Original Research Article

Socioeconomic and Age-incidence of Breast Cancer: Modeling Using Artificial Intelligence Technique

Purpose: The majority of women presenting with breast cancer it are not possible to identify specific risk factors. Age is the major factor on breast cancer incidence. Also, poverty status can be classified. Because of the weakness of the underlying empirical data in many countries, a number of the indicators presented here are associated with significant uncertainty. The fuzzy logic inference method as an artificial intelligence technique is proposed for modeling data. Methods: In our situation it is very difficult to use classical logic to model a system with the available knowledge. Classical logic does not allow working with uncertainty in the information when knowledge about the behavior of the systems is imprecise. A fuzzy system was constructed with three inputs parameters and one output expressing the number of cases. Results: The result of the fuzzy program so far, is a numeric and symbolic terms of number of breast cancer recorded; using the fuzzy inputs data in the universe of discourse (poor, near poor or non-poor), age and period. Conclusion: Once the established system, it allows to predict the impact of each input and its effect on the output parameter. Assessing the degree of impact allows us to define the set the factor that has the greatest impact in the fight against breast cancer. The result is the contribution of the set of input variable, taking into account inaccuracies and the complexity involved in the process.

Key Words: Breast cancer, Epidemiology, Incidence factors, Fuzzy logic

INTRODUCTION

The global burden of cancer continues to increase largely because of the aging and growth of the world population alongside an increasing adoption of cancer-causing behaviors, particularly smoking, in economically developing countries [1], [2],[3]. Several risk factors for breast cancer have been well documented. [4],[5].

However, for the majority of women presenting with breast cancer it is not possible to identify specific risk factors. Age specific incidence rates, increase rapidly with age until about 45-50 years of age, after which they continue to increase but at a slower rate. In addition to age, a few other risk factors, socioeconomic status is associated with relative risks [6]. Age is the major factor on breast cancer incidence based on the incidence data for 1987-2010.

The impact of incidence is steadily increasing from 50 to 74 years old. This burden reflects also in the black women in USA, having similar impact on age incidence as in the white women, although the incident rate is slightly lower. A projection for breast cancer suggests a continuous growth in the number of incident cases that would increase [7]. For most cancer sites there is a linear log–log relationship between incidence and age. This relationship does not hold for breast cancer, and certain 'key' breast cancer risk factors suggest that breast tissue does not 'age' in step with calendar time. A quantitative description of 'breast tissue

age' is suggested which brings the age–incidence curve of breast cancer into line with the common log–log cancers and explains quantitatively the known key risk factors [8]. For results classified by poverty status, population estimates were grouped into three categories according to the percent of the population in the census-tract living below the federally defined poverty threshold. The conventional view that breast cancer is a "disease of affluence" is increasingly at odds with the empirical evidence and lived experiences of poorer women and women of color diagnosed with breast cancer. The amount of missing information may vary between subgroups and can change over time [9],[10],[11],[12]. Where necessary the estimates provided have been derived from multiple sources, depending on each indicator and on the availability and quality of data. Statistical and health information systems are weak and the underlying empirical data may not be available or may be of poor quality. Because of the weakness of the underlying empirical data in many countries, a number of the indicators presented here are associated with significant uncertainty [13]. In this situation it is very difficult to use classical logic to model a system with the available knowledge. Classical logic does not allow working with uncertainty in the information when knowledge about the behavior of the systems is imprecise. The proposed system utilizes fuzzy logic.

METHODS

Fuzzy logic inference

The fuzzy logic approaches, a sub-field of intelligent systems, are being widely used to solve a wide variety of problems in medical, biological and environmental applications. One of the most important areas of application of fuzzy set theory as developed by Zadeh [14] is Fuzzy Rule-Based System. These fuzzy logic systems constitute an extension of the classical rule-based systems, because they deal with "if-then" rules whose antecedents and consequences are composed of fuzzy logic statements, instead of classical logic ones. In a broad sense, a fuzzy rule-based system is a rule-based system where fuzzy logic is used as a tool for representing different forms of knowledge about a problem, as well as for modeling the interactions and relationships that exist between its variables. Due to this property, fuzzy logic principles have been successfully applied to a wide range of problems in different domains for which uncertainty and vagueness emerge in varying ways. Fuzzy modeling [15], fuzzy control [16], and fuzzy classification [17] are the most common applications.

Fuzzy logic deals with reasoning on a higher level, using linguistic information acquired from domain experts. The above-mentioned capabilities make fuzzy logic a very powerful tool to solve many ecological problems, where data may be complex or in an insufficient amount. The fuzzy logic concept provides a natural way of dealing with problems where the source of imprecision is an absence of sharply defined criteria rather than the presence of random variables [18]. The fuzzy approach considers cases where linguistic uncertainties play some role in the control mechanism of the phenomena concerned [19]. Fuzzy inference systems (FIS) are powerful tools for the simulation of nonlinear behaviors with the help of fuzzy logic and linguistic fuzzy rules [20], [21].

For example, there is not a straight-line relationship between the socioeconomic, age and breast cancer incidence.

In this study, we take to decision algorithms using the engine that makes inferences on a fuzzy rule system. For all the algorithms presented below there is a common rule form for rules that associate an observation vector.

a= (a(1), a(2), …. a(n)) with a diagnosis. Further, we assume the following general form of the kth rule in the system (k = 1, 2, …K):

If a(1) is A1k AND …AND a(n) is An,k THAN b is Bk where Aik, are fuzzy sets (whose membership functions are designated by (μAi,k) that correspond to the nature of particular observations (for simplicity we assume the sets to be triangular fuzzy numbers) whereas k B is a discrete fuzzy set defined on the result set, with the Bk membership function. The particular decision algorithms to be used in plant effect have in common both the inference engine and the procedure for rule system derivation from the learning set. In the proto-formal deduction rule, the syllogism:

Q_1 A's are B's AND Q_2(A&B)'s are C's THAN Q_1Q_2A's are (B&C)'s [22].

Fuzzy Logic Modeling

A most studies interest exists for evaluating effect of age and socioeconomic impact of the breast cancer. In our case, we can introduce the relationship between these parameters as input and risks of damage by breast cancer as output variable Tables 1.

Tables 1. Mammography Prevalence (%) within the Past Two Years by Age and Poverty Status*, US, Selected Years 1987-2010

	40-49			50-64			65 years and over		
Year	**Poor**	**Near poor**	**Non-poor**	**Poor**	**Near poor**	**Non-poor**	**Poor**	**Near poor**	**Non-poor**
1987	19	18	44	15	24	45	13	20	35
1990	32	39	69	30	40	72	31	39	61
1991	33	44	70	37	50	73	35	42	63
1993	36	48	70	47	47	79	40	48	71
1994	43	48	70	46	48	78	44	49	73
1998	45	47	73	53	62	83	52	58	71
1999	51	53	77	63	65	83	58	60	77
2000	47	44	76	62	68	87	55	60	82
2003	51	54	72	58	64	85	57	63	73
2005	42	50	74	50	59	84	49	56	73
2008	47	47	73	57	59	84	49	59	78
2010	48	46	74	55	57	83	51	56	75

*Poor persons are defined as below the poverty threshold. Near poor persons have income of 100%-199% of the poverty threshold. Non-poor persons have an income 400% or more than the poverty level [23].

A fuzzy logic system is established. The input variables are age, socioeconomic status and periods. The output variable is the number of recorded cases. Figure 1.

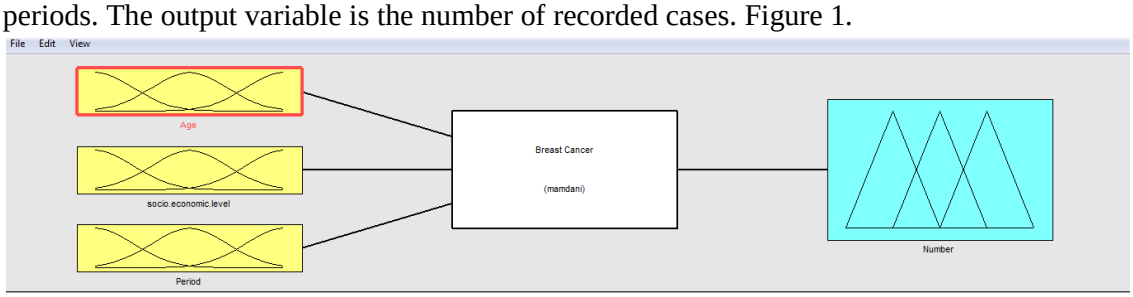


Figure 1. System block diagram with three inputs and one output

Fuzzyfication of Inputs

The data for the inputs were classified into three linguistic categories:

The input ‘Age’: The variable ‘age’ is fuzzyfied on [First tranche, Second tranche and Third trance] Figure 2.

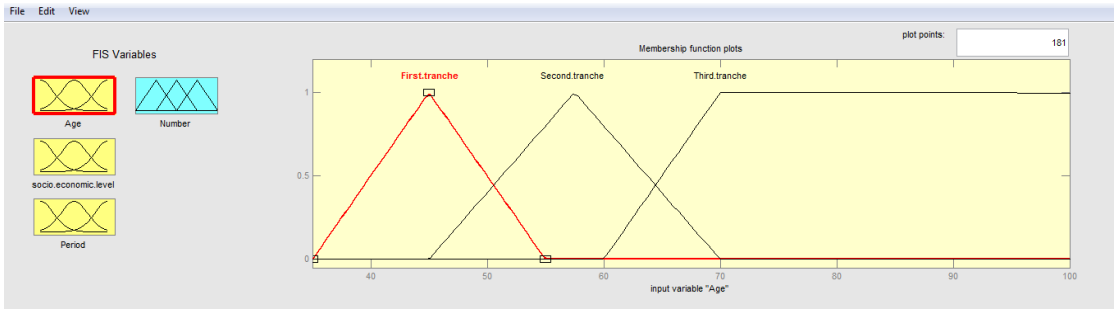


Figure 2. Fuzzyfication of input variable “Age”

The input ‘socioeconomic level: This variable is fuzzyfied on three membership function [Poor, Near poor and Non-poor]. Figure3.

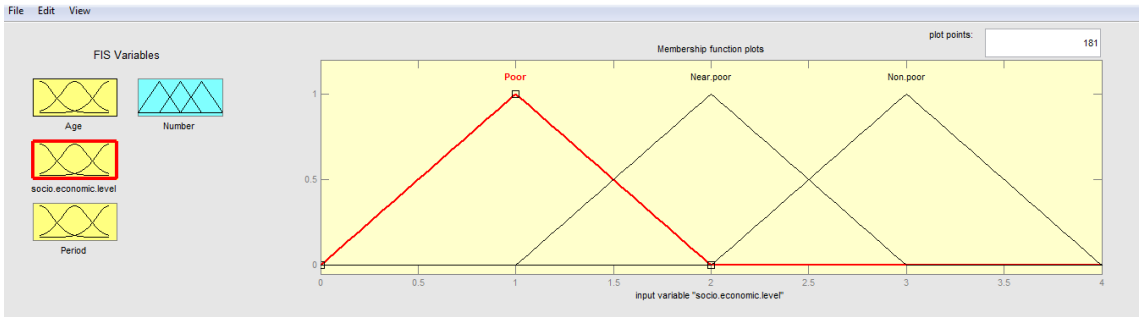


Figure 3. Fuzzyfication of input variable “Socioeconomic level”

The input ‘Period’: This variable is fuzzyfied on [Period 1 to Period 10]. Figure 4.

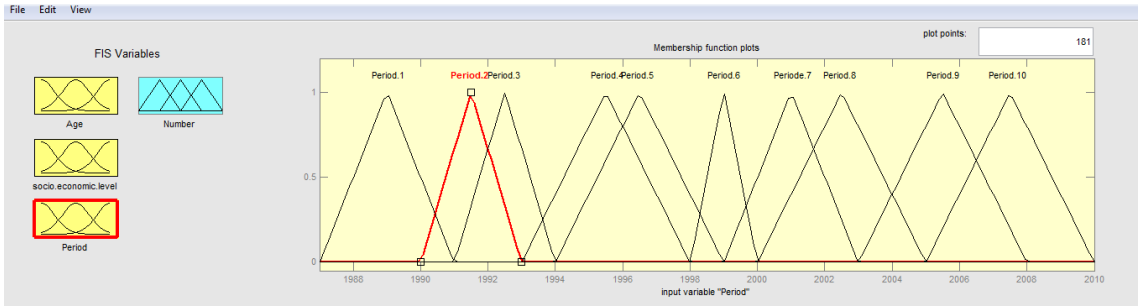


Figure 4. Fuzzyfication of input variable “Period”

Fuzzyfication of Output

The data for the output (the number of recorded cases of breast cancer) was classified into three linguistic categories: Low, Medium and High. Figure 5.

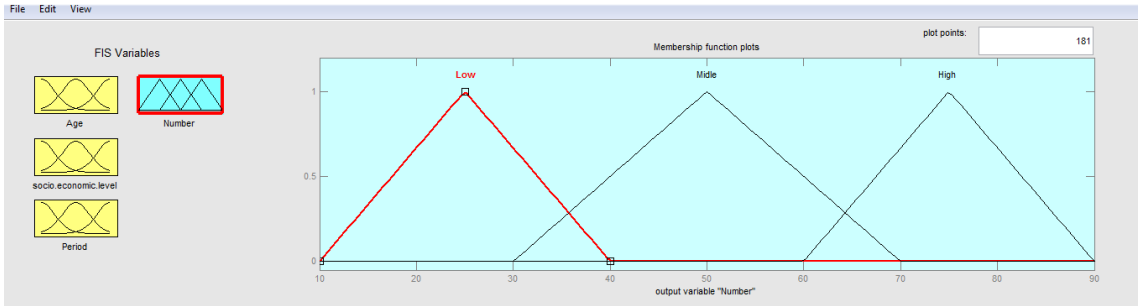
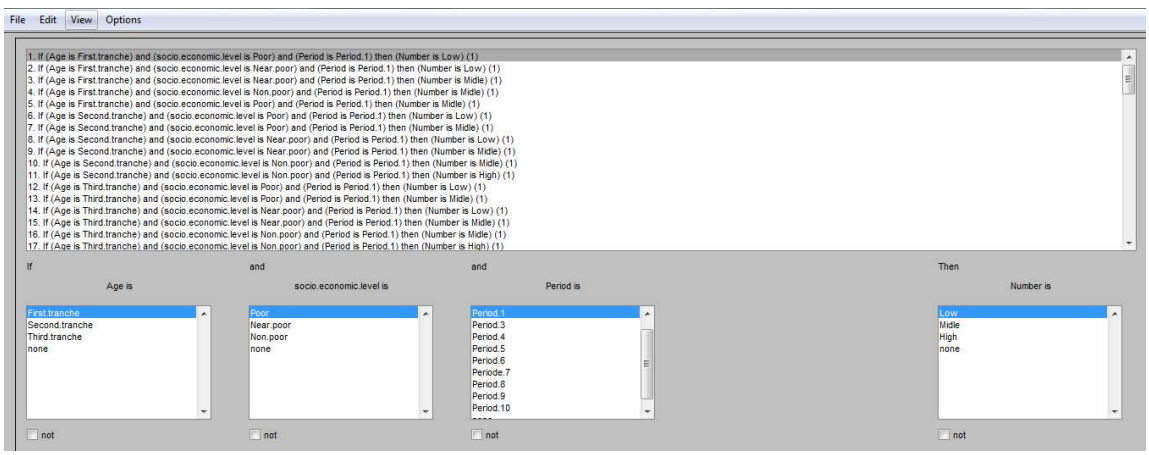


Figure 5. Fuzzyfication of output variable Number"

Fuzzy Rules

The rules determined by the choice of the fuzzy membership function are defined for each input variable. In general form, each fuzzy rule is written as were A_1 and A_2 are the fuzzy sets that describe the nature of the inputs, such as poor, near poor, or non-poor. The linguistic control rules of this system are given by:

If X_1 IS $X_1(1)$ and X_2 IS $X_2(2)$ and…X_n IS $X_n(n)$ Than Y_1 is $Y_1(1)$



Defuzzifier

This system has tow outputs that describes the impact on breast cancer incidence. In fact explains the contribution of each factor on rate that induces directly effect. The crisp value outputs are given by the defuzzyfication process after estimating its inputs values. In this system we used centre of average method which has the mathematical expression that is: (ΣSi.Ri)/(Ri).

In the defuzzyfication, the exact expression is obtained with "centroïd" method according to validity degree. The outputs values according to the inputs values obtained from the designed fuzzy engine system.

RESULTS

The factors effect on degree of impact on different aspect like environmental, economic or health system is based on fuzzy logic model. It is designed for measurement of different parameters. This system consists of tow inputs variables. The rule base of this system is used to determine the outputs parameters values: (low); (medium); or (high) for each energy type, according to the inputs values.

MATLAB-simulation is used by applying rules. Figure 6 shows an example using the MATLAB-rule viewer and simulation result.

At each level corresponding to the use of factors (age, socioeconomic level and period), the result can be displayed to the linguistic term output and the number of cases of the corresponding breast cancer. It is sufficient to set at random values of the input variables to automatically and instantaneously read the number of cases in the system output. This is also expressed in (low, medium or high).

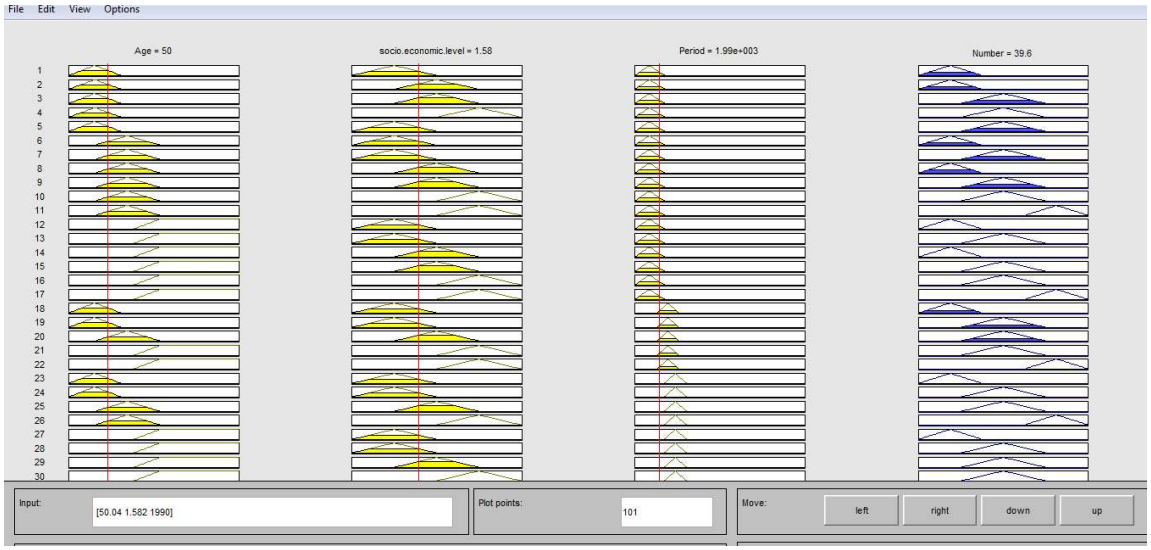


Figure 6. Application Example: attribution random variable inputs and direct reading of the output variable

The number of recorded cases of breast cancer can also be obtained based on other variable as expressed in Figure 7.

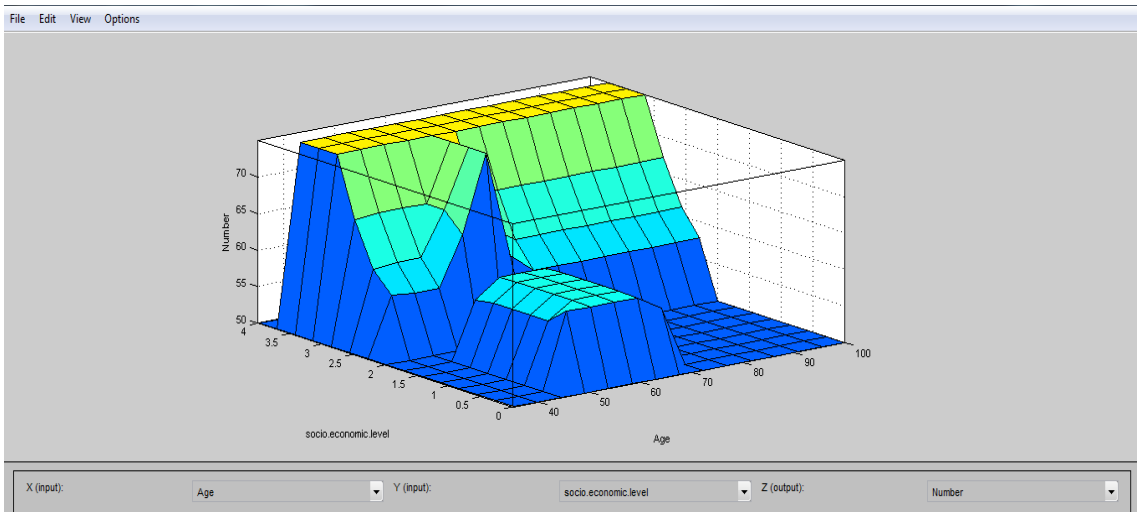


Figure 7. Application Example: attribution random variable inputs and direct reading of the output variable

CONCLUSION

The artificial intelligent system using fuzzy logic method could extend our understanding of factors affects on breast cancer. The intelligent software created in this study could be used for giving an idea about the future development of breast cancer. Like other factors are

involved in the process and that are not considered in this study, in addition to imprecision already considered in the data analysis, the system is expandable to other factors.

The goal of this study is to design and perform a pilot investigation which will provide preliminary data. Modern methods of computational intelligence such as fuzzy logic are used to achieve the medical complex analysis. The result of the fuzzy program so far, is a numeric and symbolic terms of number of breast cancer recorded; using the fuzzy inputs data in the universe of discourse (poor, near poor or non-poor). As the input parameters are characterized by uncertainty, we believe that this tool is very adequate. We emphasize that our fuzzy system is not meant to replace or substitute for an experienced physicians; on the contrary, we envisage that the fuzzy logic system should be viewed as a decision support in the most accurate. Once the established system, it allows to predict the impact of each input and its effect on the output parameter. Assessing the degree of impact allows us to define the set the factor that has the greatest impact in the fight against breast cancer. The result is the contribution of the set of input variable, taking into account inaccuracies and the complexity involved in the process.

REFERENCES

1. CA Cancer J Clin 2011;61:69–90. 2011 American Cancer Society, Inc.
2. Hamdi Cherif M. et al., Cancer estamation of incidence and survival in Algeria 2014. Nobel Research. 2015. 3(9): 100-104.
3. Hadmdi Cherif M. et al., Time trends of cancer incidence in Setif Algeria 1986-2010: An observational study. BMC Cancer. 2014. 14:637.
4. (IARC, 2008; Lacey et al., 2009). IARC (2008). World cancer report 2008. Lyon, International Agency for Research on Cancer.
5. Lacey JV Jr. et al. Breast cancer epidemiology according to recognized breast cancer risk factors in the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial Cohort. BMC Cancer, 2009. 9, 84.
6. Kelsey JL. A review of the epidemiology of human breast cancer. Epidemiol Rev. 1979;1:74-109.
7. María Paula Curado, Breast cancer in the world: Incidence and mortality. Salud Publica Mex 2011;53:372-384.
8. Pike M. C. , Krailo M. D. , Henderson B. E. , Casagrande J. T. & Hoel D. G. Hormonal' risk factors, 'breast tissue age' and the age-incidence of breast cancer. Nature 303, 767 - 770 (30 June 1983); doi:10.1038/303767a0
9. Kogevinas M, Pearce N, Susser M, Bofetta P, eds. Social Inequalities in Cancer. Lyon, France: International Agency for Research on Cancer; 1997. IARC scientific publication 138.
10. Rose DP, Royak-Schaler R. Tumor biology and prognosis in black breast cancer patients: a review. Cancer Detect Prev.2001;25:16–31. [Medline]
11. Anglin MK. Working from the inside out: implications of breast cancer activism for biomedical policies and practices. Soc Sci Med.1997;44:1403–1415. [CrossRef] [Medline]
12. Krieger N. Exposure, susceptibility, and breast cancer risk: a hypothesis regarding exogenous carcinogens, breast tissue development, and social gradients, including black/white differences, in breast cancer incidence. Breast Cancer Res Treat.1989 ;13:205–223.
13. World Health Organization 2015. WHO Library Cataloguing-in-Publication Data. World health statistics 2015. NLM classification: WA 900.1.
14. Zadeh L. Fuzzy sets. Information and Control 1(965); 8:338–353.
15. Pedrycz, W. Fuzzy modelling: paradigms and practice. Kluwer Academic Press; 1996.
16. Driankov, D.; Hellendoorn, H.; Reinfrank, M. An introduction to fuzzy control. Springer-Verlag; (1993)
17. Chi, Z.; Yan, H.; Pham, T. Fuzzy algorithms: with applications to image processing and pattern recognition. World Scientific; (1996)

18. Bouharati S.; Benmahammed K.; Harzallah D. and El-Assaf Y.M. Application of artificial neuro-fuzzy logic inference system for predicting the microbiological pollution in fresh water. Journal of Applied Sciences 2008. 8(2): 309-315.
19. Demir F, Korkmaz KA. Prediction of lower and upper bounds of elastic modulus of high strength concrete. Constr Build Mater. 2008. 22(7):1385–93.
20. Mamdani, E. H. Application of the fuzzy logic to approximate reasoning using linguistic synthesis. IEEE Transactions and Computers, 1977. C-26, 1182–1191
21. F. Allag, R. Zegadi, S. Bouharati, L. Tedjar, I. Bouharati. Dynamic of air pollution and its effect on newborns: Analysis using fuzzy logic inference system. Wulfenia Journal, part no. 2. 2013. Pp. 18-25.
22. Li-Xin W. Modelling and control of hierarchical systems with fuzzy system. Automatica J. IFAC., 1997. 33: 1041-1053.
23. National Center for Health Statistics.Health, United States, 2012: With Special Feature on Emergency Care. Hyattsville, MD. 2013. Library of Congress Catalog Number 76–641496. Cited in: American Cancer Society. Breast Cancer Facts & Figures 2013-2014. Atlanta: American Cancer Society, Inc. 2013.