International Journal of Computational Intelligence Techniques
ISSN: 0976-0466 & E-ISSN: 0976-0474, Volume 4, Issue 1, 2013, pp.-114-117.
Available online at http://www.bioinfopublication.org/jouarchive.php?opt=&jouid=BPJ0000221

VISUALIZATION OF RISK IN BREAST CANCER USING FUZZY LOGIC IN MATLAB ENVIRONMENT

LATHA K.C.[1]*, MADHU B.[2], AYESHA S.[1], RAMYA R.[3], SRIDHAR R.[4] AND BALASUBRAMANIAN S.[5]

[1]Department of Water & Health, JSS University, Mysore- 570 015, Karnataka, India.
[2]Department of Community Medicine, JSS Medical College, Mysore-570 015, Karnataka, India.
[3]Department of Anatomy, JSS University, Mysore- 570 015, Karnataka, India.
[4]Department of Computer Science, Rama Krishna Mission Vidyalaya, Coimbatore- 641 020, TN, India.
[5]Director Research, JSS University, SS Nagar, Mysore- 570 015, Karnataka, India.
*Corresponding Author: Email- latha_tanvi23@yahoo.com

Received: April 19, 2013; Accepted: May 09, 2013

Abstract- Breast cancer is the global burden, the recorded breast cancer data from hospitals can be used to develop a fuzzy decision support system to help in risk stratification of breast cancer, which helps as a decision supporting system. This system uses age and cancer state as a input variable and one output variable as risk status which is further divided into Not serious, Serious and Very serious. This parsimonious system helps in the assessment of the breast cancer risk. Despite the latest technological developments, the method and criteria used to quantify are reliable risk estimate and still subjective. The present study introduces a set of fuzzy rules that can be used to process the relevant data from breast cancer cases in order to assess breast cancer risk prognosis which is qualitative.

Keywords- Breast cancer, Fuzzy logic, MATLAB, Seriousness, Decision supporting system, Fuzzy Rules, Tumor, Node, Metastasis

Citation: Latha K.C., et al. (2013) Visualization of Risk in Breast Cancer Using Fuzzy Logic in MATLAB Environment. International Journal of Computational Intelligence Techniques, ISSN: 0976-0466 & E-ISSN: 0976-0474, Volume 4, Issue 1, pp.-114-117.

Introduction

Cancer is a global killer, taking 7.6 million lives across the world each year. The impact of cancer on all populations is devastating. Cancer is a global epidemic, affecting all ages and all socio-economic groups. Breast cancer is the leading cancer among women. One in ten of all new cancers diagnosed worldwide each year is a cancer of the female breast and it is the most common cancer in women in both developing and developed countries. It is also the principal cause of death from cancer among women globally**.** Cancer of the breast in women is a major health burden worldwide. In India breast cancer has replaced cervical cancer by leading.

Machine learning tools in medical diagnosis is increasing due to the improved effectiveness of classification and recognition systems to help medical experts in diagnosing breast cancer [14].

Geocomputation has facilitated data collection and statistical report generation but however advanced autonomous techniques in exploratory analysis were not widely adopted [6]. For making any decision, enormous data are required from different sources. Since the data involves certain uncertainties, the analysis of data through conventional statistical methods affects the process of decision making. In addition, it is a time consuming procedure for any database.

Recently, medical databases have accumulated large quantities of information about patients and their medical conditions. A single disease may manifest itself quite differently with different intensities depending on the patient. The existing intelligent techniques are unable to derive conclusions when the diagnosis of disease involves several levels of uncertainties and imprecision and it is inherent to medicine. The computational intelligence techniques like decision trees and neural networks are important tools for the rule extraction and data understanding [19]

The application of fuzzy logic in medicine and bioinformatics has received much appreciation [16]. Guo and Neagu [8] have applied fuzzy analysis in data mining technique to formulate the decision tree models using k-means algorithm. Due to complexity of biomedical classification problems, a more realistic target using fuzzy logic can be a hold to an effective Decision Support System for binary classification [9]. The decision supporting system is designed using fuzzy logic technique.

Fuzzy logic produces more-realistic answers by replacing the inflexible “yes/no” by a topical adjustment in the form of a “more or less” and by introducing linguistic nuances into the process of decision [15].

Fuzzy Logic in Breast Cancer

Fuzzy logic is an extension of Boolean logic that replaces binary truth values with degrees of truth. It was first described by Zadeh in 1965, and its application in 1969 for early diagnosis domain for Fuzzy Logic. Blechner (2005) has reviewed that Fuzzy Logic allow membership values between 0 and 1, arguably it can provide a more realistic representation of data that was inherently noisy and imprecise.

The diagnosis of breast cancer involves several levels of uncertainty which manifests itself quite differently, depending on the patient and surrounding environment and intensity [16]. Architecture for medical knowledge information systems permit fuzzy logic across several medical information sources to assist in improving public health status.

The breast cancer disease is associated with many factors like age, diet, marital status, cancer stage, education, occupation, family grade, treatment, heredity, number of children, environmental factors, diet, etc.[17] The breast cancer stage and other factors have different degrees of association and uncertainty. Therefore, the focus of this section is the prediction of susceptibility of a patient suffering from breast cancer. This problem was studied on the basis of the parameters taken from the doctor's diagnostic reports using fuzzy logic with the objectives: To identify the breast cancer risk region based on TNM (Tumour, Number of Lymph Nodes and Metastasis) staging using Fuzzy Logic. To predict the intensity range of the breast cancer using Fuzzy Logic in MATLAB environment.

Data Used and Methodology

The breast cancer seriousness can be predicted by extracting the knowledge from the data. The extent of disease from the records was observed after scrutiny. Out of 940 data the complete history of the breast cancer patient was obtained from 843 breast cancer cases. The parameters which were taken for the study in breast cancer data are location, age, sex, diet, marital status, education, occupation, family grade, number of children, stage of diagnosis Tumour [T], Number of Lymph Nodes [N] and Metastasis [M], treatment, heredity and period of illness. This scrutinized data was used for the analysis.

From the medical databases, a clear association between clinical, pathological and sociological factors and recurrence-free survival in breast cancer patients can be observed. The most widely applied intelligent techniques are Neural Networks, genetic algorithms, fuzzy techniques and knowledge based expert system. Therefore for the present study we have adopted Fuzzy logic as an Intelligent Technique which is demonstrated as a clinical decision supporting system (CDSS) for non-communicable diseases.

Using this aforementioned data, the breast cancer risk assessment in Southern region of Karnataka (districts like Mysore, Mandya, Chamrajnagar, Hassan, and Coorg) was carried out.

Fuzzy Logic

Fuzzy logic is a problem-solving control system methodology that lends itself to implementation in systems ranging from simple, small, embedded micro-controllers to large, networked, multi-channel Personal Computer or workstation-based data acquisition and control systems. Fuzzy rule can be applied to build a classifier, a model used for prediction, or it can be applied to form a decision support system.

Membership Functions in the Fuzzy Logic Tool

A fuzzy set is an extension of a classical set. If X is the universe of discourse and its elements are denoted by x, then a fuzzy set A in X is defined as a set of ordered pairs.

$$A = \{x, \mu_A(x) \mid x \varepsilon. \quad (1)$$

$\mu_A(x)$ is called the membership function of x in A. The membership function maps each element of X to a membership value between 0 and 1. The Fuzzy Logic Toolbox includes 11 built-in membership function types. These 11 functions are, in turn, built from several basic functions viz., piecewise linear functions, the Gaussian distribution function, the sigmoid curve, quadratic and cubic polynomial curve. By convention, all membership functions have the letters 'mf' at the end of their names. The membership functions are formed using straight lines. The simplest is the Triangular membership function, and it has the function name 'trimf'. It is nothing more than a collection of three points forming a triangle. The Trapezoidal membership function, 'trapmf', has a flat top and is just a truncated triangle curve. These straight line membership functions have the advantage of simplicity.

Two membership functions are built on the Gaussian distribution curve: a simple Gaussian curve and a two-sided composite of two different Gaussian curves. The two functions are 'gaussmf' and 'gauss2mf'. The generalized bell membership function is specified by three parameters and has the function name 'gbellmf'. Gaussian and Bell membership functions are popular methods for specifying fuzzy sets because of their smoothness and concise notation but they are unable to specify asymmetric membership functions, which are important in certain applications. The sigmoidal membership function is either open left or right. Asymmetric and closed (i.e. not open to the left or right) membership functions can be synthesized using two sigmoidal functions, so in addition to the basic 'sigmf', we also have the difference between two sigmoidal functions, 'dsigmf', and the product of two sigmoidal functions 'psigmf'.

Fuzzy Logic has been used in cancer prognosis; being non crisp it can act as a natural ally of a physician in prognostic decision making process [10]. Phillips [13] has used that fuzzy logic model predicted breast cancer using volatile biomarkers in their breath. Therefore, in addition to the above mentioned data mining technique Fuzzy Logic was applied in the present study to interpret the risk and soften the accuracy or interpretability of the breast cancer.

The stages were found to be very important than the other variables which is included in the [Table-1]. The data was further classified into TNM in the database and adopted for the Fuzzy Logic. A rule based Fuzzy Expert System (FES) is developed in this study which uses the laboratory data and helps by to create awareness about breast cancer which plays a crucial role in early decision or as decision supporting system.

Table 1- Grouping of Stage in Cancer

Sl. No.	Stages	Seriousness	Clinical Classification
1	IV	Very Serious	Any T, Any N, M1
2	III A	Serious	T0N2M0
			T1N2M0
			T2N2M0
			T3N1,N2M0
	III B		T4N0,N1,N2M0
	III C		Any TN3M0
	II A		T0N1M0
			T1N1M0
			T2N0M0
	II B		T2N1M0
			T3N0M0
3	I	Not Serious	T1N0M0
	Zero		Tis*N0M0

In recent times, the soft computing technique was found to be an intellectual research area adopted in computational biology. There are several methods for the commercially available techniques in soft computing such as Bayesians statistics, genetic algorithm, principle components etc., which were applied for many challenging

problems in medicines and engineering. Koperski and Han [11] have adopted fuzzy logic for making a prognostic decision in breast and prostate cancers.

Later several methods were adopted for the classification of the biological data sets using fuzzy inference system. Recently, Bellaachia and Guven [3] have used three data mining techniques like C4.5 decision tree algorithm, neural network and Naïve Bayes for the breast cancer survivability studies. Among this, C4.5 algorithm provides a better performance than the other two techniques adopted by them. Fuzzy logic has been applied to study the breast cancer stage, based on Tumor size (T), Lymph Node number (N) and Metastases (M), which have different degree of association and uncertainty The present study has been undertaken to analyze the range of risk in breast cancer integrated with MATLAB environment for Fuzzy logic.

For the design process, breast cancer tumour size, number of nodes and the metastasis were used as input parameters and the breast cancer risk was obtained as an output. For fuzzification of these parameters the linguistic variables Very Serious (VS), Serious (S) and Not Serious (NS) were used. For the inference mechanism, the Mamdani max-min inference was applied.

The units of the used parameters are: Tumour (Size), Lymph Node (Number) and Metastasis (Yes or No) [Fig-1]. Parts of the developed fuzzy rules using MATLAB platform are shown in the [Table-2]. Totally four Fuzzy rules were formed and the output was shown. The performed Fuzzy rules predict the seriousness of the breast cancer at their respective stage.

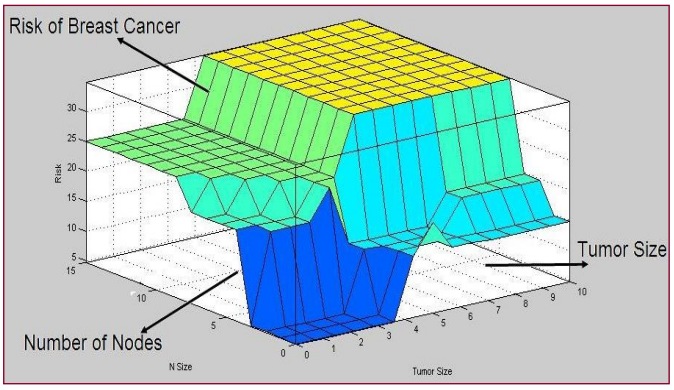

Fig 1- 3D Surface Visualization of TNM with respect to Risk in Breast Cancer

Table 2- Developed Coding for the Prediction of Seriousness in MATLAB platform

Rule Number	Tumour	Node	Metastatis
1	Very Small	Not Involved	No
2	Present	Involved	No
3	Present	Involved	No
4	Present	Involved	Yes

Rule 1: If Tumour= very small and Node = 0, then Metastasis = No, i.e., if the patient's Tumour size was very small, no involvement of the nodal region, with no metastasis, then the patient's breast cancer risk was found to be Not serious.

Rule 2: If Tumour= small and Node = 1or more, then Metastasis = No, i.e., if the patient's Tumour size is small, nodal region is involved, with no metastasis, then the patient's breast cancer risk was observed as Serious.

Rule 3: If Tumour= present and Node = 2 or more, then Metastasis = Yes, i.e., if the patient's Tumour size and the nodal region both are involved, with metastasis, then the patient's breast cancer risk was predicted to be Very Serious.

Rule 4: If Metastasis is yes then the Scale is Very Serious

Fuzzification of the used parameters is made by the membership functions of the Fuzzy Logic. From the [Table-3] the predicted range of the risk of the breast cancer by FES is similar to the data observed in the literature. This system is observed to be very good for testing and learning process in the cancer epidemic, which can be used to process the relevant data from breast cancer cases in order to give a breast cancer risk prognosis. The breast cancer risk reviewed through Fuzzy logic is shown in the [Fig-2].

Table 3-Comparison of FES with Literature

Tumour (T)	Lymph Node (N)	Metastatis (M)	Literature				FES	
			T	N	M	Risk	value	Risk
1	0	0	<2 cm	-	-	NS	1.02	NS
0	1	0	-	1-3 nodes	-	S	1.22	S
1	1	0	<2 cm	1-3 nodes	-	S	1.8	S
2	0	0	>2-5 cm	-	-	S	1.93	S
2	1	0	>2-5 cm	1-3 nodes	-	S	2.11	S
3	0	0	>5 cm	-	-	S	2.22	S
0	2	0	-	4-9 nodes	-	VS	2.57	VS
1	2	0	>2 cm	4-9 nodes	-	VS	2.96	VS
2	2	0	>2-5 cm	4-9 nodes	-	VS	3.05	VS
3	1,2	0	>5 cm	1-9 nodes	-	VS	3.82	VS
4	0,1,2	0	CW/S	>1 node	-	VS	3.92	VS
01-04	3	0	Any	>9 nodes	-	VS	4.5	VS
01-04	01-03	1	Any	>1 node	yes	VS	4.78	VS

Range : Risk; 0-1.02 : NS (Not Serious); 1.22-2.22 : S (Serious); 2.57-4.78 : VS (Very Serious)

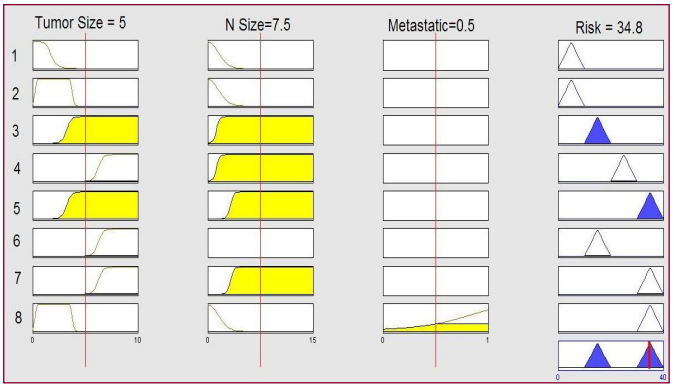

Fig. 2- Fuzzy Rule Viewer of the Breast Cancer Risk for the TNM

Many biological variables were characterized by the non-linear characteristics that are better represented by an interval than by a binary process [2]. This aspect makes the fuzzy logic a potential and useful method to manage biological data. Fuzzy modeling approach represents a promising means of predicting the phenotypic heterogeneity in colorectal cancer presentation in mutation carriers. It also enables the formulation of clinical risk scores, thereby allowing individualization in the prevention strategies [5].

Results

Test System and Discussion

For testing purpose, we have tested in consultation with oncologist the data of breast cancer and the risk factors using an algorithm developed by us in MATLAB and the results are as follows. The tumor classification is mentioned in the [Table-1] were given as input in the fuzzy system and fuzzy output was presented in [Table-

3]. Which reveals that the value 3.92 to 3.94 is Serious, for Very Serious the fuzzy value starts from 2.57 to 4.78, if the fuzzy value is less than 1.02, it is considered as Not Serious. Where age was used as a input parameter along with the cancer stage. Age is an important variable in breast cancer. The input variable has two fuzzy sets below age 50, premenopausal and above age 50, post-menopausal. Similarly for education code and family grade were also performed and the results are presented in [Table-4] and [Table-5]. From the table it indicates that Seriousness and Very Seriousness conditions are more in illiterates and who have just completed matriculation when compared with the educated and highly educated people. In case of Family Grade Low income group shows more number by leading, followed by medium income group while in case of high income group only one Very serious condition out of 513 cases. However in both the analysis Not Serious condition was not counted due very few cases.

Table 4- Education Qualification and FES

Education Code	No of .Serious Conditions	No. of Very Serious Conditions
Illitrate	47	57
10^{th} - 12^{th}	98	82
Degree and above	39	36

Table 5- Family Grade and FES

Family Grade	No of Serious conditions	No of Very Serious conditions
Low	147	110
Medium	56	62
High	-	1

The presented study links the age, education and family grade of the patient to detect the seriousness in order to understand the breast cancer risk. The fuzzy rules mentioned earlier is reasonably agreed with the clinicians assessment, our algorithm in MATLAB environment can be used to screen the programs to automatically assign breast cancer risk to patients who needs priority, attention and medical care. However for future research it is necessary to incorporate multiple level fuzzy systems that are incorporating two or three risk variables, which have more validation. The study can be extended further for breast cancer using axillary nodes classification and disease extend, so that to improve the prognostic risk estimation for multi level fuzzy system, which will incorporate all relevant risk factors mentioned by[7]. This study can be extended to other cancers too also incorporating molecular data which will provide a fruit full diagnostic method. Our procedure can be used as a Clinical Decision Supporting System.

Conclusion

Conventional computational methods for problem solving arrive at conclusion through a series of procedures and noise reductions. However, Fuzzy Logic provides a simple way to attain definite conclusions based upon vague, ambiguous, imprecise, noisy or missing input information. Fuzzy Logic incorporates a simple, rule-based 'IF X and Y then Z' approach to solving problem rather than attempting to model a system mathematically.

The study has outlined, discussed and resolved the Fuzzy Expert System for prognostic decision making in the seriousness of the breast cancer. This paper describes a design for a Fuzzy Expert System for the determination of the possibility of the diagnosis of the breast cancer, which can be used for the scope. This system can be developed further by increasing the knowledge rules from one side and by adding the neural network from the other side.

This study proposed a computational algorithm to assess the seriousness of the occurrences using the stages of diagnosis.

References

[1] Amir A., Evans D.G., Shenton A., Lalloo F., Moran A., Boggis C., Wilson M. and Howell A. (2003) *Journal of Med. Genet.,* 40, 807-814.

[2] Baxt W.G. (1994) *Cancer Letters*, 77, 85-93.

[3] Bellaachia A. and Guven E. (2006) 6*th SIAM International Conference on Data Mining*.

[4] Blechner M.D. (2005) *Behaviour of Various Machine Learning Models in the Face of Noisy Data*, Harvard-MIT Division of Health Sciences and Technology, Final Project.

[5] Brand M.R., Jones D.D., Lynch T.H., Brand E.R., Watson P., Ashwathnayaran R. and Roy K.H. (2006) *World Journal of Gastroenterology*, 12(28), 4485-4491.

[6] Castro V.E. and Lee I. (2001) 6*th International Conference on Geocomputation.*

[7] Ekeh A.P., Alleyne R.S., Duncan O.A. (2000) *Journal of National Medical Association*, 92, 372-374.

[8] Guo G. and Neagu D. (2005) *International Journal of Computational Intelligence and Applications*, 5(3), 321-333.

[9] He Y., Tang Y., Zhang Y.Q. and Sunderraman R. (2006) *International Journal of Data Mining and Bioinformatics,* 1(1), 3-18.

[10] Khan U., Shin H., Choi J.P. and Kim M. (2008) 7*th Australasian Data Mining Conference*, Glenelg, Australia, 87.

[11] Koperski K. and Han H. (1995) 4*th International Symposium on Large Spatial Databases*, 47-66.

[12] Pach F. and Abonyi J. (2006) *Proceedings of World Academy of Science, Engineering and Technology*, (13), 45-50.

[13] Phillips M., Cataneo N.R., Ditkoff B.A., Fisher P., Greenberg J., Gunawardena R., Kwon C.S., Tietje O. and Wong C. (2006) *Breast Cancer Research and Treatment, Springer*.

[14] Sahan S., Polat K., Kodaz H. and Gunes S. (2007) *Computers in Biology and Medicine*, 37(3), 415-423.

[15] Schneider J., Peltri G., Bitterlich N., Philipp M., Velcovsky H.G., Morr H., Katz N. and Eigenbrodt E. (2003) *Clin. Exp. Med.*, 2, 185-191.

[16] Torres A. and Nieto J.J. (2005) *Journal of Biomedicine and Biotechnology*, 91908.

[17] Valarmathi S., Harathi P.B., PrashanthiDevi M., Guhan P. and Balasubramanian S. (2008) *Geoinformation Technology for Better Health*, 141-144.

[18] Yijun S., Goodison S., Jian L., Liu L. and Farmerie W. (2007) *Bioinformatics*, 23(1), 30-37.

[19] Zurada J.M., Duch W. and Setiono R. (2004) *Proceeding of the IEEE*, 92(5), 771-805.