the mean velocity, and possibly its deviation. In [3] the authors performed laboratory experiments and obtained the rolling velocity of the cells. Our model is capable of reconstructing these velocities. In Fig. 5a we can see



**Fig. 5** **a** Velocity reconstruction over the time period of 3 s. The mean value of experimental data from [3] is the same as the mean value for simulated velocity from our simplified model. **b** Averaged velocity of the cell over a certain time period depends on the wall shear stress. *Data A* (experimental) and *Data B* (model) are taken from [15]

the measured velocities depicted by green symbol cross and the velocities obtained from our model depicted with red symbols plus.

The second verification of the model tests whether our model can capture the dependence of cell velocity on the different fluid shear stresses. Namely, the question is whether one set of parameters gives the same values of averaged cell velocity for different fluid shear stresses.

In [15] the authors investigated the mechanics of leukocyte adhesion to endothelial cells using in vitro side-view flow assay. The authors showed how the averaged velocity of the cell depends on shear stress. We demonstrate that we are able to capture this dependence. The experiment was performed for shear stresses ranging from 0 to 2 Pa, for each stress we tracked the movement of the cell and we computed the averaged velocity over the time 1ms. The results are presented in Fig. 5b. For details about the simulation setup we refer to [8].

### 4.3 Computational Scalability

One of our long term goals, motivated by biomedical applications, for more details see Sect. 6.4, is to simulate large numbers of moving cells. Currently, the state of the art in this area is at the order of  $10^4$  of cells, but the models used for these simulations sometimes make significant simplifications, e.g. in [41] only one way coupling between cells and fluid is considered. In [55], authors show results of simulations of approximately 50000 cells at 45 % hematocrit, including aggregation of cells, use simple continuous model and take into account only the stretching and bending stiffness of the object. There have been even larger simulations done [47], but even though they deal with individual deformable cells, they model them as very simple ellipsoidal sacs and with no aggregation.

**Table 1** Simulation time (fluid computed on GPU) in minutes for computation on 1, 2, 4 and 8 cores

Cores	Cells		
		512	1024
1	38.0	163.9	754.9
2	18.8	92.1	295.8
4	9.9	34.4	166.3
8	7.1	29.2	98.5

Currently, using our Object-in-fluid framework [9], we are able to simulate few thousand cells. The code is parallelized and on our hardware (Nvidia Tesla M2090 with 64GB RAM and 2 Intel Xeon E5-2609 @ 2.40 GHz CPUs) we can run up to 8 threads. Moreover, the fluid can be computed using GPU (240 cores with 4 GB GDDR3 global memory in our case). While the initialization of fluid on GPU simulation takes longer than in the CPU simulation, once it is initialized, there is a significant speed-up during computation.

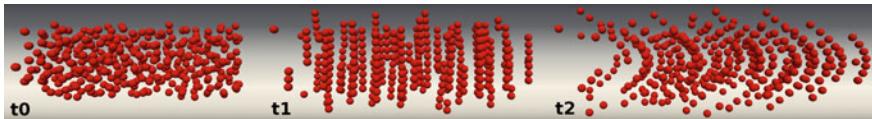
To illustrate computational scalability of the problem in terms of number of parallel computational cores, we set up a simulation of a channel with dimensions  $240 \times 120 \times 120 \mu\text{m}^3$  (3 456 000 fluid nodes). We have uniformly distributed 512, 1024 and 2000 cells (each 2 microns in diameter) in this channel, so that cell density is constant over the whole simulation box. In Table 1 we can clearly see that increasing number of computational cores decreases the computational time.

In another simulation experiment that we have performed [12], we looked at breakdown of simulation time by the individual model parts. We saw that while for single cell simulation, the fluid calculation is much more time consuming (takes about 90 % of the total simulation time on CPU, 10 % of the time on GPU), once more cells are included in the same channel under the same conditions, the computation of their elasticity (linearly dependent on number of cells), cell-fluid interactions (linear dependence) and cell-cell interactions (quadratic dependence unless treated very carefully, for more details see Sect. 6.3) consume significant amounts of time.

To conclude, currently the most efficient way to run multiple-cell simulations using our framework is to use single GPU for fluid, multiple CPUs for computations of elastic objects and template approach during initialization (identical objects are created by copying the same template).

#### 4.4 Magnetic Particle Chaining in Parabolic Flow

In 2001 Hayes et al. [28] observed soft-magnetic particles in a microfluidic channel through a microscope. With an applied magnetic field the particles self-organize to chains which are driven by the fluid. Due to the small dimensions of the channel



**Fig. 6** Magnetic particles are randomly placed in a microfluidic channel with a parabolic flow (time step  $t_0$ ). At time step  $t_1$  an external magnetic field perpendicular to the flow magnetizes the particles and chaining occurs. Turning the field off at time step  $t_2$  breaks the chains and the particles follow the local fluid velocity

and a pressure induced flow, a parabolic velocity profile is created (zero velocity at the channel boundary, maximum velocity of around 2 mm/s at the channel center). As soon as the applied field is turned off, the particles sense the local laminar flow velocity and the chains break. As the particles at the channel center are flowing faster, parabolic flow profile can be observed almost immediately. Shortly afterwards, the particles are randomly mixed again.

We were able to recreate this experiment by using the Object-in-fluid framework (Fig. 6). The simulation setup is chosen in such a way, that the magnetic particles can be represented by mass points as explained in Sect. 2.3 (diameter of particles is much smaller than the lattice grid length of the fluid).

## 5 Understanding Membrane Stresses—Insight from Simulation Model

Promising preliminary results concerning the energy-like characteristic described in Sect. 4.1 indicate the possibility of measuring the local stresses inside the membrane. Similarly, we can measure the relative areal strain of individual mesh triangles during the motion of cells inside microfluidic chambers. This way, we can predict whether given specific parameters of the flow and geometry of the channels will cause the rupture of cells.

The membrane of red blood cell can withstand a finite strain, beyond which it ruptures. The generally accepted threshold value is 4 % for the total areal strain [19]. The membrane however does not stretch uniformly. Locally, the areal strain can go beyond the threshold 4 % without rupturing the cell. To study this phenomenon, no experimental data is available. Using our model, we can set up the simulation in which a cell undergoes 4 % areal strain with nonuniform local strains. This simulation gives us better understanding of this behavior. In [51], we have designed an in-silico experiment where the total areal strain reached 4 %.

In the experiment, the cell undergoes a rotating motion in shear flow while experiencing the 4 % increase of the total surface.

There are several causes of rupture of biological cells. One such reason may be when the membrane is locally exposed to high areal strain. When the value of areal strain is high enough, the cell ruptures immediately. We denote this value  $\tau_{imm}$ .

In this case, time exposure to areal strain  $\tau_{imm}$  is not significant. Another reason for the rupture may include the possible cumulative damage of the membrane during the process. Therefore the quantity characterizing this kind of damage is the integral of the areal strain over the time and we denote it by  $\Upsilon = \int_{time} \tau(s)ds$ .

The motivation for this type of membrane degradation comes from [19, 35]. The authors claim that the exposure time of areal strain plays significant role in the cell rupture.

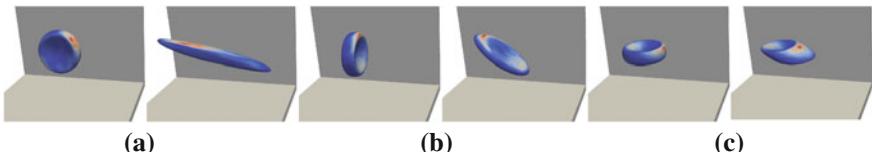
We tracked down the locations on the membrane where these quantities reached their maximal values. We suggest that the cumulative load of local areal strain may be an appropriate measure to determine the location where the cell ruptures. Our simulations show that even with the 4 % global areal strain, some local parts of the membrane can be strained significantly more. There are several triangles in the mesh that are stretched by more than 100 %. More quantitative results are available in [51].

## 5.1 Localization of the Most Stretched Parts

In Fig. 7, we can see the snapshots from the simulations. Here we want to show the shape of the cell exactly at the moment of the cell rupture, that is at the moment when the total areal strain reaches 4 %. This time instance occurs at different times for different initial rotations of cell. We present the initial position of the cell and the position at the moment of the rupture for the three different rotations.

The membrane coloring is according to values of  $\Upsilon$ . This way, we can localize the parts of the membrane with the highest stress measured by  $\Upsilon$ .

With presented computational experiments we have demonstrated that the deformation and rupture of the cell depends on the original rotation of the cell. Our model can be used to study global and local cell behavior in flow in general, not just shear flow. The change of shape and change of global area in flow can be used in further calibrations of the cell model.



**Fig. 7** Three pairs of figures (a), (b) and (c), show the cell at the beginning of the simulation and at the end of the simulation for each of the three considered rotations. The simulations were stopped when the maximum relative deviation of global area reached 4 %. The colors from blue (lowest value) to red (highest value) correspond to  $\Upsilon$  at the time of rupture

## 6 Prospective Developments

While many features of the model have already been implemented in the Object-in-fluid framework [9], which is a part of a scientific open-source software package ESPResSo [4], the module is still under active development as we are continuously adding new capabilities. In this section, we outline the preliminary work for further enhancements of the framework as they follow from our work in various research directions: interaction of neighboring membranes, cell aggregation, and seeding of large simulations. Afterwards, we present three direct applications of our simulation model, in which one can gain biological insight using computational modeling.

### 6.1 *Interactions of Neighboring Membranes*

The deformation and motion of individual cells is significantly influenced by other cells in their vicinity, especially in fluids with relatively high ratio of solid content. In blood, the solid content reaches more than 40 % of the total volume. In general, cell-cell interaction can be treated by introducing repulsive forces between two neighboring blood cells. In more detail, there is a nonlinear potential defined between each couple consisting of two points, one point being a mesh point from one cell and the second point being the mesh point from the other cell. Once a cell comes close to another cell, repulsive forces start to act between the corresponding mesh points.

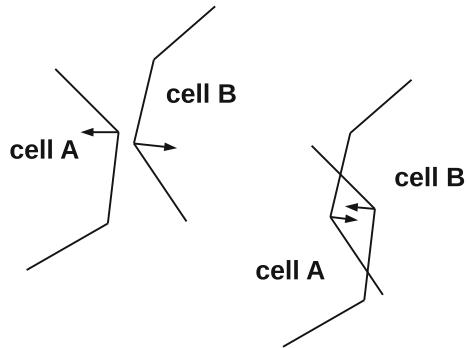
The implementation using such potential has its disadvantages. Due to the discrete time steps, one needs to be sure that the mesh points from two different cells never come too close, otherwise the meshes can overlap. Once this happens, the repulsive potential will eventually support the overlapping and the two meshes crash into each other.

To overcome this drawback, we are in the process of implementing direction-dependent repulsive forces. The force direction would not be solely dependent on the distance between the two respective mesh points (each from a different cell), but it would also depend on the outward vector of the membrane at these points. This way, the forces can be defined to always point in the direction away from collision and thus two cells will not overlap, see Fig. 8.

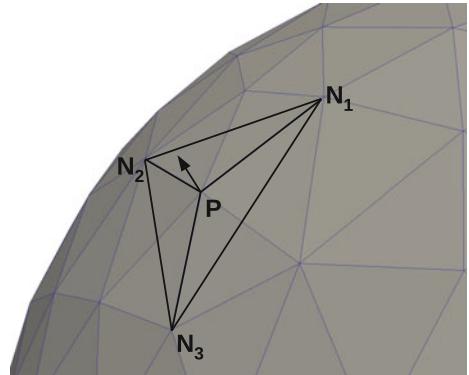
For this, we are working on a combination of two interactions—one bonded and one non-bonded. The bonded interaction is defined between each mesh point and some of its neighbors. We do not take all neighbors because nodes have variable number of neighbors and it is sufficient to choose three that are best spatially distributed. In this interaction, we then use the three neighbors to compute the outward vector at given node (see Fig. 9).

The cell-cell interaction is then defined between two points that come close enough and is inversely proportional to their distance. In case it is determined using their outward vectors that the membranes have already crossed, repulsive forces are applied to correct this situation.

**Fig. 8** Direction dependent repulsive forces ensure that the cells are always repelling, regardless the possible overlap



**Fig. 9** For each mesh point  $P$ , we look at all its neighbors and select three of them  $N_1$ ,  $N_2$ ,  $N_3$  that are most spatially distributed. From these we then compute the outward pointing unit vector



## 6.2 Cell-Free Layer and Cell Aggregation

Various hemorheologic effects are caused by confinement. *In vivo*, this corresponds to passage through small blood vessels (diameters smaller than  $30\mu\text{m}$ ), in microfluidic chambers it concerns flow through small openings. In this kind of flows, a cell-free layer is formed at the boundary caused by lift forces related to the walls and elasticity of RBCs [25]. Also the famous Fåhraeus [23] and Fåhraeus and Lindqvist [24] effects are tightly related to the existence of the cell-free layer: Due to the lateral migration of the RBCs towards the centerline, the cells move faster than the average flow. The mass balance then requires that the hematocrit in the center is larger than on the sides and in small tubes the flow rate strongly depends on the tube diameter.

Another peculiarity of RBCs is that they have been observed to form aggregates at small shear rates (below a few  $\text{s}^{-1}$ ). These aggregates, can be either one-dimensional, called rouleaux or 3D clusters [45] and are the major determinant of the shear thinning property of blood. The protein fibrinogen seems to be responsible for aggregation as it can bind to RBCs and connect them, however aggregation is a reversible process and these connections can be broken if shear conditions change.

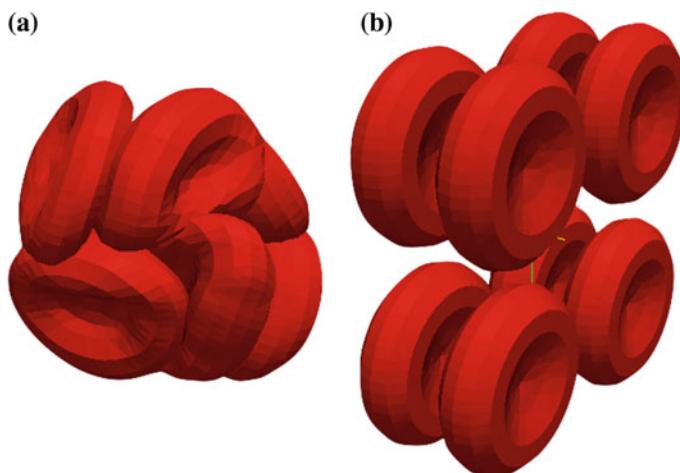
To capture these features in a model requires setting up proper potentials for cell-cell interactions. In case of aggregates, the interaction potential needs to exhibit “capturing” behavior strong enough that forces, e.g. fluid flow, acting on one cell of the aggregate will not easily separate it from the cluster, but rather be distributed and influence all connected cells. The cell-free layer also requires calibration of the dense flows. Both of these should be possible in the future using the Object-in-fluid module.

### 6.3 Large Simulations Seeding

Compared to simulations with just few cells, for dense simulations with many cells it is very important to pay attention to initial phase. It is highly impractical to deterministically deform cells and place them into the simulation domain so that they form relatively packed suspension (e.g. Fig. 10a).

One alternative is to place regularly spaced cells in relaxed shape into the channel (e.g. Fig. 10b) and let the flow evolve until they assume the typical irregular packed distribution. A disadvantage of this approach is that it takes a very long time—typically even longer than the actual simulation time once the initial distribution is known.

Another method, proposed in [36], is to “grow” the cells. It starts by specifying random location and orientation for each cell. The number of cells is determined from required final hematocrit, i.e. if the total volume of the simulation domain is 200000 fl and required hematocrit is 30 %, we know that 60000fl should be occupied by red blood cells. Taking the the volume of healthy RBC to be  $V_{RBC} = 100fl$ , we then know that the simulation should contain 600 cells.



**Fig. 10** Illustration of seeding red blood cells for simulation: **a** densely packed, **b** regularly placed (color in online)

However, these are not put in the simulation box at their full volume. They start at about 30 % of their final volume and slowly (at the order of  $10^{-3} V_{final}/dt$ ) grow to their final size. The big advantage is that when the membrane interaction forces (e.g. as described in Sect. 6.1) are turned on, the cells naturally deform in this process and thus a reasonably accurate dense suspension is created.

Open question is whether the seeds in this process need to be biconcave discoids, as the RBCs are, and slowly grow both their surface and volume. Alternatively, they could be regularly packed spheres whose volume is equal to (or slightly less than) the desired RBC volume and they would mainly increase the surface area. This notion is motivated by the biological process of erythrocyte maturation. Upon leaving the bone marrow, the young RBCs, at this stage called reticulocytes, have spherical shape of roughly the same volume as regular RBCs and only as they mature (in 24–36 h), they increase their membrane area and gradually obtain the biconcave discoid shape [26].

Another thing to consider from computational point of view are the cell-cell interactions. In principle, any pair of cells can interact via a repulsive potential described in more detail in Sect. 6.1, however, for a large number of cells (even for hundreds or thousands) this becomes unfeasible, since there are (for 1000 cells) half a million pairs of cells in each of which the mesh nodes interact point wise (e.g. 1000 mesh nodes on one cell with 1000 mesh nodes on another cell). To avoid unnecessary evaluation of (zero) interactions for cells that are too far from each other, we have implemented a neighbor list for each cell which is updated periodically during the simulation, so that it includes all the cells in vicinity of the given cell and the interactions are only evaluated for these.

## ***6.4 Optimization of Microfluidic Devices***

Advances in the previously described sections will allow us to tackle our long-term goals, since the overall direction of this research concerns development of microfluidic devices. On one side, computational model can help with the optimization of existing devices. Currently used devices are mostly designed using experimental techniques, or, in some cases using simplified models involving fluid and structure modeling. Complex three dimensional computational model described in previous sections allows us to simulate the device and to test numerous different geometries. Thus, simulations open the possibility to pick the one with the best performance and to improve the existing design.

On the other side, computational model is extremely useful in the initial design of devices. When designing experimental setup, the biologists are interested in numerous problems. They need to have a good estimation on the range of parameters for the experiment. They would like to know which parameters influence the result of the experiment the most, so that they can be more careful when adjusting these parameters. The simulations can give the answers to all of these questions without actually performing the experiments.

In the following, we describe three areas where we see direct application of our computational model. In all three cases, simulations cover at least a part of the microfluidic device. The simulation thus gives us a perfect view of the performance of the whole device, on its efficacy. It also determines the sensitivity to the initial parameters.

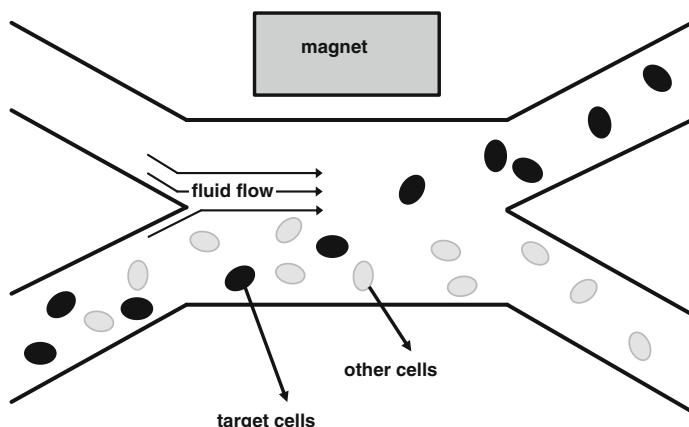
#### 6.4.1 Magnetic Cell Sorting

Cell separation is capable of isolating cells of interest from a complex background. The general principle of microfluidic devices for continuous particle flow separation is that a force applied across a microscale channel displaces particles as they move through the channel, with the magnitude of the displacement corresponding to some property of the particle. In magnetic cell sorting, it is the magnetic force.

Magnetic separation is achieved by selectively attaching magnetic beads to cells or cell populations, then altering their path using micro-ferromagnetic objects. Alternatively, the native magnetic susceptibility of cells can replace the role of magnetic beads. While most biological matter is composed of water, a weakly diamagnetic substance, certain cells contain paramagnetic material. For example, magnetic changes in red blood cells have been used to separate diseased cells and cells with congenital abnormalities from normal cells [42]. Separation of red blood cells from whole blood by native magnetic susceptibility was first demonstrated in 1975 by Melville [38].

Separation by selective attachment of magnetic beads was demonstrated in 1977 by Molday [40]. This technique is primarily useful in selecting certain populations of cells that can be differentiated by the chemicals on their surfaces [31].

The principle of magnetic cell separation is described in Fig. 11. In case of magnetic bead separation a preprocessing step must be done in order to attach magnetic



**Fig. 11** Schematics of magnetic cell separation

beads to target cells. In case of native magnetic susceptibility, this step can be skipped. There is a mixture of target cells and the other cells on the inflow. Permanent magnets are placed at certain locations and they attract target cells so their trajectory is directed into the expected outflow channel.

#### 6.4.2 Size Based Cell Sorting

Size-based cell separation, is achieved using a method called deterministic lateral displacement, in which particles of different sizes follow one of two paths, through a micro-fabricated array of posts. This method can be used as a sort of continuously operable filter, separating large particles from small particles, or it can be used to measure the size of objects.

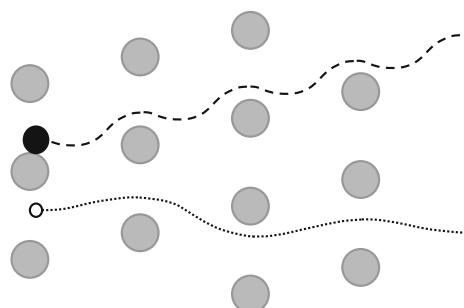
The mechanism of separation is physical displacement by obstacles arrayed in the flow. Figure 12 depicts typical particle behavior. The posts are offset perpendicularly to the main fluid flow so that the resulting array is tilted. This structure makes the fluid asymmetrically bifurcate around the posts. The whole flow is thus divided into horizontal “zigzag” streams.

Particles in this flow behave in two ways depending on their size relative to the stream next to the post. Particles with small radii, will not be displaced by the micro-post and will remain bound within one stream as they follow a “zigzag” trajectory that weaves through the posts in the array. Particles whose radii are larger than the stream width, denoted by the dark gray particle with dotted trajectory in Fig. 12, cannot fit within the stream next to the post as they travel through a gap and will be “bumped” into the adjacent stream. Such particle will travel along the array tilt angle instead of with the fluid and other small particles.

#### 6.4.3 Rare Cell Trapping

Rare cell capture devices face the challenge of high efficacy due to the very low starting concentrations of such cells. Many microfluidic devices take advantage of the 3D structure of channels to increase the surface area available to be coated with

**Fig. 12** Schematics of size-based cell separation



the antibody. Du et al. [16] demonstrated the efficacy of this technique in straight microchannels by differentially capturing human mammary epithelial cells and breast cancer cells by use of epithelial membrane antigen and epithelial growth factor receptor. They achieved 30 % capture rate.

The same technique in S-shaped microchannels showed capture rate of 50–83 % [56]. Recent work by Wang et al. on silicon nanopillars indicated that the topology of the microdevice itself may contribute greatly to the efficiency of rare cell capture [54]. In fact, major areas for improvement in rare cell capture were indicated to be increasing viability and capture efficiency/purity of directly processed biological samples, as a majority of current studies only process spiked cell lines or pre-diluted/lysed samples [46].

Here, computer modeling can play a significant role. With our complex model we can simulate most important physical and biological phenomena during the process of cell isolation or cell trapping. We can process different sizes of microposts, different shapes of microchannels and by this we elucidate the dependence of efficacy on the inner geometry of cell trapping devices.

## 7 Conclusion

Understanding the rheology of blood has been a scientific challenge for almost a century. It is vital for life and our knowledge of it can significantly impact diagnostics and treatment. Recent advances in modeling blood flow, including our work, bring us closer to this ultimate goal, however there are still many open questions. The good news is that today's computational capabilities bring us to the point when using good models we can simulate blood on the level of individual elastic cells and thus tackle these questions and obtain the needed insight.

**Acknowledgments** This work was supported by the Slovak Research and Development Agency under the contract No. APVV-0441-11. The work of Ivan Cimrák was also supported by the Marie-Curie grant No. PCIG10-GA-2011-303580.

## References

1. Ahlrichs, P., Dunweg, B.: Lattice-Boltzmann simulation of polymer-solvent systems. *Int. J. Mod. Phys. C* **8**, 1429–1438 (1998)
2. Akoun, G., Yonnet, J.: 3D analytical calculation of the forces exerted between two cuboidal magnets. *IEEE Trans. Magn.* **20**(5), 1962–1964 (1984). doi:[10.1109/TMAG.1984.1063554](https://doi.org/10.1109/TMAG.1984.1063554)
3. Alon, R., Chen, S., Puri, K., Finger, E., Springer, T.: The kinetics of L-selectin tethers and the mechanics of selectin-mediated rolling. *J. Cell Biol.* **138**(5), 1169–1180 (1997)
4. Arnold, A., Lenz, O., Kesselheim, S., Weeber, R., Fahrenberger, F., Roehm, D., Košovan, P., Holm, C.: ESPResSo 3.1—molecular dynamics software for coarse-grained models. In: Griebel, M., Schweitzer, M. (eds.) *Meshfree Methods for Partial Differential Equations VI*.

- Lecture Notes in Computational Science and Engineering, vol. 89, pp. 1–23. Springer, New York (2013)
5. Basu, H., Dharmadhikari, A.K., Dharmadhikari, J.A., Sharma, S., Mathur, D.: Tank treading of optically trapped red blood cells in shear flow. *Biophys. J.* **101**, 1604–1612 (2011)
  6. Chen, S., Doolen, G.: Lattice-Boltzmann method for fluid flows. *Annu. Rev. Fluid Mech.* **30**, 329–364 (1998). doi:[10.1146/annurev.fluid.30.1.329](https://doi.org/10.1146/annurev.fluid.30.1.329)
  7. Cimrák, I.: Cell-in-fluid research group (2015). <http://cell-in-fluid.fri.uniza.sk>
  8. Cimrák, I.: A simplified model for dynamics of cell rolling and cell-surface adhesion. In: AIP Conference Proceedings, 1648, 210005 (2015). doi:[10.1063/1.4912490](https://doi.org/10.1063/1.4912490)
  9. Cimrák, I., Gusenbauer, M., Jančigová, I.: An ESPResSo implementation of elastic objects immersed in a fluid. *Comput. Phys. Commun.* **185**(3), 900–907 (2014). doi:[10.1016/j.cpc.2013.12.013](https://doi.org/10.1016/j.cpc.2013.12.013)
  10. Cimrák, I., Gusenbauer, M., Schrefl, T.: Modelling and simulation of processes in microfluidic devices for biomedical applications. *Comput. Math. Appl.* **64**(3), 278–288 (2012)
  11. Cimrák, I., Jančigová, I., Bachratá, K., Bachratý, H.: On elasticity of spring network models used in blood flow simulations in ESPResSo. In: Bisschoff, M., Oñate, E., Owen, D., Ramm, E., Wriggers, P. (eds.) III International Conference on Particle-based Methods—Fundamentals and Applications Particles 2013, pp. 133–144 (2013)
  12. Cimrák, I., Tóthová, R., Jančigová, I.: Recent advances in mesh-based modeling of individual cells in biological fluids. In: 2014 10th International Conference on Digital Technologies (DT), pp. 25–31 (2014)
  13. Dao, M., Li, J., Suresh, S.: Molecularly based analysis of deformation of spectrin network and human erythrocyte. *Mater. Sci. Eng. C* **26**, 1232–1244 (2006)
  14. Derby, N., Olbert, S.: Cylindrical magnets and ideal solenoids. *Am. J. Phys.* **78**(3), 229–235 (2010). doi:[10.1119/1.3256157](https://doi.org/10.1119/1.3256157)
  15. Dong, C., Lei, X.X.: Biomechanics of cell rolling: shear flow, cell-surface adhesion, and cell deformability. *J. Biomech.* **33**(1), 35–43 (2000)
  16. Du, F., Baune, M., Thoming, J.: Insulator-based dielectrophoresis in viscous media—simulation of particle and droplet velocity. *J. Biomech.* **65**, 452–458 (2007)
  17. Dunweg, B., Ladd, A.J.C.: Lattice-Boltzmann simulations of soft matter systems. *Adv. Polym. Sci.* **221**, 89–166 (2009)
  18. Dupin, M., Halliday, I., Care, C., Alboul, L.: Modeling the flow of dense suspensions of deformable particles in three dimensions. *Phys. Rev. E Stat. Nonlinear Soft Matter Phys.* **75**, 066707 (2007)
  19. Evans, E., Waugh, R., Melnik, L.: Elastic area compressibility modulus of red cell membrane. *Biophys. J.* **16**(6), 585–595 (1976)
  20. Fedosov, D.: Multiscale modeling of blood flow and soft matter. Ph.D. thesis, Brown University (2010)
  21. Feng, Z.G., Michaelides, E.E.: Proteus: a direct forcing method in the simulations of particulate flows. *J. Comput. Phys.* **202**, 20–51 (2005)
  22. Fischer, T.M.: Shape memory of human red blood cells. *Biophys J.* **86**(5), 3304–3313 (2004). doi:[10.1016/S0006-3495\(04\)74378-7](https://doi.org/10.1016/S0006-3495(04)74378-7)
  23. Fåhræus, R.: The suspension stability of the blood. *Physiol. Rev.* **9**, 241–274 (1929)
  24. Fåhræus, R., Lindqvist, T.: The viscosity of blood in narrow capillary tubes. *Am. J. Physiol.* **96**, 562–568 (1931)
  25. Goldsmith, H.L.: Red cell motions and wall interactions in tube flow. *Fed. Proc.* **30**, 1578–1590 (1971)
  26. Gronowicz, G., Swift, H., Steck, T.L.: Maturation of the reticulocyte in vitro. *J. Cell Sci.* **71**, 177–197 (1984)
  27. Gusenbauer, M., Kovacs, A., Reichel, F., Exl, L., Bance, S., Ozelt, H., Schrefl, T.: Self-organizing magnetic beads for biomedical applications. *J. Magn. Magn. Mater.* **324**(6), 977–982 (2012). doi:[10.1016/j.jmmm.2011.09.034](https://doi.org/10.1016/j.jmmm.2011.09.034)
  28. Hayes, M.A., Polson, N.A., Garcia, A.A.: Active control of dynamic supraparticle structures in microchannels. *Langmuir* **17**(9), 2866–2871 (2001). doi:[10.1021/la001655g](https://doi.org/10.1021/la001655g)

29. Holm, C., Arnold, A., Lenz, O., Kesselheim, S.: ESPResSo Documentation (2014). <http://espressomd.org/wordpress/documentation>
30. Hosseini, S.M., Feng, J.: A particle-based model for the transport of erythrocytes in capillaries. *Chem. Eng. Sci.* **64**, 4488–4497 (2009)
31. Inglis, D.W., Riehn, R., Austin, R.H., Sturm, J.C.: Continuous microfluidic immunomagnetic cell separation. *Appl. Phys. Lett.* **85**, 5093–5095 (2004)
32. Jančigová, I., Cimrák, I.: Non-uniform force allocation for area preservation in spring network models. *Int. J. Numer. Meth. Biomed. Engng.* In print
33. Jančigová, I., Tóthová, R.: Scalability of forces in mesh-based models of elastic objects. In: ELEKTRO 2014: 10th International Conference, pp. 562–566. IEEE (2014)
34. Krishnamurthy, S., Yadav, A., Phelan, P.E., Calhoun, R., Vuppu, A.K., Garcia, A.A., Hayes, M.A.: Dynamics of rotating paramagnetic particle chains simulated by particle dynamics, Stokesian dynamics and lattice-Boltzmann methods. *Microfluid. Nanofluid.* **5**(1), 33–41 (2007). doi:[10.1007/s10404-007-0214-z](https://doi.org/10.1007/s10404-007-0214-z)
35. Lee, S., Ahn, K., Lee, S., Sun, K., Goedhart, P., Hardeman, M.: Shear induced damage of red blood cells monitored by the decrease of their deformability. *Korea-Aust. Rheol. J.* **16**(3), 141–146 (2004)
36. MacMeccan, R.M.: Mechanistic effects of erythrocytes on platelet deposition in coronary thrombosis. Ph.D. thesis, Georgia Institute of Technology (2007)
37. Melle, S., Calderón, O.G., Rubio, M.A., Fuller, G.G.: Microstructure evolution in magnetorheological suspensions governed by mason number. *Phys. Rev. E* **68**(4), 041503 (2003)
38. Melville, D., Paul, F., Roath, S.: Direct magnetic separation of red cells from whole blood. *Nature* **255**, 706 (1975)
39. Mills, J.P., Qie, L., Dao, M., Lim, C.T., Suresh, S.: Nonlinear elastic and viscoelastic deformation of the human red blood cell with optical tweezers. *Mol. Cell. Biomech.* **1**(3), 169–180 (2004)
40. Molday, R.S., Yen, S.P.S., Rembaum, A.: Applications of magnetic microspheres in labeling and separation of cells. *Nature* **268**, 437–438 (1977)
41. Nakamura, M., Bessho, S., Wada, S.: Spring network based model of a red blood cell for simulating mesoscopic blood flow. *Int. J. Numer. Method Biomed. Eng.* **29**(1), 114–128 (2013). doi:[10.1002/cnm.2501](https://doi.org/10.1002/cnm.2501)
42. Paul, F., Melville, D., Roath, S., Warhurst, D.C.: A bench top separator for malarial parasite concentration. *IEEE Trans. Magn.* MAG-17, 2822–2824 (1981)
43. Peskin, C.S.: The immersed boundary method. *Acta Numerica* **11**, 479–517 (2002)
44. Philippova, O., Barabanova, A., Molchanov, V., Khokhlov, A.: Magnetic polymer beads: recent trends and developments in synthetic design and applications. *Eur. Polym. J.* **47**(4), 542–559 (2011). doi:[10.1016/j.eurpolymj.2010.11.006](https://doi.org/10.1016/j.eurpolymj.2010.11.006)
45. Popel, A.S., Johnson, P.C.: Microcirculation and hemorheology. *Annu. Rev. Fluid Mech.* **37**, 43–69 (2005)
46. Pratt, E., Huang, C., Hawkins, B., Gleghorn, J., Kirby, B.J.: Rare cell capture in microfluidic devices. *Chem. Eng. Sci.* **66**(7), 1508–1522 (2011)
47. Rahimian, A., Lashuk, I., Veerapaneni, S., Chandramowlishwaran, A., Malhotra, D., Moon, L., Sampath, R., Shringarpure, A., Vetter, J., Vuduc, R., Zorin, D., Biros, G.: Petascale direct numerical simulation of blood flow on 200k cores and heterogeneous architectures. In: Proceedings of the 2010 ACM/IEEE International Conference for High Performance Computing, Networking, Storage and Analysis, SC '10, pp. 1–11. IEEE Computer Society, Washington, DC, USA (2010). doi:[10.1109/SC.2010.42](https://doi.org/10.1109/SC.2010.42)
48. Rejniak, K.A.: Investigating dynamical deformations of tumor cells in circulation: predictions from a theoretical model. *Front. Oncol.* **2**, 111 (2012)
49. Skalak, R., Branemark, P.: Deformation of red blood cells in capillaries. *Science* **164**, 717–719 (1969)
50. Taylor, C.A., Hughes, T.J., Zarins, C.K.: Finite element modeling of blood flow in arteries. *Comput. Methods Appl. Mech. Eng.* **158**(1–2), 155–196 (1998). doi:[10.1016/S0045-7825\(98\)80008-X](https://doi.org/10.1016/S0045-7825(98)80008-X)

51. Tóthová, R., Cimrák, I.: Local stress analysis of red blood cells in shear flow. In: AIP Conference Proceedings, 1648, 210003 (2015). doi:[10.1063/1.4912488](https://doi.org/10.1063/1.4912488)
52. Tóthová, R., Jančigová, I., Cimrák, I.: Energy contributions of different elastic moduli in mesh-based modeling of deformable object. In: ELEKTRO 2014: 10th International Conference, pp. 634–638. IEEE (2014)
53. Tsubota, K., Wada, S.: Elastic force of red blood cell membrane during tank-treading motion: consideration of the membrane's natural state. *Int. J. Mech. Sci.* **52**, 356–364 (2010)
54. Wang, L., Lu, J., Marchenko, S., Monuki, E., Flanagan, L., Lee, A.: Dual frequency dielectrophoresis with interdigitated sidewall electrodes for microfluidic flow-through separation of beads and cells. *Electrophoresis* **30**, 782–791 (2009)
55. Xu, D., Kaliviotis, E., Munjiza, A., Avital, E., Ji, C., Williams, J.: Large scale simulation of red blood cell aggregation in shear flows. *J. Biomech.* **46**(11), 1810–1817 (2013). doi:[10.1016/j.jbiomech.2013.05.010](https://doi.org/10.1016/j.jbiomech.2013.05.010)
56. Xu, Y., Philips, J., Yan, J., Li, Q., Fan, Z., Tan, W.: Aptamer-based microfluidic device for enrichment, sorting, and detection of multiple cancer cells. *Analytical Chemistry* **81**, 7436–7442 (2009)

# Usage of Industrial Computed Tomography for Evaluation of Custom-Made Implants

Radovan Hudák, Jozef Živčák, Teodor Tóth, Jaroslav Majerník  
and Martin Lisý

**Abstract** Development of additive technologies and biocompatible materials facilitated their use in the custom-made implants manufacture. Verification of custom-made implants manufactured using the additive manufacturing technologies is the key task to be fulfilled prior to the clinical application of an implant. It consists of parameters verification within individual steps, from a software design, through manufacturing, surface finishing, up to finalization of a medical product. The article presents possible uses of a 3D printing and the computed tomography (Metrotom 1500, Carl Zeiss, Germany) for the verification of selected parameters of customized implants manufactured using the Direct Metal Laser Sintering (DMLS) technology with the EOSINT M280 equipment (EOS GmbH, Germany) from the biocompatible titanium alloy Ti-6Al-4V (Grade 5). The article describes the possibilities of the computed tomography use in the verification of implant shapes and external dimensions, as well as internal structure. The internal structure means the implant porosity assessment.

---

R. Hudák (✉) · J. Živčák · T. Tóth

Faculty of Mechanical Engineering, Department of Biomedical Engineering and Measurement,  
Technical University of Košice, Letná 9, 042 00 Košice, Slovakia  
e-mail: radovan.hudak@tuke.sk

J. Živčák

e-mail: jozef.zivcak@tuke.sk

T. Tóth

e-mail: teodor.toth@tuke.sk

J. Majerník

Faculty of Medicine, Department of Medical Informatics, University in Košice, Trieda SNP 1, 040  
11 Košice, Slovakia  
e-mail: jaroslav.majernik@upjs.sk

M. Lisý

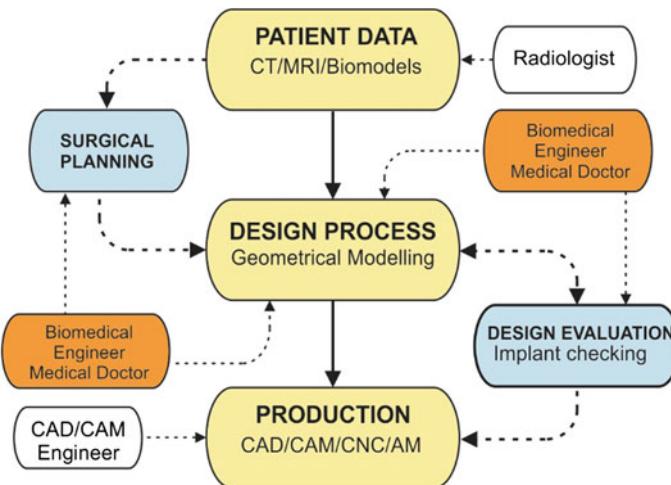
CEIT Biomedical Engineering, Tolstého 3/6, 040 01 Košice, Slovakia  
e-mail: martin.lisy@ceit-ke.sk

## 1 Introduction

The development of modern imaging methods, computer-aided processing of three-dimensional (3D) data (CAD/CAM), and new manufacturing technologies (3D printing or additive manufacturing—AM) facilitated the massive growth in the production of implants and replacements for human body parts which copy anatomical structures relatively accurately [1].

Diagnostic methods most frequently used for the collection of medical data serving for the custom-made implants modelling include the computed tomography (CT) and the magnetic resonance imaging (MRI). Quality of the resulting model depends on the accuracy of diagnostic equipment and the data resolution. Reduction of the scan spacing (slice index) using the CT enables production of more slices along the diagnosed section, which increases the resolution of the obtained data. The longer the scanning period, the higher the resolution of the image; on the other hand, however, it is necessary to consider a patients exposure to the radiation, cost-depending scanning time, as well as a patients comfort [1].

An important part of the custom-made implant designing and manufacture process is the process of design evaluation (Fig. 1) consisting of the verification and validation processes. The article describes possible uses of modern technologies, such as the computed tomography, for the purpose of verification of the processes and validation of custom-made implants manufactured using the DMLS additive technology with the EOSINT M280 equipment (EOS GmbH, Germany). The case study will be presented using a custom-made cranial implant [2].



**Fig. 1** Flowchart of design and manufacturing of custom medical implant [2]

## 2 Design and Manufacturing of Custom Made Implant

Custom-made cranial implant was manufactured using the DMLS technology, from the Ti-6Al-4V titanium alloy (Grade 5). The manufacture as such was preceded by the collection of clinical CT data of a particular patient in the DICOM format which were used to create a point cloud and subsequently a CAD model [1].

Successful implant manufacture requires the elaboration of an optimal methodology of procedures and operations so that they are reproducible for the highest possible number of specific cases. This methodology includes the proposals of individual solutions, whereas the entire sequence is customised to meet the requirements imposed with regard to the final product. The comprehensive methodology proposal should take into consideration that this issue is not purely within the responsibility of a technician; it depends, above all, on the cooperation between the technician and a physician/surgeon [1].

The overall process of the custom-made implant manufacture consists of the following steps:

1. generation of a damaged section 3D model—using the CT, MRI,
2. computer-assisted implant design,
3. implant design verification,
4. implant manufacture using appropriate technology,
5. manufactured implant verification.

Computed tomography enters the designing and manufacture process as the technology facilitating the generation of input data (DICOM data) which are used as the basis for the implant designing and subsequently at continuous and final inspections of the implant [1].

In the initial modelling stage, following the collection of a patients DICOM data, the input data in the format are imported and converted into the working environment of the Mimics software (Materialise, Belgium). Prior to work with CT images in the format, it is necessary to determine proper orientation of the given images with regard to the human body planes. Afterwards, soft tissues are removed and a skull model is created (Fig. 2) [1].

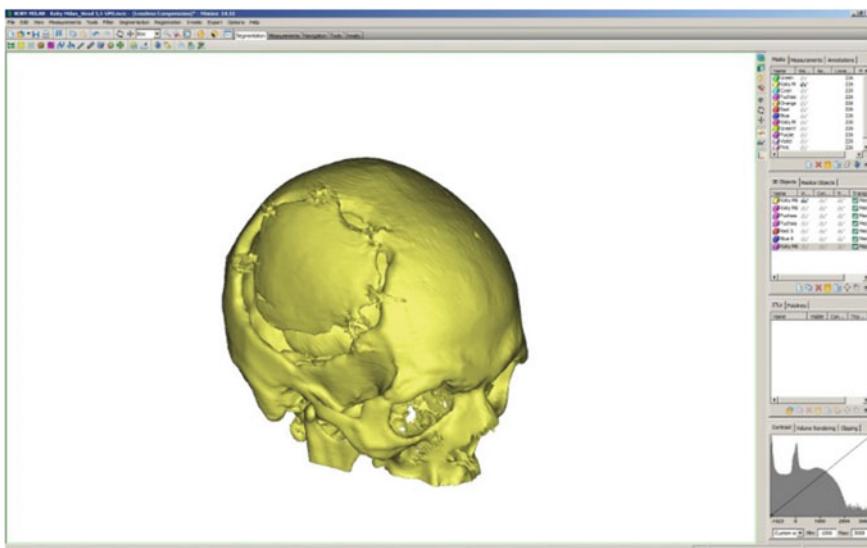
Figure 3 shows a referential 3D model of a particular patients skull with a detailed image of the skull defect.

Subsequently, the cranial implant is created by gradual software-aided removing or remodeling of the 3D referential model, as specified by a doctors requirements and on the basis of a patients needs [1].

Cranial implant creation uses the so-called vertical skeletal planes that serve as the basis for creating a mirror image of the healthy skull section at the defect locations, or of a missing section.

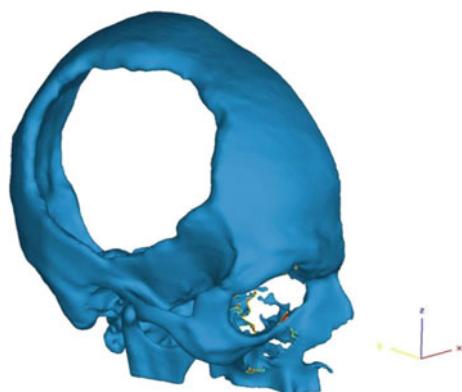
The cranial implant design visualization based on vertical and horizontal skeletal planes with regard to the defect margins is presented in Fig. 4 [1].

The following step consisted of suggesting the positioning of the fixation mechanism, considering the sutures of the separated skull section to the skull (Fig. 5) [1].



**Fig. 2** Skull model after soft tissue removal [1]

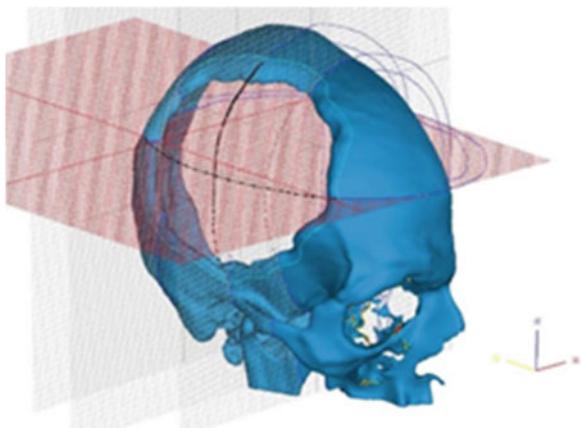
**Fig. 3** Referential CAD skull [1]



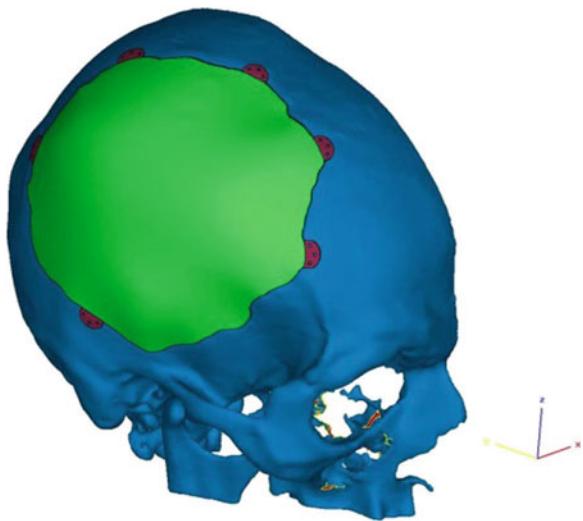
On the above mentioned fixation mechanisms, the lengths of screws protruding into both cortical bone components were defined in accurately determined positions with regard to the skull [1].

The CAD model finalisation and subsequent CAM modelling were followed by the custom-made cranial implant manufacturing using the DMLS technology [1] (Fig. 6).

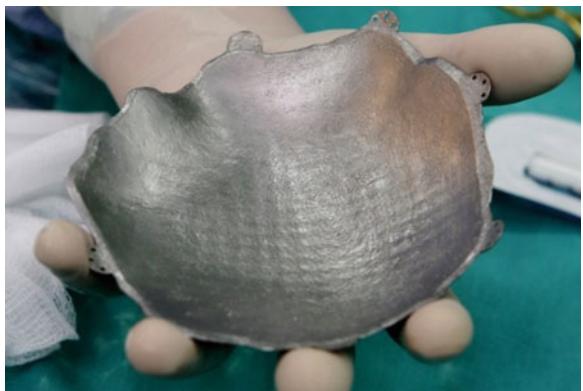
**Fig. 4** Vertical and horizontal skeletal planes for implant modeling [1]



**Fig. 5** Suggested fixation mechanisms [1]



**Fig. 6** Cranial implant made of Ti-6Al-4V titanium alloy (Grade 5) [1]



### 3 Verification of the Custom—Made Cranial Implant

Unlike conventional engineering applications where majority of objects or their parts can be described using elementary 2D and 3D elements (plane, circle, roller, ...), individual implants are mostly formed of curves. Therefore, the preparation of any drawings as well as the verification of the manufactured implant by conventional measuring technologies (CMM, conventional gauges, ...) is difficult to perform. Currently, the verification of a manufactured implant can be carried out using 3D scanners and the industrial computed tomography. The use of 3D scanners is appropriate for the shape inspection, for example the shape of an implants outer surface. In case of the comprehensive inspection of the entire implant, it is appropriate to use the computed tomography that provides the data regarding the surface, as well as the content, and is less time-consuming [2, 3].

Computed tomography can be used in the three following areas:

- shape deviations inspection,
- dimensional analysis,
- thickness measurement,
- defectoscopy.

Some autors use computed tomography for evaluation of porosity in a laser sintered artifacts or implants.

Lonard et al. in study “Assessment by X-ray CT of the effects of geometry and build direction on defects in titanium ALM parts” demonstrated that X-ray computed tomography is a powerful tool for fully characterizing, in 3D, the typical defects seen in titanium ALM components. Not only can the whole specimen be examined, but the exact size, shape, maximum dimension and location of the pores can be obtained whilst it is impossible from 2D metallographic sections [4].

Slotwinski and Garboczi in “Porosity of Additive Manufacturing Parts for Process Monitoring” describe the usage of ultrasonic porosity sensor for 0,2 % change in porosity for CoCr alloy and compare it with CT [5–7].

Girardin et al. in “Characterization of Porosity in a Laser Sintered MMCp Using X-Ray Synchrotron Phase Contrast Microtomography” use for characterization of porosity in a laser sintered metal matrix composite x-ray synchrotron [8].

For shape validation only the autors use various devices like coordinate measuring machines and 3D scanners. The CMM machines measure only in few defined points and we dont obtain the whole geometry of measured object. With 3D scanners we obtain the whole surface geometry but with worster accuracy. The accuracy of scanners is given in tenths of a millimeter and the accuracy of CT is in microns.

For deviation analysis Drstvensek et al. in “Applications of Rapid Prototyping in Cranio-Maxillofacial Surgery Procedures” use for cranio-maxillofacial implat shape validation the GOM ATOS II optical scanner [9].

Salmi use in “Medical applications of additive manufacturing in surgery and dental care” Carl Zeiss C700 coordinate measuring machine for validation of plastic 3D model and the occlusal splints are verify by 3D scanners [10].

Podshivalov et al. in “Design, Analysis and Additive Manufacturing of Porous Structures for Biocompatible Micro-Scale Scaffolds” use microCT for verification of micro-scale bone scaffold printed by additive manufacturing [11].

Bauza et al. in “Study of accuracy of parts produced using additive manufacturing” use tomograph Zeiss Metrotom 1500 and coordinate measuring machine Zeiss Contura G2 for verification of two artifacts built from stainless steel on an EOS M270 machine [12].

Matilainen in his work “Benchmarking of laser additive manufacturing process—bachelor thesis” describe the specimens created for verification of different additive manufacturing processes. For verification of LAM machine he design own artifact from EOS PH1 material (stainless steel) and provide dimensional verification of artifact [13].

### ***3.1 Shape Deviation Inspection***

As it is often difficult to assess particular dimensions of custom-made implants, an alternative method used is the comparison of the manufactured implant and the CAD model. Unlike particular values obtained by measurements, the outcome of such comparison is the map of deviations documenting the differences between the manufactured implant and the model. The advantage is the provision of the spatial distribution of deviations, not only the data representing the selected locations. The comparison of the scanned component and the CAD model requires their alignment. They are most frequently aligned applying the “Best Fit” method which uses the least squares principle, i.e. the deviations between the scan and the model are mathematically segmented. This method is not appropriate in cases when the obtained scan and the 3D model significantly differ at some locations, as the calculation might, in an effort to minimize the deviations, shift the alignment and thus the result must not necessarily correspond to the reality.

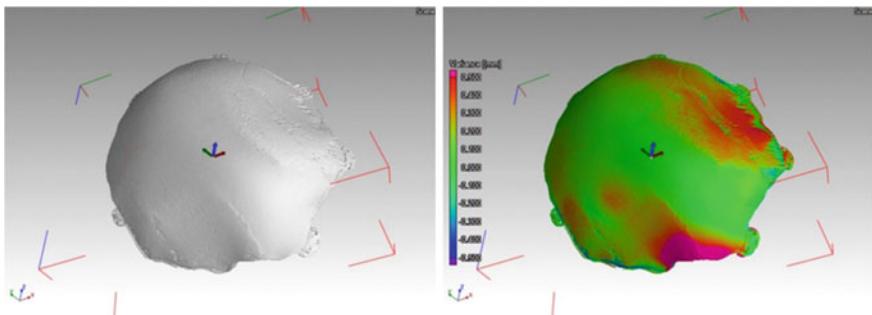
In Fig. 7, the left part, represents the obtained scan of the cranial implant and the right part shows the comparison with the CAD model. More significant deviations are visible in the scanning non-homogeneity areas. Locations intended for the implant attachment to the skull are within the tolerances.

Figure 8, the left part, represents the comparison of the hip socket prosthesis and the CAD model. Significant deviations are caused by the noise resulting from the change of the implant cumulative thickness. Subsequently, the implant was locally treated.

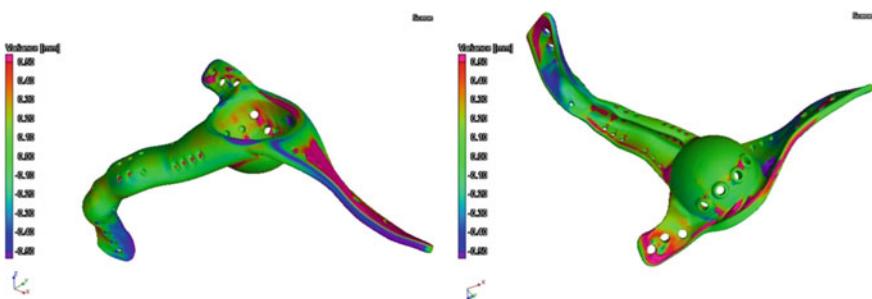
#### **Fitting methods**

Figure 9 represents the comparison of two identical objects applying the “Best Fit” method. 3D imaging displays only halves of the objects for the visualization purposes. In this case, the shape deviation is zero.

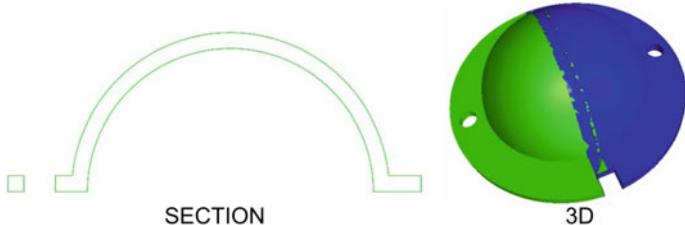
In case of significantly different objects (Fig. 10, applying the “Best-Fit” method, their mutual shift (left) is clearly visible, as a result of the height change



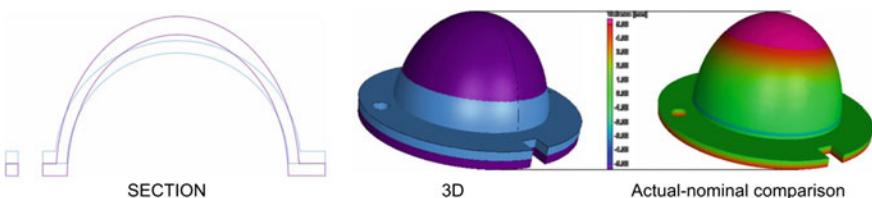
**Fig. 7** Cranial implant scan (*left*), comparison (*right*)



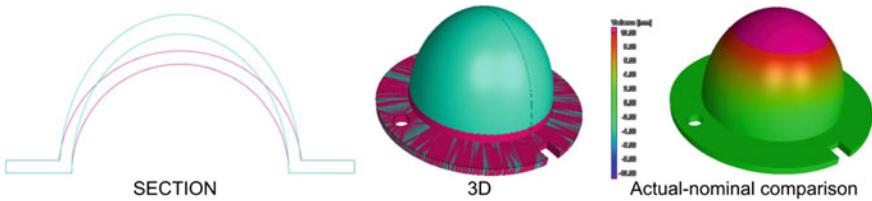
**Fig. 8** Hip socket implant including a part of the Pelvis



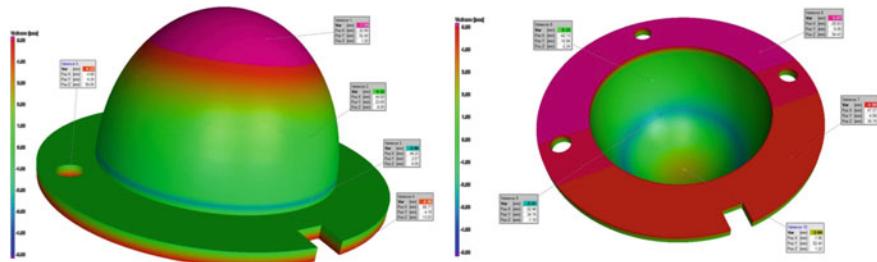
**Fig. 9** Comparison of identical objects



**Fig. 10** ‘Best-fit’ comparison



**Fig. 11** “Face-fit” comparison



**Fig. 12** Dimensional analysis

compensation. Consequently, this method is not suitable for the analysis. The right part of Fig. 10 represents a visible red band on the lower circular section, signaling the shift of the objects.

Figure 11 represents the comparison based on the “Face-Fit” principle, when individual objects are aligned according to the selected references. In this case, it was the alignment to the bottom plane and the groove. The left part Fig. 11. represents the cross section of the overlapping objects; the central part shows the overlap in 3D; and the right part shows the comparison of the scanned object and the 3D model, whereas the green areas fall within the tolerances, red areas are marginal, and purple areas are beyond the tolerance.

### 3.2 Dimensional Analysis

In case it is required to express numerically the values of deviations at any selected location, it is possible to use a tool enabling, after the objects are aligned and compared, the calculation and representation of a deviation at a particular location Fig. 12. If the dimensions between the components (distances) or the dimensions of individual components are required, it is appropriate to use a specialized tool, e.g. Calypso by the ZEISS Company. The dimensional analysis, however, is interesting especially in case of functional dimensions, for example the head diameter etc. in endoprostheses.

Another option how to inspect the product is to inspect the wall thickness; for this purpose it is necessary to enter the range of the assumed thickness of the object wall

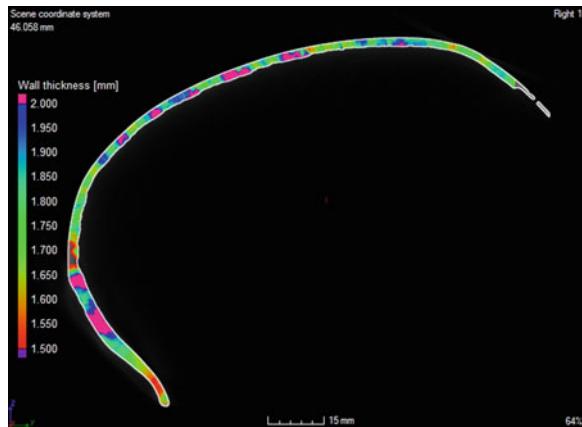
(the minimum and maximum value) and the material requirements. The thickness is evaluated as their perpendicular distance with certain permitted angle for the search thereof. In this case, the limits were set to the range of (1.5; 2.0) mm and the search angle was 30°.

### 3.3 Thickness Measurement

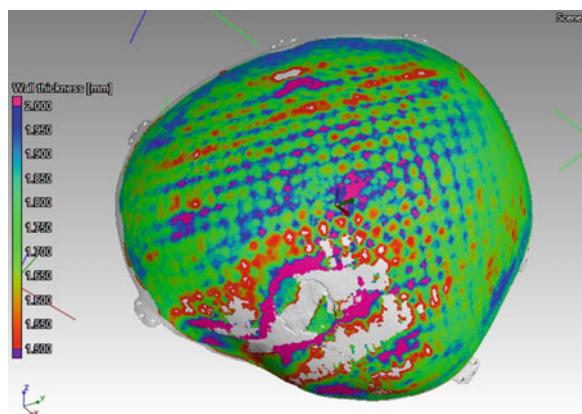
Figure 13 shows the implant cross section with the wall thickness colour coding. The cross section contains visible local changes in the wall thickness [1].

Figure 14 shows the change in the implant wall thickness along its entire surface. Clearly visible is the square mesh on the surface caused by the remains of the supporting material after it was removed from the implant inner side. The remaining material does not affect the implant functionality and thus its impact on the implant functionality is negligible [1].

**Fig. 13** Implant cross section with colour-coded wall thickness



**Fig. 14** Implant wall thickness change along the entire surface [1]



### 3.4 Defectoscopy

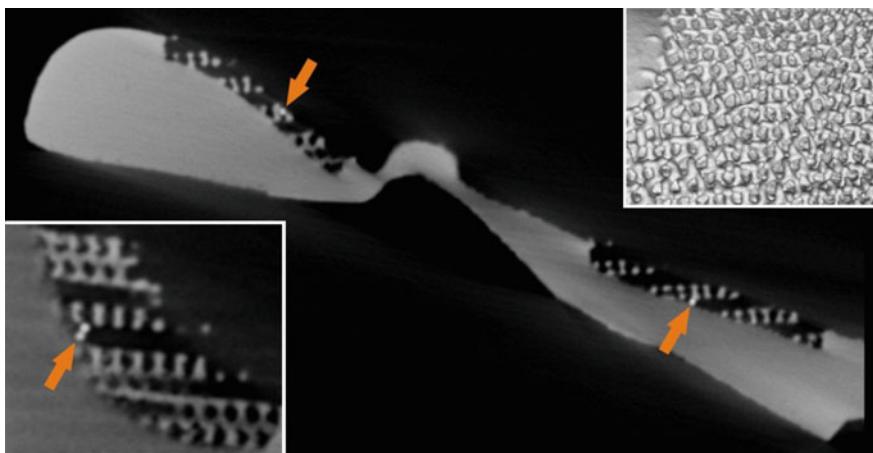
The defectoscopy means the identification of air bubbles (pores) or foreign materials (inclusion) in the basic material. Air bubbles can be formed due to imperfect manufacture, as the laser fails to fuse the building material at some locations. With regard to the fact that the implant is manufactured from the powder as the basic material, the probability of foreign material presence is minimal. Implant manufacturing is followed by the post-processing that includes also sand-blasting to reduce the surface roughness. In case of the manufacture of implants containing trabecular structures (porous structures facilitating the tissue grow through), these structures might trap the particles of the material intended for the sand-blasting. Its density is different; therefore, it can usually be identified as a significantly brighter point.

Figure 15 shows the implant cross section, while the arrows indicate the particles of the material intended for the sand-blasting. In the top right corner there is a detail of the trabecular structure, and the bottom left corner contains a detail of the inclusion from a different view.

To verify the ability to identify the defects during the manufacture process using the industrial computed tomography, an artefact containing artificial cavities was designed and subsequently manufactured.

When designing the artefact, it was necessary to consider several variables entering the manufacture process:

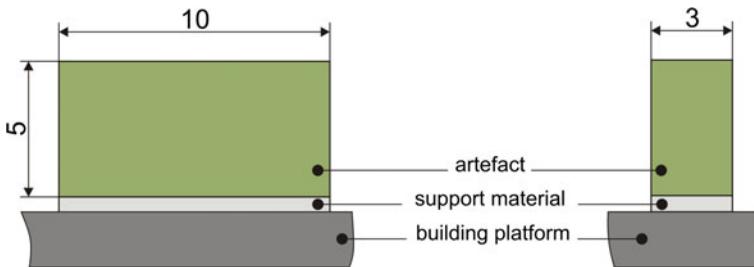
1. Material density—it affects the quantity of the energy required for the artefact scanning and, at the same time, it limits the minimum distance between the artefact and the X-ray tube (spot/voxel ration) and thus also the overall achieved artefact magnification on the detector.



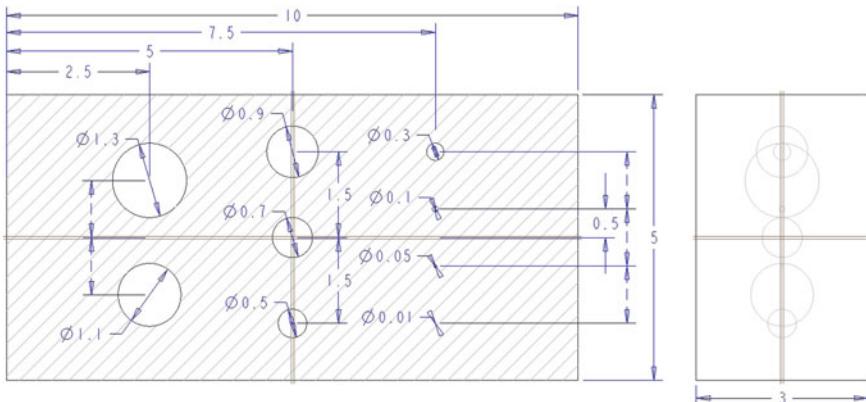
**Fig. 15** Inclusion (sand-blasting material)

2. Artefact size—it is limited by the combination of the linear attenuation coefficient which depends on the material density and the thickness of the radiated material. Penetration of a larger sample requires more energy. Sample size is also limited by the distance from the X-ray tube and thus also the maximum artefact magnification on the detector. Proposed artefact is shown in Fig. 17.
3. Artefact orientation in the manufacture artefact orientation affects the support application and the support removal method. The orientation as such does not affect the cavities inside the artefact (Fig. 16).
4. Sizes of pores (cavities) artefact designing was carried out using the cavities of spherical shape with the diameters of 0.01 mm to 1.3 mm, evenly distributed in the artefacts volume (Fig. 17).

The artefact was manufactured in cooperation with the CEIT Biomedical Engineering s.r.o company using the Eosint M280 machine from the EOS GmbH company, and the basic material used was Ti64 titanium alloy. The artefact drawing is shown in Fig. 17.



**Fig. 16** Artefact orientation in the building platform



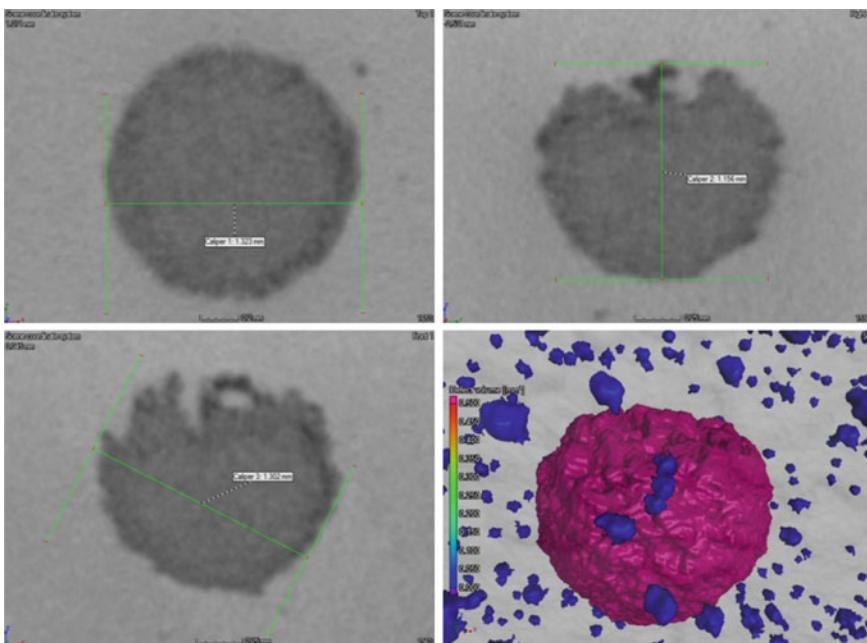
**Fig. 17** Artefact drawing

As the manufacturing is carried out layer by layer, all potential cavities are filled with unfused basic powder material. The assessment was carried out using the VGStudioMAX 2.2 software by the Volume Graphics GmbH Company.

Porosity is assessed on the basis of deviations of contrast between the material depicted in the images in grey tones and the air which is black. In case of the designed artefact, the assessment will be focused on the differences between two shades of grey, where the fused material is brighter, and thus of the density higher than the unfused material. Figure 18 shows a pore in three cross section planes and its reconstruction into 3D.

Artefact scanning was carried out using the Carl Zeiss Metrotom 1500 which was calibrated prior to the artefact scanning (detector calibration, centering, geometrical calibration, calibration of axis). Prior to the scanning, the assessed artefact was placed inside the preparation from extruded polystyrene which ensured the required stability during the scanning. The scanning consisted of two stages.

In the first stage, the artefact was scanned in the distances of 60, 200, and 600 mm from the X-ray radiation source. The scanning parameters were adjusted to the shortest possible distance between the artefact and the X-ray tube (60 mm) where a compromise between the output and the scanning distance was required. For this reason, a lower output and a longer scanning time (2,000 ms) were chosen. For other distances (200, 600 mm) the identical scanning parameters were maintained.



**Fig. 18** Pore representation

**Table 1** Input data table

Measurement	Distance (mm)	Voltage (kV)	Current ( $\mu\text{A}$ )	IT (ms)	Gain	Voxel ( $\mu\text{m}$ )
SN1	60	165	100	2000	16	16.39
SN2	200	165	100	2000	16	53.5
SN3	600	165	100	2000	16	159.54
SN4	200	190	280	1000	8	53.5
SN5	600	190	280	1000	8	159.54

Distance distance of object from X-Ray source

IT integration time

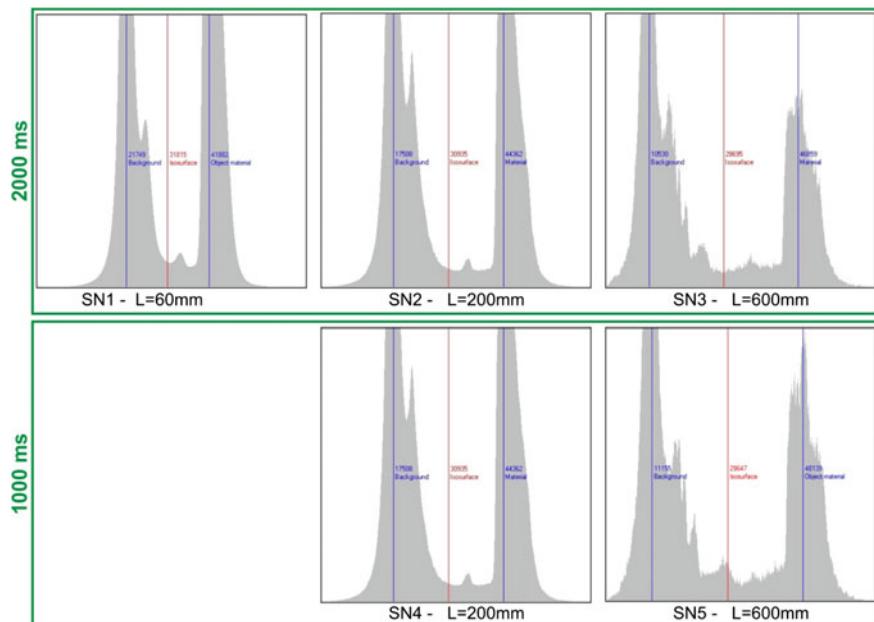
In the second scanning stage, the parameters were optimized for the distances of 200 and 600 mm; as a result, with the distance of 60 mm, the scanning was not possible. With regard to larger distances between the artefact and the sources, scanning parameters could be modified, whereas the applied output was increased and the scanning time was reduced down to 1,000 ms.

Scanning parameters for both stages are shown in Table 1. The data indicate that the voxel for 600 mm is approximately 10 times higher than with the distance of 60 mm, which means that while the artefact thickness of 3 mm with the distance of 60 mm consists of 183 voxels, with the distance of 600 mm it is only 18.8 voxels. This difference in number of voxels is demonstrated by the amount of obtained artefact details.

Figure 19 represents the impact of the distance on the histogram in the computer-assisted artefact reconstruction in the VGStudio SW, where the top line represents stage 1 (scanning time 2,000 ms, SN1, SN2, SN3 scans) and the bottom line represents stage 2 (scanning time 1,000 ms, SN4, SN5 scans). Due to increasing distance and thus decreasing number of voxels that form the artefact, changes occur in the histogram. With the distance of 600 mm, individual shades of grey are more significantly “depicted” in the histogram. With the distance of 600 mm, a significantly smaller powder identifying “hill” is visible than with the distances of 60 and 200 mm.

When comparing the histograms for scans with various adjustments of SN2 versus SN4 (distance of 200 mm), only a small difference is visible between the values of individual parameters (air, material, isosurface). With the distance of 600 mm and SN3 versus SN5 scans, the difference is visible also in histograms, where in the histogram for SN5, individual shades of grey are even more significantly highlighted. The histograms are shown in Fig. 19.

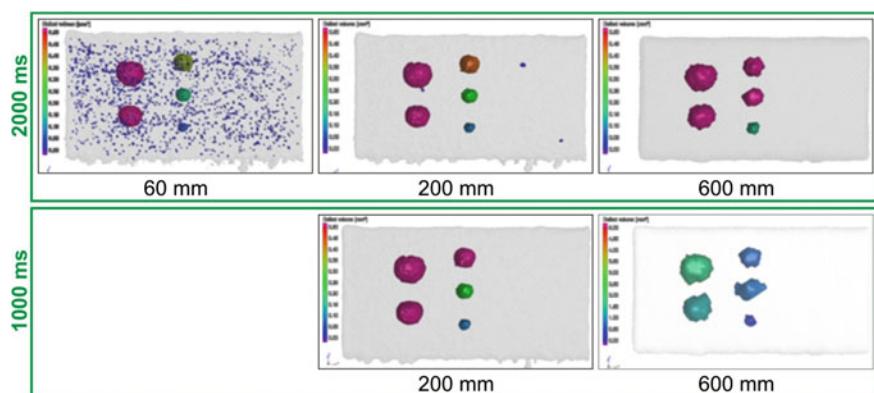
The results for the scanning time of 2,000 ms show that with the distance of 60 mm, the smallest assessed pore size is 0.3 mm and the smallest identifiable pore size is 0.06 mm. Smaller assessed pores cannot be identified, it means that the manufacture process is probably not able to manufacture an artificial cavity of the given size. With the scanning distance of 200 mm, the smallest identifiable pore size is 0.3 mm and with the distance of 600 mm it is 0.5 mm. Natural pores cannot be detected with the given distances.



**Fig. 19** Histograms for each measurement settings (distance from X-ray source and integration time)

With the scanning time of 1,000 ms, the diameter of the smallest identifiable pore for both scanning distances is 0.5 mm. In case of 600 mm distance, however, a significant error occurs in the pore reconstruction.

Figure 20 shows the output of the porosity assessment in the distances of 60, 200, and 600 mm for individual scanning stages.



**Fig. 20** Comparison of results

## 4 Conclusion

The article suggests the possibilities of new technologies in the field of custom-made implants verification with the aim to avoid implant failure inside a patients body and protect thus not only the patient but also the manufacturer. Modern technologies, such as the computer tomography, play an important role in the implant evaluation process. They facilitate analyzing the dimensions and the shape of implants, as well as the internal material structure, with the accuracy depending on the scanning system parameters, dimensions, and shape of an implant and its position during the scanning.

Together with the simulations using CAD/CAM software and with proper application they can provide a strong tool in the process of custom-made implants verification and validation, and in the implementation in the medical ISO standards they can represent an efficient tool for implementation of additive technologies in the medical manufacture.

The results indicate that with the adjustment of appropriate input parameters for the alignment of the generated scan and the CAD model, this technology is suitable for the analyses of shape deviations, compared to the CAD model. Subsequent use of tools facilitates the generation of deviations at a random implant spot, and after the assessment requirements (tolerances) are determined, it can be assessed whether the output complies with the requirements. Dimensional analysis is not a priority in case of cranial implants, as such implants contain the minimum amount of assessable dimensions. Greater significance is assigned to the surfaces where the implant is attached to the bone, as well as the tags for the attachment to the bone. A dimension which is important in case of an acetabular cup implant is its diameter.

Computed tomography is a tool appropriate also for the identification of possible pores and inclusions in the implant. In both cases, it is necessary to consider the scanning distance, because when it increases, the equipments resolution decreases. Tiny pores in the scanning distance of 60 mm are in line with the standard specified by the equipment manufacturer. In case of control of inclusions inside the trabecular structure, a suitable choice of the scanning distance and scanning parameters is required to achieve the best possible identification. The above mentioned indicates that sometimes a particular area must be scanned several times.

**Acknowledgments** Presented paper was supported by the project Research of New Diagnostic Methods in Invasive Implantology, MSR-3625/2010-11, Stimuli for Research and Development of/ Ministry of Education, Science, Research and Sport of the Slovak Republic and project KEGA 036TUKE-4/2013, Implementation of New Technologies in Design and Fabrication of Implants in Biomedical Engineering and Related Scientific Fields and Creation and promotion of Technologies in diagnostics of components and junctions with computed tomography (ITMS:26220220038) supported by the Research & Development Operational Programme funded by the ERDF.

## References

1. Hudák, R., Živčák, J., Tóth, T., Majerník, J., Lisý, M.: Evaluation of custom-made implants using industrial computed tomography. In: Digital Technologies 2014: The 10th International Conference, 9–11 July 2014, Zilina, Piscataway, IEEE, pp. 83–87 (2014). ISBN 978-1-4799-3301-3
2. Tóth, T., Hudák, R.: Computed tomography—its development, principle and image artifacts. *Acta Mech. Slovaca.* Ro. **17**(4), 40–47 (2013). ISSN 1335-2393
3. Kruth, J.P., Bartscherb, M., Carmignatoc, S., Schmittd, R., De Chiffree, L., Weckenmannf, A.: Computed tomography for dimensional metrology. *CIRP Ann. Manufact. Technol.* **60**(2), 821–842 (2011). <http://www.sciencedirect.com/science/article/pii/S0007850611002083>
4. Léonard, F., Tammas-Williams, S., Prangnell, P., Todd, I., Withers, P.: Assessment by X-ray CT of the effects of geometry and build direction on defects in titanium ALM parts. In: Conference on Industrial Computed Tomography (ICT), 19–21 Sept 2012, Wels, Austria (2012). <http://www.ndt.net/article/ctc2012/papers/91.pdf>
5. Slotwinski, J.A., Garboczi, E.J.: Porosity of additive manufacturing parts for process monitoring. In: 40th Review of Progress in Quantitative Non-Destructive Evaluation, 22–26 July 2013, Baltimore, MD (2013). [http://www.nist.gov/manuscript-publication-search.cfm?pub\\_id=914546](http://www.nist.gov/manuscript-publication-search.cfm?pub_id=914546)
6. Slotwinski, J.A.: Additive manufacturing at NIST, the science of digital manufacturing. <http://cba.mit.edu/events/13.03.scifab/Slotwinski.pdf> (2013). Accessed 7 Mar 2013
7. Brown, C., Lubell, J., Lipman, R.: Additive Manufacturing Technical Workshop Summary Report (2013)
8. Girardin, E., Renghini, C., Dyson, J., Calbucci, V., Moroncini, F., Albertini, G.: Characterization of porosity in a laser sintered MMCp using X-ray synchrotron phase contrast microtomography. *Mater. Sci. Appl.* **2**, 1322–1330 (2011)
9. Drstvensek, I., Hren, N.I., Strojnik, T., Brajlih, T., Valentan, B., Pogacar, V., Hartner, T.Z.: Applications of rapid prototyping in crano-maxillofacial surgery procedures. *Int. J. Biol. Biomed. Eng.* Issue <http://www.naun.org/main/NAUN/bio/bio-19.pdf>
10. Salmi M.: Medical applications of additive manufacturing in surgery and dental care. Doctoral Dissertation, Aalto university (2013). <https://aaltodoc.aalto.fi/bitstream/handle/123456789/11854/isbn9789526054964.pdf?sequence=1>
11. Podshivalov, L., Gomes, C.M., Zocca, A., Guenster, J., Bar-Yoseph, P., Fischer, A.: Design, analysis and additive manufacturing of porous structures for biocompatible micro-scale scaffolds. *Proc. CIRP* **5**, 247–252. ISSN 2212–8271. doi:[10.1016/j.procir.2013.01.049](https://doi.org/10.1016/j.procir.2013.01.049)
12. Bauza, M.B., Moylan, S.P., Panas, R.M., Burke, S.C., Martz, H.E., Taylor, J.S., Alexander, P., Knebel, R.H., Bhogaraju, R., O’Connell, M.T., Smokovitz, J.D.: Study of accuracy of parts produced using additive manufacturing. In: 2014 ASPE Spring Topical Meeting: Dimensional Accuracy and Surface Finish in Additive Manufacturing Berkeley, CA, United States, 13–16 April 2014. [http://www.nist.gov/manuscript-publication-search.cfm?pub\\_id=915785](http://www.nist.gov/manuscript-publication-search.cfm?pub_id=915785)
13. Matilainen, V.: Benchmarking of laser additive manufacturing process bachelor thesis. Lappeenranta University of Technology (2012). <http://www.doria.fi/bitstream/handle/10024/84863/BSc%20Thesis%20Matilainen%20fv.pdf?sequence=1>

# Assessment of Mortality Risk for Patients Undergoing Colorectal Surgery Using Regression Modeling

Kateřina Janurová and Lubomír Martínek

**Abstract** Since its beginning in the 1990s, laparoscopic surgical technique has partly or even totally replaced the classical open technique in some branches of surgery. The surgical resection of colorectal cancer, which this paper is engaged in, is no exception. However, besides the well-known benefits of this technique, there are also less known disadvantages, which can significantly influence morbidity. In such a situation we can ask, which technique will guarantee longer survival time to the patient. The medical survival censored data of 866 patients were evaluated by the Cox proportional hazards model in order to answer this question and to find the other parameters, which can influence the survival time of the patient. The surgical techniques were compared separately for patients with surgical resection of colon and for patients with surgical resection of rectum, because these two types of techniques are inherently different. Survival analysis performed by the Cox proportional hazards model led in both cases to the same conclusion, namely that there is no statistically significant difference in survival times between the two groups of patients operated by different surgical techniques.

## 1 Introduction

This chapter provides survival analysis of medical survival data of 866 patients, who underwent surgical resection of colorectal cancer at the University Hospital of Ostrava during the years 2001–2012. Currently, there are used two basic surgical techniques for the colectomy: either classical (open) or laparoscopic surgical technique. The first goal of the article is therefore the comparison of those techniques

---

K. Janurová (✉)  
IT4Innovations National Supercomputing Center,  
VŠB Technical University of Ostrava, Ostrava, Czech Republic  
e-mail: katerina.janurova@vsb.cz

L. Martínek  
The Clinic of Surgery, University Hospital Ostrava, Ostrava, Czech Republic  
e-mail: lubomir.martinek@fno.cz

for finding the type of surgery which guarantees longer overall survival time, if any. The consensus of European association of endoscopic surgery for colon carcinoma mentions, that there is no difference between morbidity of laparoscopic and open operations of colon [14], although the results of various medical studies regarding the comparison of morbidity and mortality after both types of surgeries of colon are commonly available and proved the merits of laparoscopic surgical technique, e.g. [1, 2, 7, 9, 10, 13].

The second goal is a development of complex survival model in order to find any other characteristics, which can influence survival time of patients. The characteristics of interest were chosen by the surgeons from the clinic of surgery of University Hospital of Ostrava and they are described in detail in the Sect. 4.

Data coming from the medical study are right-censored, in the form of vectors  $(t_1, c_1, \mathbf{x}_1), \dots, (t_n, c_n, \mathbf{x}_n)$ , where  $t_i$  is either a time of death or a time in which the observation of  $i$ th patient is stopped,  $c_i = 1$  (resp.  $c_i = 0$ ) is censoring indicator, according to death (resp. stopping time) occurring first and  $\mathbf{x}_i$  is the vector of other patients characteristics (covariates).

Analysis was performed using semiparametric Cox proportional hazards model separately for patients with surgical resection of colon and for patients with surgical resection of rectum.

## 2 Basic Relations in Survival Analysis

Survival analysis is a set of advanced statistical methods for the analysis of right-censored data. In the medical field of action we usually concentrate on estimating of survival function instead of the distribution function, because we are typically interested in describing how long the patients live rather than how quickly they die. Survival function represents the probability that the survival time of patient is greater than some specified time  $t$ , denoted according to [8]  $S(t)$ . The distribution function, denoted as  $F(t)$ , on the other hand, represents the probability that the survival time is less than or equal to some value  $t$ , so

$$S(t) = P(T > t) = 1 - P(T \leq t) = 1 - F(t). \quad (1)$$

The risk of death at some time  $t$  is expressed by the hazard function, denoted  $h(t)$ , which can be obtained from the probability that the patient dies at time  $t$ , under the condition of surviving to time  $t$  as

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}, \quad t \geq 0. \quad (2)$$

The relationship between hazard and survival function is given by the formula

$$h(t) = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{S(t)} = \frac{-d \log S(t)}{dt}, \quad (3)$$

where  $f(t)$  is the probability density function, which can be expressed as  $f(t) = F(t)' = -S(t)'$  for  $t \geq 0$ .

From (3), the survival function can be obtained as

$$S(t) = \exp \left[ - \int_0^t h(u) du \right] = \exp [-H(t)], \quad t \geq 0, \quad (4)$$

where  $H(t)$  is called integrated or cumulative hazard function, which expresses the risk of death from the beginning of follow-up until the time  $t$ .

### 3 The Cox Proportional Hazards Model

#### 3.1 Introduction

In medical studies, we are often interested in comparing survival under the “new” treatment with the “standard” treatment rather than a complete description of survival time. In those cases, the model is needed whose parameters can be used for comparison of the relative survival experience of the two treatment groups as well as the model who can adjust for other patient characteristics at the same time. The semiparametric regression model, who has a fully parametric regression structure but its dependence on time is left unspecified, is well suited for this type of problem. The form of the hazard function of a semiparametric regression model is given by

$$h(t|\mathbf{x}) = h_0(t)r(t|\mathbf{x}, \boldsymbol{\beta}), \quad (5)$$

where the vector  $\mathbf{x} = (x_1, x_2, \dots, x_p)$  denotes the  $p$  patient characteristics (covariates) and  $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)$  is the vector of unknown parameters. The hazard function (5), is the product of the two functions; the baseline hazard function  $h_0(t)$ , which depends on the survival time but not on covariates and a function  $r(\mathbf{x}, \boldsymbol{\beta})$ , which is not depended on time and which characterizes the effect of the covariates. Cox [4] proposed the parameterization of  $r(\mathbf{x}, \boldsymbol{\beta})$  as  $\exp(x_1\beta_1 + x_2\beta_2 + \dots + x_p\beta_p)$  what yielded the hazard function as follows:

$$h(t|\mathbf{x}) = h_0(t) \exp(x_1\beta_1 + x_2\beta_2 + \dots + x_p\beta_p) = h_0(t) \exp(\mathbf{x}'\boldsymbol{\beta}). \quad (6)$$

Under the Cox model, given by (6), the ratio of the hazard functions for a patient with set of covariates denoted  $\mathbf{x}^*$  compared to a patient with set of covariates denoted  $\mathbf{x}$  is

$$HR(\mathbf{x}^* : \mathbf{x}) = \frac{h(t|\mathbf{x}^*)}{h(t|\mathbf{x})} = \frac{h_0(t) \exp(x_1^*\beta_1 + x_2^*\beta_2 + \dots + x_p^*\beta_p)}{h_0(t) \exp(x_1\beta_1 + x_2\beta_2 + \dots + x_p\beta_p)} = \exp \{(\mathbf{x}^* - \mathbf{x})'\boldsymbol{\beta}\}. \quad (7)$$

The hazard ratio ( $HR$ ) represents the instantaneous risk over the study time period. It is independent on survival time; as a consequence Cox regression model is proportional-hazards model with proportionality as an important assumption of the model. In the medical study  $HR$  characterizes the measure of effect that describes the relationship between a predictor patients covariate of interest and survival time, after adjusting for the other patients covariate identified in the study and included in the model over the entire time period. For example, let  $x$  denote the parameter of the group covariate with group equal to 1 receiving treatment and group equal to 0 receiving placebo. If we put the treatment group in the numerator of (7), we get the  $HR$  of  $\exp(\beta)$ . A  $HR$  equal to one means, that there is no effect in the treatment. A  $HR$  greater than one, on the other hand, means, that the treatment group has larger risk of death in comparison with the placebo group. Similarly, a  $HR$  less than one implies, that the treatment group has lower risk of death than the placebo group. Looking at the relationship between hazard and survival function (4), we can see, that if the  $HR$  is less than one, the ratio of corresponding survival probabilities is larger than one. Consequently, the treatment group has larger probability of survival at any given time  $t$ , after adjusting for the other covariates. Under the Cox model, the survival function can be expressed as follows

$$S(t|\mathbf{x}) = [S_0(t)]^{\exp(\mathbf{x}'\boldsymbol{\beta})}, \quad (8)$$

where  $S_0(t) = \exp^{-H_0(t)}$  is the baseline survival function.

### 3.2 Fitting the Cox Proportional Hazards Model

In fitting the Cox proportional hazards model, our primary interest lies in finding the vector of unknown parameters  $\boldsymbol{\beta}$ . The popular approach was suggested by Cox [5], in which he obtained an expression he called “partial likelihood function” that is not depended on  $h_0(t)$  but only on the parameters of interest. Assuming that there are no ties in the survival times, the partial likelihood function is given by the expression

$$l(\boldsymbol{\beta}) = \prod_{i=1}^m \frac{\exp(\mathbf{x}'_{(i)}\boldsymbol{\beta})}{\sum_{j \in R(t_{(i)})} \exp(\mathbf{x}'_j\boldsymbol{\beta})}, \quad (9)$$

where the product is taken over the  $m$  distinct ordered survival times, the summation in the denominator is over all patients with survival or censored time greater than or equal to the time  $t_{(i)}$  (patients in the risk set at time  $t_{(i)}$ ), denoted  $R(t_{(i)})$  and  $x_{(i)}$  denotes the value of the covariate for the patient with ordered survival time  $t_{(i)}$ .

Parameter estimates are then obtained by maximizing the log partial likelihood function with respect to the parameter of interest, because it is computationally simpler than maximizing (9). The log partial likelihood function is

$$L(\boldsymbol{\beta}) = \log l(\boldsymbol{\beta}) = \sum_{i=1}^m \left\{ \mathbf{x}'_i \boldsymbol{\beta} - \ln \left[ \sum_{j \in R(t_{(i)})} \exp(\mathbf{x}'_j \boldsymbol{\beta}) \right] \right\} \quad (10)$$

and the maximum partial likelihood estimators are obtained by differentiating the right hand side of (10) with respect to each covariate  $\beta_i$ ,  $i = 1, \dots, p$  (yielding  $p$  equations), setting equal to zero and solving for the unknown parameters. The equation for the  $k$ th unknown parameter  $\beta_k$  is

$$\frac{\partial L(\boldsymbol{\beta})}{\partial \beta_k} = \sum_{i=1}^m \left\{ x_{ik} - \frac{\sum_{j \in R(t_{(i)})} x_{jk} \exp(\mathbf{x}'_j \boldsymbol{\beta})}{\sum_{j \in R(t_{(i)})} \exp(\mathbf{x}'_j \boldsymbol{\beta})} \right\}, \quad (11)$$

where the value of covariate  $x_k$  for the patient with observed survival time  $t_{(i)}$  is denoted by  $x_{ik}$ . The maximum partial likelihood estimator of vector  $\boldsymbol{\beta}$  is denoted  $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p)$  and the point estimate for  $HR$  is

$$\hat{HR}(\mathbf{x}^* : \mathbf{x}) = \exp \left\{ (\mathbf{x}^* - \mathbf{x})' \hat{\boldsymbol{\beta}} \right\}. \quad (12)$$

We can construct (1) 100 % confidence intervals for the point estimate of  $HR$  as

$$\exp \left\{ (\mathbf{x}^* - \mathbf{x})' \hat{\boldsymbol{\beta}} \pm z_{1-\alpha/2} \sqrt{(\mathbf{x}^* - \mathbf{x})' \hat{\text{Var}}(\hat{\boldsymbol{\beta}}) (\mathbf{x}^* - \mathbf{x})} \right\}, \quad (13)$$

where  $\hat{\text{Var}}(\hat{\boldsymbol{\beta}})$  is the estimator of the covariance matrix of the maximum partial likelihood estimator, obtained as an inverse of the observed information matrix evaluated at the maximum partial likelihood estimator

$$\hat{\text{Var}}(\hat{\boldsymbol{\beta}}) = I(\hat{\boldsymbol{\beta}})^{-1}. \quad (14)$$

The diagonal elements of the  $p$  by  $p$  observed information matrix are

$$\frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_k^2} \quad (15)$$

and the off-diagonal matrix elements are

$$\frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_k \partial \beta_l}. \quad (16)$$

As mentioned, the partial likelihood function method described above is based on the assumption, that we have no ties in the survival times. Treatment of ties was discussed for example by Breslow [3] or Efron [6], who provided approximations of partial likelihood.

### ***3.3 Assessing the Cox Proportional Hazards Model Assumptions***

The vital assumption to the interpretation and use of a fitted Cox proportional hazards model is the proportional hazards. Proportional hazards means that the hazard ratio is constant over the entire time period, i.e. the hazard function of one patient is proportional to the hazard function of the other patient.

This assumption can be tested by methods based on the Schoenfeld residuals. The Schoenfeld residuals are defined for each observed time of death, [12]. If the proportional hazards assumption holds for a particular covariate then the Schoenfeld residual for that covariate will not be related to survival time.

The test is calculated by correlating the Schoenfeld residuals for a particular covariate with a suitable transformation of time, usually based on  $\ln(t)$ ,  $\text{rank}(t)$  or the Kaplan-Meier estimate of the survival function  $K(t)$ . The null hypothesis is that the correlation is equal to zero. Rejection of null hypothesis concludes that proportional hazards assumption is violated.

In the assessment of model adequacy, it is also important to determine whether there are any patients that have an undue influence on the estimates of the parameters or on fit of the model. Patients with this influence are said to be influential observations.

Another issue in assessment of model adequacy is detection of nonlinearity. Nonlinearity means that the functional form in the parametric part of the model is specified incorrectly. For detecting nonlinearity the martingale residuals may be plotted against covariates for the detection. If the nonlinearity is detected in the covariate, the fractional polynomials method [11] can be used to suggest the transformation of the scale of covariate.

## **4 Evaluation of the Effect of Covariates on Survival Time of Patients with Surgery of Colon**

### ***4.1 Description of Covariates***

From the initial amount of patients characteristics, the surgeons had selected small subset of eleven interesting covariates, whose effect on survival time was further

**Table 1** Description of the covariates obtained from the FN Ostrava study, 866 patients

Covariate	Description	Codes/Values
<i>conversion</i>	Type of the surgery	1 = open
		2 = laparoscopic
		3 = conversion from laparoscopic to open
<i>stage</i>	Stages of cancer	1–4 from best to worst
<i>gender</i>	Gender	f = Female
		m = Male
<i>asa</i>	Anesthesiology indicator of physical condition of the patient	1–4 from best to worst
<i>care</i>	Type of care	p = palliative care
		c = curative care
<i>grading</i>	Tumor grade	1 = Grade 1
		2 = Grade 2
		3 = Grade 3
<i>diagnosis</i>	Place of tumor in the colon	mm
<i>age</i>	Age at hospital admission	Years
<i>bmi</i>	Body mass index	kg/m <sup>2</sup>
<i>bl</i>	Blood loss during surgery	ml
<i>rpn</i>	Ratio of examined positives nodes	%

examined by the Cox proportional hazards model. The description of selected covariates is given in Table 1.

Table 2 shows the results of unicovariate proportional hazards analysis of discrete covariates in relation to survival time (in days). The results of the same analysis, but in this case for continuous covariates, are shown in Table 3.

The *p*-value of the partial likelihood ratio test indicated, that the effect of covariates *gender* and *diagnosis* was not significant at the 20 % level, therefore these covariates were not included in the multivariable model.

The method of purposeful selection of covariates [8] was chosen for the model building process, because it is the only method, which allows to the data analyst full control over all modeling steps. For the approximation of partial likelihood function was chosen the Breslow formula [3]. With this method we were able to develop the preliminary main effect model, described in Table 4.

The categories 1 and 2 as well as the categories 3 and 4 of the covariate *asa* were merged together during the model building process to create a new binary covariate *asa\_3+4*. Furthermore the categories 1, 2 and 3 of covariate *stage* were merged in a new binary covariate *stage\_4* and analogously the merged categories 1 and 2 of the covariate *grading* formed one category for a new binary covariate *grading\_3*. These changes were done, because the new binary covariates yielded a simpler model and

**Table 2** Estimated median time to death with 95 % confidence interval estimates and partial likelihood ratio test *p*-values for categorical covariates, 524 patients with surgery of colon

Covariate	Category (events, <i>n</i> )	Median time to death (95 % CIE)	Partial LR test <i>p</i> -value
<i>conversion</i>	1 (160, 237)	1323 (1067, 1600)	0.001
	2 (132, 264)	2056 (1732, 3487)	
	3 (14, 23)	1397 (724, -)	
<i>stage</i>	1 (22, 79)	3487 (3112, -)	0.001
	2 (60, 149)	-(2442, -)	
	3 (72, 127)	1875 (1423, 2775)	
	4 (124, 139)	435 (354, 585)	
<i>gender</i>	f (121, 212)	1619 (1284, 2556)	0.722
	m (185, 312)	1707 (1419, 2021)	
<i>asa</i>	1 (20, 46)	-(1469, -)	0.011
	2 (128, 237)	1933 (1423, 2953)	
	3 (140, 214)	1439 (1071, 1732)	
	4 (18, 27)	1180 (409, -)	
<i>care</i>	c (152, 354)	3100 (2462, -)	<0.001
	p (154, 170)	398 (331, 503)	
<i>grading</i>	1 (103, 176)	1933 (1492, 2665)	0.048
	2 (143, 272)	1850 (1497, 2658)	
	3 (33, 45)	1165 (499, 2027)	

the decreased number of codes had not changed the value of coefficients for any of the other covariates in the model.

The further step in the model building process was the examining of the scale of the two continuous covariates included in the preliminary main effect model, *age* and *rpn*.

## 4.2 Using Fractional Polynomials to Examine the Scale of Continuous Covariates

The assumption of linearity in log hazard of all continuous covariates was examined by the fractional polynomials method. The results of the analysis, which suggested the scale transformation for covariate *age*, are shown in Table 5.

The linear model was evaluated as significantly better than the model without the covariate *age*, then the one-term model was evaluated as significantly better than the linear model and finally the two-term model was evaluated as significantly better than the model with one-term. The two-term (2, 2) fractional polynomial model was assessed as the one with numerically smallest value of  $-2\log$  partial likelihood

**Table 3** Estimated median time to death with 95 % confidence interval estimates and partial likelihood ratio test *p*-values for continuous covariates, 524 patients with surgery of colon

Covariate	Interval (events, <i>n</i> )	Median time to death (95 % CIE)	Partial LR test <i>p</i> -value
<i>diagnosis</i>	(18.0, 18.2] (74, 138)	1853 (1427, 3252)	0.747
	(18.2, 18.5] (90, 133)	1284 (882, 1732)	
	(18.5, 18.7] (105, 195)	2027 (1492, 2980)	
	(18.7, 19.0] (37, 58)	1245 (768, 2021)	
<i>age</i>	(26, 60] (77, 147)	2036 (1469, 3469)	<0.001
	(60, 67] (64, 125)	2442 (1271, -)	
	(67, 75] (83, 134)	1731 (1067, 2021)	
	(75, 97] (82, 118)	1229 (859, 1472)	
<i>bmi</i>	(15.4, 23.7] (87, 132)	1070 (703, 1357)	0.006
	(23.7, 26.2] (83, 134)	1497 (1026, 2056)	
	(26.2, 29.6] (65, 127)	2147 (1677, -)	
	(29.6, 45.7] (71, 131)	1898 (1449, 2556)	
<i>bl</i>	(0, 1] (164, 274)	1732 (1372, 2147)	0.171
	(1, 200] (76, 146)	1561 (1303, 2665)	
	(200, 2000] (66, 104)	1397 (962, 1875)	
<i>rpn</i>	(0, 1] (97, 248)	3469 (2665, -)	<0.001
	(1, 25] (47, 79)	1554 (962, 2462)	
	(25, 100] (57, 75)	872 (503, 1110)	

**Table 4** Estimated coefficients, standard errors, *z*-scores, two-tailed *p*-values and 95 % confidence interval estimates for the preliminary main effect proportional hazards model, patients with surgery of colon

Covariate	Coeff.	Std. Err.	<i>z</i>	<i>p</i> >   <i>z</i>	95 % CIE
<i>age</i>	0.029	0.008	3.622	<0.001	0.013, 0.044
<i>rpn</i>	0.015	0.003	5.517	<0.001	0.010, 0.020
<i>asa_3+4</i>	0.378	0.153	2.467	0.014	0.078, 0.679
<i>stage_4</i>	1.529	0.175	8.757	<0.001	1.187, 1.871
<i>grading_3</i>	0.589	0.225	2.625	0.009	0.149, 1.029
Log-likelih.	-1036.273				

among the 44 two-term models fit. Therefore we decided to proceed with the two transformation of covariate *age* as follows

$$\text{agefp1} = \left( \frac{\text{age}}{10} \right)^2 \quad (17)$$

**Table 5** Summary results of fractional polynomials analysis of *age*, patients with surgery of colon

Scale of covariate <i>age</i>	Log-Likelihood	<i>G</i> for Model x Linear	Approx. <i>p</i> -value	Powers
Not in the model	-1034.113			
Linear (1 df)	-1036.273	0.000	<0.001 <sup>a</sup>	1
<i>J</i> = 1 (2 df)	-1033.148	6.250	0.012 <sup>b</sup>	3
<i>J</i> = 2 (4 df)	-1028.968	14.610	0.015 <sup>c</sup>	2; 2

<sup>a</sup>Compares linear model to model without *age*

<sup>b</sup>Compares the best *J* = 1 model to one with *age* linear

<sup>c</sup>Compares the best *J* = 2 model to the best *J* = 1 model

**Table 6** Estimated coefficients, standard errors, *z*-scores, two-tailed *p*-values and 95 % confidence interval estimates for the preliminary main effect proportional hazards model, patients with surgery of colon

Covariate	Coeff.	Std. Err.	<i>z</i>	<i>p</i> >   <i>z</i>	95 % CIE
<i>agefp1</i>	0.029	0.008	3.622	<0.001	0.013, 0.044
<i>agefp2</i>	0.029	0.008	3.622	<0.001	0.013, 0.044
<i>rpn</i>	0.015	0.003	5.517	<0.001	0.010, 0.020
<i>asa_3+4</i>	0.378	0.153	2.467	0.014	0.078, 0.679
<i>stage_4</i>	1.529	0.175	8.757	<0.001	1.187, 1.871
<i>grading_3</i>	0.589	0.225	2.625	0.009	0.149, 1.029
Log-likelih.	-1036.273				

and

$$\text{agefp2} = \left(\frac{\text{age}}{10}\right)^2 \cdot \left(\frac{\text{age}}{10}\right). \quad (18)$$

The results of fitting the model containing transformation (17) and (18) of covariate *age* are shown in Table 6.

The results of fractional polynomial analysis of covariate *rpn* supported the assumption of linearity in the log hazard.

The next step in the model development process was selection of interactions.

#### 4.3 Interactions in the Model

The interaction terms were formed, except for the covariate *age*, as an arithmetic product of the pair of covariates from Table 6. Interaction terms containing covariate *age* were created using the two polynomial transformations (17) and (18), which yielded one extra degree of freedom in chi square distribution in partial likelihood ratio test. The list of possible interactions, the degrees of freedom of the interaction

and the *p*-value for the partial likelihood ratio test that compared the models with and without the interaction are shown in Table 7.

It can be seen from the *p*-values in Table 7, that no interaction was significant at the traditional 5 % level; therefore they are not included in the model. The final step in the model development process was checking for adherence to key model assumption.

#### 4.4 Model Assumption Checking

Assessment of the proportional hazards assumption was performed by the score test based on the scaled Schoenfeld residuals using functions of survival time mentioned in Sect. 3.3. The *p*-values of the score test, shown in Table 8, indicated that there was no evidence of the hazard being nonproportional in any of the model covariates.

The analysis of the influential observation shown in Fig. 1 suggested that none of the observations was significantly influential individually (all changes in the coefficients were smaller than 20 %).

#### 4.5 Interpretation of the Final Model

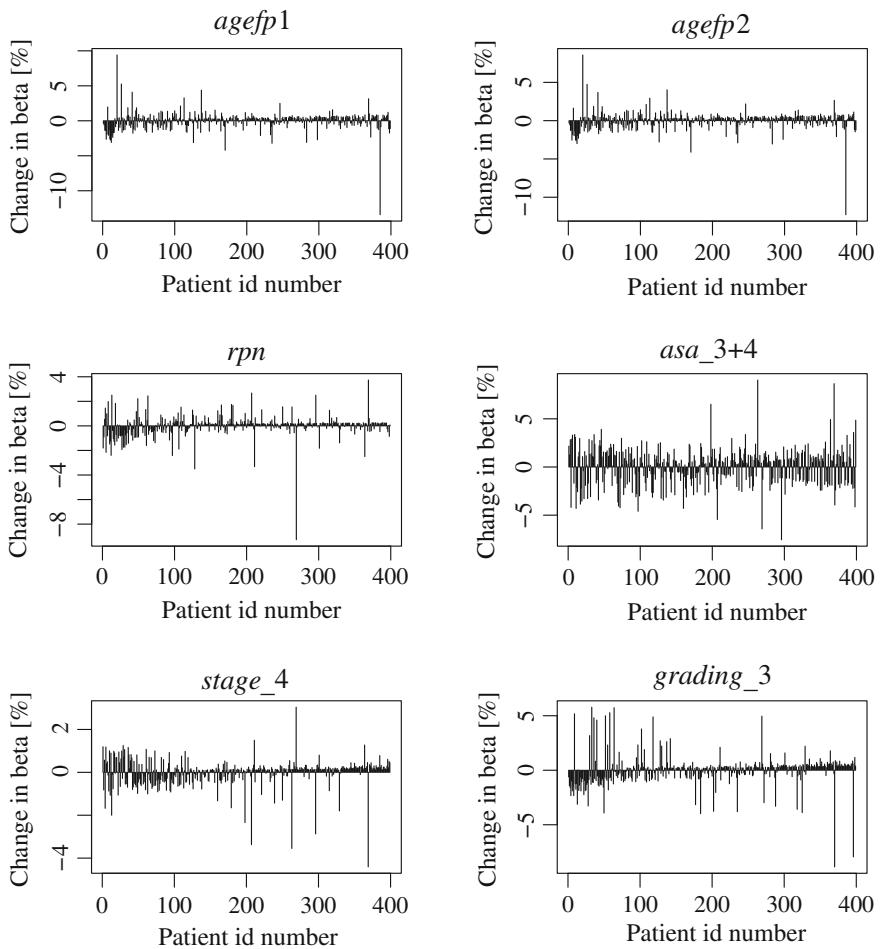
The list of estimated hazard ratios of the final model obtained from Table 6 is showed in Table 9. Value of the estimated hazard ratio for a 10 % increase in *rpn* is 1.18 with interpretation that patients, who had at the beginning of the study 10 % higher ratio of positive nodes, are dying at 18 % higher rate than patients at the lower ratio of

**Table 7** Interaction covariates, degrees-of-freedom (df), and *p*-values for partial likelihood ratio test comparing model with and without interaction, patients with surgery of colon

Interaction	Covariate	df	<i>p</i> -value
<i>age</i>	<i>rpn</i>	2	0.687
	<i>asa_3+4</i>	2	0.890
	<i>stage_4</i>	2	0.365
	<i>grading_3</i>	2	0.711
<i>rpn</i>	<i>asa_3+4</i>	1	0.971
	<i>stage_4</i>	1	0.398
	<i>grading_3</i>	1	0.885
<i>asa_3+4</i>	<i>stage_4</i>	1	0.919
	<i>grading_3</i>	1	0.336
<i>stage_4</i>	<i>grading_3</i>	1	0.804

**Table 8** Estimated coefficients, standard errors, z-scores, two-tailed  $p$ -values and 95 % confidence interval estimates for the main effect proportional hazards model, patients with surgery of colon

Covariate	df	$t$ p-value	$\ln(t)$ p-value	$S_KM(t)$ p-value	rank( $t$ ) p-value
<i>agefp1</i>	1	0.529	0.733	0.787	0.921
<i>agefp2</i>	1	0.558	0.685	0.830	0.968
<i>rpn</i>	1	0.760	0.159	0.448	0.343
<i>asa_3+4</i>	1	0.515	0.444	0.451	0.415
<i>stage_4</i>	1	0.329	0.144	0.183	0.146
<i>grading_3</i>	1	0.441	0.168	0.475	0.504
Global	6	0.825	0.317	0.690	0.622



**Fig. 1** Analysis of influential observations for each covariate in the final model from Table 8

**Table 9** Estimated hazard ratios and 95 % confidence interval estimates for *rpn*, *asa\_3+4*, *stage\_4* and *grading\_3*, patients with surgery of colon

Covariate	$\hat{HR}$	95 % CIE
<i>rpn</i>	1.180	1.117, 1.245
<i>asa_3+4</i>	1.489	1.097, 2.020
<i>stage_4</i>	4.503	3.200, 6.344
<i>grading_3</i>	1.925	1.238, 2.996

positive nodes. The increased rate of dying as little as 12 % or as high as 25 % is also consistent with the data, as indicated by the 95 % confidence interval in Table 9.

The estimated hazard ratio for *asa\_3+4* covariate is 1.49. This means that patients, who were at the beginning of the study classified in category 3 or 4 in covariate *asa*, are dying at a 49 % higher rate than patients included in category 1 or 2. The 95 % confidence interval in Table 9 suggests that the rate could be as much as a 2.02-fold or as little as a 1.10-fold increase in the rate of dying.

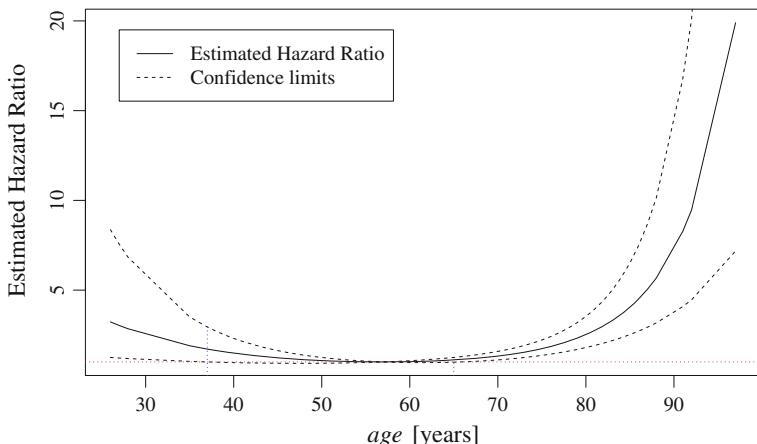
Value of the estimated hazard ratio for *grading\_3* is 1.93 with interpretation that patients, who were at the beginning of the study classified in category 3 in covariate *grading*, are dying at a 93 % higher rate than patients included in category 1 or 2. The increased rate of dying as little as 24 % or as high as 200 % is also consistent with the data, as indicated by the 95 % confidence interval in Table 9.

The estimated hazard ratio for covariate *stage\_4* is 4.50. Patients, who were at the beginning of the study classified in category 4 in covariate *stage*, are dying at a rate 4.5 times higher than patients included in category 1, 2 or 3. The 95 % confidence interval in Table 9 suggests that the rate could be as little as a 3.20-fold or as much as a 6.34-fold increase in the rate of dying.

The covariate age is modeled with two non-linear terms (17), (18); as a consequence the estimated hazard ratio will depend on the values of age being compared. Therefore the comparison of hazard ratio is done relative to its minimum value in about 57 years. Selected values of estimated hazard ratio and its 95 % confidence interval estimates are listed in Table 10.

**Table 10** Estimated hazard ratios and 95 % confidence interval estimates of the effect of *age* compared with age = 57 years, patients with surgery of colon

<i>age</i>	$\hat{HR}$	95 % CIE
30	2.522	1.135, 5.602
40	1.490	0.961, 2.311
50	1.075	0.927, 1.247
60	1.015	0.964, 1.067
70	1.327	1.111, 1.585
80	2.522	1.800, 3.533
90	7.264	3.745, 14.089



**Fig. 2** Estimated hazard ratio and 95 % confidence limits of covariate *age*

The overall change in estimated hazard ratios is plotted in Fig. 2. It can be seen that the rate of dying relative to the minimum in 57 years decreases prior to this value and after overcoming this value increases in an asymmetric manner. Sorting of all values of estimated hazard ratio showed, that the value 1 is contained in the confidence limits for covariate age between the values 37 and 65. Therefore we can conclude that there is a significant increase in the rate of dying for age less than 37 and greater than 65.

## 5 Evaluation of the Effect of Covariates on Survival Time of Patients with Surgery of Rectum

### 5.1 Description of Covariates

For the effect evaluation on survival time were in the case of patients with surgical resection of rectum selected the following covariates: *stage*, *gender*, *asa*, *conversion*, *age*, *bmi* and *bl*. Their description is given in Table 1. The univariable analysis of each covariate in relation to survival time (in days) is shown in Tables 11 and 12.

The covariates not included in the multivariable analysis were *gender* and *bl*, since they were not significant in the univariable analysis at the 20 % level.

For the model building process was again chosen the method of purposeful selection of covariates and Breslow approximation of partial likelihood function. Similarly

**Table 11** Estimated median time to death with 95 % confidence interval estimates and partial likelihood ratio test *p*-values for categorical covariates, 342 patients with surgery of rectum

Covariate	Category (events, <i>n</i> )	Median time to death (95 % CIE)	Partial LR test <i>p</i> -value
<i>conversion</i>	1 (79, 138)	1588 (1270, 3316)	0.180
	2 (111, 190)	1309 (1014, 1739)	
	3 (6, 14)	2304 (1614, -)	
<i>stage</i>	1 (11, 55)	-(-, -)	<0.001
	2 (28, 74)	-(2453, -)	
	3 (65, 112)	1614 (1238, 2808)	
	4 (59, 64)	376 (262, 521)	
<i>gender</i>	f (64, 115)	1351 (956, -)	0.912
	m (132, 227)	1530 (1238, 2304)	
<i>asa</i>	1 (20, 46)	-(1270, -)	0.011
	2 (97, 181)	1741 (1351, 3567)	
	3 (75, 109)	1014 (855, 1572)	
	4 (4, 6)	488 (302, -)	

**Table 12** Estimated median time to death with 95 % confidence interval estimates and partial likelihood ratio test *p*-values for continuous covariates, 342 patients with surgery of rectum

Covariate	Interval (events, <i>n</i> )	Median Time to Death (95 % CIE)	Partial LR test <i>p</i> -value
<i>age</i>	(25, 58] (47, 95)	2702 (1309, -)	0.004
	(58, 65] (47, 82)	1739 (1129, 3581)	
	(65, 72] (46, 83)	1507 (971, -)	
	(72, 84] (56, 82)	887 (717, 1303)	
<i>bmi</i>	(15.6, 24.2] (57, 86)	844 (564, 1201)	0.018
	(24.2, 26.6] (41, 85)	2549 (1360, -)	
	(26.6, 30.0] (50, 89)	2361 (1309, 3567)	
	(30.0, 42.6] (47, 81)	1538 (901, 2798)	
<i>bl</i>	(0, 0.1] (82, 121)	956 (599, 1360)	0.313
	(0.1, 200] (27, 66)	2083 (1508, -)	
	(200, 500] (46, 89)	2145 (1419, -)	
	(500, 2500] (41, 66)	1065 (861, 2361)	

as in the case of patients with surgery of colon, the categories 1 and 2 of the covariate *asa* were merged in the process of model development to create a new binary covariate *asa\_3+4*. The preliminary main effect model is shown in Table 13.

**Table 13** Estimated coefficients, standard errors,  $z$ -scores, two-tailed  $p$ -values and 95 % confidence interval estimates for the preliminary main effect proportional hazards model, patients with surgery of rectum

Covariate	Coeff.	Std. Err.	$z$	$p >  z $	95 % CIE
<i>age</i>	0.020	0.010	2.053	0.040	0.001, 0.040
<i>bmi</i>	-0.039	0.020	-1.929	0.054	-0.078, 0.001
<i>stage_2</i>	0.754	0.357	2.111	0.035	0.054, 1.454
<i>stage_3</i>	1.433	0.330	4.335	<0.001	0.785, 2.081
<i>stage_4</i>	2.989	0.339	8.805	<0.001	2.324, 3.655
<i>asa_3+4</i>	0.372	0.184	2.023	0.043	0.012, 0.732
Log-likelih.	-781.633				

## 5.2 Using Fractional Polynomials to Examine the Scale of Continuous Covariates

The scale of the two continuous covariates included in the preliminary main effect model, *age* and *bmi*, was examined by the fractional polynomial analysis, which supported the assumption of linearity in log hazards for both covariates. The further step in the model development process was selection of interactions.

## 5.3 Interactions in the Model

The possible interactions, the degrees of freedom of the interaction and the  $p$ -value for the partial likelihood ratio test that compared the models with and without the interaction are listed in Table 14.

The  $p$ -values in Table 14 indicated that there were no statistically significant interactions which should be included in the model. The final step in the model development process, assessing the adequacy of the model, is provided in the following section.

**Table 14** Interaction covariates, degrees-of-freedom (df), and  $p$ -values for partial likelihood ratio test comparing model with and without interaction, patients with surgery of rectum

Interaction	Covariate	df	$p$ -value
<i>age</i>	<i>bmi</i>	1	0.637
	<i>asa_3+4</i>	1	0.890
	<i>stage</i>	1	0.266
<i>bmi</i>	<i>asa_3+4</i>	1	0.973
	<i>stage</i>	1	0.298
<i>asa_3+4</i>	<i>stage</i>	1	0.864

**Table 15** Estimated coefficients, standard errors,  $z$ -scores, two-tailed  $p$ -values and 95 % confidence interval estimates for the main effect proportional hazards model, patients with surgery of colon

Covariate	df	$t$ p-value	$\ln(t)$ p-value	$S_{KM}(t)$ p-value	rank( $t$ ) p-value
<i>age</i>	1	0.677	0.344	0.523	0.505
<i>bmi</i>	1	0.301	0.618	0.621	0.694
<i>stage_2</i>	1	0.473	0.899	0.650	0.690
<i>stage_3</i>	1	0.212	0.641	0.250	0.260
<i>stage_4</i>	1	0.084	0.232	0.062	0.061
<i>asa_3+4</i>	1	0.448	0.278	0.268	0.255
Global	6	0.677	0.560	0.247	0.229

## 5.4 Model Assumption Checking

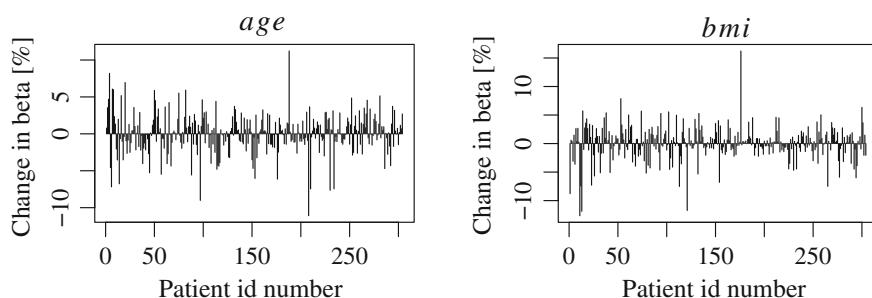
The results of the score test of scaled Schoenfeld residuals using functions of survival time are shown in Table 15. The  $p$ -values indicated that there was no evidence of the hazard being nonproportional in any of the model covariates.

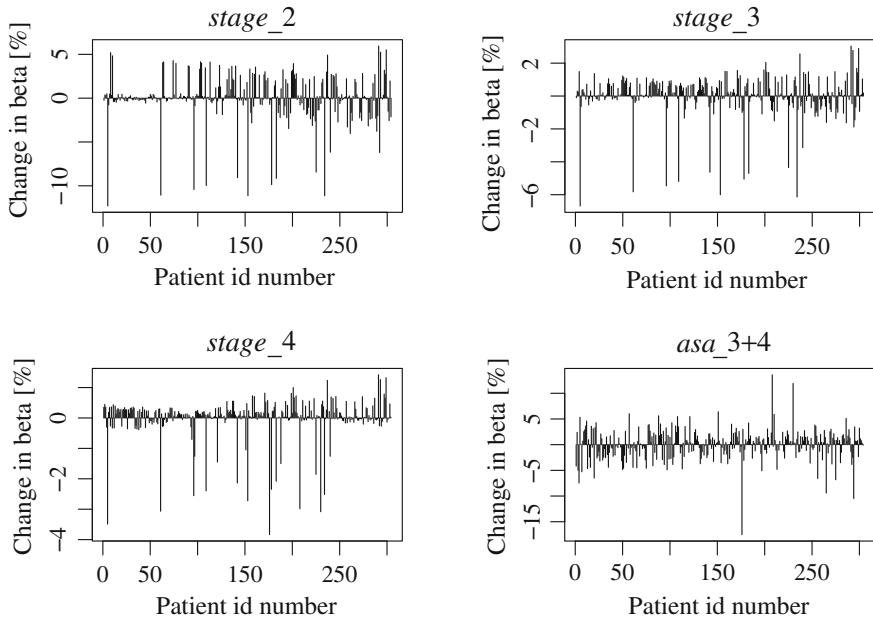
None of the observations was evaluated as influential, because all changes in the coefficients were smaller than 20 % see Figs. 3 and 4.

## 5.5 Interpretation of the Final Model

Estimated hazard ratios of the final model obtained from Table 13 are showed in Table 16.

The estimated hazard ratio for a 10 % increase in *age* is 1.22. This means that patients with a 10 % higher age are dying at 22 % higher rate than patients with lower age. The increased rate of dying as little as 1 % or as high as 49 % is also consistent

**Fig. 3** Analysis of influential observations for each covariate in the final model from Table 13



**Fig. 4** Analysis of influential observations for each covariate in the final model from Table 13

**Table 16** Estimated hazard ratios and 95 % confidence interval estimates for *age*, *bmi*, *stage\_2*, *stage\_3*, *stage\_4*, *asa\_3+4*, patients with surgery of rectum

Covariate	$\hat{H}R$	95 % CIE
<i>age</i>	1.221	1.010, 1.492
<i>bmi</i>	0.677	0.458, 1.010
<i>stage_2</i>	2.126	1.055, 4.282
<i>stage_3</i>	4.190	2.192, 8.009
<i>stage_4</i>	19.869	10.214, 38.650
<i>asa_3+4</i>	1.451	1.012, 2.080

with the data, as suggested by the 95 % confidence interval in Table 16. Because the model is linear in covariate *age*, this result holds over the observed range of *age*.

The presence of covariate *bmi* in the model is constrained by its significant influence on values of parameter of other covariates included in the final model. As can be seen from the 95 % confidence interval for hazard ratio, which contains the value 1. The conclusion is, that covariate *bmi* has no effect on survival time of the patients.

Value of the estimated hazard ratio for covariate *stage*, category = 2 is 2.13 with interpretation that patients, who were at the beginning of the study classified in category 2 in covariate *stage*, are dying at a 113 % higher rate than patients included in category 1. The increased rate of dying as little as 6 % or as high as 328 % is

also consistent with the data, as indicated by the 95 % confidence interval. Patients included in category 3 in covariate *stage*, are dying at a rate 4.19 times higher than patients included in category 1. The increased rate of dying could be as little as a 2.20-fold or as much as a 8.01-fold. Patients included in category 4 in covariate *stage*, are dying at a rate 19.87 times higher than patients included in category 1. The 95 % confidence interval suggests that the rate could be as little as a 10.21-fold or as much as a 38.65-fold.

The estimated hazard ratio for covariate *asa\_3+4* is 1.45 with interpretation that patients, who were at the beginning of the study classified in category 3 or 4 in covariate *asa*, are dying at a 45 % higher rate than patients included in category 1 or 2. The increased rate of dying as little as 1 % or as high as 108 % is also consistent with the data.

## 6 Conclusion

Multivariable analysis of right-censored medical survival data of 866 patients, who underwent surgical resection of colorectal cancer at the University Hospital of Ostrava during the years 2001–2012, was performed to compare survival under two different types of surgical techniques in order to answer the questions which type of surgical technique guarantees longer overall survival time and what are the other characteristics which can affect the survival time of the patients. The data were analyzed separately for patients with surgical resection of colon and for patients with surgical resection of rectum.

Using the Cox proportional hazards model with Breslow approximation of partial likelihood function, we concluded that the type of the surgical technique does not affect the survival time of both the patient with surgical resection of colon and the patient with surgical resection of rectum. This result contrasts with the results of medical studies cited in the introduction, but on the other hand fully corresponds with consensus of European association of endoscopic surgery for colon carcinoma.

Regarding the patients with surgical resection of colon, the Cox proportional hazards model evaluated as significantly influencing survival time covariates age of patient at hospital admission, stage of cancer, anesthesiology indicator of physical condition of the patient, tumor grade and ratio of examined positives nodes. For patients with surgical resection of rectum, Cox proportional hazards model evaluated as significantly influencing survival time covariates age of patient at hospital admission, body mass index, stage of cancer and anesthesiology indicator of physical condition of the patient.

**Acknowledgments** This Paper has been prepared in the framework of the IT4Innovations Centre of Excellence project, reg. no. CZ.1.05/1.1.00/02.0070 supported by Operational Programme ‘Research and Development for Innovations’ funded by Structural Funds of the European Union and the Government of Czech Republic.

## References

1. Abraham, N.S., Young, J.M., Solomon, M.J.: Meta-analysis of short-term outcomes after laparoscopic resection for colorectal cancer. *Br. J. Surg.* **91**, 1111–1114 (2004)
2. Bonjer, H.J., et al.: Laparoscopically assisted vs open colectomy for colon cancer. *Arch. Surg.* **142**, 298–303 (2007)
3. Breslow, N.E.: Covariance analysis of censored survival data. *Biometrics* **30**, 89–100 (1974)
4. Cox, D.R.: Regression models and life-tables (with discussion). *J. R. Stat. Soc. Ser. B* **34**, 187–220 (1972)
5. Cox, D.R.: Partial likelihood. *Biometrics* **62**, 269–276 (1975)
6. Efron, B.: The efficiency of Coxs likelihood function for censored data. *JASA* **72**, 557–565 (1977)
7. Fleshman, J., et al.: Laparoscopic colectomy for cancer is not inferior to open surgery based on 5-year data from the COST study group trial. *Ann. Surg.* **246**, 655–662 (2007)
8. Hosmer, D.W., Lemeshow, S., May, S.: Applied Survival Analysis. Regression Modeling of Time-to-Event Data. Wiley, Hoboken (2008)
9. Jackson, T.D., et al.: Laparoscopic versus open resection for colorectal cancer: a meta-analysis of oncologic outcomes. *J. Am. Coll. Surg.* **204**, 439–446 (2007)
10. Kitano, S., et al.: A multicenter study on laparoscopic surgery for colorectal cancer in Japan. *Surg. Endosc.* **20**, 1348–1352 (2006)
11. Royston, P., Altman, D.G.: Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling (with discussion). *Appl. Stat.* **43**, 429–467 (1994)
12. Schoenfeld, D.: Partial residuals for the proportional hazards regression model. *Biometrics* **69**, 239–241 (1982)
13. Schwenk, W., et al.: CShort term benefits for laparoscopic colorectal resection. *Cochrane Database of Syst Rev* (2) (2005). doi:[10.1002/14651858.CD003145.pub2](https://doi.org/10.1002/14651858.CD003145.pub2)
14. Veldkamp, R., et al.: Laparoscopic resection of colon cancer: consensus of the European Association of Endoscopic Surgery. *Surg. Endosc.* **18**, 1163–1185 (2004)

# Profiling of Software Requirements for the Pharmaceutical Enterprise Manufacturing Execution System

Vyacheslav Kharchenko, Oleksandr Gordieiev and Alina Fedoseeva

**Abstract** This chapter is devoted to analysis of existing Pharmaceutical Enterprise (PE) Manufacturing Execution System (MES) requirements. A special technique grounded on so-called Semantic Facet-Hierarchical Structures (SFHS) is suggested for the requirements profiling, development of general and particular PE MES requirements profiles taking into account software requirements of ISO/IEC 25010 “Systems and software—Systems and software Quality Requirements and Evaluation (SQuaRE)—System and software quality models” and GAMP “Good Automated Manufacturing Practice” standards. The chapter has the following structure: the Sect. 1 introduces into problem, the Sect. 2 describes features of PE and MES and corresponding standards, the Sect. 3 is dedicated to the SFHS-based technique analysis, the Sect. 4 presents case study and last one concludes and discusses future works.

## 1 Introduction

The modern society makes high requirements to computer-based systems and software in pharmaceutical enterprise. Thus the technological process of drugs production is closely related with patients risks, special attention should be to profiling of the manufacturing execution system requirements. Manufacturing execution system (MES)—specialized software used in technological process of drugs production.

---

V. Kharchenko (✉)

National Aerospace University “KhAI”, 17 Chkalova Street, Kharkiv, Ukraine  
e-mail: V.Kharchenko@khai.edu

O. Gordieiev

University of Banking of the National Bank of Ukraine, 1 Andriyvska Street,  
Kyiv, Ukraine  
e-mail: alex.gordeyev@gmail.com

A. Fedoseeva

National University of Pharmacy, 53 Pushkinskaya Street, Kharkiv, Ukraine  
e-mail: fedosaa@ukr.net

There are serious requirements to reliability and safety of the instrumentation and control systems (I&Cs) and MESs as a specific type of such systems for pharmaceutical enterprise. Pharmaceutical enterprise (PE) is an industrial enterprise for drugs production. International industry software products such as Sparta Systems' TrackWise Enterprise Quality Management Software, AssurX, MARG and others, guarantee support the workflow management, tracking and regulatory reporting for all critical technological operations in process of drugs production [1–3]. Drugs—medicinal form in prepackaged, packaged and marked form according to the requirements for normative-technical documentation with the specified shelf life.

Applications of MESs are regulated by industrial standards ANSI/ISA88 and ANSI/ISA95. They define the models for batch process control [4] and the models of manufacturing operations management [5].

MESA [6] describes typical functions for PE, which covers all the operations of technological processes of drugs production. Designing of PE MES should be conducted by comprehensive assessing the quality and security of software [7, 8].

This chapter is devoted to analysis of existing PE MES requirements. A special technique grounded on so-called Semantic Facet-Hierarchical Structures (SFHS) is suggested for the requirements profiling, development of general and particular PE MES requirements profiles. The chapter has the following structure: the Sect. 2 describes features of PE and MES and corresponding standards, the Sect. 3 is dedicated to the SFHS-based technique analysis, the Sect. 4 presents case study and last one concludes and discusses future works.

## 2 An Analysis of Pharmaceutical Enterprise MES Related Standards

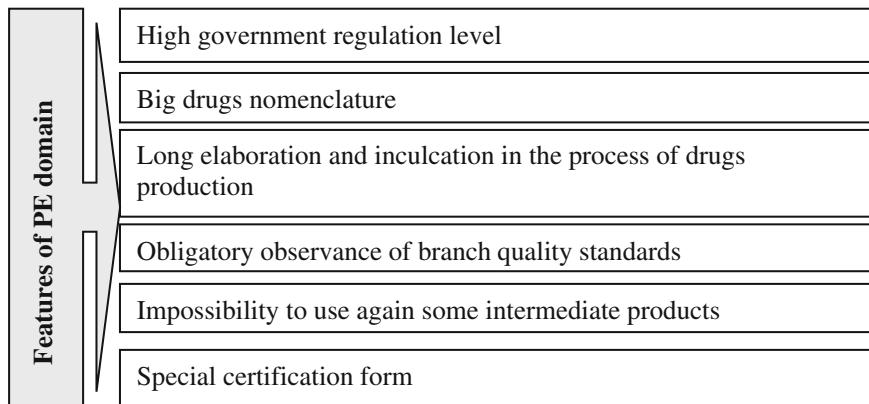
### 2.1 Features of PE MES

Nowadays more and more attention is given to application information technologies in PE. This is primarily due to the development of technological possibilities of pharmaceutical branch and also using of specialized software during technological process of drugs production.

These factors lead to the need to develop normative profiles for PE software. The software applied in PE domain should be aimed at reducing the negative effects for patients because of the failure or wrong usage of the software. The pharmaceutical industry faces the task of meeting all FDA [9] regulatory compliance requirements.

The technological process of drugs refers to:

- process-costing manufacturing type is sequence of technological stages (operations) each of which cannot be interrupted at any moment. PE are usually produced some accompanying or by-products besides drugs. Some technological stages (operations) in drugs manufacturing can be repeated recursively;



**Fig. 1** Features of PE domain

- batch manufacturing type is application of thousands of recipes, ten thousands components in technological process of drugs, whose interaction is necessary to monitor (separately and ready-made drugs). Special recipes in a small amount are often used.

The technological process of drugs production must be implanted according to the technological regulation of drugs production and GMP rules.

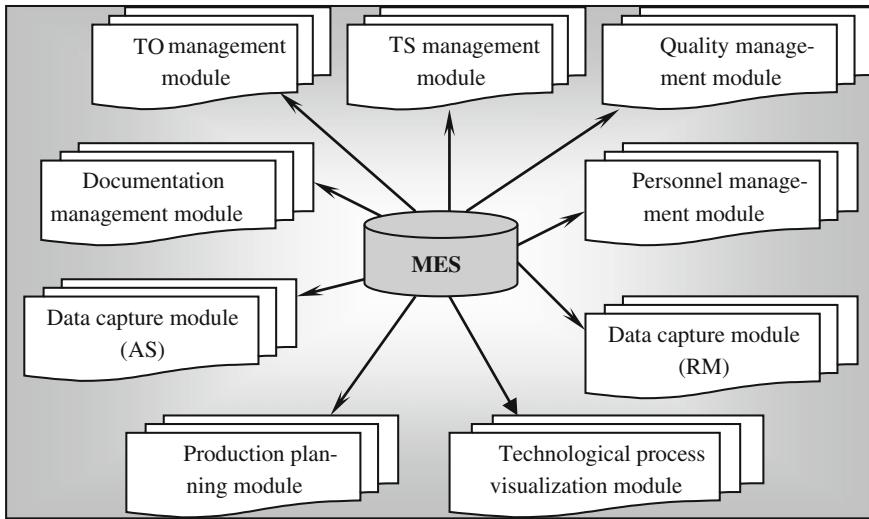
Features of PE domain represented at Fig. 1.

MES for PE embody:

- organization and detail realization electronic document flow in PE;
- tools to control the quality of technological stage, technological operation and drugs or semi-manufactured goods. Technological stage (TS) is a complex of technological operations resulting in the information of intermediate product (at the end-product final step), which is determinate quantitatively and characterized by quality. Technological operation (TO) is a part of the technological process of drugs links one of the using equipment;
- control of industrial stocks and batches output;
- confirmation of drugs quality with electronic passport of drugs batch record, which contains information about technological process of drugs, including the data about operators and used materials. Electronic passport of drugs batch record (EBR) is an information about technological process including operation and used RM data;
- validation and revalidation of technological process of drugs production;
- optimization of technological process of drugs production.

The most important principle of MES project is modularity (Fig. 2), because the technological process of drugs production refers to Process-Costing manufacturing type.

For the correct and safe functioning of software in the critical PE MES it is required to analyze the existing international normative documents (standards, guidelines,



**Fig. 2** Modular structure MES for PE

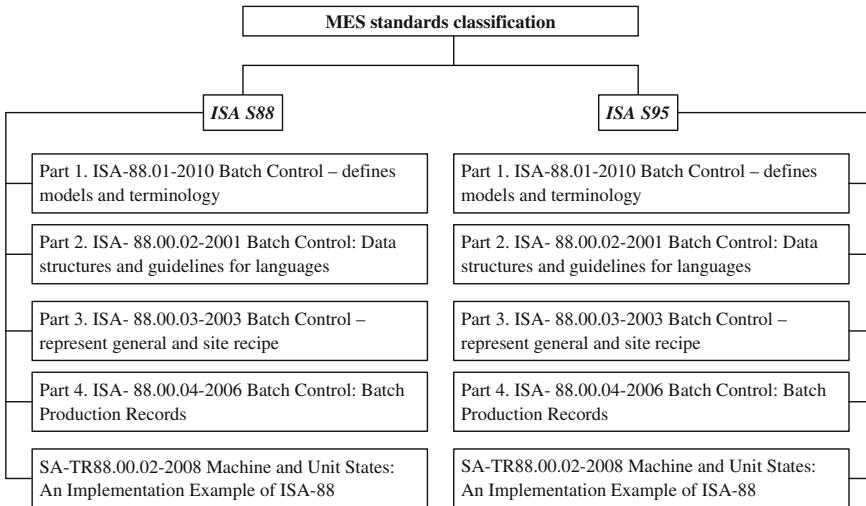
directives and technical reports) in all domains related to the technological processes of drugs production.

A lot of normative documents governing the technological process of drugs production with using special software can be divided into two groups: pharmaceutical enterprise standards and software engineering standards. The first group of standards includes regulations under which the technological process of drugs production, validation, revalidation and guaranteeing drugs quality are carried out.

The second group of standards includes regulations under which development, inculcation and accompaniment software processes are carried out. Today these two groups of standards are separately, but there is a need for general profile standards creation, which should be followed in technological process of drugs production in PE with MES.

## 2.2 PE MES Standards

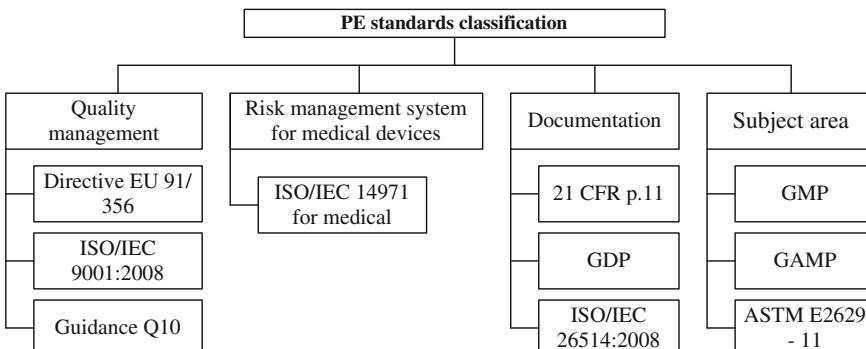
MES development for PE is regulated by normative acts of MESA International (Manufacturing Enterprise Solutions Association), and also of standards international Society of Automation (ISA) [10]: ISA S95 “Enterprise Control System Integration” and ISA S88 “General and Site Recipe Models and Representation”. MES standards were developed with the participation of leading automation systems manufacturers (Siemens, Rockwell Automation, Schneider Electric) and IT-companies (IBM, Oracle, SAP AG).

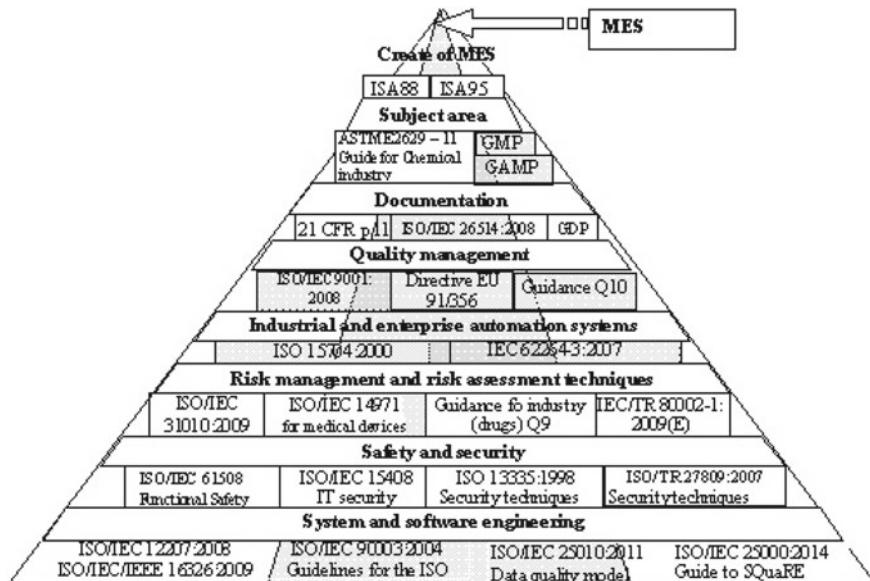
**Fig. 3** MES standards classification

ISA S88 and ISA95 standards describe MES components development. Special attention in these standards is paid implementation features of the technological processes at enterprises.

According the specifications of ISA S95 standard the pharmaceutically relevant data are to be handled by the PE MES. ISA S88 standard contains models and terminology for controlling batch processes in PE. The standards differ in terms of their purpose, which means manufacturing companies will increasingly make use of both standards.

Pharmaceutical enterprise standards and software engineering standards MES standards classification is represented on Figs. 3 and 4. All PE standards are divided into groups. Each group consists of the main categorized standards: Quality

**Fig. 4** PE standards classification



**Fig. 5** Profile forming base for PE MES

management, Risk management system for medical devices, Documentation and Subject area. Each category corresponds to main requirements and guidelines.

### 2.3 Profile-Forming Base for PE MES

High-quality PE MES development requires a thorough analysis of the normative base of branches such as pharmaceutical manufacturing, validation of automated system in pharmaceutical enterprise, system and software engineering, documentation, risk management and assessment techniques. There are its own regulatory acts in each of these brunches.

International standards, guidelines, directives and technical reports, selected on the basis of the criteria [11] and listed above brunches were collected into profile-forming standards base. Figure 5 presents profile-forming base for PE MES.

## 3 SFHS-Based Technique for the Requirements Profiling

To develop MES software profile the standard requirements should be analyzed, selected and added as needed taking into account features of PE. For that SFHS-based technique is used.

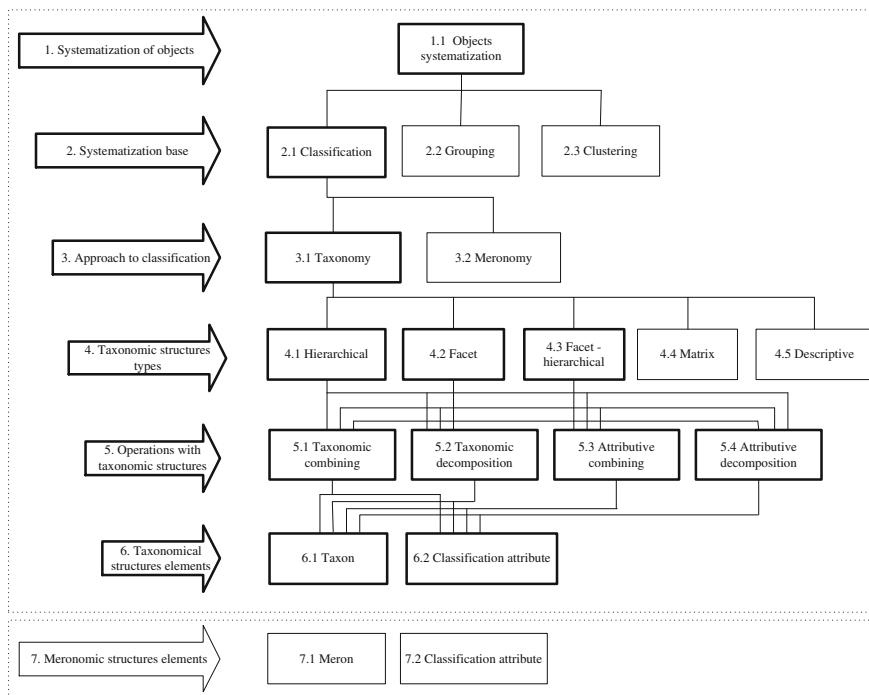
### 3.1 Taxonomy Terms

Connections between concepts and represent in the facet-hierarchical structure form have been established (Fig. 6). Classification attribute identified by a number, and taxons identified by two numbers: the first is the number of classification attribute, and the second is the number of taxons.

Depending on the base (2) three types of ordering have been selected: the systematization of discrete (classification 2.1), continuous (grouping, 2.2) and self-organizing or procedurally interacting (clustering 2.3) of objects.

Classification is twofold. It is expressed in the approach to the classification of objects (3). There are two approaches to classification: taxonomic (taxonomy) (3.1), which is based on the decomposition of objects into classes, is characterized by more or less resemblance classification sets; meronomic (meronomy) (3.2) is based on the partition of objects into parts that have some common attribute (Fig. 6).

Following types of taxonomic structures (4) have been selected: hierarchical (4.1), a facet (4.2), facet-hierarchical (4.3) (mixed species), the matrix (4.4), descriptive (4.5). Degenerate cases of topologies are descriptive when difficult to determine of objects topology and matrix—when the topology is complex. The following operations for taxonomic structure (5) are defined:



**Fig. 6** Definitions facet-hierarchical structure

- taxonomic combining is combining of two taxonomic structures in an immutable set of classification attributes (combining in width) (5.1);
- taxonomic decomposition is partition on the basis of establishing equivalence between the elements of the taxon sets ( $T$ ) and the choice criteria ( $C$ ) (5.2);
- attributive combining is combining of two taxonomic structures at an increase of elements in set of classification attributes (5.3);
- attributive decomposition is decomposition of taxonomic structure on basis of establishment of equivalents between  $A$  and  $C$  (5.4).

Taxonomy is forming of external system out of objects. Meronomy represents them as internal structures. Taxonomic structure consists of (6) taxons (6.1) and classification attributes (6.2). Meronomic structure consists of merons (7.1) and classification attributes (7.2). Taxon is a set of objects, which are combined by some general classification attribute. Meron is a set of parts, belonging to these objects which have the same general classification attribute.

This work is oriented on formal operations with discrete objects. In this connection in future will consider one from systematization types—classification (taxonomy).

### **3.2 Variants of Taxonomic Structures**

Taxonomic structures are represented by two variants of description: set-theoretic and set-matrix. Taxonomic structure is set, consisting of three elements:  $S \{A, T, \Psi\}$ , where  $A$ —a set of classification attributes  $A = \{A_i\}_{i=1}^n$ ;  $T$ —a set of taxons  $T = \{t_i\}_{i=1}^n$ ;  $\Psi$ —relationship between elements  $A_i \in A \nparallel T_i \in T, A\Psi T$ .

If the sequence of taxons or classification attributes is important, then this ordered set or tuple (marked as  $\langle \dots \rangle$ ), otherwise, taxons or classification attributes in taxonomic structures is unordered set usually (marked as  $\{\dots\}$ ). Taxons are unordered set, but for classification attributes sequence is important. To represent hierarchical structures ( $S_H$ ) we should use the following record:  $S_H = \{A_H, T_H, \Psi_H\}$ , where  $A_H$ —a set of classification attributes in hierarchical structure;  $T_H$ —a set of taxons in hierarchical structure;  $\Psi_H$ —relationship between elements of sets  $T_H$  and  $A_H$ , in which connection is:

- $A_H = \langle A_{Hi} \rangle_{i=1}^n$ —a set (tuple) of classification attributes in hierarchical structure;
- $T_H = \langle \dots \langle t_{j_1\dots k} \dots \rangle \dots \rangle$ —a set (tuple) of taxons in hierarchical structure—inserted tuples, which provide possibility of description of subordination hierarchies;
- $\Delta T = \langle t_i \rangle$ —subset (tuple) of taxons in hierarchical structure, which appropriate to classification attribute;
- $\Psi : \forall A_{Ii} \leftrightarrow \Delta T_{Ii} \subset T_I$ —relation between taxonomic classification attributes and a variety of taxons in a hierarchical structure.

To represent the facet structures ( $S_F$ ) we should use the following record:  $S_F = \{A_F, T_F, \Psi_F\}$ , where  $A_F$ —a set of classification attributes in facet structure;  $T_F$ —a

set of taxons in facet structure;  $\Psi_F$ —relation between elements of sets  $T_F$  and  $A_F$ , in which connection is:

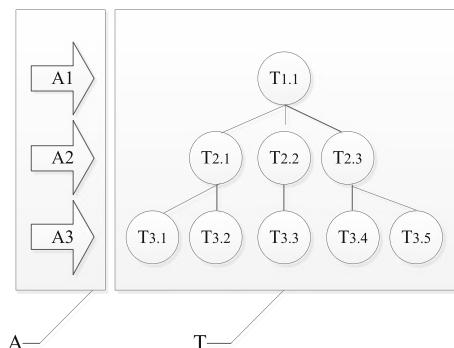
- $A_F = \{A_{Fi}\}_{i=1}^n$ —a set of classification attributes in facet structure;
- $T_F = \bigcup_{i=1}^n \Delta T_{Fi}$ —a set of taxons in facet structure, which consists of combination of each subset  $\Delta T_{Fi}$ , corresponding to classification attribute (facet row);
- $\Delta T_{Fi} = \{t_{ij}\}_{j=1}^{n_i}$ —a set of taxons, corresponding to classification attribute (facet row);
- $\Delta T_{Fi} \cap \Delta T_{Fj} = \emptyset$ —taxons subset, corresponding to different disjoint classification attributes (facet rows);
- $\Psi : \forall A_{Fi} \leftrightarrow \Delta T_{Fi} \subset T_F$ —relation between taxonomic attributes and set of taxons in facet structure.

Set-matrix description of taxonomic structure represents the set of classification attributes (A) and the set of taxons (T). The relations between elements of these sets are represented by contiguity and compliance matrixes. Contiguity matrix for hierarchical structures is needed to represent logical relations between elements of set T, i.e. for description of topology (logical relations between taxons). The compliance matrix is used to represent the compliance between elements of sets T and A. Example of representing hierarchical structures according to the model of description is depicted on Fig. 7.

Hierarchical structure (Fig. 7), consisted of sets  $A = \{A_1, A_2, A_3\}$  and  $T = \{T_{1.1}, T_{2.1}, T_{2.2}, T_{2.3}, T_{3.1}, T_{3.2}, T_{3.3}, T_{3.4}, T_{3.5}\}$  in the form of the contiguity matrix (Table 1) and the compliance matrix (Table 2) have been represented. The contiguity matrix is formed as following: “one” is an equivalent of logical relation between taxons, “zero” its absence. Compliance matrix formed as follows: “one” is compliance taxon with classification attribute in hierarchical structure, “zero” if compliance is absence.

For facet structures the compliance of elements T and A is determined the compliance matrix only, because logical relations of elements are absent in set T (Fig. 8). The facet structure (Fig. 8), consisted of sets  $A = \{A_1, A_2, A_3\}$  and  $T = \{T_{1.1}, T_{1.2}\}$ ,

**Fig. 7** Hierarchical structure

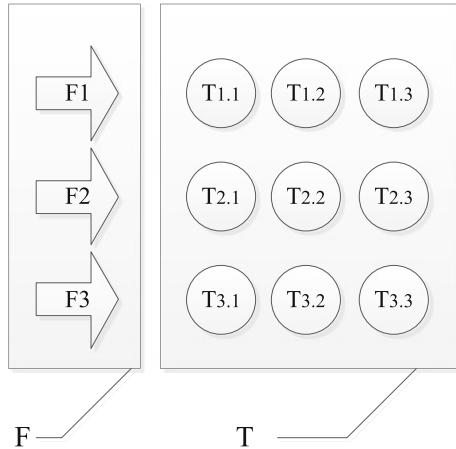


**Table 1** Contiguity matrix

	T <sub>1.1</sub>	T <sub>2.1</sub>	T <sub>2.2</sub>	T <sub>2.3</sub>	T <sub>3.1</sub>	T <sub>3.2</sub>	T <sub>3.3</sub>	T <sub>3.4</sub>	T <sub>3.5</sub>
T <sub>1.1</sub>	0	1	1	1	0	0	0	0	0
T <sub>2.1</sub>	1	0	0	0	1	1	0	0	0
T <sub>2.2</sub>	1	0	0	0	0	0	1	0	0
T <sub>2.3</sub>	1	0	0	0	0	0	0	1	1
T <sub>3.1</sub>	0	1	0	0	0	0	0	0	0
T <sub>3.2</sub>	0	1	0	0	0	0	0	0	0
T <sub>3.3</sub>	0	0	1	0	0	0	0	0	0
T <sub>3.4</sub>	0	0	0	1	0	0	0	0	0
T <sub>3.5</sub>	0	0	0	1	0	0	0	0	0

**Table 2** Contiguity matrix

	T <sub>1.1</sub>	T <sub>2.1</sub>	T <sub>2.2</sub>	T <sub>2.3</sub>	T <sub>3.1</sub>	T <sub>3.2</sub>	T <sub>3.3</sub>	T <sub>3.4</sub>	T <sub>3.5</sub>
A(F) <sub>1</sub>	1	0	0	0	0	0	0	0	0
A(F) <sub>2</sub>	0	1	1	1	0	0	0	0	0
A(F) <sub>3</sub>	1	0	0	0	1	1	1	1	1

**Fig. 8** Facet structure

T<sub>1.3</sub>, T<sub>2.1</sub>, T<sub>2.2</sub>, T<sub>2.3</sub>, T<sub>3.1</sub>, T<sub>3.2</sub>, T<sub>3.3</sub>} in the form of compliance matrix has been represented. It is identical to compliance matrix for hierarchical structure.

### 3.3 Operation of Combining

The basis of operation of combining is formation of the generalized taxonomic structure of original structures. The operation of combining consists of combining elements of sets T and A together.

To combine taxonomic structures it was suggested to use the following logical operations:  $\cup$ —combining of sets;  $\cap$ —intersection of sets;  $\subset$ —strict inclusion of one set into another;  $\vec{\cup}$ —taxonomic combining of taxons sets;  $\cup \downarrow$ —attributive combining of taxons sets;  $\cup_{FHS}$ —combining hierarchies into facet-hierarchical structure. The operations  $\cup_{FHS}$ ,  $\cup \downarrow$ ,  $\vec{\cup}$  were introduced in first for the best representation variants of taxonomic structure combining.

Taxonomic combining types for hierarchical structure have been considered:

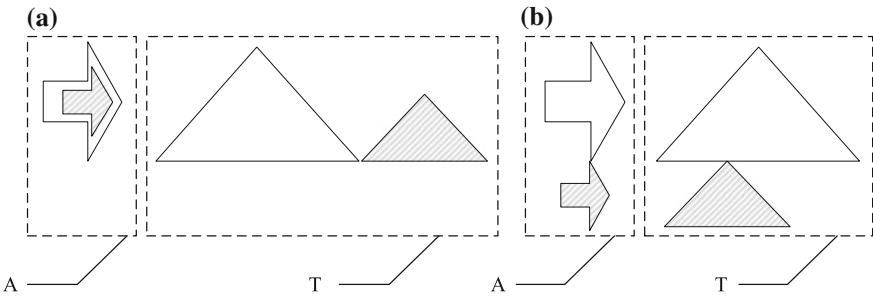
1. Taxonomic combining for hierarchical structures consists of combination of elements of taxonomic sets ( $T = T_1 \vec{\cup} T_2$ ), the set of attribute must remain unchanged ( $A = A_1$ ). In which the connection combined with a set of taxons can be without intersection of elements ( $T_1 \cap T_2 = \emptyset$ ) or with intersection of elements ( $T_1 \cap T_2 \neq \emptyset$ ) (Fig. 9a).
2. Attributive combining of hierarchical structures consists of combination of taxons sets elements ( $T = T_1 \cup \downarrow T_2$ ) and classification attributes elements sets ( $A = A_1 \cup A_2$ ). In which the connection between combination of set of taxonomic classification attributes can be without intersection of elements ( $T_1 \cap T_2 = \emptyset$ ,  $A_1 \cap A_2 = \emptyset$ ) or with intersection of elements ( $T_1 \cap T_2 \neq \emptyset$ ,  $A_1 \cap A_2 \neq \emptyset$ ) (Fig. 9b).
3. Combining of hierarchical structures to facet-hierarchical structure. It is degenerated variant of combining. It is a case when the establishment of relations is impossible between sets A and T. Combining of hierarchical structures is realized due to the formation of two facet structures. Each taxonomic structure will be separated by the facet structure. In result each hierarchical structure will be separated facet structure.

The next considered types of taxonomic combining operations for facet structures are:

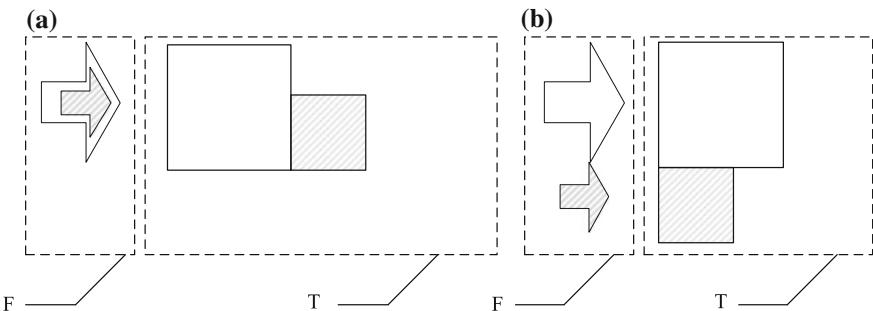
1. Taxonomic combination for facet structures consists of combining of elements of taxonomic sets ( $T = T_1 \vec{\cup} T_2$ ), set of attribute must remain unchanged ( $F = F_1$ ). In which the connection of combination of set of taxons can be without intersection of elements ( $T_1 \cap T_2 = \emptyset$ ) or with intersection of elements ( $T_1 \cap T_2 \neq \emptyset$ ) (Fig. 10a).
2. Attributive combining of facet structures consists of combining the taxons sets elements ( $T = T_1 \cup \downarrow T_2$ ) and sets of classification attributes elements ( $A = F_1 \cup F_2$ ). Combining of set of taxonomic classification attributes can be realized without intersection of elements ( $T_1 \cap T_2 = \emptyset$ ,  $F_1 \cap F_2 = \emptyset$ ) or with intersection of elements ( $T_1 \cap T_2 \neq \emptyset$ ,  $F_1 \cap F_2 \neq \emptyset$ ) (Fig. 10b).

### ***3.4 Forming of Sets of Variants Taxonomic Structures Combine***

Forming of set of variants combining of taxonomic structures is developed on the basis of analysis compliance of logical operations. As a result of analysis the



**Fig. 9** Schematic representation of combining operations types for hierarchical structures



**Fig. 10** Schematic representation of combining operations types for facet structures

taxonomic structures are determined by the following elements of operation of combining:  $\overset{\rightarrow}{\cup}$ —taxonomic combining;  $\cup \downarrow$ —attributive combining;  $\cup_{FHS}$ —combining in FHS;  $T_1 \cap T_2$ —intersection of taxons sets;  $A_1 \cap A_2$ —intersection of classification attributes.

The combinations of these elements are defined as a set of variants of combining taxonomic structures. In result of simple running over of elements we received full set of variants of combining taxonomic structures.

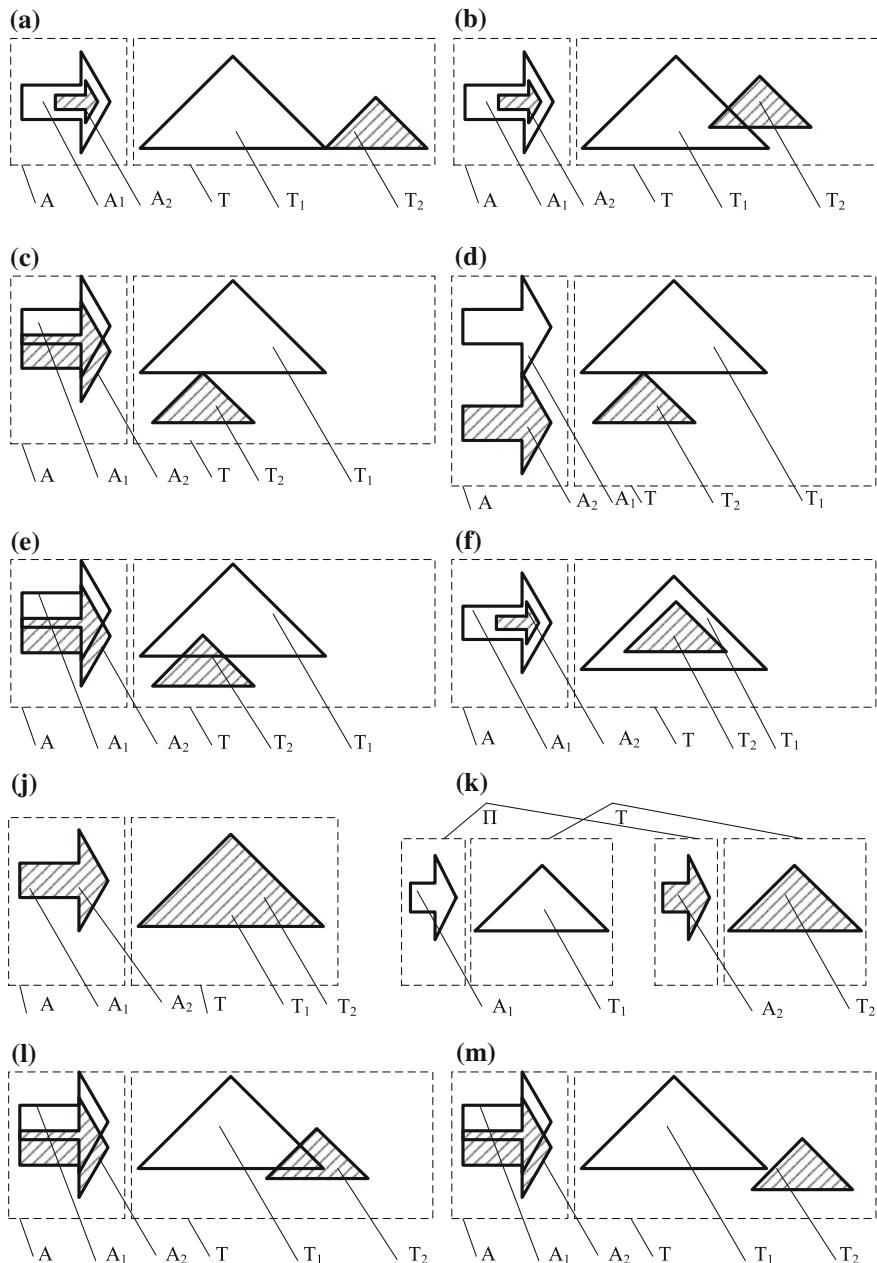
Two following additional variants of taxonomic structures combination exist: combination of hierarchical structures with inclusion of taxons set and with equality of taxons set. Description of variants of combining for hierarchical structures (Table 3) and for facet structures (Table 4) have been determined. Variants of combining for hierarchical structures are conceptually represented on Fig. 11 and for facet hierarchical structures are conceptually represented on Fig. 12.

**Table 3** Formal description of hierarchical structures combined variants

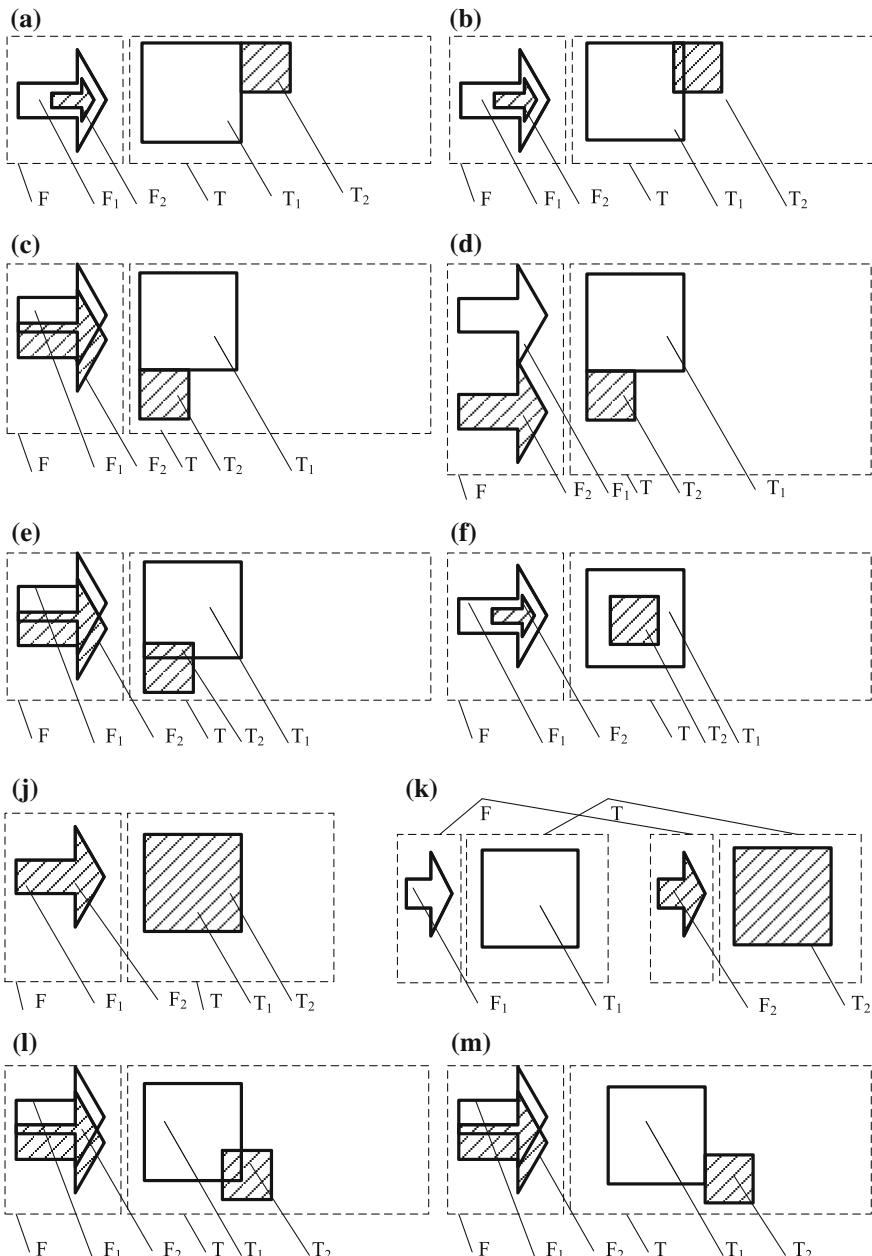
1	Taxonomic combining of hierarchical structures without intersection of taxons sets (Fig. 11a)	$(2.1)$	2	Taxonomic combining of hierarchical structures with intersection of taxons sets (Fig. 11b)	$(2.2)$
	$\begin{cases} A = A_1; \\ T = T_1 \dot{\cup} T_2, T_1 \cap T_2 = \emptyset. \end{cases}$			$\begin{cases} A = A_1; \\ T = T_1 \dot{\cup} T_2, T_1 \cap T_2 \neq \emptyset. \end{cases}$	
3	Attributive combining of hierarchical structures without intersection of taxons sets and with intersection of hierarchies attributes (Fig. 11c)	$(2.3)$	4	Attributive combining of hierarchical structures without intersection of taxons sets and without intersection of hierarchies attributes (Fig. 11d)	$(2.4)$
	$\begin{cases} A = A_1 \cup A_2, A_1 \cap A_2 \neq \emptyset; \\ T = T_1 \cup \downarrow T_2, T_1 \cap T_2 = \emptyset. \end{cases}$			$\begin{cases} A = A_1 \cup A_2, A_1 \cap A_2 = \emptyset; \\ T = T_1 \cup \downarrow T_2, T_1 \cap T_2 = \emptyset. \end{cases}$	
5	Attributive combining of hierarchical structures with intersection of taxons sets (Fig. 11e)	$(2.5)$	6	Combining of hierarchical structures with inclusion of taxons sets (Fig. 11f)	$(2.6)$
	$\begin{cases} A = A_1 \cup A_2, A_1 \cap A_2 \neq \emptyset; \\ T = T_1 \cup \downarrow T_2, T_1 \cap T_2 \neq \emptyset. \end{cases}$			$\begin{cases} A = A_1; \\ T = T_1. \end{cases}$	
7	Combining of hierarchical structures with equality of taxons sets (Fig. 11j)	$(2.7)$	8	Combining of hierarchical structures in facet-hierarchical structure (Fig. 11k)	$(2.8)$
	$\begin{cases} A = A_1 = A_2; \\ T = T_1 = T_2. \end{cases}$			$\begin{cases} A = A_1 \cup_{FHS} A_2, A_1 \cap A_2 = \emptyset; \\ T = T_1 \cup_{FHS} T_2, T_1 \cap T_2 = \emptyset. \end{cases}$	
9	Taxonomic combining and attributive combining of hierarchical structures with intersection of taxons sets (Fig. 11l)	$(2.9)$	10	Taxonomic combining and attributive combining of hierarchical structures without intersection of taxons sets (Fig. 11m)	$(2.10)$
	$\begin{cases} A = A_1 \cup A_2, A_1 \cap A_2 \neq \emptyset; \\ T = (T_1 \dot{\cup} T_2) \cup (T_1 \cup \downarrow T_2), T_1 \cap T_2 \neq \emptyset. \end{cases}$			$\begin{cases} A = A_1 \cup A_2, A_1 \cap A_2 \neq \emptyset; \\ T = (T_1 \dot{\cup} T_2) \cup (T_1 \cup \downarrow T_2), T_1 \cap T_2 = \emptyset. \end{cases}$	

**Table 4** Formal description of facet structures combined variants

1	Taxonomic combining of facet structures without intersection of taxons sets (Fig. 12a)	$(2.11)$	2	Taxonomic combining of facet structures with intersection of taxons sets (Fig. 12b)	$(2.12)$
	$\begin{cases} F = F_1; \\ T = T_1 \dot{\cup} T_2, T_1 \cap T_2 = \emptyset. \end{cases}$			$\begin{cases} F = F_1; \\ T = T_1 \dot{\cup} T_2, T_1 \cap T_2 \neq \emptyset. \end{cases}$	
3	Attributive combining of facet structures without intersection of taxons sets and with intersection of classification attributes (Fig. 12c)	$(2.13)$	4	Attributive combining of facet structures without intersection of taxons sets and without intersection of classification attributes (Fig. 12d)	$(2.14)$
	$\begin{cases} F = F_1 \cup F_2, F_1 \cap F_2 \neq \emptyset; \\ T = T_1 \cup \downarrow T_2, T_1 \cap T_2 = \emptyset. \end{cases}$			$\begin{cases} F = F_1 \cup F_2, F_1 \cap F_2 = \emptyset; \\ T = T_1 \cup \downarrow T_2, T_1 \cap T_2 = \emptyset. \end{cases}$	
5	Attributive combining of facet structures with intersection of taxons sets (Fig. 12e)	$(2.15)$	6	Combining of facet structures with inclusion of taxons sets (Fig. 12f)	$(2.16)$
	$\begin{cases} F = F_1 \cup F_2, F_1 \cap F_2 \neq \emptyset; \\ T = T_1 \cup \downarrow T_2, T_1 \cap T_2 \neq \emptyset. \end{cases}$			$\begin{cases} F = F_1; \\ T = T_1. \end{cases}$	
7	Combining of facet structures with equality of taxons sets (Fig. 12j)	$(2.17)$	8	Combining of facet structures in facet-hierarchical structure (Fig. 12k)	$(2.18)$
	$\begin{cases} F = F_1 = F_3; \\ T = T_1 = T_3. \end{cases}$			$\begin{cases} F = F_1 \cup_{FHS} F_2, F_1 \cap F_2 = \emptyset; \\ T = T_1 \cup_{FHS} T_2, T_1 \cap T_2 = \emptyset. \end{cases}$	
9	Taxonomic combining and attributive combining of facet structures with intersection of taxons sets (Fig. 12l)	$(2.19)$	10	Taxonomic combining and attributive combining of facet structures without intersection of taxons sets (Fig. 12m)	$(2.20)$
	$\begin{cases} F = F_1 \cup F_2, F_1 \cap F_2 \neq \emptyset; \\ T = (T_1 \dot{\cup} T_2) \cup (T_1 \cup \downarrow T_2), T_1 \cap T_2 \neq \emptyset. \end{cases}$			$\begin{cases} F = F_1 \cup F_2, F_1 \cap F_2 \neq \emptyset; \\ T = (T_1 \dot{\cup} T_2) \cup (T_1 \cup \downarrow T_2), T_1 \cap T_2 = \emptyset. \end{cases}$	



**Fig. 11** Graphic representation of hierarchical structures variants combining



**Fig. 12** Graphical representation of variants of facet structures combining

### ***3.5 Operation of Decomposition***

Operation of decomposition of taxonomic structures consists of decomposition of sets A and T into subsets. Some terms, which are needed for description of taxonomic structures decomposition operation have been introduced:  $T_{res}$ —a set of taxons of residual taxonomic structure;  $T_{res.i}$ —initial set of taxons of residual taxonomic structure;  $T_{res.red}$ —reduced set of taxons of residual taxonomic structure;  $T_{des}$ —a set of taxons of desired taxonomic structure;  $T_{des.i}$ —initial set of taxons of desired taxonomic structure;  $T_{des.red}$ —reduced set of taxons of desired taxonomic structure;  $T_{strt}$ —a set of taxons of starting taxonomic structure; initial taxonomic structure—structure before execution over its operation of decomposition; desired taxonomic structure—structure after execution of operation of decomposition; residual taxonomic structure—addition of structure to of initial taxonomic structure.

Following types of hierarchical structures decomposition (attributive and taxonomic decomposition) have considered. Attributive decomposition of hierarchical structures consists in choice of hierarchical structure, proceeding from semantic compliance between elements of sets A and C (set of choice criteria). Attributive decomposition cannot be used to hierarchical structures, because set A in hierarchical structures is indivisible and fixed.

Taxonomic decomposition of hierarchical structures consists of forming of set  $T_{des} \subset T$  for hierarchical structures. Set  $T_{des}$  is determined on the basis of semantic compliance between elements of sets T and C. Semantic equivalents are determined by expert. We represent taxonomic decomposition by the following stages.

Stage 1. Forming of initial data: hierarchical structure by represented of contiguity matrix, compliance matrix of elements of sets T and A, set C.

Stage 2. Establishment of equivalence between elements of taxons set and choice criteria. Initial data is given in sets of T and C. Semantic equivalents between elements of sets T and C are determined by expert.

Stage 3. Receive of sets  $T_{des}$  and  $T_{res}$ . For this:

- set  $T_{des.i}$  has been defined. Elements of set is elements of set  $T_{des}$ , with which was established compliance of elements of set C.  $T_{des.i} = \{t_i, t_{i+1}, \dots, t_n\}$ ;
- set  $T_{des.red}$  has been defined;
- set  $T_{des}$ , as a result of combining  $T_{des.i}$  and  $T_{des.red}$  has been defined:

$$T_{des} = T_{des.i} \cup T_{des.red}, T_{des.i} \cap T_{des.red} \neq \emptyset;$$

- set  $T_{res}$ , set  $T_{des.red}$  has defined. For this we define  $T_{des.i..}$ . This is a set is symmetric difference of  $T_{strt}$  and  $T_{des}$ ,  $T_{res.i} = T_{strt} \setminus T_{des}$ ;
- set  $T_{res.red}$  set  $T_{des.red}$  have been defined;
- set  $T_{res}$  on basis combining  $T_{res.i}$  and  $T_{res.red}$  have been formed:

$$T_{res} = T_{res.i} \cup T_{res.red}, T_{res.i} \cap T_{res.red} \neq \emptyset.$$

Stage 4. Correction of completeness set  $T_{des}$  for classification attributes. This is stage typical only for attributive decomposition of facet structures. Expert can precise of set  $T_{des}$  for classification attributes.

The following types of decomposition facet structures: attributive decomposition and taxonomic decomposition have been considered.

Attributive decomposition of facet structures consists of choice of facet structure, proceeding from semantic compliance of elements of sets A and C.

This type of decomposition included in all stages, which description in operation of taxonomic decomposition of hierarchical structure is:

- on the second stage determined semantic compliance between elements of sets A and C;
- on the fourth stage produced correction of completeness sets  $T_{des}$  for classification attributes.

Taxonomic decomposition of hierarchical structures consists of choice of facet structure, i.e. set  $T_{des}$ , proceeding from semantic compliance of elements sets  $T \nparallel C$ .

This type of decomposition included all stages, which description in the operation of taxonomic decomposition of hierarchical structures, excluding stage №4.

### ***3.6 Semantic Issues***

In this work the operation of combining and decomposition is considered in the context of structures (taxonomic) excluding their semantic context. The analysis of semantic context of taxonomic structures with operations of combining and decomposition will give the following opportunities:

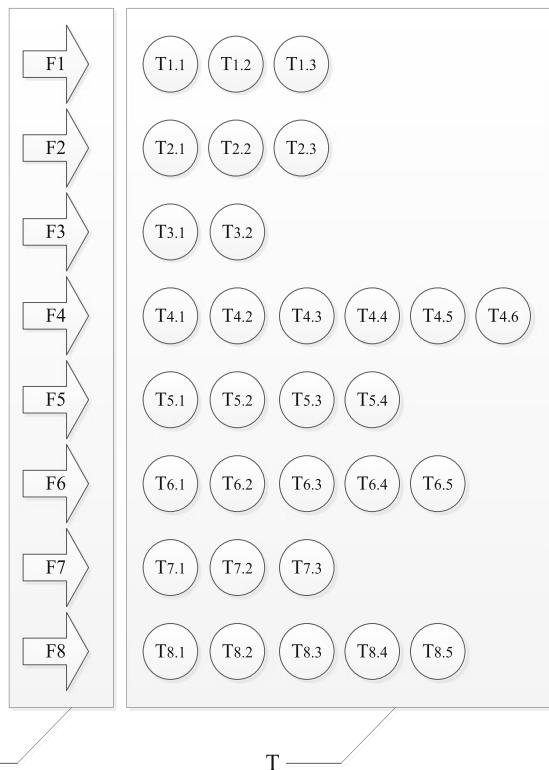
- firstly, consider taxonomic structures not only on structural level, but and on semantic level of individual nodes;
- secondly, take into account features of semantic structures for their formal description (representation);
- in the third, determine limitations at use of operations of combining and decomposition on structural level and semantic level.

### ***3.7 Profiling Sequence***

Basic stages of profiling have considered.

Stage 1. Transformation initial information in FHS. The goal of this stage is transformation of information in facet hierarchical structure. Dedication such operations in separate stage connected with absence of formalization in initial information about requirements. Such work demand of considerable time expense.

**Fig. 13** Facet structure for ISO/IEC 25010



Stage 2. Representation of FHS in set-matrix look. The basic task of this stage is representation of FHS in look of matrices (contiguity matrix and compliance matrix for hierarchical structures, compliance matrix for facet structures) and establish of variant of combining or decomposition of FHS.

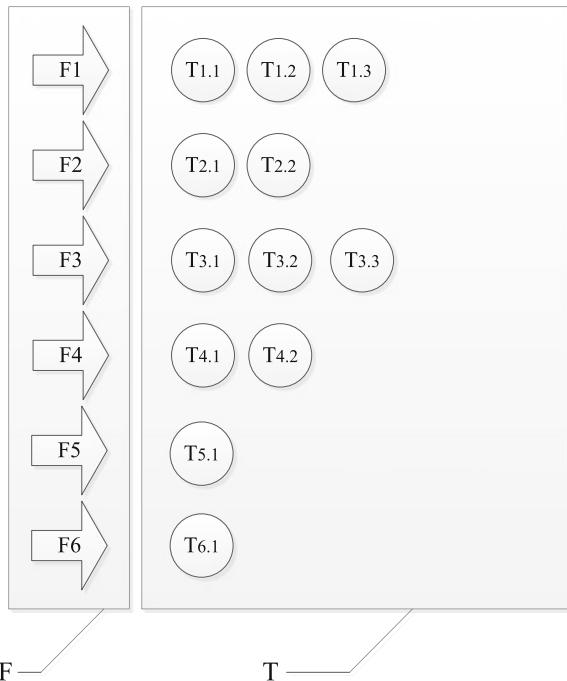
Stage 3. Choice of operation type. The basic task of this stage is determination of operation (combining and decomposition) for forming of software requirements profile.

Stage 4. Choice of profile. Stage needed, when FHS use iteratively. Basic task this opportunity of download of FHS from database.

Stage 5. Operation of combining (or decomposition) of facet and hierarchical structures. Stage needed for realization of combining (or decomposition) of FHS in according with their types and establishing of combining type (or decomposition).

Stage 6. Forming and description of requirements profile. Requirements profile on the basis of selected operation of taxonomic structure transformation has been received. Further created profile must be description for repeated his use.

**Fig. 14** Facet structure for GAMP



## 4 Case Study: Profiling of Requirements to PE MES Software

### 4.1 Profiling Procedure

Let us receive the generalized software requirements MES profile. The software requirements profile will have developed for initial data. In this case software MES requirements profile based on [12] and GAMP (Good Automated Manufacturing Practice) for software validation [13]. Let's represent short characteristic of these standards:

1. ISO/IEC 25010 standard “Systems and software—Systems and software Quality Requirements and Evaluation (SQuaRE)—System and software quality models” standard defines: (a) A quality in use model composed of five characteristics (some of which are further subdivided into subcharacteristics) that relate to the outcome of interaction when a product is used in a particular context of use. This system model is applicable to the complete human-computer system, including both computer systems in use and software products in use; (b) A product quality model composed of eight characteristics (which are further subdivided into subcharacteristics) that relate to static properties of software and dynamic properties

**Table 5** Compliance matrix of facet structure

	F <sub>1</sub>	F <sub>2</sub>	F <sub>3</sub>	F <sub>4</sub>	F <sub>5</sub>	F <sub>6</sub>	F <sub>7</sub>	F <sub>8</sub>
T <sub>1.1</sub>	1	0	0	0	0	0	0	0
T <sub>1.2</sub>	1	0	0	0	0	0	0	0
T <sub>1.3</sub>	1	0	0	0	0	0	0	0
T <sub>2.1</sub>	0	1	0	0	0	0	0	0
T <sub>2.2</sub>	0	1	0	0	0	0	0	0
T <sub>2.3</sub>	0	1	0	0	0	0	0	0
T <sub>3.1</sub>	0	0	1	0	0	0	0	0
T <sub>3.2</sub>	0	0	1	0	0	0	0	0
T <sub>4.1</sub>	0	0	0	1	0	0	0	0
T <sub>4.2</sub>	0	0	0	1	0	0	0	0
T <sub>4.3</sub>	0	0	0	1	0	0	0	0
T <sub>4.4</sub>	0	0	0	1	0	0	0	0
T <sub>4.5</sub>	0	0	0	1	0	0	0	0
T <sub>4.6</sub>	0	0	0	1	0	0	0	0
T <sub>5.1</sub>	0	0	0	0	1	0	0	0
T <sub>5.2</sub>	0	0	0	0	1	0	0	0
T <sub>5.3</sub>	0	0	0	0	1	0	0	0
T <sub>5.4</sub>	0	0	0	0	1	0	0	0
T <sub>6.1</sub>	0	0	0	0	0	1	0	0
T <sub>6.2</sub>	0	0	0	0	0	1	0	0
T <sub>6.3</sub>	0	0	0	0	0	1	0	0
T <sub>6.4</sub>	0	0	0	0	0	1	0	0
T <sub>6.5</sub>	0	0	0	0	0	1	0	0
T <sub>7.1</sub>	0	0	0	0	0	0	1	0
T <sub>7.2</sub>	0	0	0	0	0	0	1	0
T <sub>7.3</sub>	0	0	0	0	0	0	1	0
T <sub>8.1</sub>	0	0	0	0	0	0	0	1
T <sub>8.2</sub>	0	0	0	0	0	0	0	1
T <sub>8.3</sub>	0	0	0	0	0	0	0	1
T <sub>8.4</sub>	0	0	0	0	0	0	0	1
T <sub>8.5</sub>	0	0	0	0	0	0	0	1

of the computer system. The model is applicable to both computer systems and software products.

2. “Good Automated Manufacturing Practice” (GAMP) [13] is a standard for validation of automated systems for pharmaceutical branch. This standard regulates the main phases of systems development life cycle for PE software. According to GAMP software suppliers and software users must be responsible and carefully abide validation procedures.

**Table 6** Compliance of taxons indexes with taxon names

T <sub>1.1</sub>	Functional completeness
T <sub>1.2</sub>	Functional correctness
T <sub>1.3</sub>	Functional appropriateness
T <sub>2.1</sub>	Time-behavior
T <sub>2.2</sub>	Resource utilization
T <sub>2.3</sub>	Capacity
T <sub>3.1</sub>	Co-existence
T <sub>3.2</sub>	Interoperability
T <sub>4.1</sub>	Appropriateness recognisability
T <sub>4.2</sub>	Learnability
T <sub>4.3</sub>	Operability
T <sub>4.4</sub>	User error protection
T <sub>4.5</sub>	User interface aesthetics
T <sub>4.6</sub>	Accessibility
T <sub>5.1</sub>	Maturity
T <sub>5.2</sub>	Availability
T <sub>5.3</sub>	Fault tolerance
T <sub>5.4</sub>	Recoverability
T <sub>6.1</sub>	Confidentiality
T <sub>6.2</sub>	Integrity
T <sub>6.3</sub>	Non-repudiation
T <sub>6.4</sub>	Accountability
T <sub>6.5</sub>	Authenticity
T <sub>7.1</sub>	Adaptability
T <sub>7.2</sub>	Installability
T <sub>7.3</sub>	Replaceability
T <sub>8.1</sub>	Modifiability
T <sub>8.2</sub>	Testability
T <sub>8.3</sub>	Modularity
T <sub>8.4</sub>	Reusability
T <sub>8.5</sub>	Analyzability

## 4.2 Initialization Data

According to profiling algorithm stages has defined taxonomic structure types and conceptual represented the initial profiles. In this case initial profile represents with facet structures (Figs. 13 [12] and 14 [13]).

Initial facet structures have described in set-matrix view.

Facet structure (Fig. 13) of PE MES requirements profile [12], consists of sets  $F = \{F_1, F_2, F_3, F_4, F_5, F_6, F_7, F_8\}$  and  $T = \{T_{1.1}, T_{1.2}, T_{1.3}, T_{2.1}, T_{2.2}, T_{2.3}, T_{3.1}, T_{3.2}, T_{4.1}, T_{4.2}, T_{4.3}, T_{4.4}, T_{4.5}, T_{4.6}, T_{5.1}, T_{5.2}, T_{5.3}, T_{5.4}, T_{6.1}, T_{6.2}, T_{6.3}, T_{6.4}, T_{6.5}\}$ .

**Table 7** Compliance of classification attributes (facets) indexes with facets names

	Classification attributes name
F <sub>1</sub>	Functional Suitability
F <sub>2</sub>	Performance efficiency
F <sub>3</sub>	Compatibility
F <sub>4</sub>	Usability
F <sub>5</sub>	Reliability
F <sub>6</sub>	Security
F <sub>7</sub>	Portability
F <sub>8</sub>	Maintainability

**Table 8** Compliance matrix of facet structure

	F <sub>1</sub>	F <sub>2</sub>	F <sub>3</sub>	F <sub>4</sub>	F <sub>5</sub>	F <sub>6</sub>
T <sub>1.1</sub>	1	0	0	0	0	0
T <sub>1.2</sub>	1	0	0	0	0	0
T <sub>1.3</sub>	1	0	0	0	0	0
T <sub>2.1</sub>	0	1	0	0	0	0
T <sub>2.2</sub>	0	1	0	0	0	0
T <sub>3.1</sub>	0	0	1	0	0	0
T <sub>3.2</sub>	0	0	1	0	0	0
T <sub>3.3</sub>	0	0	1	0	0	0
T <sub>4.1</sub>	0	0	0	1	0	0
T <sub>4.2</sub>	0	0	0	1	0	0
T <sub>5.1</sub>	0	0	0	0	1	0
T <sub>6.1</sub>	0	0	0	0	0	1

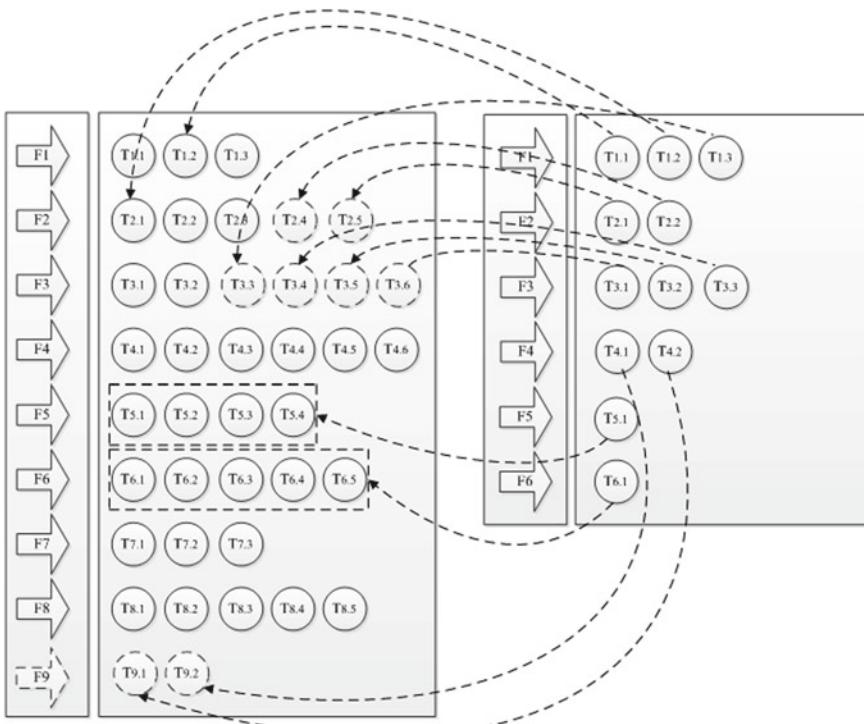
**Table 9** Compliance of taxons indexes with taxons name

	Taxon Names
T <sub>1.1</sub>	Functional correctness
T <sub>1.2</sub>	Response
T <sub>1.3</sub>	Synchronization
T <sub>2.1</sub>	Bandwidth
T <sub>2.2</sub>	Access speed
T <sub>3.1</sub>	User interface
T <sub>3.2</sub>	Interface with another systems
T <sub>3.3</sub>	Interface with another equipments
T <sub>4.1</sub>	Physical allocation
T <sub>4.2</sub>	Physical conditions
T <sub>5.1</sub>	Reliability
T <sub>6.1</sub>	Security

T<sub>7.1</sub>, T<sub>7.2</sub>, T<sub>7.3</sub>, T<sub>8.1</sub>, T<sub>8.2</sub>, T<sub>8.3</sub>, T<sub>8.4</sub>, T<sub>8.5</sub>} in the form of compliance matrix has represented (Table 5). For better understanding of profiling procedure taxons name (Table 6) and classification attributes (facets) of initial taxonomic (facet) structures have represented in Table 7.

**Table 10** Compliance of classification attributes (facets) indexes with facets names

	Classification attributes name
F <sub>1</sub>	Functional
F <sub>2</sub>	Data
F <sub>3</sub>	Interfaces
F <sub>4</sub>	Environment
F <sub>5</sub>	Reliability
F <sub>6</sub>	Security



**Fig. 15** Result of software requirements profiles profiling

Facet structure (Fig. 14) of requirements profile [13] consists of sets  $F = \{F_1, F_2, F_3, F_4, F_5, F_6\}$  and  $T = \{T_{1.1}, T_{1.2}, T_{1.3}, T_{2.1}, T_{2.2}, T_{3.1}, T_{3.2}, T_{3.3}, T_{4.1}, T_{4.2}, T_{5.1}, T_{6.1}\}$  in the form of compliance matrix has represented (Table 8). For better understanding of profiling procedure taxons name (Table 9) and classification attributes (facets) of initial taxonomic (facet) structures have represented in Table 10.

### 4.3 Profiling Results

Initial software requirements profiles have represented in the form of facet structures. In this connection for forming of generalized software requirements profile has used of operation of combining facet structures. As a result of operation combining software requirements profiles, which represented in the form of initial facet structures have received generalized software requirements profile (Fig. 15).

## 5 Conclusions

In this chapter international standards, which describes the PE MES requirements software have considered.

SFHS-based technique for formal description and formation of profiles requirements has proposed. Within the framework of SFHS-based technique set of terms and connection between them have represented (Fig. 6). Possible variants of combining hierarchical (Fig. 11) and facet structures (Fig. 12) have reviewed. SFHS-based technique can be applied to develop profiles requirements to PE MES taking into account their features. To illustrate the approach and technique example of generalized forming process profile to the PE MES requirements was analyzed.

Future work will be connected with the development of tools to support the requirements profiling process and case-based assessment of PE software quality and safety.

## References

1. Enterprise Quality Management Software for the Pharmaceutical Industry [Electronic resource]. <http://www.assurx.com/industries/pharmaceutical.htm>. AssurX, Inc. (2014)
2. Enterprise Quality Management Software for Pharmaceutical and Biotechnology Industries [Electronic resource]. <http://www.spartasystems.com/industries/pharmaceutical-biotech>. Sparta Systems, Inc. (2014)
3. Pharmaceutical Manufacturing software [Electronic resource]. <http://www.margcompusoft.com/manufacture.html>. Marg Compusoft Pvt. Ltd. (2014)
4. ANSI/ISA88. Standard for batch processing industries [Text]. ANSI/ISA-88.00.01 (2010)
5. ANSI/ISA 95. International standard for the integration of enterprise and control systems [Text]. ANSI/ISA95 (2001)
6. Manufacturing Enterprise Solutions Association (MESA) [Electronic resource]. <http://www.mesa.org>. MESA International (2014)
7. Kharchenko, V., Netkachova, K., Sklyar, V., Siora, A., Andrashov, A.: Safety case-oriented assessment of critical software: several principles and elements of techniques [Text]. Monographs of system dependability, vol. 2, pp. 11–25. Dependability of Networks, Wroclaw, Oficyna Wydawnicza Politechniki Wroclawskiej (2010)
8. Case-assessment of critical software systems. V. 3. Safety [Text]. In: Kharchenko, V. (ed.) Kharkiv, 301 p. National aerospace university KhAI (2012) (In Russian)
9. Fda, U.S. Food and Drug Administration [Electronic resource]. <http://www.fda.gov> (2014)

10. International Society of Automation [Electronic resource]. <http://www.isa.org>
11. Fedoseeva, A., Kharchenko, V.: Development of normative profile for MES software of the pharmaceutical enterprise [Text]. In: Proceedings of the Section of Young Researchers and Scientists, pp. 21–28, Zilina (2014)
12. ISO/IEC 25010: Systems and software engineering—systems and software quality requirements and evaluation (SQuaRE)—system and software quality models [Text]. ISO/IEC JTC1/SC7/WG6 (2011)
13. ISPE Good Automated Manufacturing Practice (GAMP) [Text]. Guide by International Society for Pharmaceutical Engineering, ISPE, 128 p, USA (2005)

# Application of Survival Analysis on Analysing the Association Between Chromosomal Aberrations and Carcinoma

Veronika Kubíčková

**Abstract** This paper focuses on the analysis of data that come from 219 radon-mine workers who underwent cytogenetic analysis (or analyses) of peripheral blood lymphocytes. The main focus lies on the connection between the results of the analysis (the occurrence of various types of chromosomal aberrations—chromatid and chromosome changes and breaks) and the incidence of carcinoma, but we also analyzed the association with other explanatory variables, such as age, smoking, and the level of exposure to radon. The Cox analysis was performed separately for the incidence of any type of cancer and for the incidence of lung cancer only as dependent variables. In the first case, we found two suitable models. Both models utilise age, smoking status, the level of exposure and the frequency of chromatid changes. In addition to these variables, the first model contains the frequency of chromatid breaks and the second model contains overall frequency of chromosomal aberrations. As for the lung cancer incidence, we found only two significant factors, the level of exposure to radon and the fact whether or not the subject ever smoked. The other purpose of the paper was to compare our results with the results of the study of Šmerhovský et al. [8], as our study was an expansion to this study. For the overall incidence of cancer, the results were not markedly different from the aforementioned study. Regarding the lung cancer incidence, a significant association of chromosomal aberration frequency and chromatid breaks frequency had been found in the original study, while our findings showed that none of the aberrations were significant for the lung cancer incidence.

**Keywords** Cox model · Chromosomal aberrations · Cancer · Radon exposure

---

V. Kubíčková (✉)

Department of Applied Mathematics, VŠB—Technical University of Ostrava,  
Ostrava, Czech Republic

e-mail: veronika.kubickova1@gmail.com

## 1 Introduction

In this paper we perform the analysis of medical survival data of 219 miners of the Cínovec ore mine who were exposed to radon. The subjects underwent one or more cytogenetic analyses that detected the occurrence of structural chromosomal aberrations in peripheral blood lymphocytes, specifically chromosomal breaks and changes and chromatid breaks and changes. These types of aberrations have been conceptualized as a biomarker of early effects of carcinogenesis, but the long-term effects of their increased frequency remain still uncertain and are not much explored [8].

MUDr. Šmerhovský et al. designed a retrospective epidemiologic study that was purposed to study the association between the occurrence of these types of aberrations and the risk of cancer. The results of the study were published in 2002 and can be found in [8]. During the last 12 years, new information has appeared (i.e., new cases of cancer and new deaths for other reasons occurred), so we include the new information and carry out a new analysis.

The Cox model, which is one of the most useful tools of the survival analysis, is used for the analysis. First, we consider a diagnosis of any type of carcinoma or a death from it as a failure. Death from other reasons and the fact that the subject survived until the end of study (May 31, 2011) are treated as censoring (i.e., the failure itself could not be observed, but we know that the subject had no failure until the censoring time). Secondly, we restrict the failure only to the carcinoma of the lung. Censoring in this case (in addition to the cases mentioned above) includes death from other types of cancer. In both cases, we assume that the censoring is independent of survival time.

## 2 Theoretical Tool

### 2.1 Introducing Survival Analysis

The survival analysis focuses on analysing the time interval between a specified moment (*origin*) and the occurrence of a specified event (*failure*). The variable representing the length of the time interval (usually denoted by  $T$ ) is a nonnegative number. The main interest of the analysis lies in the connection between  $T$  and some explanatory variables. Further we assume that  $\mathbf{X} = (X_1, \dots, X_q)^T$  is a vector of explanatory variables independent of time and  $\mathbf{Z} = (Z_1, \dots, Z_p)^T$  a vector of functions derived from  $\mathbf{X}$  (e.g. the explanatory variables themselves, indicator functions, interactions, etc.). Small letters  $\mathbf{x}$  and  $\mathbf{z}$  denote particular realizations of  $\mathbf{X}$  and  $\mathbf{Z}$ .

As mentioned above, the failure is not always observed. The range of phenomena that cause this include so called (emphright/left/interval) *censoring* and *truncation* but in this article we deal with the right censoring only; dealing with other types does not require substantial modifications of the process. The subject is said to be right

censored at time  $t$  if the only information we have is that he had not failed until  $t$ , i.e. the failure could have occurred anytime in the time interval  $(t, \text{now})$ . The data come in the form of  $(t_i, f_i)$ ,  $i = 1, \dots, n$ , where  $t_i$  are observed times of events (failure or censoring) and  $f_i$  are censoring indicators ( $f_i = 1/0$  for a failure/censoring observed at  $t_i$ ).

For the validity of the survival analysis methods, it is important that the censoring mechanism is *independent*, which means that the risk of failure for every subject of the risk set in every time  $t$  is the same regardless of the occurrence of censoring.

Functions commonly used to describe survival data are the *survivor function* and the *hazard function*.

*Survivor function*  $S(t; \mathbf{x})$  indicates the probability of surviving the time  $t$  for all the subjects with explanatory variables  $\mathbf{x}$ :

$$S(t; \mathbf{x}) = \mathbb{P}(T > t; \mathbf{x}). \quad (1)$$

*Hazard function*  $\lambda(t; \mathbf{x})$  indicates the risk of failure at time  $t$  for those subjects whose values of explanatory variables are  $\mathbf{x}$  and who have lived until  $t$ , i.e., it indicates the conditional probability of failure in the infinitesimal time interval  $[t, t + dt]$  given that the subject has not failed or has not been censored before  $t$ . It can be easily seen that hazard function can take nonnegative values only. (This section sources from [6], Chaps. 1 and 2).

## 2.2 Cox Model

Cox model is popular among the experts because it provides reasonable estimates of regression parameters for a wide class of problems without the necessity to specify a baseline hazard function. Another advantage of this model results from the use of exponential function. The condition of nonnegativity of the hazard function is automatically ensured by using the exponential. Otherwise, some restrictions on explanatory variables would be necessary. (See [7], pp. 96–97, for more details).

The Cox model assumes the hazard function in the form

$$\lambda(t|\mathbf{x}) = \lambda_0(t) \exp(\mathbf{z}^T \boldsymbol{\beta}), \quad (2)$$

where  $\lambda_0$  is an unknown function, called *baseline hazard function*, common for all the subjects in the study, and  $\boldsymbol{\beta}$  is a vector of unknown regression coefficients.

As our interest lies in the connection of the failure time with the regression variables, we can use the concept of *partial likelihood* to eliminate the nuisance parameter  $\lambda_0(\cdot)$  and infer on the regression parameters  $\boldsymbol{\beta}$ . ([6], Chap. 4).

We will quickly introduce the basic idea of the implementation of the partial likelihood on the survival data. The sources for this are ([6], Sect. 4.2), and [2].

First, we assume that there are no ties among the failure times, i.e. there are  $k$  observed failures at times  $t_1^0 < t_2^0 < \dots < t_k^0$  (and we denote  $t_0^0 = 0$  a  $t_{k+1}^0 = \infty$ ). Further, let the identifier  $(i)$  denote the subject that failed at  $t_i^0$ . The rest of  $n - k$  subjects are censored, always  $m_i$  subjects censored in the interval  $[t_i^0, t_{i+1}^0)$ , at the times  $t_{i,1}^0, \dots, t_{i,m_i}^0$ , identifier  $(i, j)$  denotes the subject censored at  $t_{i,j}^0$ .

We also assume that the censoring is independent and *noninformative*, i.e., the likelihood of a subject being censored in a time interval  $[t, t + dt)$  (assuming that the set surviving the time  $t$  is known) is independent of the vector  $\beta$  (thus independent of the covariates vector  $\mathbf{x}$ ).

For  $i = 1, \dots, k + 1$  let us denote:

$$Y_i = \{t_i^0, t_{i-1,j}^0, (i-1, j); j = 1, \dots, m_{i-1}\}, \quad (3)$$

$$S_i = \{(i)\}. \quad (4)$$

The sequence  $Y^{(i)} = (Y_1, \dots, Y_i)$  contains all the failure times smaller or equal to  $t_i^0$ , all the censoring times preceding  $t_i^0$  and the identifiers of the subjects censored at those times. The sequence  $S^{(i)} = (S_1, \dots, S_i)$  contains the identifiers of the subjects who failed before or at the time  $t_i^0$ .

The likelihood can be written in the following form:

$$\prod_{j=1}^k f(y_j | \mathbf{y}^{(j-1)}, \mathbf{s}^{(j-1)}; \lambda_0(t), \beta) \prod_{j=1}^k f(s_j | \mathbf{y}^{(j)}, \mathbf{s}^{(j-1)}; \beta). \quad (5)$$

The information about  $\beta$  lost by omitting the first product is not essential, as we assume that the censoring is noninformative and the distribution of the failure times can be explained by the form of the baseline hazard function. By omitting the first product we obtain the partial likelihood function

$$L(\beta) = \prod_{j=1}^k f(s_j | \mathbf{y}^{(j)}, \mathbf{s}^{(j-1)}; \beta) = \prod_{j=1}^k L_j(\beta). \quad (6)$$

$L_j$  denotes the conditional probability of the fact that it is exactly the subject  $(j)$  that fails in the infinitesimal time interval  $[t_j^0, t_j^0 + dt_j^0)$  on the condition specifying the set of subjects at risk at the time  $((t_j)^0)^-$ . The risk set is denoted  $R(t_j^0)$ . The conditional probability equals the ratio of the probability that exactly the subject  $(j)$  fails in the interval  $[t_j^0, t_j^0 + dt_j^0)$  and the probability that anyone of the set  $R(t_j^0)$  fails in the interval.

As we assume the Cox model, i.e. the hazard function in the form  $\lambda(t|\mathbf{x}) = \lambda_0(t) \exp(\mathbf{z}^T \beta)$ , we can write:

$$\begin{aligned}
L(\boldsymbol{\beta}) &= \prod_{j=1}^k \frac{\lambda_0(t_j^0) \exp(\mathbf{Z}_{(j)}^T \boldsymbol{\beta})}{\sum_{l \in R(t_j^0)} \lambda_0(t_l^0) \exp(\mathbf{Z}_l^T \boldsymbol{\beta})} \\
&= \prod_{j=1}^k \frac{\exp(\mathbf{Z}_{(j)}^T \boldsymbol{\beta})}{\sum_{l \in R(t_j^0)} \exp(\mathbf{Z}_l^T \boldsymbol{\beta})}.
\end{aligned} \tag{7}$$

We can see that the resulting likelihood function does not contain the baseline hazard function. For obtaining the partial likelihood estimate of the coefficient vector  $\boldsymbol{\beta}$ , we proceed as usual using our result instead of the standard likelihood function.

*Note:* In practice, ties in the failure times can occur (usually due to rounding-off or grouping of the observations). The occurrence of ties brings some difficulties, as we assume continuous distribution, but can be handled for example using Breslow or Efron approximation (please see ([6], Sect. 4.2) for details).

### 3 Computational Tool

Statistical software R was used to implement the model computationally. We mostly used the package `survival` that contains many functions that facilitate the implementation of the survival analysis into R. The bibliographic source for this section is [9].

The basic function is `Surv(times, status)`, which compiles the (right) censored data into the form that the other functions from the package `survival` can process. The vector `times` includes the times of the events (failures or censorings), the vector `status` (of the same length as `times`) states whether the event was a failure (mostly the value TRUE or 1) or a censoring (FALSE or 0).

The function `coxph` is designed to create and test the Cox model for submitted data. The most important arguments are:

- `formula`—the formula for the model, in the form “survival object as returned by `Surv`” ~ “regressors”
- `data`—the data frame which interprets the variables used in the others arguments
- `subset`—logical vector that determines the rows of the data frame that should be taken into account,
- `init`—vector of the initial approximation of the parameters, default value is the null vector,
- `ties=c(''efron'', ''breslow'', ''exact'')`—specification of the ties-handling method, the default option is the Efron method.

Example: Let us have fictional data:

```

times=c(1,8,13,17,23,23,25,29,30,30,31,31,31,31,32,34,35,37,39,40),
status=c(0,1,1,0,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1),
x=as.factor(c(1,1,1,2,1,2,2,1,2,1,2,2,1,1,2,1,1,2,2,1,1))

```

where the vector `times` contains the times of the complete cure from a disease (i.e. failure) or some kind of right censoring, the vector `status` is as described above, and the vector `x` determines the kind of cure given to each patient.

The Cox model for a dichotomous factor (assuming the values 1 and 2) is in the form:

$$\lambda(t|x) = \lambda_0(t)e^{\beta I(x=2)}, \quad \text{where } I \text{ is the indicator function.}$$

By calling the function `coxph` we can get a basic concept about the parameters.

```
> Cox1=coxph(Surv(times,status)~x)
> Cox1
      coef  exp(coef)   se(coef)      z      p
x2     -0.0882     0.916     0.489 -0.181  0.86
Likelihood ratio test=0.03  on 1 df, p=0.857 n= 20
```

Each row of the output contains the estimate of the parameter for the variable (in this case one), estimates of its exponential (i.e. the hazard ratio) and standard deviation, the value of the Wald statistics  $z$  (for the hypothesis that the parameter is equal to 0) and the  $p$ -value for the hypothesis (against two-sided alternative).

For more details we can call:

```
> summary(Cox1),
```

which gives us additional information about the confidence interval for each estimated hazard ratio, the determination index for the model and the values of the statistics of the score, Wald and likelihood ratio tests for the hypothesis that the parameter vector is null.

## 4 Analysis

### 4.1 Data Description

The cohort consists of 219 subjects. 66 of them were diagnosed with or died of cancer, 40 subjects were censored during the study due to death from other causes. Remaining 113 were censored at the end of the study. As for the lung cancer, there were 28 people diagnosed with this type of carcinoma, 60 subjects were censored during the study and the rest of 133 subjects were censored at the end of the study.

The following variables were taken into account:

- `age`—continuous variable, the age of the subject as he entered the study (at the first cytogenetic analysis),
- `abc`—continuous variable, the average frequency of aberrant cells,

**Table 1** The distribution of the continuous covariates (columns 4–8 indicate appropriate quantiles)

Variable	# NAs	# 0s	min	0.05	0.33	0.5	0.67	0.95	max
age	0	0	24.8	27.4	35.0	39.2	45.6	56.0	70.3
abc	0	1	0.0	1.22	2.0	2.4	2.92	5.37	9.0
b1	0	9	0.0	0.33	1.0	1.2	1.55	3.14	9.0
b2	0	9	0.0	0.25	0.89	1.0	1.25	2.0	3.6
ch1	0	141	0.0	0.0	0.0	0.0	0.1	0.4	1.67
ch2	0	158	0.0	0.0	0.0	0.0	0.0	0.39	1.0
exp	37	0	2.0	4.1	11.0	13.0	18.0	32.0	40.0
WLM	37	0	0.6	3.12	11.19	32.23	73.4	423.4	613.3

- b1—continuous variable, the average frequency of chromatid breaks,
- b2—continuous variable, the average frequency of chromosome breaks,
- ch1—continuous variable, the average frequency of chromatid changes,
- ch2—continuous variable, the average frequency of chromosome changes,
- exp—continuous variable, the length of employment at the mine,
- WLM—continuous variable, the total level of exposure to radon during the whole period of employment (so called *working level month*),<sup>1</sup>
- smk—dichotomous variable that indicates whether or not the subject ever smoked. The subjects that did not provide this information were considered as smokers, because smoking was extremely prevalent among miners in the observed period (see [8]).

*Notes:* Most of the subjects underwent the cytogenetic analysis multiple times. Therefore the variables abc, b1, b2, ch1 and ch2 indicate the *arithmetic average* of the frequencies of the respective types of aberration detected during individual analyses.

Every cell that contains one or more of the aforesaid aberrations is declared aberrant. It follows that the variable abc correlates with the sum of the variables b1, b2, ch1 and ch2 (Spearman's correlation coefficient 0.987). The most frequent type of aberration is the chromatid break b1 (Spearman's coefficient with abc 0.79). The changes of both types are rare. For more details about the distributions of the explanatory variables, see Tables 1 and 2.

As expected, the variable describing the length of employment in the mine exp is strongly correlated to the variable WLM, which indicates the level of exposure (Spearman's correlation coefficient 0.7).

---

<sup>1</sup>One working level WL is defined as any combination of radon and radon daughters in 1 l of mine air, which will result in the emission of  $1.3 \times 10^5$  MeV of  $\alpha$  energy (definition taken from [8, p.168]).

**Table 2** The distribution of the discrete covariate

Variable	Number of NAs	Non-smokers	Smokers
smk	55	77	87

**Table 3** Cox model for overall incidence of cancer I

Variable	HR	p-value	95 % confidence int. for HR
age	1.110	$6.5 \times 10^{-8}$	[1.069, 1.153]
WLM	1.002	$1.1 \times 10^{-2}$	[1.001, 1.004]
smk	2.225	$9.7 \times 10^{-3}$	[1.214, 4.080]
b1	1.537	$5.6 \times 10^{-3}$	[1.134, 2.083]
ch1	5.778	$1.5 \times 10^{-2}$	[1.410, 23.670]

**Table 4** Cox model for overall incidence of cancer II

Variable	HR	p-value	95 % confidence int. for HR
age	1.106	$1.3 \times 10^{-7}$	[1.065, 1.148]
WLM	1.002	$7.3 \times 10^{-3}$	[1.001, 1.004]
smk	2.238	$9.0 \times 10^{-3}$	[1.223, 4.096]
abc	1.305	$2.7 \times 10^{-2}$	[1.031, 1.651]
ch1	5.184	$3.6 \times 10^{-2}$	[1.116, 24.074]

## 4.2 Overall Cancer Incidence

First, simple models for each variable separately were created and the significance of the parameters was tested. The following variables appeared significant: age, WLM (level of significance 0.001), exp (significance 0.01) and abc and b1 (significance 0.1)

Subsequently we created a model containing all of the covariates (without interactions) and then, with the help of the likelihood ratio test, we found two most suitable submodels.

The first model shows the age at the first analysis, the level of exposure to radon, the frequencies of chromatid breaks and changes and the status of smoking as the relevant explanatory variables. The information about hazard ratios, intervals of confidence and p-values in this model is provided in Table 3.

The second model is very similar, except for it contains the overall frequency of chromosomal aberrations instead of the frequency of chromatid breaks as a covariate. For detailed information about this model see Table 4.

**Table 5** Cox model for the lung cancer as a failure

Variable	HR	p-value	95 % confidence int. for HR
WLM	1.007	$8.1 \times 10^{-10}$	[1.004, 1.009]
smk	2.576	$3.8 \times 10^{-2}$	[1.056, 6.281]

### 4.3 Lung Cancer Incidence

In the case of lung cancer as a failure, we proceeded in the same way as described in the previous section. Regarding the simple models for each variable separately, following variables appeared to be significant: age, WLM (level of significance 0.001), exp (significance 0.01) and b1 (significance 0.1).

Only two of these covariates were significant in the model including all the covariates—smoking smk and the level of exposure WLM. The most appropriate model contains the same two variables—variables describing the level of exposure to radon and the status of smoking. Details can be found in the Table 5.

## 5 Conclusion

We used the Cox analysis to investigate on the connection between the occurrence of cancer and the explanatory variables. Our main interest lied in the frequencies of various types of chromosomal aberrations, other variables were for example the age at the beginning of the study, the status of smoking, the level of exposure to radon, etc.

First, we derived a suitable Cox model for the diagnosis of any type of cancer as a failure. There were two resultant models:

$$\lambda(t; \cdot) = \lambda_0(t) \exp(0.105 \text{age} + 0.002 \text{WLM} + 0.8 \text{smk} + 0.43 \text{b1} + 1.75 \text{ch1}), \quad (8)$$

$$\lambda(t; \cdot) = \lambda_0(t) \exp(0.101 \text{age} + 0.002 \text{WLM} + 0.806 \text{smk} + 0.266 \text{abc} + 1.646 \text{ch1}), \quad (9)$$

where variable age describes the age at the beginning of the study, WLM describes the level of exposure, smk describes the status of smoking and b1/ch1/abc describe the frequency of chromatid breaks/chromatid changes/overall frequency of aberrant cells.

All of the parameter estimates are positive, which means that higher value of any of the explanatory variables in practice indicates higher occurrence of cancer.

For further interpretation we use so-called *hazard ratio* HR. It indicates how many times a risk of failure in a fixed time interval is increased (or decreased in case of negative parameter) when the respective covariate is increased by one unit.

As for the aberrations, the overall occurrence of cancer seems to be significantly connected to the total frequency of aberrant cells (HR = 1.3 with 95 % confidence

interval  $I = [1.031, 1.651]$ ) and/or to the frequency of chromatid breaks ( $HR = 1.53$ ,  $I = [1.134, 2.038]$ ) and to the frequency of chromatid changes ( $HR > 5$  in both models, level of significance 0.05, although not very frequent, this type of aberration cannot be omitted from the models based on the likelihood ratio test).

Regarding the remaining covariates, variables showing significant relation to the occurrence of cancer are age ( $HR = 1.1$ , p-value  $\ll 0.001$ ), smoking ( $HR = 2.2$ , p-value  $< 0.1$ ) and level of exposure to radon measured in WLM ( $HR = 1.002$ , p-value  $< 0.05$ ).

Secondly, we derived a suitable model for the occurrence of lung cancer as an explanatory variable. Contrary to the previous study [8], no type of chromosomal aberration was found to be significant in the resulting model:

$$\lambda(t; \cdot) = \lambda_0(t) \exp(0.006 \text{WLM} + 0.946 \text{smk}). \quad (10)$$

The significant variables were the level of exposure measured in WLM with the hazard ratio 1.007 (p-value  $\ll 0.001$ ) and the status of smoking with the  $HR = 2.576$  and p-value  $< 0.05$ ). Smoking seems to be such a strong predictor of lung cancer that it overshadows the influence of other variables.

**Acknowledgments** The authors would like to thank The National Institute of Public Health for providing the data and for the kind permission to publish this article.

This work was supported by the FEECS VŠB—Technical University of Ostrava (Project No. SP2014/42).

## References

1. Cox, D.R.: Regression model and life tables (with discussion). *J. R. Statist. Soc. B* **34**, 187–220 (1972)
2. Cox, D.R.: Partial likelihood. *Biometrika* **62**, 269–276 (1975)
3. České rudné hornictví v letech 1945–2000, Závod Cínovec [online]. <http://www.hornictvi.info/cetení/4500/06.htm>, [cited 8.3.2013]
4. Fleming, T.R., Harrington, D.P.: Counting Processes and Survival Analysis. Wiley, New York (1991)
5. ICD-10 Code Translator [online]. <http://www.aapc.com/icd-10/codes/index.aspx>, [used 3.12.2012]
6. Kalbfleisch, J.D., Prentice, R.L.: The Statistical Analysis of Failure Time Data. Wiley, 2nd edn. (2002)
7. Klein, M., Kleinbaum, D.G.: Survival Analysis. A Self-Learning Text, 2nd edn. Springer, New York (2005)
8. Šmerhovský, Z., et al.: Increased risk of cancer in radon-exposed miners with elevated frequency of chromosomal aberrations. *Mutat. Res.* **514**, 165–176 (2002)
9. Therneau, T.: Package “survival” [online]. Version 2.37-4, Feb. 2013. <http://cran.r-project.org/web/packages/survival/survival.pdf>, [downloaded 15.3.2013]
10. Veraverbeke, N.: Topics on Survival Analysis. Lectures organized by the Department of probability and mathematical statistics MFF UK. Prague, April 2012

# Modelling the Results of the Phadiatop Test Using the Logistic and Ordinal Regression

Pavlína Kuráňová

**Abstract** This study was based on examination Phadiatop onerous test at the Clinic of Occupational and Preventive Medicine in order to save money and not make unnecessary testing. The aim of this study was to assess the outcome of the test Phadiatop only under close personal or family history of each patient. This estimation was used statistical methods specifically logistic and ordinal regression. The most important findings are that Phadiatop test result does not imply eczema; it is a different immune response and the disease is not relevant in personal or family anamnesis. The patient was based on a family and personal anamnesis, in assigning only two groups (healthy or sick) correctly classified with a probability of 75 %. The test sensitivity is about 77 % and the diseases influencing the results are asthma and allergic rhinitis. The success rate of classifying each patient into one of the five Phadiatop test groups according to the seriousness of diseases was about 68 %. Also a testing based on age groups of the patients was done using this database. The presence of the positive Phadiatop test was the most common for people born between 1972 and 1981, where the genetic predispositions for a positive Phadiatop test results are about 39 %.

## 1 Introduction

The atopy rate of inhabitants of the Czech Republic is increasing. Atopy could be understood as a personal or family predisposition to become, mostly in childhood or adolescence, hyper-sensitive to normal exposure of allergens, usually proteins. These individuals are more sensitive to typical symptoms of asthma, eczema, etc. The Phadiatop test is used as a measure of atopy.

According to disease severity, results of the Phadiatop test are divided into the six following groups: Groups 0 and I indicate none or weak form of atopy and the remaining groups (II, III, IV, V, and VI.) indicate increasing severe forms of atopic

---

P. Kuráňová (✉)

Department of Mathematical Methods in Economics,  
VŠB—Technical University of Ostrava, Ostrava, Czech Republic  
e-mail: pavlina.kuranova@vsb.cz

symptoms. Unfortunately, the Phadiatop test is expensive, so we try to predict the results of the test on the basis of a detailed family and personal anamnesis. The knowledge of the results of the Phadiatop test is very important especially for diagnosis of allergic dermatitis and also for the professional medical care for travelers [1–3].

Various laboratories around the world deal with the Phadiatop test results and the atopy. In Swedish population, there are interesting results such as [4, 5]. The connection between a positive Phadiatop test, sex, age and smoking is described in [6], which proves that the atopy falls down with the age and that smoking tobacco is connected with higher IgE and influences atopy negatively. The connection of ordinal regression and the Phadiatop test result is shown in [7], where Linnerberg et al. prove the connection of alcohol consumption and positive Phadiatop test. They proved that the connection is universal and it is not specific for certain groups of people. Information obtained from personal and family anamneses and the age of each patient were used for detecting the presence of asthma, allergic rhinitis, eczema or other forms of allergy (contact allergy, food, etc.). Family and personal anamnesis of each patient were evaluated by medical expert. Then, we created and verified a mathematical model for the accurate classification of patients into one of the five groups of the Phadiatop test.

In this paper we discuss the ordinal regression approaches for obtaining the results of Phadiatop test based only on family and personal anamnesis. In this paper we deal with the prediction of each patient into one of five groups of the Phadiatop test based on ordinal regression. Next, we describe the connections of genetic predispositions for the atopy according to the age group of inhabitants. Database of patients comes from 2010 to 2012 [8–11].

## 2 Logistic Regression as a Tool for Discrimination

The logistic regression was not originally created for the purpose of discrimination, but it can be successfully applied for this kind of analysis [12–14]. A logistic regression model, which is modified for the purpose of discrimination, is defined as follows. Let  $Y_1, \dots, Y_n$  is a sequence of independent random variables with alternative distributions, whose parameters satisfy:

$$\begin{aligned} P(Y_i = 1|X_i = x_i) &= \frac{e^{\beta_0 + \beta' x}}{e^{\beta_0 + \beta' x} + 1}, \\ P(Y_i = 0|X_i = x_i) &= \frac{1}{e^{\beta_0 + \beta' x} + 1}, \end{aligned} \quad (1)$$

for  $i = 1, \dots, n$ , where  $\beta' = (\beta_1, \dots, \beta_p)'$  is unknown  $p$ -dimensional parameter and  $X_1, \dots, X_n$ , are  $(p+1)$ -dimensional random vectors  $(\beta_1, \dots, \beta_p)$ . This model can be called a learning phase, in which both values  $X_i$  and  $Y_i$  are known for each object (i.e. it is known to which group each object belongs to). Based on this knowledge, we try to predict parameters  $\beta_1, \dots, \beta_p$  and thus we try to estimate function  $\pi(x)$ , where

$$\pi(x) = P(Y = 1|X = x) = \frac{e^{\beta_0 + \beta' x}}{e^{\beta_0 + \beta' x} + 1}. \quad (2)$$

Another object for which the classification is unknown is assigned to one of two groups according to the value of decision function  $\pi(x)$ .

The object will be included in the first group if  $\pi(x) > 0.5$ . Otherwise, the object will be included in the second group. The main advantage of this model is that it does not require conditions for distributions of random vectors  $X_1, \dots, X_n$ . However, the model assumes a very specific form of probability  $P(Y = 1|X = x)$  and we should verify the significance of the relationship

$$\pi(x) = P(Y = 1|X = x) = \frac{e^{\beta_0 + \beta' x}}{e^{\beta_0 + \beta' x} + 1}. \quad (3)$$

### 3 Introduction to Ordinal Regression

Many variables of interest are ordinal. That means that you can rank the values, but the real distance between categories is unknown. Diseases are graded on a scale from the least severe to the most severe. Survey respondents choose answers on a scale from “strongly agree” to “strongly disagree”. Ordinal Regression allows you to model the dependence of a polytomous ordinal response on a set of predictors, which can be factors or covariates [12, 13].

Phadiatop test, the explained variable, is ordinal; its  $s \geq 2$  categories can be objectively organized. Based on this reality, the definition of logit can be done. If we construct the model from the category of the neighbouring row, the ordinal logits can be defined as

$$\ln \frac{\pi_j}{\pi_{j-1}}, \quad j = 1, 2, \dots, s - 1. \quad (4)$$

Using the value of the distributive function of this division ( $F_j$ ), resp. its supplement into one ( $1 - F_j$ ), the cumulative logit can be defined as

$$\ln \frac{F_j}{1 - F_j} = \ln \frac{P(Y \leq y_j)}{P(Y > y_j)} = \ln \frac{\pi_0 + \pi_1 + \dots + \pi_j}{\pi_{j+1} + \pi_{j+2} + \dots + \pi_{s-1}}, \quad j = 1, 2, \dots, s - 2. \quad (5)$$

And the regressive function using the cumulative logit as

$$\ln \frac{F_j}{1 - F_j} = x' \beta_j, \quad j = 0, 1, \dots, s - 2. \quad (6)$$

The parameters  $\beta_{0j}$  are threshold (limit) parameters for the individual categories of the quantity  $Y$  and they represent the logarithm odds, that  $Y$  acquires maxim  $j$ th category, not any higher. According to the way of defining the cumulative logit, in

this case  $\beta_{0,0} \leq \beta_{0,1} \leq \dots \leq \beta_{0,s-2}$ . The plus coefficients in the vector  $\beta_j$  then means that the growing values of the explaining variables cause the dominance of the lower ones, or the dominance of the higher categories of the quantity  $Y$  falls down, and the other way round. In order to get the usual parameters interpretation, model (6) where

$$x' \beta_j = \beta_{0j} + \sum_{i=1}^k \beta_{ij} \cdot x_i, \quad (7)$$

is often recorded as

$$\ln \frac{F_j}{1 - F_j} = \beta_{0j} - \sum_{i=1}^k \beta_{ij} \cdot x_i, \quad j = 0, 1, \dots, s-2. \quad (8)$$

For example, when the odds of the explained variable gains the value from the  $j$ th category and not from any of the previous categories, then the regressive model can be defined as

$$\ln \frac{\pi_j}{\pi_0 + \pi_1 + \dots + \pi_{j-1}} = x' \beta_j, \quad j = 1, 2, \dots, s-1. \quad (9)$$

## 4 Results Gained Through Logistic Regression

The tested biomedical data are from the University Hospital of Ostrava, Department of Occupational and Preventive Medicine, Ostrava, Czech Republic. Two databases were used for testing. The older one contains 1027 records and logistic regression was used there. The newer database contains 1974 records and ordinal regression was used there. Each coefficient has a 99 % confidence level. The logistic regression is used in order to predict the results of the Phadiatop test. The medical database contained information on 1027 patients. Patients in Group 0 have Phadiatop test 0 or I (no visible symptoms), so no treatment was necessary. The remaining patients with Phadiatop test II–VI are members of Group 1. Medical treatment is necessary for these patients.

Logistic regression does not have any requirements for the data arrangement, but we need a specific format of the data for the logistic regression. For this particular case we have one dependent variable  $Y$ , Phadiatop ( $Ph$ ), which depends on two independent variables of personal anamnesis ( $OA$ ) and family anamnesis ( $RA$ ). Variable  $Y$  can be 0 or 1, according to the membership of a patient to Group 0 or Group 1, respectively. Values of these independent factors were obtained from medical experts. The expert severity scores for personal and family anamneses are presented in Table 1. Here, the category “Others” represents the score of various kinds of allergies (food allergies, etc.). The independent variables of personal and

**Table 1** Expert severity score of personal and family anamnesis

Factor	Asthma	Allergic rhinitis	Eczema	Others
Score	10	8	6	4

family anamnesis of the logistic regression were obtained as a sum of these scores for each patient from the database.

In the first step, we try to create a regression model by assuming all supplied data for the training phase (Case A). Then, the developed model is tested using the same data. We used data corresponding to all 1027 patients. We obtained the following logistic model:

$$\ln \left( \frac{Ph}{1 - Ph} \right) = -1.54347 + 0.212376 \cdot OA + 0.0146104 \cdot RA \quad (10)$$

Prediction results of the model (10) are summarized in Table 2, column Case A. In summary, the results of Phadiatop test were incorrectly predicted for 220 patients, which give us the following error prediction rate of the model:

$$E_p = \frac{220}{1027} = 0.21422 \quad (11)$$

The error rate was statistically evaluated by exact confidence limits: With 90 % probability we can say that the relative error of the model is on the interval [0.193; 0.236]. Thus, the results of the Phadiatop test were incorrectly predicted only for every fourth or fifth patient. In the second step, we created a learning group as a random sample from 90 % of the database (926 patients). To verify the correctness of model assumptions, the logistic model was created using the learning group. We obtained the following updated model:

$$\ln \left( \frac{Ph}{1 - Ph} \right) = -1.56672 + 0.21122 \cdot OA + 0.0198823 \cdot RA \quad (12)$$

For testing this updated model, we analyzed the remaining data set, i.e. 10 % of the database (1027 patients), which were not assumed for the training phase (Case B). The results are summarized in Table 2, column Case B. Of course, decreasing the

**Table 2** Prediction results of regression models

	Case A	Case B
Number of correctly classified patients	807	78
Number of incorrectly classified patients	220	24
Number of patients predicted for Group 0	233	21
Number of patients predicted for Group 1	331	33
Exact confidence limits for the model prediction rate	[0.193; 0.236]	[0.168; 0.315]

number of tested patients produces wider confidence interval, but these first results remain very promising.

We also provided analysis of variance for the second model. A statistically significant relationship between variables at 99 % confidence level was proved there.

We evaluated coefficients of OA and RA using the Pearson Chi-square significance test. Variable OA (personal anamnesis) is statistically significant at the 95 % confidence level. On the other hand, p-value for RA variable (family anamnesis) is larger than 0.05. Thus, RA variable is not statistically significant and may be excluded from the model. This result is maybe caused by the insufficient information on family anamnesis in the database.

## 5 Testing Based on Logistic Regression Using the Personal Anamnesis Only

Further tests were done only for personal anamnesis OA, this was divided into 4 independent variables of asthma, allergic rhinitis, eczema and others. A logistic regressive model was created and verified for these data using the theory of logistic regression.

The quality of a probabilistic prediction model can be characterized by terms: Sensitivity, Specificity, Positive and Negative Predictive Values [15].

*Sensitivity* reflects the ability of the model to detect those with disease, whereas *specificity* illustrates the ability of a model to distinguish who do not have disease:

$$\text{Sensitivity}(SE) = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}} = \frac{TP}{TP + FN}. \quad (13)$$

$$\text{Sensitivity}(SE) = \frac{\text{true negatives}}{\text{true negatives} + \text{false positives}} = \frac{TN}{TN + FP} \quad (14)$$

Moreover, *positive predictive value* (PPV) is defined as the number of individuals for whom the model test is positive and they have a disease divided by all of those who have a positive model result:

$$PPV = \frac{TP}{TP + FP} \quad (15)$$

Finally, *negative predictive value* (NPV) is defined as the number of individuals who do not have disease divided by all of those who have a negative model result:

$$PPV = \frac{TN}{TN + FN} \quad (16)$$

An ideal prediction model has neither false negatives, nor false positives. Thus,  $SE = SP = PPV = NPV = 100\%$  for an ideal prediction model. In the first step, we have tried to create a prediction model of the Phadiatop test by assuming all of the supplied data corresponding to 1027 patients. Then, the developed model has been tested using the same data. We have obtained the following global prediction model:

$$\ln \left( \frac{PhModel}{1-PhModel} \right) = -1.4856 + 0.1528 \cdot asthma + 0.3156 \cdot allergic\ rhinitis + 0.0765 \cdot ekzema + 0.1549 \cdot others. \quad (17)$$

Our model has one dependent variable (PhModel) and four independent variables: asthma, allergic rhinitis, eczema and others.

The prediction results of the global model (17) are summarized in Tables 3 and 4. For example, there were 166 “true positive” patients with the positive Phadiatop test ( $PhTest = 1$ ) correctly recognized by the prediction model as  $PhModel = 1$ . Similarly, there were 641 “true negative” patients, i.e. patients with negative Phadiatop test ( $PhTest = 0$ ) correctly recognized by the prediction model as  $PhModel = 0$ .

A negative result from the Table 3 is very good at reassuring that a patient has no serious atopic syndrome, ( $NPV = 80\%$ ) and at this initial screen correctly identifies 92 % of those who do not have the serious atopic syndrome (the specificity). All of the incompletely filled records in the medical database belong to patients with  $PhTest = 0$ . In the previous section, these incomplete records have been re-placed by zeros, because it had been expected that all positive allergic symptoms were stored in the database. In contrary to the previous section, this section brings prediction results of a case when all of the incomplete filled database records have been removed from the database. By this pre-processing process, i.e. by removing these incomplete filled records, the reduced database has had finally 376 records. For these records, we have created a new prediction model (18):

$$\ln \left( \frac{PhModel}{1-PhModel} \right) = -0.9765 + 0.1251 \cdot asthma + 0.2675 \cdot allergic\ rhinitis + 0.0113 \cdot ekzema + 0.0623 \cdot others \quad (18)$$

**Table 3** Prediction results of the logistic regression model (17)

	PhModel = 1	PhModel = 0
$PhTest = 1$	166 (TP)	165 (FN)
$PhTest = 0$	55 (FP)	641 (TN)

**Table 4** Prediction results of the logistic regression model (17). 1027 patients

Sensitivity (SE)	Specificity (SP)	Positive predictive value (PPV)	Negative predictive value (NPV)
0.50	0.92	0.75	0.80

**Table 5** Prediction results of the logistic regression model (18). 376 patients

	PhModel = 1	PhModel = 0
PhTest = 1	166 (TP)	49 (FN)
PhTest = 0	55 (FP)	106 (TN)

**Table 6** Prediction results of the logistic regression model (18). 376 patients

Sensitivity (SE)	Specificity (SP)	Positive predictive value (PPV)	Negative predictive value (NPV)
0.77	0.66	0.75	0.80

The prediction results of this model are presented in Tables 5 and 6. For example, there have been observed 166 “true positive” patients. It means that the diseased patients with PhTest = 1 have been correctly recognized by the prediction model as PhModel = 1. Furthermore, the number of patients with a positive PhTest is  $166 + 49 = 215$ .

As a result, a sensitivity of  $166/215 = 77\%$  means that the model correctly recognizes 77% actual positives i.e. 77% diseased people are correctly recognized as being ill. The prediction model is also good at confirming the diseased people (PPV = 75%).

The percentage of the model sensitivity indicates that the prediction outcome is a slightly better than the prediction outcome for the complete database of 1027 patients. By removing the incomplete records, only the fully relevant data is analyzed.

Note, that although the database has been significantly reduced, results corresponding to the column “PhModel = 1” of Tables 5 and 7 are the same.

The prediction model of the reduced database (18) has also been a subject of statistical tests; see Table 7.

Table 7 shows that the variables eczema and/or others are statistically insignificant and may be excluded from the prediction model. By excluding these statistically insignificant variables we obtain the following prediction model Model (19):

$$\ln \left( \frac{\text{PhModel}}{1 - \text{PhModel}} \right) = -0.7927 + 0.1189 \cdot \text{asthma} + 0.2538 \cdot \text{allergic rhinitis} \quad (19)$$

**Table 7** Test of the statistical significance of the Model (18)

Factor	Chi-square	Df	P-Value
Asthma	13.9485	1	0.0002
Allergic rhinitis	56.2333	1	0.0000
Eczema	0.0356	1	0.8503
Others	0.6814	1	0.4091

Even this prediction model has been a subject of the accuracy testing for a patient inclusion into two Phadiatop groups. Based on these results we can conclude that the reduced model (19) works exactly the same as in the case where we have four variables (18).

## 6 The Obtained Results by Ordinal Regression

The database from University hospital in Ostrava was used for modelling for the results using the ordinal regression. The database contained 1974 records, these records were checked and the final database contained 1132 records after excluding the incomplete data. The database also contains complete data about the family anamnesis of each patient. Separate patients are divided into groups according to the type of illness (skin disease, profession-caused asthma, and other allergies); also a category of control group for travelers and patients without any apparent illness is included. The information about the database are included in Table 8.

### 6.1 The Control Group Data Evaluation

This group was thus examined with the Phadiatop test. The control group contains 541 records and similarly to the previous testing, groups 0 and I are united as well as the groups V and VI (the value of the Phadiatop test VI is present only 4 times in the group). Based on these results it is evident that the whole family anamnesis in this group is not important and all the coefficients can be excluded from the model. This supports the idea of doctors that the eczema is not related to the seriousness of the Phadiatop test and it is a different immunological reaction. The relationship between atopy and allergic contact dermatitis deals with such [16].

**Table 8** Information on the disease division in the database

	Ph 0	Ph I	Ph II	Ph III	Ph IV	Ph V	Ph VI	Sum
Professional asthma (PA)	53	29	11	15	5	2	0	115
Skin problems (Skin)	105	34	18	25	12	5	3	202
Other allergy (Alergo)	107	61	26	38	32	9	1	274
Control group	245	77	46	91	54	24	4	541
Sum	510	201	101	169	103	40	8	1132

**Table 9** Evaluation and verification of the independent variables for ordinal regression

	Estimate	Wald's test	Significance
OA asthma	-1.146	17.399	0.000
OA allergic rhinitis	-2.504	139.842	0.000
OA eczema	-0.222	0.657	0.001
OA others	-0.415	4.203	0.000
RA asthma	-0.038	0.025	0.152
RA allergic rhinitis	-0.306	2.404	0.038
RA eczema	0.171	0.322	0.570
RA others	-0.004	0.000	0.900

**Table 10** Prediction results for the ordinal regression, control group, (541 records)

	Ph Model 0 + I	Ph Model II	Ph Model III	Ph Model IV	Ph Model V+VI
PhTest 0 + I	292	0	28	8	0
PhTest II	34	0	8	4	0
Ph Test III	40	0	42	9	0
Ph Test IV	15	0	27	10	2
Ph Test V + IV	2	0	20	5	1

According to the values based on the Wald's test results it is noticeable that the *RA eczema* and *RA others* are not important for the model and thus they can be excluded from the model, as seen in Table 9.

The results of the patient's classification into one of the five groups of the Phadiatop test for the control group are included in the Table 10. The model classified correctly 345 patients out of 541 patients, thus the correctness ratio is 0.64. Nevertheless, also in this model there are no patients classified into the group II. This supports the idea of connecting the group II with another category. The patients are already predicted in the group V in this model. The low values of the prediction may be caused by the low number of the patients with the Phadiatop value V and VI.

It was proved earlier that the groups 0 and I are equal and thus can be united. For the purposes of this analysis, also the groups V and VI were united, because in the group VI there were only 8 records (Tables 10 and 11).

## 6.2 Evaluation of Patients with a Known Disease Category Classification

The group of diseased patients is divided into patients with profession-caused asthma (PA), skin diseases (Skin) and allergies (Alergo). It is necessary to examine each

**Table 11** The division of patients into the groups according to the disease type

	Ph 0 + I	Ph II	Ph III	Ph IV	Ph V + VI	Sum
Professional asthma (PA)	82	11	15	5	2	115
Skin problems (Skin)	139	18	25	12	8	202
Other allergy (Alergo)	168	26	38	32	10	274
Sum	389	55	78	49	20	591

**Table 12** Estimates of the independent variables for profession-caused asthma

	Estimate	Wald's test	Significance
OA asthma	-0.565	1.386	0.239
OA allergic rhinitis	0.750	1.011	0.315
OA eczema	16.438	—	—
OA others	-0.352	0.590	0.442
RA asthma	-0.469	0.791	0.374
RA allergic rhinitis	-0.711	1.829	0.176
RA eczema	-0.927	1.412	0.235
RA others	-0.203	0.684	0.767

**Table 13** Estimates of the independent variables for skin problems

	Estimate	Wald's test	Significance
OA asthma	-0.283	0.189	0.664
OA allergic rhinitis	-1.657	18.013	0.000
OA eczema	0.837	1.370	0.541
OA others	-0.816	4.689	0.030
RA asthma	-0.215	0.354	0.552
RA allergic rhinitis	-0.267	0.348	0.591
RA eczema	-0.112	0.102	0.750
RA others	0.556	1.283	0.257

group separately and focus on the examination of the skin diseases group. The exact numbers of patients in individual groups of the Phadiatop test are mentioned in the following table.

The following (Tables 12, 13 and 14) show the estimated independent variables according to the diseases.

**Table 14** Estimates of the independent variables for other allergies

	Estimate	Wald's test	Significance
OA asthma	-0.576	4.882	0.027
OA allergic rhinitis	-1.383	27.220	0.000
OA eczema	-1.200	15.212	0.000
OA others	-0.464	3.158	0.076
RA asthma	-0.301	1.133	0.287
RA allergic rhinitis	0.113	0.135	0.713
RA eczema	0.227	0.498	0.480
RA others	-0.283	0.797	0.372

**Table 15** The evaluation of patients classification into a particular Phadiatop test group

	Ph Model 0 + I	Ph Model II	Ph Model III	Ph Model IV	Ph Model V
PhTest 0 + I	297	0	59	31	2
PhTest II	43	0	8	3	0
Ph Test III	52	0	19	6	1
Ph Test IV	25	0	11	10	3
Ph Test V + IV	9	0	6	5	0

Models for classification of each patient into a particular Phadiatop test groups according to their disease were made. Nevertheless, all the results were much worse than those for classification of patients according to the “control” group data.

The control group model proved to be the most suitable model for Phadiatop testing throughout the diseases. The model based on the Control group was proved to be the best result. Further patients testing will be done based on this model. Phadiatop groups 0 + I and V + VI are united for the following calculations. The final model was created from a Control group of patients and was tested on completely different data, records from ill patients.

For practical use of this model, the other data will be tested, thus, all the disease patients (groups of Profession-caused asthma, Skin, Alergo) together 591 records. The groups are described in Tables 6 and 9. The testing results are described in the following Table 15.

These results support the good predictable ability of the model. In previous cases the model correctly classified patients only into groups 0 + I. In this case the patients are predicted into all Phadiatop test groups; again, the best success rate is for the healthy patients 297 out of 389.

Using the ordinal regression for classification causes a very small difference in the probability of classifying patient into this group and not into a “neighbouring” group. The model is based on the Control group data and is used for testing on different data. The probability of correct classification of patients despite the classification into the neighbouring group is shown in Table 16.

**Table 16** Comparison of the prediction results into individual Phadiatop test groups

Group	Ph M 0 + I	Ph M II	Ph M III	Ph M IV	Ph M V
Together in the group	389	55	78	49	20
Number of predicted patients	426	0	104	55	6
Correctly predicted	297	0	19	10	0
The result including the mistake	297	52	25	24	5

- Correctly classified patients:  $297 + 19 + 10 = 326$ .
- Patients classified into the neighbouring group:  $(43 + 9) + 6 + (11 + 3) + 5 = 77$ .
- Total success rate of the classification:  $\frac{326+77}{597} = \frac{403}{597} = 0.68$ .

The result shows that the best prediction qualities were for the patients division into groups “negative” (Phadiatop groups 0 + I) and “positive” (Phadiatop groups II–VI). The model classifies the patients into all Phadiatop test groups. If we consider the probable incorrect classification into a neighbouring group, the prediction results improve. A significant improvement appears at the “positive” Phadiatop test groups.

## 7 The Phadiatop Test Results Evaluation Based on the Age and Genetic Predispositions

As it was mentioned above, the newer database contains more records than the previous one; it also includes the age of the patients. Based on this information, a proportional representation of different age groups was done for the Phadiatop test results (see Table 12). This evaluation was done for the diseased patients (with specific problems). For clarity, only the positive (II–VI) and the negative (0 and I) Phadiatop tests are mentioned.

Based on the mentioned age groups and the positive and negative Phadiatop tests division it is clear that a proportional representation of the diseased patients is more or less equal. The only group that differs and has the most positive patients is the age group of 31–40 years (born 1981–1972), thus, the young patients.

According to the available information, the genetic predisposition for atopy (positive Phadiatop test II–VI) should be about 30 % for the inhabitants of the Czech Republic. For this analysis, a patient with genetic predispositions was that one who filled in the statistically significant positive diseases of the Phadiatop test results into his family anamnesis. Thus, a patient with genetic predispositions included into his/her family anamnesis:

- Positive asthma or allergic rhinitis and other records then did not need to be taken into account.
- Positive eczema and other diseases at once, if the asthma and rhinitis were negative.

**Table 17** Comparison of the prediction results into individual Phadiatop test groups

Age	Phadiatop 0 + I	Phadiatop II–VI	Ratio of the positive	Genetic predisposition	Predisposition for the positive
To 1982	36	38	0.19	26	17
1981–1972	79	56	0.28	65	30
1962–1971	85	37	0.18	45	11
1952–1961	101	38	0.19	41	12
Before 1952	88	33	0.16	56	13
Sum	389	202	1	233	83

Table 17 (the column Genetic predisposition) shows that there are 233 patients with the genetic predisposition for the positive Phadiatop test and the most patients with this predisposition are in the age group between 30 and 40 years. Thus, the proportional representation of the patients with genetic predisposition is  $\frac{233}{591} = 0.39$ . This proves the division of the data file, where there are 202 positive patients out of 591, about 34 %. Based on the family predispositions, 39 % of all patients should be in the positive Phadiatop test group. Nevertheless, there are  $\frac{83}{202} = 0.41$  genetically predisposed patients with the positive Phadiatop test (202 records).

## 8 Conclusion

The Phadiatop test is a powerful but an expensive technique developed for screening of allergic sensitization. For this reason we have constructed a probabilistic model, which evaluates the data from a personal and family anamnesis obtained from patients with suspected allergic diseases.

Logistic regression was used for the prediction based on the database from 2009. The family anamnesis proved to be statistically non-significant variable. Family anamnesis was incomplete in the database. In the prediction model, just the personal history of patients allergic diseases is evaluated. Only asthma and allergic rhinitis seem to be significant factors for the Phadiatop test prediction model, because all of the remaining attributes have been shown to be statistically insignificant and therefore these insignificant attributes have been removed from the prediction model. The prediction model has interesting properties: the statistical testing over a real patient database indicates that 77 % diseased people have been correctly recognized as being ill. Thus, approximately only every fourth sick patient has been wrongly classified as being healthy by the developed prediction model. The prediction result at the former researches was about 75 %, but the patient was classified into one of the two groups only (diseased or healthy) and the database contained only 376 records.

Ordinal regression was used for testing the newer database from 2010 to 2012. Ordinal logistic regression was used for evaluating the Phadiatop test prediction.

This database included 1132 relevant records; patients were described using the family and personal anamnesis. The database was also divided into a control group and groups with various health problems. Based on the examination, family anamnesis seems to be unimportant and it would be appropriate to continue using the personal anamnesis only. Ordinal regression was used for the examination. The importance of the patient's classification was an unexpected result. The complete database was examined as well as the control group database only. It was even suitable to exclude the parameter of the eczema from the personal anamnesis in this case.

The predictive performance of our complete database (1132 records) and also the control group database (541 records) is about 64 %. However, the final model obtained from the data control group of records that has been tested on completely different patient records has predictive power about 68 %. The result seems to be satisfactory when we consider that we predict patient's classification into one of the five groups. This result is very important for the savings of the public health sector, because the Phadiatop test is expensive. Next interesting result were gained by testing the diseased group (with specific symptoms – 591 records altogether) is the fact that most extended occurrence of the positive Phadiatop test result was in the age group of 31–40 and this is then the group with the biggest number of patients who are genetically predisposed for the positive Phadiatop test result. Our conclusion proves the presumption that about 30 % of population has the genetic predisposition for the positive Phadiatop test. Based on our calculation, the proportional representation is about 39 %.

The future research will focus on the connection between genetic predispositions to the results of the Phadiatop test and as exact classification into the Phadiatop test groups as possible. A detailed analysis of the results will also be important for the future research of biomedical relations, which are hidden in the given biomedical database.

**Acknowledgments** This paper was done thanks to cooperation with The University Hospital of Ostrava, the Department of Clinic of Occupation and Preventive medicine. This work was supported by the International Visegrad Fund's Standard Grant No. 21320401.

## References

1. Williams, B., Siegel, C., Portnoy, J.: Efficacy of a single diagnostic test for sensitization to common inhalant allergens. *Ann. Allergy Asthma Immunol.* **86**(2), 196–202 (2001)
2. Hajduková, Z., Vantuchová, Y., Klimková, P., Makhoul, M., Hromádka, R.: Atopy in patients with allergic contact dermatitis. *J. Czech Phys. Occup. ther.* **2**, 69–73 (2009). In Czech
3. Hajduková, Z., Pólová, J., Kosek, V.: The importance of Atopy Investigation in the Department of Travel Medicine, Allergies: magazine for continuous education in allergy and clinical immunology, No. 2 (2005), pp. 106–109
4. Sigurs, N., et al.: Asthma and allergy patterns over 18 years after severe RSV bronchiolitis in the first year of life. *Thorax* **65**.12 (2010), pp. 1045–1052
5. Wuthrich, B., et al.: Prevalence of atopy and pollinosis in the adult population of Switzerland (SAPALDIA study). *Int. Arch. Allergy Immunol.* **106**(2), 149–156 (2009)
6. Wuthrich, B., et al.: IgE levels, atopy markers and hay fever in relation to age, sex and smoking status in a normal adult Swiss population. *Int. Arch. Allergy Immunol.* **111**(4), 396–402 (2009)

7. Linneberg, A., et al.: Immunoglobulin E sensitization to cross-reactive carbohydrate determinants: epidemiological study of clinical relevance and role of alcohol consumption. *Int. Arch. Allergy Immunol.* **153**(1), 86–94 (2010)
8. Spiewak, R.: Atopy and contact hypersensitivity: a reassessment of the relationship using objective measures. *Ann. Allergy Asthma Immunol.* **95**(1), 61–65 (2005)
9. Kuráňová, P.: The processing of the medical data with use of logistic regression, reliability & risk analysis: theory & applications (R&RATA). *J. Int. Group Reliability* **24**(01), 65–72 (2012)
10. Kuráňová, P., Praks, P., Hajduková, Z.: Improving the probabilistic prediction model of the phadiatop test by data pre-processing. In: Mendel 2012: 18th International Conference on Soft Computing. Brno University of Technology, Brno, pp. 500–504 (2012) ISSN 1803–3814. ISBN 978-80-214-4540-6
11. Kuráňová, P., Hajduková, Z.: The use of logistic and ordinal regression for the prediction of the phadiatop test results. In: Sborník konference the Ninth International Conference on Digital Technologies 2013, Ilina, pp. 111–115 (2013) ISBN: 978-1-4799-0923-0. doi:[10.1109/DT.2013.6566297](https://doi.org/10.1109/DT.2013.6566297)
12. Menard, S.: LOGISTIC REGRESSION From Introductory to Advanced Concepts and Applications. SAGE Publications Inc., California (2009)
13. Hosmer, D.W., Lemeshow, S.: Applied Logistic Regression. Wiley, New York (2000)
14. McCullagh, P., Nelder, J.A.: Generalized Linear Models, 2nd edn. Chap-man & Hall/CRC, Boca Raton (1989)
15. Powers, D.A., Xie, Y.: Statistical Methods for Categorical Data Analysis. Academic Press, San Diego (2000)
16. Vidal, C., et al.: Evaluation of the phadiatop test in the diagnosis of allergic sensitization in a general adult population. *J. Investig. Clin. Immunol.* **15**(2), 124–130 (2005)

# Complex Data Management in MRI Results Processing

Michal Kvet, Monika Vajsová and Karol Matiaško

**Abstract** Cancer is one of the most serious problem of current medicine. Effective diagnostics and proper treatment brings the possibility for patients to become healthy. Global management of MRI results—processing, visualizing is the main part of the developed project. Moreover, these data must be stored effectively to monitor the progress over the time. This document describes the principles of processing, detecting and storing data in column level temporal system with emphasis on index structures.

## 1 Introduction

Cancer is one of the most serious problems of current medicine. Its severity affecting primarily the patient's life, but also its surroundings and society. The aim of many medicine professionals is to improve diagnostic and treatment methods. Adequate treatment of early diagnosed disease gives patients the possibility of becoming healthy, it improves the quality of life and increases the possibility to return to normal life and, of course, reduce the costs of treatment.

The field of radiology has been changed using enormous technical development not only devices using ionizing radiation during recent years. Magnetic resonance imaging is one of the diagnostic methods that provide two-dimensional and three-dimensional images of organs of the human body. It is unique and increasingly used primarily for its ability to create quality images without the use of radiation, respectively to show the structure of the human body that other methods have failed to describe sufficiently. One of the important factors that significantly affects the

---

M. Kvet (✉) · M. Vajsová · K. Matiaško

Univerzity of Zilina, Zilina, Slovakia

e-mail: Michal.Kvet@fri.uniza.sk

M. Vajsová

e-mail: Monika.Vajsova@fri.uniza.sk

K. Matiaško

e-mail: Karol.Matiasko@fri.uniza.sk

treatment process is based on the body's response to treatment. It is therefore necessary to monitor the effect of treatment over time. For this reason it is necessary to store individual MRI results to allow comparison of the images.

However, the apparatus performing MRI results is still not a common part of all hospitals all over the world, patients come from wide areas and data transfer is complicated based on the complete structure and size. Following text deals with the brain tumour detection methods based magnetic resonance imaging, describes principles as well as visualisation methods and structures for storing the results in database. It brings new approach for data over time modelling and managing. Temporal database focused on historical and actual data processing is the root for the development. The most important task is to find the area of interest—detect anomalies or tumour, allowing reducing the amount of data stored. Thus, only areas with detected anomalies, respectively areas with increased values of processed markers are stored and later managed and monitored. Significant part of the measured data is not necessary to store and deal with, because there is no tumour cells at all.

## 2 Diagnostic Methods

Brain tumours are usually diagnosed after symptoms damaging brain functions. The basic diagnostic methods are imaging methods: computer tomography (CT)—based on the principle of X-rays, positron emission tomography (PET) based on the principle of monitoring increased sugar metabolism by tumour cells. Magnetic resonance imaging (MRI) is the diagnostic method used to create two-dimensional and three-dimensional images of the organs. The method has been used in health care since 1980, it is a unique and increasingly used mostly for its ability to create high-quality images without the use of radiation.

The other tests are neurologic, ear, eye exams, blood tests, examination of cerebrospinal fluid and other specialized tests. Definitive diagnosis can be made by stereotactic neurosurgical techniques—fine needle biopsy under CT guidance for the purpose of histological examination [13, 15].

## 3 Physical Principles of MRI

Magnetic resonance imaging is based on the principle of measuring changes in the magnetic moments of nuclei of elements with odd atomic number placed in a strong static magnetic field after application of radiofrequency pulses. Due to the rotation atom nuclei around axes (spin), the core magnetic field (magnetic moment) is formed. Hydrogen atom contains one proton in the nucleus and in the tissues as abundant. If the examined tissue is put in a strong magnetic field, layout proton spins in one direction. In practice, there are two opposite directions, one of them predominates—so the resulting magnetic moment is therefore in one direction. In this state, the magnetic

moment of protons takes place two types of movements—both rotates around axes (spin) while moving around an imaginary cone shell, which is called precession. If the radiofrequency pulse (electromagnetic waves in the radio band short wave) is applied in the tissue with a frequency consistent with the proton precession frequency, resonance occurs according to the principle of magnetic moments deflection from the original direction by a certain angle and also synchronizes all proton precession. After the pulse occurs, everything returns to its original state. Time for that is known as relaxation time.

The time required to return biased magnetic moment is known as  $T1$  relaxation time, “get out of sync” precession is known as  $T2$  relaxation time. Both depend primarily on the composition of the material studied in the vicinity of protons. Signal, which is obtained after a series of different radio pulses, has the same character—it is electromagnetic wave that can be registered by receiving coils (basically antennas) and measures its size. A series of pulses required to obtain a measurable signal is referred as a sequence. Proton resonance frequency depends on the intensity of the external magnetic field. If this field gradient coils with appropriate modifications, we can obtain information about the location from which the signal arrives.

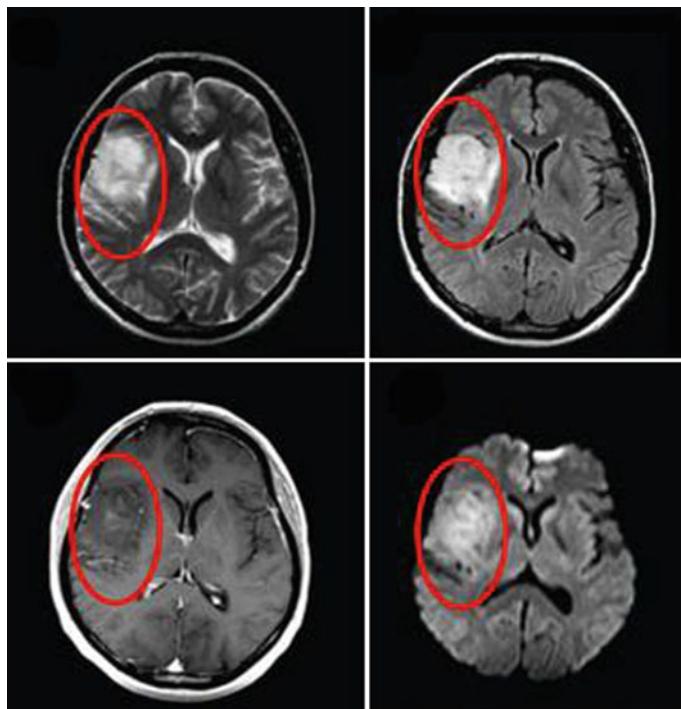
The principle of MRI is based on the measurement of magnetic moments of nuclei of hydrogen atoms in the water and fats.

During the patient examination layer images are made using various types of sequences informing about the difference of relaxation times  $T1$  and  $T2$  ( $T1$  or  $T2$ -weighted sequence) or the number of protons (weighted sequence by proton density). It is necessary to compare the signal strength of the same place in different types of sequences to get the complete information. The examination takes 20–60 min demanding on the speed of particular device and the number of displayed sequences. It is most often used in neuroradiology, to display diseases of the brain, spinal cord and spine. In addition, it can be used to diagnose musculoskeletal disorders; it can display all the components of bone and bone marrow by ligaments, tendons, muscles and cartilage. Increasingly, MRI is also used to view the organs in the abdomen, chest and neck. The advantage of MRI is more detailed view of soft structures, the ability to view in any plane, and the absence of radiation. MRI cannot be used for patients with pacemakers; implanted metallic materials in the body complicate scanning due to possibility of heating or displacement of implanted material [3, 14, 16, 17].

## 4 Intracranial Tumours

Intracranial tumours are a diverse group of tumours that differ in localization, symptoms, histological composition. The occurrence of each species depends primarily on age. Impaired functions of the brain and brain nerves and even increased intracranial pressure are the most common symptoms of the brain tumour, in general—Fig. 1.

The most common symptoms are called paretic symptoms—limb movement disorder, numbness, vision, speech or mental changes. Another group of symptoms are the symptoms resulting from local brain tissue irritation manifested as different



**Fig. 1** Brain tumour [17]

types of seizures. Syndrome of increased intracranial pressure is referred to a set of symptoms, which include mainly headache, vomiting and visual disturbances [13, 15, 17].

## 5 Tumour Markers

Tumour marker is a substance, usually protein, the occurrence of which indicates the presence of cancer. This substance can be almost exclusively produced by the tumour cells, thus the healthy cells do not produce them, but the markers can also be part of the normal cells, the presence of which in abnormal quantity or time also signifies the tumour. In general, we can observe the presence of these markers either within the tissue, which can distinguish between normal and damaged tissue or in body fluids (e.g. serum, urine, cerebrospinal fluid), in which these substances are released from the tissues. We use the term marker for a substance that occurs directly in the tumour tissue for this purpose. Markers in this sense may be either protein, or cell surface proteins that are part of the cellular metabolism (such as enzymes, hormones). Most tumours are characterized by the presence of one or several more or less specific

markers. An overall picture of the production of these is captured by tracking the presence of multiple markers—although with a lower sensitivity and specificity.

Searching for cancer in the population by monitoring the presence of markers in the blood is not used, because the tests have a relatively high sensitivity but lower specificity. The importance of monitoring the presence of tumour markers on the contrary is essential for patients with proven presence of malignancy. At the time of diagnosis is important to know the values of markers in the blood and in the tumour tissue.

## 6 Input Data Stream

Standard input application file is CSV (Comma Separated Value) file for each patient. Each file contains 58 columns—Fig. 2. The first column contains the ID of the row, the other three columns define the position of the measured element in three dimensional space ( $x$ ,  $y$ ,  $z$ ). The remaining 54 columns contain values of markers (*Ala*, *Cho*, *Gln*, *Glu*, *Ins*, *Lac*, ...). The treatment processing is not uniform in the location, so for each marker position is also stored error value (*Ala\_SD*, *Cho\_SD*, *Gln\_SD*, *Glu\_SD*, *Ins\_SD*, *Lac\_SD*).

Figure 3 shows the structure of the input file.

Another data export approach is based on the XML file—Fig. 4.

Standard data manipulation is therefore based on the file systems. If there is requirement to monitor the progress over the time, all data valid between the defined

ID	Position	Marker values	Marker error
----	----------	---------------	--------------

**Fig. 2** Input data file—one row definition

ID	x	y	z	Ala	Ala_SD	Cho	Cho_SD	Cr	Cr_SD	Gln	Gln_SD	Glu
0	0	0	0	51,3039	46,8067	27,7085	2,1205	28,6834	19,5322	15,5723	16,6582	52,8449
1	0	0	1	36,5340	35,0655	27,2899	18,2572	85,4350	10,5313	5,5565	7,3897	84,4359
2	0	0	2	58,4186	35,7065	62,3876	5,0895	33,4464	14,6610	46,0266	0,1893	85,0778
3	0	0	3	48,4223	23,2602	76,1001	13,7149	33,1611	31,8647	1,0293	0,3090	29,5363
4	0	0	4	48,6166	18,2845	37,0915	14,0463	19,4888	22,4259	89,6615	18,9182	15,5740
5	0	0	5	51,9815	5,2764	22,5499	18,1794	49,6230	30,8967	20,9275	13,1466	97,6898
6	0	0	6	53,0866	39,5707	7,6320	9,2921	64,0502	44,2271	16,8434	16,4744	79,6427
7	0	0	7	23,1725	31,7302	66,1116	19,1768	54,3023	46,3964	65,5412	10,0933	9,0380
8	0	0	8	50,5113	5,2766	50,7428	12,4796	10,7679	45,4544	77,4557	3,7255	43,9231
9	0	0	9	53,7951	6,4782	56,4071	20,9098	97,1896	2,4305	13,9624	2,3828	63,1823
10	0	1	0	19,8978	19,0205	98,1441	13,7131	36,7476	21,4265	63,3130	18,5634	82,4009
11	0	1	1	35,3263	40,6950	53,4020	12,1414	3,1538	7,1546	94,3463	11,9663	32,2252

**Fig. 3** Input data file

**Fig. 4** Input data file—one data row definition using XML

```

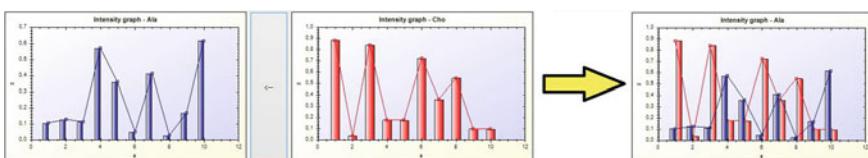
<MRI_data date="01-02-2014">
    <data_row id="0" x="0" y="0" z="0">
        <marker_Al>
            <value> 44.6508 </value>
            <error> 9.8641 </error>
        </marker_Al>
        <marker_Cho>
            <value> 84.9545 </value>
            <error> 11.1195 </error>
        </marker_Cho>
        <marker_Cr>
            <value> 10.3121 </value>
            <error> 35.4200 </error>
        </marker_Cr>
        ...
    </data_row>
    ...
</MRI_data>

```

time frames are loaded. Thus, it is necessary to store all the images. Newer approach is based on the database—one row contains the all measured data of the patient at defined time and position. Thus, it is possible to select only required data. This approach is based on row level temporal system. In term of time, it is the significant improvement, but the disc space requirements remain unchanged. Moreover, if some markers were not measured, the database would contain many undefined values.

## 7 Working Modes

The apparatus consists of the three main working modes—graphs section, two-dimensional model and three-dimensional model. Moreover, these data models can be compared over the time to see the progress of the treatment and reaction of the patient's body. Standard approach consists of two markers monitoring, one can be specific for the disease or can be used as the calibration marker. Graph section—Fig. 5—monitors the changes in one layer. Using the fusion of two markers creates possibilities to see the differences and comparisons.



**Fig. 5** Intensity graph—actual and default marker

## 8 Two-Dimensional (2D) Model

2D model section consists of four basic display parts important for user consideration:

- Values of the first marker in the selected two-dimensional slice with the appropriate values of the error of the measurement in that location.
- Values of the second marker in the selected two-dimensional slice with the appropriate values of the error of the measurement in that location. It can be referential marker for the specific tumour type.

All these data are projected in the specific datagrids and are colourly displayed based on the values. The rate of the colour differences between the intervals and categories can be set in the Colour module, where the number and borders of the intervals are chosen. The default setting is a spectrum of colours from white to red for marker values and a range from the green to red for measurement error influenced by the circumstances and also by the used MRI device—Fig. 6.

There are two possibilities, how to perceive the error of the measurement reflected to the values of the markers in the datagrids. The difference is in the method, which displays the values. If the value of the measurement error at a given voxel exceeds the maximal limit, we accept, the shown value can be blue indicating the potentially incorrectly measured value or the original value obtained on the basis of color intervals can be shown. In this case, we do not consider the data processing to reduce measurement error. If we need to reduce the value of the measurement error, the module for error processing is used. The display mode can be directly linked to the acquired images (DICOM data) got during the MRI scan to create interactive way of evaluating and displaying the anomalies—Fig. 7.

The selected position in the grid is shown after clicking on any cell (blue arrow in the Fig. 8) or interactive model. At the same time, the user gets the information about the value of the selected marker at that position (orange arrow in the Fig. 8), inclusion to the interval (green arrow—Fig. 8) and also the value of the error (gray arrow—Fig. 8). If the marker value is higher than the limit (critical value), exclamation mark is shown representing the location with extremely high value of the marker.

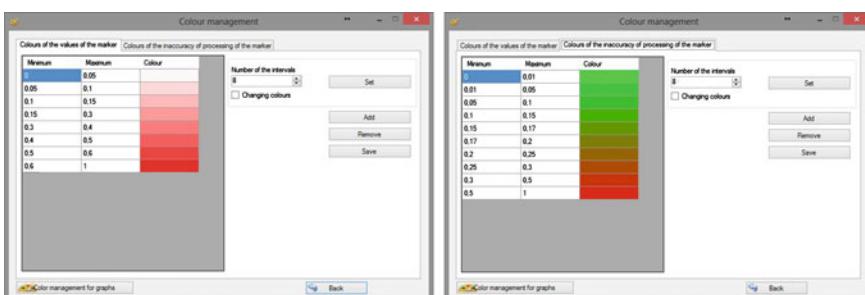
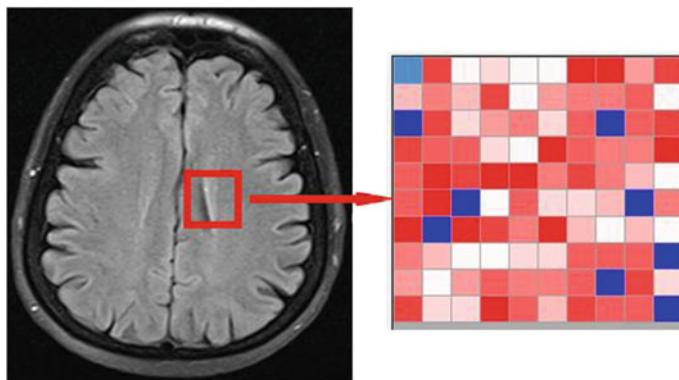
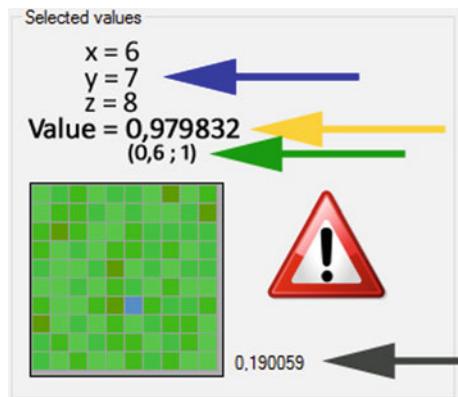


Fig. 6 Color spectrum



**Fig. 7** 2D model and the slice of the brain

**Fig. 8** Marker values

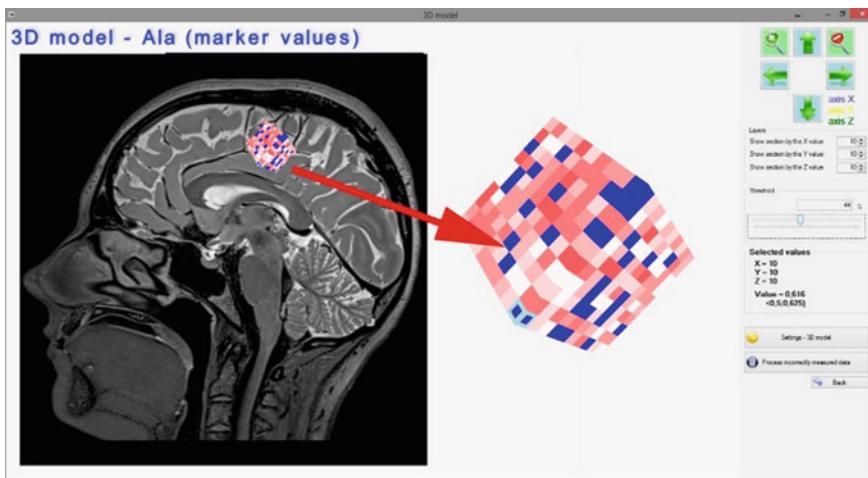


## 9 Three-Dimensional (3D) Model

3D model is based on picture recreation from the DICOM data models and images and processes marker values with highlighting anomalies—Fig. 9. Anomaly does not strictly determine tumour, but indicates uncommon values of markers in the quantity or time position.

Most important parameters and attributes of the application are shown in Fig. 10. Critical value—expresses the maximum value of the marker, which is still considered to be acceptable and is not considered as a potential anomaly.

Tolerance value—expresses the maximal acceptable difference between the different results of the MRI scan over the time. This value is used to monitor tumour growth (e.g. a comparison before and after the surgery). Threshold value—expresses the maximal value of measurement error, which is still accepted by the user. If they want to reduce the value of the error, it can calculate the new value of the error marker and surroundings.



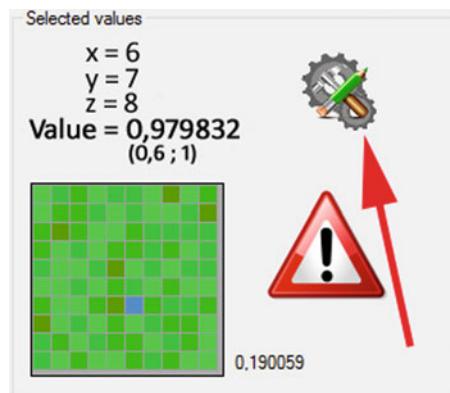
**Fig. 9** 3D model



**Fig. 10** Actual application settings

Measured values can have sometimes high measurement error and therefore it is possible and even necessary to replace them with new values, calculated from the neighborhood, based on specific properties, diffusion and other parameters. If the value is changed, this information is also stored (red arrow—Fig. 11).

**Fig. 11** Processed data



## 10 Error Processing

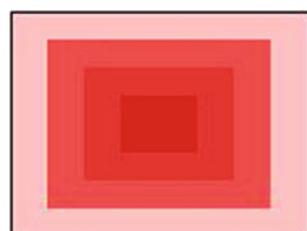
Data obtained from the magnetic resonance imaging examination of the patient may not have the required quality at each voxel. Even, in some locations, the marker value can be even undefined—that location was not measured at all or with the poor quality. Loaded data must be cleaned to show relevant data and to locate the anomaly, although there is no potentially risk area. Vice versa, there can be also the other problem, that the data do not directly locate the anomaly or the tumour, although there is a problem. The user sets the threshold value, which indicates the maximal error (in percentage), which is accepted. All the values with the error exceeding the limit cannot be considered relevant. However, if the doctor needs to get the values also in those areas, our application offers the possibility of new values of the elements calculating. There are a lot of sophisticated methods to do this based on conditions, thresholds, relationships between the markers and so on. The easiest way of processing is to replace the value based on the surroundings—linear approximation—Fig. 12. When we realize the fact of the diffusion, the method is suitable for fast management.

Thus, each element has 26 elements (or less) with a width = 1. New value of the element is created by summing the values of all surrounded elements (specified by the width) and divided by the number of these elements. It is considered correct only if the value of the error is reduced (value of the error is calculated the same way as the new value of the element). There is important to emphasize that it is not necessary to reduce the error value below the threshold. The main point is that the error value is diminished. Following formula refers to the way of calculating the new value of voxels, sets  $A$ ,  $B$ ,  $C$  define the elements in the axes [11].

$$x_{ijk} = \frac{\left( \sum_{a=-1}^1 \sum_{b=-1}^1 \sum_{c=-1}^1 x_{(i+a)(j+b)(k+c)} \right) - x_{ijk}}{|A| * |B| * |C| - 1} \quad \exists x_{(i+a)(j+b)(k+c)} \quad (1)$$

If the new value is defined by the wider surroundings, weights of the value elements are added to the formula (the closer element has the higher weight).  $w_{(i+a)(j+b)(k+c)}$  expresses the weight of the element  $x_{(i+a)(j+b)(k+c)}$ —the value of the element at the position  $(i + a)(j + b)(a + c)$ . The weight of the element  $x_{ijk}$  is 0, because it is considered to be untrustworthy.

**Fig. 12** Surroundings cells

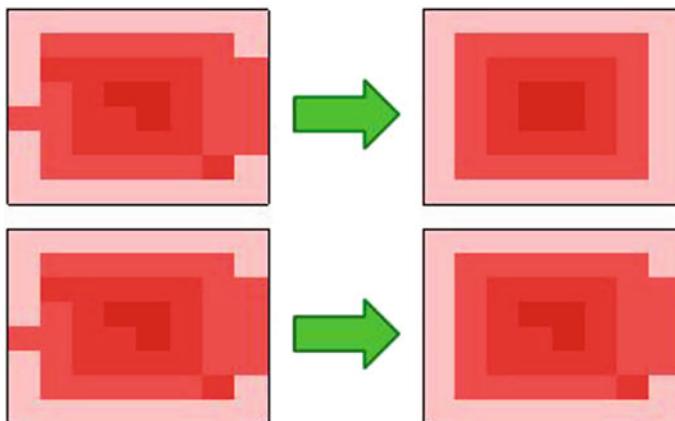


$$x_{ijk} = \frac{\sum_{a \in A} \sum_{b \in B} \sum_{c \in C} x_{(i+a)(j+b)(k+c)} * w_{(i+a)(j+b)(k+c)}}{|A| * |B| * |C|} \exists x_{(i+a)(j+b)(k+c)} \quad (2)$$

However, the diffusion does not model the values very well and in this field, it can be really dangerous to replace the value, which is not 100 % correct. The problem is just not completely uniform diffusion due to different structure and properties, that's why the proposed method is good, but must be corrected after the processing using the other specific methods. The problem is described—Fig. 13. The left part shows the real image of the brain slice, the top image in the right part shows the result after the linear approximation. The last one is more relevant, uses wide range of centres of the approximation. It also considers the relationships between the other markers. The correctness of these developed supporting techniques is about 80 % and these techniques are still improved and new ones are developed [7, 9].

Elements for processing can be selected manually from a table (datagrid), or all the elements according to the criteria are selected. The following figure shows the unprocessed data elements—their positions (position  $x, y, z$ ), current values and error values with the corresponding color category—Fig. 14.

After successful processing (successful processing means the reduction of the error value), the new values and info about processing are displayed—Fig. 15.



**Fig. 13** Real and processed image

x	y	z	Value	Error value	Process
1	1	1	0.059	0.286	Process
1	1	5	0.012	0.293	Process
1	1	6	0.464	0.441	Process
1	1	9	0.602	0.272	Process
1	1	10	0.338	0.299	Process
1	2	1	0.249	0.249	Process
1	2	2	0.211	0.499	Process
1	2	4	0.562	0.437	Process

**Fig. 14** Voxels before the processing

x	y	z	Value	Error value	Process
1	1	1	0.059	0.286	Process
1	1	5	0.012	0.293	Process
1	1	6	0.464	0.441	Process
1	1	9	0.602	0.272	Process
1	1	10	0.338	0.299	Process
1	2	1	0.249	0.249	Process
1	2	2	0.211	0.499	Process
1	2	4	0.562	0.437	Process

**Fig. 15** Voxels after the processing

## 11 Results Comparison Over Time

Tumour progression monitoring in time, as well as the reaction of the body and the tumour tissue to the treatment is a basic requirement. As already mentioned in the introduction, there are two possibilities for loading data, either the results of each examination are stored separately (CSV, XML) or all the results of the patient are stored together, in our case, it is a temporal database. If the patient is examined, their results can be added to this database to enable track the progress over time. Either all the measured data can be stored, or only those, that are in specified location, or only the values, the difference between actual and historical data exceeds the value ?. Doctor can compare these values of the marker to see the progress colourly displayed and rated.

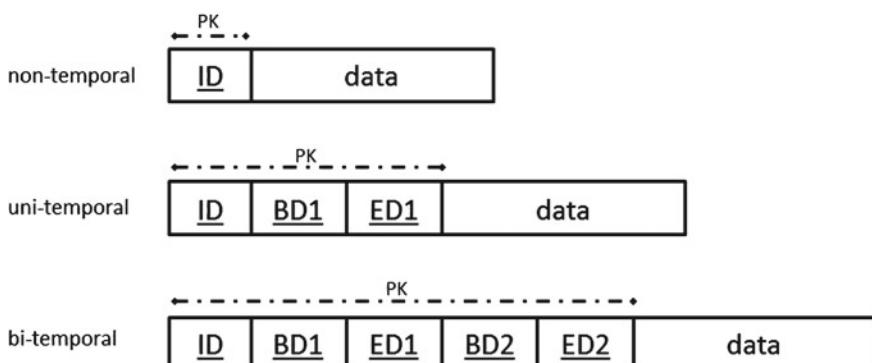
## 12 Results Comparison Over Time

Storing the results of magnetic resonance imaging in different spectrum with complete markers for a particular type of the tumour is not easy and requires sophis-

ticated methods for optimizing data and reducing disc space requirements. Storing each image in specific structures, whether in (conventional) databases or in files, is not appropriate based on the management and optimization. Moreover, synchronization of such structure and approach is very problematic. The main problem occurs, if there is necessity to share data, send them to a central repository (server), or send them to another hospital. Network requirements are in that case too high and the surgeon has to wait a long time until the required data are available. In addition, doctor needs to complete data for decision making without decreased quality. Temporal database defines new paradigm for data manipulation in comparison with the existing conventional systems. Rows with different values of the primary key (*PK*) can represent one object at different time frames. Data manipulation operators (in this case, it is mostly *Insert* statement) must specify not only the object itself (identifier of the patient), but also the processed time period. This approach must be therefore carefully monitored mostly by using procedures, functions and triggers. Moreover, the time definition can have different forms (interval definition, start time definition), which are in detail described in [1, 2, 5, 12].

Row in the temporal systems can be defined in three different ways using time, which are described in the Fig. 16. *ID* is a unique identifier; *PK* refers to a primary key. Uni-temporal table uses *BD1* and *ED1*—a pair of columns defining the beginning and end value of the period—validity. The extension of this model is bi-temporal, which uses also attributes *BD2* and *ED2*, to define the second time interval—transaction time. Transaction validity in this medical system can be expressed by processing measurement error on the basis of the values of the other markers as well as the results of the MRI over time. All these systems store the image in one row for the whole result of the processing. Therefore, they are commonly called object level temporal data [1, 2, 4, 10].

The question is how to manage these data? Is this approach effective? And what are the requirements?



**Fig. 16** Structure of temporal data file [8]

## 13 Requirements for Temporal System

System requirements can be divided into four groups with special aspects [7]:

1. Aspect of usability is based on the easy manageable methods. The aim is to provide easy access to the latest results of the MRI, as well as the results at any time in the past. Moreover, it should be possible to get changes based on the historical treatment and make projections for the future.
2. Aspect of performance is based on the correctness of the results—the reduction of measurement errors, respectively automatic replacement of incorrect data evaluation. Time management and results should be the same in performance rate and form when accessing actual or historical data.
3. Aspect of data structure—this is a special requirement of medical disc storage optimization based on the relevance definition. Only values in the monitored area are necessary to be stored. If the markers do not change their values or there cannot be certainly located the anomaly, there is no reason for storing and processing them. System focuses therefore only on the objects of interest, positions, where the anomaly is or can be located. Therefore, the Epsilon ( $\varepsilon$ ) value can be defined, which expresses the minimal change of values of markers in the defined area, which should be stored.  
If the value  $\varepsilon = 0$ , all the measured data are stored in the database.
4. Transaction management, which contains the error processed data, got using the other marker values, but also using timed data—MRI results over the time.

## 14 Classification for Temporal Databases

Current issues in this area have not use any classification for temporal model of databases. From this point of view we suggested classification rule in this form:  $\alpha/\beta/\gamma$  / where:

- $\alpha$  represents kind of DBS
  - N no database system support (e.g. file system only)
  - R Relational DBS (RDBS)
  - X Object relational DBS
  - O Object oriented DBS
  - U unspecified type of DBS
- $\beta$  kind of the temporal structure
  - N Non-temporal
  - U Uni-temporal
  - B Bi-temporal
  - M Multi-temporal
  - E Difference defined value of the attribute (Epsilon ( $\varepsilon$ ) temporal data)

- γ kind of transaction processing (Online transaction processing—OLTP)
  - N Nontransactional
  - L OLTP only with logs
  - O OLTP with temporal objects
  - A OLTP with temporal attributes of the object

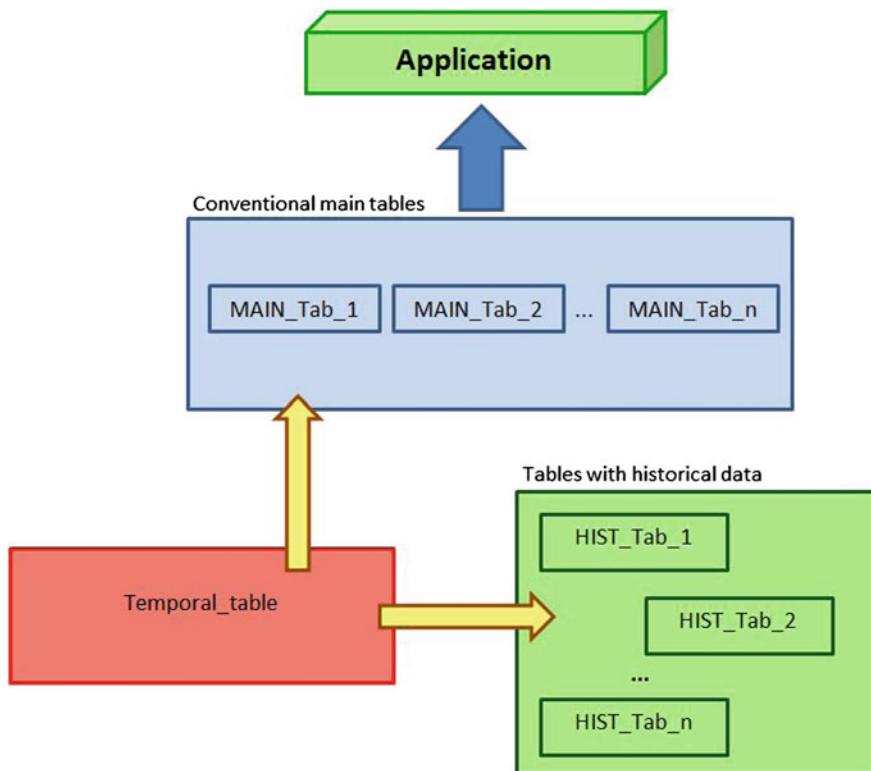
For the purpose of medical data processing, the best solution is R/E/A which represents temporal database using RDBS with the OLTP support in attribute level. In this meaning, transactions represent the error reduction processing.

## 15 Column Level Temporal System

Object level temporal system does not fulfil the performance requirement because of the duplicities. It is necessary to store the whole image regardless the values of the changes. Better way for modelling is therefore column level, which manages not the whole objects (one position with all the marker values), but only attributes (markers). If one or more attributes are changed, only the information is stored in the database. Moreover, if the data are stored in the database only if the change is located in the place of interest—potential anomaly, the size of the structure is significantly reduced.

Figure 17 shows the principle of the database management system. The basic part of the system is the temporal table that contains references to the changed attribute values in a defined time. This table also contains the reference to the previous valid value of the attribute. Thus, if the value is changed, information about the *Update* is stored in the temporal table and historical table is inserted into the table containing historical values. Each temporal column (marker, whose evolution over time is important to us) has its own historical table.

Figure 18 shows the structure of the model, which can be divided into three groups. The first one consists of the *Tab\_MRI\_results*, which includes not just the region of the interest, but a complete image of the brain. Primary key consists of the *ID* obtained from the input file. Non-temporal attributes consist of the positional data (*x*, *y*, *z*). Marker values themselves are stored in temporal columns, because these values are important to manage over time. The value of each marker is automatically associated with the error of the measurement and processing for the defined voxel. The second part is the *Temporal\_table*, which consists of the next mentioned attributes [6, 8, 10]—see also Fig. 18.



**Fig. 17** Temporal model for historical data processing [7]

- *ID\_change*
- *ID\_trans*—references the transaction validity of the marker value. Transaction validity is changed, if reduction of the measurement error is necessary.
- *ID\_previous\_change*—references the last change of the voxel identified by *ID*. This attribute can also have NULL value that means, the data have not been updated yet, so the data were inserted for the first time in past and are still actual.
- *ID\_orig*—carries the information about the identifier of the row that has been changed.
- *ID\_column, ID\_row*—hold the referential information to the old value of attribute (if the DML statement was *Update*). Only *Update* statement of temporal column sets not null value.
- BD the begin date of the new state validity of an object.

Last structural group contains a historical values of the attributes. Model shown in Fig. 18 contains only standard markers, but this concept can be extended by the definitions of the new marker values—it is necessary to extend only section that processes historical values (by adding new table for each marker) and also expands the table *Tab\_MRI\_results* by 2 columns for each new marker. Access methods and principles are described in [7, 8, 11].

Tab\_MRI\_results

ID	x	y	z	Ala	Cho	Gln	Ala_SD	Cho_SD	Gln_SD
A	1	1	1	Ala_1	Cho_1	Gln_1	A <sub>SD</sub> _1	Ch <sub>SD</sub> _1	G <sub>SD</sub> _1
B	1	1	2	Ala_2	Cho_2	Gln_2	A <sub>SD</sub> _1	Ch <sub>SD</sub> _1	G <sub>SD</sub> _1
C	1	1	3	Ala_3	Cho_3	Gln_3	A <sub>SD</sub> _1	Ch <sub>SD</sub> _1	G <sub>SD</sub> _1

Temporal\_table

ID_change_pk	ID_trans	ID_previous_change	ID_orig_NN	ID_column	ID_row	BD_NN
1	1	NULL	A	NULL	NULL	15.1.2013
2	1	NULL	B	NULL	NULL	15.1.2013
3	1	NULL	C	NULL	NULL	15.1.2013
4	1	1	A	Ala	36	15.7.2013
5	1	2	B	Ala	37	15.7.2013
6	3	4	A	Cho	75	1.8.2013
7	4	6	A	Cho	76	1.8.2013
8	5	7	B	Cho	77	16.8.2013
9	5	5	B	Gln	42	16.8.2013

Ala\_hist\_tab

ID	Value	Error
...	...	
36	Ala <sub>hist</sub> _36	E_Al <sub>hist</sub> _36
37	Ala <sub>hist</sub> _37	E_Al <sub>hist</sub> _37

Cho\_hist\_tab

ID	Value	Error
...	...	
75	Cho <sub>hist</sub> _75	E_Cho <sub>hist</sub> _75
76	Cho <sub>hist</sub> _76	E_Cho <sub>hist</sub> _76
77	Cho <sub>hist</sub> _77	E_Cho <sub>hist</sub> _77

Gln\_hist\_tab

ID	Value	Error
...	...	
42	Gln <sub>hist</sub> _42	E_Gln <sub>hist</sub> _42
43	Gln <sub>hist</sub> _43	E_Gln <sub>hist</sub> _43

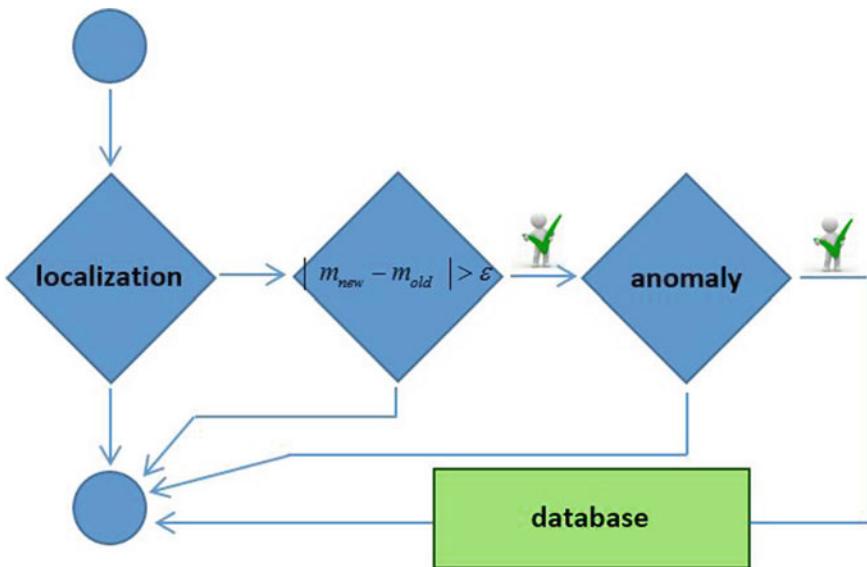
Fig. 18 Temporal structure (column level) [7]

## 16 Epsilon Column Level Temporal System

The solution for the actual and historical data management is column level temporal system. Voxel is not represented as one object (the whole markers definition), but each marker is separately defined with respect to the validity in time. One advantage of this solution is principle of separation the actual data from historical. It allows existing systems to operate without changes. Each marker value progress is pipelined list of changes by definition temporal validity. However, this system can be even improved by the definition of the minimal valid change of the marker value—Epsilon ( $\varepsilon$ ). The relevant data are only those, that characterize potential anomaly or the change of the marker (positive or negative) is significant. Epsilon ( $\varepsilon$ ) value in medical MRI systems is defined separately for each marker and can differ based on the location and position.

In this case, however, it is necessary to monitor also the sensitivity and specificity of markers to reduce the incorrect positivity. Figure 19 shows the flowchart of the Epsilon ( $\varepsilon$ ) temporal database. New values are inserted into the temporal database only if the (potential) anomaly is detected and the change of the marker value is relevant.

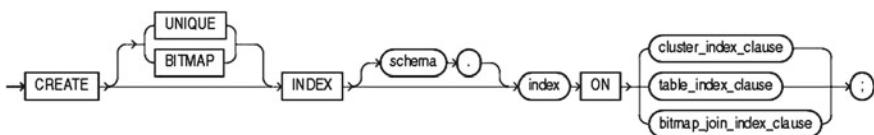
$$|m_{new} - m_{old}| > \varepsilon \quad (3)$$



**Fig. 19** Flowchart [8]

## 17 Index Structures

Database systems define an index as an optional structure associated with a table or table cluster that can sometimes speed data access—Fig. 20. By creating an index on one or more columns of a table, you gain the ability in some cases to retrieve a small set of randomly distributed rows from the table. Indexes are one of many means of reducing disk I/O. If a heap-organized table has no indexes, then the database must perform a full table scan to find a value. For example, without an index, a query for object with  $ID = 200$  in main table (*Tab\_MRI\_results*) with monitoring changes over time would require the database to search every row in every table block (*temporal\_table*) for this value. This approach does not scale well as data volumes increase [19].



**Fig. 20** Index definition [19]

Indexes are schema objects that are logically and physically independent of the data in the objects with which they are associated. Thus, an index can be dropped or created without physically affecting the table for the index [18].

The absence or presence of an index does not require a change in the wording of any SQL statement. An index is a fast access path to a single row of data. It affects only the speed of execution. Given a data value that has been indexed, the index points directly to the location of the rows containing that value. Database management system automatically maintains the created indexes—changes (*Insert*, *Delete*, *Update*) are automatically reflected into index structures. However, the presence of many indexes on a table degrades the performance because the database must also update the indexes. Moreover, the optimizer does not have to use them, if the full scan would be faster or if the index is not suitable for the query based on the conditions. Thus, if the user forces the system to use the index, the performance rate can be much worse than without their use. Each index has 2 properties—visibility (invisible index = maintained, but not used by the optimizer) and usability (unusable index = not maintained and not used by the optimizer). Information about the indexes can be found in `user_indexes` data table.

## 18 Experiments

The aim of the developed structure is to provide fast, reliable approach to the actual and historical data objects—data, which were valid in the past. Moreover, it should also be possible to get data after the *Delete* operation, thus, to manage also objects, which are not valid now, but were used in the past.

The optimization is based on the performance level—size of the whole structure and time to get required data (actual state of the whole database and the object changes monitoring over the time). In comparison with classical column level temporal system described in [8, 10], extended column level temporal system requires one extra block (64 kB) for `temporal_table`. However, the performance rate is much better.

This section deals with the index performance and size of the global structure. Eight indexes have been developed, which were compared to declare the quality and usability for this system. Notice, some attributes even decrease the performance of the system, although the optimizer is not forced to use them by hints. The system uses none or one developed index for `temporal_table` (Fig. 18):

- Ind1: System without indexes.
- Ind2: `ID_orig`.
- Ind3: `ID_orig`, `ID_previous_change`.
- Ind4: `ID_orig`, `BD`.
- Ind5: `ID_orig`, `ID_previous_change`, `BD`.
- Ind6: `UNIQUE ID_orig`, `ID_previous_change`.
- Ind7: `BD`, `ID_orig`.
- Ind8: `BD`, `ID_orig`, `ID_previous_change`.

Experiment results were provided using Oracle 11g. The parameters of used computer are:

- Processor: Intel Core i7-3612QM, CPU @ 2.10 GHz
- Operation memory: 16 GB
- HDD: 1 TB, 5400 rpm

The measured values are shown in Fig. 21. Although only operation *Insert* and *Update* is significant for medical data processing (medical results processing), all of the DML operations have been evaluated. All the time values are expressed in seconds.

Complete number of each operation was 10 000 (*Insert*, *Update*). Minimal number of operations on the object was 3, maximal number was 24, the average value was 5,4795 (Fig. 21).

Figure 22 shows the experiments results for DML operations—*Insert*, *Update* and *Select*. The operation *Select* is divided into three groups:

- Get object state for all objects (*Select*).
- Get object state for object with minimal changes (*Smin*).
- Get object state for object with maximal changes (*Smax*).

Figure 22 shows the performance rate on the graphs.

Another important parameter is the complex size of the database with all the objects and index. When using index *Ind2–Ind7*, there are required 32 blocks, if using index *Ind8*, there are required 48 blocks (one block is 64 kB).

The experiment results shows potential of the indexes *Ind2*, *Ind3*, *Ind4*, *Ind5* and *Ind6*, which have significant performance improvement in comparison with not-indexing system. Moreover, *Ind7* and *Ind8* do not provide any performance improvement.

## 19 Results

This project is medically oriented aimed on displaying and processing results of MRI, which can obtained by viewing the presence of certain specific markers in abnormal amount or time. The treatment of the patient is long-termed.

	Ind1	Ind2	Ind3	Ind4	Ind5	Ind6	Ind7	Ind8
Insert	8,7860	9,2713	9,8587	9,6187	9,3070	9,5007	9,5333	9,5753
Update	41,3800	18,6040	19,0387	18,6560	17,6423	18,6703	42,9513	42,6517
Delete	30,4880	12,7317	13,1843	12,3610	12,6097	12,4063	31,0890	31,6797
Restore	37,0390	14,0167	13,9707	14,1097	13,7083	14,3283	37,0180	37,2927
Select	49,2590	4,0610	4,0113	3,9500	4,0127	4,0317	97,5830	104,4483
Smin	0,0047	0,0007	0,0007	0,0003	0,0010	0,0007	0,0100	0,0113
Smax	0,0047	0,0010	0,0010	0,0007	0,0007	0,0007	0,0100	0,0110

**Fig. 21** Experiment results



**Fig. 22** Performance results

Monitoring the presence of the markers is important during diagnostics, during the treatment, too. Decreasing values of markers indicate good treatment effect. If the values do not change or are rising, it indicates the resistance of the disease to a particular treatment. Monitoring the presence of markers is used after cancer treatment, too. If the values begin to rise, it means that the number of tumour cells in the body is again increasing, suggesting either the tumour reappearance or the extension of the tumour in the form of metastases. Accordingly, it may proceed with treatment as soon as possible. Therefore, it is necessary to store the images and results over the time effectively with emphasis on size and access time. Temporal database approach based on column level offers good possibilities and is compared with the existing structures. Effective designed indexes described in the previous section improves the performance, too.

## 20 Conclusion

Cancer is one of the most serious medical problems worldwide. Early diagnostics and adequate treatment is a key role for treatment of the patient.

The aim is to create a universal solution for results of magnetic resonance imaging processing; to detect and locate brain tumours and provide output in the form of tables, graphs, 2D and 3D models based on user requirements. The aim of the application is to show the presence of markers—substances that are present in the tumour tissue. The application was designed to show the presence of the markers over the time. Therefore, it is necessary to define structures for complex medical data management.

Temporal management offers processing object valid data and their changes over the time. MRI results management requires complex management and access to the whole information about the patient results and body apostrophes reaction to the treatment. Temporal data management used today does not cover the complexity of the data management; it does not provide sufficient power to manage large volumes of medicine data. Therefore, the column level temporal system for these data has been developed. Epsilon ( $\varepsilon$ ) temporal structure offers storing only values, which have significant change. This approach is very important because of the data optimization requirement with a focus of data transmission and server management.

Time processing and measurement errors reducing require extension of the definition of this concept, which is based on results processing comparisons with emphasis on different markers, their relationships and dependencies. Thus, the transaction validity is used to model process of error reduction.

**Acknowledgments** This publication is the result of the project implementation: Centre of excellence for systems and services of intelligent transport II., ITMS 26220120050 supported by the Research & Development Operational Programme funded by the ERDF and is also supported by the project VEGA 1/1116/11—Adaptive data distribution. Center of translational medicine, ITMS 26220220021 supported by the Research & Development Operational Programme funded by the ERDF. Podporujeme výskumné aktivity na Slovensku/Projekt je spolufinancovaný zo zdrojov EÚ.



## References

1. Date, C.J.: Date on Database. Apress, Calif (2006)
2. Date, C.J., Darwen, H., Lorentzos, N.A.: Temporal Data and the Relational Model. Morgan Kaufmann, San Francisco (2003)
3. Hornak J.: The Basics of MRI, Interactive Learning Software. Henietta, New York (2008)
4. Jensen, C.S.: Introduction to Temporal Database Research
5. Johnston, T., Weis, R.: Managing Time in Relational Databases, Morgan Kaufmann, San Francisco (2010)
6. Kvet, M., Lieskovsk, A., Matiako, K.: Temporal data modelling, 2013. In: IEEE Conference ICCSE, pp. 452–459, 26–28 April 2013
7. Kvet, M., Matiako, K.: Epsilon temporal data in MRI results processing, 2014. In: IEEE Conference Digital Technologies, pp. 209–217, 9–11 July 2014
8. Kvet, M., Matiako, K.: Uni-temporal modelling extension at object versus attribute level, 2013. In: IEEE Conference EMS 2013, pp. 7–11, 20–22 Nov 2013
9. Kvet, M., Kvet, M., Matiako, K.: Application for brain tumour imaging. In: IEEE Conference IWSSIP 2014, pp. 47–50, 12–15 May 2014
10. Kvet, M., Matiako, K., Kvet, M.: Transaction management in fully temporal system, In: IEEE conference UKSim 2014, pp. 147–152, 26–28 March 2014
11. Kvet, M., Meina, J., Matiako, K.: Algorithm for brain tumour detections. Acta Electrotechnica et Informatica **12**(2), 45–50 (2012)

12. Mat, J.: Transformation of relational databases to transaction-time temporal databases. In: ECBS-EERC, pp. 27–34 (2011)
13. Nekula, J.: Radiologie. Univerzita Palackho, Olomouc (2005). ISBN: 024410117
14. Pianykh, O.: Digital Imaging and Communications in Medicine. Springer, Berlin (2008). ISBN: 978-3-540-74570-9
15. Zeman, M.: Speciln Chirurgie. Galn (2006). ISBN: 8072622609
16. <http://www.mri-portal.com/verejnost.php>
17. [http://www.solen.sk/index.php?page=pdf\\_view&pdf\\_id=4455&magazine\\_id=10](http://www.solen.sk/index.php?page=pdf_view&pdf_id=4455&magazine_id=10). Accessed 1 Sept 2014
18. Oracle documentation [http://docs.oracle.com/cd/E11882\\_01/server.112/e25789/indexiot.htm#CNCPT721](http://docs.oracle.com/cd/E11882_01/server.112/e25789/indexiot.htm#CNCPT721). Accessed 1 Sept 2014
19. Oracle documentation [http://docs.oracle.com/cd/B12037\\_01/server.101/b10759/statements\\_5010.htm](http://docs.oracle.com/cd/B12037_01/server.101/b10759/statements_5010.htm). Accessed 1 Sept 2014

# Biometrics for Biomedical Applications

Kenneth Lai, Steven Samoil, Svetlana N. Yanushkevich  
and Gaetan Collaud

**Abstract** This chapter focuses on the emerging applications of biometrics in biomedical and health care solutions. It includes surveys of recent pilot projects, involving new sensors of biometric data and new applications of human physiological and behavioral biometrics. It also shows the new and promising horizons of using biometrics in natural and contactless control interfaces for surgical control, rehabilitation and accessibility.

## 1 Introduction

Biometric technologies emerged over 30 years ago for the needs of forensic sciences, as well as for automated security access, and is defined as a means to identify or verify humans by their physiological or behavioral traits. Since then, it has paved the way to a vast variety of new applications, such as everyday personal device access, fitness monitoring, health care patient monitoring. An example is the Spree, a high-tech fitness headband, that performs body temperature measurement, heart rate monitoring and movement tracking. It aims to capture real-time data in order to optimize every workout and accurate performance measurement in one single

---

K. Lai (✉) · S. Samoil · S.N. Yanushkevich

Department of Electrical and Computer Engineering, Biometric Technologies Laboratory,  
University of Calgary, 2500 University Drive NW, Calgary, AB T2N 1N4, Canada  
e-mail: kelai@ucalgary.ca

S. Samoil  
e-mail: ssamoil@ucalgary.ca

S.N. Yanushkevich  
e-mail: syanshk@ucalgary.ca

G. Collaud  
École d'Ingénieurs et d'Architectes de Fribourg, Haute école spécialisée de Suisse Occidentale,  
Bd de Pérolles 80, Box 32, 1705 Fribourg, Switzerland  
e-mail: gateancollaud@gmail.com

and easy-to-use device [49]. Also recently introduced is a smartphone-enabled ECG heart monitor AliveCor, which is a build into a “skin” for the iPod [2].

An example of analysis of human biometrics is remote monitoring of patients’ physiological and behavioral patterns in hospitals or care units [43]. Another example is a gesture-controlled environment in a hospital, that allows a nurse or a surgeon to scroll a patient’s file on the monitor without removing their gloves or touching a screen or a keyboard for sanitary reasons [21]. Another example is enhanced accessibility for a wheelchair patient that can use the movement of their face to control a cursor on the screen of his or her personal device [24].

Another fast developing area is biometric-based natural interfaces, controlled by gestures, gazes and facial expressions [29, 35]. The inverse task of analysis of biometrics, modeling, or simulation, of biometric data, is used for applications such as augmented reality and virtual environments. Synthetic biometric data is generated, in particular, when a doctor needs to model the outcomes of a surgical procedure, or to be trained using 3D graphical model of parts of the human body. The related aspects of virtual reality are useful for rehabilitation purposes, for example, for patients that are placed in a virtual environment for body movement training in physiotherapy.

## 2 Biometric Data

Biometric data, or biometrics, include visual, infrared, and acoustic data for identification of both physical appearance (including aging and intentional changes, or camouflage), physiological characteristics (temperature, blood flow rate), and behavioral features (voice) of humans. Other examples of biometrics are odor, gait, keystroke pattern, signature, fingerprints, ear, iris, and retina. Gait, in particular, is used to determine the gender of a walker, pregnancy, affliction of the legs or feet, and drunkenness [16]. Electro-Cardiogram (ECG) signal is another biometrics that is used in human biometrics monitoring in health care [1].

The IR spectrum provides information to recognize certain skin diseases [20, 56], emotion [51], psychological features and gender [16]. For example, a high temperature spot in the IR image of the face, caused by Severe Acute Respiratory Syndrome, are important to detect early for further immediate isolation of those infected, to prevent another outbreak of such an epidemic. The thermodynamic relation between the blood flow rate at the skin level, blood temperature at the body core, and the skin temperature is used to convert IR intensity to temperature, and, therefore, to the blood flow rate. Also, the breathing rate can be estimated during the dialog of the patient or user with personnel [44].

In this chapter, we focus on facial biometrics, as well as gesture biometrics, captured by using multi-spectral sensors.

### 3 Facial Biometrics in Biomedical Applications

Apart from application of facial biometrics in security systems, where facial recognition is performed for the purposes of human identification or verification, other information, acquired from facial images in various spectra, find its application in the biomedical domain. Examples include face recognition for people with “face blindness” (prosopagnosia), evaluation of facial expressions for pain detection or mood analysis, and facial temperature estimation (from IR facial images).

#### 3.1 Face Recognition

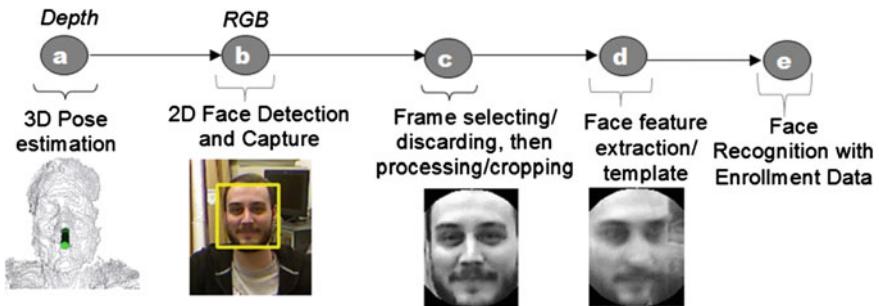
Facial recognition technology is a well-developed domain. Various applications of this technology include intelligent surveillance, user identification or verification, and interview supporting. The most recent applications include multimedia, such as recognition of faces in photo and on-line social group databases. The most advanced, recent example is an under-development wearable computer technology such as eyeglasses, by the industry giant Google. It proposes to use existing portable devices, such as a smart phone or tablet, and combine it with a head-mounted display system. It intends to assist anyone who are affected by poor “face memory” (most of us) or simply want to excel in remembering faces. The collection of a face database can be accomplished during the very first meeting with the acquaintances. The secondary important application of such device is to use it as a rehabilitation tool for people affected by a neurological disorder that has impaired their ability to recognize familiar faces (prosopagnosia). Since prosopagnosia is suggested to affect up to 2 % of the population, the development of a face recognition system tailored to patients with prosopagnosia is a priority in the field of clinical neuroscience [11].

“Soft” facial biometrics, such as expression, can help improve the face recognition applications which become confused when identifying the same person smiling, aged, with various accessories (moustache, glasses), and/or poorly lit. Facial recognition tools can be improved by training on a set of generated facial expressions and appearance/environment variations generated from real facial images [17].

#### 3.2 Face Recognition in RGB-D Spectra

Since the development of the inexpensive RGB-Depth (RGB-D) sensor Kinect for Windows [35], many machine vision projects have used these devices for human face and gesture recognition.

RGB-D sensors are a unique tool that are consistently used to capture both texture (RGB) and shape (depth) information. One of the many applications is to use depth data to create a model to perform head pose estimation. An instance of



**Fig. 1** Sample scheme of using depth and RGB for face recognition

combining depth for pose estimation and texture for face recognition provides an optimal performance in time and accuracy. Head pose estimation operates using only depth information and actively prunes images that are not frontal views; while, face recognition are performed on specific frontal view face images providing both higher accuracy in recognition and reduce in time to perform a recognition task.

Figure 1 represents an application that uses depth data along with RGB to perform face recognition [25]. Using the Kinect sensor, the video data stream is split into RGB and depth components; depth is utilized solely for pose estimation, and RGB is needed for texture analysis in order to perform face detection and recognition.

Based on the scheme shown in Fig. 1a, the depth data is accessed by a head pose estimation algorithm. The project [25] utilized, in particular, a random regression forest algorithm, based on work by Fanelli et al. [15]. A head pose algorithm operates using pre-trained data to create a generic face model that is then attempted to be fitted to a depth model. Depending on the fitting, a pose vector is created from the center of the head towards the nose. The pose vector indicates where a head is facing. A frontal view can be established by approximating the relations between the pose vector and the origin of the camera. Figure 2 displays the head pose vector.

Following the schema, the pose vector is then used to approximate the frontal view. Figure 3 represents the selected frontal view images based on the pose estimation.



**Fig. 2** Head pose algorithm, green cylinder represents the pose vector (color in online)



**Fig. 3** Sample faces with detection frontal views; the frames with crossed out faces failed frontal view detection

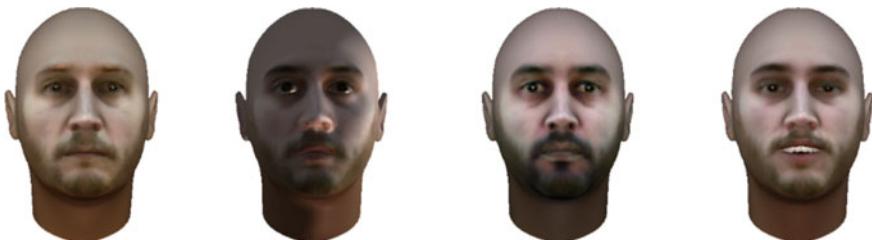
The selected frontal frames are passed on to face detection and capture (Fig. 1b). For each frontal frame, a face detection algorithm is applied to extract faces (Fig. 1c). Facial features are extracted from each face in step Fig. 1d. Finally, in Fig. 1e, face recognition is performed on the extracted facial features.

Overall, applying a depth-based pruning method greatly reduces the number of frames the face recognition and face detection requires. In addition, face recognition accuracy increases, because using frontal view frames are beneficial for most face recognition algorithms.

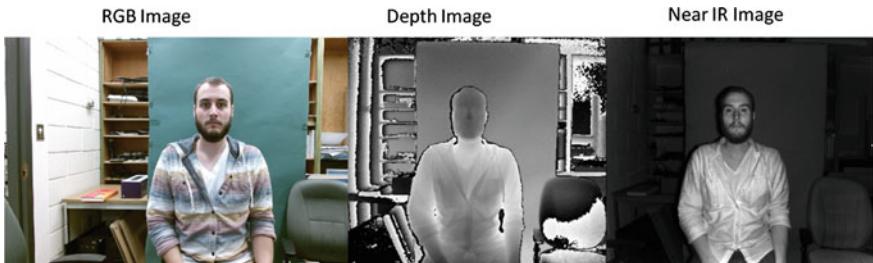
### 3.3 Face Modeling

The reverse process to face analysis is face reconstruction, a classic problem of criminology, but also of reconstructive medicine, or surgery.

A 3D face model includes two constituents: a face shape model (represented by a 3D geometric mesh) and a skin texture model (generated from 2D images). The main advantage of 3D face modeling is that the effect of variations in illumination, surface reflection, and shading from directional light can be significantly decreased. For example, a 3D model can provide controlled variations in appearance while the pose or illumination is changed (Fig. 4).



**Fig. 4** Modeling of aging, drunk, a poorly lit face, and facial accessories



**Fig. 5** Sample images of the different possible Kinect outputs

Also, the estimations of facial expressions can be made more accurately in 3D models compared with 2D models.

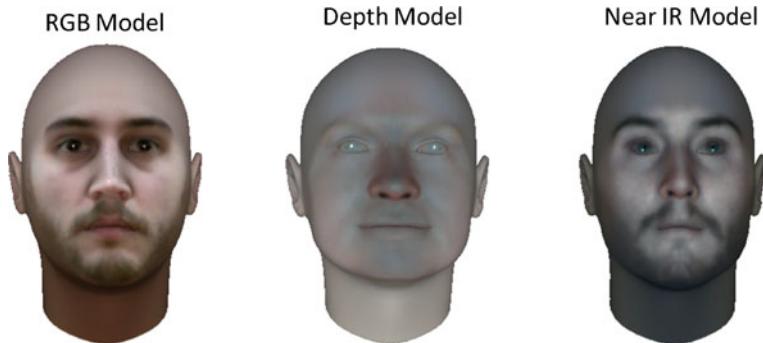
A face shape is modeled by a polygonal mesh, while the skin is represented by texture map images. Any individual face shape can be generated from the generic face model by specifying 3D displacements for each vertex. Synthetic face images are rendered by mapping the texture image on the mesh model.

This approach can also be used to assist modeling faces in visible and infrared bands (Fig. 5) [4]. The texture maps can model the hemoglobin and melanin content (in visible spectrum) and the temperature distribution of the facial skin. These maps are the output of the face analysis and modeling module. This information is used for evaluating the physical and psycho-emotional state of a person.

### 3.4 Modeling Assisted with RGB-D Data

The most recent approach to 3D modeling of faces takes advantage of advances in sensor technologies: 3D modeling of faces can be done using a variety of different image types including infrared, near infrared, depth, and RGB images. Using the Kinect v2, infrared, depth, and RGB images can be obtained simultaneously. Examples of the three types of images taken in the Biometric Technologies Laboratory at the University of Calgary are included in Fig. 5.

RGB images are the normal color images that can be taken with almost any camera. Near infrared images are taken using the light spectrum that is almost at the infrared point. This allows the camera to see in the dark allowing for usage of motion tracking and other features in a dark environment. The Depth image is generated by creating a 3-dimensional map of the room and then showing the distance to the object as different shades of grey. Using these images and modeling software, such as the FaceGen Modeler software [14], a 3D model of the photo subject can be generated. The results of this are shown in Fig. 6.



**Fig. 6** Models created using the different possible Kinect outputs as input to FaceGen

### 3.5 Modeling the Aging Face

A survey of age synthesis and estimation techniques was conducted in 2010 by Fu et al. [18]. One of the methods of age synthesis discussed by the authors was a paper written by Lanitis et al. on creating automatic ways of simulating aging faces. The method used by Lanitis et al. was to have a set of training images for various different ages to create a model representing the ages. To age an image of a face the simulation technique requires that the age of the image is known. Using the known age the image can be modified to represent the new age using the model created from the training images [28]. Figure 7 shows an example of modelling the aging process using FaceGen. Previous research has also been conducted into age estimation by Fu and Huang [19] and also by Guo et al. [23].

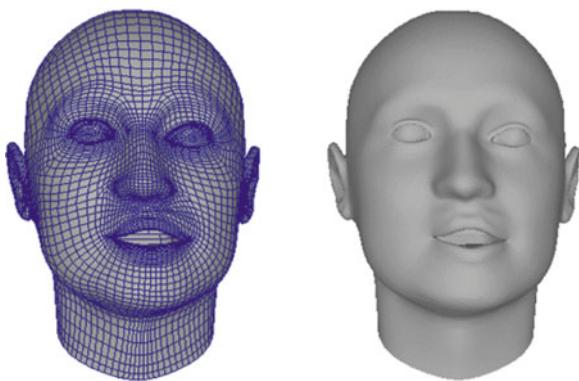
### 3.6 Facial Expression Analysis and Synthesis

Facial expressions are formed by about 50 facial muscles [12] and are controlled by dozens of parameters in the model. The facial expression can be identified once the



**Fig. 7** Aging modeling (neutral facial expression) using the FaceGen application

**Fig. 8** An example picture is given here using data from [31] of a 3D model (generated using FaceGen) showing a person with Right Bell's Palsy and a surprised expression



facial action units are recognized. This task involves facial feature extraction (eyes, eyebrow, nose, lips, chin lines), measuring geometric distances between the extracted points/lines, and then facial action units recognition based on these measurements.

Psychologists distinguish two kinds of short-term facial expressions: *controlled* and *non-controlled* facial expressions [41, 42, 51]. Controlled expressions can be fixed in a facial model by generating control parameters, for example, a type of smile. Non-controlled facial expressions are very dynamic and are characterized by short time durations. Visual pattern analysis and classification can be carried out in 100 msec and involves a minimum of 10 synaptic stages from the retina to the temporal lobe (see, for example, [45]).

In [31], 3D models were used as a way to model facial nerve disorders such as Bells Palsy. This was done by creating a 3D model and then editing the mesh. By changing certain points of the mesh the 3D model could be changed so that it was very similar to the face of a real person with the condition. In the project, originally done by [31], the models were created using the FaceGen Modeler SDK. Using the SDK allowed access to the underlying polygons and their vertices. Shifting specific vertices allowed the creation of a 3D model that closely represented different facial nerve disorders. The ability to move parts of the face allows quite a bit of freedom in the ability to simulate many different situations [31]. An example of the model and the underlying mesh for Right Bell's Palsy can be seen in Fig. 8.

## 4 Gesture Biometrics in Biomedical Applications

Besides facial gesture, body and hand gesture analysis falls into domain of behavioral biometrics analysis. Examples include gesture controlled interfaces, sign-language reading, as well as analysis of body posture or gait.

#### 4.1 Gesture-Based Control Based on RGB-D Sensor

3D sensors can be used for developing a gesture controlled interface on a computer.

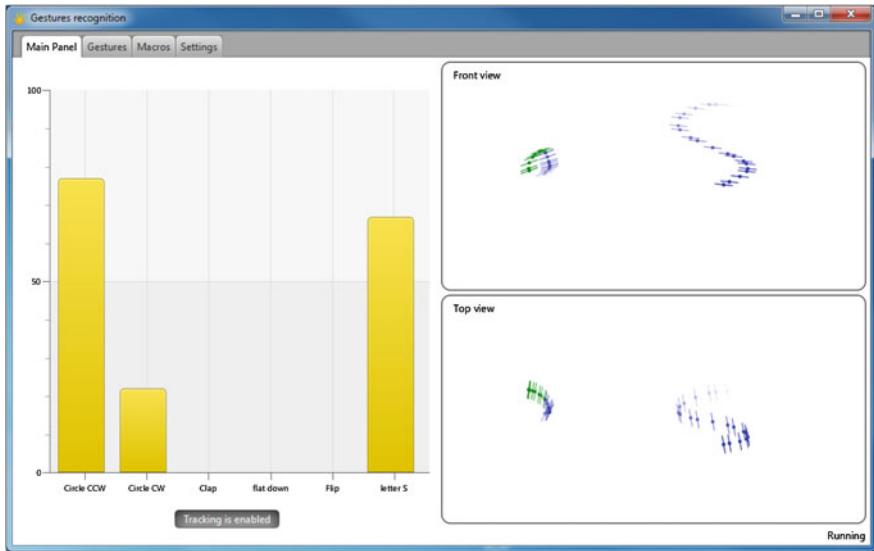
An example of such realization is described in [26]. A PrimeSense Carmine RGB-D camera, a hand tracking library created by 3Gear Systems [55], and an algorithm to match gestures were deployed to investigate if such depth camera can provide enough accuracy. The camera has to be placed at the top of the user space, as shown in Fig. 9.

The hands tracking library provides raw data, such as position and rotation of each joint of the fingers on both hands. In the project, a filter was designed that extracts and samples the relevant features. For example, instead of using an absolute position, velocity of movements was used. This allowed to easily match gestures in each part of the scene. Next, these features were passed to the matching algorithm, based on the Dynamic Time Warping (DTW). The DTW algorithm is based on calculating the Euclidean distance for each samples of two sequences (gestures), and finding the shortest distance between the sequences. If the distance value is zero, it means that the sequences are equal (which is never the case due to noise). The complexity of the DTW is  $O(n^2)$ , and after optimization is reduced to  $O(n)$ . This means that the DTW can be computed in real-time at a high sample rate.

When a new gesture is recorded, only relevant data is extracted (either position or rotation, or pose of left/right hands). Using the trained gestures, the system is able to detect which gestures were performed, and then control a predetermined



**Fig. 9** Depth camera position for gesture recognition [55]



**Fig. 10** Gesture recognition application’s main panel

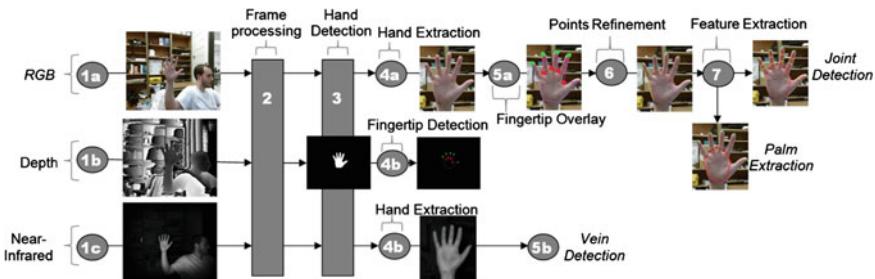
action (for example, opening a favorite site in a web browser). The interface of the implemented algorithm is shown in Fig. 10; it displays the distance to each of the gestures in real-time (the chart on the left), and the current position of the hands (the chart on the right).

In addition, by pointing at the screen with the index finger, we can use our hands to control the mouse, click and scroll, or trigger specific actions by doing gestures and do precise manipulations with the mouse.

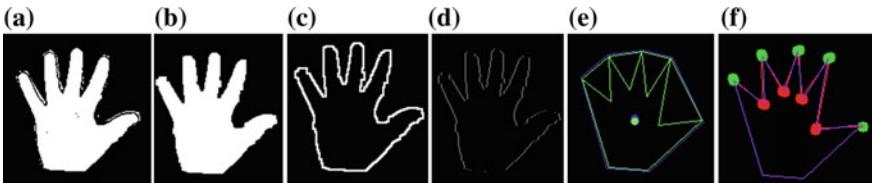
#### 4.2 Hand Recognition Using RGB-D Sensor

In [57], the Kinect sensor was used to perform 3D hand registration using depth data, and further hand identity recognition by its geometrical measurements, such as length of fingers. A combination of near-IR and some of the visible light spectrum were used to help do identification of a dorsal hand (back of the hand) using the vein patterns was proposed in [7].

In [46], an RGB-Depth (RGB-D) camera such as the Kinect v2 [36] was used for hand detection and recognition. The RGB, depth, and near-infrared (near-IR) spectra provided by the camera, were used to analyze palm print, hand shape, finger joint location, possibly and vein patterns. Extraction of the hand is first done using depth data. The frames with the best palm position are selected, and then correlated into the synchronized RGB and near-IR frames for further processing of the related information in each spectra. Using the hand location information the palm can be



**Fig. 11** Proposed hand biometric system using multi-spectral sensors



**Fig. 12** Clockwise: **a** original binary image **b** image after dilation, erosion, and opening **c** result of gradient operation **d** extracted contour **e** polygon, convex hull, and center of mass **f** final image with finger tips in green and points of inflection in red

extracted in the RGB data for use in palm recognition. The suggested design of a system that captures RGB, Depth as well as near-infrared (near-IR) data is shown in Fig. 11.

The approach chosen in [46], detects fingertips in the depth image of a hand using a convexity defect detection algorithm [32]; the result of such detection is shown in Fig. 12.

The fingertip location is translated to the RGB image. Using the valley points between the fingers, the wrist points and the midway pink-wrist point a polygon can be created that mostly encloses the palm of the hand. Next, recognition of the palm is performed using Principle Component Analysis and K-Nearest-Neighbors for the classification. It can also be applied for the vein pattern recognition.

#### 4.3 Gesture-Based Control Using Leap Motion Sensor

The Leap Motion sensor is a relatively new device, introduced by Leap Motion, Inc. in 2013, that has potential to be used in many different medical applications. The device itself is a small sensor that connects to a computer that can be used for gesture control and hand tracking (Fig. 13).

Scopis Medical has integrated the Leap Motion sensor into their surgical navigation system as a way to control the interface in a touchless manner. This surgical

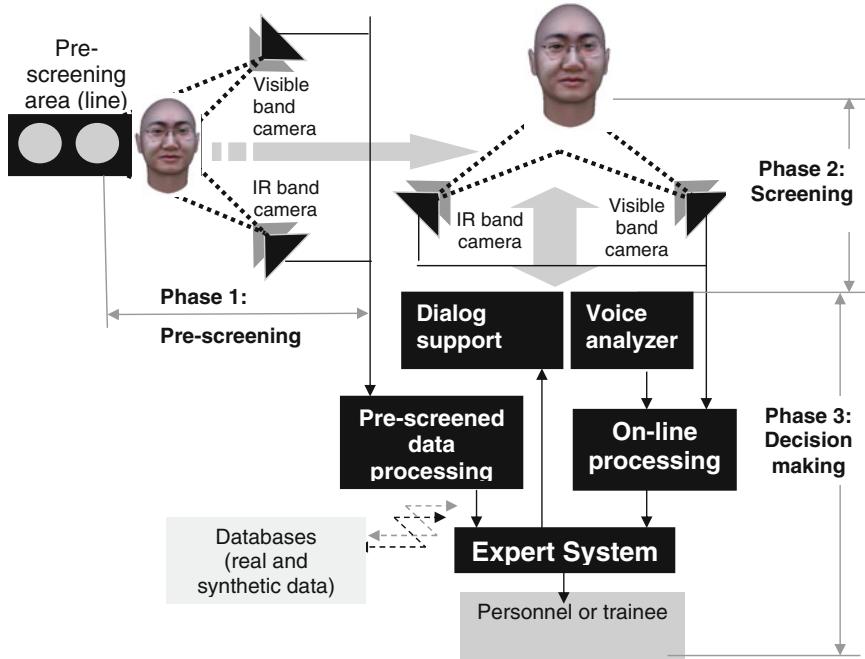
**Fig. 13** Leap Motion sensor set up and demonstrating hand tracking on desktop computer in the Biometric Technologies Lab



navigation software allows the surgeon to browse images and information such as CT and MRI scans without needing to touch a physical device [48]. In a similar vein, TedCas has also integrated the Leap Motion controller into their software to allow doctors the ability to browse medical images while still remaining sterile [53]. There is also software under development in the dental field for browsing dental records and showing 3D models to patients proposed by Dr. Cazenave. In addition to the previous applications, Leap Motion, Inc. has also started the LEAP AXLR8R project with the goal of supporting developers to create new applications for the Leap Motion. Included in these applications are several with a medical focus. MotionSavvy intends to provide real time translation of American Sign Language, Diplopia hopes to help return depth perception to patients with lazy eye, Sterile Air is developing an operating system for use in a sterile surgical environment, and Ten Ton Raygun is developing a game for use in physical rehabilitation [29].

## 5 Biomedical Decision-Making Support Systems

A registration procedure is common in the process of person identification, such as immigration control, patient pre-screening in hospitals, and other places where secure physical admission is practiced. Such systems can be enhanced with intelligent support using additional biometric data measured at a distance, such as patient's temperature, facial expression, voice pitch, gait and other behavioral patterns. The primary sources of information for such data acquisition include infrared and visual band images of the screened person, acoustic and odor etc. The screening and monitoring system can analyze this data and provide the analysis results to the medical personnel, in order to support them in decision making regarding patient's admission and care procedure.



**Fig. 14** A biometric-based screening system: data on individuals is sensed during pre-screening (before admission point; this data in a visible and IR bands is then analyzed to extract preliminary information (phase 1) and help the personnel to make a decision (phase 2)

The idea of the modeling of biometric data for decision-making support has been explored, in particular, at the Biometric Technologies Laboratory at the University of Calgary (<http://www.ucalgary.ca/btlab>). The prototype system (Fig. 14) consists of the following modules: *cameras* in visible and infrared bands placed at two levels of observation, *processors* of preliminary information and online data, an *expert system* to support conversation based on preliminarily obtained information, and a *personal file* generating module. Two-level surveillance is used in the system: surveillance of the line, or admission area, and next to the personnel desk. A facial model of a tested person is captured, and can be manipulated to mimic the changes in visual and infrared bands caused by physiological changes during the questioning period. These can be compared against generic models of change created based on statistical data. This approach can also be used to train personnel involved in questioning procedures. The decision-making support system utilizes synthetic models for rendering the biometric data to support decision-making.

## 5.1 Databases of Biometrics

The collection of large databases of biometric data, such as faces, is troublesome for many researchers due to the protection of personal information. Imitation of biometric data allows the creation of databases with tailored biometric data without expensive studies involving human subjects [60]. An example of a tool used to create databases for fingerprints is SFinGe developed at the University of Bologna [52]. The generated databases were included in the Fingerprint Verification Competition FVC2004 and perform just as well as real fingerprints [54].

There are also several other databases available for use with synthetic biometrics. One such example is the CUHK Face Sketch Database (CUFS). In this database there are photos of a person and then a sketch drawn by an artist. The example of a use for this database provided with it is for use in research with creating synthetic sketches [58].

Synthetic databases have also been created for MRI images of the brain. The BrainWeb: Simulated Brain Database is a database created for use in studying MRI images of the brain. The creators of the database used an MRI simulator to create synthetic sets of MRI data of a brain. A use case for this database presented by the creators is for the evaluation of different image analysis techniques [58].

## 5.2 Medical Personnel Assisting and Training

Based on the aforementioned RGB-D sensor Kinect, Microsoft created an application called Touchless Interaction in Medical Imaging [34]. The project is aiming at implementing a virtual medical imaging environment, in order to give surgeons an ability to obtain a virtual peek inside the human body. Such modeling and rendering would enable a surgeon, who cannot use a mouse or keyboard because they are unsterile and pose a risk of infection, to use simple hand gestures to change, move, or zoom in on MRIs, CT scans and other medical images. In initial development, the Kinect for Windows based system has thrilled surgeons who have seen it and who believe it could help make surgery faster and more accurate. The hope is that these systems will deliver better outcomes to patients when fully field-tested and approved.

Simulators of biometric data are emerging technologies for educational and training purposes (health care, immigration control, banking service, police, justice, etc.) A simulator is understood as a system for modeling specific conditions for taking in and processing biometric data. An example of such a system is a simulator for training customs officers (supported by a signature imitator, face imitator, and fingerprint imitator) [60].

In [62], the biometric-based access control system is used as a training system (with minimal extension of tools) without changing of the place of deployment. The goal of training is to develop the user's decision making skills based on biometric information collected during screening (non-invasive biometrics, such as video

or thermal video monitoring) or during the facility check-in procedure (may be invasive, such as taking fingerprints or iris scan). The training system directly benefits users of biometric-based systems covering a broad spectrum of social activities, including surveillance and control, border control, hospitals, important public events, banking, etc.

This system must be capable of simulation, in particular, simulation of extreme scenarios aimed at developing the particular skills of the personnel. Such modeling of extreme situations requires developing specific training methodologies and techniques, including virtual environments.

### ***5.3 Accessibility for Disabled Patients***

Researchers at UC Irvine [63] developed a Kinect-Wheelchair Interface Controlled (KWIC) Trainer, that helps children with disabilities learn how to use a powered wheelchair. This training device allows a caregiver to use natural gesture commands to assist a child being trained, using Kinect's gesture control.

An application that uses head pose and gesture biometrics for patients with inability to use their hand to direct the regular computer mouse, was developed at Nouse [38]. Nouse is an application for tracking head motion and, specifically, nose on the face, using a web-camera on the computer. The Nouse application is used to control the cursor on the computer screen in front of the patient. This is accomplished by tracking the user's nose and head motions.

### ***5.4 Rehabilitation Applications***

A study was conducted by Lange et al. [27] into the development of a game for use in balance rehabilitation using a Kinect sensor for human motion tracking and recognition. While statistical results were not provided in the paper, overall impressions of the patients were provided. In general it appears the users of their developed system much preferred using it over more traditional rehabilitation techniques (such as going to the gym) [27].

### ***5.5 Avatar Systems***

Virtual environments refer to interactive systems that provide the user with a sense of direct physical presence and direct experience. These systems include visual and feedback techniques, and allow natural interaction by manipulation of computer-generated objects [5]. An example of such systems is a conversational avatar-based kiosk for automated interviewing [39]. Such systems interview individuals and detect

changes in arousal, behavior, and cognitive effort by using psycho-physiological information systems. The avatar-based kiosks use heterogeneous sensors to detect both physiological and behavioral biometrics during interactions, and they affect their environment by influencing human behavior using gender and demeanor, messages, and recommendations.

Research on behavior and animation has also been conducted related to creating robots that interact more naturally with humans. The goal of that work is to create a robot software application that could keep memories and act in a more social manner by including facial recognition. The facial recognition was used so that the robot was able to recognize people that interacted with it. In the 2010 paper by Mavridis et al., a robot was used to interact with people and use images it found on Facebook to recognize who the person was; results of identifying a person correctly were reported at 50% [33].

## 6 Privacy of Biometrics

There are privacy concerns, especially in biomedical and health care applications, on using a customer's or patient's data for research, training or routine service purposes.

The issue of protecting privacy in biometric systems has inspired the direction of research referred to as *cancelable biometrics* [3]. Cancelable biometrics is aimed at enhancing the security and privacy of biometric authentication through the generation of "altered" biometric data, that is, synthetic biometrics. Instead of using a true object (finger, face), the fingerprint or face image is intentionally distorted in a repeatable manner, and this new print or image is used.

## 7 Conclusion

Biometric technologies is a unique area that combines various methodologies, concepts, and techniques from many areas of the natural and social sciences, in particular: signal and image processing, data mining, pattern recognition, virtual environment design, human-machine interaction, integrated knowledge-intensive decision-making, communication, and psychology. In many biomedical applications, analysis of biometric data (in particular, analysis of biomedical data) is combined with modeling of biometric data. New applications emerged recently as a result of advances in technology, including better RGB, Depth and Infrared sensoring systems and hardware acceleration of time-consuming signal processing.

**Acknowledgments** The authors would like to thank the National Science and Engineering Research Council (NSERC) (support via Discovery grant "Biometric intelligent interfaces"), Queen Elizabeth II Scholarship, and the Department of Electrical and Computer Engineering of the University of Calgary for their continuous support of this research.

## References

1. Agrafioti, F., Gao, J., Hatzinakos, D.: Heart biometrics: theory, methods and applications. In: Yang, J., (ed.) *Biometrics: Book 3*, Intech, pp. 199–216 (2011)
2. Alivecor. <http://www.alivecor.com/home>. Accessed Jan 2014
3. Bolle, R., Connell, J., Pankanti, S., et al.: *Guide to Biometrics*. Springer, New York (2004)
4. Boulanov, O.R., Gavrilova, M.L., Poursaberi, A., et al.: Biometric-based intelligent agent systems. IADIS Int. Conf. Intell. Syst. Agents, Rome, Italy **24–26**, 162–164 (2011)
5. Burdea, G.C., Coiffet, P.: *Virtual Reality Technology*, 2nd edn. Wiley, New York (2004)
6. Can, A., Steward, C.V., Roysam, B., et al.: A feature-based, robust, hierarchical algorithm for registering pairs of images of the curved human retina. *IEEE Trans. Anal. Mach. Intell.* **24**(3), 347–364 (2002)
7. Chen, K., Zhang, D.: Band selection for improvement of dorsal hand recognition. In: International Conference on Hand-Based Biometrics, pp. 1–4, 17–18 Nov 2011
8. Claes, P., Liberton, D.K., Daniels, K., et al.: Modeling 3D Facial Shape from DNA. *PLOS Genet.* **10**(3), e1004224 (2014). doi:[10.1371/journal.pgen.1004224](https://doi.org/10.1371/journal.pgen.1004224)
9. Cui, J., Wang, Y., Huang, J., et al.: An iris image synthesis method based on PCA and super-resolution. In: International Conference on Pattern Recognition, pp. 471–474, 23–26 Aug 2004
10. Du, Y., Lin, X.: Realistic mouth synthesis based on shape appearance dependence mapping. *Pattern Recognit. Lett.* **23**(14), 1875–1885 (2002)
11. Duchaine, B., Nakayama, K.: Developmental prosopagnosia: a window to content-specific face processing. *Curr. Opin. Neurobiol.* **16**(2), 166–173 (2006)
12. Ekman, P., Rosenberg, E.L., (eds.): *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*. Oxford University Press, Oxford (1997)
13. Eveland, C.K., Socolinsky, D.A., Wolff, L.B.: Tracking human faces in infrared video. *Image Vis. Comput.* **21**, 579–590 (2003)
14. FaceGen. <http://www.facegen.com/>. Accessed Nov 2013
15. Fanelli, G., Dantone, M., Gall, J., et al.: Real time head pose estimation with random regression forests. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 617–624, 20–25 June 2011
16. Foster, J.P., Nixon, M.S., Prügel-Bennett, A.: Automatic gait recognition using area-based metrics pattern. *Recogn. Lett.* **24**, 2489–2497 (2003)
17. Franke, K., Ruiz-del-Solar, J.: Soft-biometrics: soft-computing technologies for biometric-applications. In: Pal, N.R., Sugeno, M. (eds.) *Advances in Soft Computing AFSS*, pp. 171–177. Springer, Berlin (2002)
18. Fu, Y., Guo, G., Huang, T.S.: Age synthesis and estimation via faces: a survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **32**(11), 1955–1976 (2010)
19. Fu, Y., Huang, T.S.: Human age estimation with regression on discriminative aging manifold. *IEEE Trans. Multimedia* **10**(4), 578–584 (2008)
20. Fujimasa, I., Chinzei, T., Saito, I.: Converting far infrared image information to other physiological data. *IEEE Eng. Med. Biol. Mag.* **10**(3), 71–76 (2000)
21. GestureTek Health. <http://www.gesturetekhealth.com>. Accessed March 2014
22. Google Glasses. <https://developers.google.com/glass>. Accessed May 2014
23. Guo, G., Fu, Y., Dyer, C., et al.: Image-based human age estimation by manifold learning and locally adjusted robust regression. *IEEE Trans. Image Proces.* **17**(7), 1178–1188 (2008)
24. Jiang, H., Duerstock, B.S., Wachs, J.P.: A machine vision-based gestural interface for people with upper extremity physical impairments. *IEEE Trans. Syst. Man Cybern. Syst.* **44**(5), 2168–2216 (2014)
25. Lai, K., Samoil, S., Yanushkevich, S.N.: Multi-spectral facial biometrics in access control. In: IEEE Symposium on Computational Intelligence in Biometrics and Identity Management, pp. 102–109, 9–12 Dec 2014

26. Lai, K., Samoil, S., Yanushkevich, S.: Application of biometric technologies in biomedical systems. In: International Conference on Digital Technologies, pp. 207–216, 9–11 July 2014
27. Lange, B., Chang, C., Suma, E., et al.: Development and evaluation of low cost game-based balance rehabilitation tool using Microsoft Kinect sensor. In: IEEE International Conference on Engineering in Medicine and Biology Society, pp. 1831–1834, 30–3 Sept 2011
28. Lanitis, A., Taylor, C.J., Cootes, T.F.: Toward automatic simulation of aging effects on face images. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(4), 442–455 (2002)
29. Leap Motion Incorporated. Introducing the 10 LEAP AXLR8R Teams. <https://developer.leapmotion.com/blog/introducing-the-10-leap-axlr8r-teams/> Accessed March 2014
30. Lefohn, A., Budge, B., Shirley, P., et al.: An ocularist's approach to human iris synthesis. *IEEE Mag. Comput. Graph. Appl* **23**(6), 70–75 (2003)
31. Lo, B., Lee, H., Ing, M., et al.: Modeling of Facial Nerve Disorders. Undergraduate Project Report, Biometric Technologies Laboratory, University of Calgary (2006)
32. Maisto, M., Panella, M., Liparulo, L., et al.: An accurate algorithm for the identification of fingertips using an RGB-D camera. *IEEE J. Emerg. Sel. Top. in Circuits Syst.* **3**(2), 272–283 (2013)
33. Mavridis, N., Petychakis, M., Tsamakos, A., et al.: FaceBots: steps towards enhanced long-term human-robot interaction by utilizing and publishing online social information. *Paladyn* **1**(3), 169–178 (2010)
34. Mantis, H., Taylor, A.: Imaging the body: embodied vision in minimally invasive surgery. In: Proceedings of Human Factors in Computing Systems (2013). doi:[10.1145/2470654.2466197](https://doi.org/10.1145/2470654.2466197)
35. Microsoft Kinect. <http://www.microsoft.com/en-us/kinectforwindows/>. Accessed Dec 2013
36. Microsoft Kinect for Windows. <http://www.microsoft.com/en-us/kinectforwindows/default.aspx>. Accessed March 2014
37. Moriyama, T., Xiao, J., Kanade, T., et al.: Meticulously detailed eye model and its application to analysis of facial images. In: IEEE International Conference on Systems, Man and Cybernetics, pp. 629–634, 10–13 Oct 2004
38. Nouse. <http://www.nouse.ca/en/>. Accessed March 2014
39. Nunamaker Jr, J.F., Derrick, D.C., Elkins, A.C., et al.: Embodied conversational agent based Kiosk for automated interviewing. *J. Manage. Inf. Syst.* **28**(1), 17–48 (2011)
40. Oliveira, C., Kaestner, C., Bortolozzi, F., et al.: Generation of signatures by deformation. In: Murshed, N.A., Bortolozzi, F. (eds.) *Adv. Doc. Image Anal.*, pp. 283–298. Springer, Berlin (1997)
41. Oliver, N., Pentland, A.P., Berard, F.: LAFTER: a real-time face and lips tracker with facial expression recognition. *Pattern Recognit.* **33**(8), 1369–1382 (2000)
42. Pantic, M., Rothkrantz, L.J.M.: Automatic analysis of facial expressions: the state-of-the-art. *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(12), 1424–1445 (2000)
43. Patel, S., Park, H., Bonato, P., et al.: A review of wearable sensors and systems with application in rehabilitation. *J. Neuro Eng. Rehabil.* **9**, 21 (2012). doi:[10.1186/1743-0003-9-21](https://doi.org/10.1186/1743-0003-9-21)
44. Pavlidis, I., Levine, J.: Thermal image analysis for polygraph testing. *IEEE Eng. Med. Biol. Mag.* **21**(6), 56–64 (2002)
45. Rolls, E.T.: Toward automatic simulation of aging effects on face images. *Behav. Process.* **33**(1–2), 113–138 (1994)
46. Samoil, S., Lai, K., Yanushkevich, S.: Multispectral hand biometrics. In: International Conference on Emerging Security Technologies, pp. 24–29, 10–12 Sept 2014
47. Sanchez-Avila, C., Sanchez-Reillo, R.: Iris-based biometric recognition using dyadic wavelet transform. *IEEE Aerosp. Electron. Syst. Mag.* **17**(10), 3–6 (2002)
48. Scopis GmbH. Touchless control of a surgical navigation system. <http://www.scopis.com/en/news/news/details/archive/2013/may/03/article/beruehrungslose-steuerung-eines-klinischen-navigationssystems/>. Accessed July 2013
49. Spree. <http://spreesports.com>. Accessed March 2014
50. Sproat, R.W. (ed.): *Multilingual Text-to-Speech Synthesis: The Bell Labs Approach*. Kluwer Academic Publishers, Norwell (1997)

51. Sugimoto, Y., Yoshitomi, Y., Tomita, S.: A method for detecting transitions of emotional states using a thermal facial image based on a synthesis of facial expressions. *Robot. Auton. Syst.* **31**(3), 147–160 (2000)
52. Synthetic Fingerprint Generator. <http://bias.csr.unibo.it/research/biolab/sfinge.html>. Accessed March 2014
53. TedCas Medical Systems.: TedCas integrates leap motion controller with medical imaging systems. <http://www.tedcas.com/en/node/1562>. Accessed March 2014
54. The Fingerprint Verification Competition FVC2004. <http://bias.csr.unibo.it/fvc2004/databases.asp>. Accessed March 2014
55. ThreeGear. <http://www.threegear.com/>. Accessed June 2014
56. Tsumura, N., Ojima, N., Sato, K., et al.: Image-based skin color and texture analysis/synthesis by extracting hemoglobin and melanin information in the skin. *ACM Trans. Graph.* **22**(3), 770–779 (2003)
57. Wang, C., Liu, H., Liu, X.: Contact-free and pose-invariant hand-biometric-based personal identification system using RGB and depth data. *J. Zhejiang Univ. Sci. C* **15**(7), 525–536 (2014)
58. Wang, X., Tang, X.: Face photo-sketch synthesis and recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **31**(11), 1955–1967 (2009)
59. Yamamoto, E., Nakamura, S., Shikano, K.: Lip movement synthesis from speech based on hidden markov models. *Speech Commun.* **26**(1–2), 105–115 (1998)
60. Yanushkevich, S.N., Stoica, A., Srihari, S.N., et al.: Simulation of biometric information: the new generation of biometric systems. In: International Workshop on Modeling and Simulation in Biometric Technology, pp. 87–98, 22–23 June 2004
61. Yanushkevich, S.N., Stoica, A., Shmerko, V.P., et al.: *Biometric Inverse Problems*. Taylor and Francis/CRC Press, Boca Raton (2005)
62. Yanushkevich, S.N., Stoica, A., Shmerko, V.P.: Fundamentals of biometric-based training system design. In: Yanushkevich, S.N., Wang, P., Srihari, S., et al. (eds.) *Image pattern recognition: synthesis and analysis in biometrics, Machine Perception and Artificial Intelligence*, vol. 67, pp. 365–406. World Scientific
63. Zondervan, D.K., Reinkensmeyer, D.J.: Kinect-wheelchair interface controlled (KWIC) robotic trainer for powered mobility. In: International Conference of the IEEE Engineering in Medicine and Biology Society (2012)

# Multimedia Support of Medical Education Utilizing On-Line Archive of Video Lectures

Jaroslav Majerník and Pavol Jarčuška

**Abstract** Recent trends in higher education utilize advances of information and communication technologies and web services more intensively than ever to support distance education forms. The reasons raised mainly because of the needs to minimize working load of teachers, to create reusable teaching resources and to satisfy students requirements to obtain electronic and on-line available study materials. Similarly to other branches, we applied such techniques to offer our students but also to all interested persons the latest knowledge from the area of infectious diseases, their prevention and treatment. Therefore, we realized specialized and scientific sessions that were live streamed, archived and shared through video gallery on our faculty portal.

## 1 Introduction

Various teaching methods can be used to offer students new information and knowledge. It is true also for digitally oriented ones already applied in almost any scientific area [1, 2], including medicine with the aim to increase also the public knowledge and to ensure better prevention behaviours [3–5]. Considering the grade of used technology, the distance education can be organized through different concepts including distance learning, distributed learning, online learning, e-learning, virtual education, web-based learning, computer-based training, and blended or hybrid learning [6–8].

Higher education curriculums uses lectures as a cornerstone of teaching strategies. They represent the most dominant method of teaching large groups of students [9] and are very important to introduce new topics and to prepare audience for practical part of education as well as for later practice. Usually, the key facts and concepts are

---

J. Majerník (✉)

Faculty of Medicine, Department of Medical Informatics, Pavol Jozef Šafárik University  
in Košice, Trieda SNP 1, 040 11 Košice, Slovakia  
e-mail: jaroslav.majernik@upjs.sk

P. Jarčuška

Faculty of Medicine, Department of Infectology and Travel Medicine, Pavol Jozef Šafárik  
University in Košice, Rastislavova 43, 041 90 Košice, Slovakia  
e-mail: pavol.jarcuska@upjs.sk

clearly summarized and provided to the students while interacting with their teachers [10, 11]. Thanks to the multimedia the teachers can enhance both the learning experience and understanding of difficult scientific problems [12, 13]. On the other hand, the students expect to have a wide range of learning e-sources with convenient and flexible access. Influenced by social media services they also prefer to use on-line tools to self-study and to make effective collaboration and communication [14–16]. Evaluation of experiences based on using a video-based learning environment revealed that the students consider broadcasted or steamed instruction video as effective for carrying out self-evaluations. The teachers also considered streaming video as useful education tool for all the reflection processes of their students. On the other hand, they also indicated some shortcomings [17, 18]. Other studies that examined the final grade and satisfaction level differences among students taking specific courses using three different methods: face-to-face in class, satellite broadcasting, and live video streaming indicated no statistically significant differences. Such results provide evidence to support distance education and to use it as a viable, convenient and flexible alternative delivery mode capable of extending learning opportunities.

Sharing of medical knowledge over a long distance has relatively long history and is usually referred to telemedicine applications. Based on telemedicine, it is possible to increase not only the level of medical education, but also to realize remote medicine and emergency medical services. Here, the aims in medical education are usually oriented to support remotely accessible lectures, real clinical examinations and interventions, simulations based on virtual patients or techniques of motor skills development [19–21]. Most of such activities are moderated by clinical professionals with years of practical experiences, what is one of the most important factors to increase interest in medical education. These activities are also the chance for students to obtain first clinical information, to imagine real problems the physicians have to solve when working with their patients and to understand differences in individual cases the students will deal throughout their later professional career. Such experiences are extremely useful and represent inseparable part of medical learning programs.

Because of changes in curriculum and related transferring processes based on more effective utilization of modern e-learning tools in education at our faculty, we decided to create a methodology that combines advantages of face-to-face and distance education into the teaching approach useful for our students, but also for other persons interested in infectology topics. The amount of medical information to be transmitted to the users, including texts, audio and video data requires equivalent technologies, network infrastructure and encoding techniques. Therefore, we used only the tools with no special technological requirements at the side of remote users.

## 2 Educational Video Delivery Methods

Basically, and from technological and historical point of view, there were various methods used to distribute educational content to remote students all around the

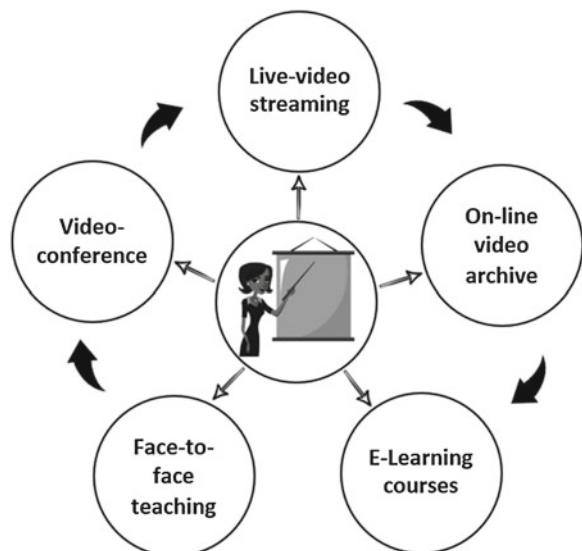
world. In almost all cases, this process was referred to the delivery of video where the following three main technological profiles were applied:

- videoconference—interactive two-way video and audio. This is probably the best “simulation” of the real classes and the face-to-face forms of education. The teaching events are realized in real-time, communication runs synchronously, but depending on the quality of internet connection, usually the high-speed networks are required.
- instructional television fixed service (ITFS)—broadcasted (satellite, cable and/or terrestrial) as one-way video and two-way audio. It was widely employed at many colleges and universities since the 80s, where the distance students were able to make a phone call to the teacher to ask questions and interact with him/her.
- streaming—one-way video and audio. Using streaming technology the live-lecture video (with audio) is broadcasted using network infrastructure and the students may watch it anywhere they are. Archived materials together with web-streaming functionalities can be used to deliver delayed education content to the distance learners through the universities’ web sites. The students are usually allowed to interact with the teacher via e-mail and/or web-based discussion boards (asynchronous communication).

Respecting our aims and accessibility of technologies we integrated videoconferences and live streaming in to the model of education that is shown in Fig. 1.

Live broadcasts of scientific and educational sessions were captured and then processed, archived and shared as on-line video-clips to be accessible anytime and anywhere. The structure of our methodology respects two main requirements. The first is the ability to distribute live education events to the almost unlimited number

**Fig. 1** Methods integrated in the system intended to deliver medicine related educational materials to undergraduate and postgraduate students



of users and the second one represents accessibility for the users having no special technological equipment.

### 3 Broadcasting of Live Events

Most of the infectology related lectures were broadcasted using video streaming technologies. In general, these technologies combine videoconferencing and the Internet so that the live or recorded events can be distributed via the web. Then, the students can view the streams on their own computers wherever it suits them rather than having to take part in a face-to-face lecture. The use of video streaming as a sole teaching strategy essentially remains a non-interactive medium and may therefore have certain disadvantages that are comparable to traditional lectures. Nevertheless, it represents innovative teaching and learning resources with perspectives to be used also in mobile devices.

Thanks to the availability of high-bandwidth network infrastructure the live video streaming enables teachers to deliver high-quality video and audio presentations while enabling students to view, interact, and connect with their teachers and/or classmates. This opposes that video streaming removes interaction and human contact between students and teachers. The real truth is that the advances in methods of data compression and extension of computer networks have significantly increased the ability to interconnect teachers and their students across the world. Teachers and lecturers may perform teaching remotely using live video streaming over the Internet even more easily than ever before. All participants, like teachers, students and/or parents can be involved in collaboration with each other and simplify both the learning and communication processes. Two main concepts of video streaming on the Internet can be specified. These include:

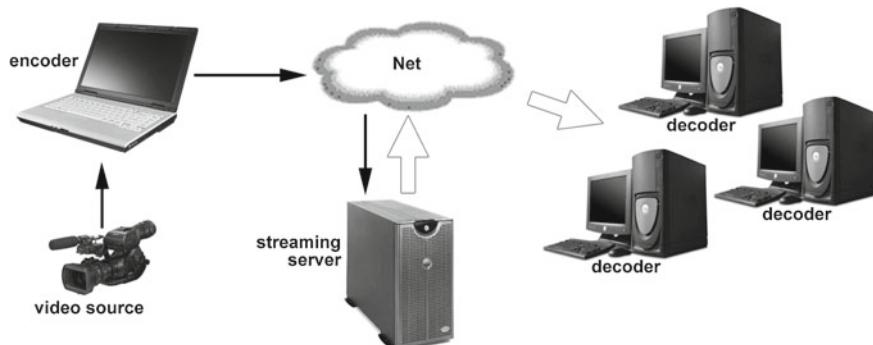
- true streaming—also called as streaming media or streaming servers, where no waiting to watch event is involved as the media (specialized servers) are able to broadcast live events. An additional advantage of true streaming is that the education content can be broadcasted at different bit rates. Users can start watching live events almost immediately after they start playing it. Time delay, due to the server processing and type of network connection, according to our experiences, moves about 15 s. In the case of archived records, the file is also played almost as soon as it is downloaded. Basically, the content is sent to the users as stream so they watch it as it arrives to the computer.
- downloading—also termed as http streaming. It is a method where the video file (containing video, audio, animations, texts etc.) is usually offered to the users as embedded hyperlink on the web page. Users download entire content/file on their computers as first and then it can be opened and viewed. Depending on size of file it takes some time to be downloaded completely, so the users have to wait for whole file. Although it cannot be used for live events, the advantage is that the user can quickly skip already viewed or for him/her uninteresting parts of

lectures, presentations and/or other teaching courses. Combination of streaming and downloading methods creates something like progressive downloading, where the downloaded file can be played as its individual parts are received. In contrast to true streaming this may result in non-continuous playing.

To stream live events using true streaming methods it is necessary to have streaming server, that prepares, encodes and transmits video to the users. There are several file formats that can be used and some of them are offered by RealNetworks (Real Media), Apple (QuickTime), Adobe (Adobe Flash) or Microsoft (Windows Media). Considering our previous experiences and our technical equipment we decided to use RealNetworks Helix technology to steam live education events. We prefer to use it as it meets all of our technical requirements. Except of many other advantages it is also because of ability to handle traffic loads and ability to detect users' connection speeds and to supply appropriate files automatically. Thanks to its stability we use it as an effective tool of distance education and also as a tool to offer the newest research results to the wide group of remote users.

Helix, similarly to other server solutions, uses three main component to handle live events. These are encoder, streaming server and decoder. Encoder is the source of video signal. It is responsible for digital compression of all captured video and audio data that are sent to the streaming server. Streaming server forwards video data to the remote users. It produces different file formats and streams them on different bit rates. And finally, the decoder is presented by the player that is responsible for decoding of received video files and for playing them in computer of remote user. Principle scheme of the video streamed communication and interconnection of above mentioned parts of true streaming is shown in Fig. 2.

The advantages of Helix technology include platform independence (Linux, Windows, and Solaris) and support of various codecs. The encoder at the transmitting side is represented by RealProducer processing and encoding input video and audio signals. The necessary equipment consists of video camera, microphone, speakers and computer with connection to the Internet. Helix server receives data from encoder and



**Fig. 2** Principle of true live video streaming

distributes streamed content to the remote users. Free RealPlayer is used as decoder at the side of students, watching live education event. In this way, the students as consumers of education content do not need to have any special ICT products or equipment to watch live events.

## 4 Video Gallery of Educational Lectures

Individual live video streams are broadcasted as free to join events, so everybody interested in particular topics is able to watch them. However, there are often various objective reasons why some of the sessions cannot be viewed when broadcasted. Therefore, the raw video records are used to prepare archive of audiovisual lectures including DVD movies, compressed video formats for web as well as interactive presentations. Educational outputs are processed to be available for students in both on-line and off-line forms. In this manner, the streams can be archived and shared together with additional education materials.

The process to create usable audiovisual educational works include scene capturing, editing and rendering of teaching suitable parts, adding additional content and comments, conversion to the requested video/multimedia format and publication of final materials as it is shown in Fig. 3.

Educational outputs are processed to be available in both on-line and off-line forms. In this manner, the streams were archived and shared through faculty's web portal together with additional education materials. Here, the video lectures (containing video, audio, animations, texts etc.) are distributed using so called web streaming or progressive downloading. Although it cannot be used for live events, the advantage is that the user can quickly skip already viewed or for them uninteresting parts of lectures, presentations and/or other teaching courses.

There are several standards used to share video over the Internet while enabling high degree of compression retaining reasonable video quality. However, there are also some limitations in the transmission of high quality medical information using these formats through Internet and mobile communications. Despite of this, the

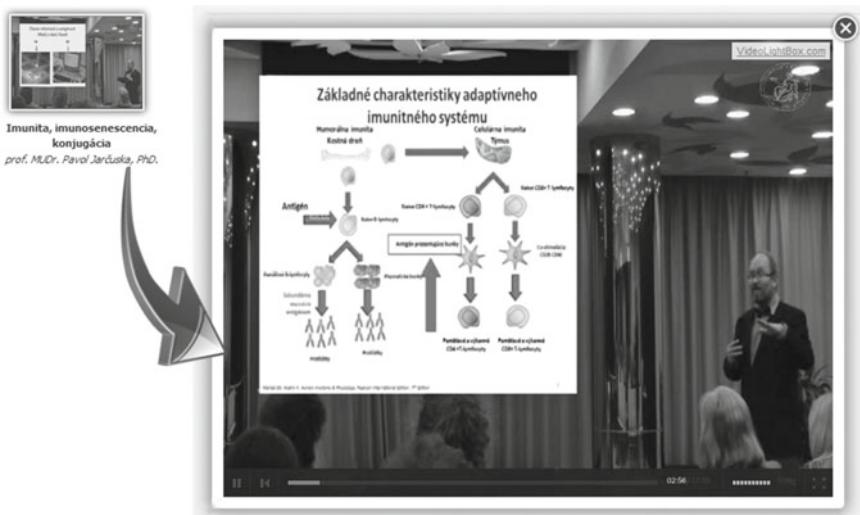


**Fig. 3** Educational video processing scheme

most preferred formats are MPEG-4 and FLV. MPEG-4 is widely used for digital television, in smart phones or as streaming media. We decided to use FLV as the Flash was the most relevant target platform for many years, and because it uses progressive downloading, can be downloaded from start to finish over HTTP, plays as the file is downloaded (no need to wait), is low cost and effective and is useable also for slow connections. Nevertheless, both containers have their pros and cons. Output FLVs were compressed to get optimal quality to file size ratio, adopting characteristics of video size, bitrate, frame rate, aspect ratio and audio compression. The values  $640 \times 480$  px, 512 bps, 15 fps and MP3 VBR audio were applied as we recognized them to be sufficient for our web-shared video lectures. It is because of optimal combination of the dynamic movie of the speaker (usually slow motions) and the static frames of his/her presentation. An example of FLV player embedded in our web portal is shown in Fig. 4.

The faculty's web portal of multimedia support in the education of clinical and health care disciplines (<http://portal.lf.upjs.sk>) was selected as the most suitable platform to share already prepared multimedia outputs and as the best way to offer them to the students and public. Also the links to the Moodle e-learning courses are integrated in this portal because of existing interconnection between MEFANET activities.

We created a video gallery based on principles of Video LightBox to be able to share movies in a well-arranged and attractive way. Video LightBox is an intuitive wizard program, available for free for non-commercial use, allowing adding videos to almost any website with stylish popup video effects. This beautiful product helps to generate source codes to embed particular video to the website in a few clicks.



**Fig. 4** Video thumbnail and FLV player embedded in portal's article

In general, it is necessary just to add video, to specify the template for video popup and video thumbnail appearance and to publish the result. Except of thumbnail, each educational video is equipped with the title and the name of the speaker/author. An example of final video gallery published and accessible with no restrictions is shown in Fig. 5.

Sekcia I - Fyziológické osidlenie človeka baktériami, najčastejšie bakteriálne infekcie

Sekcia II - Bakteriálne infekcie, ktoré sú menej časté alebo sa v praxi na nich menej myslí

Sekcia III - Mechanizmy prenosu infekcií v komunité a v nemocničach

			
Hemolytico uremicus syndróm u dospelého pacienta s hnačkou <i>doc. MUDr. Igor Stankovič, CSc.</i>	Malária - stále prekvapením <i>MUDr. Mária Kopčová</i>	Dirofilariáza - ochorenie nielen trápor a subtrápor <i>MUDr. Lúcia Plešecák, PhD.</i>	Nové trendy v liečbe vírusových infekcií <i>prof. MUDr. Pavol Jarčuška, PhD.</i>
			
Biofilm a jeho úloha pri infekcii <i>doc. MUDr. Milan Nikš, CSc.</i>	Molekulárna diagnostika papilomavírusov <i>MUDr. Katarína Domaracká, PhD.</i>	Nové trendy HIV infekce v ČR <i>doc. MUDr. Marie Starýková, CSc.</i>	Eliminácia rizika prenosu vírusu hepatitidy B krvou darovanou v oblasti ... <i>RNDr. Eleonóra Šépkániová</i>
			
Klieštami prenášané infekcie na Slovensku <i>RNDr. PaedDr. Mária Nováková, PHD.</i>	Chronická hepatitída B u dietetá'a po transplantácii pečene <i>MUDr. Daniela Hudačková</i>	Coxsackie v skautskom tábore <i>MUDr. Diana Vološinová, PhD.</i>	
<u>Sekcia IV - Antibakteriálne liečivo používané v komunité</u>			
<u>Sekcia V - Antibakteriálne liečivo používané v nemocničiach</u>			
<u>Sekcia VI - Antimykotiká, antivirotyká a antiparazitiká</u>			
<u>Sekcia VII - Respiračné infekcie</u>			
<u>Sekcia VIII - Uroinfekcie, neuroinfekcie, gastrointestinálne infekcie, infekcie kože a mäkkých tkániv</u>			
<u>Sekcia IX - Ako a kedy majú pacienti používať antibiotiká?</u>			
<u>Sekcia X - Sepsa</u>			
<u>Sekcia XI - Nemocničné infekcie a rezistenčia na antiinfektiva v nemocničiach</u>			
<u>Sekcia XII - Opatrenia v prevencii nosokomijných infekcií</u>			
<u>Sekcia XIII - Antivirotyká liečba najčastejších infekcií - krok pred alebo krok za antibiotickou liečbou?</u>			
<u>Sekcia XIV - Prezentácia odborných odporučení MZ SR v liečbe mykotických infekcií</u>			
<u>Sekcia XV - Zhadnotenie vzdelávacej aktivity</u>			

**Fig. 5** Video gallery integrated in faculty's web portal and used to share educational infectology video lectures

Individual FLV files can be stored in external servers or directly in the portal's repository. We did not specify any restrictions to access individual lectures as these should be available for wider public. However, respecting principles of the portal, the access rights can be specified in the same way as it is in the case of any other educational attachments. In this case, the content can be set to be available for various groups of users, including:

- nonregistered anonymous users,
- registered anonymous users who accept the terms of use within his/her registration,
- users of MEFANET network, i.e. students or teachers from any Czech or Slovak medical faculty,
- users of local university or faculty, whose affiliation to that university/faculty has been verified at the portal or via the local information system of that university/faculty,
- users to whom attachments are made available only on the author's explicit consent.

In order to have a clear and well-arranged list of video lectures, we separated closely related themes into 15 sessions. These were classified as sessions to cover most relevant infectology topics including the most frequent bacterial infections, antibacterial pharmaceutics, respiratory infections, nosocomial infections, resistance on anti-infective pharmaceutics, preventions against infections, antivirotic treatment of the most frequent infections, etc. Currently, there are lectures in all of these sessions published on the portal and their number gradually increases with each new processed video lecture. The sessions are presented as simple hyperlinks and their content is shown or hidden only after the user will select it. In this way, it is possible to filter the video lectures, to save the space in the gallery and to show the users only the materials they are interested in. The list of individual sessions as are organized within our project is summarized in Table 1.

The first column in the Table 1 shows the titles of particular sessions, while the second one refers to the number of audiovisual lectures and overall time of these lectures and the third column refers to their mean time. We prefer to create short educational clips, rather than entire film covering whole topic. This allows time to be spent on both the introduction of the topic and the dissemination of the key points. Moreover, there has been a lot of research conducted that has shown that the average attention span lasts anywhere from 7 to 15 min. Currently, there are 146 audiovisual lectures with overall time of 27 h 57 min and 15 s shared on the faculty's web portal. Mean time for these outputs is  $0:11:29 \pm 0:04:54$ . All of these lectures are opened and available for free and playable immediately after being selected. The mean time between the time of lecture selection and the time when this lecture is displayed and starts to be played at the side of remote users was  $1.286 \pm 0.302$  s as it was analyzed in wired 100 Mb network infrastructure outside of the university network. However, this time may vary depending on type of the network type at the side of remote users and its stability. The delay can be significantly increased especially in the case of limited and slow wireless connections.

**Table 1** Summary of videolectures published in video gallery categorized according to the infection topics

Session topic	No.; duration	Duration (mean ± SD)
Physiological bacterial settlement, most frequent bacterial infection	07; 2:09:02	0:18:26 ± 0:10:06
Rare bacterial infections	07; 1:16:19	0:10:54 ± 0:03:44
Mechanisms of infections transmission in community and in hospital	11; 2:14:42	0:12:15 ± 0:06:39
Antibacterial pharmaceutics used in community	01; 0:11:05	0:11:05 ± NaN
Antibacterial pharmaceutics used in hospital	04; 0:38:02	0:09:31 ± 0:01:07
Antimycotics, antivirotics and antiparasitics	09; 1:42:37	0:11:24 ± 0:02:44
Respiratory infections	05; 1:04:34	0:12:55 ± 0:06:50
Urinary tract infection, neuro-infections, gastrointestinal infections and skin infections	35; 5:26:34	0:09:20 ± 0:02:20
How and when to use antibiotics	09; 1:55:08	0:12:48 ± 0:04:42
Sepsis	09; 1:22:53	0:09:13 ± 0:01:59
Nosocomial infections and resistance on antiinfective pharmaceutics	05; 0:42:41	0:08:32 ± 0:01:49
Preventions against nosocomial infections	05; 1:11:34	0:14:19 ± 0:04:41
Antivirotic treatment of the most frequent infections	37; 7:38:17	0:12:23 ± 0:04:56
Recommendations in treatment of micotic infections	01; 0:13:01	0:13:01 ± NaN
Selected topics and panel discussion	01; 0:10:46	0:10:46 ± NaN

## 5 Conclusion

We realized web based approach to disseminate medical educational content including latest information about infectious diseases, their treatment and prevention to the undergraduate and postgraduate students at medical faculties, as well as to the clinical professionals and specialists and to the wide range of interested population. The combination of traditional teaching methods and new technological innovations brought advanced teaching and learning tools for our teachers and students. Here,

the live education activities were broadcasted via web services as opened and free to join events. The teams of teachers and lecturers were composed of physicians and professionals with long-time skills in education of clinical disciplines.

Information about regularly organized sessions were announced at the faculty webpage as well as at the webpages of other education institutions, scientific societies, and health chambers to allow all interested persons watching them in real-time. Then, the captured video records were combined with additional educational content and shared using our faculty's portal with no restrictions as well as in infectology e-learning course. The users are allowed to communicate with teachers using on-line discussion tool or directly during face-to-face classes to improve students centered teaching approach. First individual responses indicated positive acceptance of this approach from both the teachers as well as their students, especially because of accessibility of teaching materials. Further questioner based survey is needed to be realized to validate this first subjective feedbacks.

**Acknowledgments** Results presented in this work were obtained with the support of the national grant KEGA 005UPJS-4/2012 (50 %) and in the framework of the project "Support research and development in the Moravian-Silesian Region 2014 DT 1—International research teams" (RRC/07/2014), financed from the budget of the Moravian-Silesian Region (50 %).

## References

1. Masats, D., Dooly, M.: Rethinking the use of video in teacher education: a holistic approach. *Teach. Teach. Educ.* **27**(7), 1151–1162 (2011)
2. Wiens, P.D., Hessberg, K., LoCasale-Crouch, J., DeCoster, J.: Using a standardized video-based assessment in a university teacher education program to examine preservice teachers knowledge related to effective teaching. *Teach. Teach. Educ.* **33**, 24–33 (2013)
3. Goodman, K., Nowacki, A., Wu, J., Hickner, J.: Utilization and impact of pre-office visit video health maintenance education. *Patient. Educ. Couns.* **85**(2), e65–e68 (2011). <http://www.sciencedirect.com/science/journal/07383991/85/2>
4. Komenda, M., Schwarz, D., Feberová, J., Stipek, S., Mihal, V., Dusek, L.: Medical faculties educational network: multidimensional quality assessment. *Comput. Meth. Program. Biomed.* **108**(3), 900–909 (2012)
5. Penhaker, M., Krawiec, J., Krejcar, O., Novak, V., Bridzik, R.: Web system for electrophysiological data management 2010. In: 2nd International Conference on Computer Engineering and Applications, pp. 404–407 (2010). ISBN: 978-0-7695-3982-9
6. Bernera, E.S., Adams, B.: Added value of video compared to audio lectures for distance learning. *Int. J. Med. Inf.* **73**, 189–193 (2004)
7. Kay, R., Kletskin, I.: Evaluating the use of problem-based video podcasts to teach mathematics in higher education. *Comput. Educ.* **59**(2), 619–627 (2012)
8. Reisslein, J., Seeling, P., Reisslein, M.: Video in distance education: ITFS versus web-streaming—evaluation of student attitudes. *Internet High. Educ.* **8**(1), 25–44 (2005)
9. McKinney, A.A., Page, K.: Podcasts and videostreaming: useful tools to facilitate learning of pathophysiology in undergraduate nurse education? *Nurse Educ. Pract.* **9**, 372–376 (2009)
10. Abdous, M.: E-learning quality assurance: a process-oriented lifecycle model. *Qual. Assur. Educ.* **17**, 281–295
11. Toth, T., Zivcak, J.: A comparison of the outputs of 3D scanners. *Procedia Eng.* **69**, 393–401 (2014)

12. Černý, M., Penhaker, M.: Wireless body sensor network in health maintenance systems. *Elektronika ir Elektrotechnika* **9**, 113–116 (2011)
13. Sloan, D., Stratford, J., Gregor, P.: Using multimedia to enhance the accessibility of the learning environment for disabled students: reflections from the skills for access project ALT-J. *Res. Learn. Technol.* **14**(1), 39–54 (2006)
14. Feberová, J., Dostálková, T., Polásková, A., Seydlova, M.: Webbased interactive learning programs for dentistry concept and its evaluation. *MEFANET J.* **1**(1), 5–10 (2013)
15. Lim, S., et al.: Getting closer and experiencing together: antecedents and consequences of psychological distance in social media-enhanced real-time streaming video. *Comput. Hum. Behav.* **28**, 1365–1378
16. Shneiderman, B., Preece, J., Pirolli, P.: Realizing the value of social media requires innovative computing research. *Commun. ACM* **54**(9), 34–37 (2011)
17. Abdous, M., Yoshimura, M.: Learner outcomes and satisfaction: a comparison of live videotransmitted instruction, satellite broadcast instruction, and face-to-face instruction. *Comput. Educ.* **55**, 733–741 (2010)
18. Torres-Ramirez, M., García-Domingo, B., Aguilera, J., de la Casa, J.: Video-sharing educational tool applied to the teaching in renewable energy subjects. *Comput. Educ.* **73**, 160–177 (2014)
19. Bolekova, A., Spakovska, T., Kluchová, D., Toth, S., Vesela, J.: Nitrogen innervation of small intestine in rats after ischemic-reperfusion injury. *Gastroenterologie a Hepatologie* **67**(3), 202–206 (2013)
20. Macura, D., Macurova, A.: Bounded solutions of the nonlinear differential systems. *Int. J. Pure Appl. Math.* **70**(5), 755–760 (2011)
21. Foertsch, J., Moses, G., Strikwerda, J., Litzkow, M.: Reversing the lecture/homework paradigm using eTEACH web-based streaming video software. *J. Eng. Educ.* **91**, 267–275 (2002)

# Evaluation of Postural Stability Using Motion Analysis Techniques

Jaroslav Majerník and Jozef Živčák

**Abstract** The most of the posture stability tests are based on comparison of clinical trials realized in two different conditions, i.e. with and without visual information about surrounding environment. Such tests are performed with eyes open, fixed on certain point in front of the subject, and with eyes closed. The acquired results give us opportunity to understand individual diseases of vestibular system in more details. The main aim of this research study was to integrate the human motion analysis methods to evaluate postural stability and balance control mechanisms in patients with vestibular diseases. A marker-free analysis system was used to detect and evaluate motion of anatomical landmarks in single video camera records. Pilot study was realized at the group of 22 patients to verify both the system functionality and the methodology suitability. All here included patients were evaluated separately as first, because of their different impairments affecting postural control. Then, the results were analysed within the group of here included patients as well. The first results showed that the method was efficient and proved quantitative changes in posture stability parameters as well as their dependence on visual perception. Also, the results obtained in patients with similar symptoms convinced us that the increasing number of analysed subjects and thus increasing database of patients data could help clinicians to identify the background of closely related groups of posture disorders as well as to improve the patients health status by supporting of decision making process aimed to select optimal treatment.

---

J. Majerník (✉)

Faculty of Medicine, Department of Medical Informatics, Pavol Jozef Šafárik University in Košice, Trieda SNP 1, 040 11 Košice, Slovakia  
e-mail: jaroslav.majernik@upjs.sk

J. Živčák

Faculty of Mechanical Engineering, Department of Biomedical Engineering and Measurement, Technical University in Košice, Letná 9, 042 00 Košice, Slovakia  
e-mail: jozef.zivcak@tuke.sk

## 1 Introduction

Human posture stability depends on combination of various internal and external factors and these are referred to the system of human postural control [1]. Its main role is to counteract the effect of gravity and to maintain postural balance under varying terrestrial conditions. However, unwanted postural sways are very often caused by the diseases of posture control system. Postural control of human beings relates to the both visual and vestibular input [2, 3]. Therefore, the stability of bipedal posture can be strongly improved by having visual information about surrounding environment. Considering human movements, it is possible to recognize stationary and moving visual information. Stationary visual information has a stabilizing effect on posture, while the effect of moving visual information destabilizes the posture. To maintain well-balanced posture it is necessary to have an ability of continuous stabilization [4]. Such process usually fatigues the human body.

Patients suffering from vertiginous diseases often have experiences with tilting of their bodies and/or with falls. The understanding of posture control system could be used to understand such postural sway. Control of balance, or postural stability, is essential in all static and dynamic activities [5]. Maintaining balance is more challenging when there are external disturbances from motion of the floor, such as when standing or walking in a moving train, bus, aircraft or ship [6]. Healthy subjects who experience a sudden postural disturbance while standing recover postural stability using appropriate movement strategies involving agonist and antagonist muscle groups according to the environmental context [7]. Postural stability is affected by both the position and the displacement of centre of mass (COM) as well as base of support at any given moment [8]. Here, the postural control assembles the synergies featuring the interplay of visco-elastic and reflexive muscle dynamics with adaptive mechanisms that reflect both anticipatory and compensatory components [9]. The integrity of these control mechanisms depends on multimodal sensory feedback from visual, vestibular, and somatosensory sources. These receptors report changes in position and velocity of the body posture, but such feedback information isn't usually automatically available and one has to use particular stimulus with a primary effect on the sensory inputs to be able to define it [10, 11].

To control human posture it is also necessary to know an expected quality of intended motion/motions as this feedforward information is used in posture stabilization process. A desired stable position of human body directs a controller within the central nervous system (CNS) to produce motor commands that drive particular muscles. These produce joint torques to compensate the potential posture destabilizing effects caused by patient's disease and/or by the gravitational force [12]. Eye and body movements, subjective vertical orientation and motion sickness all depend on the way the human CNS deals with the gravito-inertial force. However, the main cause of vertigo is the indistinguishable acceleration due to movement or gravity [13, 14]. Several studies have shown that patients with a variety of peripheral and central vestibular disorders show abnormal sensory organization testing results during computerized dynamic posturography [15–17]. Posturography has been used

in a number of studies to measure changes in standing and gait as well [18]. The frequently observed stability disorders result from unilateral vestibular deficiency. Vestibular input appears to have a minor role in the control of quiet standing and also for maintenance of fundamental locomotor balance, but the patients with loss of vestibular function perform poorly in such tasks as standing on one foot, heel/toe or while walking. Several studies confirmed that in acute stages, after a unilateral labyrinthine lesion, the walking trajectory has been shown to deviate to the lesion side as body sway increase in the frontal plane. To study the ability to maintain an erecting and stable posture it is very important to consider the visual, proprioception and vestibular system. The result of more balanced postures of individual body segments (head, trunk, lower and upper limbs) also affects the final balance posture of whole body.

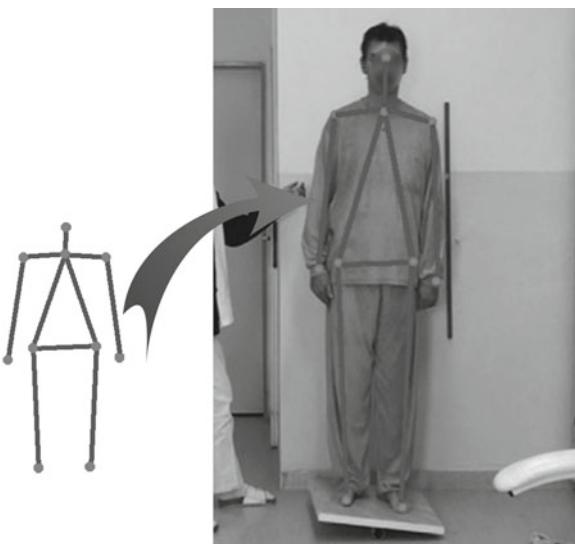
All above mentioned facts confirm that the human posture stability can be affected by many internal and/or external factors. The study of these factors requires more sophisticated methods than is the simple visual evaluation [19, 20]. Such evaluation is not only a subjective approach, but it also does not allow registering of all closely related movements accompanied with efforts to maintain posture stability. Therefore, the objective methods are developed and improved to make the diagnostics and treatment processes more precise and more effective [21].

Most of the clinical research studies use stabilometrical methods with investigation of body COM movements to evaluate patients' posture stability. However, stabilometrical diagnostics evaluates the posture as a complex motion and without distinction of body segments. In this case, it is not possible to assess their effect on maintenance of erecting and stable posture, and also to understand the background of such compensatory movements. Despite of these preferences, there are also alternative methods used to objectivise characteristics of balanced posture. These methods include also human motion analysis techniques that are capable to quantify almost any motions the patients realise during tests oriented on maintaining of posture stability. Therefore, we used our marker-free human motion analysis method to capture motion activities in patients with impairments in postural control, to analyse dynamics of individual body segments and to get the most significant characteristics that can help clinicians to diagnose health status of their patients and to make decisions for further treatment.

## 2 Materials and Methods

Our markerfree motion analysis system MAFRAN was used to analyse motion activities performed by the patients during posture stability tests. The system tracks the motion trajectories in video records and thus analyse human motion data without the need to use any special clothing, contrast background or without the need to make any arrangements in video records. Therefore, no special preparation was needed to be done before the individual tests were captured. This also makes the methodology laboratory independent. The analysis reports kinematical data of human motion in

**Fig. 1** Adaptation of the model to the patient's video record



one plane, as it is designed to process data obtained using single camera records. However, the 3D reconstruction can be also obtained synchronizing two or more cameras.

Here, a model of human body consisting of ten anatomical landmarks was designed to analyse motion in posture stability tests. Anatomical landmarks used to define human body model include forehead, neck, shoulders, right and left anterior superior iliac spine, wrists, and tip-toes. This model was created according to the requirements of clinicians from ORL department and respecting aims of the study. Adaptation of this model to the patient's body captured in video record is shown in Fig. 1.

The position of the model was tracked across the sequence of frames and then the reconstructed trajectories were used to visualize motion of individual anatomical landmarks and adjacent body segments and to calculate other kinematical characteristics, including distances, angles, velocities and accelerations. This computation was based on patients' anthropometrical data noticed together with other patients' data in motion analysis reports.

A group of twenty-two patients (8 males and 14 females, age =  $42.32 \pm 15.13$ ) treated at Department of ORL and Head and Neck Surgery participated in this study. The patients were selected by clinicians and they had various vestibular diseases including those in acute stage. The experimental protocol was approved by the local ethical committee and all here included patients were informed about the study and provided their written informed consent. Experimental protocol was designed to monitor and analyse both the motion of individual body segments and the effect of visual feedback on patients' posture stability. Therefore, two types of tests were realized. The patients were asked to stand on stable platform with arms placed at their

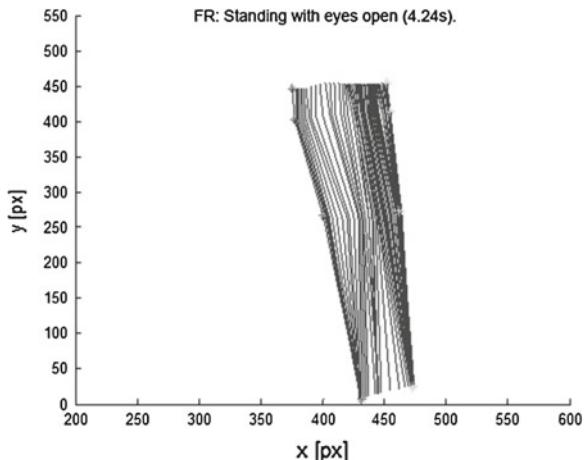
sides and maintain a comfortable posture during the first test. The second test was realized on unstable platform where the rotation axis was parallel to sagittal plane allowing only right and left side movements—RL platform (Fig. 1). Here, the patients were asked to maintain stable posture too. Except of others, the test showed various reflex strategies the patients use to withstand the perturbations. All the tests were realized with eyes opened (EO) and eyes closed (EC), so the patients performed quiet standing on stable platform with EO, standing on stable platform with EC, standing on RL platform with EO and standing on RL platform with EC. Each test was realized with the aim to maintain stable posture during 20 s. 1 min rest period was applied between individual tests.

### 3 Results

All the patients were analysed individually as first. It was because of various diseases affecting postural control of here included patients and because of the need to specify individual treatment procedures to improve their health status. Then the data were analysed within the group to find significant changes or dependences.

The first test on stable platform was successfully performed by all patients, even in conditions with EO and EC during 20 s period except of one 54 years old male patient, who was not able to stand independently for more than 4 s because of acute stage of his vertiginous disease. This patient was supported by the clinician to prevent falls before the test. When the test started and there was no support the patient always went to fall down to the right side no matter which side was used by the clinician to support him when needed. A stick figure of main axes in lower extremities, trunk and head captured in this patient is shown in Fig. 2.

**Fig. 2** Stick figure presenting axes of three main body parts (lower limbs, trunk and head) in a 54 year old male patient in acute stage of disease



**Table 1** Mean distances passed by individual body parts while maintaining posture stability on stable platform

	Distance with EO M ± SD (cm)	Distance with EC M ± SD (cm)
Head	12.03 ± 10.91	17.90 ± 19.38
Neck	9.82 ± 11.43	16.51 ± 19.55
Pelvis	6.16 ± 8.68	12.31 ± 20.58
R wrist	8.07 ± 12.59	10.57 ± 15.70
L wrist	10.73 ± 13.73	16.16 ± 26.09
R tip-toe	4.80 ± 13.44	6.97 ± 16.18
L tip-toe	8.08 ± 23.51	9.76 ± 23.49

Such stick figures showed inclinations of main body parts that were involved in maintaining of posture stability and allowed us to make graphical comparisons of their oscillations through covered areas as well as quantitative comparisons through numerical values of distances, angles, velocities and accelerations.

Table 1 summarizes mean distances realized by individual body parts during quiet standing on stable platform. As it can be seen in Table 1, the most preferred strategies employed by the patients to maintain posture stability on stable platform used compensatory movements of the head, neck and wrists. Comparing differences in passed distances showed us an influence of visual input on postural control. A paired t-test was performed to determine the effect of visual input on posture control.

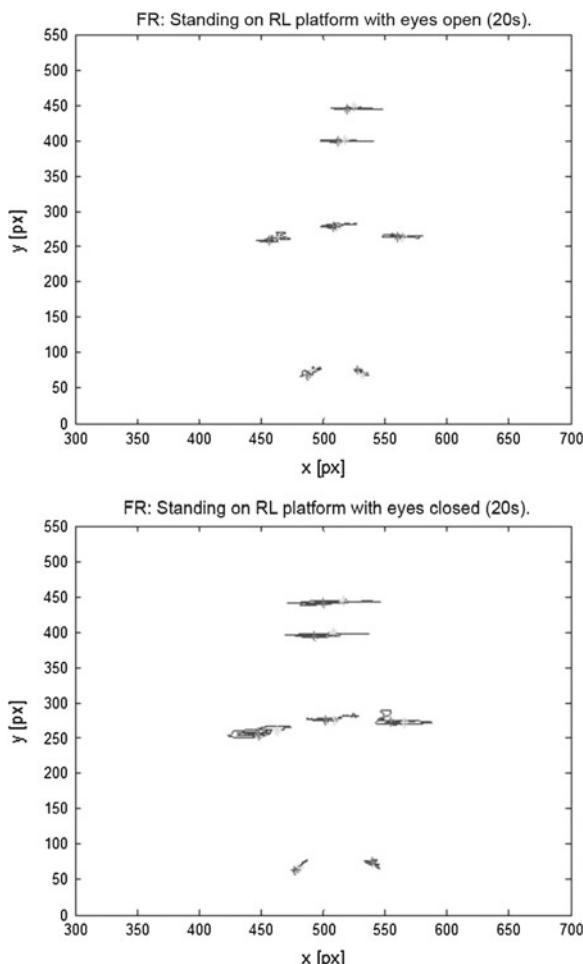
The mean change of the head distance with EO and EC while standing on stable platform ( $M = 5.861$ ,  $SD = 12.77$ ,  $N = 22$ ) was significantly greater than zero and  $t(21) = -2.153$ , two-tail  $p = 0.043$  provided evidence that the absence of visual input increases compensatory motions of the head. A 95 % C.I. about mean distance change of the head is (0.20, 11.52). The mean change of the neck distance with EO and EC while standing on stable platform ( $M = 6.683$ ,  $SD = 13.39$ ,  $N = 22$ ) was significantly greater than zero and  $t(21) = -2.341$ , two-tail  $p = 0.029$  provided evidence that the absence of visual input increases compensatory motions of the neck. A 95 % C.I. about mean distance change of the neck is (0.75, 12.62). The mean change of the pelvic distance with EO and EC while standing on stable platform ( $M = 6.143$ ,  $SD = 13.16$ ,  $N = 22$ ) was significantly greater than zero and  $t(21) = -2.190$ , two-tail  $p = 0.040$  provided evidence that the absence of visual input increases compensatory motions of the pelvic. A 95 % C.I. about mean distance change of the pelvic is (0.31, 11.98). Mean changes in passed distances of other body parts: right wrist ( $t(21) = -1.015$ , two-tail  $p = 0.322$ ), left wrist ( $t(21) = -1.729$ , two-tail  $p = 0.099$ ), right tip-toe ( $t(21) = -2.050$ , two-tail  $p = 0.053$ ) and left tip-toe ( $t(21) = -1.196$ , two-tail  $p = 0.245$ ) while standing on stable platform with EO and EC were not significantly greater than zero providing the evidence that the absence of visual input do not increase compensatory motions in these body parts.

Similarly, more evident changes considering absence of visual information were registered in tests on unstable RL platform. Six of the patients included in this study were not able to perform the test and to maintain posture stability during 20s with EC or even with EO.

Figure 3 shows the trajectories of 18 years old female patient during the tests on unstable RL platform. The unilateral vestibular disease caused that the patient had problems to maintain posture stability and the absence of visual feedback worsened the status.

The most preferred strategies employed by the patients to maintain posture stability on unstable RL platform with EO used compensatory movements of the right and left wrists. The most preferred strategies employed by the patients to maintain posture stability on unstable RL platform with EC used compensatory movements

**Fig. 3** Trajectories of motion (head, neck, pelvis, wrists and tip-toes) in 18 years old female patient during standing on unstable RL platform with eyes open (*up*) and with eyes closed (*down*)



**Table 2** Mean distances passed by individual body parts while maintaining posture stability on unstable RL platform

	Distance with EO M ± SD (cm)	Distance with EC M ± SD (cm)
Head	178.66 ± 68.21	265.02 ± 135.27
Neck	157.12 ± 58.66	235.55 ± 99.79
Pelvis	102.26 ± 27.66	158.56 ± 49.45
R wrist	280.68 ± 178.38	376.36 ± 344.34
L wrist	390.28 ± 257.53	559.01 ± 519.53
R tip-toe	121.63 ± 52.86	159.02 ± 72.61
L tip-toe	131.67 ± 57.36	176.47 ± 95.87

of wrists as it was with EO, but individually there was registered increased motion of head and neck. Table 2 summarizes mean distances realized by individual body parts during standing on unstable RL platform.

The mean change of the head distance with EO and EC while standing on unstable RL platform ( $M = 86.364$ ,  $SD = 105.560$ ,  $N = 16$ ) was significantly greater than zero and  $t(15) = -3.273$ , two-tail  $p = 0.005$  provided evidence that the absence of visual input increases compensatory motions of the head. A 95 % C.I. about mean distance change of the head is (30.11, 142.62). The mean change of the neck distance with EO and EC while standing on RL platform ( $M = 78.427$ ,  $SD = 83.085$ ,  $N = 16$ ) was significantly greater than zero and  $t(15) = -3.776$ , two-tail  $p = 0.0018$  provided evidence that the absence of visual input increases compensatory motions of the neck. A 95 % C.I. about mean distance change of the neck is (34.15, 122.70). The mean change of the pelvic distance with EO and EC while standing on unstable RL platform ( $M = 56.300$ ,  $SD = 49.553$ ,  $N = 16$ ) was significantly greater than zero and  $t(15) = -4.545$ , two-tail  $p = 0.000$  provided evidence that the absence of visual input increases compensatory motions of the pelvic. A 95 % C.I. about mean distance change of the pelvic is (29.90, 82.71). The mean change of the right tip-toe distance with EO and EC while standing on unstable RL platform ( $M = 37.390$ ,  $SD = 69.602$ ,  $N = 16$ ) was significantly greater than zero and  $t(15) = -2.149$ , two-tail  $p = 0.048$  provided evidence that the absence of visual input increases compensatory motions of the right tip-toe. A 95 % C.I. about mean distance change of the right tip-toe is (0.30, 74.48). Mean changes in passed distances of other body parts: right wrist ( $t(15) = -1.709$ , two-tail  $p = 0.108$ ), left wrist ( $t(15) = -1.856$ , two-tail  $p = 0.083$ ) and left tip-toe ( $t(15) = -1.632$ , two-tail  $p = 0.123$ ) while standing on unstable RL platform with EO and EC were not significantly greater than zero providing evidence that the absence of visual input do not increase compensatory motions in these body parts. Comparing inclination angles in main body axes obtained in tests on unstable RL platform we did not find statistically significant changes, but the patients tends to fall down rather to the right side (EO: 49.60 %, EC 51.08 %) than to the left one (EO: 43.22 %, EC: 41.81 %).

## 4 Conclusion

In this study, the usability of marker-free human motion analysis method in monitoring of patients' posture control mechanisms was verified. Using this method, the strategies to maintain posture stability and the influence of visual feedback on posture control in 22 patients were evaluated. Patients suffered from various impairments of postural control as well as the patients in acute phase of vestibular neuritis were included in this study. The patients were tested on stable and unstable platforms with and without visual input. The most preferred strategies to maintain posture stability on stable platform used compensatory motions of the head and neck, while these were replaced by corrective motions of patients' upper extremities (analysed as motions of right and left wrists) in the tests on unstable RL platform. The absence of visual feedback significantly changed posture control mechanisms that were performed by the head, neck and pelvic during standing on stable platform and by the head, neck and wrists during standing on unstable RL platform. We also recognised that the patients with vestibular neuritis were more unstable, but fallings from RL platform were registered only in patients with central vestibular or proprioceptive diseases. Older patients and/or patients in acute stage of disease were not able to perform tests on unstable RL platform at all. All obtained data were stored in database to be able to increase number of participants as well as to realise further kinematical analyses.

Utilization of optical motion analysis system helped us to obtain qualitative and quantitative information about the maintenance of erected and stable posture in patients included in this study and thus to objectivize changes in posture control. These changes are very important, especially during acute phase of vestibular diseases, for clinical examinations and planning of the most convenient vestibular trainings and treatment procedures applied in individual patients. As we supposed, the changes of information from proprioceptive, vestibular and visual systems that participate on body orientation in space, disturbed the posture stability primarily on unstable RL platform. The level of disturbance depended on affected part of vestibular system and on the severity of patient's vestibular dysfunction.

**Acknowledgments** Results presented in this work were obtained with the support of the national grant KEGA 020UK-4/2014 (50 %) and in the framework of the project "Support research and development in the Moravian-Silesian Region 2014 DT 1—International research teams" (RRC/07/2014), financed from the budget of the Moravian-Silesian Region (50 %).

## References

1. Laurens, J., et al.: Visual contribution to postural stability: interaction between target fixation or tracking and static or dynamic large-field stimulus. *Gait & Posture* **31**, 37–41 (2010)
2. Bos, J.E., Bles, W.: Theoretical consideration on canalolith interaction and an observer model. *Biol. Cybern.* **86**, 191–207 (2002)
3. Park, S., Horak, F.B., Kuo, A.D.: Postural feedback responses scale with biomechanical constraints in human standing. *Exp. Brain Res.* **154**, 417–427 (2004)

4. Simsik, D., Galajdova, A., Hrabska, I., Molcan, M.: Posture disorders diagnostics using videoanalysis. *Int. J. Rehabil. Res.* **27**, suppl. 1 (2004). ISSN 0342–5282, 82–83
5. Brenton-Rule, A., D’Almeida, S., Bassett, S., Carroll, M., Dalbeth, N., Rome, K.: The effects of sandals on postural stability in patients with rheumatoid arthritis: an exploratory study. *Clin. Biomech.* **29**, 350–353 (2014)
6. Sari, H.M., Griffin, M.J.: Postural stability when walking: effect of the frequency and magnitude of lateral oscillatory motion. *Appl. Ergon.* **45**, 293–299 (2014)
7. Jeon, H.-S., Hwang, S., Woo, Y.-K.: The effect of ankle and knee immobilization on postural control during standing. *Knee* **20**, 600–604 (2013)
8. Ching-Lung, L., Chun-Liang, L., Ching-Kun, Ch.: Stabilizing postural control for emulated human balancing systems. *Int. J. Eng. Sci.* **46**, 1120–1135 (2008)
9. Yeha, T.T., Bouleta, J., Cluff, T., Balasubramaniam, R.: Contributions of delayed visual feedback and cognitive task load to postural dynamics. *Neurosci. Lett.* **481**, 173–177 (2010)
10. De Nunzo, A.M., Nardone, A., Schieppati, M.: Head stabilization on a continuously oscillating platform: the effect of a proprioceptive disturbance on the balancing strategy. *Exp. Brain Res.* **165**, 261–272 (2005)
11. Johansson, R., Magnusson, M., Fransson, P.A.: Galvanic vestibular stimulation for analysis of postural adaptation and stability. *IEEE Trans. Biomed. Eng.* **42**, 282–291 (1995)
12. Badke, M.B., Shea, T.A., Miedaner, J.A., Grove, C.R.: Outcomes after rehabilitation for adults with balance dysfunction. *Arch. Phys. Med. Rehabil.* **85**(2), 227–233 (2004)
13. Bronstein, A., Woollacott, M.: Clinical Disorders of Balance, Gait and Posture, National Hospital for Neurology and Neurosurgery. Elsevier Science, London (1986)
14. Molcan, M., Zivcak, J., Simsik, D., Majerník, J.: 3D analysis of human locomotion in subjects with unilateral vestibular deficiency. In: 5th European Congress of Oto-Rhino-Laryngology Head and Neck Surgery, Rodos-Kos, Hellas, September 11–16, p. 139 (2004)
15. Černý, M., Drska, R., Penhaker, M.: Automated measurement of electrochemical sensors. *IFMBE Proc.* **39**, 1521–1524 (2013)
16. Majerník, J., Molcan, M., Majerníkova, Z.: Evaluation of posture stability in patients with vestibular diseases. In: SAMI 2010—8th International Symposium on Applied Machine Intelligence and Informatics, Proceedings, pp. 271–274 (2010), art. no. 5423722
17. Majerník, J., Simsik, D.: Marker-Free, Analysis of Human Gait, EMBEC’05, November 20–25, Prague, Czech Republic, IFMBE Proceedings, vol. 11 (2005). ISSN 1727-1983
18. Penhaker, M., Stula, T., Augustynek, M.: Long-term heart rate variability assessment. *IFMBE Proc.* **35**, 532–535 (2011)
19. Macurova, A., Macura, D.: Asymptotic properties of the variables of the roughness surface. *Appl. Surf. Sci.* **269**, 134–137 (2013)
20. Molcan, M., Majerník, J., Simsik, D.: 3D motion analysis in evaluation of posture. *Eur. Arch. Oto-Rhino-Laryngol.* **264**, S199, Supplement 1 (2007). ISSN 0937-4477
21. Salami, A., Dellepiane, M., Crippa, B., Baretini, L., Mora, R.: Visual-vestibular interaction test in the diagnosis of vertigo in children. *Int. J. Pediatr. Otorhinolaryngol.* **72**, 1–7 (2008)

# Some Issues on an Object-Oriented Programming Language for *Physarum* Machines

Krzysztof Pancerz and Andrew Schumann

**Abstract** The main goal of the paper is to describe essential aspects of a new object-oriented programming (OOP) language, called the *Physarum* language, for *Physarum* machines. A *Physarum* machine is a biological computing device experimentally implemented in the plasmodium of *Physarum polycephalum*, a single cell organism, also called true slime mould. The main attention is focused on a language specification based on selected high-level models describing behaviour of *Physarum* machines as well as a computer tool created for parsing the *Physarum* language statements and producing a machine code that is, in case of *Physarum* machines, a spatial distribution of stimuli.

## 1 Introduction

A *Physarum* machine is a programmable amorphous biological computing device experimentally implemented in the plasmodium of *Physarum polycephalum*, also called true slime mould [1]. *Physarum polycephalum* is a single cell organism belonging to the species of order Physarales, subclass Myxogastromycetidae, class Myxomycetes, and division Myxostelida. The plasmodium of *Physarum polycephalum* (see Fig. 1) spread by networks can be programmable. In propagating and foraging behaviour of the plasmodium, we can perform useful computational tasks. This ability was firstly discerned by Nakagaki et al. (cf. [7]).

Solving computational tasks by means of *Physarum* machines is one of the main goals of the *Physarum* Chip Project: Growing Computers from Slime Mould [2] funded by the Seventh Framework Programme (FP7). In this project, we are going

---

K. Pancerz (✉) · A. Schumann

University of Information Technology and Management, Rzeszów, Poland

e-mail: kpancerz@wsiz.rzeszow.pl

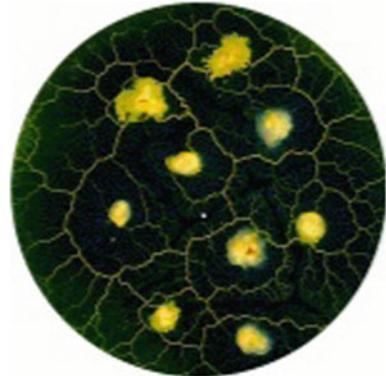
A. Schumann

e-mail: aschumann@wsiz.rzeszow.pl

K. Pancerz

University of Management and Administration, Zamość, Poland

**Fig. 1** The plasmodium of *Physarum polycephalum* [12]



to construct an unconventional computer on programmable behaviour of *Physarum polycephalum*. The *Physarum* machine comprises an amorphous yellowish mass with networks of protoplasmic veins, programmed by spatial configurations of attracting and repelling stimuli. When several attractants are scattered in the plasmodium range, the plasmodium forms a network of protoplasmic veins connecting those attractants as well as original points of the plasmodium. The plasmodium looks for attractants, propagates protoplasmic veins towards them, feeds on them and goes on. As a result, a transition system is built up. Therefore, *Physarum* motions can be treated as a kind of a natural transition system [1] with states presented by attractants. We extend a set of states of the transition system to those presented by original points of the plasmodium. Events are presented by plasmodium transitions between attractants as well as original points of the plasmodium. In a real-life implementation of *Physarum* machines, attractants are sources of nutrients or pheromones, on which the plasmodium feeds. In case of repellents, the fact that the plasmodium of *Physarum* avoids light and some thermo- and salt-based conditions is used.

To program computational tasks for plasmodium transitions, we are developing a new object-oriented programming language [9, 14], called the *Physarum* language. In general, object-oriented programming (OOP) is a programming paradigm based on the concept of the object that has data fields describing the object and associated procedures, called methods, for manipulating the data. We can distinguish two main approaches in OOP programming languages: class-based and prototype-based languages [4]. The prototype-based approach is less common than the class-based one, although, it has a great deal to offer. This model is also called class-less or instance-based programming because prototype-based languages are based upon the idea that objects that represent individuals can be created without reference to class-defining. In this approach, the objects that are manipulated at runtime (the objects that make it an “object-oriented” approach) are the prototypes. JavaScript, the very popular now prototype-based language, has been an inspiration to us and we have implemented a number of its mechanisms in the *Physarum* language. For example, there are inbuilt sets of prototypes corresponding to both the high-level models used for describing behaviour of *Physarum polycephalum* (e.g., ladder diagrams, transition

systems, timed transition systems, Petri nets) and the low-level model (distribution of stimuli). According to the prototype-based approach, objects are created by means of a copy operation, called cloning, which is applied to a prototype. Objects can be instantiated (cloning) via the keyword *new* using defined constructors. Methods are used to manipulate features of the objects and create relationships between objects.

We have defined our *Physarum* prototype-based language for describing the behaviour of given systems using some well-known abstract models. In the paper, we are focused on two of them, namely, transition systems (including timed transition systems) as well as Petri nets. These models are the basis for programming *Physarum* machines. For the programming purpose, a compiler embodied in our tool translates the high-level models into the spatial distribution (configuration) of stimuli (attractants and/or repellents) for *Physarum* machines. Transition system and Petri net models of behaviour of *Physarum* machines are described in Sect. 2. Four basic forms of *Physarum* motions are considered, namely, direction, fusion, splitting, and repelling/inaction (cf. [1]). Next, in Sect. 3, the implementation of described high-level models in the created *Physarum* language is depicted.

## 2 Selected High-Level Models of Behaviour of *Physarum* Machines

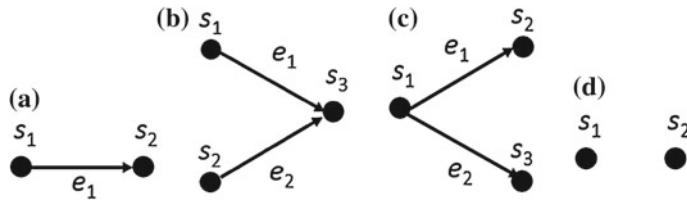
Most modern programming languages are based on high-level abstractions enabling us to deal with tedious and difficult details at a lower level in an easier way. Moreover, created models are closer to reality and more intelligible for humans. In our approach, the starting point in programming the behaviour of the *Physarum* machine is a high-level model describing *Physarum* motions. As it was mentioned in Sect. 1, we are interested, in this paper, in two well-known abstract models, transition systems (including timed transition systems) and Petri nets. In this section, we show how to use such high-level models to describe four basic forms of *Physarum* motions:

- *direct*—direction, i.e., a movement from one point, where the plasmodium is located, towards another point, where there is a neighbouring attractant,
- *fuse*—fusion of two plasmodia at the point, where they meet the same attractant,
- *split*—splitting the plasmodium from one active point into two active points, where two neighbouring attractants with a similar power of intensity are located,
- *repel*—repelling of the plasmodium or inaction.

It is worth noting that four basic forms can be treated as fundamental components which can be used to build or describe more complex systems.

Transition systems are a simple and powerful tool for explaining the operational behaviour of models of concurrency. Formally, a transition system is a quadruple  $TS = (S, E, T, I)$ , cf. [8], where:

- $S$  is the non-empty set of states,
- $E$  is the set of events,
- $T \subseteq S \times E \times S$  is the transition relation,



**Fig. 2** Transition system models of **a** *direct*, **b** *fuse*, **c** *split*, and **d** *repel*

- $I$  is the set of initial states.

Usually transition systems are based on actions which may be viewed as labelled events. If  $(s, e, s') \in T$ , then the idea is that  $TS$  can go from  $s$  to  $s'$  as a result of the event  $e$  occurring at  $s$ . Any transition system  $TS = (S, E, T, I)$  can be presented in the form of a labelled graph with nodes corresponding to states from  $S$ , edges representing the transition relation  $T$ , and labels of edges corresponding to events from  $E$ .

In Fig. 2, transition system models of four basic forms of *Physarum* motions, mentioned earlier, are shown. Formally, we have:

- *direct*:  $TS_d = (S_d, E_d, T_d, I_d)$ , where

- $S_d = \{s_1, s_2\}$ ,
- $E_d = \{e_1\}$ ,
- $T_d = \{(s_1, e_1, s_2)\}$ ,
- and  $I_d = \{s_1\}$ ,

see Fig. 2a,

- *fuse*:  $TS_f = (S_f, E_f, T_f, I_f)$ , where

- $S_f = \{s_1, s_2, s_3\}$ ,
- $E_f = \{e_1, e_2\}$ ,
- $T_f = \{(s_1, e_1, s_3), (s_2, e_2, s_3)\}$ ,
- and  $I_f = \{s_1, s_2\}$ ,

see Fig. 2b,

- *split*:  $TS_s = (S_s, E_s, T_s, I_s)$ , where

- $S_s = \{s_1, s_2, s_3\}$ ,
- $E_s = \{e_1, e_2\}$ ,
- $T_s = \{(s_1, e_1, s_2), (s_1, e_2, s_3)\}$ ,
- and  $I_s = \{s_1\}$ ,

see Fig. 2c,

- *repel*:  $TS_r = (S_r, E_r, T_r, I_r)$ , where

- $S_r = \{s_1, s_2\}$ ,
- $E_r = \emptyset$ ,
- $T_r = \emptyset$ ,

– and  $I_r = \{s_1\}$ ,

see Fig. 2d.

Each state corresponds to either an original point of the plasmodium or an attractant. Especially, initial states of transition systems can be presented by original points, where protoplasmic veins originate from. Edges represent plasmodium transitions between attractants as well as original points of the plasmodium.

The behaviour of *Physarum* machines is often dynamically changed in time. It is assumed, in transition systems mentioned earlier, that all events happen instantaneously. Therefore, in [13], we have proposed to use another high-level model, based on timed transition systems [5]. In the timed transition systems, timing constraints restrict the times at which events may occur. The timing constraints are classified into two categories: lower-bound and upper-bound requirements.

Let  $N$  be a set of nonnegative integers. Formally, a timed transition system  $TTS = (S, E, T, I, l, u)$  consists of:

- an underlying transition system  $TS = (S, E, T, I)$ ,
- a minimal delay function (a lower bound)  $l : E \rightarrow N$  assigning a nonnegative integer to each event,
- a maximal delay function (an upper bound)  $u : E \rightarrow N \cup \infty$  assigning a nonnegative integer or infinity to each event.

In *Physarum* machines, timing constraints can be implemented through activation and deactivation of stimuli (attractants and/or repellents).

Petri nets introduced by Petri [11] are a formal tool used to model discrete event systems. In [15], we proposed to use Petri nets with inhibitor arcs (cf. [3]) to model behaviour of *Physarum polycephalum*. The inhibitor arcs test the absence of tokens in a place and they can be used to disable transitions. This fact can model repellents in *Physarum* machines. Repellents play a role of elements blocking plasmodium transitions between attractants. A transition can only fire if all its places connected through inhibitor arcs are empty (cf. [16]).

Formally, a marked Petri net with inhibitor arcs is a five-tuple

$$MPN = \{Pl, Tr, Ar, w, m\},$$

where:

- $Pl$  is the finite set of places (marked graphically with circles),
- $Tr$  is the finite set of transitions (marked graphically with rectangles),
- $Ar = Ar_O \cup Ar_I$  such that  $Ar_O \subseteq (Pl \times Tr) \cup (Tr \times Pl)$  is the set of ordinary arcs (marked graphically with arrows) from places to transitions and from transitions to places whereas  $Ar_I \subseteq Pl \times Tr$  is the set of inhibitor arcs (marked graphically with lines ended with small circles) from places to transitions,
- $w : Ar \rightarrow \{1, 2, 3, \dots\}$  is the weight function on the arcs,
- $m : Pl \rightarrow \{0, 1, 2, \dots\}$  is the initial marking function on the places.

In describing the Petri net behaviour, it is convenient to use for any  $t \in Tr$ :

- $I_O(t) = \{p \in Pl : (p, t) \in Ar_O\}$ —a set of input places connected through ordinary arcs to the transition  $t$ ,
- $I_I(t) = \{p \in Pl : (p, t) \in Ar_I\}$ —a set of input places connected through inhibitor arcs to the transition  $t$ ,
- $O(t) = \{p \in Pl : (t, p) \in Ar_O\}$ —a set of output places connected through ordinary arcs from the transition  $t$ .

In the proposed approach, we have additionally assumed the following limits for the Petri net:

- $w(a) = 1$  for each  $a \in Ar$ ,
- $m(p) \leq 1$  for each  $p \in Pl$  (the capacity limit).

If  $m(p) = 1$ , then a token (i.e., a black dot) is drawn in the graphical representation of the place  $p$ . Assuming limits as the ones above, a transition  $t \in Tr$  is said to be enabled if and only if  $m(p) = 1$  for all  $p \in I_O(t)$ , i.e., the token is present in all input places  $p$  connected with the transition  $t$  through the ordinary arcs, and  $m(p) = 0$  for all  $p \in I_I(t)$ , i.e., the token is absent in all input places  $p$  connected with the transition  $t$  through the inhibitor arcs, and  $m(p) = 0$  for all  $p \in O(t)$ , i.e., the token is absent in all output places  $p$  of the transition  $t$ . If the transition  $t$  is enabled, we say that it can fire. A new marking function  $m' : Pl \rightarrow \{0, 1, 2, \dots\}$  defines the next state of the Petri net after firing the transition  $t$ :

$$m'(p) = \begin{cases} m(p) - 1 & \text{if } p \in I_O(t) \text{ and } p \notin O(t), \\ m(p) + 1 & \text{if } p \in O(t) \text{ and } p \notin I_O(t), \\ m(p) & \text{otherwise.} \end{cases}$$

Further, for any place  $p \in Pl$ ,  $m(p)$  denotes both initial marking and any new marking that appears after firing a transition.

In Figs. 3 and 4, Petri net models of four basic forms of *Physarum* motions, mentioned earlier, are shown. Formally, we have:

- *direct*:  $MPN_d = \{Pl_d, Tr_d, Ar_d, w_d, m_d\}$ , where:

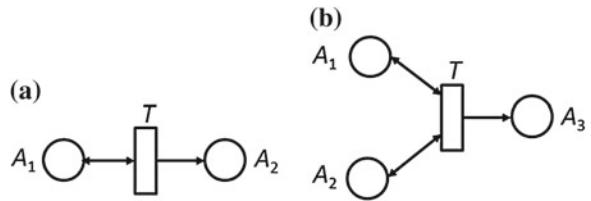
- $Pl_d = \{A_1, A_2\}$ ,
- $Tr_d = \{T\}$ ,
- $Ar_d = Ar_O^d \cup Ar_I^d$ ,  $Ar_O^d = \{(A_1, T), (T, A_1), (T, A_2)\}$ ,  $Ar_I^d = \emptyset$ ,
- $w_d(a) = 1$  for all  $a \in Ar$ ,
- and  $m_d(p) = 0$  for each  $p \in Pl$ ,

see Fig. 3a,

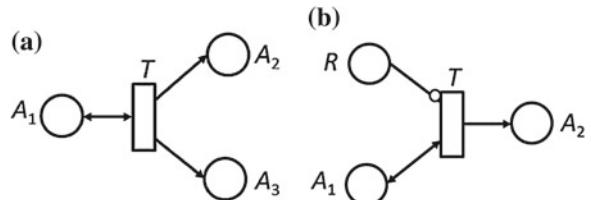
- *fuse*:  $MPN_f = \{Pl_f, Tr_f, Ar_f, w_f, m_f\}$ , where:

- $Pl_f = \{A_1, A_2, A_3\}$ ,
- $Tr_f = \{T\}$ ,
- $Ar_f = Ar_O^f \cup Ar_I^f$ ,  $Ar_O^f = \{(A_1, T), (T, A_1), (A_2, T), (T, A_2), (T, A_3)\}$ ,  $Ar_I^f = \emptyset$ ,
- $w_f(a) = 1$  for all  $a \in Ar$ ,
- and  $m_f(p) = 0$  for each  $p \in Pl$ ,

**Fig. 3** Petri net models of  
**a** direct and **b** fuse



**Fig. 4** Petri net models of  
**a** split and **b** repel



see Fig. 3b,

- *split*:  $MPN_s = \{Pl_s, Tr_s, Ar_s, w_s, m_s\}$ , where:

- $Pl_s = \{A_1, A_2, A_3\}$ ,
- $Tr_s = \{T\}$ ,
- $Ar_s = Ar_O^s \cup Ar_I^s$ ,  $Ar_O^s = \{(A_1, T), (T, A_1), (T, A_2), (T, A_3)\}$ ,  $Ar_I^s = \emptyset$ ,
- $w_s(a) = 1$  for all  $a \in Ar$ ,
- and  $m_s(p) = 0$  for each  $p \in Pl$ ,

see Fig. 4a,

- *repel*:  $MPN_r = \{Pl_r, Tr_r, Ar_r, w_r, m_r\}$ , where:

- $Pl_r = \{A_1, A_2, R\}$ ,
- $Tr_r = \{T\}$ ,
- $Ar_r = Ar_O^r \cup Ar_I^r$ ,  $Ar_O^r = \{(A_1, T), (T, A_1), (T, A_2)\}$ ,  $Ar_I^r = \{(R, T)\}$ ,
- $w_r(a) = 1$  for all  $a \in Ar$ ,
- and  $m_r(p) = 0$  for each  $p \in Pl$ ,

see Fig. 4b.

It is worth noting that in all figures including Petri net models, to simplify them, we have used bidirectional arcs between input places and transitions instead of arcs from input places to transitions and from transitions to input places. A bidirectional arc causes that the token is not consumed (removed) from the input place after firing a transition. This fact has a natural justification, i.e., firing a transition does not cause deactivation of the attractants and disappearance of the plasmodium from the origin point. The plasmodium grows to build a dendritic network of veins.

In the proposed Petri net models of *Physarum* machines, we can distinguish three kinds of places:

**Table 1** The meaning of tokens in places representing control stimuli

Token	Meaning
Present	Stimulus activated
Absent	Stimulus deactivated

**Table 2** The meaning of tokens in places representing output stimuli

Token	Meaning
Present	Stimulus occupied by plasmodium of <i>Physarum polycephalum</i>
Absent	Stimulus not occupied by plasmodium of <i>Physarum polycephalum</i>

- Places representing *Physarum polycephalum*.
- Places representing control stimuli (attractants or repellents).
- Places representing output stimuli (attractants).

In the *Physarum* language, the kind of a place is determined by the role played by it (see Sect. 3).

For each kind of places, we adopt different meaning (interpretation) of tokens. The meaning of tokens in places representing *Physarum polycephalum* is natural, i.e., the token in a given place corresponds to the presence of the plasmodium of *Physarum polycephalum* in an original point, where it starts to grow. The meaning of tokens in places representing control stimuli is shown in Table 1, whereas the meaning of tokens in places representing output stimuli is shown in Table 2. In case of control stimuli, we are interested in whether a given stimulus is activated or not. In case of output stimuli (attractants), we are interested in whether a given attractant is occupied by the plasmodium of *Physarum polycephalum*. Transitions in Petri net models represent the flow (propagation) of the plasmodium from the origin points to attractants as well as between attractants.

### 3 *Physarum* Language Implementation of High-Level Models

In this section, we describe implementation of high-level models, described in Sect. 2, in the created *Physarum* language. A grammar of the language has been described by means of the Java Compiler Compiler (JavaCC) [6] tool. To describe the syntax of the language, we use the Extended Backus-Naur Form (EBNF) notation, cf. [10]. In the EBNF notation, we use the following nonterminals:

- ID for the identifier,
- STRING for the sequence of characters,

**Table 3** Main prototypes, corresponding to transition system and timed transition system models, defined in the *Physarum* language, and their selected methods

Prototype	Selected methods
TS.State	<i>setDescription</i> , <i>setAsInitial</i>
TS.Event	<i>setDescription</i> , <i>setTimingConstraints</i>
TS.Transition	

- INT for the integer value,
- INT\_WITH\_INFINITY for the integer value or infinity symbol (*inf*).

In case of transition system and timed transition system models, the main prototypes defined in the *Physarum* language and their selected methods are collected in Table 3.

The fragment of a grammar describing the *Physarum* language used to build transition system and timed transition system models of *Physarum* machines has the following EBNF form:

```

TransitionSystem = "#TRANSITION_SYSTEM",
{TransitionSystemExpression, ";"};

TransitionSystemExpression = ID, (TransitionSystemCreation
| TransitionSystemManipulation);

TransitionSystemCreation = "=", "new", (TSStateCreation
| TSEventCreation | TSTransitionCreation);
TSStateCreation = "TS.State", "(", STRING, ")";
TSEventCreation = "TS.Event", "(", STRING, ")";
TSTransitionCreation = "TS.Transition",
"(", ID, ",", ID, ",", ID, ")";

TransitionSystemManipulation = ".",
(TSLabeledElementManipulation | TSStateManipulation
| TSEventManipulation);
TSLabeledElementManipulation = "setDescription",
"(", STRING, ")";
TSStateManipulation = "setAsInitial";
TSEventManipulation = "setTimingConstraints",
"(", INT, ",", INT_WITH_INFINITY, ")";

```

Below, we show implementation of transition system models of four basic forms of *Physarum* motions, in the *Physarum* language, described in Sect. 2:

- *direct*:

```
#TRANSITION_SYSTEM
s1=new TS.State("s1");
s1.setAsInitial;
s2=new TS.State("s2");
e1=new TS.Event("e1");
t1=new TS.Transition(s1,e1,s2);
```

- *fuse*:

```
#TRANSITION_SYSTEM
s1=new TS.State("s1");
s1.setAsInitial;
s2=new TS.State("s2");
s2.setAsInitial;
s3=new TS.State("s3");
e1=new TS.Event("e1");
e2=new TS.Event("e2");
t1=new TS.Transition(s1,e1,s3);
t2=new TS.Transition(s2,e2,s3);
```

- *split*:

```
#TRANSITION_SYSTEM
s1=new TS.State("s1");
s1.setAsInitial;
s2=new TS.State("s2");
s3=new TS.State("s3");
e1=new TS.Event("e1");
e2=new TS.Event("e2");
t1=new TS.Transition(s1,e1,s2);
t2=new TS.Transition(s1,e2,s3);
```

- *repel*:

```
#TRANSITION_SYSTEM
s1=new TS.State("s1");
s1.setAsInitial;
s2=new TS.State("s2");
```

In case of Petri net models, the main prototypes defined in the *Physarum* language and their selected methods are collected in Table 4.

The fragment of a grammar describing the *Physarum* language used to build Petri net models of *Physarum* machines has the following EBNF form:

**Table 4** Main prototypes, corresponding to Petri net models, defined in the *Physarum* language, and their selected methods

Prototype	Selected methods
TS.Place	<i>setDescription</i> , <i>setRole</i>
TS.Transition	<i>setDescription</i>
TS.Arc	<i>setAsInhibitor</i> , <i>setAsBidirectional</i>

```

PetriNet = "#PETRI_NET", {PetriNetExpression, ";"};

PetriNetExpression = ID, (PetriNetCreation
| PetriNetManipulation);

PetriNetCreation = "=". "new", (PNTransitionCreation
| PNPlaceCreation | PNArcCreation);
PNTransitionCreation = "PN.Transition", "(" , STRING, ")";
PNPlaceCreation = "PN.Place", "(" , STRING, ")";
PNArcCreation = "PN.Arc", "(" , ID, ", ", ID, ")";

PetriNetManipulation = ".", (PNArcManipulation
| PNPlaceManipulation | PNNodeManipulation);
PNArcManipulation = "setAsInhibitor" | "setAsBidirectional";
PNPlaceManipulation = "setRole", "(" , STRING, ")";
PNNodeManipulation = "setDescription", "(" , STRING, ")";

```

Below, we show the implementation of Petri net models of four basic forms of *Physarum* motions, in the *Physarum* language, described in Sect. 2:

- *direct*:

```

#PETRI_NET
A1=new PN.Place("A1");
A1.setRole(PN.OUTPUT_STIMULUS);
A2=new PN.Place("A2");
A2.setRole(PN.OUTPUT_STIMULUS);
T=new PN.Transition("T");
a1=new PN.Arc(A1,T);
a1.setAsBidirectional;
a2=new PN.Arc(T,A2);

```

- *fuse*:

```

#PETRI_NET
A1=new PN.Place("A1");
A1.setRole(PN.OUTPUT_STIMULUS);

```

```

A2=new PN.Place("A2");
A2.setRole(PN.OUTPUT_STIMULUS);
A3=new PN.Place("A3");
A3.setRole(PN.OUTPUT_STIMULUS);
T=new PN.Transition("T");
a1=new PN.Arc(A1,T);
a1.setAsBidirectional;
a2=new PN.Arc(A2,T);
a2.setAsBidirectional;
a3=new PN.Arc(T,A3);

```

- *split*:

```

#PETRI_NET
A1=new PN.Place("A1");
A1.setRole(PN.OUTPUT_STIMULUS);
A2=new PN.Place("A2");
A2.setRole(PN.OUTPUT_STIMULUS);
A3=new PN.Place("A3");
A3.setRole(PN.OUTPUT_STIMULUS);
T=new PN.Transition("T");
a1=new PN.Arc(A1,T);
a1.setAsBidirectional;
a2=new PN.Arc(T,A2);
a3=new PN.Arc(T,A3);

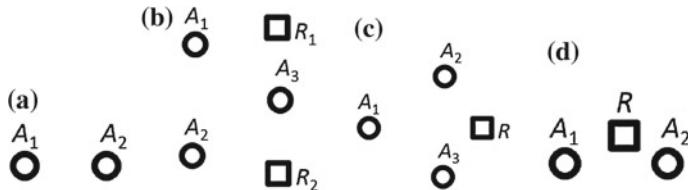
```

- *repel*:

```

#PETRI_NET
A1=new PN.Place("A1");
A1.setRole(PN.OUTPUT_STIMULUS);
A2=new PN.Place("A2");
A2.setRole(PN.OUTPUT_STIMULUS);
R=new PN.Place("R");
A3.setRole(PN.CONTROL_STIMULUS);
T=new PN.Transition("T");
a1=new PN.Arc(A1,T);
a1.setAsBidirectional;
a2=new PN.Arc(T,A2);
a3=new PN.Arc(R,T);
a3.setAsInhibitor;

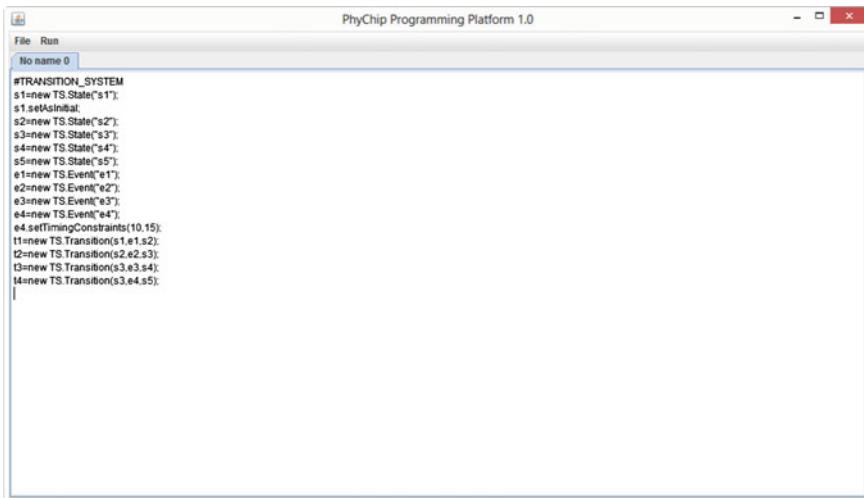
```



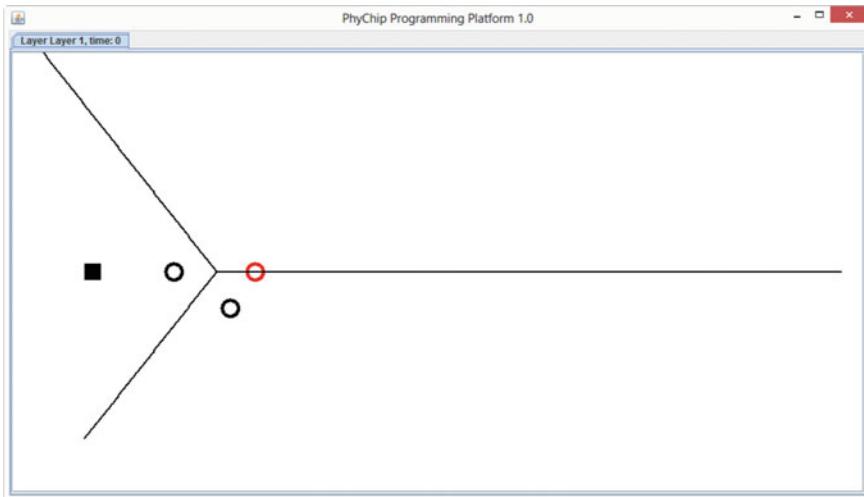
**Fig. 5** Exemplary distribution of stimuli for **a** *direct*, **b** *fuse*, **c** *split*, and **d** *repel*

Each high-level model is translated into the low-level language, i.e., spatial distribution of stimuli (attractants and/or repellents). Such distribution can be treated as a program for the *Physarum* machine. Exemplary distribution of stimuli for each of four basic forms of *Physarum* motions is shown in Fig. 5, where  $A_1, A_2, A_3$  are attractants,  $R, R_1, R_2$  are repellents. In case of fusion and splitting, proper propagation of the plasmodium is supported by repellents.

A computer tool created for parsing the *Physarum* language statements and producing a machine code, that is a spatial distribution of stimuli, has been created with a graphical user interface in the Java environment, see Figs. 6 and 7.



**Fig. 6** A window with the program code



**Fig. 7** A window with the resulting code

## 4 Conclusions

In the paper, we have shown essential aspects of a new object-oriented programming (OOP) language, called the *Physarum* language, created for *Physarum* machines. Two well-known abstract models have been used to describe basic forms of *Physarum* motions. Special attention has been focused on language specification and implementation of considered models. The main direction for further work is to propose to use some other high-level models, based on different formalisms (methodologies), for describing and programming behaviour of *Physarum* machines. Among the most interesting ones, there are  $\pi$ -calculus, spatial logic, probability theory, and game theory.

**Acknowledgments** This research is supported by FP7-ICT-2011-8.

## References

1. Adamatzky, A.: *Physarum Machines: Computers from Slime Mould*. World Scientific, Singapore (2010)
2. Adamatzky, A., Erokhin, V., Grube, M., Schubert, T., Schumann, A.: Physarum chip project: growing computers from slime mould. *Int. J. Unconv. Comput.* **8**(4), 319–323 (2012)
3. Agerwala, T., Flynn, M.: Comments on capabilities, limitations and ‘correctness’ of Petri nets. In: Proceedings of the 1st Annual Symposium on Computer Architecture (ISCA’1973), pp. 81–86. Atlanta, USA (1973)
4. Craig, I.: *Object-Oriented Programming Languages: Interpretation*. Springer, London (2007)

5. Henzinger, T.A., Manna, Z., Pnueli, A.: Timed transition systems. In: de Bakker, J., Huizing, C., de Roever, W., Rozenberg, G. (eds.) *Real-Time: Theory in Practice*. Lecture Notes in Computer Science, vol. 600, pp. 226–251. Springer, Berlin (1992)
6. JavaCC. <http://java.net/projects/javacc/>
7. Nakagaki, T., Yamada, H., Toth, A.: Maze-solving by an amoeboid organism. *Nature* **407**, 470–470 (2000)
8. Nielsen, M., Rozenberg, G., Thiagarajan, P.: Elementary transition systems. *Theor. Comput. Sci.* **96**(1), 3–33 (1992)
9. Pancerz, K., Schumann, A.: Principles of an object-oriented programming language for Physarum polycephalum computing. In: Proceedings of the 10th International Conference on Digital Technologies (DT'2014), pp. 273–280. Zilina, Slovak Republic (2014)
10. Pattis, R.E.: A notation to describe syntax. <http://www.cs.cmu.edu/pattis/misc/ebnf.pdf>
11. Petri, C.A.: Kommunikation mit automaten. Schriften des IIM nr. 2, Institut für Instrumentelle Mathematik, Bonn (1962)
12. Physarum Chip: Growing Computers from Slime Mould. <http://www.phychip.eu>
13. Schumann, A., Pancerz, K.: Timed transition system models for programming Physarum machines: Extended abstract. In: Popova-Zeugmann, L. (ed.) *Proceedings of the Workshop on Concurrency, Specification and Programming (CS&P'2014)*, pp. 180–183. Chemnitz, Germany (2014)
14. Schumann, A., Pancerz, K.: Towards an object-oriented programming language for Physarum polycephalum computing. In: Szcuka, M., Czaja, L., Kacprzak, M. (eds.) *Proceedings of the Workshop on Concurrency, Specification and Programming (CS&P'2013)*, pp. 389–397. Warsaw, Poland (2013)
15. Schumann, A., Pancerz, K.: Towards an object-oriented programming language for Physarum polycephalum computing: a petri net model approach. *Fundam. Inf.* **133**(2–3), 271–285 (2014)
16. Verbeek, H., Wynn, M., van der Aalst, W., ter Hofstede, A.: Reduction rules for reset/inhibitor nets. *J. Comput. Syst. Sci.* **76**(2), 125–143 (2010)

# A Tool for Computer-Aided Diagnosis of Psychological Disorders Based on the MMPI Test: An Overview

Krzysztof Pancerz, Olga Mich, Andrzej Burda and Jerzy Gomuła

**Abstract** The goal of the paper is to summarize our research, conducted for over the last five years, on computer-aided diagnosis of patients screened with the MMPI (Minnesota Multiphasic Personality Inventory) test. The MMPI test delivers psychometric data, in the form of the so-called profiles (thirteen descriptive attributes corresponding to three validity and ten clinical scales), enabling us to diagnose selected psychological disorders. The notable effect of conducted research is a new computer tool, called Copernicus, aiding diagnosis of psychological disorders based on MMPI profiles. The paper is focused on outlining the functionality of the created tool.

## 1 Introduction

A psychological (mental) disorder is a pattern of behavioural or psychological symptoms that impact multiple life areas and/or create distress for the person experiencing these symptoms. The classification and diagnosis of psychological disorders is an important concern for both mental health providers and mental health clients. The MMPI (Minnesota Multiphasic Personality Inventory) test [25] is the most widely used and researched clinical assessment tool for supporting diagnosis of psychological disorders. Originally, it was developed in the late 1930s by a physiological

---

K. Pancerz (✉) · O. Mich · A. Burda  
University of Management and Administration, Zamość, Poland  
e-mail: kpancerz@wszia.edu.pl

A. Burda  
e-mail: aburda@wszia.edu.pl

K. Pancerz  
University of Information Technology and Management, Rzeszów, Poland

J. Gomuła  
Medan Foundation, The Andropause Institute, Warsaw, Poland  
e-mail: jerzy.gomula@wp.pl

J. Gomuła  
Cardinal Stefan Wyszyński University, Warsaw, Poland

psychologist S.R. Hathaway and a neuropsychiatrist J.C. McKinley at the University of Minnesota. The test was later revised and updated to improve accuracy and validity (cf. [29]). Today, it is a frequently used clinical testing tool and it is one of the most researched psychological tests in existence. In our research, we have used data coming from the MMPI-WISKAD test. The MMPI-WISKAD personality inventory is a Polish adaptation of the American inventory. The test originally was translated by Chojnowski (as WIO) [4] and elaborated by Ptużek (as WISKAD) in 1950 [34].

In years 1998–1999, a team of researchers consisting of Duch, Kucharski, Gomuła, Adamczak created two independent rule systems devised for the nosological diagnosis of persons that may be screened with the MMPI-WISKAD test [6]. In the literature, we can also find descriptions of some other computer tools for classification of MMPI profiles, e.g., based on the Fortran program [18] or “Panda” [32]. For over the last five years, our research has been focused on creating a new computer tool called the Copernicus system. The main goal of this tool is to support clinical psychologists, especially researchers, in differential and clinical diagnosis based on the overall analysis of profiles of patients examined by means of personality inventories. All results of the computer-aided diagnosis, as well as auxiliary analyses, are presented to the users using a modern graphical interface. The development of the Copernicus system (consecutive versions) was also described earlier in [10, 12], and [23]. In this paper, we summarize research on analysis and classification of MMPI profiles by means of the Copernicus system. The MMPI profiles, as well as their format in the Copernicus system, are described in Sect. 2. Main features of the created computer tool are depicted in Sect. 3. Section 4 delivers description of results of experiments with classification of MMPI profiles using selected well-known machine learning approaches.

## 2 The MMPI Data

In case of the MMPI test, each case (patient)  $x$  is described by a data vector  $a(x)$  consisting of thirteen descriptive attributes, i.e.,  $a(x) = [a_1(x), a_2(x), \dots, a_{13}(x)]$ . The meaning of attributes in such a data vector is as follows (see [29]):

- $a_1$  corresponds to the scale *L (Lie)* assessing naive attempts to place oneself in a morally and culturally favourable light by denying moral imperfections,
- $a_2$  corresponds to the scale *F (Infrequency)* assessing the tendency to claim highly unusual attitudes and behaviours as a function of severe psychopathology, a patient’s seeking to place himself or herself in an unfavourable light, or a patient’s difficulties completing the inventory,
- $a_3$  corresponds to the scale *K (Correction)* assessing the tendency to control and limit the disclosure of distress, discomfort, and problems relating to others,
- $a_4$  corresponds to the scale *1.Hp (Hypochondriasis)* measuring the tendency to manifest physical symptoms as an expression of emotional discomfort, to be preoccupied with one’s health, and to reject nonmedical explanations for such symptoms,

- $a_5$  corresponds to the scale 2.D (*Depression*) measuring aspects of symptomatic depression: dysphoria, distress, pessimism, low morale, inhibition, intropunitive-ness, physical discomfort and vegetative symptoms, and problems in thinking,
- $a_6$  corresponds to the scale 3.Hy (*Hysteria*) measuring the tendency to develop physical symptoms under stress, to experience pain, and to deny social friction or discord with others,
- $a_7$  corresponds to the scale 4.Ps (*Psychopathic Deviate*) measuring alienation, social disinhibition, and the tendency to come into conflict with family, authorities, and others through rebellion, exploitation, misconduct, poorly developed conscience, and the lack of internalized moral standards,
- $a_8$  corresponds to the scale 5.Mf (*Masculinity/Femininity*) measuring broad patterns of interest, activities, attitudes, and sentiments that tend to follow gender stereotypes,
- $a_9$  corresponds to the scale 6.Pa (*Paranoia*) measuring personal/moral rigidity, interpersonal sensitivity, resentment, and ideas of being misunderstood, mistreated, persecuted, or controlled by others, and the tendency to construe the actions, intentions and motives of others as unfair, degrading, or hostile,
- $a_{10}$  corresponds to the scale 7.Pt (*Psychasthenia*) measuring the tendency to express stresses through tension, anxiety, apprehensiveness, worry, phobias, obses-sions, rumination, compulsions, and fears of losing control, with wilful and inflex-ible efforts to control such symptoms,
- $a_{11}$  corresponds to the scale 8.Sc (*Schizophrenia*) measuring severe alienation, self-contempt, apathy, cognitive disruption, inertia, feelings of unreality, alien impulses, and motor and sensory impairment,
- $a_{12}$  corresponds to the scale 9.Ma (*Hypomania*) measuring a rapid and ener-getic personal tempo, hyperarousal, hyperactivity, stimulation-seeking, euphoria, imperviousness, undercontrol, and rebellious impulses, versus lethargy, slowness, submissiveness, vulnerability, scrupulousness, and occasionally, depression,
- $a_{13}$  corresponds to the scale 0.It (*Social introversion*) measuring introversion, shyness, social anxiety, social timidity and awkwardness, and social avoidance, versus extroversion, outgoingness, social comfort and skill, social intrepidity, and social stimulation-seeking.

MMPI is composed of over 500 questions. On the basis of these questions, scores can be calculated for three validity scales (attributes:  $a_1$ ,  $a_2$ , and  $a_3$ ) and ten clinical scales (attributes from  $a_4$  to  $a_{13}$ ).

Values of attributes are expressed by the so-called T-scores [T]. The T-scores scale, which is traditionally attributed to MMPI, represents the following parameters: offset ranging from 0 to 100 T-scores, average equal to 50 T-scores, standard deviation equal to 10 T-scores. The scores are expressed as K-corrected T-Scores. The scales 1.Hp, 4.Ps, 7.Pt, 8.Sc, and 9.Ma are corrected by adding multiples of the scale K to them.

For machine learning algorithms, we have used data in a tabular form (see an example in Table 1) which is formally called a decision system (decision table)  $S = (U, A, d)$  in the Pawlak's form (cf. [33]).  $U$  is a set of cases (patients),  $A$  is a set of descriptive attributes corresponding to scales, and  $d$  is a decision attribute

**Table 1** An example of a decision table including values of MMPI scales for patients

ID	$a_1$	$a_2$	$a_3$	$a_4$	$a_5$	$a_6$	$a_7$	$a_8$	$a_9$	$a_{10}$	$a_{11}$	$a_{12}$	$a_{13}$	<i>class</i>
1	55	65	50	52	65	57	63	56	61	61	60	51	59	<i>norm</i>
2	62	66	46	70	85	71	64	59	66	76	75	55	62	<i>neur</i>
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
$N$	38	104	45	62	91	68	59	46	111	91	98	65	63	<i>dev6</i>

determining a nosological type (see Sect. 4) to which a patient is classified or the reference class.

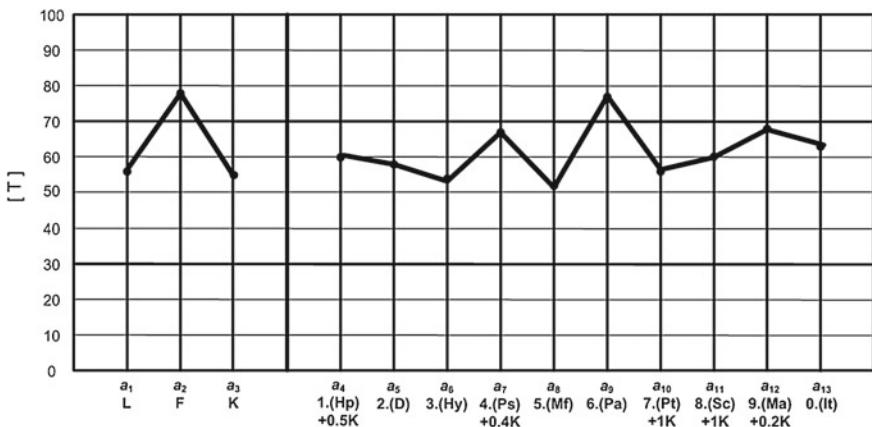
Data vectors can be represented in a graphical form as the so-called MMPI profiles. The profile always has a fixed and invariable order of its constituents (attributes, scales). Let a patient  $x$  be described by the data vector:

$$a(x) = [56, 78, 55, 60, 59, 54, 67, 52, 77, 56, 60, 68, 63].$$

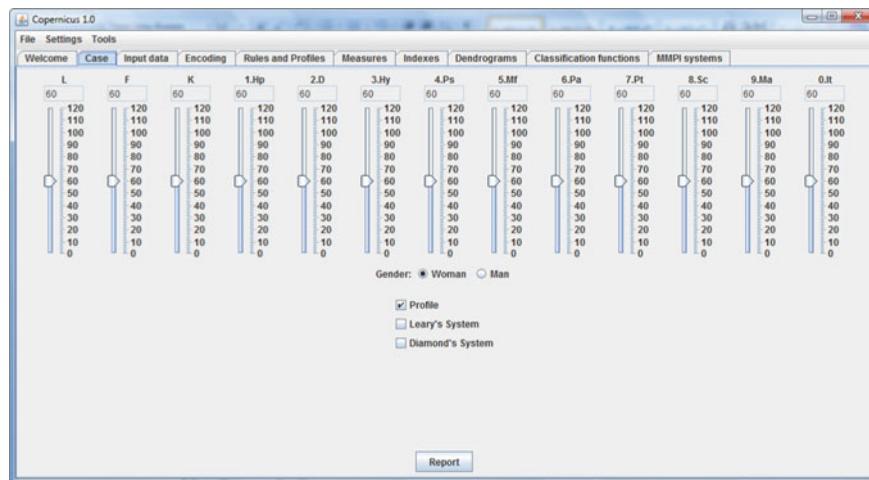
Its profile is shown in Fig. 1.

In the Copernicus system, the input data, i.e., MMPI profiles, can be provided by the user, for classification and analysis purposes, in two ways:

1. By means of the input text file. Using this way, a number of MMPI profiles can be further simultaneously analyzed and/or classified.
2. Manually, using a proper interface (see Fig. 2). Using this way, a single MMPI profile can be further analyzed and/or classified.



**Fig. 1** MMPI profile of a patient (example); suppressors  $+0.5K$ ,  $+0.4K$ ,  $+1K$ ,  $+0.2K$ —a correction value from raw results of scale  $K$  added to raw results of selected clinical scales



**Fig. 2** The Copernicus's panel for providing a single MMPI profile

### 3 A Computer Tool Aiding Diagnosis of Patients Screened with the MMPI Test

The Copernicus system supporting clinical psychologists in differential and clinical diagnosis based on the overall analysis of profiles of patients screened with the MMPI test is a tool designed for the Java platform. The main features of the tool are the following:

- multiplatforming—thanks to the Java technology, the application works on various software and hardware platforms,
- user-friendly interface—the interface is designed in order to make it possible to use in the medical environment (see some screenshots presented in the remaining part of this section),
- the module of data visualization—it allows presenting data in a clear and comprehensible way, for example, in a graphical way for a person who must make a reasonable diagnostic decision,
- modularity—the project of the application and its implementation takes into consideration modularity in order to make it possible to extend in the future and enlarge its usage on the diagnosis based on different personal inventories, not only MMPI.

We can distinguish three main parts of the Copernicus system:

- *Knowledge base.* The knowledge base embodied in the Copernicus system consists of a number of rule sets generated by different data mining and machine learning tools, such as:

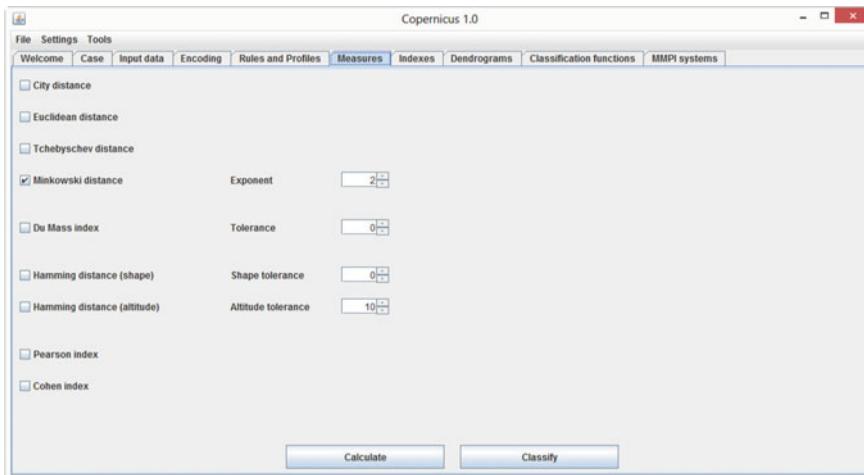
- The Rough Set Exploration System (RSES)—a software tool featuring a library of methods and a graphical user interface supporting a variety of rough set based computations [2].
- WEKA—a collection of machine learning algorithms for data mining tasks [17, 36].
- The GhostMiner System [1]—a tool providing, among others, several different types of data mining algorithms, especially FSM Neuro-Fuzzy System and SSV Decision Tree.
- The STATISTICA Data Miner [19]—a part of the STATISTICA Data Analysis and Data Mining Platform.
- NGTS—a system developed to generate decision rules using the algorithm called GTS (General-To-Specific) [20].
- TreeSEEKER—a system containing several algorithms to generate decision trees [24].
- RuleSEEKER—a tool for generation and optimization of rule sets [30].
- BeliefSEEKER—a belief network and rule induction system [16].
- A computer tool for generating fuzzy decision trees [27].
- *Multiway classification engine.* One of the main tasks of building expert systems is to search for efficient methods of classification of new cases. The classification in Copernicus is made on the basis of several methodologies. We can roughly group them into the following categories: dissimilarity measure based classifiers, index based classifiers, classification functions, rule based classifiers, and decision tree based classifiers. For each methodology, the most popular algorithms have been selected and implemented.
- *Visualization engine.* In the Copernicus system, a special attention has been paid to the visualization of analysis of MMPI data for making a diagnostic decision easier. The Copernicus system enables the user to visualize among others: classification results, clusters of patients' profiles, decision trees with tracking decision paths for examined patients, specialized diagrams (e.g. Diamond's diagram, Leary's diagram, etc.).

In the rest of this section, we describe the main classification ways available in the Copernicus system. For more details, we refer readers to [23] and some of our earlier papers listed in References.

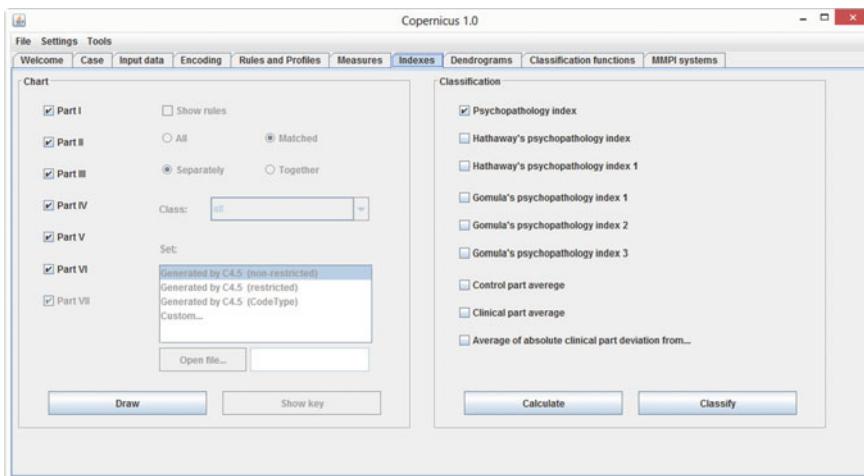
In the Copernicus system, different measures of dissimilarities between cases (patients to be classified) and class patterns (defined in the specialized literature) are used (see Fig. 3):

- standard (minimum distance metrics),
- correlation indexes (Pearson index, Cohen index),
- indexes corresponding to profile shapes and altitudes (Du Mass index, shape Hamming distance, altitude Hamming distance).

Moreover, we have implemented six psychopathology indexes: the standard psychopathology index, two Hathaway's indexes as well as three Gomuła's indexes (see Fig. 4). Ranges and weights corresponding to selected indexes are described in [10].

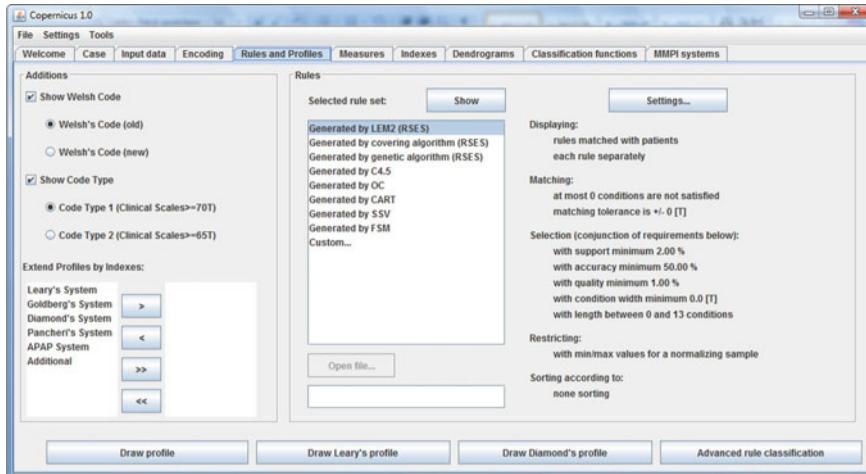


**Fig. 3** The Copernicus's panel for dissimilarity measure based classification



**Fig. 4** The Copernicus's panel for index based classification

Another approach implemented in the Copernicus system is based on classification functions. The classification function technique is one of the most classical forms of classifier design [5]. After calculating classification values for a given case, we assign to it a decision class for which the classification function has the greatest value. We have defined (using the discriminant analysis module from the STATISTICA package) as many classification functions as many decision classes there are distinguished in the training data. Coefficients (corresponding to thirteen attribute values and a free term) of classification functions (for men and women) for each decision class are presented in [10].



**Fig. 5** The Copernicus's panel for rule based classification

For the classification purpose, the users can also select a number of rule sets embodied in the Copernicus system (see Sect. 2). Moreover, the users can select their own rule sets. The rule-based analysis of MMPI data is one of the most important parts of the Copernicus system because the knowledge representation in the form of rules is the closest method to the human activity and reasoning, among others, in making the medical diagnosis. On the basis of rules, a proper diagnostic decision for the case can be made. Aggregation factors, implemented in the Copernicus system, enable the users to select only one main decision among decisions pointed by rules used for classification of the case. The rule-based analysis as well as classification of MMPI profiles using the Copernicus system are widely described in [7–9, 11, 13, 14, 21, 28]. The Copernicus's panel for rule-based classification is shown in Fig. 5.

A user can apply some other machine learning approaches for analysis and classification of MMPI profiles, among others:

- decision trees (see [22]),
- fuzzy decision trees (see [26]),
- ant-based clustering (see [31]),
- neural networks based on a multilayer perceptron (see [28]).

Results of selected experiments made by means of approaches given above are described in the next section. For more experiments, the reader is referred to our earlier papers listed in References. Especially, in [28], we summarized our searching for the optimal classifier for MMPI profiles.

## 4 Selected Experiments with Classification of MMPI Profiles

In our research, for testing classification approaches, we have used input data with classes (nosological types) assigned to patients by specialists. Our data have been categorized in two ways:

- Twenty classes, such as the reference (*norm*) class and nosological types: neurosis (*neur*), psychopathy (*psych*), organic (*org*), schizophrenia (*schiz*), delusion syndrome (*del.s*), reactive psychosis (*re.psy*), paranoia (*paran*), (sub)manic state (*man.st*), criminality (*crim*), alcoholism (*alcoh*), drug addiction (*drug*), simulation (*simu*), dissimulation (*dissimu*), and six deviational answering styles (*dev1*, *dev2*, *dev3*, *dev4*, *dev5*, *dev6*).
- Seven more general classes, called macro-classes, such as the reference (*norm*) class, as well as not-validated (*nval*), dependence (*depend*), organic (*org*), neuroticism (*neurot*), psychotism (*psychot*), and sociopathy (*socio*).

In Tables 2 and 3, we have collected results of experiments with classification of MMPI profiles using three well-known machine learning approaches:

**Table 2** Detailed accuracy by class for experiments with twenty classes

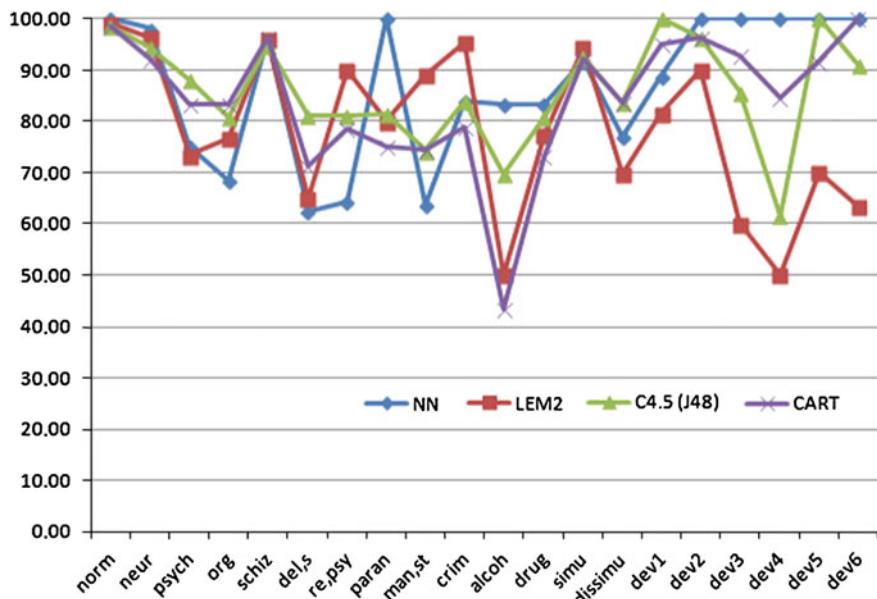
Class	NN	LEM2	J48 (C4.5)	SimpleCART
<i>norm</i>	100.00	99.13	98.50	98.50
<i>neur</i>	98.08	96.33	94.50	92.20
<i>psych</i>	75.00	73.33	88.10	83.30
<i>org</i>	68.42	76.67	80.80	83.30
<i>schiz</i>	96.19	96.12	94.80	96.30
<i>del.s</i>	62.50	65.00	81.00	71.40
<i>re.psy</i>	64.29	90.00	81.10	78.40
<i>paran</i>	100.00	80.00	81.30	75.00
<i>man.st</i>	63.64	89.00	74.30	74.30
<i>crim</i>	84.00	95.42	84.00	79.00
<i>alcoh</i>	83.33	50.00	69.60	43.50
<i>drug</i>	83.33	77.33	80.80	73.10
<i>simu</i>	91.84	94.49	92.60	92.10
<i>dissimu</i>	76.92	69.67	83.70	83.70
<i>dev1</i>	88.89	81.67	100.00	95.10
<i>dev2</i>	100.00	90.00	96.20	96.20
<i>dev3</i>	100.00	60.00	85.70	92.90
<i>dev4</i>	100.00	50.00	61.50	84.60
<i>dev5</i>	100.00	70.00	100.00	91.70
<i>dev6</i>	100.00	63.33	90.90	100.00

**Table 3** Detailed accuracy by class for experiments with seven macro-classes

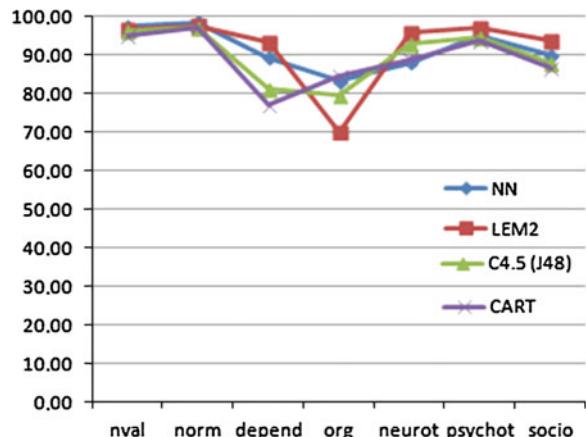
Class	NN	LEM2	J48 (C4.5)	SimpleCART
<i>nval</i>	97.50	96.72	96.20	95.10
<i>norm</i>	98.46	97.69	97.00	97.40
<i>depend</i>	89.47	93.33	81.30	77.30
<i>org</i>	83.33	70.17	79.50	84.60
<i>neurot</i>	88.14	96.00	93.10	89.00
<i>psychot</i>	95.00	97.03	94.60	93.70
<i>socio</i>	90.00	93.88	88.00	86.80

- Neural networks. We have used neural networks based on a multilayer perceptron implemented in the STATISTICA Data Miner [19].
- Rough sets. We have used the LEM 2 algorithm [15] implemented in the Rough Set Exploration System (RSES) [2].
- Decision trees. We have used two algorithms implemented in the WEKA system [17, 36]: J48 (an open source implementation of the C4.5 algorithm [35]) and CART [3].

We have been interested in classification accuracy by class. Figures 6 and 7 show a proper comparison of classification accuracy for experiments with twenty classes and seven macro-classes, respectively. It is easy to see that, in both cases, the presented

**Fig. 6** Comparison of the detailed accuracy by class for experiments with twenty classes

**Fig. 7** Comparison of the detailed accuracy by class for experiments with seven macro-classes



approaches deal differently with classification of cases belonging to individual classes (macro-classes). There is no best approach proving itself to each class. In general, it is known that there have not been designed universal machine learning and data mining methods which could be applied to each kind of data, giving the expected results. Therefore, there is a need to build a hybrid classifier combining a wide range of approaches.

## 5 Conclusions

In the paper, we have summarized information about the Copernicus system a tool for computer-aided diagnosis of psychological disorders. For over the last five years, we have tested various approaches for analysis and classification of MMPI profiles and the majority of them have been implemented in the tool. Currently, the Copernicus system delivers a variety of instruments enabling the users to examine MMPI data from different perspectives. Searching for other instruments will be continued. Especially, we are interested in a hybrid classifier combining a wide range of approaches.

## References

1. The GhostMiner System. [http://www.fqs.pl/business\\_intelligence/products/ghostminer](http://www.fqs.pl/business_intelligence/products/ghostminer)
2. Bazan, J.G., Szczyka, M.S.: The rough set exploration system. In: Peters, J., Skowron, A. (eds.) Transactions on Rough Sets III. LNAI, vol. 3400, pp. 37–56. Springer, Berlin (2005)
3. Breiman, L., Friedman, J., Olshen, R., Stone, C.: Classification and Regression Trees. Chapman & Hall, Boca Raton (1993)
4. Chojnowski, M.: Multiphasic Personality Inventory (in Polish). Polish Academy of Sciences, Psychometry Laboratory, Warsaw (1964)

5. Cios, K., Pedrycz, W., Swiniarski, R., Kurgan, L.: Data Mining. A Knowledge Discovery Approach. Springer, New York (2007)
6. Duch, W., Kucharski, T., Gomuła, J., Adamczak, R.: Machine Learning Methods in Analysis of Psychometric Data. Application to Multiphasic Personality Inventory MMPI-WISKAD (in Polish). Toruń (1999)
7. Gomuła, J., Paja, W., Pancerz, K., Mroczek, T., Wrzesień, M.: Experiments with hybridization and optimization of the rules knowledge base for classification of MMPI profiles. In: Perner, P. (ed.) Advances on Data Mining: Applications and Theoretical Aspects. LNAI, vol. 6870, pp. 121–133. Springer, Berlin (2011)
8. Gomuła, J., Paja, W., Pancerz, K., Szkola: A preliminary attempt to rules generation for mental disorders. In: Proceedings of the International Conference on Human System Interaction (HSI 2010). Rzeszów, Poland (2010)
9. Gomuła, J., Paja, W., Pancerz, K., Szkola: Rule-based analysis of MMPI data using the Copernicus system. In: Hippe, Z., Kulikowski, J., Mroczek, T. (eds.) Human-Computer Systems Interaction. Backgrounds and Applications 2. Part II. Advances in Intelligent and Soft Computing, vol. 99, pp. 191–203. Springer, Berlin (2012)
10. Gomuła, J., Pancerz, K., Szkola, J.: Analysis of MMPI profiles of patients with mental disorders—the first unveil of a new computer tool. In: Grzech, A., Świątek, P., Brzostowski, K. (eds.) Applications of Systems Science, pp. 297–306. Academic Publishing House EXIT, Warsaw, Poland (2010)
11. Gomuła, J., Pancerz, K., Szkola, J.: Classification of MMPI profiles of patients with mental disorders—experiments with attribute reduction and extension. In: Yu, J., et al. (eds.) Rough Set and Knowledge Technology. LNAI, vol. 6401, pp. 411–418. Springer, Berlin (2010)
12. Gomuła, J., Pancerz, K., Szkola, J.: Computer-aided diagnosis of patients with mental disorders using the Copernicus system. In: Proceedings of the International Conference on Human System Interaction (HSI 2011). Yokohama, Japan (2011)
13. Gomuła, J., Pancerz, K., Szkola, J.: Copernicus—an expert system supporting differential diagnosis of patients examined using the MMPI test: an index-rule approach. In: Traver, V., Fred, A., Filipe, J., Gamboa, H. (eds.) Proceedings of the International Conference on Health Informatics (HEALTHINF 2011), pp. 323–328. Italy, Rome (2011)
14. Gomuła, J., Pancerz, K., Szkola, J.: Rule-based classification of MMPI data of patients with mental disorders: Experiments with basic and extended profiles. Int. J. Comput. Intell. Syst. **4**(5), 1022–1031 (2011)
15. Grzymala-Busse, J.: A new version of the rule induction system LERS. Fundam. Inf. **31**, 27–39 (1997)
16. Grzymala-Busse, J., Hippe, Z., Mroczek, T.: Deriving belief networks and belief rules from data: a progress report. In: Peters, J., Skowron, A. (eds.) Transactions on Rough Sets VII. Lecture Notes in Computer Science, vol. 4400, pp. 53–69. Springer, Berlin (2007)
17. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The WEKA data mining software: an update. SIGKDD Explor. **11**, 10–18 (2009)
18. Hatcher, W.E.: Automated classification of MMPI profiles into psychotic, neurotic or personality disorder types. Comput Program. Biomed. **8**(1), 77–80 (1978)
19. Hill, T., Lewicki, P.: Statistics Methods and Applications. StatSoft, Tulsa (2007)
20. Hippe, Z.: Machine learning: a promising strategy for business information processing? In: Abramowicz, W. (ed.) Business Information Systems, pp. 603–622. Academy of Economics Editorial Office, Poznań (1997)
21. Jachyra, D., Gomuła, J., Pancerz, K.: Rule-based classification of patients screened with the MMPI test in the Copernicus system. In: Iantovics, B., Kountchev, R. (eds.) Advanced Intelligent Computational Technologies and Decision Support Systems. Studies in Computational Intelligence, vol. 486, pp. 31–42. Springer International Publishing (2014)
22. Jachyra, D., Pancerz, K., Gomuła, J.: Classification of MMPI profiles using decision trees. In: Szczuka, M., Czaja, L., Skowron, A., Kacprzak, M. (eds.) Proceedings of the Workshop on Concurrency, Specification and Programming (CS&P 2011), pp. 397–407. Pułtusk, Poland (2011)

23. Jachyra, D., Pancerz, K., Gomuła, J.: Multiway classification of MMPI profiles. In: Zaitseva, E., Levashenko, V. (eds.) *Proceedings of the Ninth International Conference on Digital Technologies (DT 2013)*, pp. 119–127. Zilina, Slovakia (2013)
24. Knap, M.: Research on new algorithms for decision trees generation (in Polish). Ph.D. thesis, AGH University of Science and Technology, Krakow (2009)
25. Lachar, D.: *The MMPI: Clinical Assessment and Automated Interpretations*. Western Psychological Services, Fate Angeles (1974)
26. Levashenko, V., Zaitseva, E., Pancerz, K., Gomuła, J.: Fuzzy decision tree based classification of psychometric data. In: Ganzha, M., Maciaszek, L., Paprzycki, M. (eds.) *Position Papers of the 2014 Federated Conference on Computer Science and Information Systems. Annals of Computer Science and Information Systems*, vol. 3, pp. 37–41. PTI, Warsaw (2014)
27. Levashenko, V., Zaitseva, E., Puuronen, S.: Fuzzy classifier based on fuzzy decision tree. In: *Proceedings of the International Conference on Computer as a Tool (EUROCON 2007)*, pp. 823–827. IEEE (2007)
28. Mich, O., Burda, A., Pancerz, K., Gomuła, J.: The knowledge base for computer-aided diagnosis of mental disorders based on psychometric tests. In: *Proceedings of the 10th International Conference on Digital Technologies (DT'2014)*, pp. 266–272. Zilina, Slovakia (2014)
29. Nichols, D.: *Essentials of MMPI-2 Assessment*. Wiley, New York (2001)
30. Paja, W., Hippe, Z.: Feasibility studies of quality of knowledge mined from multiple secondary sources. I. Implementation of generic operations. In: Kłopotek, M., Wierzchon, S., Trojanowski, K. (eds.) *Intelligent Information Processing and Web Mining. Advances in Intelligent and Soft Computing*, vol. 31, pp. 461–465. Springer, Berlin (2005)
31. Pancerz, K., Lewicki, A., Tadeusiewicz, R., Gomuła, J.: Ant based clustering of MMPI data—an experimental study. In: Yao, J., Ramanna, S., Wang, G., Suraj, Z. (eds.) *Rough Sets and Knowledge Technology. Lecture Notes in Computer Science*, vol. 6954, pp. 366–375. Springer, Berlin (2011)
32. Pancheri, P., De Fidio, D.: Dal minnesota multiphasic personality inventory al Panda: il MMPI-2 automatico. In: Technical Report
33. Pawlak, Z.: *Rough Sets. Theoretical Aspects of Reasoning About Data*. Kluwer Academic Publishers, Dordrecht (1991)
34. Plużek, Z.: Value of the WISKAD-MMPI test for nosological differential diagnosis (in Polish). The Catholic University of Lublin (1971)
35. Quinlan, J.: *C4.5: Programs for Machine Learning*. Morgan Kaufmann, San Francisco (1992)
36. Witten, I.H., Frank, E.: *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, San Francisco (2005)

# The Instance and Feature Selection for Neural Network Based Diagnosis of Chronic Obstructive Bronchitis

Sergey Subbotin

**Abstract** The problem of data dimensionality reduction is considered in the paper. The method of feature and sample selection for diagnosis model building by the precedents is proposed. It is based on feature space partitioning and evaluation of significances of terms and features. Than the measures of instance informativity are proposed to form a subsample. The example of proposed method application in the problem of biomedical data processing for chronic obstructive bronchitis diagnosis is provided.

## 1 Introduction

To automate the biomedical diagnosis problem solving we need to have a model of decision dependence from the descriptive features, characterizing the recognized instance (an observation of patient condition at a certain time). As a rule, due to the lack or inadequacy of expert knowledge in practice such model is constructed on the basis of observations or precedents (instances).

As an example of this problem we can consider the task of automation of diagnosis of chronic obstructive bronchitis on the data of clinical laboratory observations. For each patient state we have an observation characterized by set of clinical laboratory measured indicators (input features) and the expert diagnosis as an output feature. We need to build a model of dependence of diagnosis from input features. This will make possible to speed-up and to simplify the process of diagnosis of chronic obstructive bronchitis and also to extract knowledge from the data.

The one of the most popular and powerful tools for model building by precedents are artificial neural networks [1] that can learn by precedents, providing their generalization and extracting knowledge from the data. The feed-forward neural networks should be especially distinguished among the neural network paradigms because they

---

S. Subbotin (✉)

Zaporizhzhya National Technical University, Zhukovsky Street 64, Zaporizhzhya, Ukraine  
e-mail: subbotin@zntu.edu.ua

are simpler implemented on serial computers than other neural network models, and also are more interpretable and have developed methods of training by precedents.

The process of neural model building is typically time-consuming and highly iterative. This is caused by that the training time and accuracy of the neural network model are essentially dependent on the dimensionality and quality of the used training sample. Therefore, to improve the construction speed and quality of neural model it is necessary to reduce the dimension of the sample, providing the preservation of its basic properties.

On the one hand, the data dimensionality is determined by the number of instances. Since the available sample may contain the redundant examples it is necessary to select a subset of the most significant instances in order to reduce the data dimensionality.

On the other hand, the available data samples can contain a large number of redundant features. The using of sample with a large number of features can lead to the construction of redundant neural model with low generalizing properties, as well as to complicate the subsequent analysis and interpretation of the results and the decision-making process of neural network model. Therefore, to increase the generalizing properties and to reduce the structural and parametric complexity of the neural model it is appropriate to reduce the number of features by selecting the most informative subset of features from the original feature set.

The known sampling methods [2–23] and feature selection methods [24–28] are highly iterative and low speed, as well as characterized by the uncertainty of quality criteria of formed subsample.

The purpose of the work is to increase the speed and quality of the formation process and to reduce the dimensionality of training samples for neural network model building by the precedents on the example of solving the task of chronic obstructive bronchitis diagnosis.

## 2 Formal Problem Statement

Suppose we have an original data sample as a set of precedents (instances, observations, exemplars)  $\langle x, y \rangle$ , where  $x = \{x^s\}$ ,  $x = \{x_j\}$ ,  $x^s$  is  $s$ th instance of a sample,  $x^s = \{x_j^s\}$ ,  $x_j^s$  is a value of  $j$ th input feature  $x_j$ , characterizing the instance  $x^s$ ,  $x_j = \{x_j^s\}$ ,  $y = \{y^s\}$ ,  $y^s$  is an output feature value associated with the instance  $x^s$ ,  $s = 1, 2, \dots, S$ ,  $S$  is a number of instances in the original sample,  $j = 1, 2, \dots, N$ ,  $N$  is a number of features characterizing original sample.

For a given sample of precedents  $\langle x, y \rangle$  the problem of neural model synthesis can be presented as the problem of finding  $\langle F(), w \rangle: y^{s*} = F(w, x^s)$ ,  $f(F(), w, \langle x, y \rangle) \rightarrow \text{opt}$ , where the neural network model structure  $F()$  usually specified by the user in practice, and the set of controlled parameters  $w$  of the neural network model is adjusted on the base of the training sample,  $y^{s*}$  is a calculated output feature value for the  $s$ th instance on the neural model output,  $f()$  is a user criterion

characterizing the argument quality relatively to the problem being solved, opt is an optimal (desired or acceptable) value of the functional  $f()$  for the problem being solved.

The problem of subsample formation from a given sample  $\langle x, y \rangle$  is to find such a set of  $\langle x', y' \rangle : x' \subset \{x^s\}, y' = \{y^s | x^s \in x'\}, S' < S, N' = N$ , wherein  $f(\langle x', y' \rangle, \langle x, y \rangle) \rightarrow \text{opt}$ , where  $N'$  is a number of features in a subsample,  $S'$  is a number of instances in a subsample.

The problem of feature selection for a given sample  $\langle x, y \rangle$  is to find such a set of  $\langle x', y \rangle : x' \subset \{x_j\}, S' = S, N' < N$ , wherein  $f(\langle x', y \rangle, \langle x, y \rangle) \rightarrow \text{opt}$ , where  $N'$  is a number of selected features.

### 3 Review of the Literature

The sampling methods for decision-making model building by precedents in [2, 3] are divided into prototype selection methods and prototype construction methods. Here, the prototype means selected subsample relative to the original sample.

The prototype selection methods [4–15] do not modify, but only select the most important instances from the original sample. Depending on the strategy of solution forming these methods are divided into incremental methods [4, 5] (they successively add instances from the original sample to the subsample) and decremental methods [4–8] (they successively remove instances from the original sample, and obtain a subsample as a result). There are also separated such methods as noise filtering methods [6, 8–11] (they remove instances, which class labels do not equal with most of the neighbor labels), condensation methods [4–7, 12, 13] (this methods add instances from the original sample to the formed subsample, if they bring a new information, but do not add if they have the same class labels as their neighbors), and methods based on stochastic search [12, 14, 15] (they randomly form a subsample from the original sample, considering a set of variants of decisions and selecting the best of them). The common disadvantages of these methods are the high iterativity and big search time, as well as uncertainty in quality criteria selection of formed subsample.

The methods of prototype construction [12, 15–23] based on the original sample build artificial instances allow to describe the original sample. Among these methods it is possible to separate the cluster analysis based methods [18, 19, 23] (they replace the original sample by the centers of its clusters), the data squashing methods [17] (they replace the original sample instances by the artificial prototypes having weights obtained on their basis) and the neural network methods (neural network based methods) [16, 20–22] (they train a neural network on the original sample, which is then used for cluster centers extraction as instances of formed subsample). The common disadvantages of these methods are their high iterativity and a big operating time, and the uncertainty in the initial parameter setting. The methods based on a cluster analysis are characterized by disadvantages such as the uncertainty of cluster number, initial parameters, and metric selection for the clustering and training methods.

The data squashing methods form prototypes, which are difficult to interpret. The neural network methods have such disadvantages as the difficulty of prototype extraction from the neural network model, the no guarantee of receiving of acceptable neural network model a result of training, the neural network model variability, entailing nonstationarity of constructed prototypes, the orientation on a specific model, the uncertainty in setting the initial parameters of the model and training methods.

Additionally the combined methods are distinguished [3]. They combine the selection and formation of prototypes. The combined methods have the same disadvantages as methods of prototype selection and methods of prototype construction.

Since the prototype construction methods and the hybrid methods related with them are slower than the prototype selection methods, it is advisable to choose the latter as the basis for sampling problem solving.

In order to eliminate the disadvantages of these methods, it is advisable to form a sample without iterative busting of instance combinations by a certain percentage of instance selection from the original sample. This will significantly reduce the time. Herewith we also need to define the indicators to evaluate the individual instance informativity with regard to their position relatively to the interclass boundaries and to the centers of pseudo-clusters, which. This makes possible to generate a non-random sample, to estimate and guarantee the high quality of selected subsamples.

The feature selection methods [24–28] are typically separated on the filter and wrapper. The filter methods [24, 25, 27] do not depend on the model and include specific criteria that are measured on a sample, which allows to characterize the informativity of the selected features. Their disadvantages are the uncertainty of the criteria used for relation with the process of constructing the model, as well as high busting of feature combinations. The wrapper methods [26, 28] are extremely time-consuming because they are assumed with selection of feature combinations and construction the model for each combination, but they provide the highest accuracy for feature selection.

Therefore, the filter approach can be used to ensure a high-speed feature selection. At the same time to speed up the process of data dimensionality reduction it is encouraged to develop and use for the feature selection the indicators of feature informativity, which will be determined in conjunction with the indicators of informativity of instances.

## 4 The Method for Data Dimensionality Reduction

Let's break given feature space into rectangular regions limiting the range of values of each feature by its minimum and maximum values. Then the partition projections into feature axis allow to allocate feature intervals for each rectangular block. The intervals can be formed as cluster projections or as a regular grid, or on the basis of class boundaries in sample one-dimensional projections on the feature axes [29, 30].

Then each such interval can be considered as a term and it is possible to evaluate its importance for decision-making on instance belonging to the cluster with the weight of the  $k$ th term of  $j$ th feature of  $s$ th instance  $x^s$  based on a description of the corresponding interval center by the formula (1):

$$w_{C^s jk} = \exp(-(0.5(r_{jk} - l_{jk}) - x_j^s)^2), \quad (1)$$

as well as the weight of the  $k$ th term of  $j$ th feature of the  $s$ th instance  $x^s$  relatively to the description of the intercluster boundaries determined by the formula (2):

$$w_{B^s jk} = \exp(-\min((r_{jk} - x_j^s), (x_j^s - l_{jk}))^2). \quad (2)$$

Then the overall significance of the  $k$ th term of  $j$ th feature of  $s$ th instance  $x^s$  relatively to the description of the intercluster boundaries can be estimated using the weight determined as  $w_{s jk} = \max\{w_{C^s jk}, w_{B^s jk}\}$ .

Defining for each  $s$ th instance the term significances, we can also determine the term weights for the whole sample:

$$w_{jk} = \frac{S_{jk}}{SK_{jk}},$$

where  $S_{jk}$  is a number of instances in the  $k$ th term of the  $j$ th feature,  $K_{jk}$  is a number of classes, which instances hit the  $k$ th interval of  $j$ th feature values.

Knowing the term significance we can define the feature informativity evaluations by formula (3):

$$w_j = \max_k \{w_{jk}\}, \quad (3)$$

or by the formula (4):

$$w_j = \frac{1}{k_j} \sum_{k=1}^{k_j} w_{jk}. \quad (4)$$

It is also possible to use the individual evaluation of the feature informativity in the range [0, 1] defined by the indicators [29].

Based on evaluations of term and feature significance we can determine informativity evaluations for each  $s$ th sample instance by the formula (5):

$$I_1(x^s) = \frac{\sum_{j=1}^N \left( w_j \sum_{k=1}^{k_j} w_{jk} w_{jk}^s \right)}{\sum_{j=1}^N \left( w_j \sum_{k=1}^{k_j} w_{jk} \max_p \{w_{jk}^p\} \right)}, \quad (5)$$

or by the formula (6):

$$I_2(x^s) = \frac{1}{N} \sum_{j=1}^N \left( \frac{w_j}{\max_{i=1,2,\dots,N} \{w_i\} k_j} \sum_{k=1}^{k_j} \frac{w_{jk} w_{jk}^s}{\left( \max_{q=1,2,\dots,k_j} \{w_{jk}\} \right) \left( \max_p \{w_{jk}^p\} \right)} \right). \quad (6)$$

Suggested indicators (5) and (6) provide evaluation of individual informativity of instance  $x^s$  relatively to the initial sample in the range [0, 1]. The greater the value of corresponding indicator, the more valuable is an instance, and vices versa.

If necessary, the estimates (5) and (6) can be further normalized so that they will give not an absolute but relative value of instance significance in the sample (7):

$$I(x^s) = \frac{I(x^s) - \min_p \{I(x^p)\}}{\max_p \{I(x^p)\} - \min_p \{I(x^p)\}}. \quad (7)$$

In this case, the instance with the maximum individual informativity will receive evaluation equal to one, and the instance with minimal informativity will receive evaluation equal to zero. The application of (7) can be useful when it need to simplify the choice of the threshold for separating the sample by the corresponding informativity indicator.

The proposed indicators of evaluation of individual instance significance can be used in the subsample formation from the given original sample by one of the following methods:

- (1) to form a training subsample of those instances of the original sample, the normalized values of which individual informativity evaluations (7) are greater than some specified threshold;
- (2) to form a training subsample from the not more than  $S'/K$  instances of each class of the original sample with the greatest values of individual informativity evaluations;
- (3) to use a stochastic search based on evolutionary or multi-agent methods, selecting the best in a some sense combination of instances, using information about individual informativity of instances in the search operators to accelerate the search and focusing it on the most promising solutions.

The first method does not obviously determine the number of instances that will fall into the formed sample. The third method is iterative and requires the specification and use of quality indicators, the calculation of which can also be time consuming. Therefore, the second method is the most simple applicable in practice and relatively simple from a computational point of view. It is appropriate to examine together with the proposed measures.

## 5 The Feed-Forward Neural Network Architecture and Training

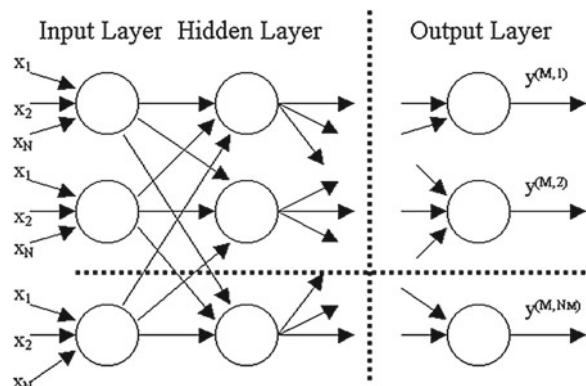
The diagnostic model  $\langle F(), w \rangle$  based on feed-forward neural network [1] can be formally presented as  $\langle M, \{N_\eta\}, \{y^{(\eta,i)}(x^{(\eta,i)})\} \rangle$ , where  $M$  is a number of layers in the neural network,  $N_\eta$  is a number of neurons in the  $\eta$ th layer, and can be functionally described as  $y_i^s = y^{(M,i)}(y^{(M-1,i)}(\dots y^{(1,i)}(x^s))), i = 1, 2, \dots, N_{\eta-1}$ ,  $w^{(\eta,i)} = \{w_j^{(\eta,i)}\}$ ,  $y^{(0,j)} = \psi^{(0,j)} = x_j^s$ ,  $N_0 = N$ ,  $x_j^{(1,i)} = x_j^s, i = 1, 2, \dots, N_{\eta-1}$ ,  $\eta = 1, 2, \dots, M, j = 1, 2, \dots, N$ .  $y^{(0,j)} = \psi^{(0,j)} = x_j^s$ ,  $N_0 = N$ ,  $x_j^{(1,i)} = x_j^s, i = 1, 2, \dots, N_{\eta-1}, \eta = 1, 2, \dots, M, j = 1, 2, \dots, N$ , where  $\{y^{(\eta,i)} = \psi^{(\eta,i)}(\phi^{(\eta,i)})\}$  and  $w = \{w^{(\eta,i)}\} = \{w_j^{(\eta,i)}\}$  are, respectively, structural blocks and parameters of diagnostic model on the base of neural network,  $\phi^{(\eta,i)}$ ,  $\psi^{(\eta,i)}$  are, respectively, discriminant (weight) and activation (transfer) functions of  $i$ th neuron of  $\eta$ th layer,  $y^{s*} = \psi^{(M,i)}(x^s)$  is computed value at  $i$ th output of neural network for instance  $x^s$  submitted to it's inputs,  $w_j^{(\eta,i)}$  is a value of  $j$ th adjusted parameter (weight) of  $i$ -th input of  $i$ th neuron of  $\eta$ th layer,  $y^s = \{y_i^s\}, i = 1, 2, \dots, N_M$ ,  $N_M$  is a number of output features,  $y_i^s$  is a value of  $i$ th output feature for instance  $x^s$ , for problems of classification  $y^s \in \{q\}, q = 1, 2, \dots, K$ ,  $K$  is a number of classes,  $K > 1$ .

The Fig. 1 shows a structure of multilayer feed-forward neural network. There are various functions may be used for weight and activation functions defining. There are most widely used functions: weighted sum

$$\psi^{(\eta,i)}(w^{(\eta,i)}, x^{(\eta,i)}) = w_0^{(\eta,i)} + \sum_{j=1}^{N_{\eta-1}} w^{(\eta,i)} x^{(\eta,i)}$$

as a weight function for  $i$ th neuron of  $\eta$ th layer, and logistic sigmoid

**Fig. 1** The structure of multilayer feed-forward neural network



$$\psi(a) = 1/(1 + \exp(-a))$$

as an activation function.

The process of feed-forward neural network training based on given training set  $\langle x, y \rangle$  aims to optimize the training function  $f$ . In the common case the training criterion  $f$  is defined as model error:  $f = E \rightarrow \min$ , where  $E$  is mean squared error:

$$E = \frac{1}{2} \sum_{s=1}^S (y^s - \psi^{(M,i)}(x^s))^2.$$

The training of feed-forward networks is usually performed using Levenberg-Marquardt gradient method with error back-propagation technique [1].

## 6 Experiments and Results

The computer program implementing the proposed method, which complements the “Automated system neural network and neuro-fuzzy model synthesis for non-destructive diagnosis and pattern classification on features” (Certificate of copyright registration no 35431 (Ukraine) from 21.10.2010) was developed to conduct experiments.

The developed software was studied in solving the task of automation of diagnosis of chronic obstructive bronchitis on the data of clinical laboratory observations.

The chronic obstructive bronchitis is a disease of the human respiratory system with progressive airways obstruction. The most important feature of chronic obstructive bronchitis is a steady progression of the disease.

It is of considerable scientific interest to obtain the model for the diagnosis of chronic obstructive bronchitis based on neural networks, as well as to identify the components of the peripheral blood, reflecting the permanence of inflammation in the bronchial and lung tissue.

The experimentally obtained data sample [31] in the study of functional and morphological properties of leukocytes and thrombocytes, vascular factor, the complement system and circulating immune complexes (CIC) in peripheral blood of 205 patients was used to construct diagnostic models. The state of patients was characterized by 28 diagnostic features:  $x_1$ —gender,  $x_2$ —age,  $x_3$ —white blood cell count, including,  $x_4$ —stabnuclear,  $x_5$ —segmentonuclear,  $x_6$ —eosinophils,  $x_7$ —monocytes,  $x_8$ —lymphocytes,  $x_9$ —phagocytic number of neutrophils,  $x_{10}$ —phagocytic number of monocytes,  $x_{11}$ —phagocytic index of neutrophils,  $x_{12}$ —phagocytic index of monocytes,  $x_{13}$ —index of completeness of phagocytosis of neutrophils,  $x_{14}$ —index of completeness of phagocytosis of monocytes,  $x_{15}$ —NBT-test (reduction reaction of nitroblue tetrazolium), which reflects the ability of neutrophils to generate reactive oxygen species,  $x_{16}$ —percentage of NBT positive neutrophils,  $x_{17}$ —NBT stimulated—reflects the reserve possibilities of neutrophils to generate

**Table 1** Fragment of the original data sample

Feature	Instance number											
	1	2	3	4	5	6	7	8	9	10	11	12
$x_1$	1	1	1	2	2	1	2	2	1	2	1	1
$x_2$	43	34	36	63	41	38	62	19	59	42	44	60
$x_3$	4.5	4.2	5.8	5.4	4.7	4.6	4.5	10	4.9	6.1	9	5.6
$x_4$	3	2	5	1	4	4	6	7	5	6	5	2
$x_5$	58	58	45	49	57	40	48	3	6	4	70	45
$x_6$	1	1	0	0	1	9	2	37	60	54	0	4
$x_7$	2	1	4	9	0	4	0	2	8	2	9	8
$x_8$	36	38	46	41	38	43	44	0	0	2	18	41
$x_9$	48	28	16	14	40	48	16	58	26	38	12	14
$x_{10}$	16	20	10	10	0	38	24	28	60	40	20	16
$x_{11}$	58	106	80	58	134	208	88	199	70	100	43	72
$x_{12}$	48	60	30	60	210	400	226	156	132	85	72	88
$x_{13}$	142	168	236	176	354	236	320	92	92	110	294	138
$x_{14}$	312	80	36	48	24	101	120	60	124	140	60	36
$x_{15}$	91	12	40	86	44	2	43	34	56	140	68	40
$x_{16}$	39	6	1	31	17	2	26	0.28	0.06	0.94	26	27
$x_{17}$	140	19	11	110	55	21	78	130	6	47	71	86
$x_{18}$	47	6	3	36	21	18	34	0.41	0.09	1.4	27	31
$x_{19}$	195	175	165	320	130	140	260	216	290	244	300	260
$x_{20}$	73	82	70	89	40	91	88	88	87	78	75	55
$x_{21}$	27	18	30	11	60	9	12	69	80	90	25	45
$x_{22}$	50	33	79	16	14	10	16	31	20	10	18	23
$x_{23}$	8	7	9	180	90	280	120	90	110	280	100	90
$x_{24}$	18	14	20	230	130	290	130	40	340	80	150	190
$x_{25}$	2	2	1	3	2	1	1	1.1	3.5	2.4	2	2
$x_{26}$	327	283	124	184	220	94	146	71	176	191	184	124
$x_{27}$	136	197	95	124	189	159	42	292	193	238	124	196
$x_{28}$	124	71	87	98	164.45	87.65	197.16	184	176	185	144.71	164
$y$	0	0	0	1	1	0	1	1	1	1	0	

reactive oxygen species in terms of their additional incentive,  $x_{18}$ —the percentage of NBTpositive neutrophils reacting to stimulation,  $x_{19}$ —platelet number,  $x_{20}$ —number of mature platelets,  $x_{21}$ —number of active platelet,  $x_{22}$ —the level of complement (CH50),  $x_{23}$ —level CIC deposited 3 % polyethylene glycol (PEG),  $x_{24}$ —level CIC deposited 4 % PEG,  $x_{25}$ —the size of the CIC,  $x_{26}$ —myeloperoxidase activity of neutrophils,  $x_{27}$ —activity of cationic proteins of neutrophils,  $x_{28}$ —vascular level of von Willebrand factor. Each instance of the sample was associated with value of the target (output) parameter—the diagnosis of the patient in (1—sick, 0—not sick). The fragment of used original sample is presented in the Table 1.

On the basis of the original sample the instance informativity evaluations were obtained and subsets of instances as a training samples were selected by the second method. To study the second method the 25, 50, 75 and 100 % (for the control) instances with the greatest values of individual significance in each class was selected from the original sample and included to the training set, respectively.

Further, for each sample a model based on a two-layer feed-forward neural network was built. It was trained using the Levenberg-Marquardt method [1]. The number of network inputs was determined by the number of features in the corresponding problem  $N$ . The second layer of the network contains one neuron. The number of neurons of the hidden (first) layer was three. All neurons of a network were used the weighted sum as weight (postsynaptic) function and logistic sigmoid as transfer function. The training method parameters were set as follows: the learning rate is 0.01, the allowable number of iterations (epochs) of the method is 1000, the target value of the error function is  $10^{-6}$ .

After neural model training process completion its final characteristics were fixed: the training time  $t_{tr}$ , and the number of spent training iterations  $ep_{tr}$ . After training each model was tested separately on the training and the whole original samples, for each of which the error was determined, respectively,  $E_{tr}$  and  $E_{all}$ . Here each error is the per cent of instances of corresponding sample for which the estimated value did not match the actual value of the output feature.

After the sample selection experiments the computed feature weights  $\{w_j\}$  were used as individual feature informativities. For each selected sample we select a feature subset of 25, 50, and 75 % of features with highest individual informativities. For each such subsample the neural model were built in the same way as in previous case, but with less number of features.

The fragment of the results of conducted experiments is presented in the Table 2.

The Table 2 shows that the use of the proposed method of instance significance determining allows in practice to select a subsample of smaller volume from of the original sample, enough to construct neural network models with the required accuracy, reducing the time to build models.

As it evident from the Table 2, with the increasing of examples number  $S$  in the formed sample the accuracy is increased (errors for formed training and for the original samples are reduced), the training time and the number of training iterations are increased, and vice versa. At the same time a significant reduction of a sample volume to the 25 % of original leads to decrease in accuracy. This can be explained by the fact that instances critical to describe the class separation can not be included to the sample of small volume.

Even a small reduction of the original sample volume in 25 % (up to 75 % of the original sample volume) yielded acceptable accuracy and reduces training time by more than 1.32 times. Reducing the volume of the original sample by a half afforded the gain in speed by 1.99 times. This confirms expediency of application of the proposed mathematical support in the neural network model building by precedents.

The closest analogue to the proposed method for determining the instance informativity is a set of indicators proposed in [30]. In contrast to the proposed in this paper, the indicators [30] characterize separately instance properties to be informative

**Table 2** The fragment of experimental results on model building by the formed samples

Data set dimensionality factors			$ep_{tr.}$	$E_{tr.} \%$	$E_{all.} \%$
$S \%$	$N \%$	$t_{tr.}, s$			
25	25	3.6	643	0	2.9
25	50	5.2	734	0	2.9
25	75	7.4	768	0	2.9
25	100	9.1	983	0	2.9
50	25	5.2	732	0	1.95
50	50	9.7	893	0	1.95
50	75	14.3	922	0	1.95
50	100	18.7	1000	0	1.95
75	25	7.5	731	0	0.49
75	50	14.9	720	0	0.49
75	75	23.4	881	0	0.49
75	100	28.1	954	0	0.49
100	25	9.5	793	0	0
100	50	18.3	931	0	0
100	75	27.6	982	0	0
100	100	37.2	1000	0	0

relatively to the external and internal borders, as well as to the class centers, which is their advantage in the problems of the data visualization and analysis. However, their disadvantages are low speed due to the need to calculate distances between instances, as well as the need and ambiguity of indicator integration to the comprehensive measures of instance informativity.

The advantage of the indicators proposed in this paper is that there is no need to calculate the distances between instances, but disadvantage is the necessary to divide the feature space. However, this disadvantage can be seen as an advantage in the case of large samples: if we use a partition that is simple from a computational point of view (for example, a regular grid) and know the minimum and maximum values of each feature than the computational cost of the proposed indicators will be less than the using of a set [30].

As it evident from the Table 2, with the sufficient reduction of number of features  $N$  in the formed sample do not provides reduction in accuracy, so for the considered task it is possible to select a small feature subset providing good accuracy.

The experiments also found that the most individually informative features are  $x_{10}$ ,  $x_{12}$ ,  $x_{15}$ ,  $x_{17}$ ,  $x_{20}-x_{23}$ , and the least informative features are  $x_4-x_6$ ,  $x_8, x_9$ ,  $x_{16}, x_{18}, x_{19}$ ,  $x_{28}$ . The obtained results are explained by the fact that the permanent inflammatory process in lung tissue involved granulocytes and activation of granulocytes, mononuclear cells and thrombocytes, consumption of complement growth of immune complexes and von Willebrand factor present. This allows identify the key factors supporting the inflammatory process. This is activation of the metabolic

activity of neutrophils and monocytes, including the generation of reactive oxygen species, thrombocytopenia and platelet activation and consumption of complement elevated level of the CIC.

The obtained results on feature selection are consistent with the published data on the cellular composition and activity of cells in the inflammation [32–34]. Consequently, the proposed method allows to select informative peripheral blood parameters, reflecting the intensity of the inflammatory process and its permanence in the broncho-pulmonary tissue in chronic obstructive bronchitis.

## 7 Conclusions

The urgent problem of mathematical support development is solved to automate the sampling and feature selection at diagnostic model building by precedents.

The method of training sample and feature selection is proposed. It determines the weights characterizing the term and feature usefulness for a given initial sample of precedents and given feature space partition. It characterizes the individual absolute and relative informativity of instances relative to the centers and the boundaries of feature intervals based on the weight values. This allows to automate the sample analysis and its division into subsamples, and, as a consequence, to reduce the training data dimensionality. This in turn reduces the time and provides an acceptable accuracy of neural model training.

The practical significance of obtained results is that the software realizing the proposed indicators is developed, as well as experiments to study their properties are conducted. The experimental results allow to recommend the proposed indicators for use in practice, as well as to determine effective conditions for the application of the proposed indicators.

The practical task of automation of chronic obstructive bronchitis diagnosis on the data of clinical laboratory observations is experimentally solved. The most informative features and instances for chronic obstructive bronchitis diagnosis were selected. This makes possible to reduce a neural network model complexity and speed-up its training process, and also to reduce the number of laboratory measurements, and, therefore, to reduce the number of measurements, and laboratory analyzes, allowing to reduce the cost and speed up the process diagnostics of chronic obstructive bronchitis.

The further studies may be focused on integration of the proposed indicators of instance and feature informativity with exhaustive and evolutionary search methods, as well as to study their applicability in solving a wide class of practical problems in technical and biomedical diagnostics, and in pattern recognition.

**Acknowledgments** The work is supported by the state budget scientific research project of Zaporizhzhya National Technical University “Intelligent information technologies of automation of designing, simulation, control and diagnosis of manufacturing processes and systems” (state registration number 0112U005350) and by the international project “Centers of Excellence for young RESearchers” of European Commission (no 544137-TEMPUS-1-2013-1-SK-TEMPUS-JPHES).

## References

1. Engelbrecht, A.: Computational Intelligence: An Introduction. Wiley, Sidney (2007)
2. Jankowski N., Grochowski M.: Comparison of instance selection algorithms I. Algorithms survey. In: Proceedings of 7th International Conference on Artificial Intelligence and Soft Computing (ICAISC-2004), Zakopane, 7–11 June, pp. 598–603. Springer, Berlin (2004)
3. Reinarz, T.: A unifying view on instance selection. *Data Min. Knowl. Disc.* **1**(6), 191–210 (2002)
4. Hart, P.E.: The condensed nearest neighbor rule. *IEEE Trans. Inf. Theory* **14**, 515–516 (1968)
5. Aha, D.W., Kibler, D., Albert, M.K.: Instance-based learning algorithms. *Mach. Learn.* **1**(6), 37–66 (1991)
6. Gates, G.: The reduced nearest neighbor rule. *IEEE Trans. Inf. Theory* **18**,(13), 431–433 (1972)
7. Kibbler, D., Aha, D.W.: Learning representative exemplars of concepts: an initial case of study. In: Proceedings 4th International Workshop on Machine Learning, Irvine, 22–25 June 1987, pp. 24–30. - Morgan Kaufmann, Burlington (1987)
8. Wilson D.L.: Asymptotic properties of nearest neighbor rules using edited data. *IEEE Trans. Syst. Man Cybern.* **2**(13), 408–421 (1972)
9. Tomek, I.: An experiment with the edited nearest-neighbor rule. *IEEE Trans. Syst. Man Cybern.* **6**, 448–452 (1976)
10. Jankowski, N.: Data regularization. In Proceedings of Fifth Conference on Neural Networks and Soft Computing, Zakopane, 6–10 June 2000, pp. 209–214. Polish Neural Networks Society, Cz stochowa (2000)
11. Broadley C.E.: Addressing the selective superiority problem: automatic algorithm/model class selection Machine Learning. In: Proceedings of Tenth International Conference, Amherst, 27–29 June 1993, pp. 17–24. Morgan Kaufmann, Burlington (1993)
12. Wilson, D.R., Martinez, T.R.: Reduction techniques for instancebased learning algorithms. *Mach. Learn.* **38**(13), 257–286 (2000)
13. Ritter, G.L., Woodruff, H.B., Lowry, S.R., Isenhour, T.L.: An algorithm for a selective nearest neighbor decision rule. *IEEE Trans. Inf. Theory* **21**(16), 665–669 (1975)
14. Domingo, C.R. Gavalda, O.: Watanabe Adaptive sampling methods for scaling up knowledge discovery algorithms. In: Proceedings of Second International Conference on Discovery Science (DS'99), Tokyo, 6–8 December 1999, pp. 172–183. Springer, Berlin (1999)
15. Skalak, D.B.: Prototype and feature selection by sampling and random mutation hill climbing algorithms. In: Proceedings of Eleventh International Conference on Machine Learning, New Brunswick, 10–13 July 1994, pp. 293–301. Morgan Kaufmann, Burlington (1994)
16. Kohonen, T.: Learning vector quantization. *Neural Networks* **1**, 303 (1988)
17. Madigan, D., Raghavan, N., DuMouchel, W et al.: Likelihood-based data squashing: a modeling approach to instance construction. *Data Min. Knowl. Discov.* **6**(12), 173–190 (2002)
18. Li, B., Chi, M., Fan, J., Xue, X.: Support cluster machine. In: Proceedings of 24th International Conference on Machine Learning, Corvallis, 20–24 June 2007, pp. 505–512. New York (2007)
19. Evans, R.: Clustering for Classification: Using Standard Clustering Methods to Summarise Datasets with Minimal Loss of Classification Accuracy. VDM Verlag, Saarbrücken (2008)
20. Sane, S.S., Ghatol, A.A.: A Novel supervised instance selection algorithm. *Int. J. Bus. Intell. Data Min.* **2**(14), 471–495 (2007)
21. Reeves, C.R., Bush, D.R.: Using genetic algorithms for training data selection in RBF networks. In: Liu, H., Motoda, H. (eds.) *Instance Selection and Construction for Data Mining*, Part VI, pp. 339–356. Kluwer, Norwell (2001)
22. Suykens, J.A., Vandewalle, J.: Least squares support vector machine classifiers. *Neural Proc. Lett.* **9**(13), 293–300 (1999)
23. Koskimaki, H., Juutilainen, I., Laurinen, P., Roning, J.: Two-level clustering approach to training data instance selection: a case study for the steel industry. In: Proceedings of International Joint Conference on Neural Networks (IJCNN-2008), Hong Kong, 1–8 June 2008, pp. 3044–3049. IEEE (2008)

24. Dash, M., Liu, H.: Feature selection for classification. *Intell. Data Anal.* **11**, 131–156 (1997)
25. Guyon, I., Elisseeff, A.: An introduction to variable and feature selection. *J. Mac. Learn. Res.* **1**(3), 1157–1182 (2003)
26. Handl, J.: Semi-supervised feature selection via multiobjective optimization/J. Handl, J. Knowles//Neural networks: International Joint Conference, Vancouver, 16–21 July 2006: proceedings. Los Alamitos: IEEE, 2006. pp. 3319–3326 (2006)
27. John, G.H., Kohavi, R., Pfleger, K.: Irrelevant features and the subset selection problem. In: Proceedings of Eleventh international conference on Machine Learning, New Brunswick, 10–13 July 1994, pp. 121–129. Morgan Kaufmann Publishers, San Francisco (1994)
28. Yu, S.: Feature selection and classifier ensembles: a study on hyperspectral remote sensing data: proefschrift ... doctor in de wetenschappen. Universitaire Instelling Antwerpen, Antwerpen (2003)
29. Subbotin, S.: The neuro-fuzzy network synthesis and simplification on precedents in problems of diagnosis and pattern recognition. *Opt. Mem. Neural Networks Inf. Opt.* **22**(12), 97–103 (2013)
30. Subbotin, S.A.: Methods of sampling based on exhaustive and evolutionary search. *Autom. Control Comput. Sci.* **47**(13), 113–121 (2013)
31. Kolisnyk, N.V., Subbotin, S.O.: Modeling of immunopathogenesis of chronic obstructive bronchitis using neural networks. In: Proceedings of II International Conference “Modern problems of biology, ecology and chemistry”, Zaporizhzhya, 1–3 October 2009. ZNU, Zaporizhzhya, pp. 124–125 (2009) (in Ukrainian)
32. Pryor, W.A., Stone, K.: Oxidants in cigarette smoke: radicals, hydrogen peroxides, peroxy nitrate and peroxy nitrite. *Ann. N.Y. Acad. Sci.* **686**, 12–28 (1993)
33. Prieto, A., Reyes, E., Bernstein, E.D. et al.: Defective natural killer and phagocytic activities in chronic obstructive pulmonary disease are restored by glycophosphopeptidal (Inmunoferón). *Am. J. Respir. Crit. Care Med.* **163**(17), 1578–1583 (2001)
34. Colvin, R.B., Bhan, A.K., Mc Cluskey, R.T. (eds.): *Diagnostic immunopathology*. Raven Press, New York, pp. 139–154 (1995)

# New Methods for the Reliability Analysis of Healthcare System Based on Application of Multi-State System

Elena Zaitseva, Vitaly Levashenko, Jozef Kostolny and Miroslav Kvassay

**Abstract** A healthcare system is complex and high-risk. Therefore reliability analysis of a healthcare system is principal step in its development and exploitation. The high-risk of a healthcare system is caused by different factors as human error, failure of devices and equipment, software fault, etc. These factors correlate with complex structure of a healthcare system that consists of technical and human parts. But as a rule in reliability engineering the analysis and estimation of technical components and human factor are implemented based on different methods that have different mathematical backgrounds. One of possible decision of this problem is development of new mathematical model, that allows to describe booth as technical components as human factors. Such model can be defined based on representation of a healthcare system as Multi-state System, for which can be define some (more than only two) performance levels.

## 1 Introduction

Reliability is a principal attribute of any system. Reliability principles are used successfully in industries to help evaluate, calculate, and improve the overall reliability of complex technical systems. As a rule complex system isn't heterogeneous and includes sub-system (components) of different types, for example, as equipment, software, human factor, organization aspects, etc. Such system can be interpreted as complex socio-technical system [20]. Reliability analysis of such system is based on

---

E. Zaitseva (✉) · V. Levashenko · J. Kostolny · M. Kvassay  
Department of Informatics, University of Zilina, Zilina, Slovakia  
e-mail: elena.zaitseva@fri.uniza.sk

V. Levashenko  
e-mail: vitaly.levashenko@fri.uniza.sk

J. Kostolny  
e-mail: jozef.kostolny@fri.uniza.sk

M. Kvassay  
e-mail: miroslav.kvassay@fri.uniza.sk

different methods of reliability engineering. There are special methods, approaches and algorithms for quantitative analysis of reliability for every of this system components (for example, human factor, equipment, software) is realized based on special approaches of reliability theory [26]. Development of reliability analysis of socio-technical system allows extending types of analyzed system. One of new types of system is healthcare system that has principal goal to assure a patient correct treatment.

Principal specific of modern healthcare systems is intensive application of information technologies [12, 25]. The Table 1 from [25] illustrates a development of healthcare information system from an immature stage to a national stage. This is explained through entities, services and infrastructures in a defined point in time. Each stage has its own characteristics that differentiate it from other stages, but all of them have to be reliable.

The application of information technologies in healthcare cause the typical structure of a healthcare system in reliability analysis terms that has been proposed in [37]. According to this structure a healthcare system is complex and high-risk. Therefore the problem of reliability analysis and estimation of this system in design and exploitation is important. One of the first investigations of reliability in medicine was [31]. In this paper author declared principal items of reliability engineering for a healthcare system as reliability analysis of medical equipment's and devices. This tendency has been developed in most works for reliability analysis of healthcare system. For example, in papers [5, 8, 29, 30] different aspects of safety and reliability of healthcare system hardware and software are analyzed. But functioning of healthcare system (correct treatment of patient) isn't caused by reliable devices and equipment only. A human factor has high influence to healthcare system reliability too. *Human reliability analysis* (HRA) methods allow examining the probability of medical errors and risk factor for correct patient treatment [19]. Human errors for a healthcare system have been considered as independent problem [9, 19]. Therefore reliability analysis of healthcare system is separated in two independent problems: (i) reliability analysis of medical devices and equipment, and (ii) human reliability analysis of medical errors. But practical work shows that these problems are not independent and must be considered. For example, some medical error can be caused by incorrect functioning of medical devices and human medical error can influence to functioning of devices and equipment [7, 9]. In papers [37], new tendency of healthcare system reliability analysis is considered: the reliability analysis has to base on joint evaluation of all principal parts (components) of healthcare system according to these papers. This tendency supposes the application of new background for a healthcare system reliability analysis. New method and mathematical background are proposed in this paper for reliability analysis of healthcare system. The important problem of this method is development of technique that permits to investigate every system component (devices, equipment, and human factor) based on united mathematical background. This technique allows estimating a healthcare system as a whole and doesn't have to separate the investigation of human factor and technical part of system (devices, equipment, and software). The theoretical conception of this method has been presented in papers [39, 40].

**Table 1** Healthcare system stage according to [25]

Stages	Entities	Department	Infrastructure
1. Hospital administration	Hospital administration	<ul style="list-style-type: none"> <li>● Patient</li> <li>● Billing</li> <li>● Wards management</li> <li>● Diagnostics management</li> <li>● MIS</li> </ul>	LAN
2. Hospital enterprise	Set of hospitals in enterprise	<p>Stage 1+</p> <ul style="list-style-type: none"> <li>● Finance</li> <li>● Materials management</li> <li>● HR management</li> <li>● Electronic claims and payments processing</li> </ul>	Internet based access with HIPAA
3. EMR basic	Hospital + Lab + Pharmacy	<p>Stage 2 +</p> <ul style="list-style-type: none"> <li>● Laboratory information system</li> <li>● Radiology information system</li> <li>● PACS</li> <li>● Pharmacy</li> </ul>	Secure HL7 based communication
4. Clinical decision support	Stage 3 + Medical colleges	<p>Stage 3+</p> <ul style="list-style-type: none"> <li>● Computerized provider order entry</li> <li>● International codification of diseases</li> <li>● Alerts/ Contraindications</li> <li>● Used for educational purposes</li> </ul>	Fully connected and paperless—SaaS ( <i>Software as a service</i> ) Model
5. Clinical research	Stage 4 + Pharma companies	<ul style="list-style-type: none"> <li>● Clinical trials</li> <li>● Clinical data research based on drug prescriptions and reactions</li> </ul>	OaaS2 (Operations as a service) Model + RaaS3 (Research as a service) Model
6. Regional	Primary healthcare centers + Epidemiological centers + Regional government	<ul style="list-style-type: none"> <li>● Telemedicine</li> <li>● Aggregation of data from various hospitals at the regional level</li> </ul>	Regional network connecting all hospitals with PHC's and Epidemiological centers
7. National	Federal government	<ul style="list-style-type: none"> <li>● Data from all regions aggregated</li> <li>● Enables healthcare planning and government initiatives towards healthcare</li> </ul>	National network connecting all associated service providers in the healthcare process

The development of conception proposed in [37] is continued in this paper. Important step of this conception is definition of mathematical model of healthcare system to calculate reliability indices and measures. This model is constructed based on the definition of number of system performance level and mathematical methods that are used for estimation of healthcare system reliability. The development of healthcare system mathematical model in term of reliability analysis is considered in this paper in detail in Sect. 2. This mathematical model is used for estimation of one of reliability analysis aspect that is investigation of influence of fixed component state change to a healthcare system performance level. This investigation in reliability analysis is known as importance analysis [15]. It is considered in Sect. 3. In Sect. 4 the example of simple healthcare system and its reliability analysis based on the proposed methods are presented. This system includes technical component and components that are conformed to human factors.

## 2 Principal Steps in Reliability Analysis of Healthcare System

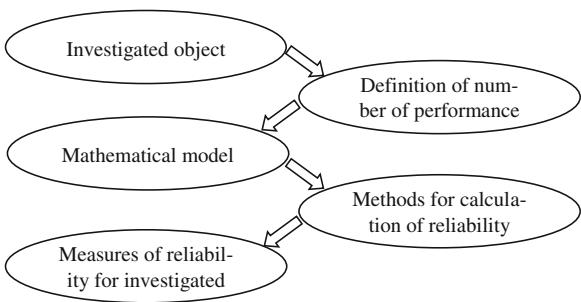
Preventable medical errors are important problem in healthcare system. The persistence of medical errors according to [9, 10] suggests that there is either an absence of reliability engineering analysis or a gap in the reliability analysis currently being performed. The decision of this problem can be implemented by changing process design of healthcare system and/or performing reliability analysis of healthcare system exploitation. Typical reliability analysis of healthcare system can include the following tasks:

- Assessment of the current reliability using past data and calculation of measures for healthcare system reliability;
- Identification of weak links and allocates higher reliability goals to them;
- Analysis of weak links within the healthcare system to predict potential failures or medical errors;
- Redesign healthcare system based on critical failures identify and using process and reliability improvement techniques that will have the most impact on the outcome;
- Verification of the design improvements and calculation of measures for healthcare system reliability;
- Define reliability specifications and document in a reliability program;
- Validation of sustainability for new design of healthcare system.

According to list of these tasks, the calculation of measures and indices for healthcare system reliability is actual problem for reliability improvement of such system. Therefore the development of new methods for estimation of healthcare system reliability is actual problem too.

The estimation of reliability of any complex system as a healthcare system includes four principal steps (Fig. 1) [42]:

**Fig. 1** The structure of a complex system reliability analysis



- the definition of number of performance levels for a system model;
- the mathematical representation of a system model;
- the development methods for the calculation of indices and measures of system reliability (for example, importance analysis);
- the measuring of the system behaviour.

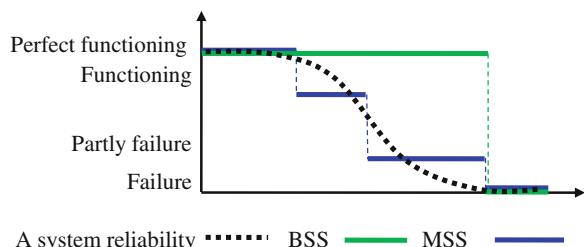
Consider specifics of these steps for the healthcare system in more details.

## 2.1 Definition of Number of Performance Levels for Healthcare System Mathematical Model

The number of performance levels for healthcare system can be defined based on the selection one of two possible mathematical models (Fig. 2): *Binary-State System* (BSS) and *Multi-State System* (MSS).

The first of them (BSS) defines only two states for the system reliability: the functioning and failure. This mathematical model is well known and widely used in reliability engineering. The system failure can be investigated in detail based on this model. However, the analysis of other performance levels before the system failure has some difficulties based on BSS. The other mathematical model that is MSS can be used to indicate some performance levels in system reliability behaviour.

**Fig. 2** Typical model in reliability analysis



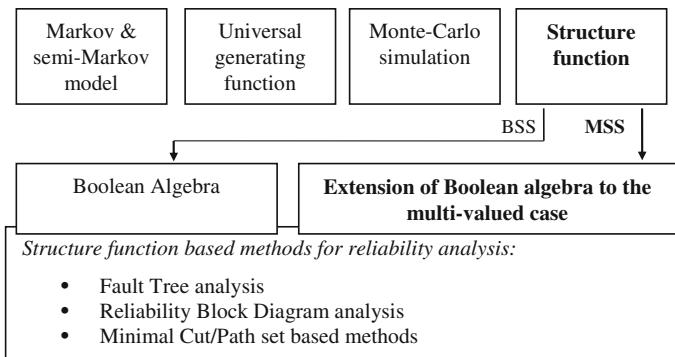
MSS reliability analysis is a more flexible to evaluating system reliability. MSS allows indicating more than two states for both as system and as component. It can be, for example, completely failed, partially failed, partially functioning and perfect functioning [17, 23]. However, dimension of MSS is dramatically large in comparison with BSS for equal number of system components. In addition, mathematical methods to MSS analysing is complex too. Therefore the application of MSS needs the development new special methods for the calculation of the measures and indices of system reliability.

The basic indices for MSS as system reliability, probability of system performance level, frequency have been considered in [17, 23]. In paper [15, 23, 36] some special indices, as *Importance Measures* (IMs), have been presented. These indices are quantitative estimation of the influence of one or some component states changes to system performance level. Effective methods for calculation of IMs have been proposed in paper [36] based on mathematical methods of Multiple-Valued Logic. The application of these methods for the estimation of healthcare system reliability is considered in this paper.

## 2.2 Healthcare System Mathematical Model

The model (mathematical representations) of healthcare system for its reliability estimation correlates with mathematical methods of reliability analysis. There are four principal groups of mathematical methods in reliability analysis (Fig. 3):

- Markov and semi-markov methods;
- Methods based on universal generation function;
- Monte-Carlo simulation;
- Structure function based methods.



**Fig. 3** Typical mathematical methods in reliability analysis

Every of these methods is used for the analysis for both as BSS as MSS and has some specifics in reliability analysis of a system. Markov model allows investigation of dynamic properties of system reliability. But Markov model dimension is increased extremely in depending of number of system components [14, 18]. The description of the system by the Universal Generating Function is used in the system reliability optimization [16]. The Monte-Carlo simulation as a rule is used for reliability assessment of system with large number of components [41]. The structure based methods were developed historically the first [3, 28]. According to these methods a system is represented and defined by the structure function. This function is defined the conformance of the system performance level and components states. As a rule for the structure function definition and representation is used Boolean functions [4, 33]. Only in some publications the structure function of MSS has been considered [13, 34]. In papers [24, 35, 36] the correlation of *Multiple-Valued Logic* (MVL) function and structure function was analyzed. The interpretation of the structure function as the MVL function allowed using the mathematical approach of MVL in the analysis of MSS structure function. The analysis the structure function and estimation of system reliability can be by such methods as fault tree analysis, minimal cut/path set analysis, etc. (Fig. 3).

The principal and important advantage of the representation of a system by the structure function is possibility to define mathematical model for a system with any structure and complexity complexity [17, 36]. Therefore the structure function of MSS is an appropriate mathematical model for healthcare system reliability analysis.

The *structure function* defines a system state (system reliability/availability/performance level) depending on the system components states. According to the definition of the structure function the system reliability in the stationary state is represented as [36]:

$$\begin{aligned}\phi(x_1, \dots, x_n) = \phi(x) : \{0, \dots, m_1 - 1\} \times \dots \times \{0, \dots, m_n - 1\} & \quad (1) \\ & \rightarrow \{0, \dots, M - 1\}.\end{aligned}$$

In (1) the  $x_i$  is the state of the  $i$ -th system component that can be defined from 0 (the component failure) to  $m_i - 1$  (the perfect component performance level):  $x_i \in \{0, \dots, m_i - 1\}$ , and the system reliability has  $M$  level from 0 (as the failure) to  $M - 1$  (as the perfect functioning). Note, the system component has different number of states:  $m_i \neq m_k$ , if  $i \neq k$  ( $i, k \in \{1, \dots, n\}$ ). The number of the system component is declared as  $n$ .

The structure function definition (1) is the definition for MSS, where system and its component have some (more than two) performance levels. The structure function of BSS is special case of MSS structure function and can be interpreted as the Boolean function [38]:

$$\phi(x_1, \dots, x_n) = \phi(x) : \{0, 1\}^n \rightarrow \{0, 1\}. \quad (2)$$

We will consider a coherent system in this paper. Such system has two principal assumptions for the structure function [17]: (a) the structure function (1) and (2) is

monotone, and (b) the system component state decrease does not improve the system reliability.

Every system component is characterized by probability of the component state:

$$p_{i,s_i} = \Pr\{x_i = s_i\}, \quad s_i \in \{0, \dots, m_i - 1\}. \quad (3)$$

### 2.3 Structure Based Method for Healthcare System Reliability Estimation

Some of methods for reliability analysis based on the structure function are shown in Fig. 1. But these methods and other structure function based methods as a rule don't allow investigating the dynamic properties of system reliability. It means that measures as system availability, reliability function and similar can be calculated. The analysis of a system performance level change has difficulties based on these methods.

In papers [35, 36] the application of Logical Differential Calculus for MSS reliability analysis has been proposed. The Logic Differential Calculus is used for analysis of dynamic properties of MVL function and this approach can be applied for analysis of dynamic behaviour of MSS that is determined by the structure function.

A system behavior and correlation of changes of components states and system reliability can be defined by mathematical tools of Logical Differential Calculus, in particular the Direct Partial Logic Derivative. The Direct Partial Logic Derivative with respect to variable  $x_i$  for the structure function (1) permits to analyse the system reliability change from  $j$  to  $\bar{j}$  when the  $i$ -th component state changes from  $a$  to  $\bar{a}$  [36, 38]:

$$\begin{aligned} & \partial\phi(j \rightarrow \bar{j})/\partial x_i(s \rightarrow \bar{s}) \\ &= \begin{cases} 1, & \text{if } \phi(s_i, \mathbf{x}) = j \text{ \& } \phi(\bar{s}_i, \mathbf{x}) = \bar{j}, \\ 0, & \text{other} \end{cases}, \end{aligned} \quad (4)$$

where  $\phi(s_i, \mathbf{x}) = \phi(x_1, \dots, x_{i-1}, s, x_{i+1}, \dots, x_n)$ ;  $\phi(\bar{s}_i, \mathbf{x}) = \phi(x_1, \dots, x_{i-1}, \bar{s}, x_{i+1}, \dots, x_n)$ ;  $s_i, \bar{s}_i \in \{0, \dots, m_i\}$ ,  $\bar{s}_i \neq s_i$  and  $j, \bar{j} \in \{0, \dots, M\}$ ,  $\bar{j} \neq j$ .

For example, consider MSS performance level change caused by the reduction of the  $i$ -th component state. It is represented as a change of the structure function value  $\phi(\mathbf{x})$  from state  $j$  into  $h$ . This change can be caused by the  $i$ -th component state change from  $s$  to  $s-1$ . In term of structure function this change is interpreted as change of the  $i$ -th variable value change from  $s$  to  $s-1$ . Therefore the Direct Partial Logic Derivative for MSS analysis is defined by the equation

$$\begin{aligned} & \partial\phi(j \rightarrow h)/\partial x_i(s \rightarrow s-1) \\ &= \begin{cases} 1, & \text{if } \phi(s_i, \mathbf{x}) = 1 \text{ and } \phi((s-1)_i, \mathbf{x}) = 0, \\ 0, & \text{other} \end{cases}. \end{aligned} \quad (5)$$

**Table 2** Importance measures

Short name	Description
SI	SI concentrates on the topological structure of the system and determines the probability of a system performance level change depending on the change of the $i$ -th component state
BI	BI of a given component is defined as the probability that such a component is critical to MSS functioning and represents loss in MSS when the $i$ -th component state reduced
CI	CI is similar to BI and take into account the probability of the $i$ -th component state reduction
CDRI	CDRI is similar to SI and take into account the probability of the $i$ -th component state reduction
DIRI	DIRI indicates the probability of a system performance level change depending on the change of any component state

The reduction of performance level and component state for BSS is defined as the system failure. Therefore the Eq. (5) for BSS is represented as:

$$\begin{aligned} \partial\phi(1 \rightarrow 0)/\partial x_i(1 \rightarrow 0) \\ = \begin{cases} 1, & \text{if } \phi(1_i, \mathbf{x}) = 1 \text{ and } \phi(0_i, \mathbf{x}) = 0 \\ 0, & \text{other} \end{cases}. \end{aligned} \quad (6)$$

Derivatives (4)–(6) allow calculation boundary states of system reliability depending on the  $i$ -th component state change that are agree with vector state:

$$\mathbf{x} = (x_1 \dots x_{i-1}, a_i \rightarrow a_i - 1, x_{i+1} \dots x_n). \quad (7)$$

The boundary states are one of basic conception in the reliability analysis and used in different mathematical methods for the computation of reliability indices and measures. For example boundary state is principal item in the investigation based on Fault Tree [6, 11, 22]. These states are considered and analysed in the method of *Failure Models and Effect Analysis* (FMEA) [11, 27]. The boundary states are used in Importance analysis for the computation of the *Importance Measures* (IM) [32, 36]. Importance analysis allows examining different aspects of reliability changes and the uncertainty in the system. IM quantifies the criticality of a particular component within the system. They have been widely used as tools for identifying system weaknesses, and to prioritise reliability improvement activities.

The most used IMs as *Structural Importance* (SI), *Birnbaum importance* (BI), *Critical importance* (CI), *Component Dynamic Reliability Indices* (CDRI) and *Dynamic Integrated Reliability Indices* (DIRI) are shown in Table 2 [17, 36].

The Direct Partial Logic Derivative is one of possible approaches for calculation of IMs [36]. In this paper we develop unify method for calculation of the IMs (Table 2)

based on the Direct Partial Logic Derivative (5). These measures can be used for the estimation of healthcare system component that has maximal influence to system functioning (system availability).

Therefore last step in the estimation of healthcare system according to Fig. 1 is the measuring of the system behaviour that is calculation of IMs.

## 2.4 Measures of Healthcare System

*Reliability function* for Binary-State System is one of basic measures of reliability and this measure is defined as probability of system function without failure during given period of time. In paper [17] this measure  $R(t)$  for MSS is interpreted as is the probability of the system being operational throughout the interval  $[0, t)$ :

$$R(t) = \Pr\{T \geq t, \phi(\mathbf{x}) > 0\}. \quad (8)$$

In stationary state instead of reliability function the measure as a system availability is used. A system availability for BSS is probability of a system functioning. But for MSS there are some levels of system performance and reliability analysis of this system needs to include estimation of probability of system to be in every of these performance state. Therefore there are some definitions of system availability for MSS. One of them allows to presented probability of MSS to be in state, that isn't less than performance level  $j$  ( $0 \leq j \leq M - 1$ ) [21]:

$$A(j) = \Pr\{\phi(\mathbf{x}) \geq j\}. \quad (9)$$

There is one more interpretation of MSS availability in [36] that in paper [17] is named as probability of MSS state. This measure is defined as probability of system reliability that is equal to the performance level  $j$ :

$$A_j = \Pr\{\phi(\mathbf{x}) = j\}. \quad (10)$$

The correlation of measures (9) and (10) is defined as:

$$A(j) = \sum_{r=1}^j A_r. \quad (11)$$

According to (11) the measure (10) is more exact and allows computing other measures as MSS availability (9) and MSS unreliability:

$$F = A_0 = 1 - \sum_{j=1}^{M-1} A_j. \quad (12)$$

Therefore we use in the development of methods for MSS reliability analysis the conception of system availability defined by (10).

The system availability  $A_j$  of the MSS is calculated based on probabilities of components states (3):

$$A_j = \sum_{\phi(\mathbf{x})=j} p_{1,s_1} \cdot p_{2,s_2} \cdot \dots \cdot p_{n,s_n}, \quad (13)$$

where  $(s_1, s_2 \dots s_n)$  are values of vector state  $\mathbf{x} = (x_1, x_2 \dots x_n)$  for which  $\phi(x_1 x_2 \dots x_n) = j$ .

The measures (9), (10) and (12) characterise system availability in general. Therefore they can be used for the availability estimation of healthcare system in stationary state. And these measures don't take into account of the influence of components states changes to a system performance level. There is other group of reliability measures that permit to investigate these aspects of system behaviour. These measures are IMs (Table 2) that are provided under importance analysis.

### 3 Importance Analyses

The importance analysis is part of reliability engineering that allows investigating the structural and topological aspects of the system in point of view of the reliability/availability. The IMs permit to indicate a system component with maximal/minimal influence to system reliability/availability. In particular, these measures indicate the probability of system performance level change caused by the change of fixed component state. Consider some of IMs (Table 2) in more details below.

In this paper investigated system is specified and it is a healthcare system. As a rule a healthcare system has specific properties that permit to consider this system as a coherent system. A coherent system has assumptions [2]:

- (a) All system components are relevant to the system;
- (b) The system structure function is monotone non-decreasing:  $\phi(x_1, \dots, 1, \dots, x_n) \neq \phi(x_1, \dots, 0, \dots, x_n)$ ;
- (c) The component state decreases to one only: from  $s$  to  $s - 1$ ;

These assumptions will be used in the definition and computation of IMs.

In addition need to note, that analysis of a coherent system performance level decrease and increase are similar. Therefore the analysis of a system performance level decrease will be considered in this paper only.

### 3.1 Structural Importance

SI is one of the simplest measures that focuses on the topological and structural properties and aspects of a system. According to the definition of this measure for BSS in paper [1], this measure determines the proportion of working states of the system in which the working of the  $i$ -th component makes the difference between system failure and its working. The generalization of the conception of SI for MSS has to take into account all possible changes of system performance levels and for MSS is defined as proportion of system state in which the change of the  $i$ -th component state makes difference in system performance level from  $j$  to  $h$  and its performance level  $j$ . Because we consider a coherent system, the component state changes from  $s$  to  $s - 1$  according to assumption (c) for a coherent system, the mathematical definition of SI is:

$$IS_i^{s,j \rightarrow h} = \frac{\rho_i^{(s,j \rightarrow h)}}{\rho_{s,j}}, \quad (14)$$

where  $\rho_i^{(s,j \rightarrow h)}$  is a number of system states when the change component state from  $s$  to  $s - 1$  results in the system reliability change from  $j$  to  $h$ ;  $\rho_{s,j}$  is number of the states for which  $\phi(s_i, x) = j$  and is calculated based on the structure function.

The number  $\rho_i^{(s,j \rightarrow h)}$  can be computed as the number of nonzero values of the Direct Partial Logic Derivative (5).

### 3.2 Birnbaum Importance

SI (14) investigates influence of system component to system performance level based on a system structure or topology only. But this measure doesn't take into account the probability of state of system components. This disadvantage can be eliminated by other importance measure as BI. BI is one of basic IMs and this measure is defined as the probability that a system is sensitive to inoperative state of the  $i$ -th system component in case of BSS [2]. In paper [38] new equation for the BI calculation has been proposed based on Direct Partial Boolean Derivatives:

$$BI_i = \Pr \{ \partial\phi(1 \rightarrow 0)/\partial x_i(1 \rightarrow 0) = 1 \}. \quad (15)$$

The Eq. (15) for calculation of BI can be generalized for MSS analysis:

$$IB_i^{s,j \rightarrow h} = \Pr \{ \partial\phi(j \rightarrow h)/\partial x_i(s \rightarrow s - 1) \neq 0 \}. \quad (16)$$

The definition of BI for MSS (16) is indicated as the probability that a system performance level  $j$  is sensitive to change from  $s$  to  $s - 1$  of state of the  $i$ -th component. In practical, the Eq. (16) is calculated as the probabilities of non-zero values of Direct Partial Logic Derivative  $\partial\phi(j \rightarrow h)/\partial x_i(s \rightarrow s - 1)$ .

### 3.3 Critical Importance

BI (16) describes the influence of a  $i$ -th component state change from  $s$  to  $s - 1$  on a system's performance level  $j$ , but doesn't take into account the probability of this component's state. CI adjusts it and is defined as the quantitative measure that the  $i$ -th system component is relevant to the system's performance level  $j$  if it has changed state from  $s$  to  $s - 1$  [36]:

$$CI_i^{s,j \rightarrow h} = BI_i^{s,j \rightarrow h} \cdot \frac{p_{i,s-1}}{A_j}, \quad (17)$$

where  $IB_i^{s,j \rightarrow h}$  is the  $i$ -th system component BI measure (16);  $p_{i,s-1}$  is probability of the  $i$ -th component state  $s - 1$  (3) and  $A_j$  is the probability of system's performance level  $j$  (10).

### 3.4 Dynamic Reliability Indices

*Dynamic Reliability Indices* (DRIs) have been introduced in paper [35]. DRIs allow the estimation of a component relevant to system failure. There are two groups of DRI: *Component Dynamic Reliability Indices* (CDRI) and *Dynamic Integrated Reliability Indices* (DIRI).

CDRI indicates the influence of the  $i$ -th component's state change from  $s$  to  $s - 1$  on a system's performance level  $j$  and is similar to the definition of SI, but CDRI includes two probabilities: (i) the probability of a system's performance level change caused by the  $i$ -th component's state change and (ii) the probability of the component state:

$$CDRI_i^{s,j \rightarrow h} = SI_i^{s,j \rightarrow h} \cdot p_{i,s-1}, \quad (18)$$

where  $SI_i$  is defined by (14);  $p_{i,s-1}$  is probability of the  $i$ -th component state  $s - 1$  (3).

DIRI has similar conception but this index indicates the influence of any component state change to the system performance level  $j$ . Therefore DIRI is defined as the probability of the performance level change from  $j$  to  $h$  that is caused by one of the system components state change (one of  $n$ ):

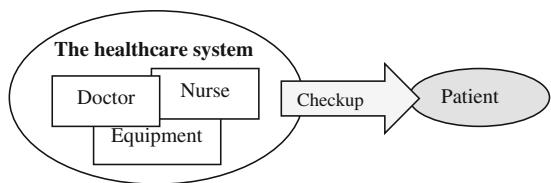
$$DIRI^{s,j \rightarrow h} = \sum_{i=1}^n CDRI_i^{s,j \rightarrow h} \prod_{\substack{q=1 \\ q \neq i}}^n (1 - CDRI_q^{s,j \rightarrow h}). \quad (19)$$

## 4 Importance Analyses of a Healthcare System

Consider the proposed methodology of a healthcare system analysis based on hand-calculation example. The investigated system is shown in Fig. 4. This system includes three components: doctor, nursing and equipment for patient checkup. The principal problem of this system is medical error that is interpreted as the system fault. There are 6 possible problems caused the medical error of this system that are most common type of problems according to [8, 9, 12]. The typical methods of reliability analysis are able to investigate the human error and technical problem in such system separately. The proposed method can be used to estimate this system as a whole.

According to proposed method the first step in the system reliability analysis is definition of number of system performance levels (Sect. 2.1). The system has five components ( $n = 5$ ) that are conformed to typical problem of patient's checkup (Table in Fig. 4). We will use three performance level for the investigate system ( $M = 3$ ) and two state for the system components states ( $m_1 = m_2 = m_3 = m_4 = m_5 = 2$ ). The values and contents of these variables are presented in Table 3. The state of this healthcare system is interpreted as: the fault (value 0) in case of a patient given wrong medication or incorrect amount (fatal error of treatment); the partial work (value 1)

**Fig. 4** The example of the healthcare system and possible medical error



Problems of checkup	Doctor	$x_1$	Misdiagnosis	Medical error
		$x_2$	Haste	
	Nurse	$x_3$	Incorrect interpretation of instructions	
		$x_4$	Inappropriate abbreviation	
	Equipment	$x_5$	Equipment problem	

**Table 3** The healthcare system components

The system components	Values of components		
	0	1	
$x_1$	Misdiagnosis	Correct diagnosis	
$x_2$	Haste	Sufficient time	
$x_3$	Incorrect interpretation of doctor instructions	Incorrect interpretation of doctor instructions	
$x_4$	Inappropriate abbreviation	Appropriate abbreviation	
$x_5$	Equipment fault	Working of equipment	

**Table 4** The structure function of the MSS for the healthcare system in Fig. 4

$x_4x_5$	$x_1x_2x_3$							
	0 0 0	0 0 1	0 1 0	0 1 1	1 0 0	1 0 1	1 1 0	1 1 1
0 0	0	0	0	0	1	1	1	2
0 1	0	0	0	0	1	1	1	2
1 0	0	0	0	0	1	1	2	2
1 1	0	0	1	1	1	2	2	2

if incorrect medical work doesn't cause a patient health; the perfect work (value 2) agrees with the correct treatment of a patient.

The second step of the method for the system reliability estimation is the definition of the mathematical model of the system and the mathematical method for the calculation of reliability indices. In the Sect. 2.2 the analysis of the possible mathematical methods and models have been considered. And the structure based methods for the analysis of system reliability have been proposed. Therefore the investigated healthcare system is represented by the structure function. This structure function is in Table 4.

Consider the importance analysis of this healthcare system. The reliability indices based on the healthcare system representation in the form of the structure function are calculated according to (10), (12) and (14)–(17).

This system availability (10) and system unavailability (12) for this system is calculated based on the structure function (Table 4) and the probabilities of the system components states (Table 5):

$$F = A_0 = p_{0,1}p_{0,2} + p_{0,1}p_{1,2}(p_{0,4} + p_{1,4}p_{0,5}) = 0.01472;$$

$$\begin{aligned} A_1 &= p_{0,1}p_{1,2}p_{1,4}p_{1,5} + p_{1,1}p_{0,2}p_{0,3} + p_{1,1}p_{0,2}p_{1,3}(p_{0,4} + p_{1,4}p_{0,5}) \\ &\quad + p_{1,1}p_{1,2}p_{0,3}p_{0,4} = 0.09424; \end{aligned}$$

$$A_2 = p_{1,1}p_{0,2}p_{1,3}p_{1,4}p_{1,5} + p_{1,1}p_{1,2}(p_{0,3} - p_{1,4} + p_{1,3}) = 0.89104.$$

**Table 5** The healthcare system components probabilities

Component state	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$
0	0.05	0.20	0.15	0.10	0.02
1	0.95	0.80	0.85	0.90	0.98

Continue the analysis of the healthcare system and calculate IMs (14)–(17). The first of these indices is SI (14). Values of SI calculated based on the Direct Partial Logical Derivatives (5). These derivatives for the healthcare system are presented in Table 5. Intermediate number of  $\rho_i^{(s,j \rightarrow h)}$  and  $\rho_{s,j}$  in (15) are calculate based on these derivatives (Table 6) and the structure function (Table 4) accordantly. Values of these numbers and SI measures for all system components are shown in Table 7.

SI (Table 7) permits to estimate the influence of doctor decision, nurse work and equipment function to a patient treatment. The fatal influence is evaluated by SIs  $IS_i^{(1,1 \rightarrow 0)}$  and  $IS_i^{(1,2 \rightarrow 0)}$  that analysis of the healthcare system failure. The measure  $IS_1^{(1,1 \rightarrow 0)}$  has maximal value, therefore the first component has important influence to the system functioning. It means that doctor's error (misdiagnosis) has maximal influence to correct treatment of patient. And nurse incorrect interpretation of doctor's instruction has not fatal influence to a patient treatment for this healthcare system structure (organization), because  $IS_3^{(1,1 \rightarrow 0)} = 0$  and  $IS_3^{(1,2 \rightarrow 0)} = 0$ .

Therefore SI allows investigate influence of fixed component state change to change of system performance level. But this measure doesn't take into account the probabilities of component state. This disadvantage is absent in importance analysis by BI (16).

For example, let us continue the analysis of the healthcare system in Fig. 4 and calculate BI for this system. This measure according to (15) is computed based on Direct Partial Logic Derivatives. These derivatives for the healthcare system are in Table 6. Consider the influence of misdiagnosis to fatal error in treatment of patient that is defined by BI for the first component  $IB_1^{(1,1 \rightarrow 0)}$  and  $IB_1^{(1,2 \rightarrow 0)}$ :

$$\begin{aligned} IB_1^{(1,1 \rightarrow 0)} &= \Pr\{\partial\phi(1 \rightarrow 0)/\partial x_1(1 \rightarrow 0) = 1\} \\ &= p_{0,2}p_{0,3} + p_{0,2}p_{1,3}p_{0,4} + p_{0,2}p_{1,3}p_{1,4}p_{0,5} + p_{1,2}p_{0,3}p_{0,4} = 0.0621 \\ IB_1^{(1,2 \rightarrow 0)} &= \Pr\{\partial\phi(2 \rightarrow 0)/\partial x_1(1 \rightarrow 0) = 1\} = p_{0,2}p_{1,3}p_{1,4}p_{1,5} + p_{1,2}p_{0,3}p_{1,4}p_{0,5} \\ &= 0.1521 \end{aligned}$$

BIs for all system components are in Table 8. These measures are calculated similar to  $IB_1^{(1,1 \rightarrow 0)}$  and  $IB_1^{(1,2 \rightarrow 0)}$  based on (16).

The analysis of data in Table 8 allows indicate the most possible problem (maximum value of BI) of the healthcare system is incorrect treatment without problem of patient health caused by misdiagnosis ( $IB_1^{(1,2 \rightarrow 1)} = 0.7056$ ). A misdiagnosis isn't importance for fatal error of patient treatment ( $IB_1^{(1,1 \rightarrow 0)}$  and  $IB_1^{(1,2 \rightarrow 0)}$ ).

CI (17) is similar to BI but takes into account the probability of the  $i$ -th component state change from  $s$  to  $s - 1$ . These measures for the healthcare system (Fig. 4) are shown in Table 9. Analysis of these measures allows defining that the doctor's haste is more possible reason of menace of a patient health ( $IC_2^{(1,1 \rightarrow 0)} = 0.5992$ ). This reason is important for non-fatal problem of a patient treatment too, because the value of CI measure is maximal for the second component in case of the system performance level change from level 2 to 1 ( $IC_2^{(1,2 \rightarrow 1)} = 0.4744$ ).

**Table 6** Direct partial logical derivatives  $\partial\phi(j \rightarrow j-1)/\partial x_1(1 \rightarrow 0)$  for the healthcare system in Fig. 4

$x_2$	0	0	0	0	0	0	0	0	1	1	1	1	1
$x_3$	0	0	0	0	1	1	1	0	0	0	1	1	1
$x_4$	0	0	1	1	0	0	1	0	0	1	0	0	1
$x_5$	0	1	0	1	0	1	0	1	0	1	0	1	0
$\partial\phi(1 \rightarrow 0)/\partial x_1(1 \rightarrow 0)$	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>0</b>	<b>1</b>	<b>1</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
$\partial\phi(2 \rightarrow 1)/\partial x_1(1 \rightarrow 0)$	<b>0</b>	<b>1</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>1</b>							
$\partial\phi(2 \rightarrow 0)/\partial x_1(1 \rightarrow 0)$	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>1</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
$x_1$	0	0	0	0	0	0	0	1	1	1	1	1	1
$x_3$	0	0	0	1	1	1	0	0	0	1	1	1	1
$x_4$	0	1	1	0	0	1	0	0	0	1	0	0	1
$x_5$	0	1	0	1	0	1	0	1	0	1	0	1	0
$\partial\phi(1 \rightarrow 0)/\partial x_2(1 \rightarrow 0)$	<b>0</b>	<b>0</b>	<b>0</b>	<b>1</b>	<b>0</b>	<b>0</b>	<b>1</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
$\partial\phi(2 \rightarrow 1)/\partial x_2(1 \rightarrow 0)$	<b>0</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>0</b>							
$\partial\phi(2 \rightarrow 0)/\partial x_2(1 \rightarrow 0)$	<b>0</b>												
$x_1$	0	0	0	0	0	0	0	1	1	1	1	1	1
$x_2$	0	0	0	1	1	1	0	0	0	1	1	1	1
$x_4$	0	1	1	0	0	1	0	0	1	0	0	1	1
$x_5$	0	1	0	1	0	1	0	1	0	1	0	1	0
$\partial\phi(1 \rightarrow 0)/\partial x_3(1 \rightarrow 0)$	<b>0</b>												
$\partial\phi(2 \rightarrow 1)/\partial x_3(1 \rightarrow 0)$	<b>0</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>0</b>	<b>0</b>							
$\partial\phi(2 \rightarrow 0)/\partial x_3(1 \rightarrow 0)$	<b>0</b>												

(continued)

Table 6 (continued)

**Table 7** Structural importance for the healthcare system

i	$\rho_i^{(s,j \rightarrow h)}$			$\rho_{s,j}$			$IS_i^{(s,j \rightarrow h)}$		
	$\rho_i^{(1,1 \rightarrow 0)}$	$\rho_i^{(1,2 \rightarrow 1)}$	$\rho_i^{(1,2 \rightarrow 0)}$	$\rho_{1,1}$	$\rho_{1,2}$	$\rho_{1,2}$	$IS_i^{(1,1 \rightarrow 0)}$	$IS_i^{(1,2 \rightarrow 1)}$	$IS_i^{(1,2 \rightarrow 0)}$
1	9	2	2	9	7	7	1.0000	0.2857	0.2857
2	2	5	0	4	6	6	0.5000	0.8333	0
3	0	3	0	4	5	5	0	0.6000	0
4	2	3	0	5	5	5	0.4000	0.6000	0
5	2	1	0	6	4	4	0.3300	0.2500	0

**Table 8** BI for the healthcare system Fig. 4

i	$IB_i^{(1,1 \rightarrow 0)}$	$IB_i^{(1,2 \rightarrow 1)}$	$IB_i^{(1,2 \rightarrow 0)}$
1	0.0621	0.7056	0.1521
2	0.0441	0.2235	0
3	0	0.2436	0
4	0.0392	0.2723	0
5	0.0360	0.1453	0

**Table 9** CI for the healthcare system Fig. 4

i	$IC_i^{(1,1 \rightarrow 0)}$	$IC_i^{(1,2 \rightarrow 1)}$	$IC_i^{(1,2 \rightarrow 0)}$
1	0.2108	0.3744	0.0085
2	0.5992	0.4774	0
3	0	0.3877	0
4	0.2663	0.0489	0
5	0.2890	0.0308	0

Consider next indices for the healthcare system. It is CDRI (18) that indicates the influence of the  $i$ -th component's state change from  $s$  to  $s - 1$  on a system's performance level  $j$ . CDRIs for the healthcare system are presented in Table 10. The comparison of SI (Table 7) and CDRI (Table 10) illustrate the influence of component state probability to the importance of the  $i$ -th component. So the influence of doctor's error (misdiagnosis) is less in case if the probability of this error is took into account:  $CDRI_1^{(1,1 \rightarrow 0)} = 0.0500$  but  $IS_1^{(1,1 \rightarrow 0)} = 1$ . CDRIs show that a doctor's haste is more possible for fatal error in treatment than misdiagnosis:  $CDRI_2^{(1,1 \rightarrow 0)} = 0.1000$ . This problem is important for non-fatal error of patient treatment too, because  $CDRI_i^{(1,2 \rightarrow 1)}$  has maximal value for second component (doctor's haste):  $CDRI_2^{(1,2 \rightarrow 1)} = 0.1667$ .

DIRI estimates the influence of any component state change to the system performance level  $j$ :  $DIRI^{1,1 \rightarrow 0} = 0.1729$ ,  $DIRI^{1,2 \rightarrow 1} = 0.2673$  and  $DIRI^{1,2 \rightarrow 0} = 0.0143$ . These indices indicate that incorrect medical work (non-fatal problem for a

**Table 10** CDRI for the healthcare system in Fig. 4

$i$	$CDRI_i^{(1,1 \rightarrow 0)}$	$CDRI_i^{(1,2 \rightarrow 1)}$	$CDRI_i^{(1,2 \rightarrow 0)}$
1	0.0500	0.0143	0.0143
2	0.1000	0.1667	0
3	0	0.0900	0
4	0.0400	0.0600	0
5	0.0066	0.0050	0

patient treatment) is more possible in case of any problem of the healthcare system components, because DIRI has maximal value for system performance level change from level two to one ( $DIRI^{1,2 \rightarrow 1} = 0.2673$ ).

Analysis of all IMs (Tables 7, 8, 9 and 10) shows that two components of the investigated healthcare system (Fig. 4) has maximal influence to a patient correct treatment and heal that are doctor's errors (misdiagnosis and haste). The misdiagnosis is more important in case of fatal problem in a patient treatment (Tables 8 and 9). But the doctor's hast is more possible for this system because measures CI (Table 9) and CDRI (Table 10) have maximal values for the second component and these measures take into account the probability of component state too. DIRIs allow indicating the incorrect medical work (non-fatal problem for a patient treatment) as more possible problem of this system. Therefore all IMs allow estimate different aspect of system performance level change depending on changes of components states.

## 5 Conclusions

Reliability analysis of a healthcare system is important problem in medicine. The application of information technologies in medicine supposes high level of reliability of medical equipment and devices. But only reliable technics can't ensure a correct patient treatment. The human factor must be included in reliability analysis too. Therefore a healthcare system is interpreted as the combination of technical and human components to assure a correct treatment of a patient [40]. This definition of a healthcare system needs the development new conception and methods of reliability analysis. These methods must to permit to investigate a healthcare system reliability based on the united background without separation of system to independent parts. The theoretical aspects of this conception have been discussed in [37, 39]. In this paper some practical positions of this conception are considered, in particular the mathematical interpretation of investigated system and some of possible techniques for reliability analysis of a healthcare system (as importance analysis).

The first and principal step in a healthcare system reliability analysis is the development of the mathematical representation of this system (development of mathematical model). The design of a healthcare system model has two specifics. The first

of them it is definition of number of a system performance levels. There are two possibilities for interpretation of investigation system depending of number of system performance level: BSS or MSS. BSS allows investigation only two performance level as working and failure that isn't sufficient for detail estimation of a healthcare system. Therefore MSS is more useful mathematical model that permits to analyse some changes in healthcare system reliability behaviour. The second specific in the modelling of investigated system is defined by mathematical methods that are used for calculation of reliability indices and measures. In this paper the structure function is considered because this mathematical interpretation of investigated system allows defining the mathematical model for system with high complexity. As result, the mathematical representation of investigated healthcare system is proposed in form of MSS structure function according to (1).

The second step in reliability analysis of healthcare system is calculation of indices and measures for quantitative reliability analysis. In this paper IMs (Table 2) are proposed for analysis and estimation of a healthcare system and algorithms for the calculation these measures are presented in Sect. 3. The simple hand calculation example of a healthcare system (Sect. 4) illustrates the efficiency of these measures application for reliability analysis.

Therefore in this paper we propose and consider the mathematical representation of a healthcare system in form of MSS structure function and analysis of this system based on IMs. These measures indicate healthcare components that are more important and principal for system correct functioning. And priority control of these components allows ensuring high level of a healthcare system and correct treatment of a patient.

**Acknowledgments** This work is supported by the grant of 7th RTD Framework Programme No 610425 (RASimAs) and grant of Scientific Grant Agency of the Ministry of Education of Slovak Republic (Vega 1/0498/14).

## References

1. Armstrong, M.: Reliability-importance and dual failure-mode components. *IEEE Trans. Reliab.* **46**(2), 212–221 (1997)
2. Birnbaum, Z.W.: On importance of difference components in a multi-component system. *Multi-Variant Anal* **2**, 581–592 (1969)
3. Butler, D.A.: A complete importance ranking for components of binary coherent systems with extensions to multi-state systems. *Naval Res. Logistics* **4**, 565–578 (1979)
4. Caldarola, L.: Coherent system with multi-state components. *Nucl. Eng. Design* **58**(5), 127–139 (1980)
5. Cohen, T.: Medical and information technologies converge. *IEEE Eng. Med. Biol. Mag.* **23**(3), 59–65 (2004)
6. Contini, S., Matuzas, V.: New methods to determine the importance measures of initiating and enabling events in fault tree analysis. *Reliab. Eng. Syst. Saf.* **96**(7), 775–784 (2011)
7. Deeter, J., Rantanen, E.: Human reliability analysis in healthcare. In: Proceedings of Symposium on Human Factors and Ergonomics in Health Care, pp. 45–51. Baltimore, USA (2012)
8. Dhillon, B.S.: Medical Device Reliability and Associated Areas. CRC Press, p. 264 (2000)

9. Dhillon, B.S.: Human Reliability and Error in Medical System. World Scientific, p. 233 (2003)
10. Dixon, N.M., Shofer, M.: Struggling to invent high-reliability organizations in health care settings: insights from the field. *Health Serv. Res.* **41**(4), 1618–1632 (2006)
11. Farcasiu, M., Prisecaru, I.: MMOSA—A new approach of the human and organizational factor analysis in PSA. *Reliab. Eng. Syst. Saf.* **123**(1), 91–98 (2014)
12. Fichman, R.G., Kohli, R., Krishnan, R.: The role of information system in healthcare: current research and future trends. *Inf. Syst. Res.* **22**(3), 419–427 (2011)
13. Hudson, J.C., Kapur, K.C.: Modules in coherent multistate systems. *IEEE Trans. Reliab.* **32**(2), 183–185 (1983)
14. Hellmich, M., Berg, H.P.: Markov analysis of redundant standby safety systems under periodic surveillance testing. *Reliab. Eng. Syst. Saf.* **133**(4), 48–58 (2015)
15. Kuo, W., Zhu, X.: Importance Measures in Reliability, Risk, and Optimization. Wiley, p. 472 (2012)
16. Li, Y., Zio, E.: A multi-state model for the reliability assessment of a distributed generation system via universal generating function. *Reliab. Eng. Syst. Saf.* **106**(1), 28–36 (2012)
17. Lisnianski, A., Levitin, G.: Multi-State System Reliability. Assessment, Optimization and Applications. World scientific, p. 358 (2003)
18. Lisnianski, A., Elmakias, D., Laredo, D., Haim, H.B.: A multi-state Markov model for a short-term reliability analysis of a power generating unit. *Reliab. Eng. Syst. Saf.* **98**(2), 1–6 (2012)
19. Lyons, M., Adams, S., Woloshynowych, M., Vincent C.: Human reliability analysis in healthcare: a review of techniques. *Int. J. Risk Saf. Med.* **16**(4), 223–237 (2004)
20. Mohaghegh, Z., Kazemi, R., Mosleh, A.: Incorporating organizational factors into Probabilistic Risk Assessment (PRA) of complex socio-technical systems: a hybrid technique formalization original research article. *Reliab. Eng. Syst. Saf.* **94**(5), 1000–1018 (2009)
21. Meng, F.C.: On some structural importance of system components. *J. Data Sci.* **7**, 277–283 (2009)
22. Merle, G., Roussel, J.M., Lesage, J.J.: Algebraic determination of the structure function of dynamic fault trees. *Reliab. Eng. Syst. Saf.* **96**(2), 267–277 (2011)
23. Natvig, B.: Multistate Systems Reliability Theory with Applications. Wiley, New York (2011). 232p
24. Reinske, K., Ushakov, I.: Application of Graph Theory for Reliability Analysis. Radio i Sviaz, Moscow, p. 208 (1988) (in Russian)
25. Rocha, A.: Evolution of information systems and technologies maturity in healthcare. *Int. J. Healthc. Inf. Syst. Inform.* **6**(2), 28–36 (2011)
26. Smith, D.J.: Reliability, Maintainability and Risk: Practical Methods for Engineers including Reliability Centred Maintenance and Safety-Related Systems. Butterworth-Heinemann Ltd, Oxford, p. 436 (2011)
27. Seyed-Hosseini, S.M., Safaei, N., Asgharpour, M.J.: Reprioritization of failures in a system failure mode and effects analysis by decision making trial and evaluation laboratory technique. *Reliab. Eng. Syst. Saf.* **91**(8), 872–881 (2006)
28. Shooman, M.L.: Probabilistic Reliability: An Engineering Approach. McGraw-Hill, New York (1968)
29. Spyrou, S., Bamidis, P.D., Maglaveras, N., Pangalos, G., Pappas, C.: A methodology for reliability analysis in health networks. *IEEE Trans. Inf. Technol. Biomed.* **12**(3), 377–386 (2008)
30. Taleb-Bendiab, A., England, D., Randles, M., Miseldine, P., Murphy, K.: A principled approach to the design of healthcare systems: autonomy versus governance. *Reliab. Eng. Syst. Saf.* **91**(12), 1576–1585 (2006)
31. Taylor, E.F.: The reliability engineer in the health care system. In: Proceedings of IEEE the 18th Annual Reliability and Maintainability Symposium, pp. 245–248, USA (1972)
32. Vaurio, J.K.: Ideas and developments in importance measures and fault-tree techniques for reliability and risk analysis. *Reliab. Eng. Syst. Saf.* **95**(2), 99–107 (2010)
33. Veeraraghavan, M., Trivedi, K.S.: A combinatorial algorithm for performance and reliability analysis using multistate models. *IEEE Trans. Comput.* **43**(2), 29–233 (1994)

34. Yu, K., Koren, I., Guo, Y.: Generalized multistate monotone coherent systems. *IEEE Trans. Reliab.* **43**(2), 242–250 (1994)
35. Zaitseva, E.: Reliability analysis of multi-state system. *Dyn. Syst. Geom. Theor.* **1**(1), 213–222 (2003)
36. Zaitseva, E.: Importance analysis of a multi-state system based on multiple-valued logic methods. In: Lisnianski, A., Frenkel, I. (eds.) *Recent Advances in System Reliability: Signatures, pp. 113–134*. Springer, Multi-state Systems and Statistical Inference, London (2012)
37. Zaitseva, E., Levashenko, V., Rusin, M.: Reliability analysis of healthcare system. In: *Proceedings of the IEEE Federated Conference on Computer Science and Information Systems*, pp. 169–175. Szczecin, Poland (2011)
38. Zaitseva, E., Levashenko, V.: Importance analysis by logical differential calculus. *Autom. Remote Control* **74**(2), 171–182 (2013)
39. Zaitseva, E., Kostolny, J., Kvassay, M., Levashenko, V., Pancerz, K.: Failure Analysis and Estimation of the Healthcare System. In: *Proceedings of the IEEE Federated Conference on Computer Science and Information Systems*, pp. 235–240. Krakow, Poland (2013)
40. Zaitseva, E., Kostolny, J., Kvassay, M., Levashenko, V., Pancerz, K.: Estimation of a healthcare system based on the importance analysis. In: Pancerz, K., Zaitseva, E. (eds.) *Computational Intelligence, Medicine and Biology: Selected Links*, pp. 3–22. Springer, London (2015)
41. Zio, E., Marella, M., Podofillini, L.: Importance measures-based prioritization for improving the performance of multi-state systems: application to the railway industry. *Reliab. Eng. Syst. Saf.* **92**(10), 1303–1314 (2007)
42. Zio, E.: Reliability engineering: old problems and new challenges. *Reliab. Eng. Syst. Saf.* **94**(2), 125–141 (2009)