Evaluation of breast cancer risk by using fuzzy logic

M. Caramihai*, I. Severin*, A. Blidaru**, H. Balan***, C. Saptefrati**
*POLITEHNICA University, Sp. Independentei, 313
Bucharest, ROMANIA
m.caramihai@ieee.org, irina_severin2003@yahoo.com
**Oncology Institute "Alex. O. Trestioreanu", Sos. Fundeni, 252
Bucharest, ROMANIA
alexandrublidaru@yahoo.com, crisoana@hotmail.com
***University of Medicine "C. Davila", Str. Dionisie Lupu, 56
Bucharest, ROMANIA,
drhoriabalan@yahoo.com

Abstract: - The decisional process of the best suited follow-up treatment for a breast cancer case is strongly connected to the correct diagnosis of the breast cancer risk. Though there are plenty of experienced doctors, top range imaginary devices and advanced radiological techniques, the methods and criteria to diagnose/quantify the detected lesion characteristics, to appreciate the breast cancer development stage and to finally conclude with a certified (most probable) risk diagnostic are still subjectively and poorly defined for most of the medical stuff. The present paper introduces a set of fuzzy rules that can process the relevant data of breast cancer cases in order to give a breast cancer risk prognostic factor qualitatively comparable to the one the expert gives.

Key-Words: - Breast Cancer, Fuzzy Techniques, Computer Aided Diagnostic, Risk evaluation, Lesion classification

1 Introduction

Breast cancer represents a major health problem at world level representing the first deadly cause for women. Cancer prevention through consequent screening programs, early discovery and timely, improved and diversified means of treatment are usually the most successful ways to reduce mortality.

During the last decade, the imaging devices registered a substantial rate of progress, triggering a routine screening that goes as fast and as accurate as possible in detecting lesion characteristics. A large range of technologies and instruments evolved based on X-rays analysis, ultrasound evaluations and magnetic resonance techniques, among which mammography, echography and the magnetic resonance imaging offer the best qualitative results in performance and input-output ratio [1].

Unfortunately, the existence of such advanced imagery instruments isn't enough. In addition to it, the correct assessment and classification of the detected characteristics, followed by a judicious decisional procedure of the breast cancer diagnostic is needed in order to apply, in a timely manner, the treatment with the most positive effect/response. If [1, 2, 3] concentrates on the selection and assessment procedures of the relevant lesion characteristics, a computer aided diagnosis procedure hasn't yet been developed.

The detection and the assessment of the characteristics for a breast cancer lesion is nowadays possible [2, 3] due to automatic software applications especially designed to analyze imagery results, like [4]. These CAD features are used to extract, quantify and to store the numerical values of the relevant characteristics in a reliable database [2].

Herewith, due to the fact that medical oncologic experts make the diagnostic decision for a breast cancer case based on past professional experience and knowledge, the intelligent techniques are the only one powerful enough to emulate the expert's choice. Due to their great behavior in the presence of noise, imprecision and incertitude, these techniques can obtain better results than the classical ones.

Created to substitute human-like, natural biological, non-linear thinking in the computerized world, intelligent techniques are the most advanced modeling techniques that can evaluate and decide based on an inference process that is similar to human thinking and judging. The most largely applied intelligent techniques [5, 6 ,7] are the fuzzy techniques, the neural network techniques, the genetic algorithms and the knowledge based systems (known also as expert systems).

The aim of this paper is to describe an intelligent procedure based on the fuzzy technique that is used for the evaluation of breast cancer risk. This diagnostic implies the definition of a set of judiciously chosen intelligent fuzzy rules concerning relevant personal patient data and evaluated cancer tumor classification.

2 Methods

2.1 Fuzzy intelligent technique

The high performance of the fuzzy intelligent technique has been demonstrated during the last recent years in various applications around the world. Its excellent results and easy implementation shown great progress and registered lots of improvements in devices/techniques from different fields of science like automated control (temperature control, control of the tube speed, auto-focus control for video cameras), form recognition (concerning fuzzy classification algorithms), measurement (processing of sensors information), medicine (cardiac stimulators control), economy (fuzzy decisional methods), cognitive psychology (fuzzy modeling of the human sight system), etc.

The greatest advantage of using fuzzy logic lies in the fact that scientists can model non-linear, imprecise, complex systems by transposing human experience, knowledge and practice in inference (or fuzzy) rules that use linguistic (or fuzzy) variables.

In other words, the definition of the fuzzy variables and rules doesn't start from the precise definition of a process, but from the observation and in-depth understanding of the components of a physical phenomenon. By using the fuzzy intelligent technique, the human experience related to the investigated process can materialize itself through new rules of deduction attached to the fuzzy logic.

Basically, the fuzzy logic procedure implies three steps, [5]:

1. Determination of a set of fuzzy linguistic variables, set A, for each of the input or output system variables that describe the observed phenomenon,
2. Definition of a set of fuzzy inference rules between input (I) and output (O) fuzzy variables (such as IF I1 is x1 AND I2 is x2 … THEN O1 is z1 …, where x1, x2,…, y1, y2,…, z1,z2,… are linguistic values),
3. Definition of a membership function for each fuzzy variable: μ(e): U -> [0,1], where 'μ(e)' is the membership function of the 'e' (can be any input or output variable) fuzzy variable and U is the variation interval for the same variable (see figure 1).

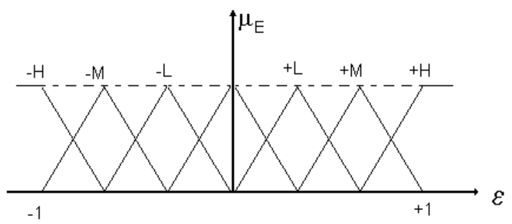


Fig.1 Graphical representation of a set of rules for error E, using 7 linguistic variables -H, -M, -L, 0, +L, +M, +H that form the set A.

The fuzzification and the defuzzification are the two bounds between the fuzzy logic system and the measured phenomenon data. Fuzzification is used for the transformation of a real value x□R (acquired from the studied process) in one of the fuzzy values from a fuzzy set for a specific fuzzy variable.

Defuzzification is the inverted process that transforms the output fuzzy variable (computed by the set of inference rules between the input variables) in a crisp, real value x□R (normally sent back to the process, as control feedback).

Herewith, there are multiple criteria for determining the fuzzification or the defuzzification process like singleton, triangular, trapezoidal or Gaussian transformation, [5] (figure 1 uses the triangular transformation).

2.2 Breast cancer lesion classification

The diagnosis procedure of breast cancer consists in the appropriate observation of the imagery results in search for characteristics that fully cater (from the expert oncologist opinion) for the cancerous development state of the detected lesion, [8].

Many screening cases have been studied and analyzed and a set of characteristics has been chosen as best defining a breast cancer tumor (lesion contrast, lesion differentiation, presence of spiculations, of angulations, existence of posterior shadow, margins, calcifications, fragmentation, etc.). Based on these characteristics oncologists evaluate the state of the breast cancer and classify it on the internationally recognized BI-RADS scale.

BI-RADS (Breast Imaging-Reporting and Data System) is an unitary system designed for helping medical professionals assess, interpret and classify mammographies, echographies and magnetic resonance imaging in a concise and unambiguous and standardized way, [9], by assigning numbers or numerical codes to different risk categories. The Assessment Categories are numbered from 0 to 5 describing with 0: Incomplete, 1: Negative, 2: Benign finding(s), 3: Probably benign, 4: Suspicious abnormality and 5: Highly suggestive of malignancy.

2.3 Potential risk factors

Scientists recognize the next seven broad categories of risk factors that predispose women for developing breast cancer:

- age,

- family history of breast cancer (mother or sister),
- hormonal factors (early onset of menstruation or late menopause),
- proliferative (reproductive) breast disease (for women who had no children – had no breastfed -, or had children after age 30),
- irradiation of the breast region at early ages,
- personal history of malignancy and
- lifestyle factors.

Though these factors and their complex interactions can't be determined in more than 70% of breast cancer cases, [10], studies shown that high-fat diets, obesity, or use of alcohol also contribute to a woman's risk profile. Additionally, according to epidemiologic studies, there is an increase in breast cancer incidence in women with a higher level of education or higher socio-economic status, possibly related to delayed child-bearing and lower parity, [11].

3 Results

3.1 Inputs and Outputs of the Fuzzy Logic

The construction of a fuzzy logic with more than two input variables extremely raises the complexity of the calculus and of the design definition. For simplicity, the present paper discusses the fuzzy logic with only two input variables and one output variable.

In this order, the fuzzy logic that evaluates the breast cancer risk (seen as permanent system output) takes in account the possible input variables from Table 1. These input variables, together with their relevance factor (in what matters the decisional process of the breast cancer risk), were decided based on the experience of medical staff.

Table 1: Possible fuzzy logic input variables

Input variable	*Relevance factor*
Age	6%
Age of first menstruation	6%
Number of invaded axilary nodes	9%
BI-RADS score	24%

As will be seen below, the two selected input variables for the fuzzy system will be always chosen considering the criterion of relevance domination. In this matter, because the BI-RADS score has the highest relevance factor, it will always be considered one of the two input variables.

The numeric variation interval for each of the input variables can be seen in Table 2. The chosen inference method in breast cancer cases, the invaded axilary nodes is a measure of the cancer infiltration in adjacent tissues. When benign tumors or incipient cancers don't or still didn't invade the lymph axilary nodes it means that tumor excision would be safer and it would usually have a lower risk of local cancer recurrence. The higher the number of infiltrated lymph axilary nodes, the higher the risk and the aggression of cancer, [12] may be noticed. In our procedure, the maximal number of possible infiltrated axilary nodes was considered to be 15.

Table 2: Numeric variation interval for input variables

Input variable	*Min value*	*Max value*
Age	10	70
Age of first menstruation	10	15
Number of invaded axilary nodes	0	15
BI-RADS score	0	5
Output variable		
Breast cancer risk	0%	50%

The breast cancer risk factor is the output of all considered fuzzy systems. The variation interval for the output variable can be seen also in Table 2, 0% corresponding to null breast cancer risk, while 50% corresponds to a high breast cancer risk.

3.2 Fuzzy Systems

The means and technique of developing a fuzzy system was described in the METHODS section and works similar in all of the researched cases. This paper presents the results of only one of the three studied fuzzy systems that have as inputs and outputs the variables contained in Table 3.

Table 3: Studied fuzzy systems

Fuzzy systems	*Input variables (I)*	*Output variables (O)*
FZ1	**I1**: Age **I2**: BI-RADS score	**O1**: Breast cancer risk
FZ2	**I1**: Age of first menstruation **I2**: BI-RADS score	**O1**: Breast cancer risk
FZ3	**I1**: Number of invaded axilary nodes **I2**: BI-RADS	**O1**: Breast cancer risk

	score	

Using the fuzzyTECH tool, [13], Figure 2 presents the block scheme for the FZ1 fuzzy logic. The input fuzzy variables are I1: Age and I2: Birads, and the output fuzzy variable is the O1:BCRisk.

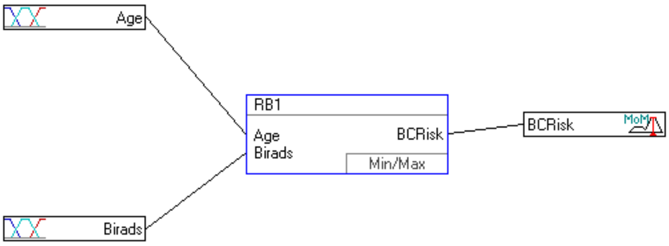


Figure.2 Researched FZ1 fuzzy logic block scheme

For each of the selected input and output variables was chosen a set of three linguistic fuzzy values – low, medium and high. The triangular fuzzification procedure is used for each of the input or output variables, converting the measured (stored or evaluated) numerical value into one of the fuzzy values. Figure 3, 4, 5 presents the visual definition of each fuzzy variable, and Table 4 presents a summary of the I/0 variables with their correspondent fuzzy values and triangular parameters

Table 4: I/0 variables with correspondent fuzzy values and triangular parameters

Variable	*Triangular Base value 1*	*Triangular Base value 2*
I1: Age		
Low	10	30
Medium	30	50
High	50	70
I2: Birads		
Low	0	2.5
Medium	1.25	3.75
High	2.5	5
O1: BCRisk		
Low	0	10
Medium	10	30
High	30	50

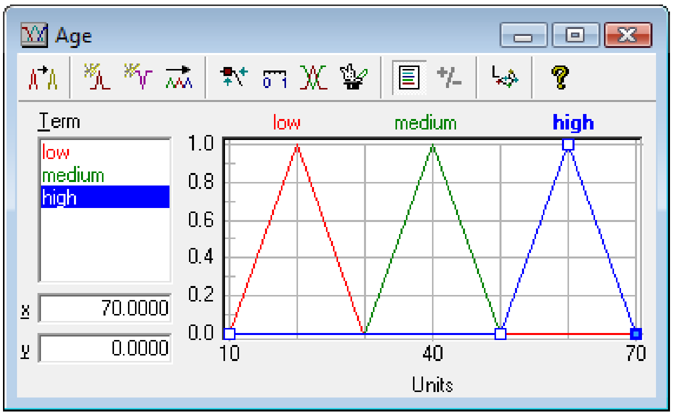


Fig.3 Input fuzzy variable definition for I1: Age

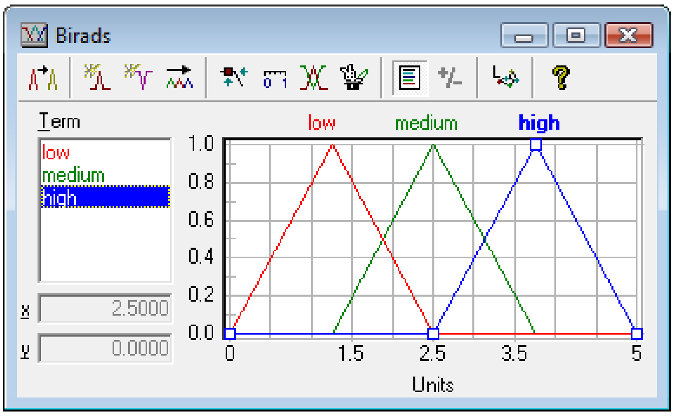


Fig.4 Input fuzzy variable definition I2: Birads

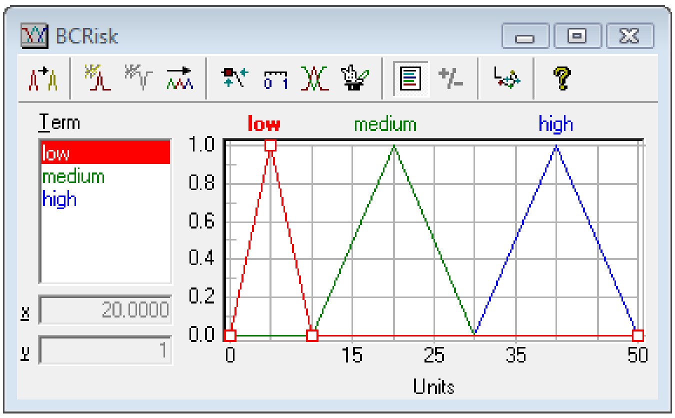


Fig.5 Output fuzzy variable definition for O1: Breast Cancer Risk

3.2 Rules of the Fuzzy Logic

The studied fuzzy system works on a fuzzy logic specified also according to the medical experience. The definition of the fuzzy inference rules for FZ1 is shown in Figure 6 as a matrix and in Figure 7 as a spreadsheet and the selected inference method for the rules is min-max.

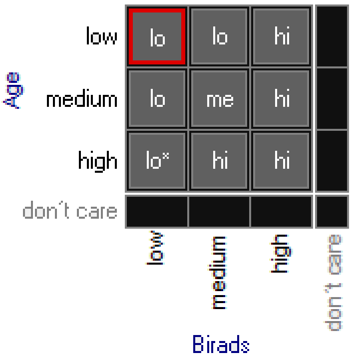
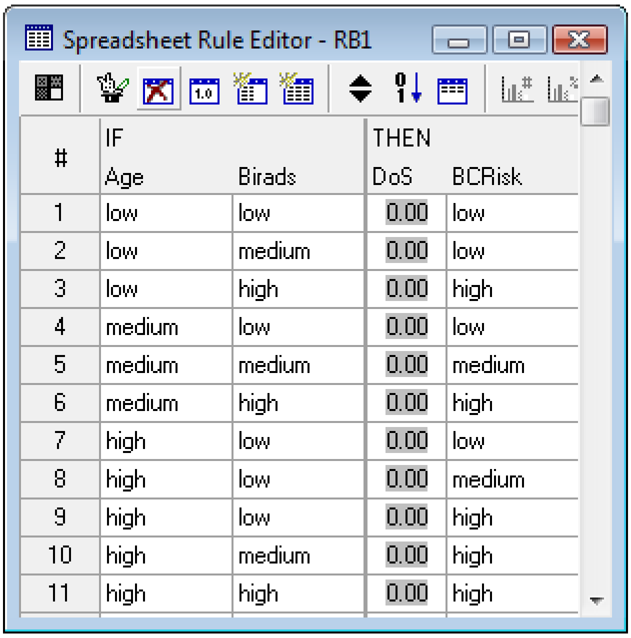


Fig.6 Fuzzy inference rules as matrix for FZ1


Spreadsheet Rule Editor - RB1

#	IF Age	Birads	THEN DoS	BCRisk
1	low	low	0.00	low
2	low	medium	0.00	low
3	low	high	0.00	high
4	medium	low	0.00	low
5	medium	medium	0.00	medium
6	medium	high	0.00	high
7	high	low	0.00	low
8	high	low	0.00	medium
9	high	low	0.00	high
10	high	medium	0.00	high
11	high	high	0.00	high

Fig.7 Fuzzy inference rules as spreadsheet for FZ1

As follows, Figure 8 and Figure 9 present the simulation results of the researched FZ1 fuzzy logic. For a breast cancer case characterized by age 22 and evaluation Birads result 1 the assessed potential risk for developing breast cancer (in the next five years) is 5%.

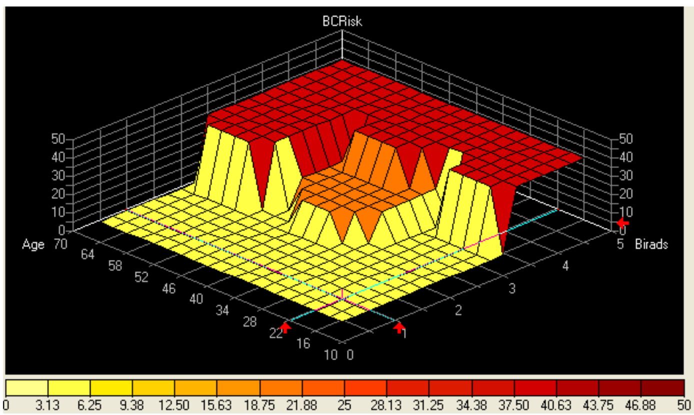


Fig.8 Simulation result for FZ1 (Age 22, Birads 1, BCRisk 5)

Similarly, in Figure 9, patient of age 58 and evaluated Birads result 3 has a determined potential risk of developing breast cancer in the next 5 years of 40%.

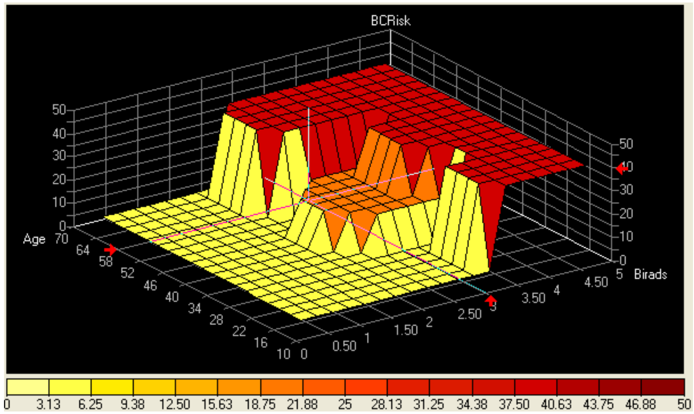


Fig.9 Simulation result for FZ1 (Age 58, Birads 3, BCRisk 40)

4 Conclusion

The fuzzy system defined in this paper makes possible the correlation between (stored and evaluated) patient data (like number of invaded axilary nodes, BI-RADS score, age and other case related data) and the prognostic risk of developing breast cancer, emulating the expert thinking and experience with the help of the computer. Though the fuzzy procedure still needs to be calibrated and validated with real data and under real conditions (for a large number of patients) prior to its use on a large scale, this procedure can be easily and successfully integrated and used in screening programs to assign, automatically, breast cancer risk factors to patients in order to highlight the cases that need prior attention and care.

The quality of this approach is thought to be successful following the positive results of such systems in other fields of science. In any case, extensive analysis and comparison of the presented system with similar CAD systems [14] is needed to prove the qualitative sprint. For sensibility improvement, our work in progress includes the research of a cascade fuzzy system that integrates all relevant information about the investigated patient.

References:

[1] M. Caramihai et. all, "Breast Cancer Treatment Evaluation based on Mammographic and

Echographic Distance Computing", World Academy of Science, Engineering and Technology, Vol.56, pp. 815-819, 2009

[2] V. Balanica et. all, “Integration of image and patient data, software and international coding systems for use in a mammography research project”, World Academy of Science, Engineering and Technology, Vol.58, pp.1002-1005, 2009

[3] M. Caramihai et. all, “Quality breast cancer treatment evaluation CAD module in a data management software”, ICCESSE, 2010, www.waset.org.

[4] DICOM Home page, Accessed at http://medical.nema.org/ on 24/08/2009.

[5] I. Dumitrache, “Ingineria Reglarii Automate”, Politehnica Press, 2005

[6] I. Dumitrache, N. Constantin, M. Dragoicea, “Retele neurale”, Matrix Press, 1999

[7] I. Dumitrache, C. Buiu, “Algoritmi genetici. Principii fundamentale si aplicatii in automatica”, Mediamira Press, 2000

[8] M. Elter, A. Horsch, “CADx of mammographic masses and clustered microcalcifications: A review,” Med. Phys. 2009, 35 (6) 2052-2068.

[9] BI-RADS, What is BIRADS? Accessed at http://www.birads.at/info.html on 24/08/2009.

[10] G. C. Keller, "Cancer Detection in the Periodic Physical Examination", Cancer, 51:2446-2447, 1983

[11] A.P. Ekeh, R.S. Alleyne, O.A. Duncan, “Role of Mammography in diagnosis of breast cancer in inner-city hospital”, Journal of the National Medical Association, 92:372-374, 2000

[12] Sobin LH, Wittekind CH, “Classification of Malignant Tumours“, eds. TNM, 6th ed. New York: John Wiley & Sons, 2002

[13] fuzzyTECH Software, Accessed at http://www.fuzzytech.com/ on 15/01/2010.

[14] I.Dumitrache, I.Mihu, M.Voinescu, "A hybrid medical decision support system", CSCS, pp. 402-406, 2007X1.