Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2013) 21: 897 – 912

doi:10.3906/elk-1108-22

Cancer risk analysis by fuzzy logic approach and performance status of the model

Atınç YILMAZ,[1,*] **Kürşat AYAN**[2]
[1]Department of Computer Engineering, Haliç University, İstanbul, Turkey
[2]Department of Computer Engineering, Sakarya University, Sakarya, Turkey

Received: 08.08.2011 • **Accepted:** 14.12.2011 • **Published Online:** 03.05.2013 • **Printed:** 27.05.2013

Abstract: Cancer is the leading life-threatening disease for people in today's world. Although cancer formation is different for each type of cancer, it has been determined by studies and research that stress also triggers cancer types. Early precaution is very important for people who have not fallen ill yet with a disease like cancer that has a high mortality rate and expensive treatment. With this study, we expound that the possibility of developing such disease may be decreased and people could take measures against it. For the 3 cancer types selected as pilot work by introducing a fuzzy logic model, the risks for acquiring these cancer types and preliminary diagnosis for the person to remove these risks are presented. After calculating the risk outcome, the effect of stress on cancer is discussed and determined. Within the study, a fuzzy logic technique that can easily be adapted to other industry studies, as well, is applied to the health industry and effective software for application is developed. Due to this type of study, people will have the chance to take measures against developing cancer and the rate of suffering from cancer may be decreased. Furthermore, the performance status of the new technique is revealed by calculating performance measurements by the outcomes of the models developed by the new type of fuzzy logic technique for 3 cancer types selected as a pilot in the Mamdani type of fuzzy logic model.

Key words: Fuzzy logic, artificial intelligence, cancer, risk analysis, preliminary diagnosis, soft computing

1. Introduction

Since the resolution of real life events is very complicated, the definition of these cases by distinct equations seems not to be possible. As a result of this, the preference for methods outside of certainty is to be soluble even if they are approximate. They will be accurate for the solution of many problems. Since mathematical equations cannot portray all natural events in a certain way, solutions in such studies are approximate to an extent. In engineering, all theories and equations express the case in an approximate manner. Events and mechanisms analyzed in a fuzzy system can be modeled by considering the related verbal and substantially uncertain information as complementary to the events, instead of equations that are accepted only in implications of certain rules and that are assumptions. Verbal human data in engineering systems along with numerical information formulated within a system must be considered.

The fuzzy logic approach gives machines the ability of processing special data of humans and of working by benefiting from their experiences and foresights. While bringing in this ability, it uses symbolical expressions instead of numerical expressions. The transfer of these symbolic expressions to the machines is based on a mathematical basis. This mathematical basis is the fuzzy logic sets theory and fuzzy logic is based on this [1]. The basis of a fuzzy logic controller is this kind of verbal expression and the logical relationships between them.

*Correspondence: atincyilmaz@halic.edu.tr

During the application of the fuzzy logic controller, the mathematical modeling of the system is not essential. The transfer of verbal expressions to the computer is based on a mathematical basis. This mathematical basis is named fuzzy sets theory and fuzzy logic. Fuzzy logic expresses multilevel procedures in the [0,1] range, unlike the 2 levels of (0,1) as in classical logic. Fuzzy logic has the ability to conduct procedures according to information that is not fully known or entered incompletely [2].

As a result of limited medical resources and ineffective usage of existing resources, every year hundreds of thousands of people in the world lose their lives due to certain diseases. Usage of numerical systems in medicine and health systems may reduce the loss of patients [3]. Mathematical models may be used almost everywhere that a decision-making problem exists.

Cancer is a genetic disease, formed as a result of growth and proliferation of cells in an uncontrolled or abnormal manner due to cells' exiting from DNA damage, and it is the leading life-threatening disease for humans in today's world. Although cancer formation is different for each type of cancer, it has been determined by studies and research conducted that stress also triggers cancer types. Stress is a bodily constraint, coming from the physical and social environment, not causing the disease directly but leading to bodily and psychological diseases due to its reduction of resistance in the human body. It is suggested that psychological stresses particularly pressure the immune system by reducing T lymphocytes. This reduction in the response of the immune system increases the frequency of infectious diseases and cancer. Stress leads to the settlement of cancerogenic cells and to their spread in the entire body by disrupting the immune system of the body. Animal studies have shown such a relationship between stress and cancer [4].

Fuzzy logic plays an important role in the field of medicine and was investigated in many medical applications. Some of the applications of fuzzy logic in medicine are the diagnosis of breast cancer, lung cancer, or prostate cancer [5,6].

Early preventive measures are very important for people who have not become sick yet in order to avoid a high mortality rate and expensive treatment in diseases such as cancer. With this type of study, the chance of developing such disease may decrease and people can take measures. To the extent that the cancer is diagnosed earlier, its treatment is that much more successful. If medicine can use techniques such as fuzzy logic from artificial intelligence methods in its own fields, in the future, many diseases such as cancer may reach a treatable level or may be prevented by early diagnosis. Thus, expensive treatments or surgeries may not even be required. Today, most of the people developing cancer apply to hospitals at advanced stages of the disease and therefore are diagnosed late. As a result of this, treatments are useless most of the time and the patient dies in a short time. Future-oriented diagnosis of cancer in healthy people is one of the most important issues that should be emphasized [5].

The purpose of the study is to determine the risks of developing types of cancers in the future for healthy people and to preliminarily diagnose it by specifying pilot cancer types to work on. Based on this, firstly, similar studies conducted by using artificial intelligence and the fuzzy logic model in medicine and in the subject of cancer are reviewed, implications are made, and incomplete issues are determined. Consequently, the cancer types to be studied will be determined and the factors influencing this cancer type will be investigated. After implications are made, these factors are used in the model formed by the composition of a fuzzy logic model. In this respect, breast cancer, lung cancer, and colon cancer are selected as pilot cancer types. The reason for selection of these cancer types is the frequency of patient numbers and the appropriateness of this kind of study for the indicated cancer types. The risks of developing breast, lung, and colon cancers within the study are revealed by using a fuzzy logic model and the opportunity to offer suggestions to the person to remove this risk is provided.

In this study, the purpose is to investigate the usability of the fuzzy logic model in the health field in the light of data held, to assess the performance difference of the proposed fuzzy logic method with respect to the Mamdani method, and to evaluate and share the results obtained. In our research, as being different from others studies conducted in the literature, the aim is not only to see the difference of performance outcomes of the prepared application, produce statistical data, and determine the risk factors affecting breast, lung, and colon cancers, but also to develop an application to work in every computer system loaded with .NET framework that can be used by doctors or potential patients for people suspected to have breast, lung, and colon cancer. Besides this, another objective is to introduce a fuzzy logic method that can produce more successful results. Moreover, it is aimed that the introduced fuzzy logic model determined for breast, lung, and colon cancers as pilot cancer types be composed of findings known without any testing or expert opinions. As a result of this, without any analysis or expert input, a person can calculate the risk status for any of the 3 cancer types conveniently with the help of software in any computer loaded with the .NET framework. We also try to construe the effect of stress, as a subject having a triggering role in every kind of disease, on cancer types within the software, differently than in other studies.

2. Fuzzy logic

Fuzzy logic aims at modeling human thinking and reasoning and at applying the model to problems according to needs. It tries to equip computers with the ability to process special data of humans and to work by making use of their experiences and insights. When human logic solves problems, it creates verbal rules such as "if <event realized> is this, the <result> is that". Fuzzy logic tries to adapt these verbal rules and the human ability to make decisions to machines/computers. It uses verbal variables and terms together with verbal rules [7].

Verbal rules and terms used in the human decision-making process are fuzzy rather than precise. Adapting human logic system to computers/machines will increase the problem-solving ability of computers/machines. Verbal terms and variables are expressed mathematically as membership degrees and membership functions. Fuzzy decision-making mechanisms use symbolic verbal phrases instead of numeric values. Transferring these symbolic verbal phrases to computers is based on mathematics. This mathematical basis is fuzzy logic.

Systems that use fuzzy logic are alternatives to the difficulty of mathematical modeling of complex nonlinear problems and fuzzy logic meets the mathematical modeling requirements of a system.

Systems that use fuzzy logic can produce effective results based on indefinite verbal knowledge, like humans. In fuzzy logic, information is verbal phrases, such as 'big', 'small', 'very', or 'few', instead of numeric values. If a system's behavior can be expressed by rules or requires very complex nonlinear processes, the fuzzy logic approach can be applied.

2.1. Fuzzy clusters

The fuzzy cluster concept is an extension of a classical cluster. In a classical cluster, an element is either within a cluster (1) or not within a cluster (0) (Figure 1). In fuzzy clusters, an element has any membership value between 0 and 1 [8].

In classical clusters, "1" represents being a member while "0" represents not being a member. In fuzzy clusters, "1" represents full membership (full membership degree), degrees between "0" and "1" represent degrees of membership, and "0" represents full nonmembership (full nonmembership degree) [9].

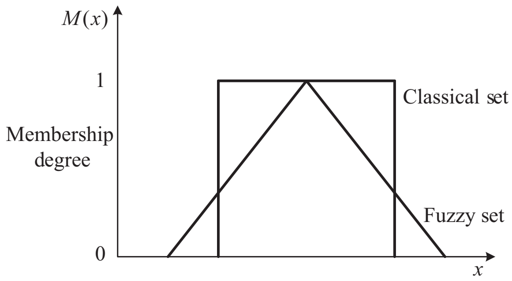

Figure 1. Indication of a classical and fuzzy cluster on coordinate system.

2.2. Linguistic variables

While variables in mathematics usually take numerical values, in fuzzy logic applications, the nonnumeric linguistic variables are often used to facilitate the expression of rules and facts.

A linguistic variable such as age may have a value such as 'young' or 'old'. However, the great utility of linguistic variables is that they can be modified via linguistic hedges applied to primary terms. The linguistic hedges can be associated with certain functions. For example, Zadeh proposed to take the square of the membership function. This model, however, does not work properly [10].

2.3. Usage of fuzzy logic in medicine and similar studies

Most of the concepts used in medicine are fuzzy. The fuzzy logic method is convenient for medical applications due to the uncertain nature of medical concepts and of the relationships between these concepts. Uncertain medical cases may be defined by fuzzy sets. Fuzzy logic suggests methods of solution production that have the ability of approximate drawing of conclusions [11]. Due to the complexity of the practice in medicine, the traditional quantitative analysis approaches are not appropriate. The sources of uncertainty can be classified as follows [6]:

- Presence of insufficient information about the patient.
- Most of the time, the patient's medical history is provided by the patient himself/herself and/or by his/her family. To a large extent, this information is generally subjective and uncertain.
- Health examination. Most of the time, physicians obtain objective data; however, in some cases, the border between normal and pathological cases is not clear-cut.
- Test results related to the laboratory and other diagnoses may also be subject to some errors and even to the patient's misconduct prior to examination.
- There might be symptoms that are faked, exaggerated, or shown as less than they really are. Patients may neglect to talk about some of the symptoms.

When studies are analyzed in detail, it is observed that artificial intelligence techniques for health sciences are applied to a large extent to diagnosis and identification. The same case also applies for cancer. For methods used in the studies reviewed, introductions and clinical findings are emphasized; findings that cannot be known without analysis and the taking of expert opinions are emphasized in the model as an introduction. Ready-made tools that are held for artificial intelligence by application software such as FuzzyTech or MATLAB were also used in the studies reviewed. Moreover, a proposed fuzzy logic model has not been developed within these studies; the Mamdani type of fuzzy logic model or techniques such as multilayer artificial nerve web were previously used in the solution of the problems.

2.4. Why use fuzzy logic?

The fuzzy logic method was preferred to current mathematical models due to its ability for modeling the obscurity in the related problem, ability to work with lower cost and easy application, higher mechanical intelligence level, ability to solve new problems by using the experience of the model within the framework of the rules defined in the model, flexible structure, compatibility for solving the insufficiently defined problems, and use of intuitive methods instead of a specific algorithm.

This study makes risk analysis for taking measures, unlike in diagnosis. Because of that, the reasons for choosing fuzzy logic in contrast to other systems are grounded on the lack of a flexible structure of other systems, operation for accurate data, and production of accurate results. Here is a list of general observations about fuzzy logic:

- Fuzzy logic is conceptually easy to understand and is a more intuitive approach. The mathematical concepts behind fuzzy reasoning are very simple.
- Fuzzy logic is flexible.
- Fuzzy logic is tolerant of imprecise data.
- Fuzzy logic can model nonlinear functions of arbitrary complexity.
- Fuzzy logic can be built on top of the experience of experts.
- Fuzzy logic is based on natural language. Because fuzzy logic is built on the structures of qualitative description used in everyday language, fuzzy logic is easy to use.

The last statement is perhaps the most important one and deserves more discussion. Natural language, which is used by ordinary people on a daily basis, has been shaped by thousands of years of human history to be convenient and efficient. Sentences written in ordinary language represent a triumph of efficient communication.

3. Proposed fuzzy logic method and application

3.1. Proposed method – fuzzy logic model

First, the Mamdani type of fuzzy logic model is developed for cancer types specified as a pilot within this study and used within the application software. Performance measurements of the Mamdani type fuzzy logic model were made for 3 cancer types by using various model cases. Since the results produced by the model will be very important for industries such as health where even the smallest detail has great importance, the need for introducing a fuzzy logic model with a higher performance arose. In this respect, the proposed fuzzy logic model approach is introduced by making modifications to the Mamdani type fuzzy logic model and by introducing new methods.

3.1.1. Rating formulas and fuzzification

Generally, in practice, making the change ranges appearing in the classical set form fuzzy is required for fuzzy set, logic, and system procedures [12,13]. For this, it is considered that all of the elements that may be present in a range have various values between 0 and 1, instead of having membership degree equal to 1. In this case, it is accepted that some elements include uncertainty. In the case that these uncertainties from nonnumerical cases arise, fuzziness is mentioned. Convenience of fuzzy sets depends on the skill of being able to form membership degree functions appropriate for different concepts. The most frequently used functions are triangle and trapezoid for ease. The display of elements pertaining to any fuzzy set by triangle membership function and trapezoid membership function on the new type fuzzy logic approach are displayed in Figures 2 and 3, respectively.

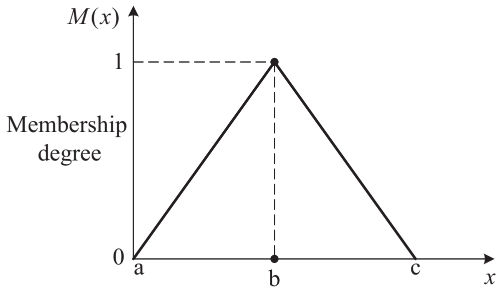

Figure 2. Triangle membership type and membership degree calculation.

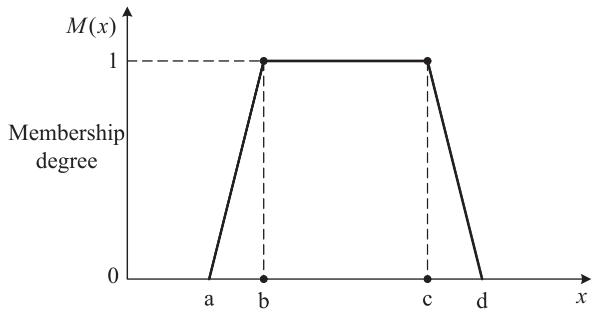

Figure 3. Trapezoid membership type and membership degree calculation.

When triangle membership function is used, fuzzification is done according to the variable's value. First of all, the variable's maximum probable value is taken and this value's trigonometric functions and Euclidian relation are formulated for the fuzzification. For the trapezoid membership function, the trapezoid splits into triangle shapes and the same formulations are applied. The fuzzification formulations of the triangle membership function and trapezoid membership function are shown in Eqs. (1) and (2), respectively.

$$
\begin{aligned}
&\text{if } a > x || c < x \Rightarrow 0\\
&\text{if } a \leq x \leq b \Rightarrow \tfrac{\max' \times [\sin(x) - \sin(a)]}{[\sin(b) - \sin(a)]}\\
&\text{if } b \leq x \leq c \Rightarrow \tfrac{\max' \times [\sin(c) - \sin(x)]}{[\sin(c) - \sin(b)]}
\end{aligned} \tag{1}
$$

$$
\begin{aligned}
&\text{if } a > x || d < x \Rightarrow 0\\
&\text{if } a \leq x \leq b \Rightarrow \tfrac{\max' \times [\sin(x) - \sin(a)]}{[\sin(b) - \sin(a)]}\\
&\text{if } b \leq x \leq c \Rightarrow 1\\
&\text{if } c \leq x \leq d \Rightarrow \tfrac{\max' \times [\sin(d) - \sin(x)]}{[\sin(d) - \sin(c)]}
\end{aligned} \tag{2}
$$

In fuzzification formulations, for the symbolized values, "x" fuzzy value reveals "max", while the entry states the maximum fuzzy set's maximum value and "a - b - c" states the membership function's threshold value.

Besides functions in the form of triangles or trapezoids being used frequently or being in other appropriate forms, subsets are required to be in forms that are overlapping with each other.

3.1.2. Rule processing unit

In fuzzy logic, rules are formulated by conditional cases in the form of 'if ... then, ... let it be'. All input variables are converted into verbal variable values, the obtained fuzzy result stage is applied based on rules for current status, and values of verbal variables are calculated at exit. On the other hand, a fuzzy rule should have verbal input and output terms in the form of 'if ... then, ... let it be' (for example, if X value is A, then let Y value be B). The 'if ...' section is called the status; the '... let it be' section is called the result or decision section. In the example of 'if X value is A, then let Y value be B', A and B are verbal words and they indicate to which status X and Y values pertain in fuzzy sets X and Y. As rules are processed in order, the result found is processed to exits indicated by the following formulas and rules for the new type of fuzzy logic approach within the rules related to entry values are made fuzzy themselves [14].

The rule processing unit formulation is shown in Eq. (3). In the formulas of the rule processing unit, "t_1, t_2, t_3" symbolizes the results to achieve "tresult , tresult", the results of the rule processing unit; x_n are the input values to calculate the fuzzy value; and n is the output set number that generates results.

$$\begin{aligned} t_1 &= \tfrac{x_1+x_2+....+x_n}{n} \quad & t_2 &= (x_1-t_1)^2+(x_2-t_1)^2+...+(x_n-t_1)^2 \\ t_3 &= \tfrac{t_2}{n} & t_{result} &= \sqrt{t_3} \end{aligned} \tag{3}$$

If more than one value exists in any of the output values for related rules, the greatest value within these values is selected.

3.1.3. Defuzzification

In practical applications, especially in engineering plans, projects, and designs, definite numerical values are required for sizing. The implications of the fuzzy variable, set, logic, and systems in artificial intelligence studies that might be fuzzy should be converted into definite numbers. All of the procedures made for conversion of fuzzy information into definite results are named defuzzification procedures [15,16].

The defuzzification process will be done by applying the following formula in the new type of fuzzy logic approach by using peak values of the related output set produced as results and output values calculated within the rules.

Defuzzification formulation is shown in Eq. (4). In the formulas of the defuzzification process, α_{ort} symbolizes the harmonic averages for defuzzification, n is the output set number that generates results, $\max_n$ is the highest value of the sets, top_n is the point where the relevant set gets to the peak point, and "result" is the outcome of the defuzzification process.

$$\begin{aligned} \alpha_{ort} &= \tfrac{n}{\frac{1}{\max_1}+\frac{1}{\max_2}+...+\frac{1}{\max_n}} \\ result &= \tfrac{(top_1\times\alpha_{ort})+(top_2\times\alpha_{ort})+....+(top_n\times\alpha_{ort})}{\max_1+\max_2+..+\max_n} \end{aligned} \tag{4}$$

3.2. Dataset

A total of 120 datasets for breast cancer, 140 datasets for lung cancer, and 110 datasets for colon cancer were provided from Oncology Services for testing the performance measurement of this presented fuzzy logic model (Figure 4). Sixty-three of the breast cancer datasets used included people who were diagnosed with breast cancer. The other 57 datasets included people who went to the hospital for examination but were not diagnosed with breast cancer. Eighty-six of the lung cancer datasets used included people who were diagnosed with lung cancer. The other 54 datasets included people who went to the hospital for examination but were not diagnosed with lung cancer. Sixty-nine of the colon cancer datasets used included people who were diagnosed with colon cancer. The other 42 datasets included people who went to the hospital for examination but were not diagnosed

	Sex	Age	Age of menarche	Age of menopause	Age of first childbearing	Alcohol consumption	Nutritional habits	Genetic relationship
1	Female	33	14	None	27	Once a week	Low-fat	None
2	Female	42	15	None	32	Twice a week	High-fat	Nephew
3	Male	52	None	None	None	Once in 15 days	High-fat	Aunt
4	Female	25	None	None	None	Once in 3 days	Low-fat	Other
5	Male	37	None	None	None	Once in 2 days	High-fat	Mother
6	Female	31	15	None	19	None	Low-fat	None
7	Female	56	14	45	22	Once in 15 days	Low-fat	None
8	Female	63	13	43	21	None	High-fat	Mother
9	Female	39	14	None	None	None	Low-fat	Unknown
10	Male	42	None	None	None	Once in 3 days	High-fat	Unknown

a) Sample breast cancer dataset

	Age	Genetic relationship	Cancer history	Inflammatory status	Physical activity	Weight (kg)/height (cm)	Cigarette consumption	Alcohol consumption	Nutritional habit
1	28	None	None	None	Regular	75/180	None	None	Well-nourished
2	45	Cousin	None	None	Low	75/190	1-2 cigarettes	Once a week	Undernourished
3	38	Uncle	Uterine	Long ago/recuperated	Regular	86/180	Half a package	One in 3 days	Undernourished
4	53	None	Colon	Short time ago/recuperated	Regular	95/190	1-2 cigarettes	Twice a week	Undernourished
5	45	None	None	None	Regular	87/169	1 package	One in 3 days	Well-nourished
6	44	Unknown	None	Long ago/recuperated	Regular	80/182	None	Once in 15 days	Well-nourished
7	50	None	None	None	Regular	80/180	1-2 cigarettes	None	Well-nourished
8	60	Grandfather	None	None	Low	70/160	Half a package	None	Undernourished
9	39	Other	Breast	None	Low	90/175	One package	Once a week	Well-nourished
10	52	None	None	Long ago/recuperated	None	89/171	Half a package	Once in 15 days	Well-nourished

b) Sample colon cancer dataset

	Sex	Age	Skin color	Cigarette consumption	Age of smoking initiation	Exposure to second-hand smoke	Occupational group	Habitat	Genetic relationship	Monthly income (TL)	Nutritional habits
1	Male	53	Swarthy	Half a package	25	Only at home	Worker	City center	None	2100	Undernourished
2	Male	36	Dark brown skin	None	None	A circle of friends	White collar worker	Inner-city building complex	None	2500	Well-nourished
3	Female	34	White skin	1-2 cigarettes	30	A circle of friends	White collar worker	Inner-city building complex	None	3000	Undernourished
4	Male	29	Swarthy	Half a package	20	Work environment – a circle of friends	White collar worker	Home with garden	None	1400	Well-nourished
5	Male	41	Light brown skin	One package	16	Always	Worker	City center	Unknown	1000	Well-nourished
6	Female	40	Dark brown skin	1-2 cigarettes	30	A circle of friends	White collar worker	City center	None	1200	Well-nourished
7	Male	51	Swarthy	More than 2 packages	15	Only at home	Worker	City center	Cousin	900	Undernourished
8	Female	47	Swarthy	None	None	A circle of friends	White collar worker	Chalet	Unknown	2000	Well-nourished
9	Female	30	Dark brown skin	1-2 cigarettes	24	A circle of friends	Worker	Flat with sea view	None	650	Well-nourished
10	Male	55	Swarthy	More than 2 packages	18	Work environment – a circle of friends	Field hand	Flat with sea view	None	3500	Well-nourished

c) Sample lung cancer dataset

	Cancer type	The result of risk	Stress durability	Tendency to stress
1	Breast	Low risk	Durable	Low tendency
2	Breast	Risk	Not durable	Tendency
3	Breast	Healthy	Highly durable	Low tendency
4	Breast	Risk	Highly durable	Low tendency
5	Lung	Risk	Highly durable	Low tendency
6	Lung	Low risk	Highly durable	High tendency
7	Lung	Healthy	Highly durable	High tendency
8	Colon	Healthy	Not durable	Low tendency
9	Colon	Risky	Tendency	High tendency
10	Colon	Low risk	Highly durable	Durable

d) Sample stress dataset

Figure 4. Sample dataset.

with colon cancer. The model's success status was presented by using both healthy and patient datasets. Thirty datasets were provided for calculating stress' trigger status on breast cancer. Nineteen of the used datasets included people who were diagnosed with breast cancer and the other 11 datasets included people who went to the hospital for examination but were not diagnosed with breast cancer. Thirty datasets were provided for calculating stress' trigger status on lung cancer. Sixteen of the datasets used included people who were diagnosed with lung cancer and the other 14 datasets included people who went to the hospital for examination but were not diagnosed with lung cancer. Thirty datasets were provided for calculating stress' trigger status on colon cancer. Eighteen of the datasets used included people who were diagnosed with colon cancer and the other 12 datasets included people who went to the hospital for examination but were not diagnosed with colon cancer.

3.3. Cancer risk analysis application

For diseases like cancer, taking preventive measures before the initiation of the disease, learning about risk status, and preliminary diagnosis for diseases with very difficult treatments and recoveries are important issues. In consideration of this case, in this study, a software that will measure the susceptibility for that cancer type and risk status for specific cancer types in healthy people or people not diagnosed with the disease is developed by selecting a fuzzy logic model from artificial intelligence techniques. In this respect, breast, lung, and colon cancers are selected as pilot cancer types. The reason for selecting the indicated cancer types is the relevancy of the factors leading to these diseases for this type of study and their substantially high incidences in the world.

Fuzzy logic models previously used were reviewed, and then solutions were provided for cancer types related to Mamdani type fuzzy logic model. After performance measurements of the models were made for breast, lung, and colon cancer diseases, modifications are made on the Mamdani type of fuzzy logic model and a new method is introduced. The proposed fuzzy logic method was used for breast, lung, and colon cancer diseases, and performance measurements were taken. The newly formed fuzzy logic method produced better results compared to the Mamdani type fuzzy logic method for the 3 cancer models. The software structure was composed by combining the advantages of the programming techniques, oriented to the object, through the C# programming language on the Visual Studio .Net 2010 platform. Within the software, 5 different visually based software programs are actualized. The first of these forms is the section where the cancer type is selected to calculate risk analysis. According to the selection made, the 2nd, 3rd, or 4th forms will be opened. The second application software calculates the risk status of the person for breast cancer by the new type of fuzzy logic method; the third application software calculates the risk status of the person for lung cancer by the new type fuzzy logic method; the fourth application software calculates the risk status of the person for colon cancer by the new type fuzzy logic method. The fifth form calculates the risk status that will be formed by the triggering of the cancer types by the stress factor based on risk results calculated for breast, lung, or colon cancers by the new type of fuzzy logic method (Figure 5).

After expert opinions about this subject of risk factors for breast cancer, lung cancer, and colon cancer diseases were obtained and literature studies were conducted, fuzzy logic models were developed for fuzzy logic cancer types. In the breast cancer model, sex, age, genetic status, menarche age, menopause age, first childbirth age, alcohol consumption, and nutritional habits were determined as factors for cancer risk [17,18]. In the lung cancer model, sex, age, skin tone, smoking, age of starting smoking, passive smoking environment, occupational status, living environment, genetic status, economic status, and nutritional habits were determined as factors for cancer risk [19]. Lastly, in the colon cancer model, age, genetic status, cancer history, inflammation status in the intestines, physical activity status, weight status, smoking, alcohol consumption, and nutritional habits were determined as factors for cancer risk [20–22].

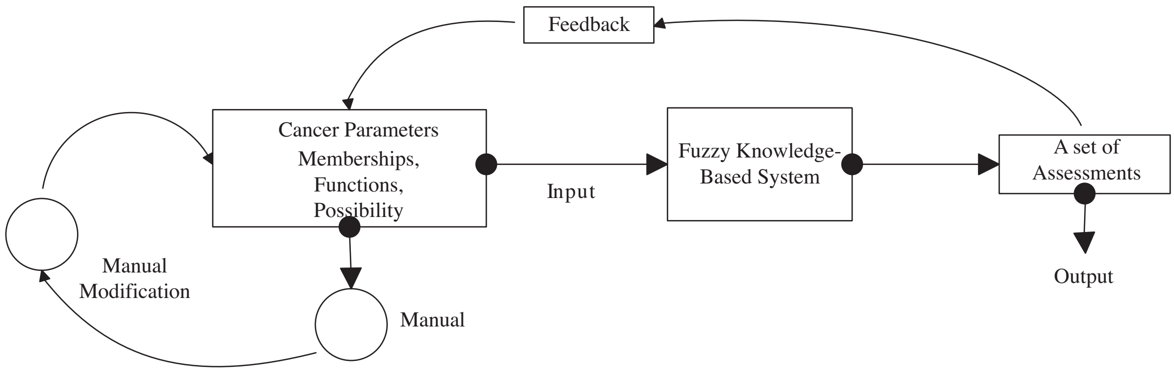

Figure 5. General architecture of the fuzzy model.

The membership degrees of all factors in the 3 models were determined. As an output result of data received from these factors, the risk status of the person for this cancer type has been analyzed within the model (excessive risk, risk, slight risk, and healthy) (Figure 6).

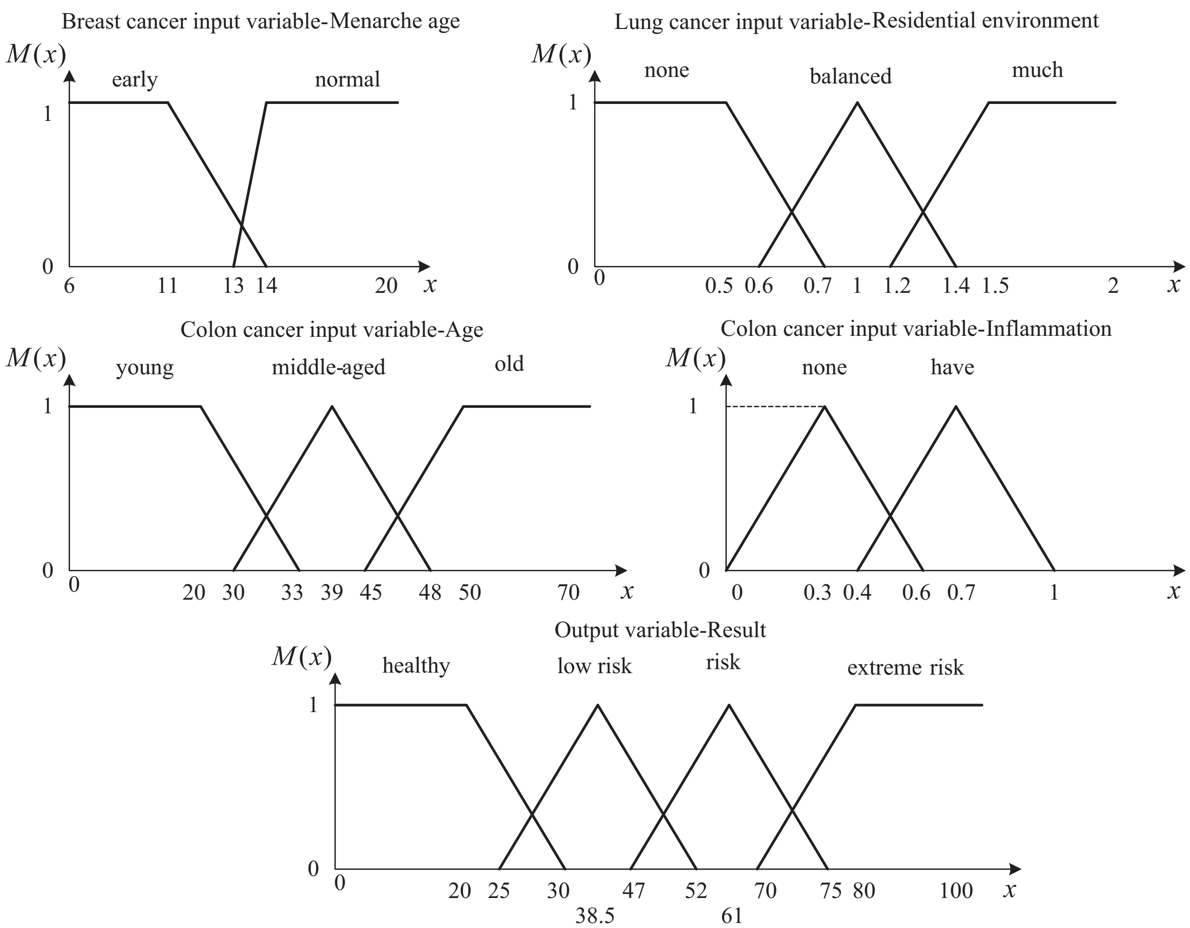

Figure 6. Examples of membership degrees.

After determination of membership functions and membership degrees for the fuzzy logic model formed for breast cancer, lung cancer, and colon cancer, rules of the fuzzy logic model were determined in light of the data held and expert opinions. A total of 115 rules for breast cancer, 180 rules for lung cancer, and 152 rules formed for colon cancer were applied, data were tested, and the best results for the fuzzy logic model were reached. In Table 1, some sample rules are presented for the 3 cancer types.

After completion of the design of fuzzy logic models, membership degrees, and setting rules, cancer analysis software is developed by the new type of fuzzy logic approach. Interfaces of software developed exist separately for breast cancer, lung cancer, and colon cancer. After entries in the software are completed by the user, by applying Eq. (1) or Eq. (2) to every entry value receiving a value within the software, the fuzziness of the values entered is realized. After entries of all fields related to the person in the software are completed and the "calculate the risk outcome" button is pressed, membership degrees for all entries are calculated one by one with the proposed fuzzy logic method and all of the rules formed for the model are checked in order. In cases included by related rules, calculation-based related output is made by the proposed fuzzy logic method. As a result of rules, the risk status of the person as pertaining to Healthy, Slight Risk, Risk, or Excessive Risk groups and calculation of these membership degrees are assigned by Eq. (3). After application of all rules, in order to calculate the outcome of the risk status of the person, by conversion of values produced as a result of rules into formulas of 14 different probabilities separately, the purification value of the values within the Healthy, Slight Risk, Risk, and Excessive Risk groups is calculated by applying Eq. (4). A person's risk status is calculated according to the resulting value; to which group the person pertains and at which value is found.

Table 1. Sample rules.

Sample rules
Age = (Young or Middle-Aged) & Genetic = None & Cancer Developed = None & Inflammation = None & Nutritional Habits = Healthy & Physical Activity = (Regular or Low) & Weight = (Small or Low) & Smoking = None & Alcohol Consumption = Sometimes → Healthy (Colon Cancer)
Age = (Young or Middle-Aged) & Genetic = None & Cancer Developed = None & Inflammation = None & Nutritional Habits = Healthy & Physical Activity = (Regular or Low) & Weight = (Small or Low) & Smoking = Normal & Alcohol Consumption = Sometimes → Low Risk (Colon Cancer)
Age = (Young or Middle-Aged) & Genetic = None & Cancer Developed = None & Inflammation = None & Nutritional Habits = Healthy & Physical Activity = (Regular or Low) & Weight = (Small or Low) & Smoking = Extreme & Alcohol Consumption = (None or Sometimes) → Risky (Colon Cancer)
Sex = Male & Age = Old & Genetic = Far & Alcohol Consumption = None → Low Risk (Breast Cancer)
Sex = Female & Age = Middle-Aged & Genetic = First Degree & Menarche Age = Age Early & Alcohol Consumption = Much → Extremely Risky (Breast Cancer)
Sex = Female & Age = Old & Folk = White & Smoking = None & Passive Smoking Environment = None & Vacation = Desk Job & Residential Environment = None & Genetic = None → Low Risk (Lung Cancer)
Sex = Male & Age = (Young or Middle-Aged) & Folk = White & Smoking = None & Passive Smoking Environment = None & Vacation = Risk Area & Residential Environment = None & Genetic = None & Economic Status = (Poor or Fair) → Risky (Lung Cancer)

It is observed that the result is better than that from the Mamdani method, which is reached by testing our system with data of patients and healthy people held for the related cancer type in the fuzzy logic model software that is formed by the new method and is conducting risk analysis for breast cancer, lung cancer, and colon cancer. First, in 94 of 120 data held for the Mamdani type fuzzy logic model, for the performance of breast cancer analysis of the system accurate results are obtained and performance measurement is ensured at a rate of 78.33%. In 97 out of 120 data held for the new type of fuzzy logic model, accurate results are obtained and the performance measurement is ensured at a rate of 80.83%. In 109 out of 140 data held first in the Mamdani type fuzzy logic model for performance of lung cancer analysis of the system, accurate results are obtained and performance measurement is ensured at a rate of 77.85%. In 112 out of 140 data held for the new type of fuzzy logic, accurate results are obtained and the performance measurement is ensured at a rate of 80%. In 89 out of 110 data held for the Mamdani type of fuzzy logic model for performance of colon cancer analysis of the system, accurate results are obtained and performance measurement is ensured at a rate of 80.80%. In 91 out of 110 data held for the new type of fuzzy logic model, accurate results are obtained and performance measurement is ensured at a rate of 82.72%.

Receiver operating characteristic (ROC) analysis is a method that is used for defining optimal sensibility and specificity probability of a medical test. In order to determine the success of the proposed fuzzy logic model and medical tests, ROC analysis was made for the 3 cancer types and ROC analysis for the Mamdani and proposed fuzzy logic methods results can be seen in Tables 2 and 3, respectively. As a result of the ROC analysis, it was observed that the accuracy rate of the model for all 3 cancer types is better than that of the Mamdani model. It was also found that the prepared medical test shows a successful performance at the end of ROC analysis. The ROC curve for the proposed fuzzy logic model can be seen in Figure 7.

Table 2. ROC analysis for the Mamdani method for cancer types.

	Sensitivity	Specificity	Positive identification	Negative identification	Accuracy
Breast cancer	0.95	0.60	0.72	0.08	0.79
Colon cancer	0.78	0.85	0.90	0.30	0.81
Lung cancer	0.91	0.56	0.76	0.18	0.78

Table 3. ROC analysis for the proposed fuzzy logic method for cancer types.

	Sensitivity	Specificity	Positive identification	Negative identification	Accuracy
Breast cancer	0.97	0.64	0.75	0.05	0.81
Colon cancer	0.80	0.88	0.92	0.28	0.83
Lung cancer	0.93	0.60	0.79	0.15	0.80
Effect of stress on breast cancer	0.85	0.55	0.76	0.33	0.74
Effect of stress on colon cancer	0.89	0.67	0.80	0.2	0.80
Effect of stress on lung cancer	0.88	0.65	0.74	0.18	0.77

In order to calculate the effect of stress on cancer, a stress model is composed by using the fuzzy logic model according to the risk outcome calculated from any of the breast cancer, lung cancer, and colon cancer models. After expert opinions about the effects of stress on breast cancer, lung cancer, and colon cancer were obtained and status was determined by literature studies, the design of the fuzzy logic model was started and the fuzzy logic model was developed for the effect of stress on cancer. In the stress–cancer model, risk outcome is calculated by the model for cancer types, and stress resistance test results and inclination towards stress were determined as factors for cancer risk. Pursuant to these, membership degrees of 3 different factors were determined. As an output result of data received from these 3 factors, risk status of the person for this cancer type has been analyzed within the model (excessive risk, risk, slight risk, and healthy) (Figure 8).

A stress resistance test of 22 questions to measure the stress resistance of the person and a stress inclination test of 26 questions to measure the inclination of the person towards stress were prepared by expert psychologists. Due to the fuzzy logic model formed by a new method for breast cancer, lung cancer, and colon cancer, the risk outcome that is calculated within the software is determined as the final entry value for the stress fuzzy logic model. The purpose hereby is to be able to compare risk outcome determined by the model with the triggering status together with the stress effect. After membership functions and membership degrees were determined for the fuzzy logic model formed for the effect of stress on cancer, the rules of the fuzzy logic model were determined in light of the data held and expert opinion. The best result was reached for the fuzzy logic model by application of 64 rules formed and by testing the data held.

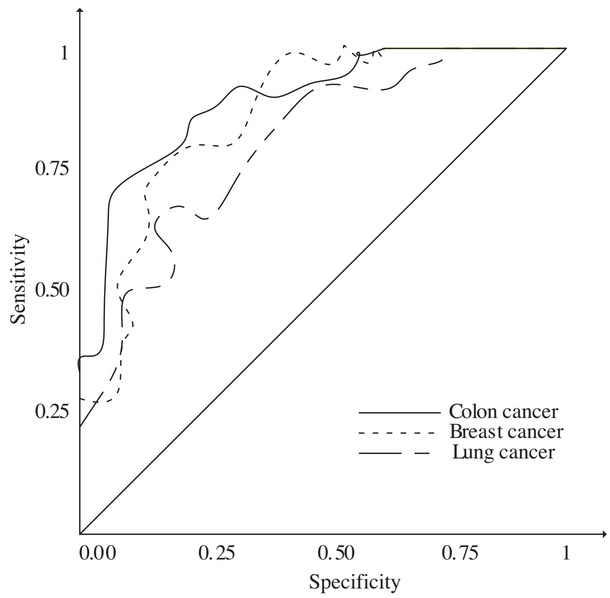

Figure 7. ROC curve for proposed fuzzy logic model.

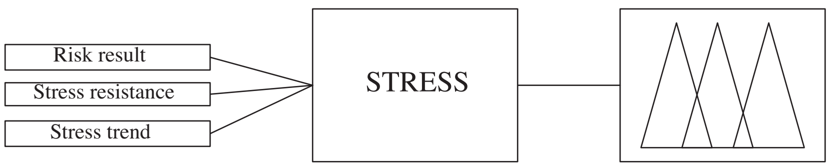

Figure 8. Stress–cancer fuzzy logic model.

After all people answered the stress resistance scale and stress inclination scale questions in the software and pressed the "calculate the stress effect" button, membership degrees for stress resistance and stress inclination entries were calculated by the proposed fuzzy logic method and all of the rules formed for the model were checked in order. In cases included by related rules, calculation oriented to related output is made by the proposed fuzzy logic method. A person's risk status is calculated according to the output resulting value; which group the person belongs to and at which value is found (Healthy, Slight Risk, Risk, Excessive Risk).

4. Results

The below results were analyzed within the study conducted for modeling fuzzy logic cancer risk analysis.

- The proposed fuzzy logic method, which has been developed by modifying the Mamdani method, was put forth. It was observed that the method gave better results when it was compared to the accuracy rate of the Mamdani method for 3 types of cancer (Table 4).

- Statistical analysis of the model tested with the available data was put forth with the ROC method used in biostatistical analysis. The accuracy rate of the ROC analysis is 0.81 for breast cancer, 0.83 for colon cancer, and 0.80 for lung cancer.

- All data used at the introduction phase of the model are findings that are known without any testing of expert opinion. Therefore, with the composed model, people will be able to learn their tendency for the related cancer types without any examination needed. At the same time, the model is also a guide for doctors for preliminary diagnosis.

- In fuzzy cancer models, the influential parameters were defined as sex and genetics for breast cancer; genetics, inflammation, alcohol, and nutrition for colon cancer; and smoking, age of starting smoking, and genetics for lung cancer.

Table 4. Performance measurement for Sugeno method, Mamdani method, and proposed method for cancer types.

Type of cancer	Performance of Sugeno method (%)	Performance of Mamdani method (%)	Performance of proposed fuzzy logic method (%)
Breast cancer	72.50	78.33	80.83
Lung cancer	66.4	77.85	80
Colon cancer	72.72	80.80	82.72
Effect of stress on breast cancer	60	73.33	76.66
Effect of stress on lung cancer	63.33	73.33	76.66
Effect of stress on colon cancer	66.66	76.66	80

5. Discussion

The difference of this study from others is not only the superiority of its performance and statistical differences. In addition to this, it is an application that can be used both by doctors or potential patients, with a user-friendly application that can be used in every computer loaded with the .NET framework, able to calculate risks to prevent breast, lung, and colon cancer. Besides this, a new method that can produce more successful results is introduced. All introductions of fuzzy logic models determined for breast, lung, and colon cancers from cancer types selected as the pilot are composed of findings that are known without any testing or expert opinions. Consequently, without any analysis or expert advice, a person can calculate the risk status for any of the 3 cancer types conveniently with the help of software in any computer. The effect of stress, as a subject having a triggering role in every kind of disease, on cancer types has been construed within the software, unlike in other studies.

A performance report has been extracted by using the Mamdani method from fuzzy logic models, and afterwards the proposed fuzzy logic method was introduced and performance differences by the renewed system were brought up. The reason why breast cancer, lung cancer, and colon cancer were selected as pilot cancer types within the study is the frequency of patient numbers for the indicated cancer types and their conformance for this type of study. The risks of developing breast, lung, and colon cancers were revealed by using the proposed fuzzy logic model within the study and the opportunity to offer suggestions to the person for removal of the risk has been provided. The reason for selecting the fuzzy logic model is the effective drawing of a conclusion of systems where fuzzy decision is used, depending on uncertain linguistic information as human logic can do.

The new obtained method was discovered by modifying the Mamdani method, which is one of the methods of fuzzy logic. While only a process is made with Euclidian relation in the Mamdani method, different trigonometric functions are also used during the defuzzification process in the new method; on the other hand, many other different methods were applied, such as calculation of harmonic average in the rule processing unit and defuzzification. Thus, it is observed that the obtained results are more successful for the system in process.

In order to measure the compatibility and the performance of the study, risk analysis was conducted in the system by using data from patients and healthy people. However, 100% accuracy in the system could not be detected since the risk status of a person having a low risk rate may change in the future with different living conditions and factors.

References

[1] L.A. Zadeh, "Fuzzy sets", Information and Control, Vol. 8, pp. 338–353, 1965.

[2] Ç. Elmas, Artificial Intelligence Applications, Seçkin Publishing, Ankara, 2007.

[3] M.G. Alvarez, "Molecular basis of cancer and clinical applications", Surgical Clinics of North America, Vol. 80, pp. 443–457, 2000.

[4] A. Bilge, O. Cam, "Women's strategies for overcoming the stress and the examination of their health beliefs as a important factor for cancer preclusion", Journal of Anatolian Psychiatry, Vol. 9, pp. 16–21, 2008.

[5] M.F. Abbod, D.G. Von Keyserlingk, D.A. Linkens, M. Mahfouf, "Survey of utilization of fuzzy technology in medicine and healthcare", Fuzzy Sets and Systems, Vol. 120, pp. 331–349, 2001.

[6] A. Torres, J.J. Nieto, "Fuzzy logic in medicine and bioinformatics", Journal of Biomedicine and Biotechnology 2006, pp. 1–7, 2006.

[7] H. Ishibuchi, T. Nakashima, T. Murata, "A fuzzy classifier system that generates fuzzy if-then rules for pattern classification problems", IEEE International Conference on Evolutionary Proceedings, Vol. 2, pp. 759–764, 1995.

[8] H. Elbi, "Psychological aspects of cancer", Journal of Turkish Psychiatry, Vol. 2, pp. 60–64, 1991.

[9] G.J. Klir, B. Yuan, Fuzzy Sets and Fuzzy Logic: Theory and Applications, Prentice Hall, New Jersey, 1995.

[10] H. Ishibuchi, T. Nakashima, T. Morisawa, "Simple fuzzy rule-based classification systems performed well on commonly used real-world data sets", Proceedings of the North American Fuzzy Information Processing Society Meeting, pp. 21–24, 1997.

[11] N.H. Phuong, V. Kreinovich, "Fuzzy logic and its applications in medicine", International Journal of Medical Informatics, Vol. 62, pp. 165–173, 2001.

[12] M.T. Tsai, P.C. Tung, K.Y. Chen, "Experimental evaluations of proportional–integral–derivative type fuzzy controllers with parameter adaptive methods for an active magnetic bearing system", Expert Systems, Vol. 28, pp. 5–18, 2011.

[13] S. Guadarrama, S. Munoz, C. Vaucheret, "Fuzzy Prolog: a new approach using soft constraints propagation", Fuzzy Sets and Systems, Vol. 144, pp. 127–150, 2004.

[14] R.R. Yager, "On the interpretation of fuzzy if then rules", Applied Intelligence, Vol. 6, pp. 141–151, 1996.

[15] R. Belohlavek, V. Vychodil, "Fuzzy attribute logic over complete residuated lattices", Journal of Experimental & Theoretical Artificial Intelligence, Vol. 18, pp. 17–25, 2006.

[16] F. Steimann, "Fuzzy set theory in medicine", Artificial Intelligence in Medicine, Vol. 11, pp. 1–7, 1997.

[17] H. Seker, M. Odetayo, D. Petrovic, R.N.G. Naguib, "A fuzzy logic based method for prognostic decision making in breast and prostate cancers", IEEE Transactions on Information Technology in Biomedicine, Vol. 7, pp. 114–122, 2003.

[18] J. Ravi, A. Ajith, "A comparative study of fuzzy classification methods on breast cancer data", 7th International Work Conference on Artificial and Natural Neural Networks, 2003.

[19] N. Van Zandwijk, "Aetiology and prevention of lung cancer", European Respiratory Monograph, Vol. 17, pp. 13–33, 2001.

[20] R.M. Brand, D.D. Jones, H.T. Lynch, R.E. Brand, P. Watson, R. Ashwathnayaran, H.K. Roy, "Risk of colon cancer in hereditary non-polyposis colorectal cancer patients as predicted by fuzzy modeling: influence of smoking", World Journal of Gastroenterology, Vol. 12, pp. 4485–4491, 2006.

[21] C.Q. Bao, C. Jin, B.H. Xu, Y.L. Gu, J.P. Li, X. Lu, "Vaccination with apoptosis colorectal cancer cell pulsed autologous dendritic cells in advanced colorectal cancer patients: report from a clinical observation", African Journal of Biotechnology, Vol. 10, pp. 2319–2327, 2011.

[22] H.P. Nguyen, V. Kreinovich, "Fuzzy logic and its applications in medicine", International Journal of Medical Informatics, Vol. 62, pp. 165–173, 2001.