Original Article
http://mjiri.iums.ac.ir
Medical Journal of the Islamic Republic of Iran (MJIRI)

Med J Islam Repub Iran. 2021(3 Apr);35.44. https://doi.org/10.47176/mjiri.35.44

Developing a clinical decision support system based on the fuzzy logic and decision tree to predict colorectal cancer

Raoof Nopour[1], Mostafa Shanbehzadeh[2], Hadi Kazemi-Arpanahi[3, 4]*

Received: 17 Jun 2020 Published: 3 Apr 2021

Abstract

Background: Colorectal Cancer (CRC) is the most prevalent digestive system- related cancer and has become one of the deadliest diseases worldwide. Given the poor prognosis of CRC, it is of great importance to make a more accurate prediction of this disease. Early CRC detection using computational technologies can significantly improve the overall survival possibility of patients. Hence this study was aimed to develop a fuzzy logic-based clinical decision support system (FL-based CDSS) for the detection of CRC patients.

Methods: This study was conducted in 2020 using the data related to CRC and non-CRC patients, which included the 1162 cases in the Masoud internal clinic, Tehran, Iran. The chi-square method was used to determine the most important risk factors in predicting CRC. Furthermore, the C4.5 decision tree was used to extract the rules. Finally, the FL-based CDSS was designed in a MATLAB environment and its performance was evaluated by a confusion matrix.

Results: Eleven features were selected as the most important factors. After fuzzification of the qualitative variables and evaluation of the decision support system (DSS) using the confusion matrix, the accuracy, specificity, and sensitivity of the system was yielded 0.96, 0.97, and 0.96, respectively.

Conclusion: We concluded that developing the CDSS in this field can provide an earlier diagnosis of CRC, leading to a timely treatment, which could decrease the CRC mortality rate in the community.

Keywords: Colorectal cancer, CRC; Fuzzy logic, Artificial intelligence, Risk analysis, Screening

Conflicts of Interest: *None declared*
Funding: *The Research Deputy of Abadan Faculty of Medical Sciences*

Cite this article as: Nopour R, Shanbehzadeh M, Kazemi-Arpanahi H. Developing a clinical decision support system based on the fuzzy logic and decision tree to predict colorectal cancer. *Med J Islam Repub Iran.* 2021 (3 Apr);35:44. https://doi.org/10.47176/mjiri.35.44

Introduction

Colorectal cancer (CRC) is a type of gastrointestinal (GI) malignancy that appears as masses are believed to arise from primarily benign adenomatous polyps; it can invade or extent to other parts of the body (1, 2). CRC has become the third commonest malignant tumor worldwide and is also the second leading cause of cancer-related deaths in women and the third for men worldwide (2-4). Additionally, it ranked the first cause of death in terms of cancer in most developing co untries (5). In recent decades, the CRC

Corresponding author: *Dr Hadi Kazemi-Arpanahi, H.kazemi@abadanums.ac.ir*

1. Department of Health Information Management, School of Allied Medical Sciences, Tehran University of Medical Sciences, Tehran, Iran
2. Department of Health Information Technology, School of Paramedical, Ilam University of Medical Sciences, Ilam, Iran
3. Department of Health Information Technology, Abadan Faculty of Medical Sciences, Abadan, Iran
4. Student Research Committee, Abadan Faculty of Medical Sciences, Abadan, Iran

↑What is "already known" in this topic:
Colorectal cancer (CRC) is one of the most fatal malignancies, and total survival rates remain unsatisfactory. The symptoms of CRC are mostly insidious at the beginning, preventing early diagnosis. Therefore, CRC is either progressive or metastasized in most of patients by the time they are diagnosed. The rapid case finding of CRC is not only a simple process but also is crucial to its treatment.

→What this article adds:
Application of fuzzy logic can develop a predictive model to stratify risk CRC patients more accurately. This may provide worthy prognostic information to support the physician for making a decision in a good manner for risk stratification before the onset of the treatment.

growth rate in Iran has been increasing annually and is considered as one of the critical public health challenges (6, 7). In Iran, according to the Ministry of Health reports, CRC occupies the third position in men and is the fourth type of prevalent cancer among women (4, 8, 9). This malignancy is highly treatable when detected in early stages (3). Therefore, initiating measures against CRC inhibition, protection, and treatment have a fundamental value. In this sense, an accurate and regular screening enables earlier detection; and then timely interventions can contribute to enhancing the likelihood of recovery, cure, and lower death (2, 10). Due to the complexity of the disease diagnosis, many factors should be considered, which may be complementary, contradictory, and competitive; thus, all health professionals attempt to decrease imprecision in diagnosis by collecting empirical data to manage patients' problems (11).

Even though there have been advancements in diagnostic procedures, a huge number of CRC patients, more than 90%, have had either advanced or metastasized CRC by the time they are diagnosed. Therefore, an accurate, timely, suitable, economical, and afordable diagnosis and regular screening could considerably improve the management of this disease (1). An application alternative for the existing screening tools is artificial intelligence (AI) screening. It is proven that the adoption of optimal AI techniques in the form of Decision Support Systems (DSS) can increase the classification accuracy capabilities (10, 12-14).

Case finding and analysis of CRC risk development can escalate the diagnostic precision and potentially decrease CRC-related death. Furthermore, by distinguishing the malignant and benign tumors early, AI can be used to decrease the needless procedures on benign tumors, reducing the total cost, procedure duration, and related problems (15-17).

The present paper proposes the use of a fuzzy logic-based CDSS to assess the risk of developing CRC. It is well recognized that ambiguity almost exists in all multifarious situations, which necessitates intellectual decision-making. In an ambiguous situation, the FL-based CDSS is more effective with qualitative data. FL offers a practically applicable approach to dealing with uncertainty and vagueness management in decision support (18, 19). The concept of FL was first introduced in 1965 by Zadeh (20). FL represents an alteration in the Boolean logic paradigm. Opposing to conventional Boolean logic, where the objects are classified with crisp memberships as true or false (true = 1; false = 0), if the value is 0, the element does not belong to the set; and if the value is 1, then the element completely belongs to the set. However, in fuzzy sets, each element has a degree of belonging to the membership that can be a real number, ranging from 0 and 1 (21-23). This logic uses the mathematical concepts to provide a framework for modeling intrinsic imprecision and uncertainties in medical practice for dealing with decision-making in prognosis, diagnosis, and treatment, especially in complex cases such as cancer (24-26). In other words, FL is suitable in situations in which the borderline between the sets is not well-defined, for example, an analyzing process done by man (27). Due to the complex and multidimensional nature of CRC, and the interrelation between modifiable and nonmodifiable risk factors in cancer genesis, progress, and exacerbation, the FL is applicable for multistage cancer prediction and prognosis (28). Thus, this study aimed at developing an FL-based CDSS for predicting CRC at early stages.

Methods

This retrospective applied-descriptive study was conducted in 2020, aiming at developing an FL-based DSS for earlier detection of CRC.

Data Collection

The dataset used in this study was obtained from the CRC cases referred to the Masoud internal clinic, Tehran, Iran, in 2017-2019. The patients' information was reviewed and extracted by 2 health information management experts.

The patients who referred to the Masoud internal clinic for screening, diagnosis, and treatment of CRC were included in this study. A total of 1162 patients (610 men and 552 women) were identified. The study limited the analysis of patients' medical records in 2 phases. First, to study individuals lacking high-risk factors of CRC, we did not consider the patients' medical records comprising those factors in Burke clinical screening guideline, and thus 382 cases were excluded. Second, 62 incomplete case records were excluded (more than 70% missing data). We collected 25 variables, including patients'characteristics (age, sex, body mass index (BMI), educational level, marital status, and ethnicity), nutritional characteristics (animal fat intake, fruit and vegetable consumption, and red meat consumption), lifestyle characteristics (physical activity, sleeping in the day, smoking consumption, alcohol drinking, and drug use), and clinical characteristics (taking Iron, calcium, vitamin D, multivitamins tablets, contraceptive pills, taking menopause hormones, history of fatty liver, hormone-therapy, metabolic syndrome, history of colonoscopy and endoscopy, and genetic risk.

Knowledge Acquisition

In this study, knowledge acquisition was conducted in 3 stages as follows.

Reduce the Data Set Dimensions

Reducing the size of the dataset is imperative to increasing the efficiency of data mining algorithms; to this end, only the most important factors were selected. In this study, The chi-square test was used for determining the most influential risk factors for developing CRC. This method can determine the correlation between the 2 qualitative variables. The correlation between the 2 different variables can be calculated by by the chi-square test and Formula 1, where Fo_i is the observed frequency and Fe_i is the expected frequency, whose calculation method has been presented in Formula 2 (29).

$$X^2 = \sum_{i=1}^{n} (Fe_i - Fo_i)^2 / Fe_i \quad (1)$$

$$\mathrm{Fe_i} = \frac{\mathrm{n_i} * \mathrm{n_j}}{\mathrm{n}} \quad (2)$$

Knowledge Representation

Two points were important in developing the C4.5 data mining algorithm: first, it is necessary to consider all the factors. Next, the algorithm's confidence factor was chosen in such a way that the performance of the algorithm doesn't decrease, so data mining was performed by the confidence factor of 0.5.

A common technique for acquiring the knowledge is extracting and representing the rules structure by IF-THEN statement, in which IF and THEN refer to the condition and the result of actions, respectively. In this study, to extract the rules from the decision tree, we first started from the root node and by reaching the leaf node, we transformed the entire navigation path into the IF-THEN structure rule. To extract the next rules, we started again from the root node and wrote the entire node in that path to the next leaf node. In addition to extracting rules from the decision tree, weights were assigned to each rule generated based on the frequency of the samples classified in each of the leaf nodes. So the fuzzy system's knowledge base rules with IF-THEN format were extracted by the C4.5 decision tree algorithm.

Evaluating the Data Mining Algorithm Performance

Based on the confusion matrix (Table 1), we used the accuracy, sensitivity, and specificity criteria with the 10 fold cross-validation for evaluating the algorithm performance (Relation 3 to 5).

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \quad (3)$$
$$Specificity = \frac{TN}{TN+FP} \quad (4)$$
$$Sensitivity = \frac{TP}{TP+FN} \quad (5)$$

Designing a Decision Support System

Due to the presence of qualitative variables and uncertainties in their values, to design a clinical DSS, we used a fuzzy method for simulation of human reasoning with uncertainty. In this study, the researcher-made method was used for the fuzzification of variables. Fuzzification was performed by consulting with 10 gastrointestinal experts familiar with the CRC factors. For example, for the fruit and vegetable consumption variable, the amount between 200 to 300 grams were considered as the low consumption and the consumption between 220 to 280 grams was more likely to be in the low consumption group. Therefore, by consulting the experts, the trapezoid membership function (180 220 280 320) was assigned for this feature of the variable. In Formula 6, f(x) shows the membership function of the x existed in the low range.

$$f_{low}(x) = \begin{cases} (x-180)/40, & 180 \le x < 220 \\ 1, & 220 \le x < 280 \\ (320-x)/40, & 280 \le x < 320 \end{cases} \quad (6)$$

The output of the fuzzy system offers a flexible solution to physicians in this way. In this study, we used MATLAB R 2013 software to implement a fuzzy inference system. The output value was between 0 to 1. If the value was lower than 0.5, we considered it as low risk, else they were classified in the high-risk group.

Table 1. The Confusion Matrix

Output		Predicted values	
		Positive (1)	Negative (0)
Actual value	Positive (1)	TP	FN
	Negative (0)	FP	TN

Table 2. Mamdani Reasoning Mechanism Used for CRC Risk Prediction

And method	MIN
Or method	MAX
Implication	MIN
Aggregation	MAX
Defuzzification	Centroid

The Mamdani inference mechanism was used in this study as the reasoning method. In the Mamdani method, the inputs and outputs of the fuzzy inference system are expressed as fuzzy-fuzzy, and the relationships between them are mapped by the rules that existed in the system knowledge base. The system parameters used in the Mamdani reasoning mechanism for CRC risk prediction are shown in Table 2. For developing FL-based DSS user interface, we used the C# programming language and the NET framework 4.5.2 in windows layout by Visual Studio software.

Evaluating Developed CDSS

Evaluating the performance of the developed CDSS confusion matrix (Table 1) was used. The accuracy, specificity, and sensitivity of the system were calculated according to Relation 1.

In this respect, 125 cases of healthy people and 125 CRC patients were compared to low-risk and high-risk categories by the system, respectively.

Results

The result of the sample isolation was 718 cases, 468 of which were used for data mining and 250 (125 CRC samples and 125 healthy samples) were used for evaluating the system performance. Grade 1 and 0 were associated with people with CRC and non-CRC, respectively. At first, 25 factors, including demographic, nutritional, personal and medical history, and epidemiological factors, were used for analysis using the chi-square test; finally, 11 variables were obtained as the most important variables at $p<0.05$ (Table 3).

Finally, the knowledge base of the system was formed by 60 rules for determining the risk of CRC with a specific weight. These rules were extracted by considering the paths that had classified samples in their leaf nodes, but the paths that were terminated by leaf nodes without any classified samples were discarded. The followings are some rules for determining the risk associated with specific weights. Grade 0 is non-CRC people associated with low-risk people and grade 1 is CRC people associated with high-risk people in this study.

1. IF (Smoking (Day) is very-low) && (Aspirin (Day/2) is low) && (Exercise (In hours) is low) && (Red meat is low) && (BMI is Fat(Low-weight)) then Class= non-CRC, Rule weight=3/389.
2. IF (Smoking (Day) is very-low) && (Aspirin (Day/2)

Table 3. Chi-square Test Result for Determining Important CRC Risk Factors

No	Symbol	Variable Name	Chi-square	p
1	Fat meal	Mean animal fat consumption in a day	28.354	<0.001
2	Family history	The risk grouping of people in different types of relative type in terms of having CRC	24.322	0.004
3	Age	Age intervals	18.377	0.006
4	Red meat	Mean red meat consumption in a day	27.354	<0.001
5	Fruits& vegetables	Mean Fruits & Vegetables consumption in a day	31.681	<0.001
6	Exercise	Exercise (In hours)	28.711	<0.001
7	Aspirin (P.Days/2)	Mean aspirin take (the half pill in a day)	24.328	<0.001
8	Aspirin (Years)	Mean aspirin take (In years)	25.364	<0.001
9	Smoking (Days)	Smoking consumption (the number of pockets in a day)	32.127	<0.001
10	Smoking (Years)	Smoking consuming (In years)	29.757	<0.001
11	BMI	Body Mass Index	18.473	0.049

Table 4. C4.5 Confusion Matrix

Predicted Values	Actual Values	
	+	-
+	253	21
-	58	136

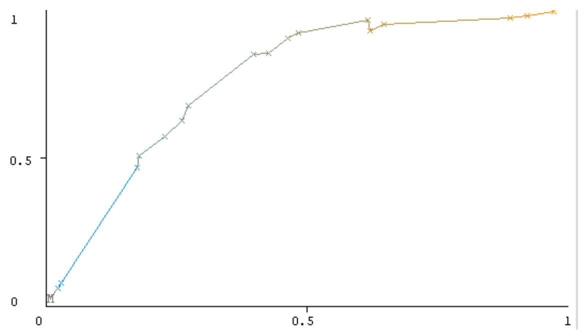

Fig. 1. The ROC of the J-48 algorithm.

is low) && (Exercise (In hours) is low) && (Red meat is low) && (BMI is Fat(Normal)) then Class= non-CRC, Rule weight=8/389.

3. IF (Smoking (Day) is very-low) && (Aspirin (Day/2) is low) && (Exercise (In hours) is low) && (Red meat is low) && (BMI is High-weight) then Class= CRC, Rule weight=4/389.

4. IF (Smoking (Day) is very-low) && (Aspirin (Day/2) is low) && (Exercise (In hours) is low) && (Red meat is low) && (BMI is Fat(Deg-1)) then Class= CRC, Rule weight=6/389.

5. IF (Smoking (Day) is very-low) && (Aspirin (Day/2) is low) && (Exercise (In hours) is low) && (Red meat is low) && (BMI is Fat(Deg-2)) then Class= CRC, Rule weight=3/389.

Based on the results of the confusion matrix (Table 4), the accuracy, specificity, and sensitivity values of the C4.5 decision tree were obtained to be 0.83, 0.70, and 0.92, respectively.

The receiver characteristics operator (ROC) curve has been depicted in Figure 1, based on the ROC; the area under the ROC curve (AUC) of the J-48 decision tree algorithm in classifying the high-risk and low-risk people was obtained to be 0.811. (The vertical and horizontal vertices in this figure represent the true positive rate (TPR) and false positive rate (FPR), respectively).

After fuzzifying the most important variables, the database of the fuzzy inference system was gathered (Table 5).

By determining the database, knowledge base, and fuzzy reasoning mechanism, the fuzzy decision support system interface was designed (Fig. 2).

After designing a fuzzy decision support system, in the test step, the evaluation was performed by the 250 samples separated from the data mining previously. The accuracy, specificity, and sensitivity of the system based on relation 2 to 4, and Table 6 were obtained to be 0.96, 0.97, and 0.96, respectively.

Discussion

This study aimed to develop a CDSS that would assist clinicians with their decisions concerning CRC risk prediction. DSS is fundamental to help health care providers in their decision-making (diagnosis, classification, etc.), especially it is applicable in serious situations such as cancer, where the decision must be made effectively and reliably. In this respect, providing a system for early cancer detection is valuable in disease therapy and inhibition. The reason why CRC was selected in this study is due to its frequency in our country. Additionally, this neoplasm tends to present at late stages and has a poor outcome, and most of the patients diagnosed in advanced stages as the results of this disease metastases to neighboring organs, the treatments are hopeless and patients die in a short time. Moreover, the detection of cancer greatly depends on the physician's expertise, but these experienced physicians are not available in all parts of the country. Thus, early CRC precaution procedures are very important for individuals who

Table 5. Fuzzy Database with Specific Fuzzy Membership Function for CRC Risk Prediction

No.	Variable Name	Variable Role	Values (Probable Values Existed in Each State)	Fuzzy Membership Functions
1	Smoking consumption (number of pockets consumed per day)	Input	Very low: <1 Low: 1-2 Medium: 2-3 High: >3	Triangular (0 0 1.25) Trapezoid (0.75 1.2 1.3 2.25) Trapezoid (1.75 2.3 2.7 3.25) Triangular (2.75 3.5 4)
2	Smoking consuming (In years)	Input	Very Low: <1 Low: 1-5 Medium: 5-10 High: >10	Triangular (0 0 1.5) Trapezoid (0.75 2.5 3.5 6) Trapezoid (4 7 9 11) Triangular (8 15 20)
3	Mean aspirin take (pill in day/2) (one pill is 80 mg)	Input	Low: <40 Medium: 40-120 High: 120-200 Very high: >200	Triangular (0 35 45) Trapezoid (30 45 115 130) Trapezoid (90 135 185 230) Triangular (180 205 220)
4	Mean aspirin take (years)	Input	Low: <1 Medium: 1-3 High: 3-5 Very high: >5	Triangular (0 0 1.25) Trapezoid (0.75 1.5 2.5 3.25) Trapezoid (2.75 3.5 4.5 5.25) Triangular (4.75 5.5 6)
5	Body mass index	Input	Low weight: <18.5 Normal: 18.5-24.9 High weight: 25-29.9 Fat (degree 1): 30-34.9 Fat (degree 2): 35-39.9 Fat (degree 3): >40	Triangular (0 15 20) Trapezoid (15 20 22 26) Trapezoid (23 26 28 31) Trapezoid (28 31 33 36) Trapezoid (33 35.5 38.5 41) Triangular (37 44 51)
6	Age	Input	Young: <45 Middle-aged: 45-64 Adult: >65	Triangular (0 25 50) Trapezoid (40 50 60 70) Triangular (60 70 90)
7	Mean fruits & vegetable consumption (Serving per day) (One serving is 100 g)	Input	Very low: <200 Low: 200-300 Medium <300-400 High: >400	Triangular (0 150 250) Trapezoid (180 220 280 320) Trapezoid (280 320 370 420) Triangular (380 450 550)
8	Mean animal fat consumption (Serving per day) (One serving is 50g)	Input	Very low: <50 Low: 50-100 Medium: 100-150 High: >150	Triangular (0 30 60) Trapezoid (40 65 85 120) Trapezoid (80 105 145 170) Triangular (140 165 200)
9	Mean red meat consumption (Serving per day) (One serving is 50g)	Input	Very low: <50 Low: 50-100 Medium: 100-150 High: >150	Triangular (0 25 55) Trapezoid (35 60 85 105) Trapezoid (80 100 140 160) Triangular (130 160 200)
10	Exercise (in hours)	Input	Low: <1 Medium: 1-2 High >2	Triangular (0 0.5 1.25) Trapezoid (0.75 1.3 1.7 2.25) Trapezoid (1.75 2.25 2.75 3)
11	Family history	Input	Non-relative (Very low): <1 Having the degree3 relative with CRC (low): 1-2 Having the degree2 relative with CRC (Medium) 2-3 Having the degree1 relative with CRC (High):>3	Triangular (0 0 1) Trapezoid (0.5 1 1.5 2) Trapezoid (1.5 2 2.5 3.5) Triangular (2.5 3.5 4)
12	Risk	target	Low Risk : <=0.5 High risk: >0.5	Triangular (0 0.3 0.55) Triangular (0.45 0.7 1)

have not fallen yet or are not diagnosed in the early stage of CRC. This could quicken and optimize the referral to specialists, diminishing the medical care expenditures and leading to reduce disease morbidity, mortality, and eventually improving the overall disease survival (12, 13, 30-33).

Recent developments in computational technologies provided more analytical capabilities to predict malignancies than traditional statistical approaches (27, 34-36). In this study, by selecting the FL model from AI techniques, we developed a DSS for the assistant physician in predicting CRC. Nowadays, multiple AI-based DSS programs have been designed for the CRC early detection, risk analysis, and screening (1, 28). The intent of applying fuzzy reasoning is due to its significant ability to handling uncertainty, imprecision, and complexity while increasing the flexibility of inference methods with approximate reasoning. Contrary to Boolean logic, FL models human expert reasoning in the form of uncertain linguistic variables. Thus, it can be stated that computers can solve problems in a way close to human experts (23, 25, 27, 37). FL provides a solution for dealing with doubts in medical decision-making practices. Using this logic for CRC prognosis or prediction has led to adherence to the best practices and guidelines (13, 38, 39).

A key component of any CDSS is the knowledge base, which uses a traditional way to represent human knowledge. These knowledge-based systems mostly use the rules in the form of IF-THEN statements, and the samples are always according to these patterns of rules (29, 40,

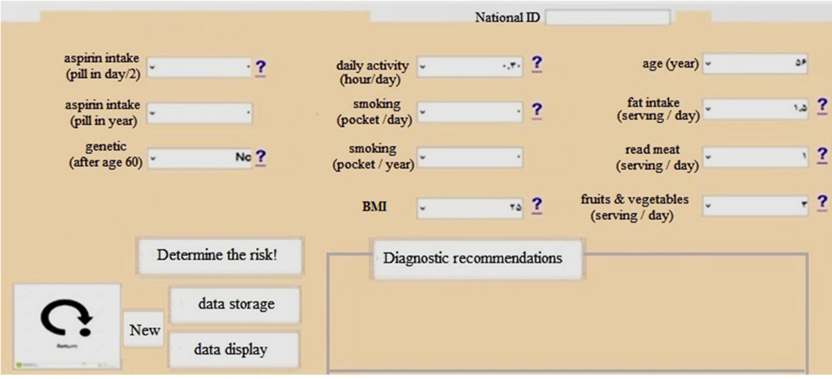

Fig. 2. The fuzzy user interface in the Visual Studio software

Table 6. The Decision Support System Confusion Matrix

System Record	Low Risk	High Risk
Non-CRC	122	3
CRC	5	120

41). Knowledge elicitation is a bottleneck in establishing these systems (42). The developed system in this study was developed based on a retrospective dataset; data miming was applied to produce the required rule base. Prediction by using data mining techniques yielded an accurate result. Data mining is the process of exploring, representing, and modeling huge volumes of data to extract hidden patterns or probable relations that provide applicable information. Mining and analyzing health data are potentially useful to perform medical evaluations to screening, prognosis or diagnosis, treatment, and survival estimation, which consequently leads to enhance clinical decision-making (22, 43-45).

Decision tree algorithms have good results in crisp domains but cannot deal with imprecision. Therefore, in this study, the C4.5 decision tree is used coupled with fuzzy modeling to deal with this impression. A decision tree with fuzzy logic was used to determine if the risk is classified as low or high in a certain individual. The role of FL is to moderate the sharp decision between attribute values in the decision tree that may result in misclassification (46). In this study, we used the C4.5 decision tree algorithm because of its good performance; also, its capabilities, such as embedding missing values before classifying the samples, downsizing and upsizing the tree by changing the confidence factor attribute, and generating rules, and ability to use data features with continuous numerical values for building the tree in addition to qualitative types, make this algorithm a better choice than other tree algorithms (47, 48).

Thus far, many previous studies have been focused on the application of fuzzy modeling in early GI malignancy prognosis, risk assessment, and survivability prediction (14, 22, 23, 49). Some of them demonstrated that applying fuzzy logic for CRC risk assessment can reduce diagnosis errors and disagreement among physicians at any level of prediction, prevention, and treatment. Yilmaz et al (2013) in their study indicated that using the fuzzy logic for CRC risk analysis yielded 0.78 of sensitivity, 0.85 of specificity, 0.90 of identification, 0.30 of negative identification, and 0.81 of accuracy (26). Santos et al (2018) in their study revealed that fuzzy sample entropy had the best performance for CRC diagnosis and differentiating benign and malignant patient groups based on histological images (AUC value = 0.983) (50). Shafi (2015) also stated that the fuzzy logic was the best AI technique to predict CRC tumor size (51). Chowdhury et al (2018) used the fuzzy logic for CRC detection at early stages and the result of their study showed that FL-based expert system, by embedding uncertainty, could be used as conventional CDSS and improve the CRC prediction accuracy (21). In our study, the result of Mamdani-type fuzzy inference in terms of accuracy, specificity, and sensitivity based confusion matrix were 0.96, 0.97, and 0.96, respectively. It also could be concluded that using this method for data analysis with uncertainty will increase the performance efficiency. Achieved results in the present study showed that the FL serves as an effective approach in dealing with the CRC risk analysis.

Lack of enough quantitative data, single center-based, the incompleteness of some data fields, the retrospective nature of the data set, the existence of some qualitative and non-objective data, and lack of some prognostic factors, such as patient and family history, are the key research limitations. It is suggested that more data mining methods, multicenter databases, and more quantitative data gathering approaches be used for increasing the performance of the CRC risk prediction. In this study, we proposed an FL-based CDSS for CRC risk prediction. Nonetheless, the great enhancement of diagnostic precision from an intelligent CDSS could be an advantage for multimodal medical data fusion and decision-making for CRC, the system acts as an ancillary diagnostic tool for the finding and forecast of the disease. This CDSS cannot be substituted by doctors in choosing the last decision for patients due to the high mortality and morbidity rates resulted from cancer.

Conclusion

This study proposes a method for predicting CRC based on FL yielding, a robust prognostic model with better accuracy than the other traditional clinical and statistical techniques for CRC patients screening. The true prediction will improve its screening and will have a positive effect on referring patients to the specialist cancer setting at the early stage of the disease. It is suggested that in the future works the researchers use new databases and new classification schemes to stabilize these accuracies with more objective, complete, accurate, and comprehensive databases. Finally, we showed that the fuzzy model can potentially be used as an optimal technique to identify the high-risk CRC people in our country.

Acknowledgement
This article is extracted from a research project supported by Abadan Faculty of Medical Sciences with the Ethical code number of IR.ABADANUMS.REC.1399.141. We appreciate the Research Deputy of Abadan Faculty of Medical Sciences who sponsored this project financially.

Conflict of Interests
The authors declare that they have no competing interests.

References

1. Chen H, Lin Z, Wu H, Wang L, Wu T, Tan C. Diagnosis of colorectal cancer by near-infrared optical fiber spectroscopy and random forest. Spectrochim Acta A Mol Biomol Spectrosc. 2015;135:185-91.
2. Sabouri S, Esmaily H, Shahidsales S, Emadi M. Survival Prediction in Patients with Colorectal Cancer Using Artificial Neural Network and Cox Regression. Int J Cancer Manag. 2020;13(1):6.
3. Ai D, Pan H, Han R, Li X, Liu G, Xia LC. Using decision tree aggregation with random forest model to identify gut microbes associated with colorectal cancer. Genes. 2019;10(2):112.
4. Goshayeshi L, Pourahmadi A, Ghayour-Mobarhan M, Hashtarkhani S, Karimian S, Dastjerdi RS, et al. Colorectal cancer risk factors in north-eastern Iran: A retrospective cross-sectional study based on geographical information systems, spatial autocorrelation and regression analysis. Geospat Health. 2019;14(2).
5. Rieger AK, Mansmann UR. A Bayesian scoring rule on clustered event data for familial risk assessment–An example from colorectal cancer screening. Biom J. 2018;60(1):115-27.
6. Taheri M, Tavakol M, Akbari ME, Almasi-Hashiani A, Abbasi M. Associations of demographic, socioeconomic, self-rated health, and metastasis in colorectal cancer in Iran. Med J Islam Repub Iran. 2019;33:17.
7. Tsoi KK, Hirai HW, Chan FC, Griffiths S, Sung JJ. Predicted increases in incidence of colorectal cancer in developed and developing regions, in association with ageing populations. Clin Gastroenterol Hepatol. 2017;15(6):892-900. e4.
8. Mozafar HS, Khodamoradi F, Salehiniya H. Associated Factors of Survival Rate and Screening for Colorectal Cancer in Iran: a Systematic Review. J Gastrointest Cancer. 2020;51(2):401-411
9. Nikbakht HA, Shokri-Shirvani J, Ashrafian-Amiri H, Ghaem H, Jafarnia A, Alijanpour S, et al. The first screening program for colorectal cancer in the North of Iran. J Gastrointest Cancer. 2019:1-7.
10. Arunkumar C, Ramakrishnan S. Prediction of cancer using customised fuzzy rough machine learning approaches. Healthc Technol Lett. 2018;6(1):13-8.
11. Ahmadi H, Gholamzadeh M, Shahmoradi L, Nilashi M, Rashvand P. Diseases diagnosis using fuzzy logic methods: A systematic and meta-analysis review. Comput Methods Programs Biomed. 2018;161:145-72.
12. Hornbrook MC, Goshen R, Choman E, O'Keeffe-Rosetti M, Kinar Y, Liles EG, et al. Early colorectal cancer detected by machine learning model using gender, age, and complete blood count data. Dig Dis Sci. 2017;62(10):2719-27.
13. Lualdi M, Cavalleri A, Battaglia L, Colombo A, Garrone G, Morelli D, et al. Early detection of colorectal adenocarcinoma: a clinical decision support tool based on plasma porphyrin accumulation and risk factors. BMC Cancer. 2018;18(1):841.
14. Scrobotă I, Băciuț G, Filip AG, Todor B, Blaga F, Băciuț MF. Application of Fuzzy Logic in Oral Cancer Risk Assessment. Iran J Public Health. 2017;46(5):612.
15. Ichimasa K, Kudo SE, Mori Y, Misawa M, Matsudaira S, Kouyama Y, et al. Artificial intelligence may help in predicting the need for additional surgery after endoscopic resection of T1 colorectal cancer. Endoscopy. 2018;50(3):230-40.
16. Wang KW, Dong M. Potential applications of artificial intelligence in colorectal polyps and cancer: Recent advances and prospects. World J Gastroenterol. 2020;26(34):5090-100.
17. Smit MA, Mesker WE. The role of artificial intelligence to quantify the tumour-stroma ratio for survival in colorectal cancer. EBio Med. 2020;61.
18. Jin S, Peng J, Li Z, Shen Q. Bidirectional approximate reasoning-based approach for decision support. Inf Sci. 2020;506:99-112.
19. Chang SS, Zadeh LA. On fuzzy mapping and control. Fuzzy sets, fuzzy logic, and fuzzy systems: selected papers by Lotfi A Zadeh: World Scientific; 1996. p. 180-4.
20. Zadeh LA. Fuzzy sets. Information and control. 1965;8(3):338-53.
21. Chowdhury T. Fuzzy Logic Based Expert System for Detecting Colorectal Cancer. IRJET. 2018;5(9):389-293.
22. Safdari R, Arpanahi HK, Langarizadeh M, Ghazisaiedi M, Dargahi H, Zendehdel K. Design a fuzzy rule-based expert system to aid earlier diagnosis of gastric cancer. Acta Inform Med.. 2018;26(1):19.
23. Wang CY, Lee TF, Fang CH, Chou JH. Fuzzy logic-based prognostic score for outcome prediction in esophageal cancer. IEEE Trans Inf Technol Biomed. 2012;16(6):1224-30.
24. Halder A, Kumar A. Active learning using rough fuzzy classifier for cancer prediction from microarray gene expression data. J Biomed Inform. 2019;92:103136.
25. Yılmaz A, Arı S, Kocabıçak Ü. Risk analysis of lung cancer and effects of stress level on cancer risk through neuro-fuzzy model. Comput Methods Programs Biomed. 2016;137:35-46.
26. Yilmaz A, Ayan K. Cancer risk analysis by fuzzy logic approach and performance status of the model. Turk J Elec Eng Comp Sci. 2013;21(3):897-912.
27. Miranda GHB, Felipe JC. Computer-aided diagnosis system based on fuzzy logic for breast cancer categorization. Comput Biol Med. 2015;64:334-46.
28. Lu W, Fu DL, Kong XX, Huang ZH, Hwang M, Zhu YS, et al. FOLFOX treatment response prediction in metastatic or recurrent colorectal cancer patients via machine learning algorithms. Cancer Med. 2020;9(4):1419-1429.
29. McHugh ML. The chi-square test of independence. Biochem Med. 2013;23(2):143-9.
30. Takamatsu M, Yamamoto N, Kawachi H, Chino A, Saito S, Ueno M, et al. Prediction of early colorectal cancer metastasis by machine learning using digital slide images. Comput Methods Programs Biomed. 2019;178:155-61.
31. Wan N, Weinberg D, Liu TY, Niehaus K, Ariazi EA, Delubac D, et al. Machine learning enables detection of early-stage colorectal cancer by whole-genome sequencing of plasma cell-free DNA. BMC Cancer. 2019;19(1):832.
32. Soodejani MT, Mirzaei H, Manesh MM, Tabatabaei SM, Ghaderi A. Incidence of Colorectal Cancer and Adenomatous Polyps After a Two-Step Screening in Isfahan Province, Iran in 2018. J Gastrointest Cancer. 2019:1-5.
33. Pourahmad S, Pourhashemi S, Mohammadianpanah M. Colorectal Cancer Staging Using Three Clustering Methods Based on Preoperative Clinical Findings. Asian Pac J Cancer Prev . 2016;17(2):823-7.
34. Baitharu TR, Pani SK. Analysis of data mining techniques for healthcare decision support system using liver disorder dataset. Procedia Comput Sci . 2016;85:862-70.
35. Kharrat N, Assidi M, Abu-Elmagd M, Pushparaj PN, Alkhaldy A, Arfaoui L, et al. Data mining analysis of human gut microbiota links Fusobacterium spp. with colorectal cancer onset. Bioinformation. 2019;15(6):372.
36. Vijayarani S, Dhayanand S. Data mining classification algorithms for kidney disease prediction. Int J Cybernet Inform. 2015;4(4):13-25.
37. Shoaip N, El-Sappagh S, Barakat S, Elmogy M. Reasoning methodologies in clinical decision support systems: A literature review. U-Healthcare Monitoring Systems: Elsevier; 2019. p. 61-87.
38. Hernández-Julio YF, Prieto-Guevara MJ, Nieto-Bernal W, Meriño-Fuentes I, Guerrero-Avendaño A. Framework for the development of data-driven Mamdani-type fuzzy clinical decision support systems. Diagnostics. 2019;9(2):52.

39. Gorgulu O, Akilli A. Use of fuzzy logic based decision support systems in medicine. Studies on Ethno-Medicine. 2016;10(4):393-403.
40. Hamedan F, Orooji A, Sanadgol H, Sheikhtaheri A. Clinical Decision Support System to Predict Chronic Kidney Disease: A Fuzzy Expert System Approach. Int J Med Inform. 2020:104134.
41. Ambika M, Raghuraman G, SaiRamesh L. Enhanced decision support system to predict and prevent hypertension using computational intelligence techniques. Soft Comput. 2020:1-12.
42. Farzandipour M, Nabovati E, Saeedi S, Fakharian E. Fuzzy decision support systems to diagnose musculoskeletal disorders: A systematic literature review. Comput Methods Programs Biomed. 2018;163:101-9.
43. Pourhoseingholi MA, Kheirian S, Zali MR. Comparison of basic and ensemble data mining methods in predicting 5-year survival of colorectal cancer patients. Acta Inform Med. 2017;25(4):254.
44. Zhang B, Liang X, Gao H, Ye L, Wang Y. Models of logistic regression analysis, support vector machine, and back-propagation neural network based on serum tumor markers in colorectal cancer diagnosis. Genet Mol Res. 2016;15(2):10.4238.
45. Nartowt B, Hart G, Muhammad W, Liang Y, Deng J. A Model of Risk of Colorectal Cancer Tested between Studies: Building Robust Machine Learning Models for Colorectal Cancer Risk Prediction. Int J Radiat Oncol Biol Phys2019;105(1):E132.
46. Domingo MJ, Gerardo BD, Medina RP, editors. Fuzzy decision tree for breast cancer prediction. Proceedings of the International Conference on Advanced Information Science and System; 2019.
47. Hssina B, Merbouha A, Ezzikouri H, Erritali M. A comparative study of decision tree ID3 and C4. 5. Int J Adv Comput Sci Appl. 2014;4(2):13-9.
48. Lakshmi K, Krishna MV, Kumar SP. Performance comparison of data mining techniques for predicting of heart disease survivability. Int J Sci Res Pub. 2013;3(6):1-10.
49. Al-Kasasbeh R, Korenevskiy N, Alshamasin M, Ionescou F, Smith A. Prediction of gastric ulcers based on the change in electrical resistance of acupuncture points using fuzzy logic decision-making. Comput Methods Biomech Biomed Eng. 2013;16(3):302-13.
50. dos Santos LFS, Neves LA, Rozendo GB, Ribeiro MG, do Nascimento MZ, Tosta TAA. Multidimensional and fuzzy sample entropy (SampEnMF) for quantifying H&E histological images of colorectal cancer. Comput Biol Med. 2018;103:148-60.
51. Shafi MA, Rusiman MS. The use of fuzzy linear regression models for tumor size in colorectal cancer in hospital of Malaysia. Appl Math Sci. 2015;9(56):2749-59.