Available online at http://www.journalcra.com

International Journal of Current Research
Vol. 4, Issue, 09, pp.072-079, September, 2012

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

ISSN: 0975-833X

RESEARCH ARTICLE

PREDICTION OF RISK IN BREAST CANCER USING FUZZY LOGIC TOOL BOX IN MATLAB ENVIRONMENT

[1,*] Valarmathi, S., [2]Ayesha Sulthana, [3]Ramya Rathan, [2]Latha, K. C., [1]Balasubramanian, S. and [1]Sridhar, R.

[1]Bharathiar University - DRDO-BU-CLS, Coimbatore, Tamil Nadu, India
[2]JSS University - Department of Water and Health, Mysore, Karnataka, India
[3]JSS University - Department of Anatomy, Mysore, Karnataka, India

ARTICLE INFO

Article History:
Received 16th June, 2012
Received in revised form
18th July, 2012
Accepted 08th August, 2012
Published online 30th September, 2012

Key words:
Breast cancer risk,
Fuzzy Expert System,
Association rule,
ID3 Algorithm.

ABSTRACT

Breast cancer risk assessment in western region of Tamil Nadu in India was carried out based on TNM (Tumor, Number of Lymph Nodes and Metastasis) staging using ID3 algorithm and Association rule. The intensity range of the breast cancer was predicted using Fuzzy logic toolbox. Out of 1,862 data the complete history of 181 breast cancer cases was obtained. The parameters involved in this study were age, sex, location, diet, year, marital status, heredity, period of illness, treatment, stages of diagnosis like Tumor [T], Number of Lymph Nodes [N] and Metastasis [M]. Eight set of fuzzy rules were used, the Mamdani max-min inference mechanism was implemented. Tumor 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), Very Serious Moderate (VSM), Serious (S) and Not Serious (NS) were used in order to give a breast cancer risk prognosis. High breast cancer risk regions in the study area were shown in map using ID3 algorithm, Coimbatore (North and South) was found to be at highest risk for breast cancer by 20% criteria in comparison with other areas. However, this 20% criteria value in the association rule showed 18 taluks including Coimbatore to be the risk regions for breast cancer. Fuzzy Expert System (FES) predicted range of the risk of the breast cancer was found to be similar with the clinical truth.

INTRODUCTION

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 [1].This approach spares many breast cancer patients from receiving complex surgical biopsies, unnecessary adjuvant treatments and its expensive medical cost [2]. In situations where experienced oncologists are not available, predictive models created with data mining techniques with acceptable accuracy can be used to support physicians in decision making [3]. Identifying salient features contributing to the incidence of the breast cancer was an essential factor for the prevention measures. Geocomputation has facilitated data collection and statistical report generation but however advanced autonomous techniques in exploratory analysis were not widely adopted [4]. 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. Hence, the spatial statistical analysis becomes inappropriate and unsuitable for data-rich environment [5]. 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. Hence, analysis of medical data is often concerned with treatment of incomplete knowledge and information. 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 compact and apprehensible predictive models through the visualization possibilities could help better human decisions. The computational intelligence techniques like decision trees and neural networks are important tools for the rule extraction and data understanding [6].

The application of fuzzy logic in medicine and bioinformatics has received much appreciation. Fuzzy analysis has been used as data mining technique to formulate the decision tree models using k-means algorithm [7, 8]. 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]. Association rule mining algorithm is the most frequently used data mining tool in rule extraction. Temporal association rules play an

***Corresponding author:** ayeshasulthanaa@gmail.com

important role in biological processes by associating the genes in transcriptional time and temporal dependencies between genes [10]. The technique ID3 (Iterative Dichotomiser: a decision tree method) with Kaplan-Meier analysis was used to determine the occurrence-free survival of breast cancer patients [11]. 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 [12].

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 L A, at the University of California, Berkeley [13]. Zadeh L A. stated that very early diagnosis would be the application domain for Fuzzy Logic [14]. Blechner M D has reviewed that Fuzzy Logic allow membership values between 0 and 1, arguably it can provide a more realistic representation of biologic, image analysis data that was inherently noisy and imprecise [15]. The diagnosis of breast cancer involves several levels of uncertainty which manifests itself quite differently, depending on the patient and surrounding environment and intensity [7]. 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, stage, treatment, heredity, environmental factors, etc. 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 following objective:

i) To identify the breast cancer risk region based on TNM (Tumor, Number of Lymph Nodes and Metastasis) staging using ID3 algorithm and Association rule and
ii) To predict the intensity range of the breast cancer using Fuzzy Logic in MATLAB environment.

Data Used and Methodology

The data for the present cross-sectional study on cancer was collected from the 2001 to 2006 records of:

i) National Cancer Research Programme (NCRP), Bangalore
ii) Government Hospitals of Coimbatore, Erode, Namakkal, Salem and Nilgiri, Tamil nadu
iii) VNCC (Cancer Centre), Oncology wing, G. Kuppuswamy Naidu Memorial Hospital, Coimbatore, Tamil Nadu
iv) Department of Oncology, Surgical Oncology and Radiation Oncology,Sri Ramakrishna Hospital, Coimbatore, Tamil Nadu

The taluks of the districts covered for the study is presented in Table 1. There are 29 taluks from the five districts in the Western region of Tamil Nadu. The cancer incidences from the Western region of Tamil Nadu were collected from hospital registers with respect to the study area. The data collection was focused on hospitals with Oncology department or special cancer wards. After removal of erroneous or incomplete records, the collected overall data on cancer cases were accounted to 12,595. From this data cases with breast cancer reports were segregated and reported as 1,862. Taluk wise data excluding Valparai (no incidence) were prepared and used for further analysis. The population exposed for each taluk was collected from the census year book, 2001 [Statistics available in the Tamil Nadu Government website (Url -1)]. Out of 1,862 data the complete history of the breast cancer patient was obtained from 181 breast cancer cases. The parameters which were taken for the study in breast cancer data are location, age, sex, diet, year, marital status, 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. Several methods (C & RT, CHAID, QUEST, C4.5 and ID3 algorithm) including COX regression model are used to investigate the survival of breast cancer patients. Each method has its own advantage and disadvantage. Recently ID3 algorithm was used for predicting breast cancer survivability [16]. In addition, association rules were also used for the same purpose. Therefore, we have adopted soft computing methods such as:

- ID3 Algorithm
- Association Rule and
- Fuzzy Logic

Using this aforementioned data, the breast cancer risk assessment in Western region of Tamil Nadu was carried out.

ID3 Algorithm

The ID3 (Iterative Dichotomiser 3) technique was used to build a decision tree based on information theory and attempts to minimize the expected number of comparisons. It was first introduced by Quinlan in 1979 and uses Information Theory invented by Shannon in 1948 [17-18]. The formal definition of entropy is shown in Definition 1. The entropy value lies between 0 and 1 and reaches a maximum when the probability was found to be the same.

Definition 1: Given probabilities p1, p2,…………….ps where ∑i=1 pi = 1, the entropy is defined as

$$H(p1,p2,\ldots,ps)= \sum_{i=1} (pi \log[1/pi]) \qquad (1)$$

The database state, D, H(D) finds the amount of order in that state. When the state was split into S new states, S = {D1, D2,….Ds}, the entropy of those states can be seen. Each step in ID3 chooses the state that orders splitting the most. The information was calculated by determining the differences between the entropies from each of the subdivided datasets. The entropies of the split datasets were weighted by the fraction of the dataset being placed in that division. The ID3 algorithm is based on Occam's razor: it prefers smaller decision trees over larger ones. However, it does not always produce the smallest tree, and is therefore a heuristic. Occam's razor is formalized using the concept of information entropy:

$$I_E(i) = - \sum_{j=1}^{m} f(i,j) \log_2 f(i,j) \qquad (2)$$

The ID3 algorithm calculates the gain of a particular split using the following formula:

$$Gain\ (D,S) = H(D) - \sum P(D_i)H(D_i) \qquad (3)$$

The information gain is used to select the most useful attribute for classification. It can be summarized through the following steps:

- Take all unused attributes and count their entropy concerning test samples
- Choose attribute for which entropy is minimum and
- Make node containing that attribute.

Association Rule

Association analysis is the widely researched topic in data mining to generate hypothesis rather than to test using statistical techniques [19]. Association rule typically consists of three parts – an antecedent (X), a consequent (Y) and a measure of the interestingness of the rule (support%, confidence% and lift), as represented in the following equation;

$$X \rightarrow Y\ (support\%, confidence\%, lift) \qquad (4)$$

The antecedent and the consequent are a set of one or more predicates. The support of a rule measures the frequency of collective occurrence of all the antecedent and consequent predicates of a rule in the dataset [20]. Each rule has an associated support and confidence, which is defined as:

Support is an estimate for $Pr[X \cap Y]$/Total number of records ……..(5)

Confidence is an estimate for $Pr[x \cap Y]/Pr[X]$ ……………..(6)

The support is the ratio of transactions that satisfy both X and Y to the number of transactions in databases. The confidence is the conditional probability of Y given X. Since users are interested in large support and high confidence, two thresholds are used to find strong association rules [21].

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. This method was studied in Wisconsin breast cancer classification problem by [22].

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.

$$A = \{x, \mu_A(x) \mid x \in X\} \qquad (7)$$

$\mu_A(x)$ 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' (Url –2). The seriousness of the breast cancer can be 339 predicted by applying the algorithm (Algorithm.docx). Fuzzy Inference System for Tumor, Lymph nodes and Metastasis is also shown in the algorithm.

Table 1.Taluks in the Districts of Western Region of Tamil Nadu

S. No	District	Taluks
1.	Old Coimbatore	Avinashi Coimbatore North Coimbatore South Mettupalayam Palladam Pollachi Tirupur Udumalpet Valparai
2.	Erode	Sathyamangalam Gobi Bhavani Perundurai Dharapuram Erode
3.	Namakkal	Paramathivelur Namakkal Rasipuram Thriuchengode
4.	Salem	Omalur Yercaud Sankagiri Salem Attur Mettur
5.	Nilgiris	Ooty Coonoor Kotagiri Gudalur

Table 2. Risk Area of the Breast Cancer for each Taluk

Field	Very Serious	Very Serious moderate	Serious	Not Serious	Total	Entropy	Weighted Sum
Avinashi	1	1	0	0	2	0.7	0.01
Bhavani	1	1	2	0	4	1.05	0.02
Coimbatore North	14	15	15	2	46	1.23	0.31
Coimbatore South	7	17	15	4	43	1.25	0.3
Coonoor	0	1	1	0	2	0.7	0.01
Dharapuram	1	1	4	0	6	0.87	0.03
Erode	7	4	8	1	20	1.21	0.13
Gobi	0	1	2	0	3	0.64	0.01
Mettupalayam	1	5	2	0	8	0.91	0.04
Mettur	0	0	1	0	1	0	0
Namakkal	0	1	3	0	4	0.57	0.01
Ooty	0	2	2	1	5	1.06	0.03
Palladam	1	2	1	0	4	1.05	0.02
Pollachi	0	4	5	0	9	0.68	0.03
Salem	2	2	1	0	5	1.06	0.03
Sathyamangalam	2	0	1	0	3	0.64	0.01
Thiruchengode	0	1	0	0	1	0	0
Tirupur	4	3	5	0	12	1.08	0.07
Udumalpet	0	1	2	0	3	0.64	0.01

Table 3. Grouping of Stage in Cancer

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

Table 4. Developed Coding for the Prediction of Seriousness in MATLAB platform

Rule Number	Tumour	Node	Metastatis
1.	Very Small	1	No
2.	Small	1	No
3.	Medium	2	No
4.	Large	2	No
5.	Medium	3	No
6.	Large	No	No
7.	Large	3	No
8.	Small	1	Yes

Table 5. 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	4.4	NS
0	1	0	-	1-3 nodes	-	S	18	S
1	1	0	<2 cm	1-3 nodes	-	S	20	S
2	0	0	>2-5 cm	-	-	S	15	S
2	1	0	>2-5 cm	1-3 nodes	-	S	14.8	S
3	0	0	>5 cm	-	-	S	15	S
0	2	0	-	4-9 nodes	-	VSM	25	VSM
1	2	0	>2 cm	4-9 nodes	-	VSM	25	VSM
2	2	0	>2-5 cm	4-9 nodes	-	VSM	25	VSM
3	1,2	0	>5 cm	1-9 nodes	-	VS	34.8	VS
4	0,1,2	0	CW/S	>1 node	-	VS	34.8	VS
1-4	3	0	Any	>9 nodes	-	VS	34.8	VS
1-4	1-3	1	Any	>1 node	yes	VS	34.8	VS

Range	:	Risk
0 – 10	:	NS (Not Serious)
11-20	:	S (Serious)
21 – 30	:	VSM (Very Serious Moderate)
31 – 40	:	VS (Very Serious)

RESULTS

ID3 Algorithm

In this study, ID3 chooses the splitting attribute with the highest gain in information, where gain is defined as the difference between how much information was needed to make a correct classification before the split versus how much information was needed after the split. It is applied for age, location, taluk, stage, year, period, marital status, treatment, heredity, sex and habitat against Very Serious (VS), Very Serious Moderate (VSM), Serious (S) and Not Serious (NS) to calculate the gain of information. The ranked histogram gives the gain of each field for the breast cancer data, which is presented as bar diagram in the Figure 1. The doctors use TNM staging which will decide the risk level of the breast cancer which is found to be a linguistic variable. The gain of the ID3 chart is shown in the Figure 2. From the graph it was found that the stage, which constitutes tumor, number of lymph nodes and metastasis play an important decision making field and was used in fuzzy logic for perception based measurement. Spatial risk area (taluk) of the breast cancer was calculated using the equation (3) and is presented in Table 2. From the Table, it was observed that Coimbatore (North and South) was found to be risk region to the breast cancer than other areas at 20% criteria. Weighted value of taluk was compared with criterion value and integrated with Map Object to visualize the results, which are presented in the Map 1. This map shows the high breast cancer risk regions in the study area using ID3 algorithm.

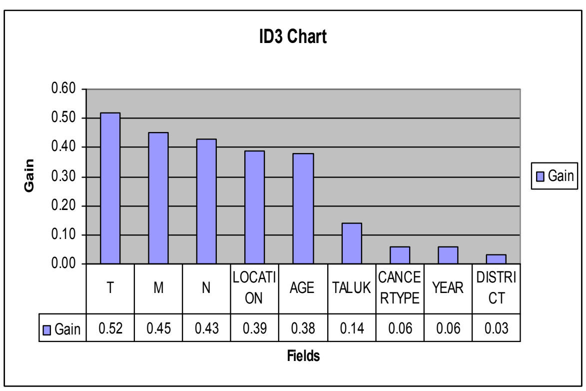


Figure 1. Gain of each Field for Breast Cancer

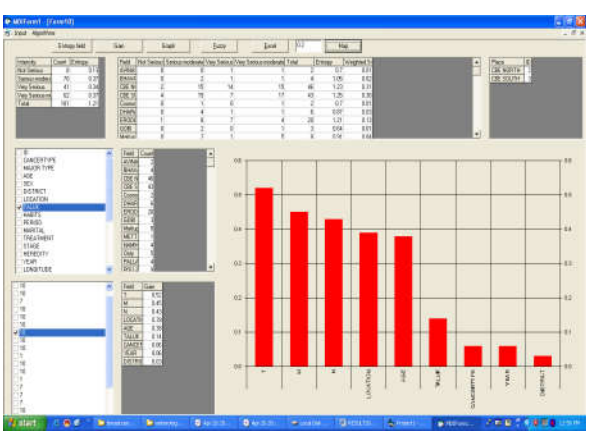
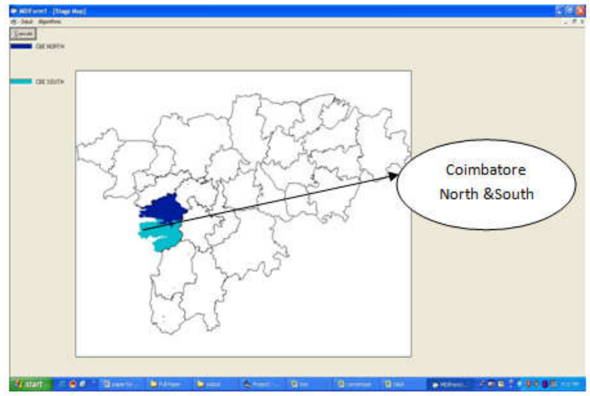


Figure 2. Gain of ID3 Chart



Map 1. High Risk of Breast Cancer in Study Area using ID3

Association Rule

Data mining methods for breast cancer survivability analysis using SEER (Surveillance Epidemiology and End Results) data were compared but the results of this analysis showed different survivability rates [23]. Hence, Bellaachia*et al.*used three data mining technique Naïve Bayes, the back-propagated neural network and the C4.5 decision tree algorithms [16]. For the present study, Association Rule in addition to ID3 algorithm was used to predict the occurrence accurately. Ranking was carried out using association rule for the breast cancer incidences. The gain of the Association Rule is shown in the Figure 3.

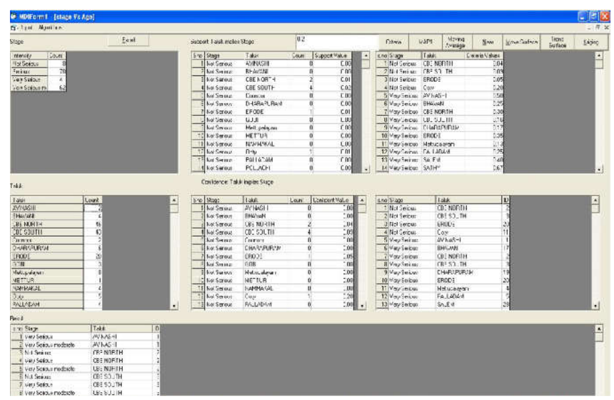
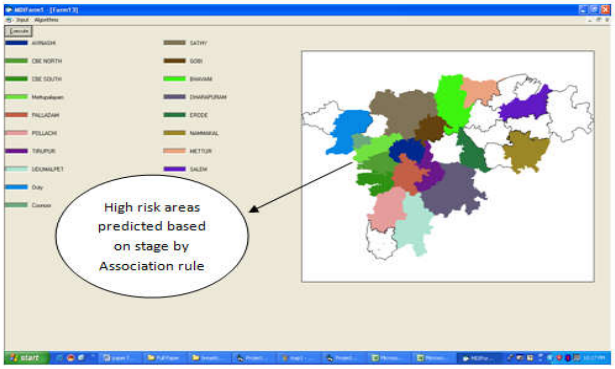


Figure 3. Gain of Association Rule

Stages such as IV, IIIA, IIIB, IIIC, IIA, IIB, I and 0 (Zero) were observed as Very Serious, Very Serious Moderate, Serious and Not Serious (Table 2) to calculate the support and confidence. This is further integrated with Map Object to view the map with the same criterion value used in ID3 algorithm. The output classification such as Very Serious, Very Serious Moderate, Serious and Not Serious can be associated with each taluk. According to the 20% criteria value in the association rule, 18 taluks from the study area were found to be the risk regions for the breast cancer which is observed in Map 2. It is clearly predicted that the stage of the breast cancer have higher ranks than the other factors. While comparing the results obtained from ID3 algorithm with Association Rule at 20% criteria, ID3 algorithm predicted Coimbatore North and South as the high risk region. However, the Association Rule showed many taluks as high risk region including. Coimbatore North and South. This showed that the significant



Map 2. Map of Risk Regions in the Study Area using Association Rule

and more precise answer is obtained through ID3 algorithm than the Association Rule. 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 [24]. Phillips et al. (2006) used fuzzy logic model to predict breast cancer using volatile biomarkers in their breath [25]. 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 3. The data was further classified into TNM in the database and adopted for the Fuzzy Logic. For further diagnosing, analyzing and understanding the breast cancer, will determine the need for biopsy (level of risk) and also predict the range of risk of breast cancer in the individual. As seen from the literature, it is quite impossible to diagnose the breast cancer only using mammography and image processing technique. Instead a rule based Fuzzy Expert System (FES) is developed in this study which uses the laboratory data and simulates an expert doctor's behavior.

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 (1995) have adopted fuzzy logic for making a prognostic decision in breast and prostate cancers [21]. Later several methods were adopted for the classification of the biological data sets using fuzzy inference system. Recently, Bellaachia *et al.* have used three data mining techniques like C4.5 decision tree algorithm, neural network and Naïve Bayes for the breast cancer survivability studies [16]. Among this, C4.5 algorithm provides a better performance than the other two techniques adopted by them. However, these algorithms will rank the breast cancer related parameters, but this is not sufficient for the diagnosis. Therefore, in our study we have incorporated a fuzzy expert system using ID3 algorithm. 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 of the breast cancer using split field provided by ID3 algorithm and integrated with MATLAB environment for Fuzzy logic.

Fuzzy logic

For the design process, breast cancer tumor 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), Very Serious Moderate (VSM), 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). Parts of the developed fuzzy rules using MATLAB platform are shown in the Table 4. Totally eight Fuzzy rules were formed and the output is shown in the Figure 4. The performed Fuzzy rules predict the seriousness of the breast cancer at their respective stage.

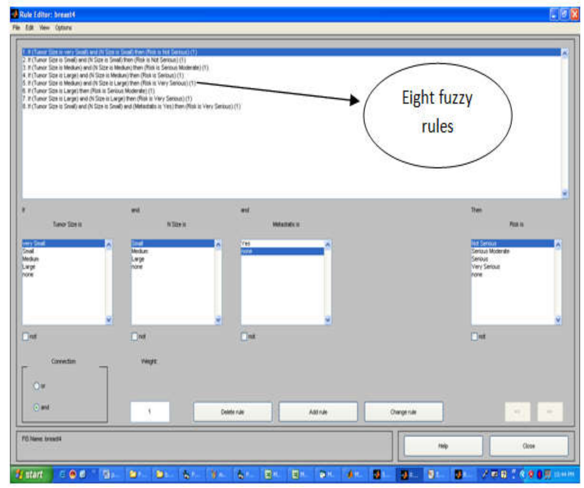


Figure 4. Fuzzy Rules

Rule 1: If Tumour= very small and Node = 1, then Metastasis = No, i.e., if the patient's Tumour size was very small and the nodal region as small, with no metastasis, then the patient's breast cancer risk was found to be serious (Stage IIA).

Rule 2: If Tumour= small and Node = 1, then Metastasis = No, i.e., if the patient's Tumour size and the nodal region was small, with no metastasis, then the patient's breast cancer risk was observed as serious (Stage IIB).

Rule 3: If Tumour= medium and Node = 2, then Metastasis = No, i.e., if the patient's Tumour size and the nodal region was medium, with no metastasis, then the patient's breast cancer risk was predicted to be very serious moderate (Stage IIIA).

Rule 4: If Tumour= large and Node = 2, then Metastasis = No, i.e., if the patient's Tumour size was large and the nodal region as medium, with no metastasis, then the patient's breast cancer risk was given as very serious moderate (Stage IIIB).

Rule 5: If Tumour= medium and Node = 3, then Metastasis = No, i.e., if the patient's Tumour size was medium and the nodal region as large, with no metastasis, then the patient's breast cancer risk was observed as very serious moderate (Stage IIIC).

Rule 6: If Tumour= large and Node = No, then Metastasis = No, i.e., if the patient's Tumour size was large and with no nodal region and metastasis, then the patient's breast cancer risk was found to be as very serious moderate (Stage IIIB).

Rule 7: If Tumour= large and Node = 3, then Metastasis = No, i.e., if the patient's Tumour size and the nodal region in the patient was large, with no metastasis, then the patient's breast cancer risk was observed as very serious moderate (Stage IIIC).

Rule 8: If Tumour = small and Node = 1, then Metastasis = Yes, i.e., if the patient's Tumour size and the nodal region as small, with metastasis, then the breast cancer risk in the patient was predicted as very serious (Stage IV). This implies that the present condition of the patient was in the final stage and very much prone to death.

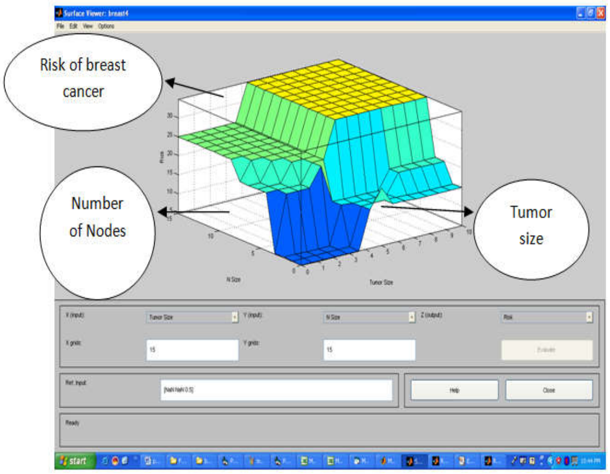


Figure 5. 3D Surface Visualization of TNM with respect to Risk in Breast Cancer

Fuzzification of the used parameters is made by the membership functions of the Fuzzy Logic. The 3D surface visualization of TNM staging with respect to risk in the breast cancer is presented in the Figure 5. The results of the developed Fuzzy Expert System are compared with the results obtained through recent literature (Table 5). From the Table 5, 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 for the medical students specializing in the cancer epidemic. The breast cancer risk reviewed through Fuzzy logic is shown in the Figure 6.

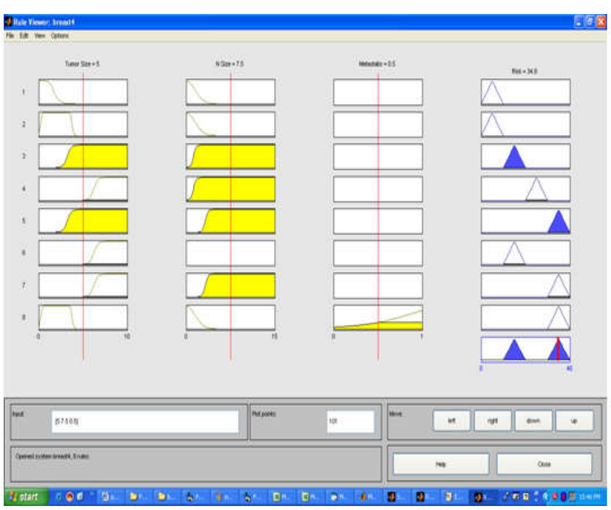


Figure 6. Fuzzy Rule Viewer of the Breast Cancer Risk for the TNM

Many biological variables were characterized by the non-linear characteristics that are better represented by an interval than by a binary process [26]. 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 [27].

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 for solving problem rather than attempting to model a system mathematically. ID3 is a decision tree constructed through criterion tests. Though ID3 is compact and intelligent, it is not incremental in itself. Hence, it sharpens the decision and identifies the influence, as additional techniques are sought. Association rule is a method to identify relationships between variables in large databases through strong rules using different measures of interests. Therefore, Fuzzy Logic was adopted as an additional integrated tool, which is a classical tool that shares the properties of ID3. Through this, the performance of the data and the precision of the outcome can be enhanced. The study has outlined, discussed and resolved the algorithms, techniques / methods adopted through soft computing methodology like ID3 algorithm, Association Rule and 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 by the specialist doctors for treatment and by the students for learning 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.

REFERENCES

Amir A, Evans D G, Shenton A, Lalloo F, Moran A, Boggis C, Wilson M and Howell A (2003) "Evaluation of Breast Cancer Risk Assessment Packages in the Family History Evaluation and Screening Programme" Journal of Med. Genet., 40: 807 – 814.

Baxt W G (1994) "Complexity, Chaos and Human physiology: the Justification for Non-Linear Neural Computational Analysis" Cancer Letters, 77: 85-93.

Bellaachia A and Guven E (2006) "Predicting Breast Cancer Survivability using Data Mining Techniques" 9th Workshop on Mining Scientific and Engineering Datasets in conjunction with the 6th SIAM International Conference on Data Mining (SDM 2006).
Biotechnology, Hindawi Publishing corporation, Article ID 91908, 1-7.

Blechner M D (2005) "Behaviour of Various Machine Learning Models in the Face of Noisy Data" Harvard-MIT Division of Health Sciences and Technology, J. Medical Decision Support, Final project.

Brand M R, Jones D D, Lynch T H, Brand E R, Watson P, Ashwathnayaran R and Roy K H (2006) "Risk of Colon

Cancer in Hereditary Non-polyposis Colorectal Cancer Patients as Predicted by Fuzzy Modeling: Influence of Smoking" World Journal of Gastroenterology, 12(28): 4485-4491.

Castro V E and Lee I (2001) "Data Mining Techniques for Autonomous Exploration of Large Volumes of Geo-referenced Crime Data" Proc. 6th International Conference on Geocomputation.

Delen D, Walker G and Kadam A (2005) "Predicting Breast Cancer Survivability: A Comparison of three Data Mining Methods" Artificial Intelligence in Medicine, 34(2): 113-127.

Guo G and Neagu D (2005) Fuzzy k-NNmodel Applied to Predictive Toxicology Data Mining" International Journal of Computational Intelligence and Applications, 5(3): 321 – 333.

He Y, Tang Y, Zhang Y Q and Sunderraman R (2006) "Adaptive Fuzzy Association Rule Mining for Effective Decision Support in Biomedical Applications" International Journal of Data Mining and Bioinformatics, 1(1): 3-18.

Hipp J, Guntzer U and Nakhaeizadeh G (2000) "Algorithms for Association Rule Mining – A General Survey and Comparison" SIGKDD Explorations, 2: 558-564.

Khan U, Shin H, Choi J P and Kim M (2008) "wFDT-Weighted Fuzzy Decision Trees for Prognosis of Breast Cancer Survivability" 7th Australasian Data Mining Conference, Glenelg, Australia, 87.

Koperski K and Han H (1995) "Discovery of Spatial Association Rules in Geographic Information Databases" Proceedings of the 4th International Symposium on Large Spatial Databases, 47-66.

Mevlut, Ture, Fusun, Tokatli, Imran and Kurt (2009) "Using Kaplan-Meier Analysis together with Decision Tree Methods (C & RT, CHAID, QUEST, C4.5 and ID3) in Determining Recurrence-Free Survival of Breast Cancer Patients" Expert Systems with Applications: An International Journal archive, 36(2): 2017-2026.

Miller H J and Han J (2001) "Geographic Data Mining and Knowledge Discovery: An Overview" Cambridge University Press: Cambridge, in press, 372.

Nam H, Lee K J and Lee D (2009) "Identification of Temporal Association Rules from Time-series Microarray Data Sets" BMC Bioinformatics, 10(S6): 1-9.

Pach F and Abonyi J (2006) "Association Rule and Decision Tree Based Methods for Fuzzy Rule base Generation" Proceedings of world academy of science, Engineering and Technology, (13): 45-50.

Phillips M, Cataneo N R, Ditkoff B A, Fisher P, Greenberg J, Gunawardena R, Kwon C S, Tietje O and Wong C (2006) "Prediction of Breast Cancer using Volatile Biomarkers in the Breath" Breast Cancer Research and Treatment, Springer.

Quinlan J R (1979) "Discovering Rules by Inducting from Large Collections of Examples" In D. Michie (ed.), Expert Systems in the Micro Electronic Age, Edinburgh University Press, 168 – 201.

Quinlan J R (1986) "Induction of Decision Trees" Mach. Learn. 1, 1: 81-106.

Sahan S, Polat K, Kodaz H and Gunes S (2007) "A New Hybrid Method based on Fuzzy-artificial Immune System and k-NN Algorithm for Breast Cancer Diagnosis" Computers in Biology and Medicine, 37(3): 415 -423.

Schneider J, Peltri G, Bitterlich N, Philipp M, Velcovsky H G, Morr H, Katz N and Eigenbrodt E (2003) "Fuzzy Logic-based Tumor Marker Profiles Improved Sensitivity of the Detection of Progression in Small-cell Lung Cancer Patients" Clin. Exp. Med., 2: 185-191.

Torres A and Nieto J J (2005) "Fuzzy logic in Medicine and Bioinformatics" Journal of Biomedicine and

Vinnakota S and Lam S N N (2006) "Socioeconomic Inequality of cancer mortality in the United States: a spatial data mining approach" International Journal of Health Geographics, 5(9): 1-12.

Yijun S, Goodison S, Jian L, Liu L and Farmerie W (2007) "Improved Breast Cancer Prognosis through the Combination of Clinical and Genetic Markers" Bioinformatics, 23(1): 30 -37.

Zadeh L A (1965) "Fuzzy sets" Information control, 8: 35-40.

Zadeh L A (1969) "Biological Application of the Theory of Fuzzy Sets and Systems" Proc. Int. Symp. Biocybernetics of the Central Nervous System, Little Brown and Co., Boston, 199-212.

Zurada J M, Duch W and Setiono R (2004) "Computational Intelligence Methods for Rule-based Data Understanding" Proceeding of the IEEE, 92(5): 771-805.

****