

Reliability Estimation of Healthcare Systems using Fuzzy Decision Trees

Vitaly Levashenko, Elena Zaitseva,
 Miroslav Kvassay,
 University of Zilina,
 Department of Informatics
 Zilina, Slovakia
 Email: {vitaly.levashenko, elena.zaitseva,
 miroslav.kvassay}@fri.uniza.sk

Thomas M. Deserno
 Uniklinik RWTH Aachen
 Department of Medical Informatics
 52057 Aachen, Germany
 Email: deserno@ieee.org

□

Abstract—Reliability is an important characteristic of any system. Healthcare systems are typical examples of such systems. In reliability engineering, such systems are considered as complex, inhomogeneous, and uncertain, and require special mathematical representations. The structure function is a suitable model representing real systems. Methods of system reliability evaluation based on the structure function are well established but deterministic. This restricts its use for uncertain or incomplete data. A structure function can be created only for system in which correlations of all components are indicated and all component states are known. In this paper, a new method for structure function construction is proposed. Incomplete data is analysed using Fuzzy Decision Trees (FDTs), where input and output attributes are interpreted as component states and values of the structure function, respectively. This method is applied to reliability analysis of healthcare system. For illustration, we considered the system laparoscopic surgery that has 4 components and 36 state vectors. In addition we evaluate proposed method by 3 benchmark's systems with 243, 108, and 512 state vectors, respectively. Two of these benchmarks have 5 components and one has 4 components. Uncertainty is simulated by randomly deleting between 5% and 90% of all state vectors before constructing the structure function. With 50% of deleted stages, the error rate is below 0.2% for all three systems. We conclude that FDT-based reliability analysis is applicable for incomplete data in medical systems, too.

I. INTRODUCTION

THE investigation in reliability engineering of healthcare system has started as analysis of medical devices and equipment [1]. Until recently, reliability quantification of equipment and devices has been a principal tendency in medicine [2, 3, 4]. Independently, the human factor in medicine has been investigated.

A special area in reliability engineering investigating the influence of human factor is named *Human Reliability Analysis* (HRA). HRA aims at identifying the potential failure of the system resulting from human errors, analyzing causes and identifying appropriate countermeasures to prevent and reduce the linked risks as much as possible. Failures in healthcare are called Medical Errors, if the

patient's condition worsens or if patients develop additional illnesses. As documented in [5], the number of deaths due to these causes is 225,000 annually, with in-hospital medical errors causing 44,000 to 98,000 deaths, and 3,000,000 injuries annually. Of course, such situation must be changed.

A first step of improvement is to investigate causes and specifics of medical error. A human error in a healthcare system is considered as independent problem of reliability analysis [5, 6, 7, 8]. Authors of papers [7, 8, 9] have indicated methods that are most useful for medical error analysis. In these papers and other investigations, authors provide the adaptation of well-known and popular HRA-methods for the medical domain. The background of HRA methods is discussed in [9].

A healthcare system includes components of different types, such as technical components (equipment/devices) and the human factor. Therefore, this system is inhomogenous in the view of reliability engineering [3, 4, 7], and the construction of mathematical representation including all types of component become challenging [3, 7, 10]. The simplest decision is obtained for system of stationary states, where the time dependence of system behavior can be ignored. One of possible representation for this condition is a structure function. This is a deterministic model and all possible component states and performance levels must be indicated and reflected in the structure function. However, complete information of healthcare systems cannot be obtained, since the observation of all situations is impossible: some of them agree with hazard of patient's health. For example, author in [3] consider reliability and safety of pacemaker application that is interpreted as system with uncertainty. Furthermore, parts of the information are deducted from expert's experiences. This information is ambiguous and unequal. It can be considered with just some reliability or confidence. Therefore, the structure function of healthcare systems must be constructed based on uncertain data.

In this paper, a novel method for constructing the structure function is proposed. We take into account the uncertainties of initial data and suggest a method that is based on the *Fuzzy Decision Tree* (FDT). FDTs are widely used in data mining for analysis of uncertain data and

□ This work was supported by the grant of 7th RTD Framework Program No 610425 (RASimAs)

decision making in ambiguities [11, 12]. FDTs can be used naturally to analyze fuzzy data. Therefore, the initial data used to construct the structure function can be fuzzy. To transform a real system's performance level and its component states into an exact model, such as the structure function, fuzziness of bounds of performance levels/states are considered. In addition, FDTs allow to take into account uncertainties caused by incompletely specified data. This is still possible if it is expensive to obtain all data about the real system's behavior, if there is only sparse data, or if data is acquired incompletely due to poor documentation.

As a rule, if the exact value of the actual data about the system behavior cannot be determined, we need to rely on more data to give additional information necessary to correct the theoretical model used [13]. An FDT allows reconstructing these data with different levels of the confidence [14]. The use of FDTs for the construction of the structure function assumes the induction of a tree that is based on the data (fuzzy and/or crisp). The values of the structure function are then defined by the FDT for all combinations of the component states.

This paper is structured as follows. Section II discusses the concept of structure function. Principal steps of the proposed method are considered in Section III. The detail description of all steps is given in Section IV. The results of method evaluation are presented in Section V, and the discussion and conclusion are given in Section VI.

II. THE STRUCTURE FUNCTION IN HEALTHCARE ANALYTICS

A. Structure Function

The concept of structure function is introduced in reliability engineering in order to mathematically describe the real system that is studied. In this case, the system is represented as a mapping that assigns a system state to every possible profile of component states. Therefore, the system performance level is defined from the states of all its components, and all possible component states as well as all performance levels must be indicated and reflected in the structure function.

The structure function allows to represent the system's reliability behavior for two typical mathematical models, i.e. the *Binary-State System* (BSS) and the *Multi-State System* (MSS). BSS permits only two states to investigate the system and its components: perfect functioning and complete failure. However in practice, many systems can exhibit different performance levels between these two extremes of full function and fatal failure [15, 16]. MSS is a mathematical model that is used to describe a system with several (more than two) levels of performance [15, 17, 18].

The concept of the structure function is used to represent BSS and MSS, and associates the space of component states and system performance levels:

$$\phi(x_1, \dots, x_n) = \phi(\mathbf{x}) : \{0, \dots, m_1 - 1\} \times \dots \times \{0, \dots, m_n - 1\} \rightarrow \{0, \dots, M - 1\}, \quad (1)$$

where $\phi(\mathbf{x})$ is the system state (performance level) from failure ($\phi(\mathbf{x}) = 0$) to perfect functioning ($\phi(\mathbf{x}) = M - 1$); $\mathbf{x} = (x_1, \dots, x_n)$ is a vector of system component's states (state vector). The state $\phi(\mathbf{x}) = 0$ represents the total failure of the component while state $\phi(\mathbf{x}) = m_i - 1$ corresponds to perfect functioning of the i -th component.

Therefore, (1) defines the MSS structure function of n components (subsystems). The state of each component can be denoted by a random variable, x_i , and every component of the MSS is characterized by probabilities of its states:

$$p_{i,s} = \Pr\{x_i = s\}, \quad s \in \{0, \dots, m_i - 1\}. \quad (2)$$

As it is shown in [10], reliability analysis of healthcare systems can be provided if the system is represented by a structure function (1). Then, the typical measures of reliability, e.g., the system availability (probabilities of the system performance levels), can be calculated by the structure function [18, 10]:

$$A_j = \Pr\{\phi(\mathbf{x}) = j\}, \quad j = 0, \dots, M - 1. \quad (3)$$

For example, consider a simple system to indicate medical errors of two components: a doctor (x_1) and diagnostic device (x_2). Suppose three levels in the doctor work: 0 is interpreted as fatal error, 1 is incorrect work (without fatal result) and 2 is perfect work. The devices can have two states only: 0 is failure and 1 is proper function. The system quantification has three levels: 0 is incorrect diagnosis with fatal consequence for the patient, 1 is incorrect diagnosis without fatal consequence for the patient, and 2 is correct diagnostics. This system (the medical error identification) is represented by the structure function in Table I. The probability of every performance level (availability) of this system is calculated according to (3):

$$\begin{aligned} A_0 &= \Pr\{\phi(\mathbf{x}) = 0\} = p_{1,0}p_{2,0}, \\ A_1 &= \Pr\{\phi(\mathbf{x}) = 1\} = p_{1,0}p_{2,1} + p_{1,1}p_{2,0}, \\ A_2 &= \Pr\{\phi(\mathbf{x}) = 2\} = p_{1,1}p_{2,1} + p_{1,2}p_{2,0} + p_{1,2}p_{2,1} \end{aligned}$$

TABLE I.
THE STRUCTURE FUNCTION OF THE MEDICAL ERROR IDENTIFICATION

The system components		$\phi(\mathbf{x})$
x_1	x_2	
0	0	0
0	1	1
1	0	1
1	1	2
2	0	2
2	1	2

There are a lot of methods to estimate different aspects of the system's reliability based on the structure function [3, 10, 15, 17, 18]. Therefore, the structure function allows for investigating the system's reliability. In some applications – such as the medical domain – the construction of the structure function is a complex problem, because the structure function (1) usually is assumed to be exact and ambiguities are not taken into account. Since data about real

systems is uncertain, the structure function may not be realistic, which is – as a rule – typical for healthcare systems [7].

B. The Structure Function for Healthcare Systems

From the point of view of reliability analysis, estimating preventable medical errors in healthcare system is the principal goal.

During the past decade, healthcare delivery has seen the introduction of ever more sophisticated and complex equipment for preventable medical errors. Furthermore, the human factor still is an important component of such systems. The persistence of medical errors according to [7] 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 the process design of healthcare system and/or performing the reliability analysis of healthcare system's exploitation. It supposes the adaptation and application of reliability analysis methods for healthcare system.

The first step in reliability estimation of any system is to determine a mathematical model (representation). Such mathematical model of a healthcare system is:

- complex, the system cannot be represented by a typical structure (series, parallel, k -out-of- n etc.) only;
- inhomogeneous, it implies different types of the system components (hardware, software, human factor);
- uncertain, it is caused by human factor influence.

In case of the structure function, complexity and inhomogeneity of healthcare systems can be realized in a mathematical representation without special algorithms [15, 19]. In the construction of the structure function for healthcare system, the uncertainty of the initial data needs adaptation and the reasons of this uncertainty must be considered. According to (1), the structure function is defined if all of possible components and system states are indicated. As a rule, it is impractical to wait until all the component states are indicated. The first factor then is incomplete specification of data, because some values of the system component's states or performance levels cannot be obtained. In substitution, extra data on the healthcare system can be obtained through expert analysis. The second factor of uncertainty of initial data is the ambiguity and vagueness of collected data values. This type of ambiguity can be caused by an expert's subjective evaluation etc. For example, two experts can set different values of system performance level for equal situation [20, 21]. Therefore, uncertainty of initial data must be considered when constructing the structure function.

Uncertainties and ambiguities of a real system have been dealt with in reliability analysis using of the likelihood concept [13, 19]. However, uncertainties that are caused solely in evaluation by an expert's experience and judgement are not random in nature. This uncertainty can't be indicated in a quantitative form by probability theory.

Fuzzy logic, however, makes it possible to define the structure function in a more flexible form than the probabilistic approach. Consequently the structure function of a healthcare system must be constructed based on fuzzy data and incomplete samples. It is a typical problem in data mining. Therefore, we propose the use of the Fuzzy Decision Tree (FDT), which has been used widely in data mining, for analysis of uncertain data, and decision making with ambiguities. Here, collected data for structure function construction can be defined and characterized by likelihood (likelihood) or confidence if there is little data about the real system's behavior. An FDT allows to reconstruct these data with different levels of likelihood (confidence) [12, 14]. Applying FDT, we propose a new method for the construction of the structure function to mathematically represent healthcare system.

FDTs imply a tree based on the data about the system, which can be fuzzy and/or crisp, and the data can be specified incompletely. The values of the structure function are then defined for all combinations of component states by the FDT: component states are interpreted as FDT attributes and the structure function value agrees with one of the M values (classes) for the system performance level.

III. A NOVEL METHOD FOR STRUCTURE FUNCTION CONSTRUCTION BY FDT

The structure function is a construct of mathematical representations that can be defined by a system structure analysis or that can be based on expert data [13, 22]. The structure function for system that includes a human factor is funded based on the evaluation by an expert's experience and judgment. Any healthcare system is a typical example of such a system.

The construction of a structure function representing expert's data requires special analysis and transformation of the initial system's data [7, 23], because expert's data is uncertain as a rule. This uncertainty can be caused by a lot of factors, but we have considered two of them. The first factor is incompletely specified data, because some values of system states or performance levels cannot be obtained. For example, it can be expensive or it might need unacceptable long time to get the data. The second factor is ambiguity and vagueness of collected data values. This type of ambiguity can be caused, for instance, by inaccuracies or errors of measurement, or subjectivity of expert's evaluations.

Therefore, the construction of the structure function must appreciate two aspects. The first is a mapping assigning the system performance level to each possible profile of component states (for example, see Table I). The second is addressed by interpreting it as classification problem for uncertain data, which is typical for data mining. One of the possible options is applying decision trees or FDT [31-34].

Our approach is based on FDTs [24, 25] and includes the following steps (Fig.1):

- Collection of data in the repository according to requests of FDT induction;

- Representation of the system model in the form of an FDT that classifies component states according to the system performance levels;
- Construction of the structure function as decision table that is created by the inducted FDT.

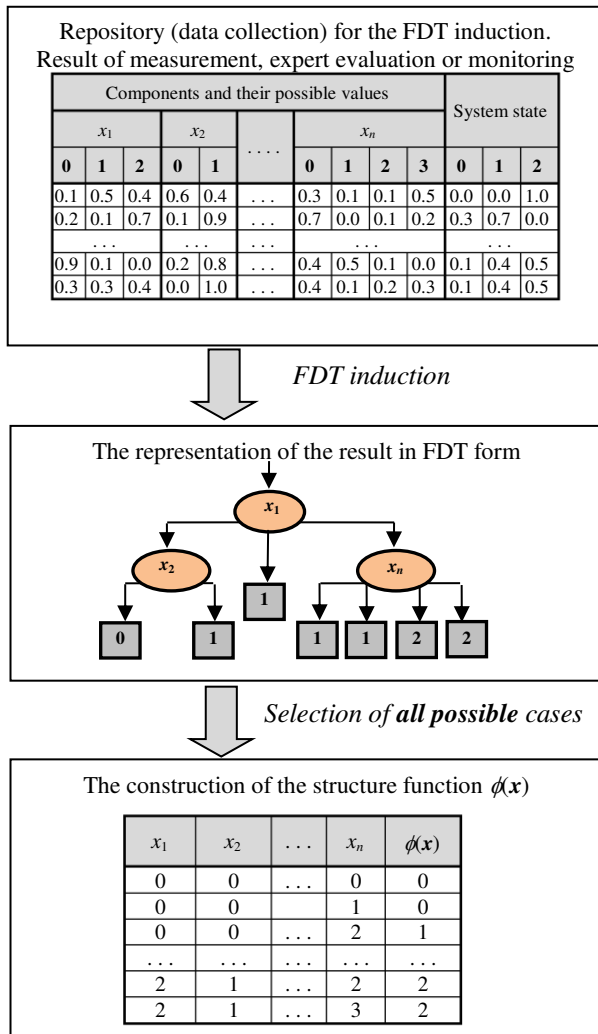


Fig. 1 Principal steps in the structure function construction based on FDT application

Therefore, the structure function is constructed as decision table that classifies the system performance level to each possible profile of components states. The decision table is formed based on a FDT that provides the mapping for all possible component states (input data) in M performance levels. FDT is inducted by uncertain data that is presented in form of specified repository.

IV. PRINCIPAL STEPS OF THE METHOD

A. The Repository for Data Collection

Collection of data in the form of a repository is provided by the monitoring or expert evaluation of values of the system's component states and the system's performance levels (Fig.1). Data for a repository is collected by

monitoring if it is possible (for example, for some devices or equipment). It is expert evaluation dominantly in case of the system with analysis of human factor that is important component of any healthcare system. Therefore data for the construction of the repository for the analysis of healthcare system collects the expert experience. The experts propose an estimation of every situation (state vector) and resulting system state (system performance level) with some possibilities (likelihoods). The transformation of expert knowledge into quantification data is implemented based general rules that used for the analysis of experts' knowledge. Some of such methods are presented and discussed in papers [20, 21]. Need to note that initial data structuration and the repository preparation is important step and can be considered as separate problem for investigation and development.

The column number is $n+1$ (for indication of n components and the system's performance level). All of the n columns are separated into m_i sub-columns, and the column for the system's performance level has M sub-columns. The sub-column is assigned with one of the values for component states or performance levels. Every row of the table represents one monitoring situation or evaluation. The table cell includes number (from 0 to 1) that interpreted as the likelihood of this value. Note that the sum of these possibilities for each value equals 1. Such data can be obtained from experts' evaluations or possibility of fuzzy clustering [26, 27]. These possibilities correspond to a membership function of fuzzy data [28]. This demand for initial data representation is caused by the method of FDT induction. Therefore, values of the i -th component state and the performance levels are defined by possibilities. These possibilities indicate ambiguity of collected data values for the analysis. Having indicated and considered the uncertainty of the monitoring data, the resulting accuracy of data analysis is increased.

For example, consider a simple laparoscopic surgery procedure [7]. It is typical healthcare system with human component. Let us indicate 4 components ($n = 4$) of this system: device (a laparoscopic robotic surgery machine [29]), two doctors (anesthesiologist and surgeon) and a nurse. Let us interpret this system as *Multi-State System* MSS and introduce the numbers of states for every component and number of performance levels of the system. Let this system has three performance levels ($M = 3$):

- 0 – non-operational (fatal medical error),
- 1 – partially operational (some imperfection),
- 2 – fully operational (surgery without any complication).

The device has two states ($m_1 = 2$):

- 0 – failure, and
- 1 – functioning.

The work of anesthesiologist is indicated by two states ($m_2 = 2$):

- 0 – non-operational (medical error),
- 1 – fully operational (without any complication).

TABLE II.
THE COLLECTED DATA FOR ANALYSIS OF LAPAROSCOPIC SURGERY SUCCESSFUL

No	x_1		x_2		x_3			x_4			$\phi(x)$		
	0	1	0	1	0	1	2	0	1	2	0	1	2
1	0.6	0.4	0.9	0.1	0.1	0.9	0.0	0.2	0.6	0.2	0.9	0.1	0.0
2	0.7	0.3	1.0	0.0	0.0	0.9	0.1	0.1	0.8	0.1	0.8	0.1	0.1
3	0.5	0.5	0.9	0.1	0.8	0.2	0.0	0.8	0.1	0.1	0.9	0.1	0.0
4	1.0	0.0	0.1	0.9	1.0	0.0	0.0	0.1	0.9	0.0	0.8	0.2	0.0
5	0.9	0.1	0.0	1.0	0.1	0.2	0.0	0.1	0.9	0.0	1.0	0.0	0.0
6	1.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0	1.0	0.0	0.0	1.0	0.0
7	1.0	0.0	0.0	1.0	0.0	0.1	0.9	0.0	0.3	0.7	0.1	0.8	0.1
8	0.0	1.0	1.0	0.0	0.0	1.0	0.0	0.0	0.6	0.6	0.1	0.9	0.0
9	0.1	0.9	0.1	0.9	0.1	0.1	0.8	1.0	0.0	0.0	0.1	0.8	0.1
10	0.3	0.7	0.9	0.1	0.0	0.0	1.0	0.0	0.5	0.5	0.0	0.1	0.0
11	0.2	0.8	0.0	1.0	0.9	0.1	0.0	0.0	1.0	0.0	1.0	0.0	0.0
12	0.0	1.0	0.0	1.0	0.1	0.9	0.0	0.8	0.2	0.0	0.0	1.0	0.0
13	0.1	0.9	0.2	0.9	0.1	0.8	0.1	0.0	0.6	0.4	0.0	0.0	1.0
14	0.2	0.8	0.0	1.0	0.0	0.1	0.9	1.0	0.0	0.0	0.1	0.8	0.1
15	0.3	0.7	0.0	1.0	0.0	0.1	0.9	0.1	0.8	0.1	0.0	0.1	0.9

The work of surgeon and the nurse can be modelled both by 3 levels ($m_3 = m_4 = 3$), i.e.:

- 0 – (the fatal error),
- 1 – (sufficient), and
- 2 – (perfect or the work without any complication).

The structure function of the system for analysis this simplified version of a laparoscopic surgery is composed of 36 situations (state vectors). The monitoring and expert analysis of this system permits to obtain some samples that represent the system behavior. In Table II, 15 samples are shown. In this table, 15 of 36 possible state vectors are indicated only. However, this information is uncertain because the data from real monitoring is incomplete and values are ambiguous. The monitoring does not allow obtaining information about all possible samples (situations) for fatal medical errors, because it would imply patient's health. So, the expert (or experts) adds some information about the system behavior, which can be uncertain, too. For example, an expert can indicate for the sample under the consideration the system performance level with likelihood only: as operational ($\phi(x) = 2$) with the likelihood of 0.0, partly operational ($\phi(x) = 1$) with likelihood 0.1 and non-operational ($\phi(x) = 0$) with the likelihood of 0.9.

Table II illustrates the correlation of monitoring data and the structure function for this system. The monitoring data can be transformed into the structure function (1) for the system based on the following rule: only the value with the highest likelihood is considered. For example, the variable x_1 in Table II has value 0 with the likelihood of 0.6 and value 1 with the likelihood of 0.4. The resultant value is defined as 0 in this case, but some of the state vectors are absent in the repository. Traditional mathematical approaches for system reliability analysis are based on the structure function, which cannot be used in this case. Therefore, the construction of a structure function (1) based on incomplete data requires a special transformation and the development of new methods.

B. Representation of system model in the form of an FDT

A decision tree (and FDT in particular) can be considered as an alternative form of the structure function. The structure function maps state vectors to each equivalence class of the system's performance levels. At the same time, a decision tree is a formalism for expressing mappings of input attributes (component's states) and output attribute/attributes (system performance level/s), consisting of an analysis of attribute nodes, which are linked to two or more sub-trees and leaves or decision nodes that are labeled with a class (in our case it is the system performance level) [14]. The outcome of the analysis is based on attribute values of a sample, where each possible outcome is associated with one of the sub-trees. A sample is classified by the starting at the root node of the tree. If this node is not a leaf, the outcome for the instance is determined and the process continues using the appropriate sub-tree. If a leaf is encountered, its label directs to the predicted class of sample. The system's component states are interpreted as values of the input attributes. The system's performance levels are considered as an instance that is classified into M classes.

FDT is one of the possible types of decision trees that permit to operate with fuzzy data (attributes) and methods of fuzzy logic. The construction of a FDT-based structure function assumes ambiguous data and that the analysis of such data can be implemented based on the methods of fuzzy logic [11, 30, 31]. The ambiguity of data values may be present in the attributes (system components states) and the exact class of the instance (system performance level).

There are different methods to induct a FDT [11, 24, 25, 31]. The principal goal of all methods is to select expanded attributes and determine the leaf node. The FDT induction is implemented based on some initial data that is interpreted as a training test. Every training sample includes n attributes A_1, \dots, A_n and an output attribute B . The construction of the structure function supposes its correlation with the FDT (Table III): the system performance level is the output

attribute and the component's states (state vectors) are input attributes. Each input attribute (component state) A_i ($1 \leq i \leq n$) is measured by a group of discrete values from 0 to m_i-1 that agree with the values of the i -th component states: $\{A_{i,0}, \dots, A_{i,j}, \dots, A_{i,m_i-1}\}$. The FDT assumes that the input set A_1, \dots, A_n is classified as one of the output value attributes B . The output attribute value B_w agrees with one of the system's performance levels and is defined by M values ranging from 0 to $M-1$ ($w = 0, \dots, M-1$).

For example, set input attributes $\{A_1, A_2, A_3, A_4\}$ and output attribute B for the success of laparoscopic surgery are indicated in Table IV according to FDT terminology. Each attribute is defined as: $A_1 = \{A_{1,0}, A_{1,1}\}$, $A_2 = \{A_{2,0}, A_{2,1}\}$, $A_3 = \{A_{3,0}, A_{3,1}, A_{3,2}\}$, $A_4 = \{A_{4,0}, A_{4,1}, A_{4,2}\}$ and $B = \{B_0, B_1, B_2\}$.

TABLE III.
CORRELATION OF THE TERMINOLOGIES OF FDT AND RELIABILITY ANALYSIS

FDT	System reliability
Number of input attributes: n	Number of system components: n
Attribute A_i ($i = 1, \dots, n$)	System component x_i ($i = 1, \dots, n$)
Attribute A_i values: $\{A_{i,0}, \dots, A_{i,j}, \dots, A_{i,m_i-1}\}$	The i -th system component state: $\{0, \dots, m_i-1\}$
Output attribute B	System performance level $\phi(x)$
Values of output attribute B : $\{B_0, \dots, B_{M-1}\}$	Values of the system's performance levels: $\{0, \dots, M-1\}$
Decision table	Structure function

A fuzzy set A with respect to an universe U is characterized by a membership function $\mu_A : U \rightarrow [0,1]$, assign a A -membership degree, $\mu_A(u)$, to each element u in U . $\mu_A(u)$ gives us an estimation of u belonging to A . The cardinality measure of the fuzzy set A is defined by $M(A) = \sum_{u \in U} \mu_A(u)$, which is the measure of the size of A .

For $u \in U$, $\mu_A(u) = 1$ means that u is definitely a member of A and $\mu_A(u) = 0$ means that u is definitely not a member of A , while $0 < \mu_A(u) < 1$ means that u is partially a member of A . If either $\mu_A(u) = 0$ or $\mu_A(u) = 1$ for all $u \in U$, A is a crisp set. The set of input attributes A is crisp for which $\mu_A(u) = 0$ or $\mu_A(u) = 1$. The values of input and output attributes are defined by the membership function. They are obtained from the monitoring data (Table II) according to the correlation shown in Table IV. Indicated attributes values are used as a training test to construct the FDT.

In this paper, we adopt the FDT induction principle to construct the structure function based on cumulative information estimates [24, 25]. The cumulative information estimates allow defining the criterion of expanded attribute selection to induct FDT with different properties. These estimates are calculated by measures of entropy and information. Entropy and information have been introduced to the information theory as a probabilistic approach. The application of these measures assumes that the sum of

possibilities of all values of every attribute equals 1 [26, 27, 28]. Note that the likelihood of attribute's value in terms of FDT induction is measured as confidence degree or degree of truth in this value.

TABLE IV.
ATTRIBUTES VALUES OF SYSTEM FOR LAPAROSCOPIC SURGERY PROCEDURE

Structure function	Attribute	Attribute values	Description of attribute values
The first component state, x_1	A_1	$A_{1,0}$	Device failure
		$A_{1,1}$	Device working
The second component state, x_2	A_2	$A_{2,0}$	Error of anesthesiologist
		$A_{2,1}$	Anesthesiologist work without complication
The third component state, x_3	A_3	$A_{3,0}$	Error of surgeon
		$A_{3,1}$	Sufficient work of surgeon
		$A_{3,2}$	Perfect work of surgeon
The fourth component state, x_4	A_4	$A_{4,0}$	Error of nurse
		$A_{4,1}$	Sufficient work of nurse
		$A_{4,2}$	Perfect work of nurse
System performance level $\phi(x)$	B	B_0	Fatal medical error
		B_1	Some imperfection
		B_2	Surgery without any complication

Figure 2 depicts the FDT for the laparoscopic surgery example. This FDT has 4 levels and includes all input attributes. This implies that all input attributes are considered to be significant for this system. The attribute A_3 is most significant as it agrees with the surgeon's work. Therefore, A_3 is associated with the FDT root (top node). This attribute can have the values $A_{3,0}$, $A_{3,1}$, and $A_{3,2}$, which are associated with branches of the FDT. Each branch agrees with a block of the output attribute values, and the confidences of every value of the output attributes are indicated. This block is a leaf if one output attribute has a sufficient level of confidence. In the other case, the FDT provides the analysis of the next input attribute. In this example, the value $A_{3,1}$ and $A_{3,2}$ of the attribute A_3 supposes the analysis of the attribute A_1 and A_2 , respectively. The sufficient value of the output attribute is defined by the user between 0 to 1. For instance in the laparoscopic surgery example, this sufficiency of the FDT has been set to 0.750 (Fig. 2).

The FDT can be transformed into classification (decision) rules. A new sample e may be classified into different classes with different confidence. Let the FDT have R leaves $L = \{l_1, \dots, l_r, \dots, l_R\}$, then each leaf $l_r \in L$ corresponds to one (r -th) classification rule. The condition part of the classification rule is a group of conditions that is represented in the form: "attribute is attribute's value". Such conditions are interconnected with an AND operator. The attributes are associated with the nodes in the path from the root to the leaf l_r .

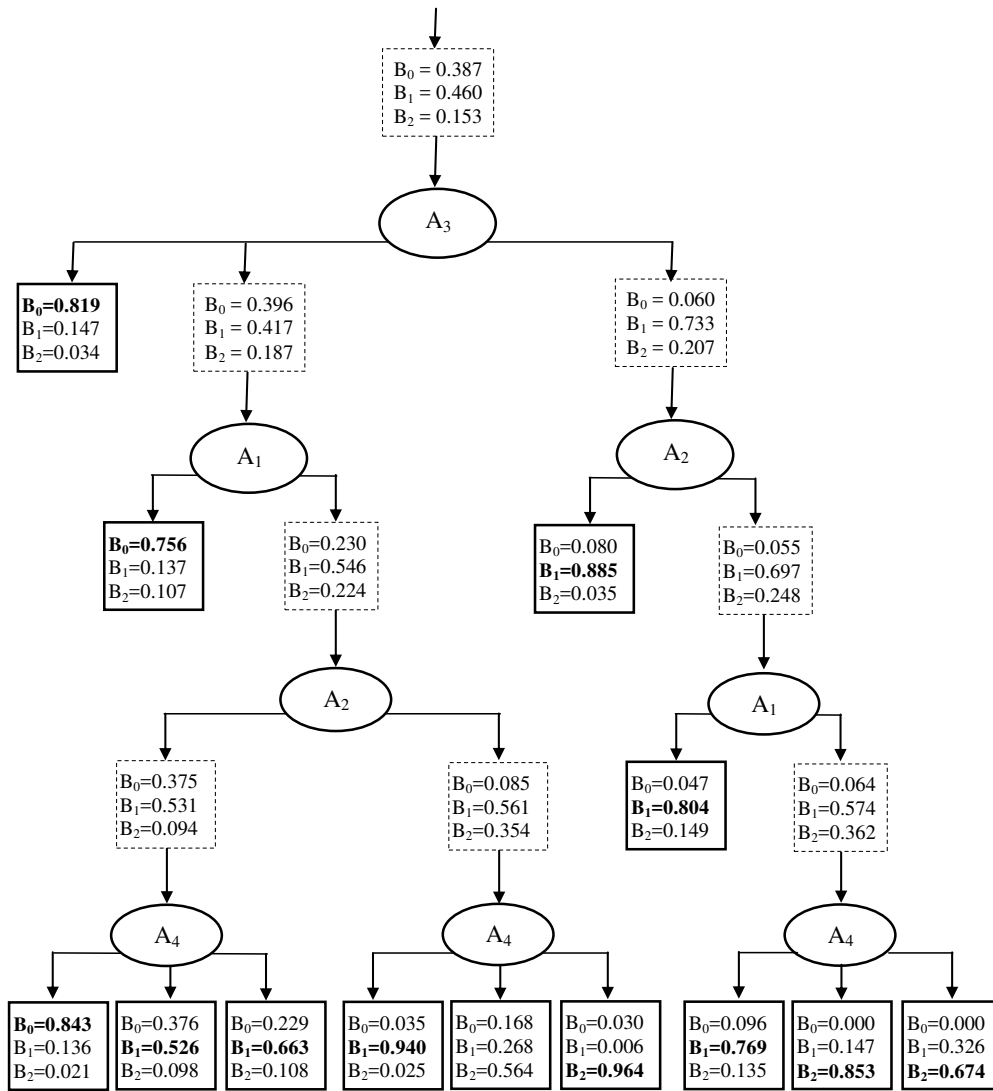


Fig. 2 The FDT for the construction of structure function to analyze laparoscopic surgery successful

The attribute's values are associated with the respective outgoing branches of the nodes in the path. Conclusions of the r -th rule are the values of class attribute B with their truthfulness vector $\mathbf{F}^r = [F_1^r, \dots, F_j^r, \dots, F_M^r]$ for each r -th leaf l and each j -th class B_j . Each value F_j^r means the certainty degree (confidence) of the class B_j attached to the leaf node l_r .

Let's consider the path $P_r(e) = \{[A_{i_1j_1}(e)]^r, \dots, [A_{i_sj_s}(e)]^r, \dots, [A_{i_sj_s}(e)]^r\}$ from the FDT root to the r -th leaf. This path $P_r(e)$ consist of S nodes which are associated with attributes $A_{i_1}, \dots, A_{i_s}, \dots, A_{i_s}$ and respectively their S outgoing branches associated with the values $A_{i_1j_1}, \dots, A_{i_sj_s}, \dots, A_{i_sj_s}$. Then the r -th rule has the following form:

$$\text{IF } (A_{i_1} \text{ is } A_{i_1j_1}) \text{ and } \dots \text{ and } (A_{i_s} \text{ is } A_{i_sj_s}) \quad (3)$$

THEN B (with truthfulness \mathbf{F}^r)

The classification rules for the laparoscopic surgery example are shown in Fig.3.

C. Construction of the structure function based on the FDT

As it was shown in Section III, the structure function agrees with the decision table. The decision table can be constructed based on the FDT or classification (decision) rules or a [25]. A decision table indicates all possible values of input attributes and agrees with the structure function. Therefore, cal the decision table is calculated for all possible values of the component's states and considered as the structure function (1).

For the laparoscopic example, assume that the state vector is $\mathbf{x} = (0 \ 1 \ 1 \ 0)$. According to the FDT (Fig. 2), the system output attribute for all possible vector states can have the value 0, 1, and 2 with the confidence of 0.387, 0.460, and 0.153, respectively. However, these confidences are under the given threshold of 0.750. Therefore, the analysis of the state vectors is implemented based on the FDT. It starts with the attribute A_3 (Fig. 2). The value of this component state is $x_3 = 1$ and the branch for the attribute value $A_{3,1}$ is

considered. The output attribute can have the value 0, 1, and 2 with the confidence of 0.396, 0.417, and 0.187, respectively. Again, all these confidences are below the threshold 0.750. Therefore, a decision about the value of the system's performance level is impossible and the analysis continues with the attribute A_1 . The estimation of this attribute is implemented using a branch with attribute value $A_{1,0}$ because the specified state vector includes $x_1 = 0$. The branch of this value has the leaf-node. Therefore, the value of the output attribute is defined as the value with the maximal confidence, which is 0.756 for the value 0. Considering this, the system performance level for the specified state vector is $\phi(x) = 0$. Therefore the fatal medical error is possible with confidence 0.756 if the level of the surgeon's work is sufficient and there is a malfunction of the laparoscopic robotic surgery machine. The analysis of other state vectors is similar and allows to obtain all possible values of the system performance level in the form of the structure function.

$r=1$: IF A_3 is $A_{3,0}$	THEN B with $F^1 = [0.819; 0.147; 0.034]$;
$r=2$: IF A_3 is $A_{3,1}$ and A_1 is $A_{1,0}$	THEN B with $F^2 = [0.756; 0.137; 0.107]$;
$r=3$: IF A_3 is $A_{3,1}$ and A_1 is $A_{1,1}$ and A_2 is $A_{2,0}$ and A_4 is $A_{4,0}$	THEN B with $F^3 = [0.843; 0.136; 0.021]$;
$r=4$: IF A_3 is $A_{3,1}$ and A_1 is $A_{1,1}$ and A_2 is $A_{2,0}$ and A_4 is $A_{4,1}$	THEN B with $F^4 = [0.376; 0.526; 0.098]$;
$r=5$: IF A_3 is $A_{3,1}$ and A_1 is $A_{1,1}$ and A_2 is $A_{2,0}$ and A_4 is $A_{4,2}$	THEN B with $F^5 = [0.229; 0.663; 0.108]$;
$r=6$: IF A_3 is $A_{3,1}$ and A_1 is $A_{1,1}$ and A_2 is $A_{2,1}$ and A_4 is $A_{4,0}$	THEN B with $F^6 = [0.035; 0.940; 0.025]$;
$r=7$: IF A_3 is $A_{3,1}$ and A_1 is $A_{1,1}$ and A_2 is $A_{2,1}$ and A_4 is $A_{4,1}$	THEN B with $F^7 = [0.168; 0.268; 0.564]$;
$r=8$: IF A_3 is $A_{3,1}$ and A_1 is $A_{1,1}$ and A_2 is $A_{2,1}$ and A_4 is $A_{4,2}$	THEN B with $F^8 = [0.030; 0.006; 0.964]$;
$r=9$: IF A_3 is $A_{3,2}$ and A_2 is $A_{2,0}$	THEN B with $F^9 = [0.080; 0.885; 0.035]$;
$r=10$: IF A_3 is $A_{3,2}$ and A_2 is $A_{2,1}$ and A_1 is $A_{1,0}$	THEN B with $F^{10} = [0.047; 0.804; 0.149]$;
$r=11$: IF A_3 is $A_{3,2}$ and A_2 is $A_{2,1}$ and A_1 is $A_{1,1}$ and A_4 is $A_{4,0}$	THEN B with $F^{11} = [0.096; 0.769; 0.135]$;
$r=12$: IF A_3 is $A_{3,2}$ and A_2 is $A_{2,1}$ and A_1 is $A_{1,1}$ and A_4 is $A_{4,1}$	THEN B with $F^{12} = [0.000; 0.147; 0.853]$;
$r=13$: IF A_3 is $A_{3,2}$ and A_2 is $A_{2,1}$ and A_1 is $A_{1,1}$ and A_4 is $A_{4,2}$	THEN B with $F^{12} = [0.000; 0.326; 0.674]$;

Fig. 3 The classification rules for the construction of structure function for the analysis of successful laparoscopic surgery

It is important to note that this method of the structure function constructing based on FDTs permits to compute (restore) data missing from the monitoring. Therefore, probabilities of system performance can be calculated according to typical methods used in reliability engineering, i.e., based on the structure function. For example, the system availability (1) can be calculated for the system that is presented the laparoscopic surgery success based on the indicated values of the component state probabilities (Table V). The availability of this system according to (3) and data

in Table V is calculated based on the structure function constructed through the FDT(Fig. 2):

$$A_0 = 0.098, \quad A_1 = 0.214, \quad A_2 = 0.688 \quad (4)$$

TABLE V.
PROBABILITIES OF THE COMPONENTS STATES

System component description	Component's states probabilities		
	$P_{i,2}$	$P_{i,1}$	$P_{i,0}$
The laparoscopic robotic surgery machine functioning, x_1	—	0.98	0.02
The anesthesiologist's work, x_2	—	0.94	0.06
The surgeon's work, x_3	0.64	0.27	0.09
The nurse's work, x_4	0.47	0.35	0.18

The values of the availabilities in (4) imply that laparoscopic surgery can be with:

- fatal medical error with probability 0.098,
- sufficient result (some complications) with probabilities 0.214 and
- perfect result 0.688 (without any complications).

Other measures can be computed by the structure function too. For example, importance measures for this system are defined according to the algorithms considered in [10, 30].

V. RESULTS

Our method has been investigated depending on the level specify of the initial data. We considered 3 benchmarks for the method evaluation [13, 15]:

- The system 1 has 3 performance levels, consist of 5 component and its structure function has 243 state vectors;
- The system 2 has 5 performance levels, consist of 4 component and its structure function has 108 state vectors;
- The system 3 has 4 performance levels, consist of 5 component and its structure function has 521 state vectors;

The structure functions are known and defined for these systems. We have used this data to examine efficiency and accuracy of proposed method for the construction of the "new" structure function based on incompletely specified data. The incompleteness is modeled by random deleting of some state vectors and assigned performance level value. The range of deleted states is changed from 5% to 90%. The "new" structure function is constructed based these incompletely specified data and compare with initial structure function.

The constructed structure functions include individual or small groups of misclassified state vectors. Therefore, we have estimated this misclassification by an error rate. The constructed structure functions and initial completely and exact specified functions are compared and the error rate is calculated as the ratio of wrong values of the structure function to the dimension of unspecified part of the function. The experiments have been iterated 1000 times for every system. The error rate for every system and for different level of unspecified component states vector is shown in Fig.4. The error rate is depended on unspecified part of the initial data (state vectors) according results presented in

Fig.4. This error increases essentially, if the unspecified part is most than 80% for all investigated systems. We obtained an insignificant growth of the error rate if the unspecified part is less than 10%. Therefore according to this investigation we can declare recommended specify of initial data between 10% and 80%. This result is typical for FDT application [11, 12]. The specification of data less than 10% isn't sufficient for the correct construction of correlations between input and output attribute. And higher level of data specification (more 80%) restricts variations of FDT induction and causes misclassification values.

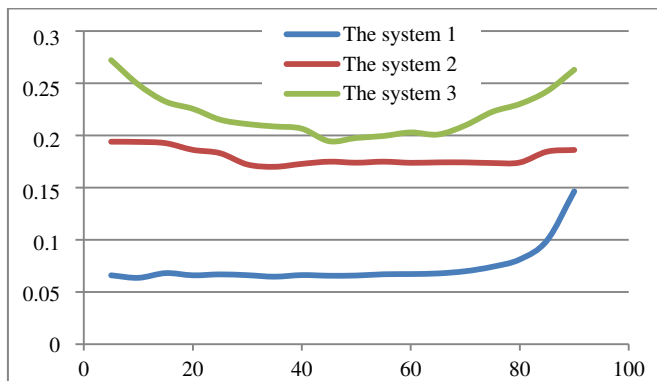


Fig. 4 The error rate for the construction of the structure function for three systems

VI. DISCUSSION

The main contribution of this paper is that we have developed novel and original method for the construction of the structure function based on incompletely specified and ambiguous data. It directly supports MSS reliability estimation and can cope with uncertain data for the analysis of system reliability/availability. This is typical problem for reliability analysis of healthcare systems, where the data cannot be obtained for all possible situations.

The analysis of the error rate for the proposed method for the construction of the structure function based on FDT shows that this method has good efficiency. This method is acceptable for the incompletely data and the incompleteness of initial data can be indicated from 10% to 85%. The constructed structure function by the proposed method has less error rate than maximal error rate in interval of the incompleteness.

But error rate is not caused by range of incompleteness of initial data only. There are some factors that can influence to error rate. First of all it is specific properties of structure function. One of very important properties is monotony of structure function that is typical for coherent system [17, 30]. The structure function of the system 1 is monotonic and the error rate of this function is minimal. The structure function of the system 2 is monotonic too. Other property of structure function is "uniformity" that mean similar number of state vectors for all system performance levels. For example, the system 2 has maximal number of state vectors

for one of performance level (46 of 108). Therefore range of incompleteness has small influence to error rate. Numbers of state vectors for every performance levels of the system 3 are similar. Therefore error rate increases for boundary range of incompleteness.

Beside the structure function properties, the system dimension (number of state vectors) influences to the error rate. This influence will be investigated in the future. Now, we can suppose that error rate decreases opposite to increasing of system dimension. The construction of the structure function of 50 state vectors is possible based on 20-30 state vectors (40-60% of defined state vectors). The structure function construction based on 5-10 state vectors (10-20%) is possible too. But level of accuracy depends on the quality of this set of state vectors. It will be essential to continue the verification and validation of our proposed method with data sets of different properties and sizes

Very important advantage of proposed method for constructing of the mathematical model of investigated system is possibility to ignore step of qualitative analysis that is typical in reliability analysis based on uncertain data. Methods for qualitative analysis (Failure mode and Effect Analysis (FMEA), Failure Mode Effects and Criticality Analysis (FMECA), Hazard Operability Study (HAZOP) etc.) are very good investigated in reliability engineering, but most of them include empirical evaluation of investigated system [8, 9, 13]. Conclusion

Our comprehensive experiments have shown that reliability estimation of healthcare systems is possible with uncertain or incomplete data if the structure function is estimated using fuzzy decision trees. This opens a wide range of clinical applications for saver healthcare.

REFERENCES

- [1] Taylor E.F. "The reliability engineer in the health care system," in *Proc IEEE the 18th Annual Reliability & Maintainability Symposium*, USA, pp.245 – 248, 1972.
- [2] Cohen T. "Medical and Information Technologies Converge," *IEEE Engineering in Medicine and Biology Magazine*, vol.23, no. 3, pp.59-65, 2004, <http://dx.doi.org/10.1109/MEMB.2004.1317983>
- [3] Dhillon B.S. *Reliability Technology, Human Error, and Quality in Health Care*, CRC Press, 190 p., 2008.
- [4] Taleb-Bendiab A, England D, et al. "A principled approach to the design of healthcare systems: Autonomy vs. governance," *Reliability Engineering and System Safety*, vol.91, no. 12, pp.1576–1585, 2006, <http://dx.doi.org/10.1016/j.res.2006.01.011>
- [5] Brall A., "Human Reliability Issues in Medical Care – A Customer Viewpoint," in *Proc. on Annual Reliability and Maintainability Symposium*, USA, pp. 46 – 50, 2006, <http://dx.doi.org/10.1109/RAMS.2006.1677348>
- [6] Catelani M., Ciani L., Risaliti C., "Risk assessment in the use of medical devices: a proposal to evaluate the impact of the human factor," in *Proc. of IEEE International Symposium on Medical Measurements and Applications*, 2014, <http://dx.doi.org/10.1109/MeMeA.2014.6860088>
- [7] Dhillon B.S., *Human Reliability and Error in Medicine*, World Scientific, 2003
- [8] Deeter J., Rantanen E., "Human Reliability Analysis in Healthcare," in *Proc. of 2012 Symposium on Human Factor and Ergonomics in Health Care*, USA, pp.45-51, 2012, <http://dx.doi.org/10.1518/HCS-2012.945289401.008>

- [9] Lyons M., Adams S., Woloshynowych M. and Vincent C., "Human reliability analysis in healthcare: A review of techniques," *Int. J. of Risk and Safety in Medicine*, vol. 16, no. 4, pp. 223–237, 2004
- [10] Zaitseva, E., Kvassay, M., Levashenko, V., Kostolny, "New Methods for the Reliability Analysis of Healthcare System Based on Application of Multi-State System," in *Applications of Computational Intelligence in Biomedical Technology*, Eds.: Bris R., Majernik J., K.Pancerz, E.Zaitseva, Springer, pp. 229-251, 2016, http://dx.doi.org/10.1007/978-3-319-19147-8_14
- [11] J. R. Quinlan, "Simplifying decision trees", *Int. J. Man-Machine Studies*, vol. 27, pp. 221-234, 1987, [http://dx.doi.org/10.1016/S0020-7373\(87\)80053-6](http://dx.doi.org/10.1016/S0020-7373(87)80053-6)
- [12] S. Mitra, K. M. Konwar, and S. K. Pal, "Fuzzy Decision Tree, Linguistic Rules and Fuzzy Knowledge-Based Network: Generation and Evaluation". *Trans. on Syst., Man Cybernetics — Part C: Applications and Reviews*, vol. 32, pp.328-339, 2002, <http://dx.doi.org/10.1109/TSMCC.2002.806060>
- [13] Aven T., Zio E., Baraldi P., and Flage R., *Uncertainty in Risk Assessment: The Representation and Treatment of Uncertainties by Probabilistic and Non-Probabilistic Methods*, Wiley, 200 p., 2014, <http://dx.doi.org/10.1002/9781118763032.ch07>
- [14] Tsiouras M. G., Exarchos T. P., and Fotiadis D. I., "A methodology for automated fuzzy model generation," *Fuzzy Sets and Systems*, vol. 159, no 23, pp. 3201-3220, 2008, <http://dx.doi.org/10.1016/j.fss.2008.04.004>
- [15] B. Natvig, *Multistate Systems Reliability Theory with Applications*, Wiley, New York, 2011, <http://dx.doi.org/10.1002/9780470977088>
- [16] E. Zio, "Reliability engineering: Old problems and new challenges", *Reliability Engineering and System Safety*, vol. 94, pp.125–141, 2009, <http://dx.doi.org/10.1016/j.ress.2008.06.002>
- [17] T. Aven, and B. Heide, "On performance measures for multistate monotone system", *Reliability Engineering and System Safety*, vol. 41, pp.259–266, 1993.
- [18] A. Lisnianski, and G. Levitin, *Multi-state System Reliability. Assessment, Optimization and Applications*. Singapore, SG: World Scientific, 2003.
- [19] H.-Z.Huang, "Structural reliability analysis using fuzzy sets theory", *Eksploatacja i Niezawodność – Maintenance and Reliability*, vol.14, no. 4, pp.284–294, 2012.
- [20] D. Ley, "Approximating process knowledge and process thinking: Acquiring workflow data by domain experts," in *Proc. 2011 IEEE International Conference on Systems, Man, and Cybernetics*, pp. 3274–3279, 2011, <http://dx.doi.org/10.1109/ICSMC.2011.6084174>
- [21] D. Yu, W. S. Park, "Combination and evaluation of expert opinions characterized in terms of fuzzy probabilities", *Annals of Nuclear Energy*, vol. 27, no 8, pp. 713–726, 2000, [http://dx.doi.org/10.1016/S0306-4549\(00\)82012-5](http://dx.doi.org/10.1016/S0306-4549(00)82012-5)
- [22] Y. Wang, and L. Li, "Effects of Uncertainty in Both Component Reliability and Load Demand on Multistate System Reliability," *IEEE Trans. on System, Man and Cybernetic – Part A: Systems and Humans*, vol. 42, no. 4, pp. 958-969, 2012, <http://dx.doi.org/10.1109/TSMCA.2011.2181163>
- [23] N. A. Stanton, and C. Baber, "Error by design: methods for predicting device usability," *Design Studies*, vol. 23, no. 4, pp.363-384, 2012, [http://dx.doi.org/10.1016/S0142-694X\(01\)00032-1](http://dx.doi.org/10.1016/S0142-694X(01)00032-1)
- [24] V. Levashenko, and E. Zaitseva, "Usage of New Information Estimations for Induction of Fuzzy Decision Trees," in *Lecture Notes in Computer Science LCNS2412*, Springer-Verlag, pp. 493-499, 2002, http://dx.doi.org/10.1007/3-540-45675-9_74
- [25] V. Levashenko, E. Zaitseva, and S. Puuronen, "Fuzzy Classified based on Fuzzy Decision Tree," in *Proc. EUROCON, Warsaw, Poland*, pp.823–827, 2007, <http://dx.doi.org/10.1109/91.227387>
- [26] R. Krishnapuram, and J. Keller, "A possibilistic approach to clustering," *IEEE Trans. on Fuzzy Systems*, vol.1, pp.98–110, 1993,
- [27] R. Kruse, Ch. Doring, and M.-J. Lesot, "Fundamentals of Fuzzy Clustering," in *Advances in Fuzzy Clustering and its Applications*, eds. J.Valente de Oliveira and W. Pedrycz, Wiley, 434 p., 2007, <http://dx.doi.org/10.1002/9780470061190.ch1>
- [28] H. Tanaka, L.T. Fan, F.S. Lai, and K. Toguchi, "Fault-tree analysis by fuzzy probability," *IEEE Trans. on Reliability*, vol. 32, pp. 453-457, 1983, [http://dx.doi.org/10.1016/S0019-9958\(65\)90241-X](http://dx.doi.org/10.1016/S0019-9958(65)90241-X)
- [29] H. R. Patel, A. Linares and J. V. Joseph., "Robotic and laparoscopic surgery: cost and training," *Surg Oncol*, vol.18, no.3, pp. 242-246, 2009, <http://dx.doi.org/10.1016/j.suronc.2009.02.007>
- [30] M. Kvassay, E. Zaitseva, and V. Levashenko, "Minimal cut sets and direct partial logic derivatives in reliability analysis," in *Proc. ESREL, Wroclaw, Poland*, pp. 241-248, 2014, <http://dx.doi.org/10.1201/b17399-37>