

Research Article

Colon Cancer: Intelligent Modeling of Risk Factors

Bouharati K¹, Hamdi-Cherif M¹, Mahnane A¹, Belkhir R², Laouamri S¹, Bouharati S^{3*}

¹Laboratory of Health and Environment, UFAS, Setif1University, Algeria

²Natural Science and Life, UFAS, Setif1University, Algeria

³Laboratory of Intelligent Systems, UFAS, Setif1University, Algeria

***Corresponding Author:** Bouharati Khaoula, Laboratory of Health and Environment. Faculty of Medicine, Ferhat Abbas University, Setif1, Algeria, Tel: +213 771-816-302; E-mail: sbouharati@univ-setif.dz

Received: 28 November 2018; **Accepted:** 10 December 2018; **Published:** 20 December 2018

Abstract

Several factors are involved in the development of colon cancer. These factors are characterized by complexity and uncertainty. The system is very complex to analyze using classical mathematical methods. Statistical analysis techniques are used. Due to the complexity of the system, these studies remain in the domain of the probable and the uncertain. A technique based on the principles of artificial intelligence is proposed. The principles of fuzzy inference are applied to the analysis of these factors. The proposed system makes it possible to take care of these uncertainty's inherent to the input variables. The result will be as accurate as possible. Just randomly set the input variables to instantly read the result at the output. This can be a tool to predict and prevent the occurrence of this type of cancer.

Keywords: Fuzzy logic; Intelligent modeling; Colon cancer; Risk factors

1. Introduction

Several factors may be considered as risk factors for colon cancer. These factors include age. The incidence of colon cancer is increasing in young adults and adults over 50 years of age. The American Association against Cancer recommend colorectal cancer screening below age 45, although the high rate increases with age [1]. Previous history of colorectal polyps can also constitute a risk factor. Studies have shown that the early diagnosis of colorectal cancers allows the identification of colonic polyps. Some patients will have an increased risk of developing this type of cancer. In general, the risk of colorectal cancer increases with age in both sexes. This type of cancer is recorded in half of diagnosed cases whose age exceeds 70 years whereas this rate is 10% among those fewer than 50 years [2]. Other risk factor is the inflammatory bowel disease. Crohn's disease and ulcerative colitis may constitute a risk of colorectal cancer. This risk increases with time and the anatomical extent of colitis as well as the degree of

inflammation. When colitis settles over time, it can be an aggravating factor in the development of colorectal cancer [3]. There is no doubt that the risk of colorectal cancer is directly associated with family ties, it is also demonstrated in families of patients with colorectal cancers. Several factors may be the cause of colorectal cancer, however, demographic factors (implying genetic factors) and dietary factors largely explain mortality by this cancer, although this remains poorly explained [4], and family history or a personal history of ovarian, endometrial or breast cancer. Studies have shown that family history is directly associated with the risk of colon cancer [5]. Other studies have analyzed the relationship between colon cancer in a person and the incidence in parents and siblings [6]. Other factors may constitute risk factors. The weight of the effect of each factor is often misunderstood. The physiological and metabolic system is very complex and varies from one person to another. The modelling of such risk factors is very difficult using classical mathematical techniques. Due to the complex nature of these factors, this study proposes the application of the principles of fuzzy inference. Fuzzy logic deals with uncertainty and where data are missing or incomplete [7]. The proposed system considers some factors measured during the period 2006 to 2014 by the National Cancer Registry of Setif in Algeria. The result will be the prediction of the appearance of this type of cancer from the values at the input of the system. As the analysis takes into account the uncertainties associated with the input variables and the inaccuracy due to the factors ignored and that have their effect, the output result will be as accurate as possible.

2. Materials and Methods

Recent techniques in computer science and artificial intelligence tend to imitate human reasoning. This is the foundation of advanced modelling [8-9]. Among these techniques, we use of the theory of fuzzy sets. The advantage of fuzzy logic is that it easily solves non-linear systems. Human expertise is translated into inference rules linking the input variables to the output variable. In recent times, the use of fuzzy analysis systems has become widespread in various fields, notably in the medical field [10]. In our case, this type of cancer is increase in people aged 45 years and older. This can be explained by the early detection and adoption of western lifestyle like physical inactivity and the richness of foods in fat [11]. This does not preclude other factors that have their effects and are ignored. This amply justifies the use of this analysis technique.

2.1 Diagram of designed fuzzy system

The proposed system consists of three input variables, a module of fuzzyfication, a database, a defuzzyfier module and an output Figure 1.

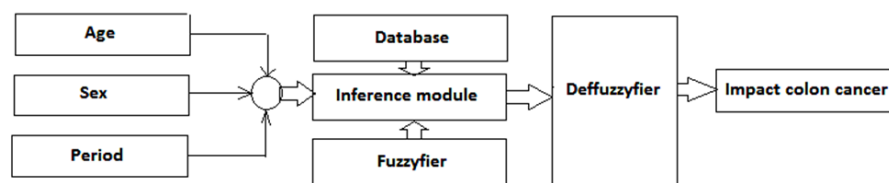


Figure 1: Structure of designed fuzzy system with three inputs, analysis module and output.

2.2 Fuzzyfication of registered input and output data

2.2.1 Input data:

- The variable “Age” expressed in numeric term must be fuzzyfied into linguistic term. For this, three triangular membership functions are created. Fuzzy intervals between two neighboring functions allow the uncertainties associated with this representation to be taken care of. The functions represent: young (0-30 years); Adult (25-55 years); Old (50 years and over).
- The variable “Sex” expressed in numeric term is not fuzzyfied. We assign a value (1) to the male sex and the value (2) to the female sex.
- The variable “Period” expressed in numeric term must be fuzzyfied. For this, three triangular membership functions are created. Fuzzy intervals between two neighboring functions allow the uncertainties associated with this representation to be taken care of. The functions represents the years between 2006 and 2014. (Table 1).

Age		0	5-	10-	15-	20-	25-	30-	35-	40-	45-	50-	55-	60-	65-	70-	75-	80-	85+
Sex	Year																		
Female	2006	-	-	1.3	-	1.2	2.7	3.4	4.3	2.5	5.9	15.5	16.1	15.4	31.5	18.2	55.1	49.8	-
	2007	-	-	-	-	-	1.3	-	4.1	2.5	5.7	3.7	29.5	15.2	23.7	26.9	38.9	23.2	46.4
	2008	-	-	1.4	1.2	-	2.6	7.8	2.0	2.4	11.0	14.3	27.6	21.9	32.1	44.1	49.1	-	87.0
	2009	-	-	-	-	-	1.2	-	5.7	4.7	8.1	13.6	8.7	34.0	32.5	8.7	23.4	-	-
	2010	-	-	-	-	-	-	4.3	5.5	4.6	7.9	22.6	25.0	31.0	32.9	26.1	-	19.0	-
	2011	-	-	-	-	-	3.5	1.4	1.7	-	7.7	6.1	12.0	33.7	82.6	17.4	-	-	35.7
	2012	-	-	-	1.3	-	5.8	4.1	3.4	-	12.0	18.1	19.8	11.0	40.6	17.1	21.1	35.1	-
	2013	-	-	-	1.3	-	1.1	1.3	6.8	2.1	7.5	32.7	23.4	27.3	32.1	33.8	10.4	52.0	-
	2014	-	-	-	-	-	2.3	-	-	2.1	17.4	11.8	23.1	27.0	31.7	83.6	41.2	34.3	-
Male	2006	-	-	-	-	1.1	-	5.0	-	2.5	3.0	3.7	21.2	27.0	18.5	20.5	44.5	25.7	-
	2007	-	-	-	-	2.2	-	4.8	2.0	2.5	5.7	10.8	9.5	-	9.4	40.7	56.2	125.1	100.1
	2008	-	-	-	-	3.2	1.2	4.6	3.9	2.4	5.5	3.5	4.3	32.3	28.4	30.7	26.7	46.3	187.0
	2009	-	-	-	-	1.1	-	-	1.9	9.2	5.3	9.9	12.3	21.6	28.9	41.0	77.5	22.5	45.0
	2010	-	-	-	2.4	-	1.2	2.8	-	6.7	-	12.7	23.8	38.0	49.0	20.7	37.9	-	-
	2011	-	-	-	1.2	-	-	-	3.4	4.3	2.5	6.0	19.2	-	39.3	21.1	-	19.8	39.7
	2012	-	-	-	-	-	2.2	2.6	6.7	2.1	15.1	-	30.3	38.5	48.3	-	61.9	39.1	39.1
	2013	-	-	-	1.2	-	-	5.2	-	8.4	2.5	11.6	37.4	38.0	47.7	20.5	48.9	38.6	77.1
	2014	-	-	-	-	-	2.3	-	-	2.1	17.4	11.8	23.1	27.0	31.7	83.6	41.2	34.3	-

Table 1: Incidence per 100.000 by age group (period) of Colon cancer (2006-2014).

2.2.2 Output data: The variable “Incidence” expressed in numeric term must be fuzzyfied into linguistic term. For this, three triangular membership functions are created. Fuzzy intervals between two neighbouring functions allow the uncertainties associated with this representation to be taken care of. We assign a value (0-40) to the low incidence, (30-70) to the average incidence and (60-100) to high incidence with reference to the recorded incidence values Table 1.

2.3 Fuzzy rules database

In general, the expression of a rule is of the form (IF ... THEN)

IF (Antecedents)...THAN (Consequence).

By referring to registered values, we assign the set of rules in this form.

Example:

IF the “Age” is young AND the “Gender” is female AND the “Year” is 2010, THEN the “Incidence” is low.

The rule base must contain all possible combinations.

2.4 Inference module

When the basis of the rules is established by reference to the measured values and the intervention of the human expert, this has made it possible to establish an intelligent application in which it is possible to predict the incidence of colon cancer. The inference of the system makes all these rules co-operate and make the correspondence between them and the output result as a consequence. The result obtained takes into account even the fuzzy intervals created between the neighboring membership functions.

[System]

Name='Colon Cancer'

Type='mamdani'

Version=2.0

NumInputs=3

NumOutputs=1

Num Rules=50

AndMethod='min'

OrMethod='max'

ImpMethod='min'

AggMethod='max'

DefuzzMethod='centroid'

[Input1]

Name='Age'

Range=[0 100]

NumMFs=3

MF2='2007': 'trimf', [1 2 3]

MF3='2008': 'trimf', [2 3 4]

MF4='2009': 'trimf', [3 4 5]

MF5='2010': 'trimf', [4 5 6]

MF6='2011': 'trimf', [5 6 7]

MF7='2012': 'trimf', [6 7 8]

MF8='2013': 'trimf', [7 8 9]

MF9='2014': 'trimf', [8 9 10]

[Input3]

Name='Sex'

Range=[0 3]

NumMFs=2

MF1='Male': 'trimf', [1 1 1]

MF2='Female': 'trimf', [2 2 2]

[Output1]

MF1='Young':'trimf',[0 15 30]

MF2='Adult':'trimf',[25 40 55]

MF3='Old':'trimf',[50 65 1000000]

[Input2]

Name='Period'

Range=[0 10]

NumMFs=9 MF1='2006':'trimf',[0 1 2]

Name='Impact'

Range=[0 100]

NumMFs=3

MF1='Low.Incidence':'trimf',[0 20 40]

*MF2='Average.Incidence':'trimf',
,[30 50 70]*

MF3='High.Incidence':'trimf',[60 80 100]

The result is to simply introduce randomly the values at the input of the system; and to read instantly the incidence at the output. Example at the Figure 2.



Figure 2: Example application random setting of the inputs and direct reading of the result at the output.

3. Results

Many studies have been devoted to analyzing the data. The tools used are often classical mathematical tools such as differential equations or statistical analyzes. The finding is that it is very difficult to encompass all the variables in a mathematical equation which makes these tools very heavy and sometimes impossible to solve. The statistical tools are still in the probable with uncertainties. This study circumvents these difficulties by using an artificial intelligence tool with fuzzy logic. The variables involved in the process are considered uncertain and expressed in linguistic terms. By this, the uncertainties are compensated. Each variable is represented by a triangular membership function over an interval ranging levels. Linguistic variables are assigned to intervals. Note that neighboring intervals overlap in fuzzy intervals. By this, uncertainties are taken care of. The output variable is linked to the input variables to consider all possible combinations. Once the system is established, it gives the possibility to introduce random variables at the input to instantly read the result at the output. This result comes from the collaboration of all the rules that link the inputs to the output.

4. Conclusion

In view of the complexity of the analyzed system, the technique used can be a tool to help prevent the onset of colon cancer. This prediction is translated into a decision model from the introduction of the input variables. Simple to use, this system supports all the complexity and inaccuracy of data. The data processed are actual values recorded during a period that may provide a basis for future years. Compensation for data deficiencies is ensured by fuzzyfication of the input and output parameters. Moreover, the intervention of human expertise is necessary in the establishment of rules of inference to overcome any insufficiency in the analysis of the data. Whatever the reliability of the results obtained, this does not prevent the use of other factors which are not considered in this study and which have an effect on colon cancer. For this, the proposed system remains extensible to the other parameters. This study can then constitute a contribution and a basis for future more global studies and can be a tool for the prediction and prevention of colon cancer.

Declaration of Interest

The authors declare that no conflict of interest.

Acknowledgments

All our gratitude is in the staff service of epidemiology of the Faculty of Medicine of the University of Setif1 in Algeria.

References

1. Owen Dyer. Colorectal cancer: US guidelines urge screening from age 45 as incidence soars in younger adults. *BMJ* 361 (2018): 2452.
2. Kolligs FT. Diagnostics and Epidemiology of Colorectal Cancer. *Visc Med* 32 (2016): 158-164.
3. Yoshino T, Nakase H, Takagi T, et al. Risk factors for developing colorectal cancer in Japanese patients with ulcerative colitis: a retrospective observational study-CAPITAL (Cohort and Practice for IBD total management in Kyoto-Shiga Links) study I. *BMJ Open Gastro* 3 (2016): 1-8.
4. Shaukat A, Dostal A, Menk J, et al. BMI Is a Risk Factor for Colorectal Cancer Mortality. *Digestive Diseases and Sciences* 62 (2017): 2511-2517.
5. Zanghieri G, Gregorio C, Sacchetti C, et al. Familial occurrence of gastric cancer in the 2-year experience of a population-based registry. *Cancer* 66 (1990): 2047-2051.
6. Sidney JW, Ann GZ, Hans G, et al. and the National Polyp Study Workgroup. Risk of colorectal cancer in the families of patients with adenomatous polyps. *N Engl J Med* 334 (1996): 82-87.
7. Bouharati I, Babouche F, Bouharati K, et al. Brain tumor identification from MRI imaging using fuzzy inference. *Radiol Diagn Imaging* 2 (2018): 1-4.
8. Dumitrache I, Constantin N, Dragoicea M. *Reteleneurale (Neural networks)*, Matrix Press (1999).
9. Dumitrache I, Buiu C. *Algoritmi genetici. Principii fundamentale esaplicatii in automatic (Genetic algorithms. Fundamental principles and applications in automatics)*, Mediamira Press (2000).

10. Javad H, Reza J, Ayoob N. Fuzzy Rule Based Diagnostic System For Detecting The Lung Cancer Disease. Journal of Renewable Natural Resources 3 (2015): 147-157.
11. Mokhtar HC, Diego S, Abbas M, et al. Time trends of cancer incidence in Setif, Algeria, 1986-2010: An observational study. BMC Cancer 14 (2014): 637.

Citation: Bouharati K, Hamdi-Cherif M, Mahnane A, Belkhir R, Laouamri S, Bouharati S. Colon Cancer: Intelligent Modeling of Risk Factors. Journal of Bioinformatics and Systems Biology 1 (2018): 015-021.



This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC-BY\) license 4.0](https://creativecommons.org/licenses/by/4.0/)