Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 6.017

IJCSMC, Vol. 7, Issue. 4, April 2018, pg.195 – 204

A BELIEF RULE BASED EXPERT SYSTEM TO ASSESS COLORECTAL CANCER UNDER UNCERTAINTY

Tanjia Chowdhury

Department of Computer science and Engineering, University of Chittagong, Bangladesh

Email: tanjiacse5@gmail.com

Abstract— Colorectal cancer (or bowel cancer) is the fatal disease for both men and women in worldwide. The risk of getting colorectal cancer increases with age, most commonly after age 50. The highest rates are seen in male, whereas the rates of women are low. In Bangladesh, 12% adult male and 3% female are affected by colorectal cancer. Colorectal cancer is diagnosed from its various signs and symptoms by physicians. Maximum patients are unable to express their signs and symptoms to the doctors clearly. For this reason it became difficult to detect the colorectal cancer accurately. From their opinions (patients) various types of uncertainty (Unknowable information) issues arises. This diagnosis process cannot provide 100 % result. In this paper, an expert system using evidential reasoning approach (RIMER) is used to detect the presence of colorectal cancer. Using this expert system, we can handle different kinds of uncertainty. This research focused on the results of three systems namely Belief Rule Based Expert System, fuzzy logic based system and human experts. The web based expert system is used instead of desktop expert system which is user friendly. The data analyzed has been accomplished by using SPSS 22 and MATLAB. The output of the BRBES provides more reliable result than fuzzy logic expert system as well as human experts.

Keywords— Colorectal cancer; RIMER; Belief Rule Base; Uncertainty; Signs and Symptoms.

I. INTRODUCTION

Colorectal cancer occurs when the normal cells of the colon and rectal are turned into abnormal cells. When the growing of the inner line of the colon or rectum is out of control then it turned into a polyp which causes cancer. The wall of the colon or rectum consists of multiple skins. Colorectal cancer starts in the inner skin (polyp) and it spreads in skins gradually. Then they can grow into the blood vessels or lymph vessels. In this way, they can spread all over of the body parts [2]. There are two types of polyps are Adenomatous polyps, Hyperplastic polyps and inflammatory polyps. Most of the colorectal cancers are sporadic (80%), which have no prior family history of the disease, others familial or hereditary (20%)

colorectal cancer which have family history of the disease [6] [7]. In Bangladesh, doctor advised the patients to performed several expensive test. Sometimes it could be difficult for them. Usually physician used the primary signs and symptoms on the patients to detect CRC. The signs and symptoms of CRC are abdominal pain, blood in the stool, Rectal bleeding, weakness and fatigue, unexpected weight loss, unexplained anaemia [6][7][14].

In medical science, a symptom is subjective while a sign is objective. Such as blood in the stool which is recognized by the doctors, nurse, family members and the patients, it's called sign. On the other hand, abdominal pain, weakness and fatigue which is sensed by the patient, it's called symptom. The traditional process consists of different kinds of uncertainties such as ambiguous, imprecision, vagueness, incompleteness, randomness and ignorance.

In this paper, three types of referential values such as high, medium and low are used for expressing the patient's condition. Vague, ambiguous, and imprecise are arises from rectal bleeding. So it can be observed that rectal bleeding cannot be expressed with 100% certainty. Because Patients cannot tell accurate time when bleeding started. As a result, different types of activity may provide vague information. Therefore, uncertainty arises due to incompleteness. In Table I we discussed different kinds of uncertainties for CRC signs and symptoms.

TABLE I. UNCERTAINTY ISSUE

CRC Factors	Uncertainty	Discussion
Diarrhoea	Ignorance, Imprecision, Incompleteness	Patients cannot explain how many weeks they are suffering by diarrhoea.
Blood in the stool	Ignorance	Sometimes patient does not notice their bowel movement and its colour.
Rectal bleeding	Ambiguous, Imprecision, vagueness	Patients cannot tell accurately when his/her bleeding is started. As a result, different types of activity may provide vague information.
Weakness and Fatigue	Imprecision, vagueness, ambiguity	Patient can avoid weakness and fatigue because they think may be it arises from overnight working. So on this purpose they do not go to doctors. .
Abdominal pain	Ignorance, vagueness, Randomness	Sometimes some patients cannot explain their pain. When the pain started and where the pain started.

The traditional way of detecting the CRC cannot provide accurate result because of uncertainty associated with signs and symptoms. Our proposed expert system can be solved that problem which is arises from traditional system.

According to American Cancer Society, Black women and men are most affected by CRC than in white men and women. CRC have several risk factors. The risk factors of CRC can be categorized in two sections, namely sporadic and others familial or hereditary. Some of risk factors are changeable but some are not changeable. Several tests are used to diagnose the CRC. Such as Guaiac-based fecal occult blood test (gFOBT) and fecal immunochemical

test (FIT), which is used to diagnosis the presence of polyp, Stool DNA test etc. sometimes it is unable to find polyps [6]. The description of Patients health condition plays a vital role for a screening test. If patients fails to express their feelings, then it arises difficulty for a screening test. Different types of problem are emerged from CRC treatment such as weakness, Nausea, diarrhoea, Sexual dysfunction etc [3] [6].

Several methods are used for diagnosis CRC. Muhammad Ammar Bin Shafi had proposed fuzzy linear regression with symmetric parameter is the best model for predicting CRC tumor size. He used twenty five independent variables to detect CRC tumor size. In my paper nine variables are used to detect a CRC. Brand et al. (2006) had proposed Fuzzy modelling which is used to analyse risk factors of colon cancer in hereditary non-polyposis CRC. Varela et al. (2006), Pereira et al. (2011) had discovered the stage of side effects for CRC patients after applying various treatment. Obi and Imianvan (2012) have taken nine symptoms to diagnosis CRC. Roslani et al. (2012) noticed one of the main factors which cause CRC, namely blood tool. There are two types of screening method guaiac-based fecal occult blood test (gFOBT) and fecal immunochemical test (FIT). He compared two methods and found that gFOBT test was positive.

Artificial neural network was also used to detect CRC. In this process, several symptoms were used but factors were not classified. This paper had focused on enormous number of variables (e.g. taken huge key factors) instead of small number of variables. Rule extraction process isn't fit for their work [11] [13].

In the previous study, different kinds of methods are used for colorectal cancer diagnosis. But belief rule based expert system is not used to detect CRC. The main goal of my research is to proper detection of colorectal cancer using rule based expert system which is not used in previous studies. This is the best method for detection of any kind of disease because it can handle various uncertainty issues. In this paper mentioned six types of uncertainties issues. Fuzzy logic as well as other systems cannot handle all of them issues whereas belief rule based expert system can handle all issues. Physician used Screening test to diagnose CRC and sometimes is not provide 100% certainty. However, Belief rule based expert system provides accurate result than fuzzy and experts system.

This paper is arranged in the following ways: Section II narrated overview of the RIMER methodology, section III narrated Implementation of Proposed methodology, section IV narrated Results and Discussion and finally section V narrated conclusion of this research.

II. RIMER METHODOLOGY

A rule based expert system is divided into two parts namely Belief rule base knowledge representation and Inference engine [9] [12].

A. Belief rule base knowledge representation

Belief Rule base knowledge representation schema has the ability to handle uncertainty. In the RIMER approach, a belief rule is expressed as IF-THEN rule. Each rule consists of two parts, namely antecedent part ant consequent part. There are various kinds of Antecedent attribute are also used. We had taken three types of referential values for each antecedent attributes namely High, Medium and Low. Rule weight and antecedent attribute weight are used as Knowledge representation parameters. Colorectal cancer assessment consists of 108 rules. For diagnosis of CRC, a belief IF-THEN rule is expressed as follows:

RK: IF (Unexpected weight loss is high) and (Unexplained anaemia is high) and (Weakness and Fatigue is low)

THEN Rectal bleeding is {(High, 0.5), (Medium, 0.3), (Low, 0.2)}

Here, the consequent attribute is "Rectal bleeding" and antecedent attribute are unexpected weight loss, unexplained anaemia and weakness and fatigue. We used "High", "Medium" and

"Low" as referential values and belief degree of each antecedent attribute are {High(0.7), Medium(0.3) and Low (0.0)} for rectal bleeding consequent. It is states that, 50% belief to bleeding as critical, 30% as median and 20% as minor. The total degree of the belief is (0.50+0.30+0.20) =1. The estimation is true.

B. Inference Procedures Using ER approach

In this procedure evaluates how various input data are transformed into final result. The steps of the inference procedures, which is used in a BRBES is given below:

1. Input Transformation

Input transformation refers to distribute input data among three referential values of antecedent [J.B. Yang, J. Liu, Wang, J, 2006]. From communication with patients and human experts (doctor), got various input data. There are three input antecedents: Sporadic (A1), Rectal bleeding (A2), Familial history (A3) and consequent colorectal cancer (A4). In this work, {High, Medium, Low} is used as a set of three referential values.

The distributed values transformed from original input data as follows:

$$S(D_i^*) = \{(h_{3i}, \gamma_{3i}), (h_{2i}\gamma_{2i})\}$$

$$\gamma_{2i} = (h_{3i} - D_i)/(h_{3i} - h_{2i})$$

$$\gamma_{3i} = 1 - \gamma_{2i} \quad (1)$$

From the figure 2, the consequent of A1 and the antecedents of A1 are D1, D2 and D3. Here showed one input data distribution for antecedent D_1. Let, the evaluation grades for each factor are h_1 (high) ‹1.00, h_2 (Medium) ‹0.5, h_3 (low) ‹0.2and D_1=0.8 and D_2=0.4 then using formula equation (1) for D_1:

$$\gamma_{2i} = \frac{1-0.8}{1-0.5} = \frac{0.2}{0.5} = 0.4, h_3 \geq 0.8 \geq 0.2$$

$$\gamma_i = (1-0.4) = 0.6,$$

$$\gamma_{3i} = 0.0$$

It has been found from the analysis of input data transformation the value of D_1 {(high, 0.6), (medium, 0.4), (low, 0)} and this referential values of each antecedent are called matching degree.

1. Calculation of Activation Weights

At first, we need to calculate the matching degree for calculating the activation of weights. The formula of matching degree [9] [12]

$$\alpha_k = \prod_{i=1}^{T_K} \left(\alpha_i^k\right)^{\overline{\gamma_{ki}}}$$

Where, $\overline{\delta_{ki}} = \frac{\delta_{ki}}{\max(\delta_{ki})}$ and i=1…….T_k

Finally, the formula for activation weight is:

$$\omega_k = \frac{\theta_k \alpha_k}{\sum_{i=1}^{L} \theta_i \alpha_i}$$

$$\sum_{i=1}^{L} \theta_i \alpha_i = \left(\theta_1 \alpha_1 + ... + \theta_n \alpha_n\right) \quad (2)$$

Here, the relative weight is denoted as θ_k of the ***k***th rule and every rule weight is 1. In the BRB while it is "1" then it's considered as high. The total number of belief rule is denoted as L.

The formula of calculating l is:

$$L = \prod_{i=1}^{T} J_i \tag{3}$$

J_i : Denotes he referential values of i[th] antecedent attribute

For example, Rule weight is 1, i.e. $\theta_i = 1$

$\alpha_1 = (0.6)^1 * (0.0)^1 = 0$, then $W_1 = \frac{1*0}{1} = 0$, Here $\sum_{i=1}^{L} \theta_i \alpha_i = 1$

2. Belief Degree Update

For diagnosis of CRC, we applied IF-THEN rule where the entire antecedent attribute had existed. The step of Belief degree update is used when some input data are missing. Using these steps we can remove ignorance [9] [12].

The formula of belief degree update is

$$\beta_{ik} = \overline{\beta_{ik}} \frac{\sum_{i=1}^{T_K} (\tau(t,k) \sum_{j=1}^{J_i} \alpha_{tj})}{\sum_{i=1}^{T_k} \tau(t,k)}$$, where β_{ik} denotes belief degree update

$$\tau(t,k) = \{1, \text{if } D_i \text{ is used in defining } R_k (t = 1....., T_k)$$

otherwise 0} (4)

TABLE II. UPDATED VALUE

Rule Id	State	High	Medium	Low
1	Initial	0.8	0.0	0.2
	Update	0.64	0.0	0.16

4. Rule aggregation

Rule aggregation measured by ER method. There are two way for aggregating input value such as Analytical and Recursive ER methods. Comparatively analytical method provides best output than recursive method. The formula of mass representation is [9] [12]

$$m_{j,k} = w_k \beta_{j.k}, j = 1....., N$$

$$m_{D,K} = 1 - w_k \sum_{J-1}^{N} \beta_{j,k}$$

$$\beta_j = \frac{m_{j,I(L)}}{1 - \overline{m}_{D,I(L)}}, J = 1....., N$$

(5)

To convert aggregated fuzzy values into numeric values following formula will be used.

$$Y_m = \sum_{n=1}^{N} D_n * \beta_n(m)$$, $\beta_n(m)$ is the utility values. (6)

III.IMPLEMENTATION OF PROPOSED METHODOLOGY

A. Architecture and Implementation

The System architecture consists of three components, namely input, process and outputs. The architectural style defines the pattern of system organization. The BRBES consists of user interface (presentation layer), inference engine (application layer) and BRB knowledge base (data processing layers) .The architecture of BRBES as shown in below Figure 1.

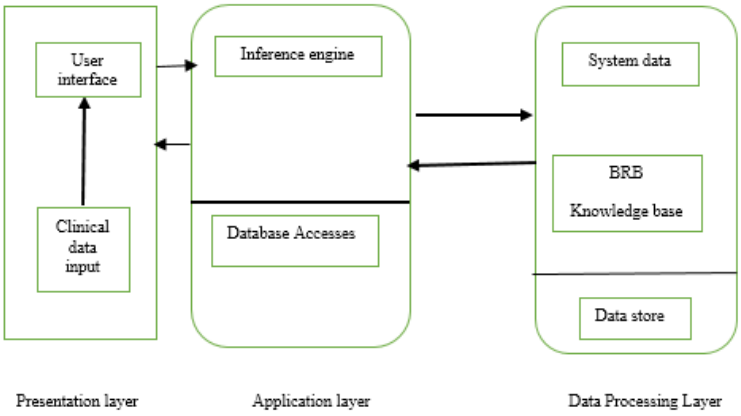

Figure1. Architecture of BRBES

We developed web based system for user interaction. Users access the expert system at any time and at anywhere without any anxiety. In front-end, JavaScript, Jquery, HTML and CSS are used to implement the system interface. In application layer, PHP is used for the inference procedure and database access because of its simplicity. In data-base management layer, MySQL is used which is back-end to store clinical data and knowledge-base.

B. Knowledge Base Construction in BRB

There are the three levels BRB structures to diagnose CRC. It consists of 4 consequent (A1 A2, A3, and A4) and 9 antecedents (D1-D9) attribute. There are 4 sub-rule bases such as A1 A2, A3, A4 and apply 108 belief rules to estimate initial rule base, where A1=27,A2=27,A3=27and A4=27.We calculate rules by using (3).We considered three referential values for each antecedent attributes.

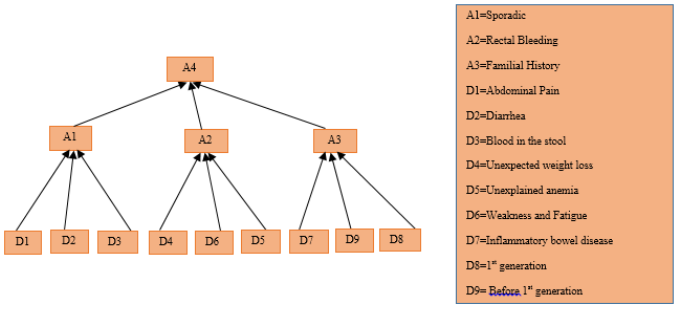

Figure 2 BRB frameworks to assess CRC

From the figure 2, the input antecedent are D1 (Abdominal pain), D2 (Diarrhea), D3 (Blood in the stool), D4 (Unexpected weight loss), D5 (Unexplained anemia), D6 (Weakness and fatigue), D7 (Inflammatory bowel disease), D8 (1^{st} generation), and D9 (Before 1^{st} generation). The BRB consists of 4 sub rule bases namely A1 (Sporadic), A2 (Rectal bleeding) and A3 (Familial history). Here, to calculate activation weight, we worked by taken rule weight is 1. Table IV Shows 1^{st} initial rule base for A1.

TABLE IV. INITIAL BELIEF RULE BASE for A1

Serial	Rule weight	IF (Antecedent)			THEN		
		D1	D2	D3	Consequent A1 (High | Medium | Low)		
0	1	H	H	H	1	0	0
1	1	H	H	M	0.8	0.2	0
2	1	H	H	L	0.8	0	0.2
3	1	H	M	L	0.6	0.4	0
4	1	H	M	M	0.4	0.6	0
5	1	H	M	H	0.5	0.3	0.2
6	1	H	L	H	0.8	0	0.2
7	1	H	L	M	0.5	0.3	0.2
8	1	H	L	L	0.2	0	0.8
9	1	M	H	H	0.8	0.2	0
10	1	M	H	M	0.4	0.6	0
		.....	.....	.....			
26	1	L	L	L	0	0	1

C. BRB Interface

System Interface is an intermediate media between user and system. The interface of the belief rule based expert system to diagnose CRC as shown in Figure 3.

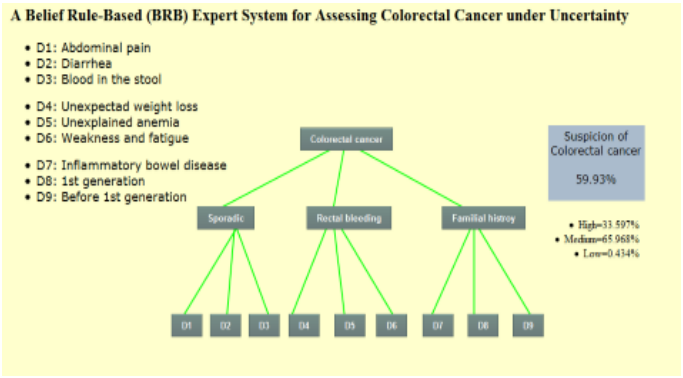

Figure 3 Interface of the system

Here, we calculated the input value of nine antecedent attribute and it has been found from the analyses, combined referential values of CRC is "(High, 33.59)", "(Medium, 65.97)", and "(Low, 0.43)".The overall result of assessment for colorectal cancer has been acquired in terms of crisp value by using (5) and obtained as 59.93%. Each of the input value is collected from the opinion of the patient as well as doctors. The possible results of CRC has been found by using (1)-(6).

IV. RESULTS AND DISCUSSION

To determine a system's accuracy with the results, real data and expert assumptions are needed. We collected the data from the Chittagong Medical Hospital, Bangladesh. In this paper, we used CRC signs and symptoms which are abdominal pain(D1), diarrhea(D2), and blood in the stool(D3), rectal bleeding(A2), unexpected weight loss(D4), unexplained anemia(D5),Weakness and fatigue(D6), inflammatory bowel disease(D7),1st generation(D8) and before 1st generation(D9) . But human assumptions are not always accurate as uncertainty issues associated with it. For this reason, we have proposed an expert system consisting belief rule knowledge based schema (BRB) with RIMER inference mechanism. For analysis, we have used various data from the CRC patients. Here, we used two numbers 1 and 0 for benchmark. We put benchmark is 1, when CRC is found and put 0 if it's not found. The comparison results of the three systems are shown in Table V. The following below figures has been generated by using SPSS 22. Fuzzy system result has been generated by using Mat Lab.

TABLE V. COMPARISON RESULT AMONG FUZZY SYSTEM, BRBES AND EXPERTS OPINION

SL. NO	D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8	D_9	Fuzzy Result	BRBES Result	Expert Opinion	Benchmark
1	H	H	H	H	H	H	H	H	H	80.44	94.25	78	1
2	H	L	L	L	H	H	H	L	M	65	77.49	70	1
3	L	H	L	M	H	L	L	H	H	45.71	64	60	1
4	L	L	L	L	H	L	L	L	L	40	27.30	30	0
5	L	M	H	L	H	L	L	H	H	38.32	35.28	41	0
6	H	H	H	L	H	L	L	L	L	67.53	74.35	70	1
7	H	M	M	L	H	L	L	M	M	80	76.80	79	1
8	H	L	L	L	H	H	L	L	M	39.23	64.20	50	0
9	H	H	H	H	H	H	L	M	M	73.56	89.99	70	1
10	L	M	M	H	H	M	M	M	H	50.78	40.01	50	0
11	H	H	M	M	H	H	H	M	L	70.03	85	73	1
12	H	M	H	H	H	M	H	H	H	86	85.10	84	1
13	H	H	H	H	H	H	L	H	M	80	90.23	78	1
14	M	H	H	H	H	M	M	H	M	58	69.56	60	1
15	H	H	H	H	H	M	L	M	M	78	80	83	1

In this paper, The Receiver Operating Characteristic (ROC) curve is used to measure the exact result of the system. It is used to compare the diagnostic performance of two or more diagnostic tests *(Griner et al., 1981).* The Area under Curve (AUC) measures the exact result of the CRC. The performance of the system can be measured by calculating the area under curve. From Table VI, the value of AUC of BRBES is 0.827 whereas the value of AUC of Expert Opinion is 0.810 and the value of AUC of fuzzy system is 0.825.As a result, from the below figure 4,5 and 6, it showed the performance of BRBES is better than the expert's opinion and fuzzy system. The experts result provides erroneous result because doctors ignore uncertainty issues associated with signs and symptoms of the CRC and fuzzy rule based expert system (FRBES) cannot manage ignorance, incompleteness and randomness.

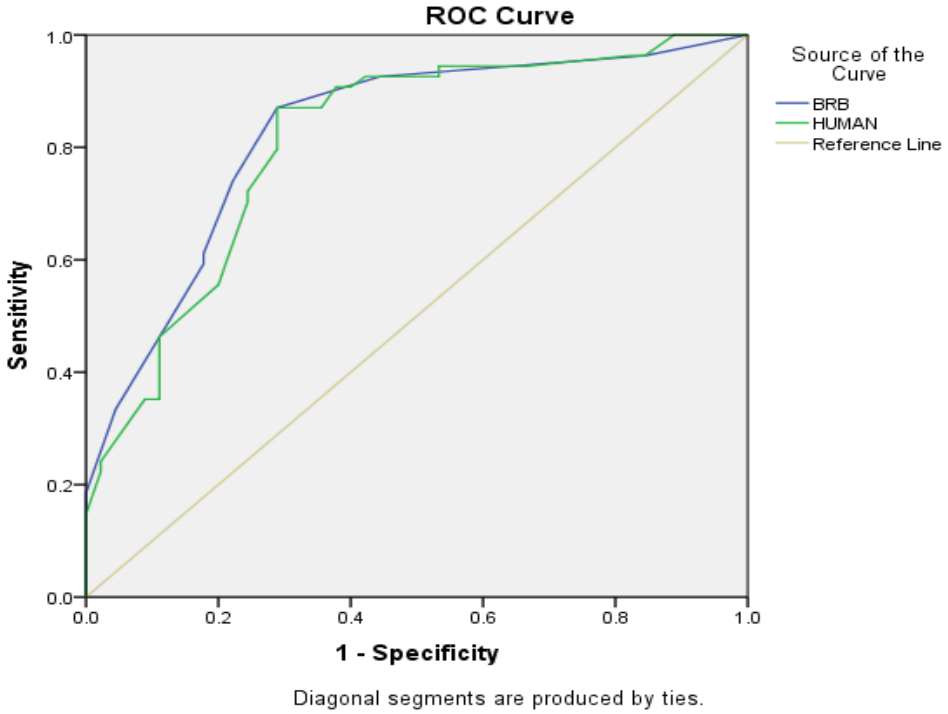

Figure 4: Comparison between BRBES output and Human (Expert's) assumption

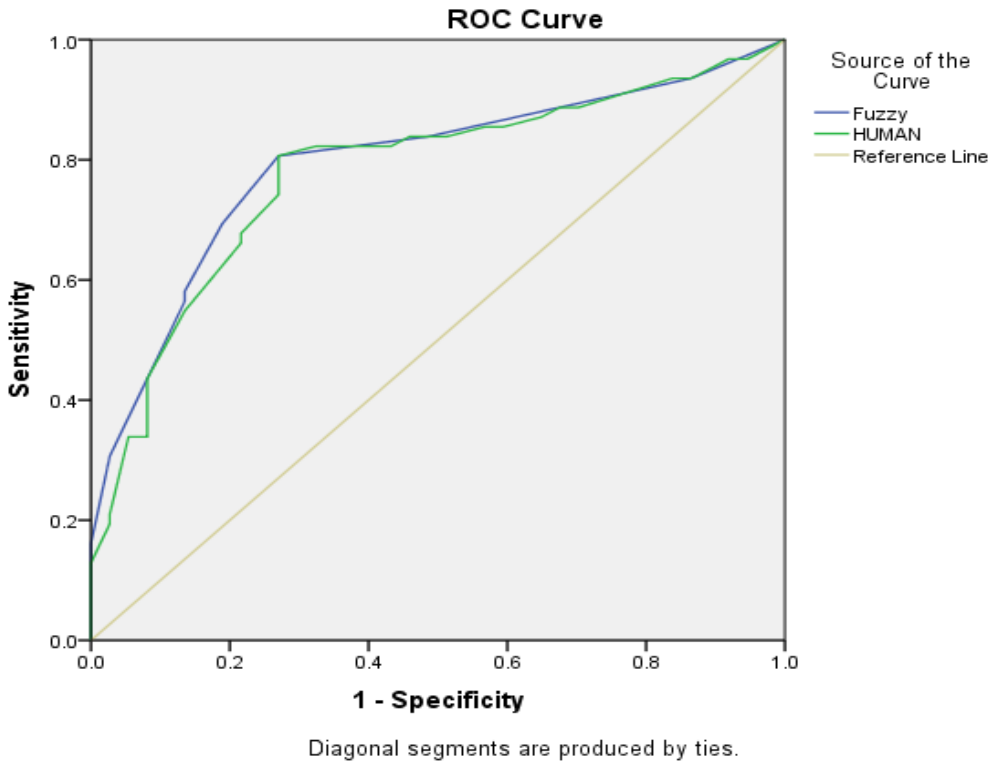

Figure 5: Comparison between Fuzzy output and Human (Expert's) assumption

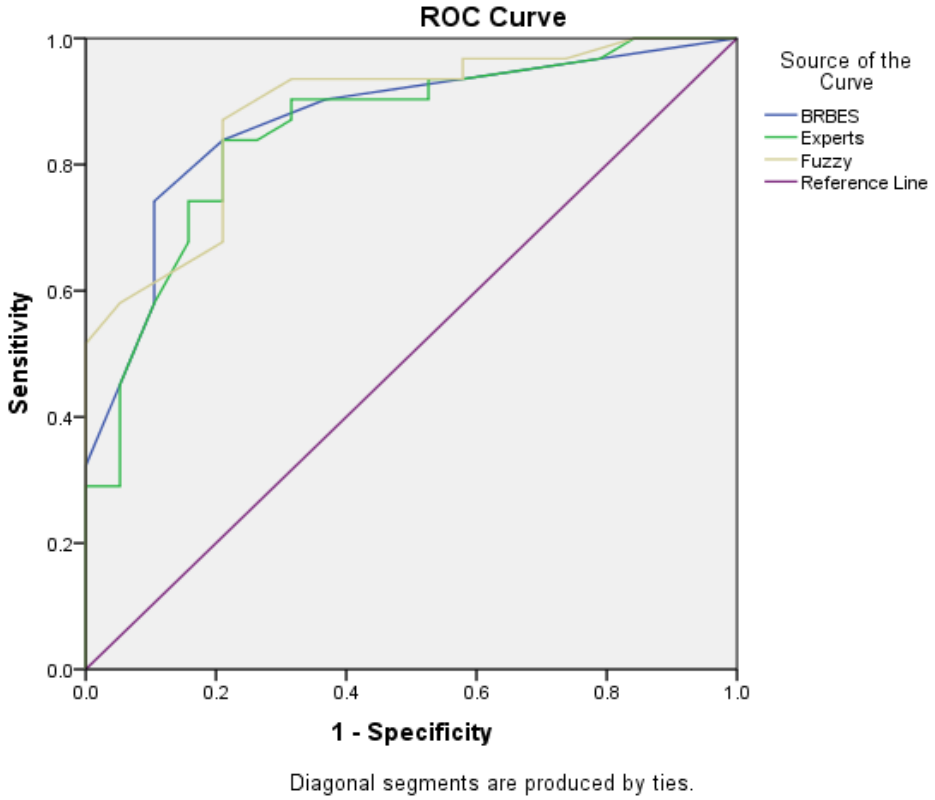

Figure 6: Combining output result of BRBES, Expert's assumption and Fuzzy system

TABLE VI. PERFORMANCE EVALUATION

Test Result Variables	Area	Asymptotic 95% Confidence Interval	
		Lower Bound	Upper Bound
BRBES	.827	.744	.910
Expert	.810	.723	.890
Fuzzy	.825	.734	.897

V. CONCLUSIONS

Bangladesh is a densely populated country. Each year two lacks people are affected by cancer. Cancer causes death. There are so many people who are not aware about their health. As we know colorectal cancer is the silent killer. So people should give attention to their health. In Bangladesh the cost of cancer treatment is very expensive. So people are afraid of it. I think early detection of cancer can remove their fears. That's why, in this paper Belief Rule Based Expert system (BRBES) is used to diagnosis colorectal cancer. This system of suspecting is more cost effective than traditional suspecting system of Colorectal Cancer. To get better result, here we are considered uncertainty issues which are manageable by using this expert system. On the other hand, previous method was not sufficient for managing all uncertainty issues. To develop this system, we are collected the input value from clinical data history and from patients. From data analysis and ROC curve, it proved that the belief rule based expert system is the best system which provides better result comparatively from the experts' result and fuzzy rule based expert system. In future we want to establish the optimization model for CRC.

REFERENCES

[1] Jackson, Peter. *Introduction to expert systems*. Addison-Wesley Longman Publishing Co., Inc., 1998.

[2] Konar, Amit. *Artificial intelligence and soft computing: behavioral and cognitive modeling of the human brain*. CRC press, 1999.

[3] Smith, Robert A., Vilma Cokkinides, and Harmon J. Eyre. "American Cancer Society guidelines for the early detection of cancer, 2006." *CA: a cancer journal for clinicians* 56.1 (2006): 11-25.

[4] Hossain, Mohammad Shahadat, et al. "A Belief Rule Based (BRB) Decision Support System to Assess Clinical Asthma Suspicion." *Scandinavian Conference on Health Informatics; August 22; 2014; Grimstad; Norway*. No. 102. Linköping University Electronic Press, 2014.

[5] Patwary, Muhammed Jamshed Alam, Subrina Akter, and Tanjim Mahmud. "An Expert System to Detect Uterine Cancer under Uncertainty." *IOSR Journal of Computer Engineering (IOSR-JCE), e-ISSN* (2014): 2278-0661.

[6] Colorectal cancer –American Cancer Society. [Online]. Available: https://www.cancer.org

[7] Colorectal cancer-ICMR. [Online]. Available: http://www.icmr.nic.in

[8] Bangladesh cancer organisations and resources. [Online]. Available: http://www.cancerindex.org/Bangladesh

[9] Chen, Yu-Wang, et al. "Inference analysis and adaptive training for belief rule based systems." *Expert Systems with Applications* 38.10 (2011): 12845-12860.

[10] HAMAD, AQEEL MOHSIN. "*LUNG CANCER DIAGNOSIS BY USING FUZZY LOGIC*." (2016).

[11] Shafi, Muhammad Ammar. *Application of fuzzy linear regression models for predicting tumor size of colorectal cancer in Malaysia's hospital*. Diss. Universiti Tun Hussein Onn Malaysia, 2015.

[12] Yang, Jian-Bo, et al. "Belief rule-base inference methodology using the evidential reasoning approach-RIMER." *IEEE Transactions on systems, Man, and Cybernetics-part A: Systems and Humans* 36.2 (2006): 266-285.

[13] Bottaci, Leonardo, et al. "Artificial neural networks applied to outcome prediction for colorectal cancer patients in separate institutions." *The Lancet* 350.9076 (1997): 469-472.

[14] Parveen, Rafia, et al. "Cancer Types and Treatment Modalities in Patients Attending at Delta Medical College Hospital." *Delta Medical College Journal* 3.2 (2015): 57-62.

