U.P.B. Sci. Bull., Series C, Vol. 73, Iss. 1, 2011 ISSN 1454-234x

EVALUATION OF BREAST CANCER RISK BY USING FUZZY LOGIC

Victor BALANICĂ[1], Ioan DUMITRACHE[2], Mihai CARAMIHAI[3], William RAE[4], Charles HERBST[5]

Procesul decizional de selecţie al tratamentului cancerului de sân este puternic relaţionat de procesul corect de evaluare şi diagnosticare a riscului de dezvoltare a cancerului pentru cazul suspectat. În ciuda dezvoltărilor tehnologice recente, metodele şi criteriile utilizate pentru cuantificarea caracteristicilor leziunilor detectate, pentru definirea stadiului de dezvoltare a cancerului şi pentru stabilirea unui diagnostic de risc final certificat, sunt încă subiective şi slab definite pentru comunităţile de clinicieni. Lucrarea de faţă propune un set de reguli fuzzy standard care pot fi utilizate în procesarea datelor relevante despre cancerul de sân studiat, indicând un factor prognostic de risc pentru cancerul de sân comparabil calitativ cu cel stabilit de medicul expert.

The decision process for selecting the best-suited follow-up treatment for a suspected breast cancer case is strongly dependent upon the correct diagnosis and assessment of the breast cancer risk. Despite the latest technological developments, the methods and criteria used to quantify the characteristics of detected lesion, so as to define the developmental stage of the breast cancer, and thus to finally arrive at a reliable (most probable) risk estimate, are still subjective and poorly defined for many clinicians. The present paper introduces a set of fuzzy rules that can be used to process the relevant data from breast cancer cases in order to give a breast cancer risk prognosis which can be qualitatively compared to that of an expert.

Keywords: Breast cancer diagnosis, fuzzy intelligent technique, CAD

[1] Eng., Dept. of Automatic Control and System Engineering, University POLITEHNICA of Bucharest, Romania, e-mail: vicord20011@gmail.com
[2] Prof., Dept. of Automatic Control and System Engineering, University POLITEHNICA of Bucharest, Romania, e-mail: idumitrache@ics.pub.ro
[3] Prof., Dept. of Automatic Control and System Engineering, University POLITEHNICA of Bucharest, Romania, e-mail: m.caramihai@ieee.org
[4], Dr., Dept. of Medical Physics, University of the Free State (UFS), Bloemfontein, South Africa, e-mail: raewid@ufs.ac.za
[5] Prof., Dept. of Medical Physics, University of the Free State (UFS), Bloemfontein, South Africa, e-mail: gnbisa@ufs.ac.za

1. Introduction

BREAST CANCER represents a major health problem worldwide being a major cause of death amongst women. Cancer management using consequent screening programs, which allow early detection and timely, optimal and varied methods of treatment, is a widely used and successful way of attempting to reduce morbidity and mortality.

During the last decade, imaging devices have undergone substantial improvements, resulting in the use of routine screening that is faster and more accurate in detecting breast lesions. A large range of technologies and instruments were developed based on X-rays imaging, ultrasound scanning 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 imaging equipment is not sufficient. In addition to this, correct assessment and classification of the detected characteristics [2, 3], followed by a judicious decisional procedure (CAD – computer aided diagnosis) determining the breast cancer diagnosis are both needed so as to promptly commence with delivery of appropriate effective treatment.

The detection and the assessment of the characteristics of breast masses is currently possible [2,3] using automatic software applications especially designed to analyze mammographic images stored in a DICOM format [4]. These image analysis methods are used to extract, quantify and store the numerical values of the relevant characteristics in a reliable database [2].

Noting also that Clinical Oncologists make diagnostic decisions about breast cancer patients based on past professional experience and knowledge, intelligent techniques are possibly the only class of automatic techniques powerful enough to emulate the expert's choice. Due to their stable behavior in the presence of noise, imprecision and uncertainty, these techniques could potentially obtain better results than classical methods.

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 judgment. The most widely applied intelligent techniques [5, 6, 7] are fuzzy techniques, neural network techniques, genetic algorithms and knowledge based systems (known also as expert systems).

The aim of this paper is to describe an intelligent procedure based on fuzzy techniques that could be used for the evaluation of breast cancer risk. This diagnostic procedure requires the definition of a set of judiciously chosen intelligent fuzzy rules concerning relevant personal patient data and automatically evaluated cancer tumour characteristics.

2. Methods

A. Fuzzy intelligent technique

The successful performance of fuzzy intelligent techniques has been demonstrated during recent years in various applications worldwide. Excellent results and easy implementation have been demonstrated. There has been marked progress in application of the technique and it has contributed to many improvements in devices and technology in several fields of science. These include 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 stimulator controllers), economy (fuzzy decisional methods), cognitive psychology (fuzzy modelling of human vision), etc.

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

In other words, the definition of the fuzzy variables and rules does not start from the precise definition of a process, but rather from the observation and in-depth understanding of the components of a physical phenomenon. By using fuzzy intelligent techniques, human experience related to the investigated process can be expressed as a set of new rules of deduction attached to the fuzzy logic system.

The fuzzy logic procedure can be summarised in 3 steps, [5]:

1. Determination of the input and output variables that describe the observed phenomenon together with the selection of their variation interval (for example, in Fig. 1, variable E is a system variable that varies in the interval [-1,1]),

2. For each fuzzy variable, define of a set of linguistic values together with their associated membership functions that map/cover the numerical range of the fuzzy variable. For example, in Fig. 1, the set of the linguistic values is A={-H,-M,-L,0,+L,+M,+H}, defined as triangular membership functions that cover the interval [-1,1],

3. 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).

The fuzzification and the defuzzification processes are the two interfaces between the fuzzy logic system and the measured phenomenon data. Fuzzification is used for the transformation of a real value, $x \in R$, acquired from the studied process, in one of the fuzzy values from a fuzzy set for a specific fuzzy variable. Defuzzification is the inverse process that transforms the output fuzzy variable (computed by the set of inference rules between the input variables) into a crisp, real value, $x \in R$, (normally sent back to the process, as control feedback).

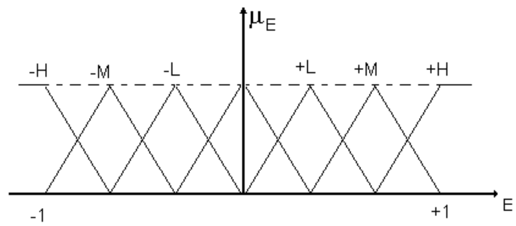

Fig..1. Graphical representation of the triangular membership functions that define the set A of linguistic values {-H,-M,-L,0,+L,+M,+H} for the fuzzy variable E

There are various criteria for determining the fuzzification or the defuzzification processes such as singleton, triangular, trapezoidal or Gaussian transformations [5]. Fig. 1 demonstrates the use of a triangular transformation.

B. Breast cancer lesion classification

The diagnostic procedure used for breast cancer identification consists of appropriate observation of the imagery results in search of characteristics that fully define, according to the expert oncologist, the cancerous development stage of the detected lesion [8].

Many screening cases have been studied and analyzed and a set of characteristics have been chosen which would best define the characteristics of a malignant breast tumour (lesion surface, lesion contrast, lesion differentiation, presence of spiculations or angulations, existence of a posterior shadow, irregular margins, calcifications, fragmentation, etc.). Based on these characteristics oncologists evaluate the state of the mass and classify it according to the internationally recognized Breast Imaging-Reporting and Data System (BI-RADS) scale.

BI-RADS is a unitary system designed for helping medical professionals assess, interpret and classify mammography, echography and magnetic resonance imaging results 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 the diagnostic status as follows; 0: Incomplete, 1: Negative, 2: Benign finding(s), 3: Probably benign, 4: Suspicious abnormality and 5: Highly suggestive of malignancy. Probability values are associated with the different categories.

C. Potential risk factors

Scientists recognize the following seven broad categories of risk factors that predispose women to developing breast cancer, [10]:

- age - risk of developing breast cancer rises dramatically with age (one in 19000 women at age of 25, one in 93 women at age of 45 and one in nine at age of 85 will develop cancer), [10]
- family history of breast cancer (mother or sister),
- hormonal factors (early onset of menstruation or late menopause),
- proliferative (reproductive) breast disease (for women that didn't have children – had no breastfed -, or had children after age 30),
- irradiation of the breast region at an early age,
- personal history of malignancy and
- lifestyle factors.

Though these factors and their complex interactions cannot be determined in more than 70% of breast cancer cases, [10], studies show 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 childbearing and lower parity [11].

3. Results

A. Inputs and Outputs of the Fuzzy Logic

The construction of a fuzzy logic system with more than two input variables markedly increases the complexity of the calculations and of the design definition. For simplicity, the present paper discusses the basic principles of the application of fuzzy logic with two input variables and one output variable.

For this order of problem, the fuzzy logic that evaluates the breast cancer risk (seen as the final output parameter of the system) can take into account two of the possible input variables from Table 1. The possible input variables together with their relevance concerning the decisional process of the breast cancer risk were presented by clinical experts and can vary slightly from clinician to clinician. As shown below, the relevance factor is only used as an overview over the importance of the variables just to have guidance concerning the variable selection.

Table 1

Possible fuzzy logic input variables decided by clinicians

Input variable	*Relevance factor*
Age	6%
Age of first menstruation	6%
Number of invaded axillary nodes	9%
Tumour Surface	24%

As seen in table 2 below, the two selected input variables for the fuzzy system are chosen giving consideration to the their relative importance. As we contrain ourselves to a fuzzy system with only two inputs and because the tumour surface area has the highest significance according to the clinicians, the considered systems that are utilised here always considered tumour surface area as one of the two input variables.

Table 2

Numerical variation interval for input/output variables(reference)

Input variable	*Min value*	*Max value*
Age (years)	30	80 years
Age of first menstruation (years)	10	15 years
Number of invaded axillary nodes	0	15
Tumour Surface Area	0	8000 pixels
Output variable		
Breast cancer risk	0%	100%

The numerical range for each of the input and output variables can be seen in Table 2. The number of invaded axillary nodes is also a measure of the cancer infiltration into adjacent tissues. When benign tumours or incipient cancers do not involve the axillary lymph nodes it indicates surgical tumor excision would be more likely to result in local control. The higher the number of infiltrated lymph axillary nodes, the greater the risk of widespread metastatic disease, and it may also be an indication of the aggression of the cancer [12]. In our algorithm, the ranges for all input and output variables are set in collaboration with the clinicians and can be changed to achieve a more accurate result.

Table 3

Considered fuzzy systems

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

The breast cancer risk factor is the output of all considered fuzzy systems (see Table 3). The output variable ranges are also given in Table 2, where 0%

corresponding to zero breast cancer risk while 100% corresponds to a high breast cancer risk. The breast cancer risk factor computed by the fuzzy system is presented after defuzzification in terms of the BIRADS score.

B. Researched Fuzzy System

The means and techniques for developing a fuzzy system described above function similarly in all of the considered cases that are listed in Table 3. This paper presents only the results of the first fuzzy system, FZ1, due to limitations in the measured and available data for test purposes.

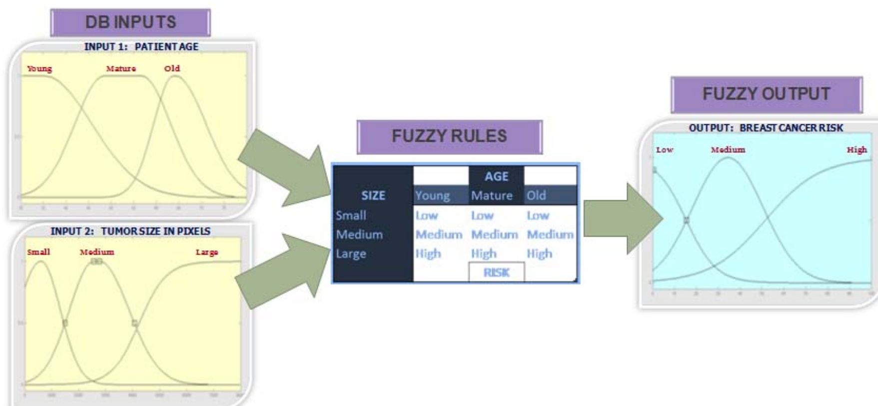

Fig. 2. Researched FZ1 fuzzy logic block scheme

The researched system was designed and investigated using Matlab [13] and fuzzyTECH, [14], software tools. Fig. 2 presents the resultant block scheme for the FZ1 fuzzy logic system. The considered input fuzzy variables are I1: Age and I2: Tumor Surface Area, and the output fuzzy variable is O1:BCRisk (Breast Cancer Risk). The Age of the patient was derived from the date of birth of the patient, the tumour surface was identified directly on the mammography using a standard segmentation algorithm described in the literature [15, 16] and the breast cancer risk was evaluated in terms of BIRADS scale and was compared to the clinical/pathological evaluation.

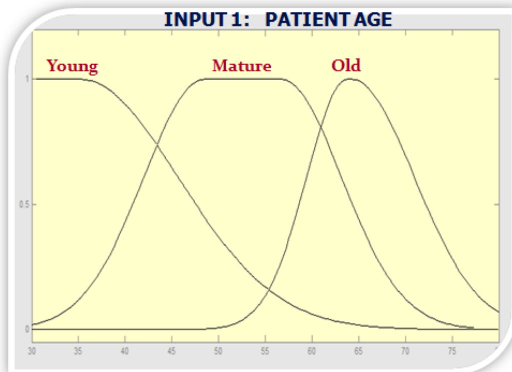

Fig. 3. Input fuzzy variable definition for I1: Age – with three possible fuzzy values: Young, Mature and Old

Each of the selected input and output variables was described by a set of three linguistic fuzzy values – low, medium and high – each defined through a gauss membership function, thus allowing the fuzzification procedure to convert the measured (stored or evaluated) numerical value into one of the fuzzy values. The Gaussians functions were not optimised to give the best possible correlation with the truth, but this can readily be done. Figs. 3, 4, 5 presents the visual definition of the fuzzy variables (tumour size, age and breast cancer risk).

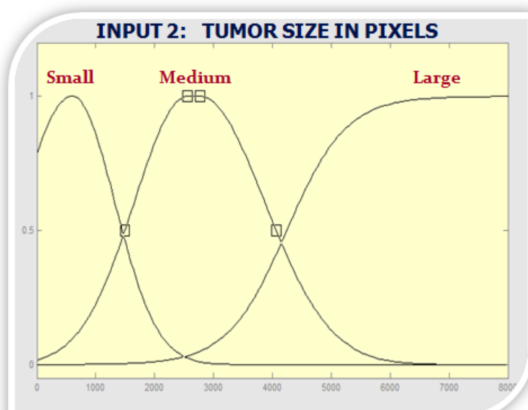

Fig. 4. Input fuzzy variable definition for I2: Tumor size in Pixels – with three possible fuzzy values: Small, Medium and Large

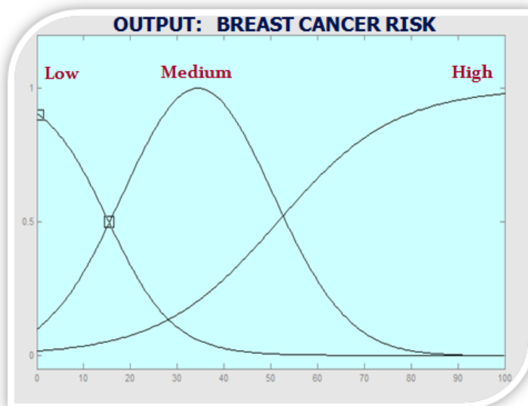

Fig. 5. Output fuzzy variable definition for O1: Breast Cancer Risk – with three possible fuzzy values: Low, Medium and High

C. Rules of the Fuzzy Logic

The researched fuzzy system uses the Clinicians' logic derived from their medical experience. The definition of the fuzzy inference rules for FZ1 is shown in Fig. 6 (each rule was equally weighted to 1) and the applied fuzzy inference type and parameters were configured in such a manner to make fuzzy decisions best fit the medical experience. As consequence, the selected inference type is Mamdani (that expects the output membership to be also a fuzzy set) and its parameters were configured appropriately like in Table 4.

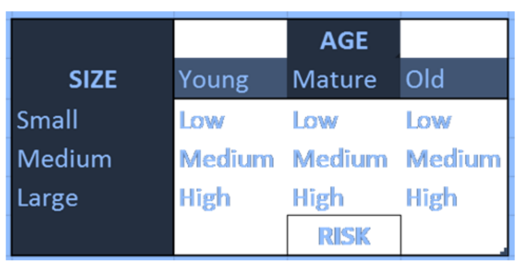

Fig. 6. Fuzzy inference rules as matrix for FZ1

Table 4

Considered fuzzy systems

Method	***Value***
AND method	prod
OR method	max
Implication	min
Aggregation	sum
Defuzzification	centroid

Taking into account the definition of the input and output variables and also of the table of rules presented above, Fig. 7 presents the result and simulation surface of the researched FZ1 fuzzy logic, which demonstrates the direct relationship between the considered inputs (age and tumour surface) and the output (the breast cancer risk).

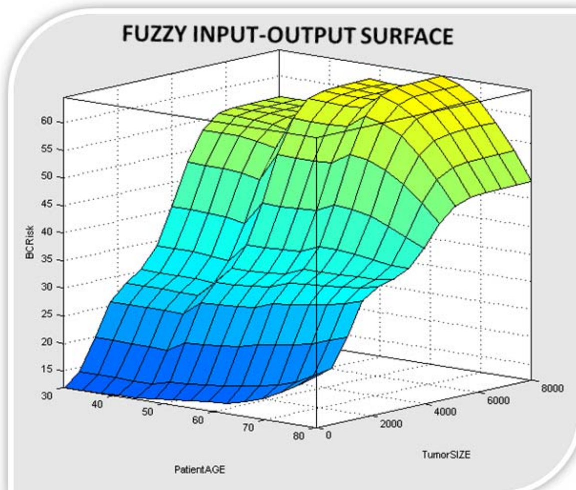

Fig.7. Resulted simulation surface for FZ1

D. Preliminary results

For testing purposes, we have used information from 60 patient records: the age of the patient, the extracted tumour segmentation result from the processed mammograms and the experts' clinical and pathological truth determination for each case.

The age and the segmentation result for each patient were given as inputs in the researched fuzzy system FZ1 and finally the fuzzy output was compared to the clinical truth. 8 presents this comparison between clinical and fuzzy system risk evaluation and suggests some agreement between clinical truth values and fuzzy evaluated values.

Future results are expected to be markedly better as optimisation of the fuzzy rule allocation and by analysing a larger number of patients should improve system performance. Furthermore, research results are expected from the next generation of multi-level fuzzy systems that allows the use of complex combinations of inputs for the evaluation of breast cancer risk.

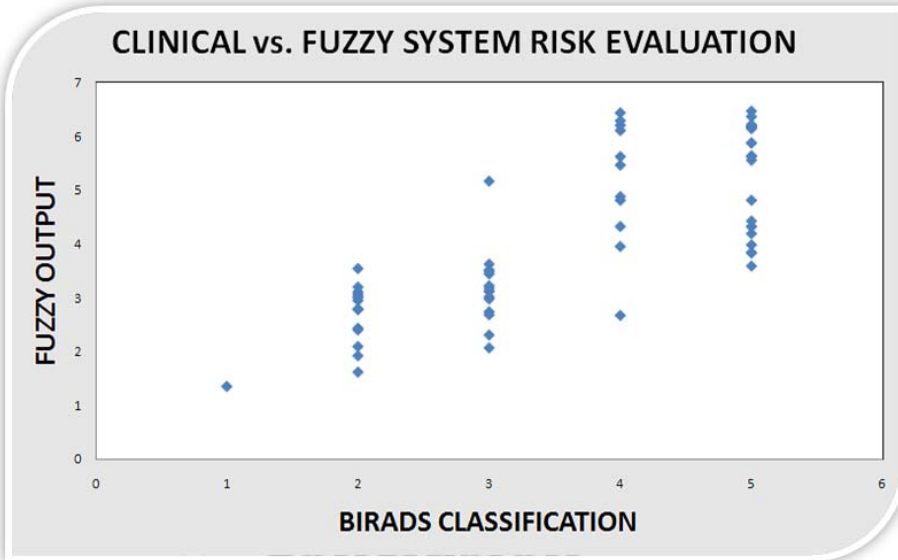

Fig..8. Comparison between clinical truth values and fuzzy evaluation in terms of BIRADS scale

4. Conclusion

Medical oncologists diagnose breast cancer based on past professional experience and knowledge. Computer-based fuzzy logic techniques are becoming powerful enough to emulate an expert's choice. Due to their robust behaviour in the presence of noise, imprecision and uncertainty, these techniques can obtain better results than classical Computer Aided Diagnostic methods.

The aim of this study is to propose a fuzzy logic technique for the prediction of the risk of breast cancer based on a set of judiciously chosen fuzzy rules utilizing patient age and automatically extracted tumour features.

The presented example links the age of the patient to the detected tumour surface in order to estimate the breast cancer risk. The defined fuzzy logic rules result in reasonable agreement with the clinician's assessment. Though it needs further calibration and validation on larger number of patients, this procedure can be easily and successfully integrated in screening programs to automatically assign breast cancer risk to patients in order to highlight the cases that need priority 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. On the other hand, extensive analysis and comparison of the presented system with similar CAD systems [17] is needed to prove the qualitative sprint

For future research, this method can be extended to include parameters like the number of invaded axillary nodes, calcifications, disease extent, etc. and so improve the prognostic risk estimation. For a more sensible and correct assesssment of breast cancer risk, our work in progress includes the research of a multi-level fuzzy system that can integrate all relevant information about the investigated patient in the calculation of a weighted breast cancer risk.

R E F E R E N C E S

[1] *M. Caramihai et. al.*, 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. al.*, 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. al.*, 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*, "ngineria Reglarii Automate (Automatic Control Engineering), Politehnica Press, 2005

[6] *I. Dumitrache, N. Constantin, M. Dragoicea*, Retele neurale (Neural networks), Matrix Press, 1999

[7] *I. Dumitrache, C. Buiu*, Algoritmi genetici. Principii fundamentale si aplicatii in automatica (Genetic algorithms. Fundamental principles and applications in automatics), 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] *Jacqueline L. Longe*,The Gale Encyclopedia of Cancer: A Guide to Cancer and Its Treatments, 2nd ed., Thomson Gale, pp. 165, 2005

[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] *L.H. Sobin, C.H. Wittekind*, Classification of Malignant Tumours, eds. TNM, 6th ed. New York: John Wiley & Sons, 2002

[13] *Matlab Software*, Accessed at http://www.mathworks.com/

[14] *fuzzy TECH Software*, Accessed at http://www.fuzzytech.com/

[15] *A. Oprea, R. Strungaru, G.M. Ungureanu*, New segmentation techniques for breast cancer detection based on mammography, 1st National Symposium on e-Health and Bioengineering, pp. 153-156, 2007

[16] *A. Sultana, M. Ungureanu, S. Pasca, R. Strungaru*, A Novel Approach in Mammogram Segmentation. Proc. Nonlinear Optics Quantum Optics, 39:271-280, 2009

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