2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)
July 6-11, 2014, Beijing, China

Fuzzy Breast Cancer Risk Assessment

Aniele C. Ribeiro, Deborha P. Silva, Ernesto Araujo

***Abstract*— A breast cancer risk assessment based on fuzzy set theory and fuzzy logic is proposed in this paper. The proposed fuzzy breast cancer system maps two controlled and two non–controlled input variables into the risk of breast cancer occurrence. Such an approach covers the age, menopause, simplified fuzzy body mass index (FBMI), and the existence of hormone replacement as input linguistic variable. The risk of a woman developing breast cancer is the outcome of the proposed fuzzy breast cancer risk analysis. The resulting fuzzy breast cancer diagnostic decision system for helping support to evaluate patients with such a clinical condition is an alternative to support healthcare professionals in such a complex health diagnosis.**

I. Introduction

CANCER has been considered as one of the most important public health problems. Whether in developed or developing countries, cancer represents about 12% of all causes of death worldwide each year, according to the World Health Organization (WHO) [1]. Further, it has estimated that for the year 2030, there will be about 27 million incident cases of cancer. Moreover, there will be 17 million of deaths and 75 million people will live annually with cancer. This estimation is still worse when taking into account that the majority of this health condition will affect those individuals with low or middle incomes.

The prevalence of cancer in women is 25% higher than in men and the most prevalent tumor is the breast cancer, accounting for 22% of new cases each year [1]. Affecting all age groups and socioeconomic areas, with increasing rates of death, the average five–year survival rate in the population of breast cancer in developed countries is about 85% meanwhile in developing countries it achieves around 60%. Although considered a cancer of relatively good prognosis if diagnosed and treated in time, death rates from breast cancer are kept in high values because such a disease is mostly diagnosed in advanced stages.

This disease is related to the process of industrialization – with the risk of illness associated with higher socioeconomic status – and other risk factors such as nulliparity, early age at menarche and late menopause, obesity – chiefly when the weight increase occurs after menopause –, alcohol consumption, and high–fat diets. In the counterclockwise direction, the physical activity and exclusive breastfeeding are protective factors. Family history, particularly in first–degree relatives before age 50, are also important risk factors for breast cancer and may indicate genetic predisposition associated with the presence of mutations in certain genes. Despite the existence of self–examination of breasts, annual mammograms for women over 40 years or above 35 years in women with a family history, it is also important to design mechanisms and systems that could contribute in anticipating or indicating the risk of this pathology.

Aniele C. Ribeiro and Deborha P. Silva are with the Post–Graduate and Research Institute (IPG) at the Faculdade de Ciências Médicas de Minas Gerais (FCMMG), Belo Horizonte, MG, Brazil (emails: aniele-cristina@hotmail.com and debbymg@hotmail.com).

Ernesto Araujo is with the Post–Graduate and Research Institute (IPG) at the Faculdade de Ciências Médicas de Minas Gerais (FCMMG), Belo Horizonte, MG, Brazil; the Health Informatics Department (DIS) at the Universidade Federal de São Paulo (UNIFESP), São Paulo, SP, Brazil; and the Inteligência Artificial em Medicina e Saúde Ltda. (IAMED), São José dos Campos, SP, Brazil, (emails: ernesto.araujo.br@gmail.com)

This paper addresses the use of fuzzy set theory and fuzzy logic to deal with imprecise and subjective variables to build a mathematical model able to assist healthcare professionals in assessing the risk of breast cancer in women. The use of fuzzy set theory and fuzzy logic has been demonstrating a powerful strategy in healthcare and medical modeling, prediction, diagnosis, and treatment [2], [3], [4], [5], [6], [7]. Fuzzy medical diagnostic decision support systems (FMDDS) are feasible tools to support physicians, healthcare professionals in dealing with cancer. The use of neuro–fuzzy systems for prostate cancer classification is proposed in [8], while employed for urological cancer management is found in [9]. A system developed for planning the dose for prostate cancer patients by using a fuzzy similarity measure and a modified Dempster-Shafer theory is applied to fuse the information in [10]. The use of fuzzy systems for predicting prostate cancer according to its pathological stage is employed in [11]. A fuzzy logic model for predicting vomiting in postoperative pediatric oncology patients is available in [12]. When concerning breast cancer health diagnoses, the classification of lesion in benign and malignant by employing a hybrid hidden Markov model–fuzzy approach in designed in [13].

The proposed model consists of a fuzzy rule–based system, taking into account four well established risk factors. Despite its importance and although several risk factors are already well established, finding out a method to establish the risk of a woman developing breast cancer correlated to its major risk factors is of vital importance. The fuzzy breast cancer risk assessment is here designed by employing two controlled and two non–controlled risk factors. This paper presents a fuzzy system for assessment and classification that maps age, menopause, hormone replacement, and a simplified fuzzy body mass index (FBMI) input variables into three classes by covering moderate, high, very high classes on the breast–cancer risk output variable, as depicted in Fig. 1. This study is restricted to females ranging from 55 to 70 years–old.

II. Fuzzy Risk Analysis for Breast Cancer

Some factors that increase the risk of developing Breast Cancer are controllable, i.e., can be avoided with behavioral

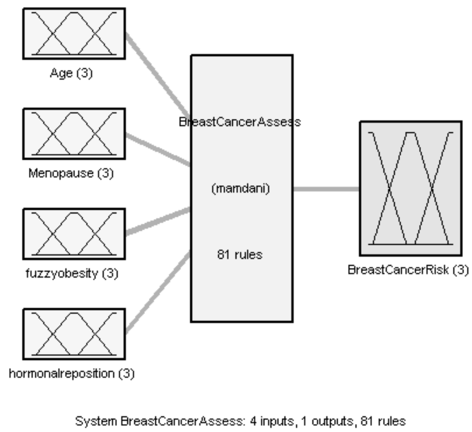

Fig. 1. Fuzzy Breast Cancer Assessment based on Controlled hormone Replacement and Fuzzy Obesity as well as Non–controlled Age and Menopause Input Linguistic Variables.

changes, either in eating habits, increased physical activity, or decreased/abandonment of smoking. These individual actions alter factors such as obesity, physical inactivity, among others. In this sense, women – or healthcare professionals – can evaluate which factors would cause a more significant influence, be it positive, be it negative, in their health conditions. The two first significant factors employed in this paper as input linguistic variables cover obesity comorbidity factor and hormone replacement. Factors such as body mass index and hormone replacement may be modified by decreasing the risk of developing breast cancer in women. In this paper, the obesity comorbidity factor is described by employing the fuzzy body mass index.

There are two other valuable factors, however, that do not depend on the human (woman) intervention. Age and menopause age affect the risk of breast cancer due to, respectively, the inherent worsening of the body as time passes and the lessening of the hormones that the body requires for its best functionalities. Due to that, these factors should be included in the design of a medical decision system to support the diagnosis of patients with breast cancer risk.

The risk input variables concerning the breast cancer analysis, assessment, and diagnosing can be mathematically written, thus, as the *fuzzy obesity*, $X_{fuzzy-obesity}$, *hormone–condition*, $X_{hormone-condition}$, *age*, X_{age}, and *age of menopause*, $X_{menopause}$, yielding a four–dimensional Cartesian input space, $X_{\textit{fuzzy–obesity}} \times X_{\textit{hormone–condition}} \times X_{age} \times X_{menopause}$. The output linguistic variable covers moderate, high, and very high classes on the breast cancer *risk*, $X_{breast-cancer-risk}$.

The input fuzzy sets given by $M_{\textit{fuzzy–obesity}}^{j_{\textit{fuzzy–obesity}}}$, $M_{\textit{hormone–condition}}^{j_{\textit{hormone–condition}}}$, $M_{age}^{j_{age}}$, and $M_{menopause}^{j_{menopause}}$ partition their respective universes of discourse X_i, for $i = 1, 2, 3, 4$, such that $j_{\textit{fuzzy–obesity}} = 1, 2, 3$, $j_{\textit{hormone–condition}} = 1, 2, 3$, $j_{age} = 1, 2, 3$, and $j_{menopause} = 1, 2, 3$ yielding a set of 81 fuzzy regions. The output linguistic terms are given by $M_{\textit{breast–cancer–risk}}^{j_{\textit{breast–cancer–risk}}}$, such that $j_{\textit{breast–cancer–risk}} = 1, 2, 3$. The membership functions, $\mu_{M_i} : X \to [0, 1]$, partition the universes of discourse, X_i, to which are associated a set of terms $T = \{M_i^1, \ldots, M_i^j, \ldots, M_i^n\}$; a linguistic term $M_i \in T$, where $c(M_i) = \{x_0 \in X_1 | \mu_{M_i}(x_0) = 1\}$ and $s(M_i) = \{x_0 \in X_1 | \mu_{M_i}(x_0) > 0\}$, respectively, denote the core and support of M_i. In this paper each linguistic term $M_i^j \in T$ is shaped according to a trapezoidal membership function. The trapezoidal membership function is given as $\mu_{M_i^j}(x_i; a, b, c, d) = max\,(min\,((x-a)\,(b-a)\,, 1, (d-x)\,(d-c)))$, where $a < b \leq c < d$, represented by the 4–tuple $\langle s1, c1, c2, s2 \rangle$, with $s(M_i^j) = [s1, s2]$.

A. Input Linguistic Variables

1) Hormone Replacement:

The hormone replacement input variable is partitioned by the linguistic terms $T_{\textit{hormone–condition}} = \{\textit{Low, Moderate, High}\}$ in three trapezoidal membership functions, distributed in $X_{hormone-condition} = [0, 10]$, as depicted in Fig. 2.

2) Age:

Three trapezoidal classes partitioning the age input variable correspond to the linguistic terms $T_{age} = \{\textit{Age–I, Age–II, Age–III}\}$ and their membership functions, distributed in $x_{age} = [55, 70]$, as illustrated in Fig. 3.

3) Menopause Age:

The set of linguistic terms, $T_{Menopause} = \{\textit{Common, Intermediate, Late}\}$, partitioning the menopause input variable corresponds to trapezoidal membership functions, distributed in $X_{menopause} = [45, 58]$ (Fig. 4).

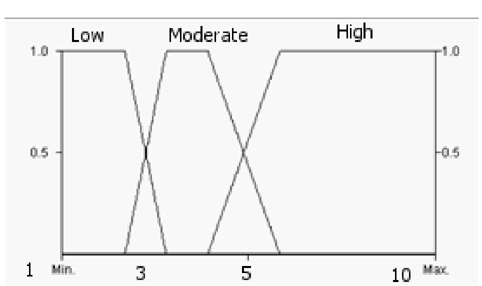

Fig. 2. Input membership functions partition the universe of discourse, $X_{\textit{Hormone–Condition}}$: Hormone Condition.

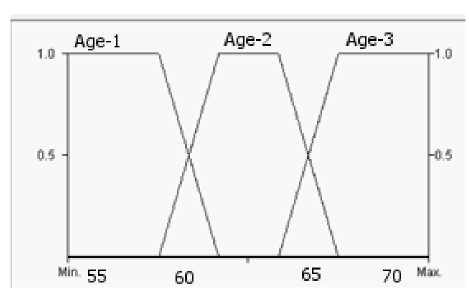

Fig. 3. Input membership functions partition the universe of discourse, X_{Age}: Age.

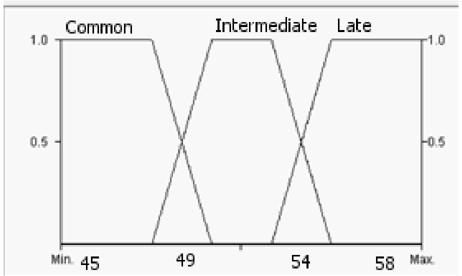

Fig. 4. Input membership functions partition the universe of discourse, $X_{Menopause}$: Menopause.

4) "Modified" Fuzzy Body Mass Index (FBMI):

Obesity is a risk factor of great importance in the diagnosis and prognosis for Breast Cancer. There are, however, diverse anthropometric indices for measuring such a comorbidity. Recommended by the World Health Organization (WHO), the Body Mass Index (BMI) is characterized by its capacity of weight excess, as given by IMC $= P/H^2$, where the body weight, P, of the individual is given in Kilograms $[Kg]$ and the square of the height, H, in $[m^2]$ [14]. The body mass index obesity classification is accomplished by using classic (Aristotelian) sets theory, such that the distinct classes of BMI in adults cover underweight (UW) when under than 18.4 kg/m^2, thin (T) when ranges from 18.5 to 24.9 kg/m^2, overweight (OW) from 25 to 29.9 kg/m^2, obesity–grade I (OI) from 30.0 to 34.9 kg/m^2, obesity–grade II (OII) from 35.0 to 39.9 kg/m^2, and (morbid) obesity–grade III (OIII) when greater than 40 kg/m^2. The BMI was first modified and treated as fuzzy sets in [15] when composing the Miyahira–Araujo Fuzzy Obesity Index (MAFOI). One of the derivate results of achieving the MAFOI is the Fuzzy Body Mass Index (FBMI) that adapts the crisp classes adopted by the World Health Organization to fuzzy sets [15].

The obesity comorbidity factor is measured in this paper by proposing a variation of the FBMI [15], [16] employed to compose the n–dimensional fuzzy input–output mapping. The simplified FBMI employed for achieving the fuzzy breast cancer assessment merges the subsets fuzzy obesity I, fuzzy obesity II, and fuzzy obesity III into a single fuzzy *obesity* class. Further, the first two fuzzy subsets of evaluation are aggregated in a fuzzy *normal* class. In so doing, this paper employs the set of linguistic terms, $T_{weight} = \{$*Normal (N), Overweight (OW), Obese (OB)*$\}$, partitioning the weight input variable, as illustrated in Fig. 5.

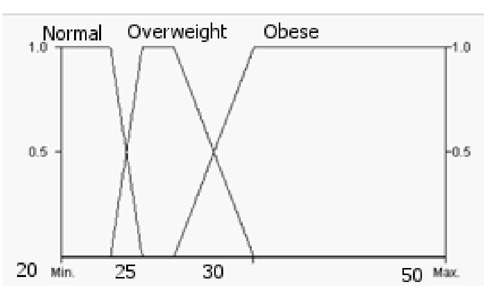

Fig. 5. Input membership functions partition the universe of discourse, $X_{Fuzzy\text{–}Obesity}$: Fuzzy Obesity.

B. Output Linguistic Variable

1) Breast Cancer Risk:

The output linguistic variable concerns the breast cancer risk. The respective membership functions and their linguistic terms are shown in Fig. 6. The output universe of discourse, Y, is partitioned into linguistic terms $T_{Breast\text{–}Cancer\text{–}Risk} = \{$*Moderate, High, Very High*$\}$ distributed in the universe of discourse with a range of $x_{Breast\text{–}Cancer\text{–}Risk} = [0, 5]$.

C. Fuzzy Breast Cancer Risk Rules

The set of 3–3–3–3 linguistic terms within the four-dimensional input premise space given by $x = [x_1, x_2, x_3, x_4]^T$ yields a set of 81 valid Mamdani fuzzy breast cancer system as given as:

$$
\begin{array}{ll}
R_1: & \text{IF } \langle \textit{Age} \text{ is } \textit{Age–I} \rangle \text{ AND} \ldots \\
 & \quad \langle \textit{Menopause} \text{ is } \textit{Common} \rangle \text{ AND } \ldots \\
 & \quad \langle \textit{Fuzzy–Obesity} \text{ is } \textit{Normal} \rangle \\
 & \quad \langle \textit{Hormone–Condition} \text{ is } \textit{Low} \rangle \\
 & \text{THEN } \langle \textit{Breast–Cancer–Risk} \text{ is } \textit{Moderate} \rangle \\
R_2: & \text{IF } \langle \textit{Age} \text{ is } \textit{Age–I} \rangle \text{ AND} \ldots \\
 & \quad \langle \textit{Menopause} \text{ is } \textit{Common} \rangle \text{ AND } \ldots \\
 & \quad \langle \textit{Fuzzy–Obesity} \text{ is } \textit{Overweight} \rangle \\
 & \quad \langle \textit{Hormone–Condition} \text{ is } \textit{Low} \rangle \\
 & \text{THEN } \langle \textit{Breast–Cancer–Risk} \text{ is } \textit{Moderate} \rangle \\
\ldots & \\
R_{80}: & \text{IF } \langle \textit{Age} \text{ is } \textit{Age–III} \rangle \text{ AND} \ldots \\
 & \quad \langle \textit{Menopause} \text{ is } \textit{Late} \rangle \text{ AND } \ldots \\
 & \quad \langle \textit{Fuzzy–Obesity} \text{ is } \textit{Overweight} \rangle \\
 & \quad \langle \textit{Hormone–Condition} \text{ is } \textit{High} \rangle \\
 & \text{THEN } \langle \textit{Breast–Cancer–Risk} \text{ is } \textit{High} \rangle \\
R_{81}: & \text{IF } \langle \textit{Age} \text{ is } \textit{Age–III} \rangle \text{ AND} \ldots \\
 & \quad \langle \textit{Menopause} \text{ is } \textit{Late} \rangle \text{ AND } \ldots \\
 & \quad \langle \textit{Fuzzy–Obesity} \text{ is } \textit{Obese} \rangle \\
 & \quad \langle \textit{Hormone–Condition} \text{ is } \textit{High} \rangle \\
 & \text{THEN } \langle \textit{Breast–Cancer–Risk} \text{ is } \textit{Very High} \rangle \ .
\end{array}
\tag{1}
$$

III. Discussion

The resulting risk surfaces for the fuzzy breast cancer assessment based on two controlled input variables intertwined

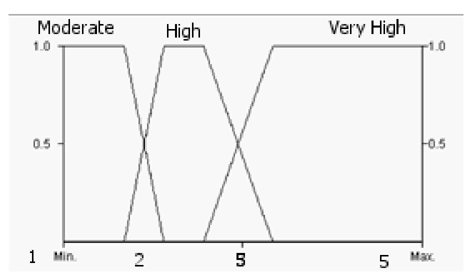

Fig. 6. Output membership functions partition the universe of discourse, $X_{Breast\text{–}Cancer\text{–}Risk}$: Breast Cancer Risk.

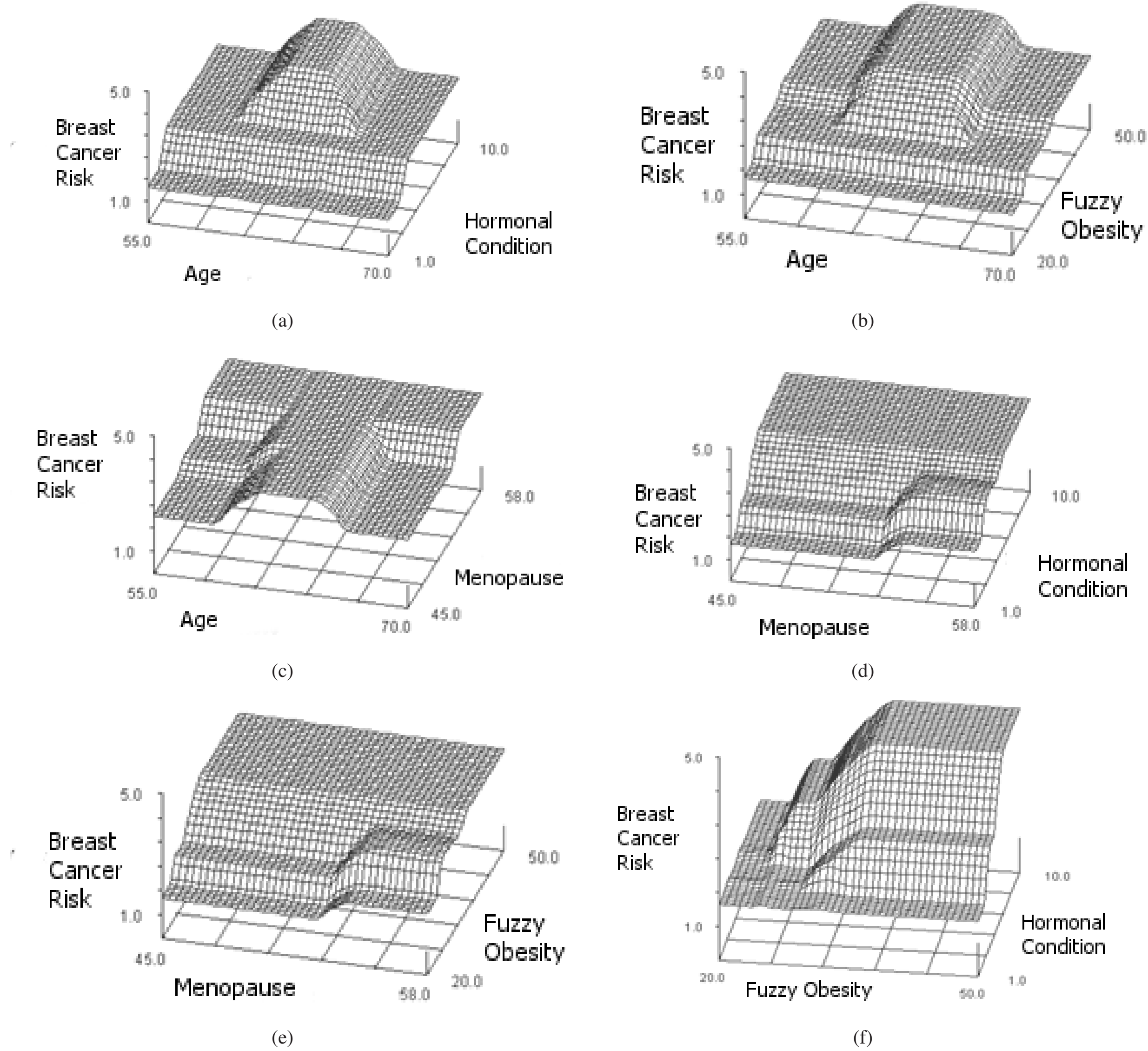

Fig. 7. Fuzzy assessment: age × hormone condition (7(a)), age × fuzzy obesity (7(b)), age × menopause (7(c)), menopause × hormone condition (7(d)), menopause × fuzzy obesity (7(e)), and fuzzy obesity × hormone condition (7(f)) for breast cancer risk analysis.

to two non–controlled input variables, in which one is a modified fuzzy body mass index, are shown in Fig. 7. The knowledge base of the fuzzy system is build up with data from the literature. Given the human visual limitations in dealing with objects in no more than three dimensions, the relationship between the input variables are shown in pairs concerning their respective severity levels.

It is the responsibility of health authorities the prevention and treatment of the population as soon as possible. The increasing access to information and technology advances in health care. The association of risk factors is cumulative, i.e., the larger the number factors present, the greater the chance of a woman developing breast cancer. To know the factors that increase the chances of developing breast cancer enables the realization of information with extensive outreach programs. The proposed fuzzy system comes to be a systematic and homogeneous diagnosis approach of patients with pre–conditions of breast cancer to reduce the mortality rate by achieving an early healthcare evaluation and a suitable interventional assistance and/or medication.

The breast cancer risk surfaces emphasize the importance of middle age (Age–II) upon the analysis. When the patient is categorized within this fuzzy class, it is possible to observe that whatever is the other input linguistic variable, i.e., hormone condition (Fig. 7(a)), level of obesity (Fig. 7(b)), or menopause (Fig. 7(c)), there is a higher prediction to achieve breast cancer. Nevertheless, when the menopause is late no matter is the age that the risk of breast cancer increases significantly (Fig. 7(c)). This characteristic is kept unaltered regardless the hormone condition (Fig. 7(d)) or the level of obesity (Fig. 7(e)), achieving the scientific reports available in the literature. On the counterclock direction, when the menopause is usual and occurs when the woman age range from 45 to about 49 the risk of breast cancer outbreak is reduced, be it compared to age (Fig. 7(c)), hormone condition (Fig. 7(d)), or level of obesity (Fig. 7(e)). Another risk factor that achieves high risk of breast cancer concerns when the hormone replacement is high regardless the level of obesity (Fig. 7(f)) or the moment in which the menopause occurs (Fig. 7(d)). Nevertheless, it is entirely dependent of the age of the woman (Fig. 7(a)). A quite similar behavior carries out when analyzing the level of obesity. Deeply dependent of the age (Fig. 7(b)), when such an input variable is compared to the menopause (Fig. 7(e)), it reaches high levels of risk when the patient is obese, regardless the age of menopause. Another important aspect that the woman and healthcare professionals should pay attention concerns the patient be obese and reaches high level of hormone replacement (Fig. 7(f)). In this sense, these two risk factors must be continuously monitored since they can be controlled to reduce the chances of developing breast cancer in women.

Finally, the use of fuzzy set theory and fuzzy logic allows not only reproducing the human thought, mainly in helping healthcare professionals in dealing with inherent imprecise and uncertain information in the diagnosis, in general, but comes to be a feasible alternative to represent the complexity of the breast cancer outbreak.

IV. Conclusion

The fuzzy medical diagnostic decision support system as herein propose becomes an alternative tool to support in risk analysis of breast cancer outbreak when taking into account the natural aging and its unfolding health conditions. The proposed fuzzy breast cancer assessment and classification maps the age, menopause age, hormone condition (replacement), and a modified fuzzy body mass index input variables into three classes by covering moderate, high, and very high classes on the risk output variable. The proposed model can provide the female population and the health authorities a health support mechanism in the prediction measurement of developing breast cancer, to reduce both the outcomes and the mortality rate.

References

[1] World Health Organization – Cancer, [Online; accessed 29–Feb–2014]. [Online]. Available: ⟨http://www.who.int/topics/cancer/en//⟩

[2] K. P. Adlassnig, "Fuzzy set theory in medicine," *Intern. Institute for Apllied Systems Analysis*, pp. 2–17, 1984.

[3] K. Sadegh-Zadeh, "Fuzzy health, illness, and disease," *Journal of Medicine and Philosophy*, vol. 25, pp. 605–638, 2000.

[4] M. F. Abbod, D. G. von Keyserling, D. A. Linken, and M. Mahfouf, "Survey of utilisation of fuzzy technology in medicine and healthcare," *Fuzzy Sets and Systems*, vol. 120, pp. 331–349, 2001.

[5] A. Torres and J. J. Nieto, "Fuzzy logic in medicine and bioinformatics," *Journal of Biomedicine and Biotechnology*, pp. 1–7, 2006.

[6] R. Seising, "On the absence of strict boundaries–vagueness, haziness, and fuzziness in philosophy, science, and medicine," *Applied Soft Computing Journal*, vol. 8, pp. 1232–1242, 2008.

[7] E. Araujo, "Fuzziness in medical measurement and approximate reasoning," in *Fuzziness and Medicine: Philosophical Reflections and Application Systems in Health Care*, ser. Studies in Fuzziness and Soft Computing, M. E. T. R. Seising, Ed. Berlin: Springer–Verlag, 2013, ch. 16, pp. 233–249.

[8] A. Keles, A. S. Hasiloglub, A. Kelesc, and Y. Aksoyd, "Neuro–fuzzy classification of prostate cancer using NEFCLASS–J," *Computers in Biology and Medicine*, vol. 37, pp. 1617–1628, 2007.

[9] M. F. Abbod, J. W. F. Catto, D. A. Linkens, and F. C. Hamdy, "Application of artificial intelligence to the management of urological cancer," *The Journal of Urology*, vol. 178, pp. 1150–1156, 2007.

[10] S. Petrovic, N. Mishra, and S. Sundar, "A novel case based reasoning approach to radiotherapy planning," *Expert Systems with Applications*, vol. 38, pp. 10 759–10 769, 2011.

[11] M. J. P. Castanho, F. Hernandes, A. M. D. Ré, S. Rautenberg, and A. Billis, "Fuzzy expert system for predicting pathological stage of prostate cancer," *Expert Systems with Applications*, vol. 40, pp. 466–470, 2013.

[12] B. S. B. Bassanezi, A. G. Oliveira-Filho, R. S. M. Jafelice, J. M. Bustorff-Silva, and A. Udelsmann, "Postoperative vomiting in pediatric oncologic patients: prediction by a fuzzy logic model," *Pediatric Anesthesia*, vol. 23, pp. 68–73, 2013.

[13] M. R. Hassan, M. M. Hossain, R. K. Begg, K. Ramamohanarao, and Y. Morsi, "Breast–cancer identification using HMM-fuzzy approach," *Computers in Biology and Medicine*, vol. 40, pp. 240–251, 2010.

[14] E. E. Calle, M. J. Thun, and J. M. Petrelli, "Body-mass index and mortality in a prospective cohort of U.S. adults," *N Engl J Med*, vol. 341, pp. 1097–1105, 1999.

[15] S. A. Miyahira and E. Araujo, "Fuzzy obesity index for surgical indication and treatment," in *Intern. Conference on Fuzzy Systems*. IEEE, 2008, pp. 2392–2397.

[16] E. Araujo and S. A. Miyahira, "Tridimensional fuzzy pain assessment," in *IEEE Intern. Conference on Fuzzy Systems*, 2011, pp. 1634–1639.