Chapter

Fuzzy Expert System for Rectal Cancer Based on Possibility Measure

Latafat A. Gardashova, Yusif R. Aliyarov and Shamil A. Ahmadov

Abstract

Intestinal infections in common and colorectal cancer in particular are quite widely spread and affect modern population in a significant manner. Therefore, they have been objects of intensive scientific research for quite a long time. It is known that the colorectal cancer's diagnostics can face some difficulties caused by the uncertainties in patients' health status and disease data. The uncertainty, in common, can be classified as probabilistic or possibilistic (fuzzy). The goal of this chapter is to analyze a fuzzy-rule-based medical expert system for the colorectal cancer's diagnostics. In the modeling, fuzzy inference based on possibility measure and knowledge extraction based on fuzzy clustering are applied. During the initial stage of the system's modeling, the applied parameters of colorectal cancer were defined by using clinical data. During the next stage, the soft-computing-based evaluation of the cancer's factors is performed. During the third stage, the applied fuzzy inference, based on possibility measure, is introduced and supported by the examples. The knowledge base of the modeled system consists of the case data obtained from 100 patients in the course of 3 years by the National Center of Oncology. The effectiveness of the modeled system was checked on the testing subset of 30 diagnoses, and 22 predictions by the expert system were defined as correct.

Keywords: colorectal cancer, IF-THEN rules, possibility measure-based inference system, tumor response, fuzzy logic, fuzzy inference, fuzzy clustering

1. Introduction

Cancer has a long and complicated history: it appeared and was recognized even in the ancient times. Modern science has carried out some significant amounts of research on tumors and their treatment. The classic triplet of the disease's treatments is surgery, radiation, and chemotherapy, which are constantly supplemented by more and more advanced methods. Modern oncology has, at its disposal, a wide arsenal of tools and methods for treating cancer: for saving human life, they help to prevent its occurrence and development; in hopeless cases, they prevent the maximum extension and ease the painful symptoms.

Due to the wide spread of oncological diseases, it is especially important to be able to detect cancer at the early stage, when it is more possible to completely heal the patient. Nowadays, cancer is the second leading cause of death in the world, after cardiovascular diseases. Cancer causes almost one in six deaths worldwide. According to the World Health Organization, the incidence of cancer in the next 20 years will increase by 70%. The State Statistical Committee also reports that in Azerbaijan, for every 100,000 people, there are over 400 patients with a malignant tumor; most of them are women. The conducted statistical data demonstrate that in our country, there is an increase in the number of patients diagnosed with cancer. Many experts believe that in a few years, malignant neoplasms will become the main cause of death worldwide, leaving cardiovascular diseases far behind. The worst thing is that the incidence of cancer is growing, but the survival rate is not increasing. In most cases, this occurs because of the late detection of the disease, as success in recovery strongly depends on the early diagnosis of asymptomatic cancer. The problem with the growing number of cancer patients should be solved not only by medicine but also by all sciences that can help in the fight against this cruel disease. This work is specifically aimed at helping oncologists in making an accurate diagnosis at early stages and possibly saving someone's life.

Nowadays, one of the most spread cancer-related infections is colorectal cancer (CRC). The statistics of this illness is studied in [1], and it has been found that CRC should be more investigated among the young generation.

In the other research [2], risk factors that affect development of CRC are analyzed. In the research, the risk for growth of cancer is defined, but patients' gender wasn't taken into consideration. Thus, a more accurate analysis of colorectal cancer is required.

Information about the illness is discussed in [3–9]. The authors used two data-driven approaches: logistic regression and neural network. The effectiveness of logistic regression in the study appeared to be near 66%; the effectiveness of neural approach was 78%. The study was performed on the data obtained from 403 patients. The results demonstrate superior effectiveness of neural networks in comparison with logistic regression when applied to cancer diagnostics. In general, neural networks have several advantages: ability to process vast amounts of information, fault tolerance, generalization ability, adaptability, and learning. In the discussed studies and applied methods, crisp statistic information was used; but data on patients are always rather inaccurate, which enables the applicability of fuzzy data.

There are several research studies on medical expert systems reported in scientific literature [10–14]. These research studies are based on linguistic information, fuzzy inference reasoning, and probability-based reasoning. However, these systems' performance is accompanied by the collateral information loss; thus, these studies possess some effectiveness limits. From this viewpoint, a possibility-measure-based fuzzy inference method seems to be more effective [15–19]. This measure-based algorithm is a kernel of information processing of the software system ESPLAN [20]. Possibility measure is a fuzzy measure and can partially operate Z number-based information. Zadeh's last theory [21] is an extension of fuzzy logic and able to represent different types of information uncertainties. Processing of information based on possibility measure might be quite effective in medicine.

The purpose of this study is to design a fuzzy rule-based expert system for diagnosis of colorectal cancer based on possibility measure and data extracted from Big Data. The rest of the paper is organized as follows. Section 2 briefly describes fuzzy c-means algorithm and the possibility-measure-based inference algorithm. Statement of the problem and its solution are given in Section 3. Finally, Section 4 concludes the paper.

2. Preliminaries

2.1 Representation of fuzzy if: then rules and possibility-measure-based inference algorithm

The framework of the knowledge base relies on the fuzzy interpretation of production rules [20]:

$$R^k : IF\ \mathrm{x_1\ is\ \tilde{A}_{k1}\ and\ x_2\ is\ \tilde{A}_{k2}\ and \ldots and\ x_m\ is\ \tilde{A}_{km}\ THEN}$$
$$\mathrm{y_1\ is\ \tilde{B}_{k1}\ and\ y_2\ is\ \tilde{B}_{k2}\ and \ldots and\ y_{kl}\ is\ \tilde{B}_{kl}}, k = \overline{1, K} \tag{1}$$

where $x_i, i = \overline{1, m}$ and $y_j, j = \overline{1, l}$ are input and output variables, rule antecedents $\tilde{A}_{kj}$ and consequents $\tilde{B}_{kj}$ are defined by using fuzzy sets, and k is the number of rules.

Inputs and outputs of the rule are linguistic data.

Main steps of the applied fuzzy-measure-based reasoning algorithm are as follows:

1. Fuzzification: inputs and outputs are defined in a linguistic manner by using trapezoidal fuzzy numbers.

2. The firing levels of the rules are calculated by using possibility measure (*Poss*):

$$\text{If the sign is " = " and } \lambda_k = \left(1 - Poss\left(\tilde{v}_k | \tilde{a}_{jk}\right)\right) \cdot cf_k, \text{ then} \tag{2}$$

$$\lambda_{jk} = \left(1 - Poss\left(\tilde{v}_k | \tilde{a}_{jk}\right)\right) \cdot cf_k. \tag{3}$$

If the sign is "≠", then *Poss* is determined as

$$Poss(\tilde{v}|\tilde{a})) = \max_y \min(\mu_{\tilde{v}}(y), (\mu_{\tilde{a}}(y)) \in [0, 1], \tau_j = \min\left(\lambda_{jk}\right). \tag{4}$$

Here, one of the main elements of logistic inference is demonstration of the object. Value of each w_i object consists of its linguistic value and the confidence degree of the linguistic value. Together, they constitute a pair, $\left(v_i, cf_i\right)$.

3. For each rule, the following computation is performed

$$R_j = \left(\min{}_j \lambda_{jk}\right) * CF_j / 100 \tag{5}$$

where CF is the confidence degree of a rule, j is the index of a rule, k is the index of the relation, and λ_{jk} is the truth degree of the kth elementary antecedent.

4. This step is to define the firing level (π) and to check $R_j \geq \pi$. If the condition holds true, then the consequent part of the rule is calculated.

5. The evaluated w_i objects have S_i value: w_i, $\left(v_i^1, cf_i^1\right)$, ... , ... , $\left(v_i^{S_i}, cf_i^{S_i}\right)$. S_i is the number of the rules in the fuzzy inference process

6. Calculation of the resulting value by using the fuzzy average value is performed as follows

$$\overline{v}_i = \frac{\sum_{n=1}^{S_i} v_i^n \cdot cf_i^n}{\sum_{n=1}^{S_i} cf_i^n} \tag{6}$$

IF $x_1 = \tilde{a}_1^j$ AND $x_2 = \tilde{a}_2^j$ AND ... THEN $y_1 = \tilde{b}_1^j$ AND $y_2 = \tilde{b}_2^j$ AND ...
IF ... THEN $Y_1 = AVR\,(y_1)$ AND $Y_2 = AVR\,(y_2)$ AND ...
This model has a built-in function, AVRG, which calculates the average value.

2.2 Fuzzy C-means algorithm

Fuzzy C-Means algorithm attempts to minimize the sum of squared errors. The algorithm is based on the iterative minimization of the following objective function [22, 23]:

$$J(W, C) = \sum_{j=1}^{k} \sum_{i=1}^{n} w_{i,j}^p dist(x_i, c_j)^2 \tag{7}$$

The following condition is satisfied for the sum of degrees of membership of a given element xi to all clusters:

$$\sum_{j=1}^{k} w_{i,j} = 1 \tag{8}$$

The following condition is satisfied for the sum of membership degrees of all elements in each cluster:

$$0 < \sum_{i=1}^{n} w_{i,j} < n \tag{9}$$

The corresponding c_j centroid for a C_j cluster is defined as:

$$c_j = \frac{\sum_{i=1}^{n} w_{i,j}^p x_i}{\sum_{i=1}^{n} w_{i,j}^p} \tag{10}$$

The fuzzy partition update formula can be obtained by minimizing the objective function with the constraint that the sum of the weights equals 1:

$$w_{i,j} = \frac{\left(1/dist(x_i, c_j)^2\right)^{\frac{1}{p-1}}}{\sum_{q=1}^{k}\left(1/dist(x_i, c_q)^2\right)^{\frac{1}{p-1}}} \tag{11}$$

3. Problem description and possible solution

Nowadays, as was previously mentioned, CRC is the second most frequent malignancy in the case of both men and women. The tumor is localized in the rectum, the farthest portion of the digestive tract, in around one-third of cases. Patients' quality of life is severely impacted by surgical therapy for rectal cancer, which results in dysfunctions that include fecal incontinence, urinary problems, and sexual issues. Increasing the number of sphincter-preserving surgeries was the primary goal of rectal cancer treatment for the past 20 years; today, the focus is on organ preservation strategies.

Neoadjuvant radiotherapy and chemotherapy are used to increase local control and overall survival in around 75–80% of instances for individuals with rectal cancer, making their care complicated. However, only 20–30% of patients who receive neoadjuvant therapy gets a full pathological response; the other 70–80% experiences a poor or ineffective response. How we can anticipate tumor response in people with rectal cancer is the key issue right now.

This variance is assumed to be influenced by the tumor's size, height from the anal margin, depth of invasion, differentiation and, of course, genetic variables such as the KRAS and BRAF mutation. The purpose of this work is to create an expert system to forecast tumor response value following neoadjuvant chemoradiation. A crucial issue is how to define the predict value of the tumor response following neoadjuvant chemoradiation. To assess tumor response using pertinent parameters is the fundamental problem. Using fuzzy rules, we can calculate the tumor response value. A compound index made up of five characteristics, each of which is judged by an expert, makes up the tumor response value following neoadjuvant chemoradiation, abbreviated as R.

The five components are: V—age, LO—localization of tumors, T—infestation rate, N—state of lymph nodes, G—mutation in the genes, and R—predict value of tumor response.

Using the abovementioned parameters, the tumor response value model can be expressed as a set of 21 rules obtained by using fuzzy C-means algorithm:

1. *If Age(V) = about 80 and Localization tumor(LO) = about 2 and Infestation rate(T) = about 2 and State of lymph nodes(N) = about 1 and Mutation in the genes(G) = about 0, THEN predict value of tumor response (R) = about 0;*

2. *If Age(V) = about 40 and Localization tumor(LO) = about 2 and Infestation rate (T) = about 2 and State of lymph nodes(N) = about 3 and Mutation in the genes (G) = about 0, THEN predict value of tumor response (R) = about 0;*

   … … … …

3. *If Age(V) = about 20 and Localization tumor(LO) = about 92 and Infestation rate(T) = about 2 and State of lymph nodes(N) = about 3 and Mutation in the genes(G) = about 1, THEN predict value of tumor response (R) = about 100*

Our goal is to use five parameters represented by fuzzy linguistic terms to describe the level of the tumor response value. Values of linguistic terms are given below as intervals. In addition, they are expressed as fuzzy data in diagrams.

Age (V):
Positive [0–40]
Medium positive [30–50]
Medium negative [40–60]

Negative [60–100]
Localization of tumors (LO):
Negative [0–5]
Weak negative [3–8]
Medium positive [6–11]
Positive [10–15]
Infestation rate (T):
Early form [0–1]
Localized [1 – 2]
Early locally developed [2 – 3]
Late locally developed [3 – 4]
Metastasized [4–4<]
Status of lymph nodes (N):
Negative [0–2]
Medium Positive [1–4]
Positive [4–20]
Mutation in genes (G):
Negative [0–0.5]
Positive [0.5–1]
TRG (tumor regression grade) (R):
Very bad [0–10]
Bad [10–20]
Sufficient [50–60]
Good [70–80]
Excellent [90–100]

Graphical representation of these linguistic terms is as shown (**Figures 1**–**6**):

From the defined linguistic terms, a knowledge base of interpretable rules is created. For instance:

Rule. *If age is about 35 and localization of tumors is middle positive and infestation rate is middle positive and state of lymph nodes is middle positive and mutation in the genes is positive, THEN predict value of tumor response is bad.*

ESPLAN shell is used for creating an expert system for rectal cancer. Below, the computer simulation is discussed.

Computer simulation. ESPLAN shell has the following modules: the module that manages all the procedure of the system; read-in and interpretation of knowledge; inference; explanation generator; knowledge base and work are service; environment interface; user interface. ESPLAN shell is realized using Prolog Artificial Intelligence language. There are functional constructions that Prolog predicates in this system. This possibility of system gives it a chance for including new functions to the program.

Representation objects and linguistic terms by using ESPLAN are given in **Figure 7**.

For example: Parameters about 50 are represented as: *about K: (D, K,1.1*K, D) = about 50:(D = 7.5, K = 47.5, 1.1 K = 52.5, D = 7.5).*

Given linguistic terms are used in ESPLAN system:
much: (D, F – D, F, 0);
more than a: (D, K + D, F, 0);
about K: (D, K,1.1*K, D);
neutral: (D, M + 2 * D, M + 3*D, D);
less than K: (0, M, K – D, D).

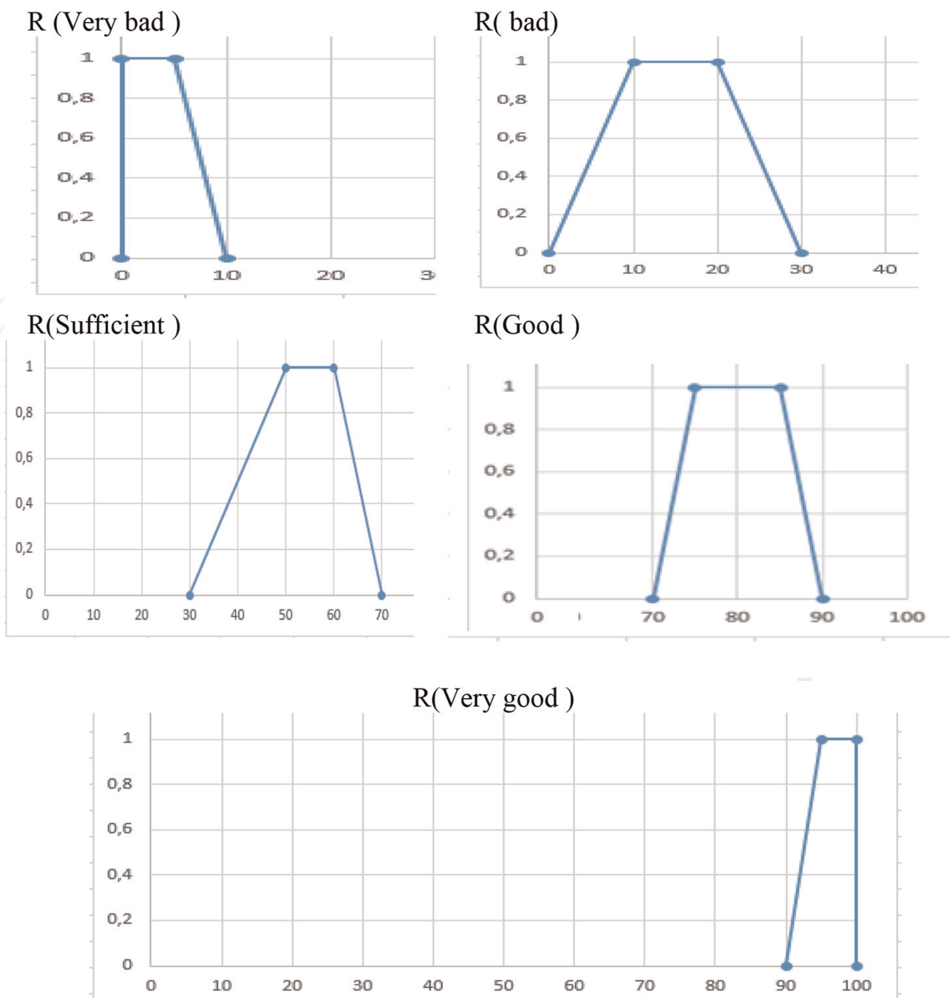

Figure 1.
Linguistic terms for Predict value of tumor response(R).

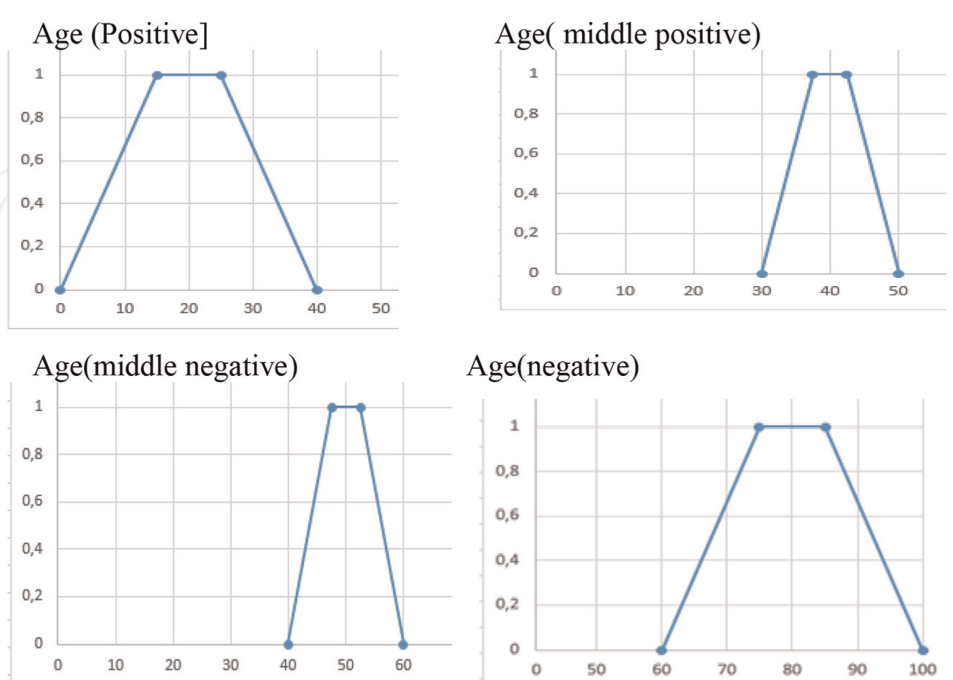

Figure 2.
Linguistic terms for Age(V).

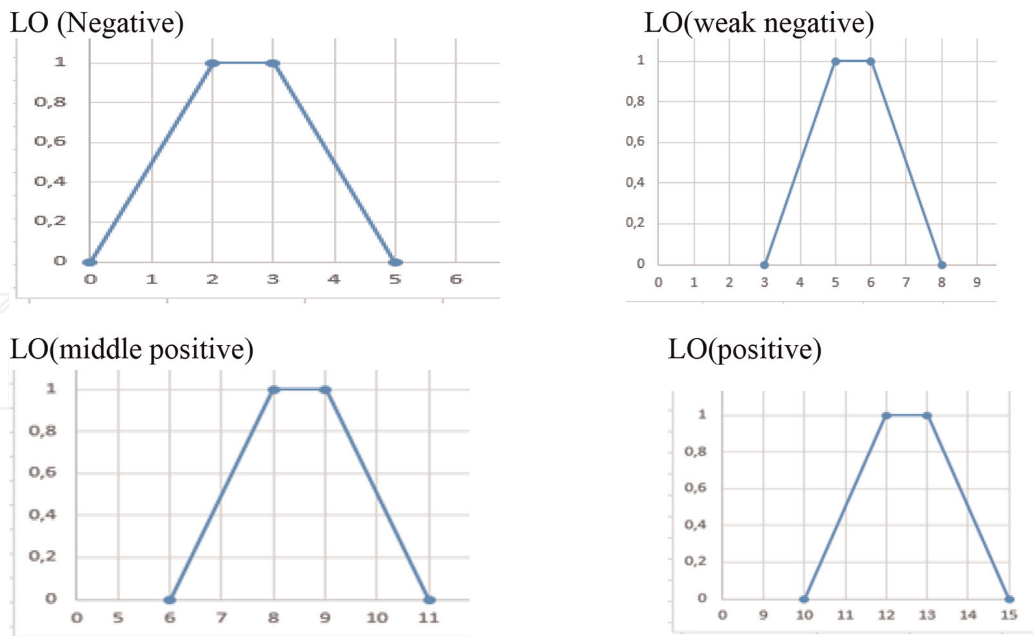

Figure 3.
Linguistic terms for Localization of tumors(LO).

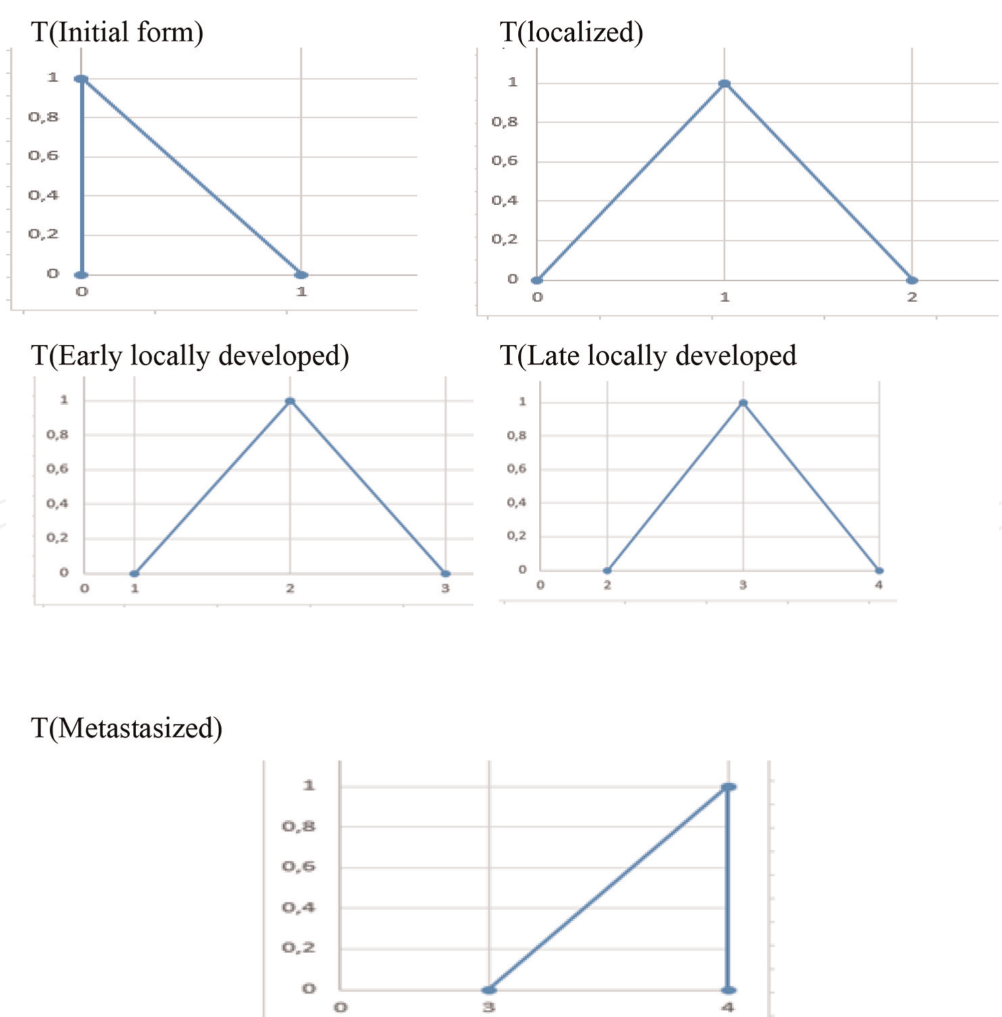

Figure 4.
Linguistic terms for Infestation rate(T).

DOI: http://dx.doi.org/10.5772/intechopen.109405

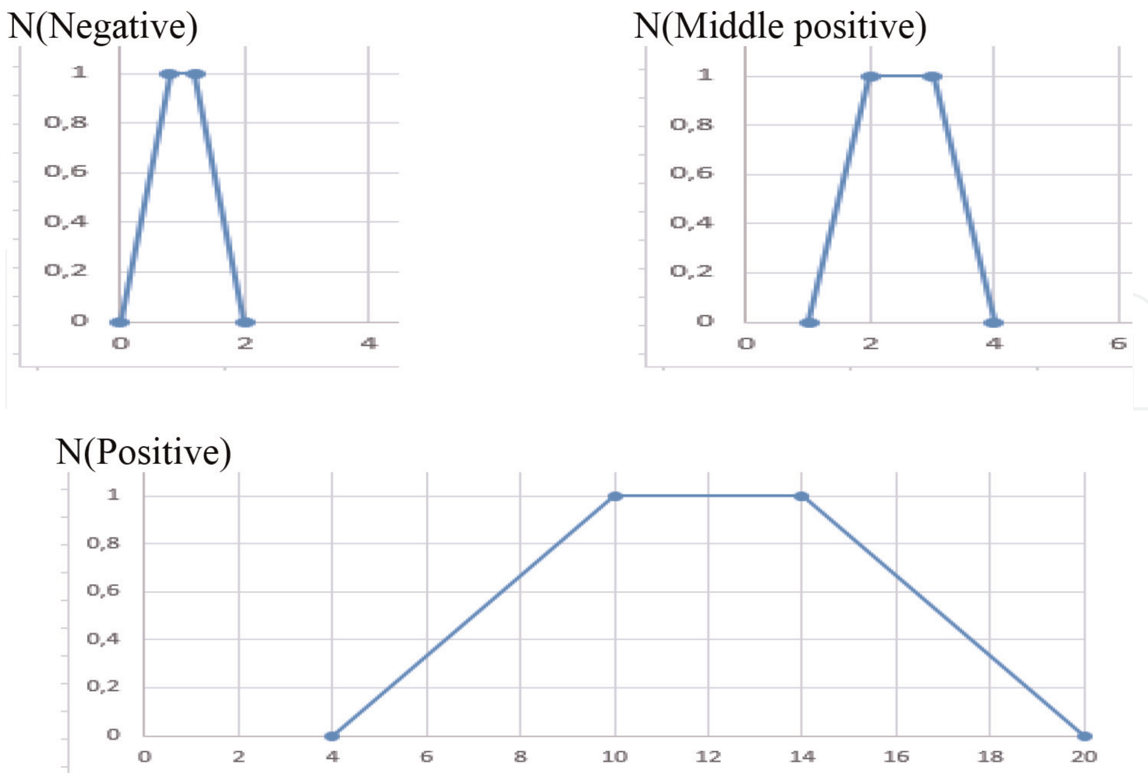

Figure 5.
Linguistic terms for State of Lymph nodes(N).

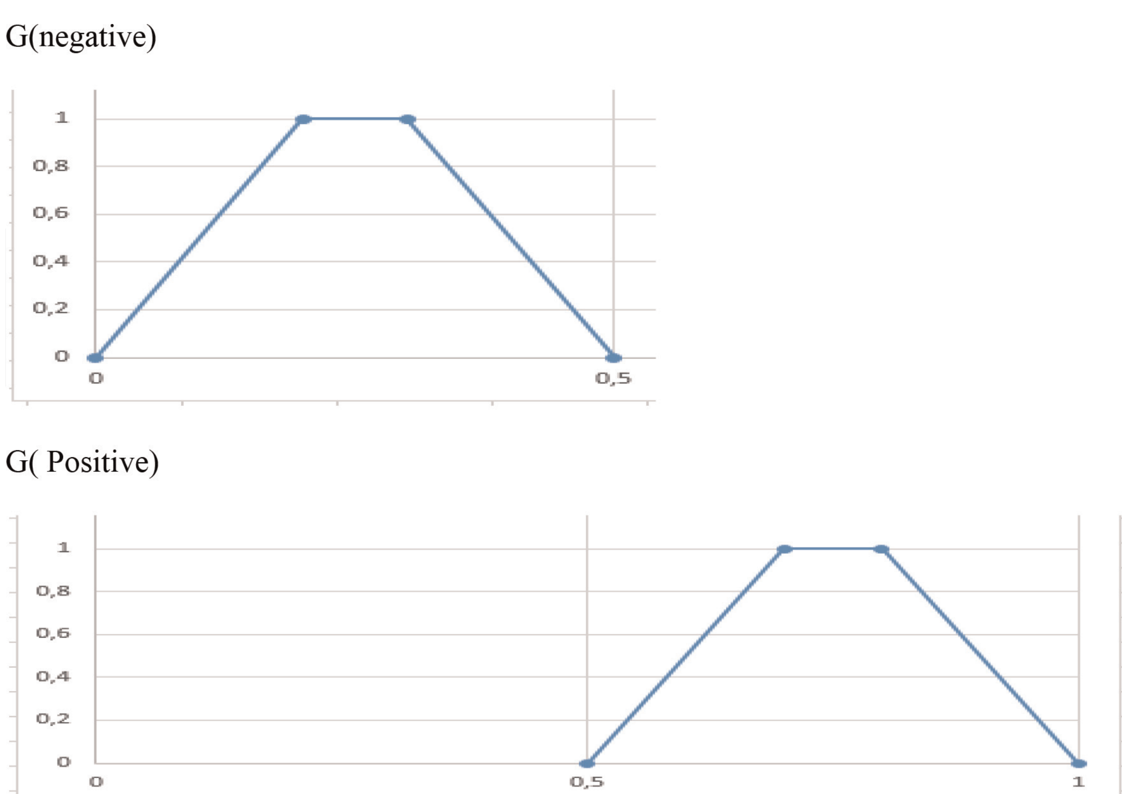

Figure 6.
Linguistic terms for Mutation in genes(G).

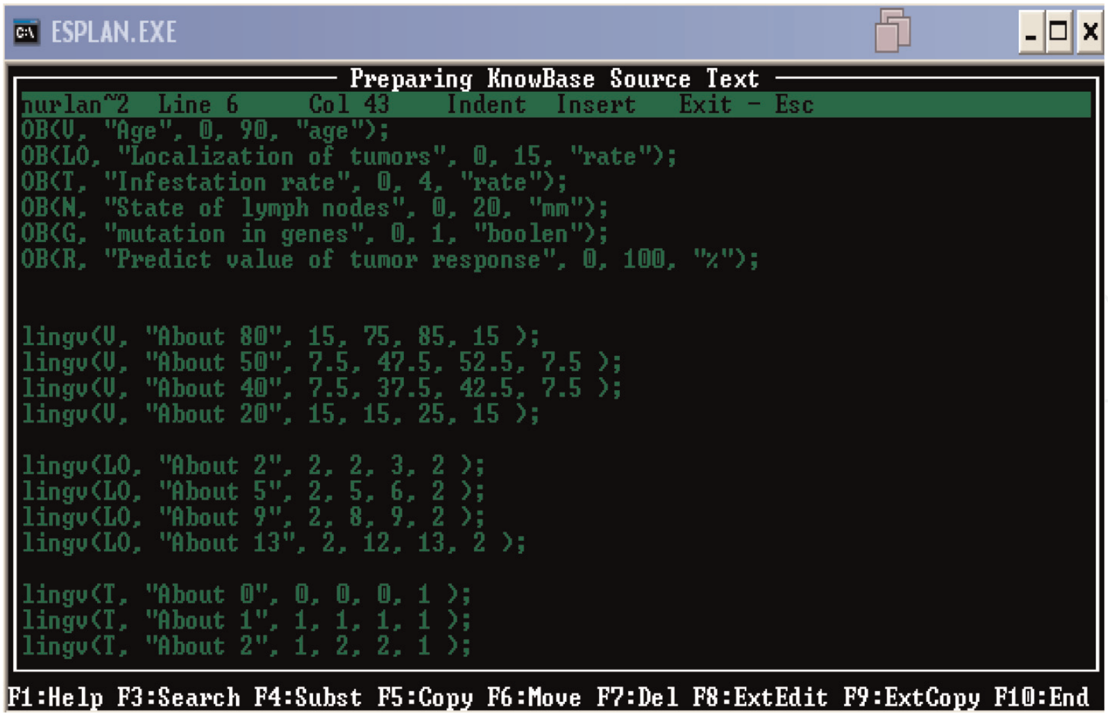

Figure 7.
Values of parameters for tumor response.

Here, minimum value is M, and the maximum value of the universe is F; D = (F-M)/5.

Demonstration of the rule is given below (**Figure 8**):

Fragment of the knowledge translation process is represented in **Figure 9**, and Fuzzy inference process is in **Figures 10–13**.

For instance,

object= “localization of tumors”,

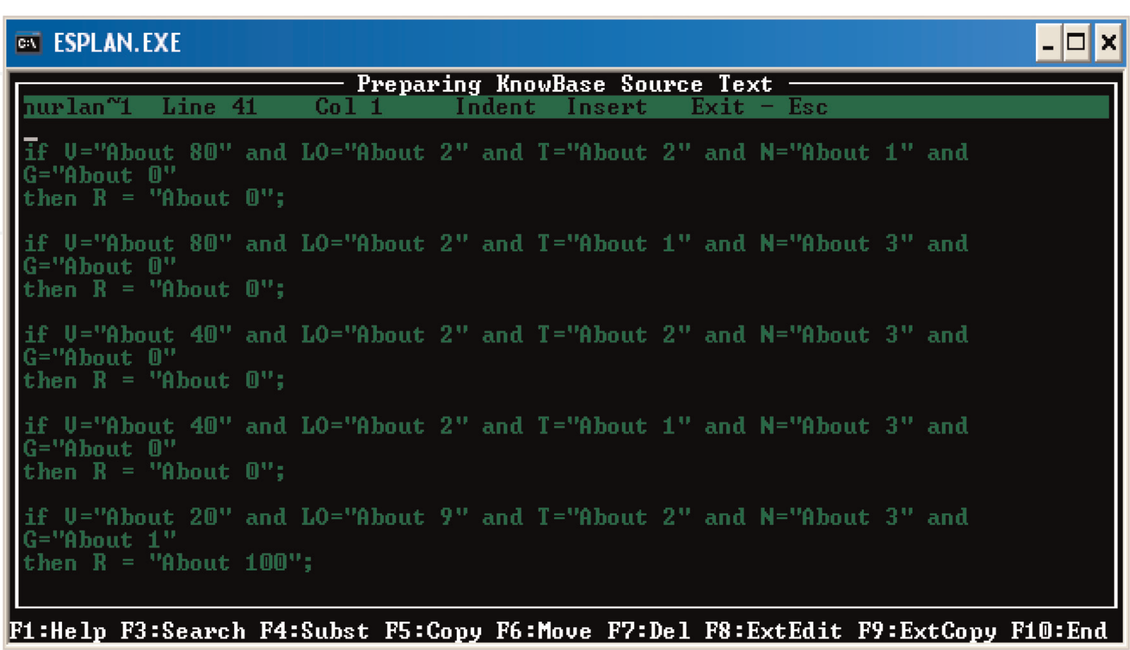

Figure 8.
Representation of the rule.

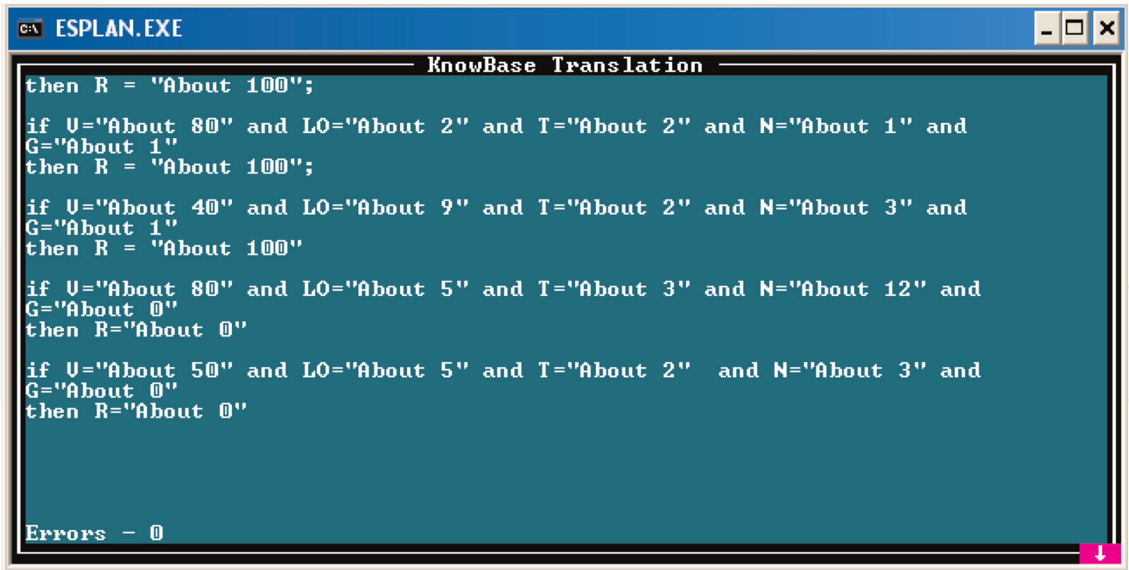

Figure 9.
Knowledge translation process.

M = minimum = 0, F = maximum = 15,

linguistic term = "about 5": About 5 = (D, M + 2*D, M + 3*D, D)

The abovementioned model is created using knowledge representation language and implemented in ESPLAN shell. Results of the performed test are:

Test 1: If age = about 38 and localization of tumors = about 4 and infestation rate = about 2 and state of lymph nodes = abou11 and mutation in the genes = about 0, THEN predict value of tumor response =?

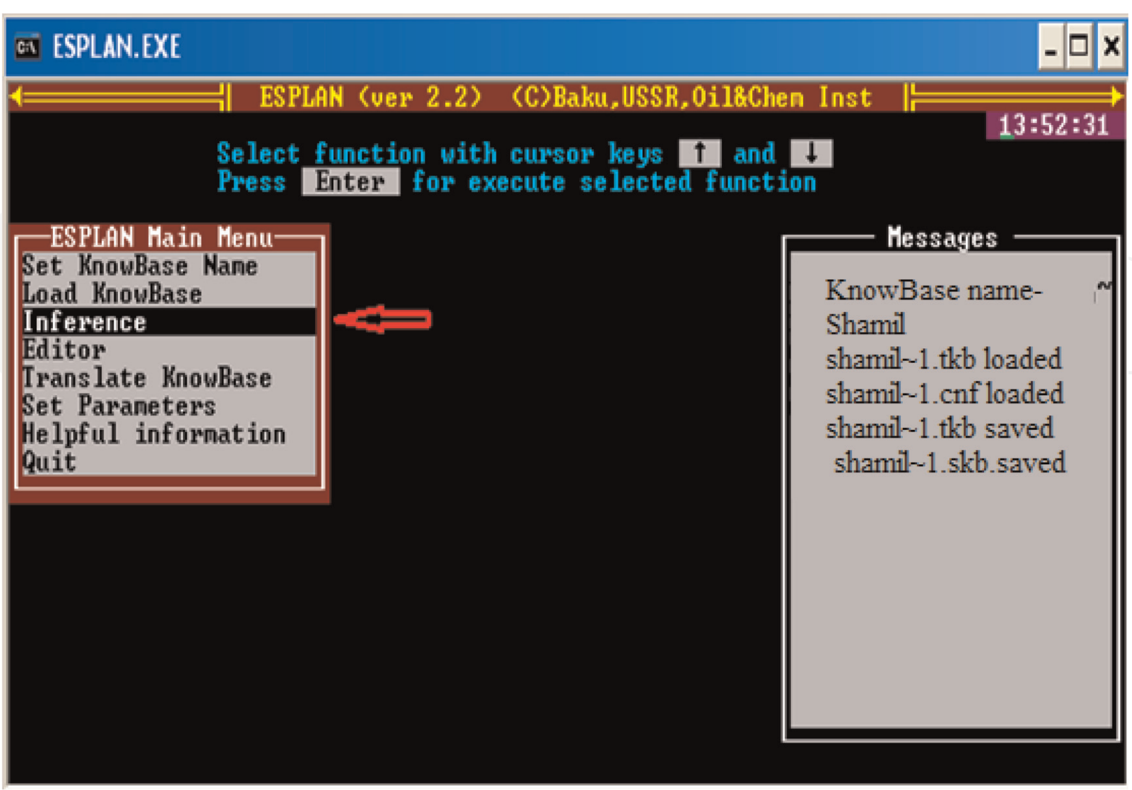

Figure 10.
Fuzzy inference process (1).

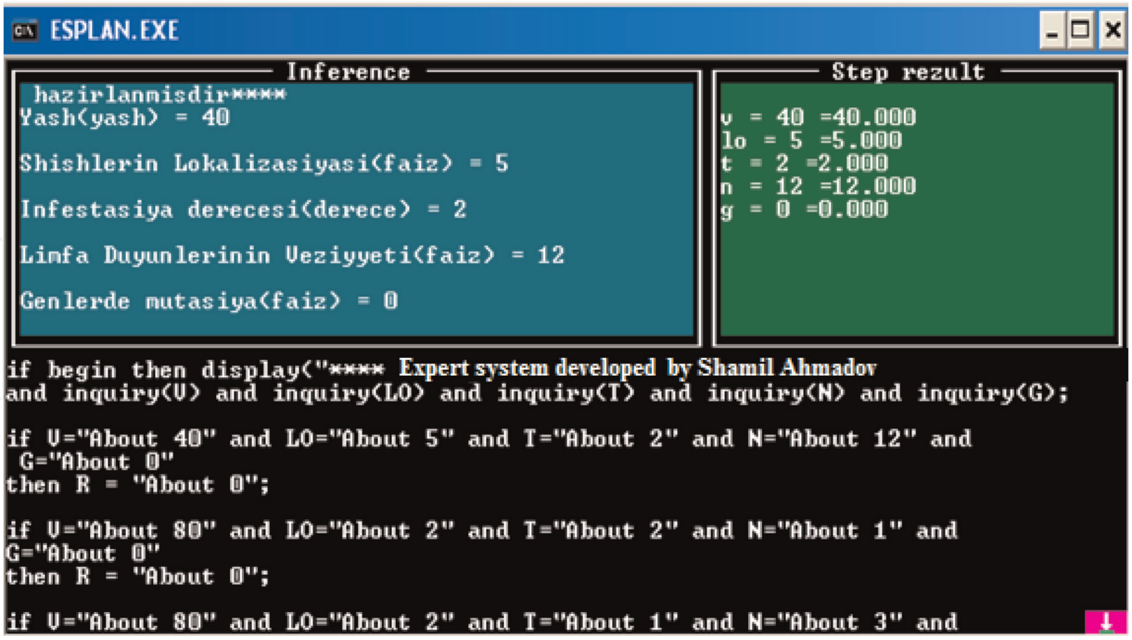

Figure 11.
Fuzzy inference process (2).

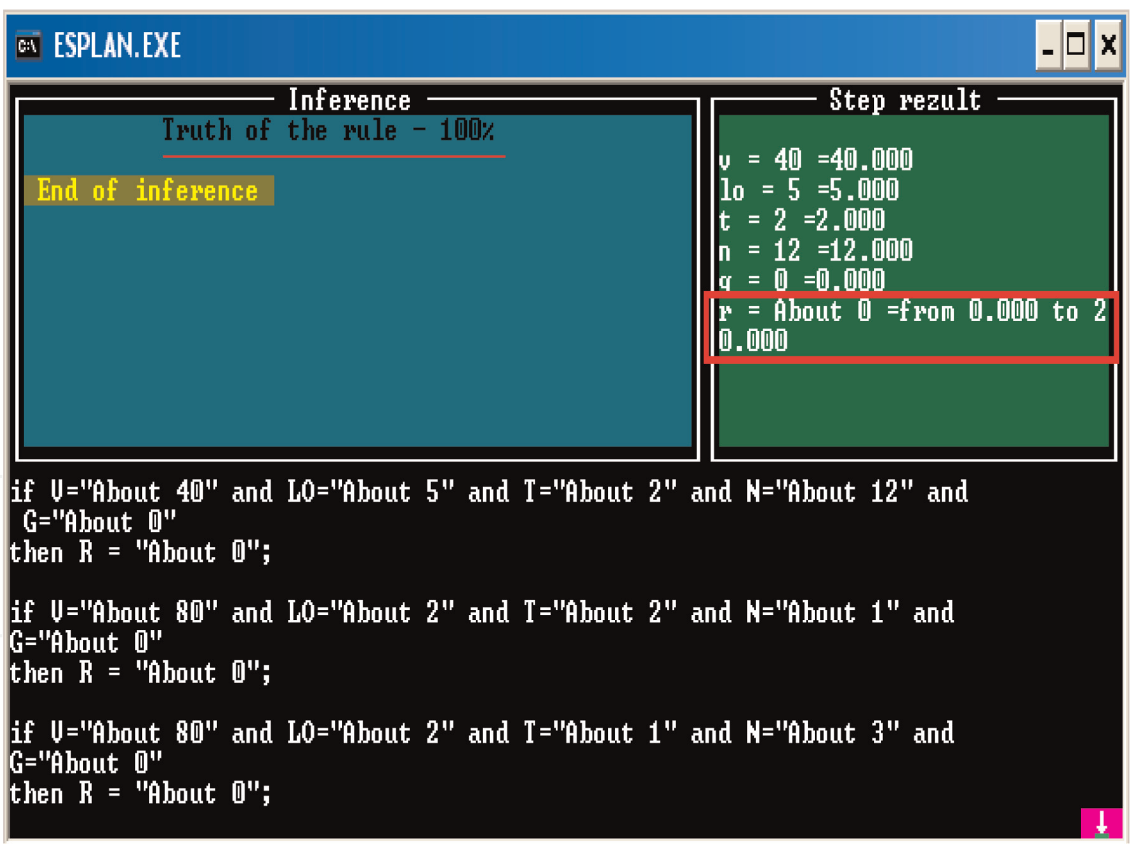

Figure 12.
Fuzzy inference result.

Result: predict value of tumor response is from 0 to 2.

The results of tests are shown in **Figures 12** and **13**.

Test 2: *Test 1: If age = middle positive and localization of tumors = weak negative and infestation rate = middle positive and state of lymph nodes = positive and mutation in the genes = negative, THEN predict value of tumor response =?*

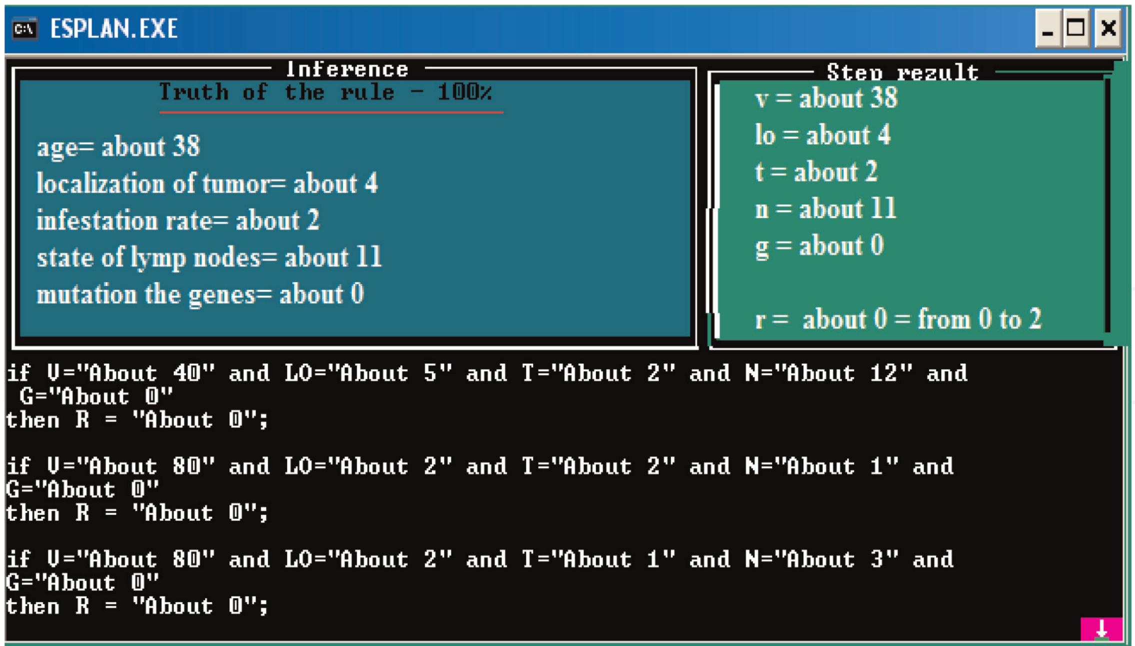

Figure 13.
Fragment of results.

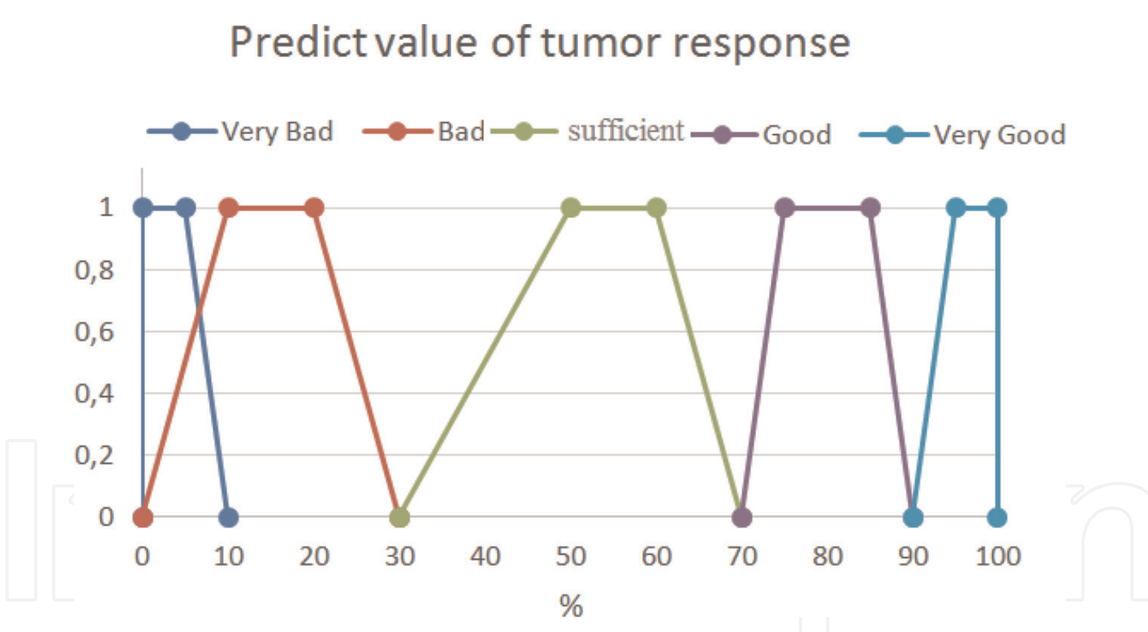

Figure 14.
Linguistic terms of tumor response.

FOR TEST 2. ANSWER:

predict value of tumor response =? ***very good.***

Representation of linguistic terms of tumor response is given in **Figure 14**.

Knowledge base is realized using possibility measure-based algorithm in ESPLAN. In this system, fuzzy logic theory is used in demonstration and operation of the linguistic terms. After fulfilling of the knowledge base of the system by adding values of several objects (for example, age is about 50), searching the solution in knowledge base is done by logical inference procedure.

There is the following opportunity of the ESPLAN shell: creating an expert system for several applications, relation with applied software package, explanation of the advices, demonstration of the results, interface with user, and so on.

The advantages of the created expert system are: working with linguistic values; possibility measure-based reasoning; realization of composition rule of inference, including knowledge base as dialog; storing knowledge about different areas in the knowledge base of the system; setting of a confidence degree for any rule (in percentage); application of external programs; and data interchange by using a file system.

4. Conclusion

The data used in this study are from the Database of National Center of Oncology. Three years of case data of 100 patients are implemented for extracting the knowledge-based rule using clustering method. Veracity of 30 diagnoses of patients was checked, and 22 from them were defined as correct. In this chapter, for the evaluation of value of tumor response, a possibility-measure-based method is used. The created expert system for rectal cancer was implemented in the ESPLAN. Several tests were performed, and the outcomes were compared to the actual patient data.

The presentation of the developed system and samples of its use in medicine demonstrate that it has a wide range of potential capabilities for making decisions based on fuzzy information under uncertain circumstances. Experimental findings demonstrate the effectiveness of the proposed intelligent system.

In the future, we are planning to study and compare different types of cancer illnesses by using soft computing tools-neural network, genetic algorithm, evolutionary computing, chaos theory, Zadeh last theory-Z-number theory, and real-life results for giving help and advice to doctors for decision-making during the treatment process. For the future works, the data that have been used in computations will be gathered from different hospitals and centers of oncology from all over the world by using internet resources.

Conflict of interest

The authors declare no conflict of interest.

Author details

Latafat A. Gardashova[1]*, Yusif R. Aliyarov[2] and Shamil A. Ahmadov[1,3]

1 Azerbaijan State Oil and Industry University, Baku, Azerbaijan

2 National Center of Oncology, Azerbaijan

3 French-Azerbaijani University, Baku, Azerbaijan

*Address all correspondence to: latsham@yandex.ru

References

[1] Siegel RL, Miller KD, Goding Sauer A, Fedewa SA, Butterly LF, Anderson JC, et al. Colorectal cancer statistics, 2020. CA: a Cancer Journal for Clinicians. 2020;**70**:145-164

[2] Amersi F, Agustin M, Ko CY. Colorectal cancer: Epidemiology, risk factors, and health services. Clinics in Colon and Rectal Surgery. 2005;**18**(3): 133-140

[3] Grumett S, Snow P, Kerr D. Neural networks in the prediction of survival in patients with colorectal cancer. Clinical Colorectal Cancer. 2003;**2**(4): 239-244

[4] Phan K, Kahlaee HR, Kim SH, Toh JWT. Laparoscopic vs. robotic rectal cancer surgery and the effect on conversion rates: a meta-analysis of randomized controlled trials and propensity-score-matched studies. Techniques in Coloproctology. 2019; **23**(3):221-230

[5] Araghi M, Soerjomataram I, Jenkins M, Brierley J, Morris E, Bray F, et al. Global trends in colorectal cancer mortality: Projections to the year 2035. International Journal of Cancer. 2019; **144**(12):2992-3000

[6] Bartos A, Bartos D, Stoian R, Szabo B, Cioltean C, Iancu I, et al. Short-term outcome and survival after multiorgan resection for locally advanced colo-rectal cancer. Identification of risk factors. Annali Italiani di Chirurgia. 2018;**89**: 229-236

[7] Tharin Z, Blanc J, Alaoui IC, Bertaut A, Ghiringhelli F. Influence of primary tumor location and resection on survival in metastatic colorectal cancer. World Journal of Gastrointestinal Oncology. 2020;**12**(11):1296-1310

[8] Ngan TT, Lan LTH, Tuan TM, Son LH, Tuan LM, Minh NH. Colorectal cancer diagnosis with complex fuzzy inference system. In: Satapathy S, Bhateja V, Nguyen B, Nguyen N, Le DN, editors. Frontiers in Intelligent Computing: Theory and Applications. Advances in Intelligent Systems and Computing. Vol. 1013. Singapore: Springer; 2020. DOI: 10.1007/978-981-32-9186-7_2

[9] Silva A, Oliveira T, Julian V, Neves J, Novais P. A mobile and evolving tool to predict colorectal cancer survivability. In: IFIP International Conference on Artificial Intelligence Applications and Innovations, Sep 2016. Springer, Cham. 2016. pp. 14-26

[10] Amato F, López A, María E, Peña-Méndez PV, Hampl A, Havel J. Artificial neural networks in medical diagnosis. Journal of Applied Biomedicine. 2013;**11**: 47-58

[11] Echeverría FR, Echeverría CR. Application of expert systems in medicine. In: Proceedings of the 2006 Conference on Artificial Intelligence Research and Development. Amsterdam, The Netherlands: IOS Press; 2006. pp. 3-4

[12] De Medeiros IB, Machado MAS, Damasceno WJ, Caldeira AM, dos Santos RC, da Silva Filho JB. A fuzzy inference system to support medical diagnosis in real time. Procedia Computer Science. 2017;**122**:167-173

[13] Fathy SK. A predication survival model for colorectal cancer. In: Proceedings of the 2011 American Conference on Applied Mathematics and the 5th WSEAS International Conference on Computer Engineering and

Applications, Jan 2011. World Scientific and Engineering Academy and Society (WSEAS). 2011. pp. 36-42

[14] Prasath V, Lakshmi N, Nathiya M, Bharathan N, Neetha P. A survey on the applications of fuzzy logic in medical diagnosis. International Journal of Scientific and Engineering Research. 2013;**4**(4):1199-1203

[15] Aliev RA, Aliev BF, Gardashova LA, Huseynov OH. Selection of an optimal treatment method for acute periodontitis disease. Journal of Medical Systems. 2012;**36**(2):639-646. DOI: 10.1007/s10916-010-9528-6

[16] Gardashova LA, Allahverdiyev RA, Saner T, Eyupoglu SZ. Analysis of the job satisfaction index problem by using fuzzy inference. Procedia Computer Science. 2016;**102**:45-50. DOI: 10.1016/j.procs.2016.09.368

[17] Mirzakhanov VE, Gardashova LA. Modification of the Wu-Mendel approach for linguistic summarization using IF-THEN rules. Journal of Experimental & Theoretical Artificial Intelligence. 2019;**31**(1):77-97. DOI: 10.1080/0952813X.2018.1518998

[18] Gardashova LA, Aliyev BF. Dental disease detection using Fuzzy logic. In: Advances in Intelligent Systems and Computing Series, ICAFS-2020. Vol. 1306. Springer. pp. 126-132. Available from: https://www.springer.com/gp/book/9783030640576

[19] Aliev BF, Gardashova LA. Selection of an optimal treatment method for acute pulpitis disease. Procedia Computer Science. 2017;**120**:539-546

[20] Aliev RA, Aliev RR. Soft Computing and its Application. New Jersey, London, Singapore, Hong Kong: World Scientific; 2001. p. 444

[21] Zadeh LA. A note on Z-numbers. Information Science. 2011;**181**(14):2923-2932. DOI: 10.1016/j.ins.2011.02.022

[22] Dunn JC. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. Journal of Cybernetics. 1973;**3**:32-57. DOI: 10.1080/01969727308546046

[23] Bezdek JC, Ehrlich R, Full W. FCM: The fuzzy c-means clustering algorithm. Computers & Geosciences. 1984;**10**(2-3):191-203. DOI: 10.1016/0098-3004(84)90020-7