Journal of Biomedical Informatics 45 (2012) 1021–1034

Contents lists available at SciVerse ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Fuzzy-probabilistic multi agent system for breast cancer risk assessment and insurance premium assignment

Farzaneh Tatari [a,*], Mohammad-R. Akbarzadeh-T [a,1], Ahmad Sabahi [b,2]

[a] Department of Electrical Engineering, Center of Excellence on Soft Computing and Intelligent Information Processing, Ferdowsi University of Mashhad, Mashhad, Iran
[b] Department of Economic and Management, Ferdowsi University of Mashhad, Mashhad, Iran

ARTICLE INFO

Article history:
Received 6 January 2012
Accepted 24 May 2012
Available online 9 June 2012

Keywords:
Breast cancer
Fuzzy probability
Fuzzy-probabilistic multi agent system
Insurance premium
Uncertainty

ABSTRACT

In this paper, we present an agent-based system for distributed risk assessment of breast cancer development employing fuzzy and probabilistic computing. The proposed fuzzy multi agent system consists of multiple fuzzy agents that benefit from fuzzy set theory to demonstrate their soft information (linguistic information). Fuzzy risk assessment is quantified by two linguistic variables of high and low. Through fuzzy computations, the multi agent system computes the fuzzy probabilities of breast cancer development based on various risk factors. By such ranking of high risk and low risk fuzzy probabilities, the multi agent system (MAS) decides whether the risk of breast cancer development is high or low. This information is then fed into an insurance premium adjuster in order to provide preventive decision making as well as to make appropriate adjustment of insurance premium and risk. This final step of insurance analysis also provides a numeric measure to demonstrate the utility of the approach. Furthermore, actual data are gathered from two hospitals in Mashhad during 1 year. The results are then compared with a fuzzy distributed approach.

1. Introduction

Risk assessment is one of the most important aspects in medical decision making. One of the risks that should be assessed is the risk of being affected by deadly diseases such as cancer where the risk is influenced by various variables. An appropriate paradigm for handling the effect of these variables can help insurance companies to better evaluate the risk of deadly diseases and provide the users suitable financial facilities for screening and preventive treatment circumstances based on a fair premium. In this order confidential risk assessments can save so many lives and subsequently decrease the treatment costs and increase the social health.

Unfortunately, recognizing all of the effective variables and their precise amount of effectiveness in a risk assessment is nontrivial. In addition, the available information is highly imprecise and can lead to uncertain conclusions. With respect to the existing uncertainties in the available information, the first source of data for risk assessment is known as "soft" data [1] that can be represented in linguistic forms (linguistic data). This data is extracted from the expert opinions, their aggregated studies and experiences. Due to the linguistic form of this information and imprecision of data, fuzzy-logic based analysis offers a promising solution paradigm [2] to handle the existing uncertainty. In addition to soft data, the statistical data is another valid data source which can be helpful. This statistical data may be incomplete when the number of observed data is not considerably large. Hence the statistical data are also accompanied by uncertainty which is caused by sparsity and insufficiency of data. We have found fuzzy probability framework as a suitable approach which enables us to enhance the reliability of our risk assessment by employing both databases of soft data and statistical data at the same time. Furthermore fuzzy probabilities enable us to show the uncertainty of the assessed risk and compute the amount of uncertainties in these fuzzy probabilities.

Fuzzy probabilities were first introduced by Zadeh [3]. Fuzzy probability (FP) theory [4] is a fuzzy approach to probability theory and is a generalized form of probability theory. In fuzzy probabilities, probability theory is complemented with an extra dimension of uncertainty provided by fuzzy set theory [5]. We divide the applications of fuzzy probabilities to two different main areas [6]. The first is the area of reliability and risk assessment. In this area fuzzy probability has been widely applied in fuzzy fault trees to assess the fault risk [7–9], risk assessment of pedestrian collisions [6], reliability assessment for pressure piping [10], risk assessment of natural hazards [11] and reliability enhancement

* Corresponding author. Address: Azadi Square, Ferdowsi University of Mashhad, College of Engineering, P.O. Box 91775-1111, Mashhad, Iran. Fax: +98 5118832731.
E-mail addresses: fa_tatari@yahoo.com (F. Tatari), akbarzadeh@ieee.org (M-R. Akbarzadeh-T), sabahi2005@yahoo.com (A. Sabahi).
[1] Address: Azadi Square, Ferdowsi University of Mashhad, College of Engineering, P.O. Box 91775-1111, Mashhad, Iran.
[2] Address: Azadi Square, Ferdowsi University of Mashhad, College of Economic, Mashhad, Iran.

1532-0464/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jbi.2012.05.004

by combining expert opinions [12]. Second is the field of decision making, where fuzzy probabilities have been employed in perception-based theory [13], optimal decision fusion [14,15], inference by aggregation [16], information retrieval [17] and inventory control [18]. Since risk is evaluated based on the uncertain data, the assessed risk is also imprecise. Therefore we consider risk as a fuzzy variable and show the likelihood of any fuzzy risk by fuzzy probabilities.

To employ and influence all the available factors in the risk assessment, we employ multi-agent systems. Multi agent system (MAS) is a distributed computational system with intelligent autonomous agents. In a MAS, agents coordinate their tasks cooperatively or competitively to reach their goals in a distributed form while the agents may be the same or different [19].

Fuzzy risk assessment is an open and highly dynamic problem because new risk parameters may be added to the old ones in our problem, also the environment is uncertain and complex. Therefore we employ multi-agent systems (MASs) which properly conforms to the features of our distributed risk assessment environment. In addition MAS structure gives the system this opportunity to add new parameters to legacy components through an agent layer. Other characteristics of MAS which make it suitable for the risk assessment problem are decentralization, result sharing and self-organization. Decentralization means the distribution of data, control or expertise [20]. In some risk assessment problems the distribution of data, control or expertise in a centralized solution is at best extremely difficult or at worse impossible especially where the parameters are numerous and are not correlated [21]. Result sharing is one the most important capabilities of a multi agent system where each agent shares his local result obtained based on his local data with other agents. The aggregation of all these local results creates a global result with a high level of confidence. Result sharing increases the confidence level in the total result and ensures the precision of the final solution [22]. Moreover, these characteristics enable the proposed flexible system to automatically continue its computations ignoring the unavailable factors. In this paper self-organization enables the designed MAS to arrange the solution procedure according to the available data without being guided or managed by an outside source.

Multi agent systems are applicable in various areas of medical sciences such as health monitoring [23–27], medical diagnosis [28–31], electronic medicine [32–34], risk assessment [35,36], medical services [37] and medical information systems [38]. In the mentioned fields, risk assessment is mostly accompanied by uncertainty, where it is usually impossible to determine all the effective factors and the percentage of their effectiveness on the final assessment.

More recently MASs have begun to emerge as an integrated solution approach to distributed computing. Research shows that domains in which data, control, expertise or resources are inherently distributed can be addressed using agent technology. A MAS consisting of multiple agents can take advantage of computational resources and capabilities that are distributed among interconnected entities. In this paper, for distributed computing of risk fuzzy probabilities, a MAS is introduced which handles fuzzy and probabilistic computing concurrently. Therefore based on the multi agent system capabilities, it is called ‘‘fuzzy-probabilistic multi agent system’’, where ‘‘fuzzy probabilistic’’ is just a reasoning feature for the employed MAS.

Actually the main reason for risk assessment is to make preventive decisions. Therefore the assessed risk is employed to calculate the insurance premium, where insurance company recommends preventive facilities for any user with a certain risk.

In this paper we introduce a general paradigm for distributed risk assessment and insurance premium assignment under uncertain conditions by integrating the capabilities of fuzzy probability and multi-agent systems synergistically which is employed for breast cancer (BC) risk assessment. We introduce an approach for fuzzy probabilities computation which is particularly helpful to model and combine imprecise probabilities derived from linguistic data and statistical data. Furthermore system can estimate the BC insurance premium based on the calculated risk which can be offered by insurance companies for preventive and screening facilities.

This paper is organized as follows. In Section 2, breast cancer risk assessment factors and models are introduced. Section 3 illustrates the performance of the proposed multi agent system. The employed soft data and statistical data are explained in Section 3.1. In Section 3.2, we describe the fuzzy probabilities calculation. We explain the procedure of breast cancer risk assessment by the proposed fuzzy-probabilistic multi agent system and ranking fuzzy probabilities in Sections 3.3 and 3.4 respectively. In Section 4, we discuss how the obtained results can be employed for defining breast cancer premiums. We demonstrate the total performance of the proposed fuzzy-probabilistic MAS in breast cancer risk assessment and insurance premium assignment through an example in Section 4.1. In Section 5, we evaluate the final results of another distributed computing method called ‘‘fuzzy MAS’’ for BC risk assessment and insurance assignment and compare the results of the proposed fuzzy-probabilistic MAS with fuzzy MAS method. Finally, conclusions are drawn in Section 6.

2. Breast cancer risk assessment factors and models

One in nine women will develop breast cancer (BC) at some point in her life [39]. Hence it is likely for any woman to be threatened by breast cancer. Based on the extensive researches on BC risk factors, we have gathered a collection of factors which have been investigated to date. Due to this fact that not all factors increase a woman’s chance of BC development equally, BC risk factors are categorized to three groups of strong, moderate and minor risk factors [40]. Here we have extended the categorization in [40] to include other several factors based on the survey of multiple articles in this field. We have shown these three groups in Table 1.

In this section we aggregate the relative researches about breast cancer risk factors thus far.

Age is the primary risk factor for BC. Overall 85% of cases is in women with 50 years of age and older, while only 5% of BC develops in women younger than 40 [40–43]. Therefore, women who are older than 50 are at high risk of breast cancer development and women who are younger than 40 are at low risk of breast cancer development.

In addition to the age another strong risk factor is family history. Women who have a family history of BC are at a higher risk for BC than those who lack such a history. Women who have first- and second- degree relatives with BC have a greater chance of developing BC [27,34,44–46].

The carriers of BRCA1and BRCA2 genes are at high risk of being affected by BC at some time in their lives [37,41]. Women who have had a prior breast biopsy that revealed a proliferative abnormality (extensive growth of the glandular breast tissue, also called hyperplasia) have an increased risk of BC [40,41].

Being younger than 12 when first menstrual period occurs increases the risk of BC development [40,41] and older age at menopause (>55) increases the risk of BC, where both of these factors are considered as moderate factors. Other moderate risk factors are age at first child’s birthday, number of pregnancies, mammogram density and exposure to radiation.

There is evidence that the more the children a woman has, the greater the protection from BC, and women with 5 or more children

Table 1
Breast cancer risk factors.

Risk factor	High risk group	Low risk group
[c]First-degree relatives with BC (mother, sisters, daughters)	⩾1	0
[c]Second-degree relatives with BC (grandmothers, aunts, nieces, cousins)	⩾1	0
[c]SNP information	Refer to Table 2	
[c]Age	>50 years	<45 years
[c]Inheriting BRCA1,2	Yes	No
[b]Age at menarche	<12 years	>14 years
[b]Age at menopause	>55 years	<45 years
[b]Age at first child birthday	>30 years	<30 years
[b]Number of pregnancies	<3	>4
[b]Mammogram density	>50%	<5%
[b]Biopsy abnormalities	Yes	No
[b]Exposure to radiation	>400mrad	<200mrad
[a]Oral contraceptive consumption period	>4 years	<2 years
[a]Alcohol consumption	>2 drinks a day	<2 drinks a day
[a]Hormone replacement therapy period	>4 years	<2 years
[a]First-degree relatives with other cancers	⩾1	0
[a]Second-degree relatives with other cancers	⩾1	0
[a]Obesity	>25	<25
[a]Vegetable and fruit consumption(serves a day)	<1	>2
[a]Physical exercises	>5 min	<15 min
[a]Race	East European, European	Asian, African

[a] Minor factors.
[b] Moderate factors.
[c] Strong factors.

are at low risk of breast cancer development in comparison with nuliparous women [41,44]. Women who have their first full term pregnancy at the age of 30 years or older have an increased risk of BC as compared women who give birth before age 30. Women who have never given birth are more likely to develop BC [40,41,43]. Mammographic density is identified on mammograms as the non-radiolucent portions of the image. These represent the fibrous and glandular tissues in the breast; where as the dark radiolucent areas are primarily fat. Women with a high percentage density in their breasts (>50%) are at higher risk of breast cancer as compared to women with low percentages of density [41,47].

High dose ionizing radiation to the chest such as repeated fluoroscopies used during lung-collapse therapy for tuberculosis or subjecting to atomic bomb explosion [48], have been shown to increase the subsequent risk of BC [49]. Radiation exposure from modern mammographic equipment is in range of 200–400 mrad, which has been calculated to have minimal impact on BC risk [50].

Minor risk factors are the factors which mostly depend on hormone replacement therapy and oral contraceptive (OC) consumption, life style, environmental factors, diet schedule and the health background of any person. Women who are obese (body mass index (BMI) > 25), are at higher risk of BC [43]. More than two serving a day of vegetable and enough physical exercises during the day decrease the risk of BC [43,51].

It has been shown that alcohol consumption ($\geqslant$2 drinks a day) increases the risk of BC [40,43]. As a woman ages, the breast grandular tissue, the tissue in which breast cancer arises, is gradually replaced by fat. Hormone therapy which includes estrogen and progestin that slows or reverses this process increases the risk of BC [40,41]. Use of oral contraceptive agents appears to increase BC risk as well women who have used these agents for a prolonged period of time (>4 years) are at higher risk of BC development [47]. Race and ethnicity are among the minor risk factors. It has been discovered that the women of Asia and Africa are at the lower risk of BC development in comparison with the American, European and eastern European women [41].

An improved understanding of genetic risk factors and their interactions with the environment would allow accurate predictions of disease and facilitate prevention through measures directed toward persons at high risk [52,53]. Although the clinical use of single, common low penetrance genes is limited, a small number of susceptibility alleles could distinguish women at high risk for breast cancer from women at low risk, particularly in the context of population screening programs. Moreover, stratifying women according to genetic risk may improve the efficiency of screening programs. Also genetic variation, rather than lifestyle or environmental factors, accounts for most of the familial clustering [54,55].

Studies in breast cancer have reported new breast cancer susceptibility loci at highly stringent levels of statistical significance (Table 2) [56–59]. These loci are all single-nucleotide polymorphisms (SNPs) with two alleles: a high-risk allele and a low-risk allele. Since there are two copies of each locus in the genome, there are three possible combinations of alleles: two low-risk alleles, one low risk and one high-risk allele, or two high-risk alleles. The risk conferred by each of these loci appears to be allele-dose-dependent with a multiplicative effect on the risk.

The specifications of these SNPs, available in Table 2, are extracted from [59,60]. Apparently the cost of a genetic test for purposes of risk profiling would be minimal as compared with the costs of a lifetime screening program.

Currently there are no risk prediction models that efficiently incorporate all known risk factors [49]. Till now there have been three risk prediction models named as Gail model [61], Claus model

Table 2
Common breast-cancer susceptibility alleles.

SNP No.	Gene[a]	Chromosome	Relative risk per allele
rs2981582	FGFR2	10q	1.26
rs3803662	TNRC9, LOC643714	16q	1.20
rs889312	MAP3K1	5q	1.13
rs3817198	LSP1	11p	1.07
rs13281615	None known	8q	1.08
rs13387042	None known	2q	1.20
rs1053485	CASP8	2q	1.13
C3435T	MDR1	21q	1.30

[a] SNP No. single-nucleotide polymorphisms number, *CASP8* denotes caspase 8, *FGFR2* the fibroblast growth factor receptor 2 gene, *LOC643714* a hypothetical protein LOC643714, *LSP1* lymphocyte-specific protein 1, *MAP3K1* mitogen-activated protein kinase 1, *TNRC9* trinucleotide repeat containing 9, and MDR1 multidrug-resistant gene.

[62] and IBIS model [63] where Gail is the earliest and IBIS is the latest. Based on the descriptions mentioned in [45], we have shown the main characteristics of these three models in Table 3.

The problem of overestimating and underestimating in these models is caused by this fact that we cannot weight the risk factors according to their effectiveness on breast cancer risk. Also if a woman wants to assess her risk based on these models, she must have the answer of all the requested factors by the program. In other words such kinds of models are not robust towards the situations where the user does not have all the information about her factors. Also if we want to update the statistical database by adding the information of new cases or new risk factors to the database, we must build a new model based on the updated database. While it is impossible to add a new part to the previous program as a complement to the old part.

Since there are still further limitations about the existing models, it is needed to develop models with fewer limitations. It is anticipated that newer risk models and methods of assessment that incorporate breast density and/or evaluation of SNPs will become available in the near future [45].

3. The proposed fuzzy-probabilistic multi-agent system

Flexible topology and collaborative structure of the MASs makes them applicable in a wide area of applications. In this paper we introduce a cooperative MAS where agents can share knowledge and distribute subtasks. By this strategy they benefit from decentralization and self-organization to solve distributed problems more effectively. In this system, agents are equipped with suitable knowledge and computational capabilities.

The proposed fuzzy-probabilistic multi-agent system (MAS) for BC risk assessment consists of several risk factor agents, each risk factor agent represents one of the BC risk factors mentioned in Table 1. There is also one interface agent which plays the role of an interface between user agent and risk factor agents. In this system each risk factor agent is provided with soft and statistical data. Therefore any risk factor agent that corresponds to the *i*th risk factor is equipped with soft data and statistical data about his own risk factor. In the following sections we will discuss that soft data of any risk factor agent is defined in the form of fuzzy sets on the universe of the corresponding risk factor. In any coop-

Table 3
Gail, Claus and IBIS models characteristics.

	Gail model [61]	Claus model [62]	IBIS model [63]
Included BC risk factors	– Age – First degree relatives with BC (mother, sisters, daughters) – Age at menarche – Age at first child birthday – Number of previous breast biopsies – Presence of atypical ductal hyperplasia – Race	– Age – First degree relatives with BC (mother, sisters, daughters) – Maternal and paternal second degree relatives with BC (grandmothers, aunts) – Ages at diagnosis of BC Extended model adjusts Claus risk for presence in the family of – Bilateral BC – Ovarian cancer – More than two relatives with BC	– Age at menarche – Age at first child birthday – Age at menopause – Number of pregnancies – Use of hormone replacement therapy (type, years of use, years since last used) – Breast disease (hyperplasia, atypical hyperplasia, Lobular carcinoma in situ (LCIS)) – Height (premenopausal) – Body mass index (BMI) (postmenopausal) *Genetic factors:* – Breast cancer and age at onset in first- , second- , and third-degree maternal and paternal relatives – Ovarian cancer and age at onset in first- and second-degree relatives – Jewish ancestry
Benefits	– Easy to access and use – Incorporates reproductive factors – Adjusts for race – Well validated on a population basis	– Accounts for moderate and strong genetic risk factors – Validated using data from a large case-control study	– Incorporates both genetic and nongenetic risk factors – Easy data entry but requires extensive family history details – Adjusts for number/age of *unaffected* first- and second- degree relatives – Provides age-adjusted risks compared to population in graph form
Limitations	– Does not incorporate second-degree relatives with breast cancer, age at diagnosis, or presence of ovarian cancer – May overestimate risk in women with nonproliferative breast lesions – Has lower accuracy for individual risk prediction – Uses older population prevalence data associated with lower baseline incidence rates	– Does not incorporate nonfamilial risk factors – Certain combinations of affected family members not accounted for – Extended model requires manual application of regression formula using baseline Claus risk	– Limited validation data – Model assumes nongenetic factors are multiplicative on risk and of same magnitude across all genotypes – Underestimates risk in women with strong family histories – Overestimates risk in women with less strong family histories – Dramatic increase in risk for women with atypia and a positive family history
Risk calculation model	Logistic regression	Logistic regression	Bayes theorem

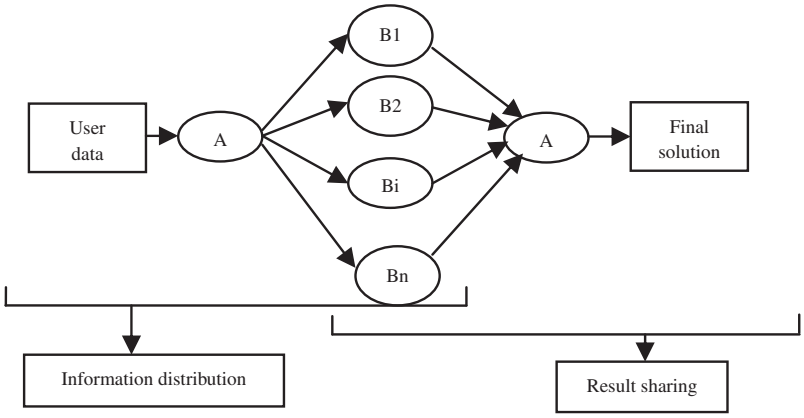

Fig. 1. Agent A is an interface agent and agent Bi ($i = 1,\ldots,n$) is a risk factor agent.

erative MAS, problem solving contains three main stages: (1) problem decomposition, (2) sub problem solution and (3) solution synthesis [20]. In this order interface agent, receives the amount of each BC risk factor from the user agent and distributes the whole information to different packages and sends any of the risk factors data to its specific risk factor agent. Then any risk factor agent based on the received amount of his risk factor and also his soft data and statistical data, computes the fuzzy probabilities of fuzzy risk (high risk and low risk) of BC development based on his own risk factor. After that each risk factor agent computes the fuzzy probabilities of risk levels, in the next step interface agent collects all the calculated fuzzy probabilities of risk factor agents and aggregates them to obtain more confidential results.

By applying fuzzy probabilities, agents cope with the existing uncertainty, at the same time applying parallel structure of MASs can save time and decrease computational complexities. Additionally regardless of the amount of information, we can consider all the available information and increase the confidence level of the aggregated decision. A general scheme of the proposed cooperative fuzzy MAS in the stages of information distribution and result sharing is outlined in Fig. 1.

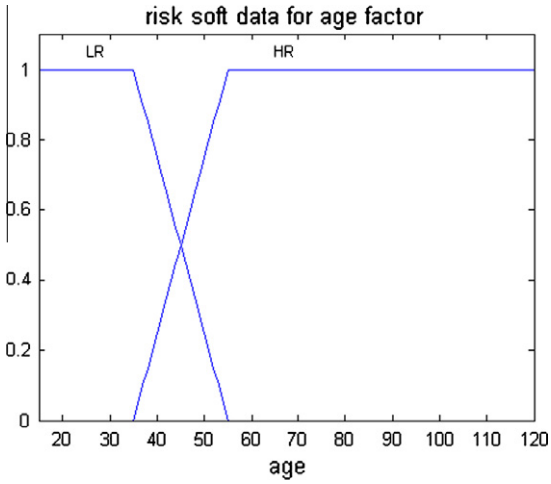

Fig. 2. Soft data for age risk factor agent.

3.1. Soft data and statistical data

As we mentioned in the previous section any risk factor agent is provided with its corresponding BC soft data and statistical data. In the BC risk assessment, the available soft data are the linguistic facts and linguistic data about risk factors (Table 1) that are collected by various experts and are available in medical literature. For example, as we mentioned in Section 2 women who are younger than 40 are at low risk of BC development and women who are older than 50 are at high risk of BC development. According to fuzzy-logic based analysis we can demonstrate this linguistic information in the form of two fuzzy sets on the universe of woman age as in Fig. 2, which forms our soft data about age risk factor.

Similarly all the linguistic information about other risk factors mentioned in Section 2 (except some factors which are naturally crisp) can be shown by two fuzzy sets of high risk and low risk.

The statistical database we employed is collected from patients with BC in Omid and Imam Reza hospitals of Mashhad during 2006–2007. Therefore the available statistical database is only from women with diagnosed BC (high risk). In this database, the information of 12 BC risk factors are available for any of these 87 women. Table 1 contains these 12 risk factors which are age, first degree relatives with BC, second degree relatives with BC, age at menarche, age at menopause, age at first child birthday, number of pregnancies, OCP consumption duration, hormone replacement therapy duration, other cancers in first- and second-degree relatives and C3435T SNP information of MDR1. The database contains the risk information of 87 BC patients where some of these patients had missing data about their risk factors. We have randomly divided the available database to two sets of training (44 patients) and testing (43 patients).

We employ both of these data, soft data and statistical data, to compute the fuzzy probabilities of high risk and low risk of BC development.

3.2. Fuzzy probabilities computations

There are several methods to compute fuzzy probabilities. In this part we explain that how any risk factor agent can employ its own soft data and statistical data to compute the fuzzy probabilities of high and low risk levels. Here we modified the procedure of fuzzy probability calculation explained in [6,11], where fuzzy probability is called possibility probability distribution. To compute fuzzy probabilities of high and low risk based on the method explained in [11], we must compute prior fuzzy probabilities where we employ soft data to calculate them based on the procedure explained in [17]. After the calculation of prior fuzzy probabilities, the statistical data are employed to convert prior fuzzy probabilities to posterior probability distributions. Finally these posterior probability distributions can be converted to posterior

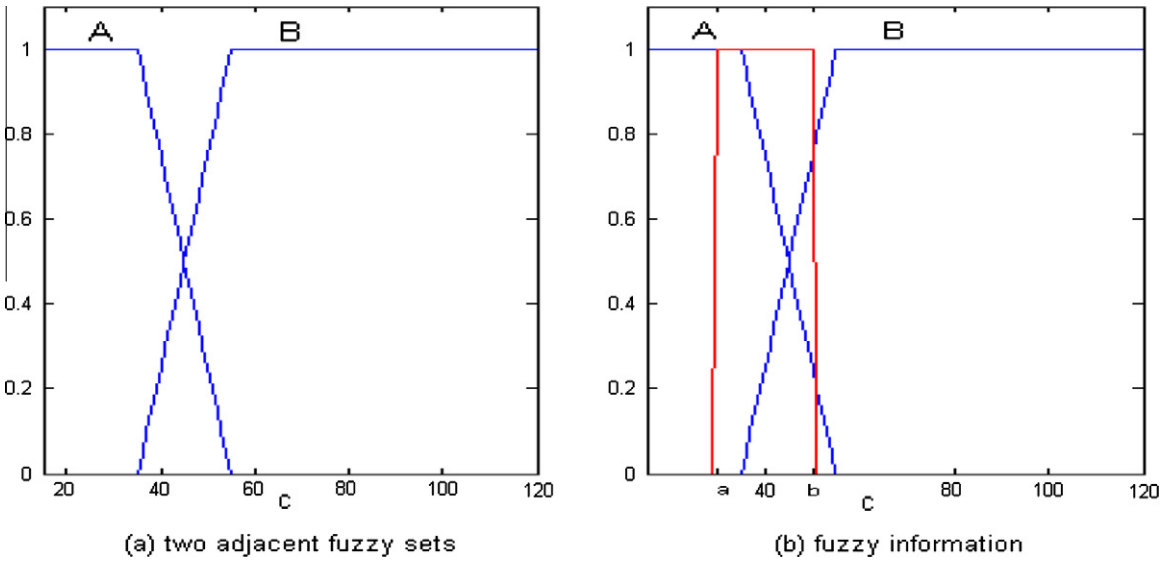

Fig. 3. (a) Two adjacent fuzzy sets and (b) fuzzy information $FI_A = A \cap FII, FI_B = B \cap FII$, (FII = $(b-a+1)$ for $c = 40$).

fuzzy probabilities. The detailed description of the employed method is as follows.

In any fuzzy computing, the nature of available data or the process of decision making to solve the problem is fuzzy. For computing the fuzzy probabilities of the occurrence of two adjacent fuzzy sets (see Fig. 3a) based on the variable measure on the universe of these two fuzzy sets, we could evaluate the membership degrees of the variable in the fuzzy sets. But employing the membership degrees is not desirable for fuzzy probabilities computations. Because the membership degrees are crisp and hence are less informative in comparison with considering a part of the fuzzy sets where the variable amount is in its center and gives more information about the situation of the variable in the fuzzy sets. To solve this problem, when the agent receives the amount of the variable and its membership degrees, it translates its information into a specific part of the two fuzzy sets. This part of fuzzy sets and its width is chosen in a way that the measured amount of the variable is located at its center. The width of this cut is called fuzzy information interval (FII $\geqslant 3$), where FII = 5 in our calculations.

This part that contains two fuzzy subsets of A and B (depicted in Fig. 3b) is called fuzzy information (FI). In this stage for computing the fuzzy probabilities of A and B based on the fuzzy information, $FI_A = A \cap FII$, $FI_B = B \cap FII$, we employ a procedure introduced in [17] as follows.

We compute S_B, S_A and S according to the obtained fuzzy information which is shown in Fig. 3b.

$$S_B = \sum_{i=1}^{n} \mu_B(x_i), \qquad S_A = \sum_{i=1}^{n} \mu_A(x_i), \qquad S = S_A + S_B \tag{1}$$

$\mu_A(x_i)$ and $\mu_B(x_i)$ are the membership degrees of x_i in FI_A and FI_B respectively, where $x_i, (a < x_i < b)$ are the integer points between a and b in Fig. 3b.

It is considered that $\pi_B(S_B/S) = 1$, $\pi_A(S_A/S) = 1$,where $\pi_A(p)$ and $\pi_B(p)$ show the possibility of each probability $p \in [0, 1]$ of A and B occurrence fuzzy probabilities respectively. Then agent considers the possibility that one of n data x_n may leave FI_B. As stated in [17] the point $x_i, (a < x_i < b)$ with a less degree of membership in FI_B, is more likely to leave. The possibility of probability for FI_B occurrence being $(S_B - \mu_B(x_i))/S$ can be calculated as

$$\pi'_B((S_B - \mu_B(x_i))/S) = \mu_A(x_i)/\mu_B(x_i) \tag{2}$$

Similarly two data points, x_i and $x_{i-1}, (\mu(x_i) < \mu(x_{i-1}))$ which have the less degree of membership values in FI_B are more likely to leave FI_B [17]. The corresponding possibility of this event is thus calculated as

$$\pi'_B((S_B - \mu_B(x_i) - \mu_B(x_{i-1}))/S) = \mu_A(x_{i-1})/\mu_B(x_{i-1}) \tag{3}$$

Similar processes are carried out until all the S_B membership values of data points $x_i, (a < x_i < b)$, leave FI_B.

Conversely it is possible for some data points to join FI_B. Suppose x_i has the minimal positive value of FI_A and it is thereby mostly possible to leave FI_A and join FI_B. The possibility of probability of FI_B occurring being $(S_B + \mu_A(x_i))/S$ can be calculated as

$$\pi'_B((S_B + \mu_A(x_i))/S) = \mu_B(x_i)/\mu_A(x_i) \tag{4}$$

Similarly if the probability of FI_B occurrence is $(S_B + \mu_A(x_i) + -\mu_A(x_{i+1}))/S$ then the possibility of the corresponding probability can be defined as follows.

$$\pi'_B((S_B + \mu_A(x_i) + \mu_A(x_{i+1}))/S) = \mu_B(x_{i+1})/\mu_A(x_{i+1}) \tag{5}$$

Similar processes are carried out until all the S_A membership values associated with FI_A join FI_B. In the last step the agent employs function $\pi = (2/(1 + \exp(-\pi'))) - 1$ to normalize $\pi' \in [0, +\infty]$ to $\pi \in [0, 1]$.

In this way any risk factor agent computes the prior fuzzy probabilities of high risk (HR) and low risk (LR) of BC development in the form of two adjacent fuzzy sets based on the received amount of their factor by employing soft data and fuzzy computing.

Low risk prior fuzzy probability and low risk posterior fuzzy probability are the same since there is no available and real statistical data about low risk women. But in the case of high risk fuzzy probabilities, we must combine the prior knowledge in the form of soft data and available statistical data of high risk women. Therefore we employ the available statistical data mentioned in Section 3.1 to obtain posterior fuzzy probability of HR from prior fuzzy probabilities.

We first divide the unit interval of probabilities in prior fuzzy probabilities into m points, $\theta_j, j = 1, \ldots, m$, and then calculate the posterior probability for each point θ_j as follows [11].

$$\begin{aligned} p'_i(\theta_j) &= \binom{n}{n_i} \theta_j^{n_1} (1-\theta_j)^{n-n_1} \dot{p}_i(\theta_j) \approx (1/c)(\gamma_j q_j + (1-\gamma_j) q_{j+1}) \\ q_j &= \theta_j^{n_i} (1-\theta_j)^{n-n_1} p_i(\theta_j) \end{aligned} \tag{6}$$

where n_i is the number of observations that satisfies the amounts of the ith risk factor among n observations (patients) in the training set and $P_i(\theta_j)$ is the possibility of θ_j probability in the prior fuzzy probability and $P'_i(\theta_j)$ is the posterior probability of θ_j probability. $\gamma_j \in [0, 1)$ is a unique value such that

$$\theta_j = \gamma_j\theta_j + (1 - \gamma_j) \cdot q_{j+1} \tag{7}$$

and c is a normalizing constant that guarantees

$$\int_0^1 p'_i(\theta)d\theta = 1 \tag{8}$$

Now is the time of transforming probability distribution to possibility distribution. Suppose a probability measure P on a set X is obtained via some statistical experiment. This probability function is a very rich piece of information, if the number of statistical experiments supporting it is high enough. According to widely accepted consistence principle, the possibility measures should dominate P in the sense that $P(A) \leqslant \pi(A)$ for all events $A \subseteq X$ and the maximally specific possibility distribution that exists must be unique [11]. As illustrated in [11] the most specific possibility distribution that approximates P from above (in the sense that $P \leqslant \pi$) can be derived quite easily [6,11]. For obtaining a possibility degree for any θ, $\pi_i(\theta)$, first $\lambda = p'_i(\theta)$ is computed. Next, the second boundary point θ' that satisfies $p'_i(\theta') = \lambda$ is found. Finally, the possibility degree $\pi_i(\theta)$ can be obtained as:

$$\pi_i(\theta) = \pi_i(\theta') = 1 - \left|\int_\theta^{\theta'} p'_i(x)dx\right|. \tag{9}$$

This possibility is equal to the area shaded grey in Fig. 4a.

The possibility probability distribution which is finally obtained is the HR posterior fuzzy probability which must be sent to the interface agent as well as the LR posterior fuzzy probability by any risk factor agent.

3.3. Breast cancer risk assessment by the proposed fuzzy-probabilistic MAS

Based on the existing soft data we quantize BC risk factors universes by two fuzzy sets of low risk (LR) and high risk (HR) except few factors that cannot be quantized by fuzzy sets due to their crisp nature, such as first- and second-degree relatives with BC, first- and second-degree relatives with other cancers, having biopsy abnormalities, SNP information and race.

Through the fuzzy computing that we explained in Section 3.2 any single factor agent computes the fuzzy probabilities of HR and LR of BC development based on the fuzzy information and available training data of his own risk factor. After that they send the fuzzy probabilities of HR and LR of BC development to the interface agent. The interface agent must aggregate these fuzzy probabilities to gain the aggregated fuzzy probabilities of HR an LR of breast cancer which are in higher level of reliability. Interface agent employs a procedure for combining fuzzy probabilities which was first proposed by Fau [6,12]. Based on the proposed method by Fau, for aggregating fuzzy probabilities, we must multiply the possibilities of all fuzzy probabilities at each probability value $p \in [0, 1]$ and normalize the products [12]. The normalized products would be the possibilities of the corresponding probabilities for the aggregated fuzzy probability.

Actually, fuzzy probabilities of strong risk factors must influence the global result more than fuzzy probabilities of moderate risk factors and also fuzzy probabilities of moderate risk factors must influence the final fuzzy probabilities more than fuzzy probabilities of minor risk factors. Therefore before receiving and combining fuzzy probabilities by interface agent, any risk factor agent weights his calculated fuzzy probabilities based on his own risk factor category (strong, moderate, minor). To improve the reliability of aggregation, 1, 0.8 and 0.6 are considered as the weights of strong fuzzy probabilities (FPs), moderate FPs and minor FPs respectively (refer to Fig. 5).

To demonstrate the effect of this reliability weighting [12], we imagine that in Fig. 6a the weight of fuzzy probability for agent 1 (a1) is 0.8 while the second agent (a2) fuzzy probability weight is one. If we employ the weights, the fuzzy probabilities will change to Fig. 6b and the fuzzy probability of their combination can be shown in Fig. 6c. In the aggregated fuzzy probability the modal value (MV) is obtained at $p = 0.485$, which shows that the resulting fuzzy probability reflects the second agent's fuzzy probability more than the first case [12].

In this order, the interface agent calculates the final fuzzy probabilities (FPs) of HR and LR of BC development by combining the weighted fuzzy probabilities of HR and weighted fuzzy probabilities of LR respectively. In this stage there are still some remaining factors which could not be defined by fuzzy sets, however according to the amount of their influence on BC risk, we can adjust the aggregated fuzzy probabilities to obtain the final fuzzy probabilities. To perform this adjustment, the interface agent asks the user about "having 1st or 2nd degree relatives with BC, having 1st or 2nd degree relative with other cancers, C3435T SNP information."

It should be noted that these factors are not modeled by HR and LR fuzzy sets since they are not inherently gradual (fuzzy) and the approach requires both HR and LR sets to be of same type, either fuzzy or crisp. For example, C3435T SNP can take on only 3 distinct values, where any of these values has a different effect on risk. For the other two parameters on relatives, since a LR case is inherently non-fuzzy, i.e. set for not having any relatives with cancer, we did not define HR fuzzy set either.

Therefore according to user's replies, the final fuzzy probabilities are tuned. For instance if the user's answers confirm that she has one 1st degree relative with breast cancer, her HR risk of BC development increases remarkably; otherwise both the HR and LR fuzzy probabilities remain unchanged. In the cases when the replies imply increase in the risk of cancer, we shift the range of HR probabilities by constant $b > 0$ (here $b = 0.3$) while the corresponding possibilities and LR fuzzy probabilities remain unchanged [46].

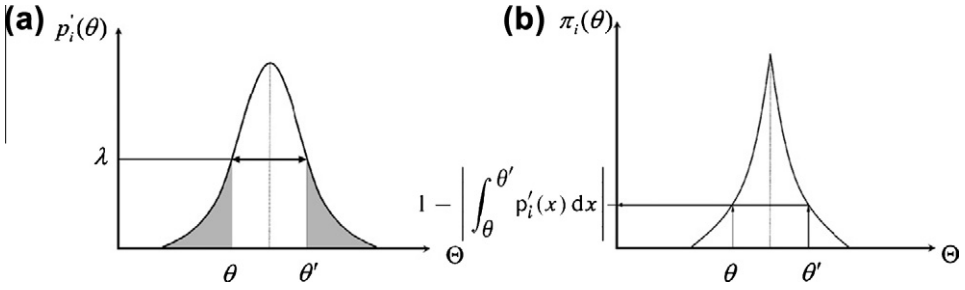

Fig. 4. (a) Probability distribution to be transformed into a possibility distribution and (b) transformed possibility.

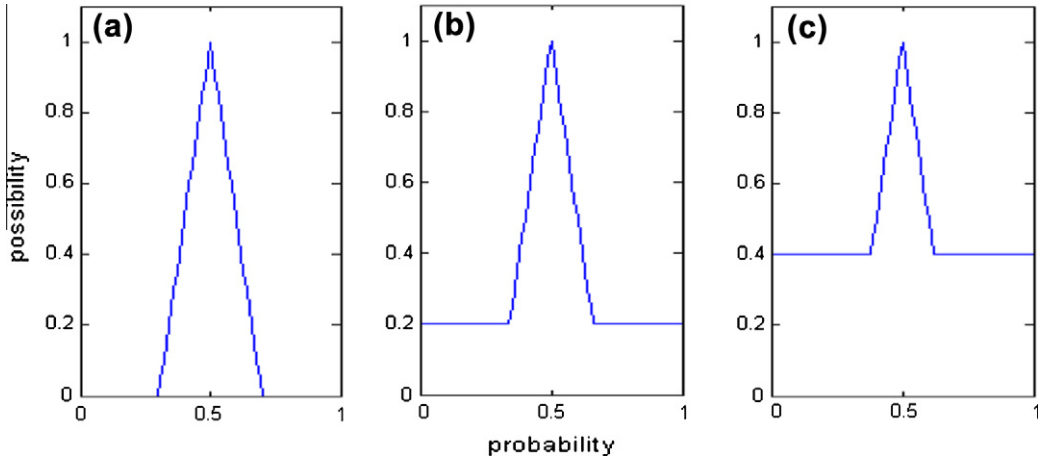

Fig. 5. (a) Fully reliable fuzzy probability, (b) 80% reliable FP, (c). 60% reliable FP.

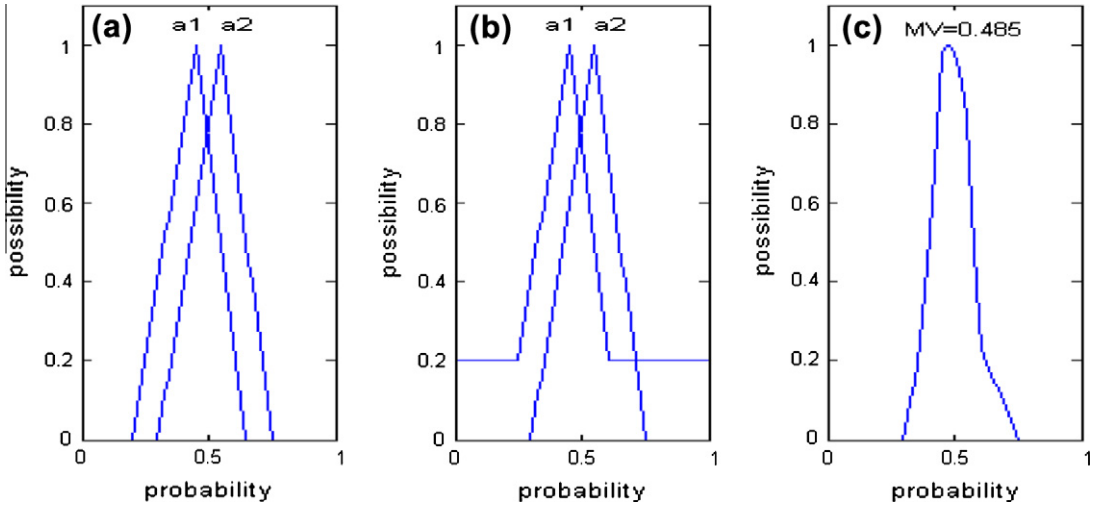

Fig. 6. (a) Two fuzzy probabilities, (b) weighted fuzzy probabilities and (c) aggregation of fuzzy probabilities.

Table 4
b Constant for the risk factors of first and second degree relatives with breast and other cancers.

b	1 Affected relative	2 Affected relatives	3 Affected relatives
1st Degree relatives with breast cancer	0.3	0.45	0.6
2nd Degree relatives with breast cancer	0.1	0.2	0.3
1st Degree relatives with other cancers	0.1	0.2	0.3
2nd Degree relatives with other cancers	0.05	0.1	0.15

In the other mentioned risk factors, we have chosen b based on the possible answers and description about their effect on BC development which are available in medical literature [24,28,40,46,51]. Chosen b for risk factors of "having 1st or 2nd degree relative with BC, having 1st or 2nd degree relative with other cancers" is shown in Table 4. In C3435T SNP risk factor, the alleles may be T or C where T-allele is high risk allele and C-allele is low risk allele. If the two alleles are T-allele then $b = 0.6$, if one allele is C and the other is T then $b = 0.3$ and when the two alleles are C then $b = 0$.

After tuning fuzzy probabilities the interface agent must evaluate whether the user is in HR or LR of BC development.

3.4. Fuzzy risk analysis based on ranking fuzzy probabilities

There are various methods for ranking fuzzy numbers for fuzzy risk analysis. By investigating the available methods, we have chosen a proper method proposed by Chen and Wang [69]. The eligibility of this method to others is that it can overcome the drawbacks of the previous methods [44,64–68] which had been employed in fuzzy risk analysis [69]. This method integrates many concepts such as the approximate area measure [70], the belief feature [71] and the signal noise ratio [72].

In this stage the interface agent is a decision maker who wants to determine the ranking order of the two fuzzy numbers $\overline{A}_i, i = 1, 2$. ($\overline{A}_1$ = fuzzy probability of high risk, $\overline{A}_2$ = fuzzy probability of low risk). The kth α_cut, $\overline{A}_i^{\alpha_k}$, of fuzzy probability $\overline{A}_i$ is defined as follows

$$\overline{A}_i^{\alpha_k} = \{p | \pi_{\overline{A}_i}(p) \geqslant \alpha_k, p \in [0,1]\}, \quad \alpha_k = k/n, k \in \{0,1,\ldots,n\} \tag{10}$$

where n denotes the number of α_cuts. The interface agent has chosen $n = 10$. The minimal value $l_{i,k}$ and the maximal value $r_{i,k}$ of the fuzzy probability $\overline{A}_i$ are defined as follows [69].

$$l_{i,k} = \inf_{p\in[0,1]} \{p | \pi_{\overline{A}_i}(p) \geqslant \alpha_k\}, \qquad r_{i,k} = \sup_{p\in[0,1]} \{p | \pi_{\overline{A}_i}(p) \geqslant \alpha_k\} \tag{11}$$

Fig. 7 shows the minimal value $l_{i,k}$ and the maximal value $r_{i,k}$ of the kth α_cut of the fuzzy probability $\overline{A}_i$. It also shows the minimal value

$l_{j,k}$ and the maximal value $r_{j,k}$ of the kth α_cut of the fuzzy number $\overline{A}_j$. The maximal barrier U and the minimal barrier L of the fuzzy probabilities $\overline{A}_i, i = 1,2.$ are defined as follows [69]:

$$U = \max_{\forall i}\{p|p \in \overline{A}_i^{\alpha}, 0 \leq \alpha \leqslant h_{\overline{A}_i}, i = 1,2\},$$
$$L = \min_{\forall i}\{p|p \in \overline{A}_i^{\alpha}, 0 \leqslant \alpha \leq h_{\overline{A}_i}, i = 1,2\} \quad (12)$$

where $\overline{A}_i^{\alpha}$ denotes the α_cut of the fuzzy probability $\overline{A}_i$ and $h_{\overline{A}_i}$ denotes the height of $\overline{A}_i$ defined as follows

$$h_{\overline{A}_i} = \sup_{p\in[0,1]}\pi_{\overline{A}_i}(p) \quad (13)$$

The signal/noise ratio $\hat{\eta}_{i,k}$ of the kth α_cut of the fuzzy probability $\overline{A}_i$ used in the proposed method is defined as follows

$$\hat{\eta}_{i,k} = (m_{i,k} - L)/(\delta_{i,k} + c), \quad m_{i,k} = (r_{i,k} + l_{i,k})/2,$$
$$\delta_{i,k} = r_{i,k} - l_{i,k} \quad (14)$$

where $m_{i,k}$ and $\delta_{i,k}$ denote the middle point and the spread of $\overline{A}_i^{\alpha_k}$ respectively, L denotes the minimal barrier of fuzzy probability $\overline{A}_i$ defined by Eq. (12), c is a parameter and $c > 0$, where the interface agent has chosen c as $U - L + 1$. The ranking index $RI(\overline{A}_i)$ of the fuzzy probability $\overline{A}_i$ is calculated as follows [69].

$$RI(\overline{A}_i) = \left(h_{\overline{A}_i}\sum_{k=1}^{n}\alpha_k \times \eta_{i,k}\right)/\left(\sum_{k=1}^{n}\alpha_k\right) \quad (15)$$

where $\alpha_k = h_{\overline{A}_i} \times \frac{k}{n}, k \in [0,1,..,n], n = 10$. The greater the value of $RI(\overline{A}_i)$ the better the ranking of $\overline{A}_i$. According to the explained procedure, the interface agent computes the ranking index of the fuzzy probabilities of high risk and low risk. Due to the higher ranking index of HR and LR fuzzy probabilities, the interface agent announces to the user that she is in high risk or low risk of breast cancer development.

Since different users may have different interpretations of high risk or low risk, we compute the HR percentage and LR percentage as follows, to numerically show the user that whether she is in HR or LR of BC development.

$$HR\_percent = (RI(HR))/(RI(HR) + RI(LR)),$$
$$LR\_percent = (RI(LR))/(RI(HR) + RI(LR)). \quad (16)$$

Additionally, we are interested to evaluate and represent how uncertain are the assessed fuzzy probabilities. To measure the amount of uncertainty of the assessed fuzzy probabilities which are also possibility distributions we employ the following definitions [73].

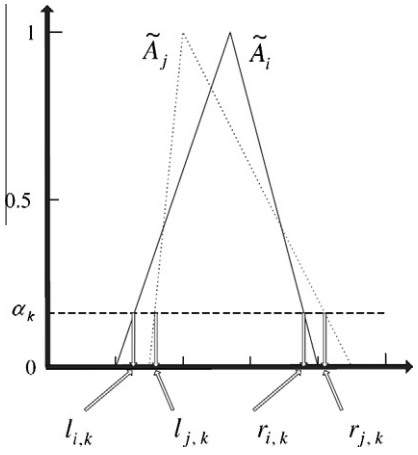

Fig. 7. Minimal values $l_{i,k}$ and maximal values $r_{i,k}$.

$$\alpha_x = \int_x \pi(x)dx, \quad \tilde{x} = \left(\int_x x\pi(x)dx\right)/(\alpha_x),$$
$$Un = \left(\int_x (x - \tilde{x})^2\pi(x)dx\right)/(\alpha_x) \quad (17)$$

where $\tilde{x}$ is called the center of gravity and Un is called the uncertainty [77]. x represents the probability which varies between [0,1] and $\pi(x)$ is the possibility of x probability. Un shows the level of uncertainty for HR and LR fuzzy probabilities. Measured uncertainty can ensure the user how much she can trust the assessed risk.

It should be mentioned that system automatically organizes its solution according to the provided data by the user. In the other words, if the user does not answer to specific risk factors, system automatically takes out the corresponding risk factors from the procedure of HR and LR fuzzy probabilities calculations and increases the uncertainty of the final calculated probabilities based on the missing risk factors. The system increases the calculated uncertainties for high risk and low risk fuzzy probabilities by a fixed amount called missing-factors-uncertainty (mfu), where $mfu = 0.5, 0.25, 0.12$ corresponds to strong, moderate and minor factors with missing data, respectively.

Finally, by calculating the HR and LR fuzzy probabilities, their percentages and their uncertainties, we prepare the fuzzy multi agent system to determine the insurance BC premium of any woman who has used the system to evaluate her BC risk. The inner structure and calculations of the proposed fuzzy-probabilistic MAS is depicted in Fig. 8.

4. Breast cancer insurance premium assignment based on the proposed method

BC risk assessment is a criterion of how the calculated risk can be decreased or controlled by preventive or screening treatments. Insurance companies can offer appropriate premiums and preventive facilities for BC development. In other words, different premiums can be assigned more appropriately to different BC risks which need different preventive and screening facilities.

There are several studies about the relationship between insurance and BC. Armstrong et al. indicates that information about increased BC is associated with increase in life insurance purchasing and compared to women who did not increase their life insurance coverage, women who increased insurance coverage were more likely to have a higher predicted life time risk of BC [74]. Lemair et al. [75] discussed that women who have learnt through genetic tests that they are at a higher risk of death by BC may purchase more insurance, which to them seems inexpensive because it is priced at a rate set for average risks and women who learn they are at lower risk may purchase less insurance. Lemair et al. analysis suggests that insurance companies could consider gathering as much information about family history as possible during the underwriting process and use BC and ovarian cancer information in setting premiums. Also they tried to determine the term insurance premium based on the available data of National Cancer Institute, where based on different age ranges, family history and genetic factors they tried to conservatively determine the BC insurance premium.

There is substantial evidence that lack of adequate health insurance coverage is associated with less access to care and poorer outcomes for cancer patients, also it is a major barrier to preventive health services and adequate treatment [76].

Therefore according to the above mentioned points we can benefit from the obtained information about fuzzy probabilities of HR and LR of BC development to help the insurance company properly adjust its BC premiums and services. Clearly high risk women are willing to pay higher premiums but low risk women prefer to pay lower premium prices. The premiums must then be tuned in

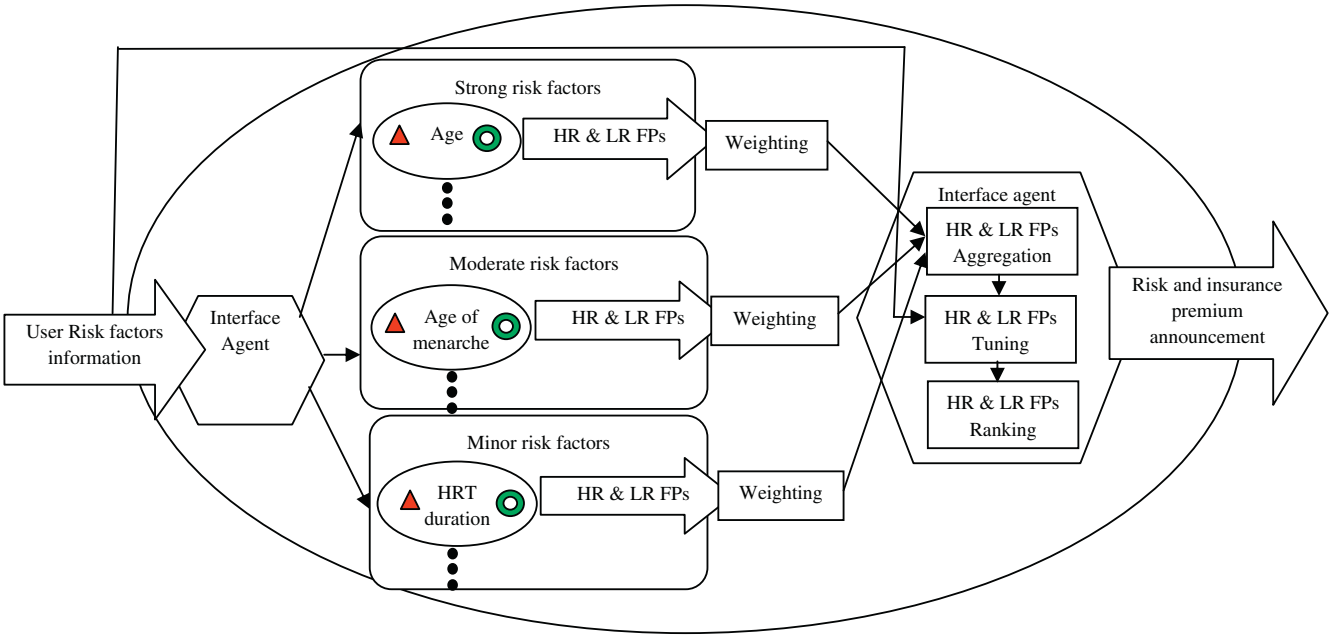

Fig. 8. Breast cancer risk assessment fuzzy-probabilistic MAS, ◎: risk factor soft data, ▲: risk factor statistical data.

a way that offers higher premiums for HR women and lower premiums for LR women, furthermore the premium formulations must guarantee that the insurance company will not have financial loss. Consequently, high risk women pay higher premium and can benefit from more preventive treatment services and low risk women who pay less premium can benefit from financial facilities for screening.

Valid statistics can show the prevalence of breast cancer among the women in any country, where we show this percentage by α (1 - α shows the percentage of women who are not affected by BC in the same country). For example, the prevalence of BC among Iranian women is 120 per 100,000 [77]. Therefore, α = (120/100,000) of women in Iran may be affected while $1-\alpha = (1-(120/100{,}000)$ of them may not be affected by BC. Therefore average price of BC premium, (*APP*) in any country, can be computed as below

$$APP = \alpha \times MEBCT + (1-\alpha) \times MEBCP \tag{18}$$

where *MEBCT* is the maximum expense of breast cancer treatment and *MEBCP* is the maximum expense of breast cancer prevention. In this paper it is assumed that *MEBCT* = 34,000,000, *MEBCP* = 3000,000, α = (120/100,000), $1-\alpha = (1-(120/100{,}000))$.

But we must tune the premium price based on the BC development risk of each woman. If in fuzzy probability ranking of HR and LR, a woman is high risk, her BC premium is then obtained as follows

$$\begin{aligned} &BC\_premium\_HR = APP \times (1 + HR\_percent) \times (1 + Un^{HR} + total\_mfu) \\ &Un^{HR} = (HR\_Un)/(HR\_Un + LR\_Un) \end{aligned} \tag{19}$$

where $(1 + HR\_percent)$ is a coefficient that increases the average premium based on the *HR_percent* (Eq. (16)). Also $(1 + Un^{HR} + total\_mfu)$ is a conservative coefficient that tunes the premium price based on the existing uncertainty to guarantee that the insurance company will not have financial loss where *total_mfu* is the sum of all missed factors uncertainties in the system which is equal to zero when there is not any missed factor.

If in FPs ranking it has been noticed that the woman is low risk, her insurance premium will be calculated as follows

$$\begin{aligned} &BC\_premium\_LR = APP \times (1 - LR\_percent) \times (1 + Un^{LR} + total\_mfu) \\ &Un^{LR} = (LR\_Un)/(HR\_Un + LR\_Un) \end{aligned} \tag{20}$$

The $(1 - LR\_percent)$ decreases the premium from the average premium for low risk woman and $(1 + Un^{LR} + total\_mfu)$ is a conservative coefficient which increases the premium based on the existing uncertainty to guarantee that the insurance company will not have financial loss. *HR_Un* and *LR_Un* are calculated by Eq. (17) based on the obtained high risk and low risk fuzzy probabilities respectively.

Preventive treatment and screening financial services may completely or partially cover mammography, magnetic resonance imaging (MRI), prophylactic ophorectomy, mastectomy and chemoprevention expenses and other treatment services which should have been considered in the maximum expenses of BC treatment and prevention in Eq. (18).

In this stage system employs different training and testing sets, extracted from our statistical database, to recommend fixed breast cancer premium prices for high risk and low risk women.

Since the system may have different results for different distributions of training and testing sets, we have validated the system for 10 different random distributions of training and testing sets. In any of these distributions, the training and testing sets contain 44 and 43 patients respectively. Each training set is given to the fuzzy-probabilistic multi agent system to calculate the HR and LR fuzzy probabilities. In addition to calculating the success average of risk assessment, the total insurance premium of the patients in the testing sets who are recognized as HR women (*TIP_HR*) in any distribution is also calculated based on Eq. (19). Similarly the total insurance premium of the patients testing sets who are recognized as LR women (*TIP_LR*) in any distribution is calculated based on Eq. (20). Since all the patients in the database are affected by breast cancer, the success average is defined as the ratio of the number of women who are recognized as HR women to the total population of testing set which is 43.

In any distribution testing, HR-insurance premium, *HR_IP*, and LR-insurance premium, *LR_IP*, are calculated as follows

$$\begin{aligned} &HR\_IP = (TIP\_HR)/(N_{HR}) \\ &LR\_IP = (TIP\_LR)/(N_{LR}) \end{aligned} \tag{21}$$

where N_{HR} and N_{LR} are the number of the women recognized as high risk and low risk in any distribution respectively. Finally the system recommends the mean of 10 available *HR_IP* s as the final insurance premium of high risk women and the mean of 10 available *LR_IP* s as the final insurance premium of low risk women. We demonstrate the described results in Table 5.

Based on the obtained results (Table 5), the mean of 10 available success averages is 88.14%. The final high and low risk insurance premiums are 0.98285e+007 and 3.9395e+005 respectively (see Table 5). We demonstrate the performance of the above fuzzy-probabilistic MAS by the below example.

4.1. Example

Assume that a user has asked the system to assess the risk of her BC development. The interface agent enquires about her risk factors (12 risk factors mentioned in Section 3.1). The user replies to the interface agent by entering her risk factors to the system as below

MATLAB 7.1 command window:
Warning: if you do not know the answer of any question press *.
User-age (enter a certain age) = 65.
Age-at-menarche (enter a certain age) = 17.
Age-at-menopause (enter a certain age) = 55.
Age-at-1st-child-birthday (enter a certain age) = 20.
Number-of-pregnancies (enter a certain number) = 6.
Hormone-usage-period (enter the period of usage by year) = 0.
Oral-contraceptive-usage-period (enter the period of usage by year) = 4.

In the next step the entered information is sent to the risk factor agents. Risk factor agents obtain the fuzzy information by choosing the parts of high risk and low risk sets based on the received information and the fuzzy information interval which is equal to 5. Afterwards they calculate the fuzzy probability of HR and LR of BC development based on their own risk factors. The interface agent combines the received fuzzy probabilities to obtain the aggregated fuzzy probabilities. It should be mentioned that when the user does not know the amount of some of her risk factors, the corresponding risk factor agents would weight these factor fuzzy probabilities by zero weight. Therefore these factors cannot influence the final fuzzy probabilities, but based on the kind of missing factor they increase the amount of uncertainties as we discussed.

In the next step, interface agent asks the user about the remaining risk factors to tune the aggregated fuzzy probabilities and obtains the final LR and HR fuzzy probabilities. These questions and the user answers are as below:

MATLAB 7.1 command window:
What is your C3435T SNP information? (CC = 1, TC = 2, TT = 3) = 3.
How many first-degree relatives with breast cancer do you have? (If you do not have any enter 0) = 0.
How many second-degree relatives with breast cancer do you have? (If you do not have any enter 0) = 1.
How many first-degree relatives with other cancers do you have? (If you do not have any enter 0) = 0.
How many second-degree relatives with other cancers do you have? (If you do not have any enter 0) = 0.

In this part if the user does not know the certain answers of some questions, the interface agent sets the b constant for those questions factors equal to 0, therefore these factors cannot have any influence on the tuning of final fuzzy probabilities but they increase the uncertainties of the assessed fuzzy probabilities. Finally by tuning the aggregated fuzzy probabilities the resulting final fuzzy probabilities are obtained and shown in Fig. 9.

In the last step the interface agent ranks the final fuzzy probabilities of HR and LR and announces the user risk based on the result of ranking as below:

"Patient is at high risk of being affected by breast cancer where the HR-percentage is 99.07 and the LR-percentage is 0.93".

The calculated uncertainties of HR fuzzy probability and LR fuzzy probability are respectively 1.2034e-005 and 1.9863e-004 where these uncertainty percentages are 0.0571 and 0.9429 respectively. System suggests the user to benefit from the insurance financial aids for prevention by proposing the following premium rate.

"You can benefit from the insurance financial aids for breast cancer prevention by paying the following premium rate"

User-premium-rate = 9,830,200.

5. Comparison and final results

In the proposed method we employ both soft data and statistical data in high and low risk fuzzy probabilities calculations. In this section the performance of the same system ignoring statistical data and employing only soft data in the process of fuzzy probabilities calculations, is evaluated. As we discussed in Section 3.2 the available soft data is employed to evaluate prior fuzzy probabilities (FPs) and the statistical data is employed to convert the prior FPs of high and low risks to posterior FPs of high and low risks. By considering just the available soft data in FPs calculations, posterior fuzzy probabilities remain the same as prior fuzzy probabilities, while the other characteristics and computations of this system is the same as the proposed fuzzy-probabilistic MAS. Since the men-

Table 5
The results of fuzzy-probabilistic MAS BC risk assessment and premium assignment based on fuzzy probabilities.

Fold number	Success average	Total HR insurance premium	Total LR insurance premium	HR-insurance premium (HR-IP)	LR-insurance premium (LR-IP)
1	0.8140	3.3547e+008	1.8755e+006	0.9585e+007	2.3444e+005
2	0.8837	3.7753e+008	1.4953e+006	0.9935e+007	2.9906e+005
3	0.9302	4.0460e+008	1.6596e+006	1.0115e+007	5.5320e+005
4	0.9535	4.0755e+008	0.6576e+006	0.9940e+007	3.2880e+005
5	0.8605	3.6502e+008	1.6218e+006	0.9865e+007	2.7030e+005
6	0.8837	3.6816e+008	2.3934e+006	0.9688e+007	4.7868e+005
7	0.8605	3.6916e+008	1.3959e+006	0.9977e+007	2.3265e+005
8	0.8372	3.5325e+008	2.8911e+006	0.9812e+007	4.1301e+005
9	0.9302	4.0300e+008	1.7894e+006	1.0075e+007	5.9647e+005
10	0.8605	3.4376e+008	3.1973e+006	0.9291e+007	5.3288e+005
Mean of each column	0.8814	3.7275e+008	1.89769e+006	0.98285e+007	3.9395e+005

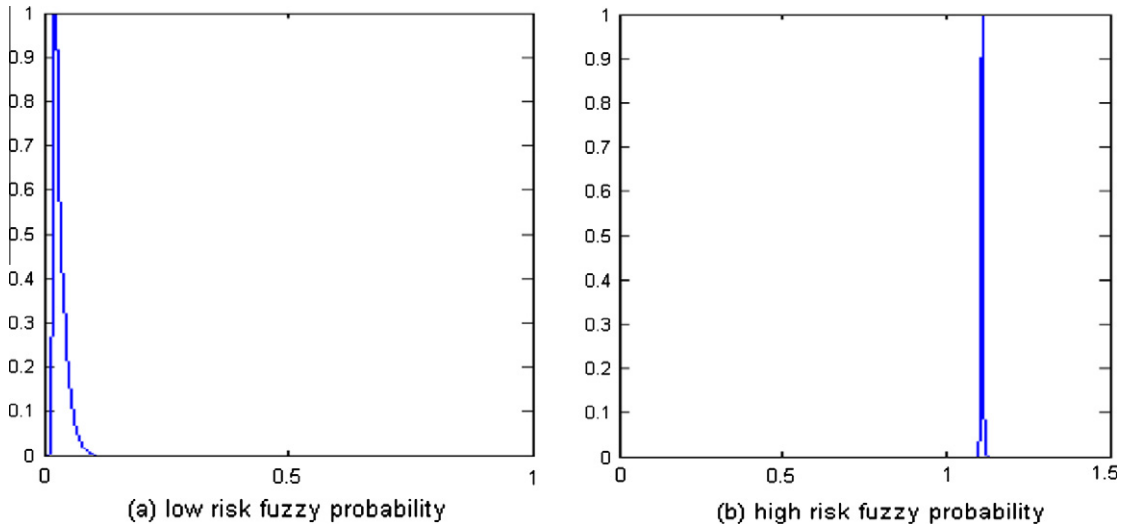

Fig. 9. (a) Low risk of BC development fuzzy probability and (b) high risk of BC development fuzzy probability.

tioned system still benefits from MAS capabilities and fuzzy computing we call it as "Fuzzy MAS" method.

To have a fair comparison, fuzzy MAS is validated for 10 different random distributions of training and testing sets. In any of these distributions, the training and testing sets contain 44 and 43 patients respectively. Table 6 shows the results of fuzzy MAS evaluation.

Based on the obtained results (Table 6), the mean of 10 available success averages is 73.49%, the final high and low risk insurance premiums are 1.1069e+007 and 0.95186e+006, which are respectively the mean of HR- and LR-insurance premiums columns in Table 6.

In Table 7 the performance of the proposed fuzzy-probabilistic MAS is compared against the performance of the mentioned fuzzy MAS method. The aim of this comparison is to evaluate the performance of the system employing both soft data and statistical data with the case of just soft data employment. Table 7 illustrates that the average success percentage of fuzzy-probabilistic MAS and fuzzy MAS methods are 88.14% and 73.49% respectively. By employing both soft data and statistical data in fuzzy-probabilistic MAS the fuzzy probabilities of HR and LR women are more precisely computed in comparison with fuzzy MAS which employs only soft data. Therefore in fuzzy-probabilistic MAS method HR and LR fuzzy probabilities and consequently HR and LR insurance premiums are more accurately calculated. These premiums ensure that the insurance company will not have financial losses based on the proposed premiums where the higher premiums of 88.14% of the population who are correctly recognized as HR women can compensate the lower premiums of the women who are mistakenly recognized as LR women. It should be mentioned that on both of these approaches, due to the wrong diagnosis of low risk women, their corresponding premiums (LR insurance premium) are incorrect.

On the other hand, in the fuzzy MAS risk assessment method, system is not able to recognize high risk women in the given test populations properly. Furthermore, due to the lack of statistical data, in fuzzy MAS method the uncertainty of HR and LR fuzzy probabilities is not decreased in the process of posterior fuzzy probabilities calculations, therefore HR and LR insurance premiums in this method are too conservatively tuned where high and low insurance are 1.1069e+007 and 0.95186e+006 respectively. Clearly in this method, the HR insurance premium of 73.49% of women who are correctly recognized as HR women cannot cover the financial loss of 26.51% of women who are mistakenly recognized as LR. It should be mentioned that the general revenue of insurance company in fuzzy-probabilistic multi-agent system is 8.7095e+006 and the insurance company revenue in fuzzy MAS method is 8.3867e+006 where the revenue is computed based on following equation:

$$revenue = SA \times (F\_HR\_IP) + (1 - SA) \times (F\_LR\_IP) \quad (22)$$

SA is the success average where in fuzzy probability and fuzzy MAS methods $SA = 0.8814$ and $SA = 0.7349$, (F_HR_IP) and (F_LR_IP) are the fixed premium rates for LR and HR women respectively which are available in Table 7.

Table 6
Results of fuzzy MAS method BC risk assessment and premium assignment (10 folds).

Fold number	Success average	Total HR insurance premium	Total LR insurance premium	HR-insurance premium (HR-IP)	LR-insurance premium (LR-IP)
1	0.6744	3.0920e+008	1.0690e+007	1.0662e+007	0.7636e+006
2	0.7209	3.4349e+008	0.9124e+007	1.1080e+007	0.7603e+006
3	0.6977	3.4806e+008	1.3296e+007	1.1602e+007	1.0228e+006
4	0.7907	3.6960e+008	0.9980e+007	1.0871e+007	1.1089e+006
5	0.7442	3.5117e+008	1.1326e+007	1.0974e+007	1.0296e+006
6	0.7209	3.4451e+008	1.2204e+007	1.1113e+007	1.0170e+006
7	0.7674	3.5962e+008	1.1961e+007	1.0898e+007	1.1961e+006
8	0.7442	3.5891e+008	0.9261e+007	1.1216e+007	0.8419e+006
9	0.7674	3.6383e+008	0.7668e+007	1.1025e+007	0.7668e+006
10	0.7209	3.4867e+008	1.2139e+007	1.1247e+007	1.0116e+006
Mean of each column	0.7349	3.49706e+008	1.07649e+007	1.1069e+007	0.95186e+006

Table 7
Comparing fuzzy-probabilistic MAS and fuzzy MAS results.

	Average success percentage	HR insurance premium	LR insurance premium	Maximum success average	Insurance company Revenue
Fuzzy-probabilistic MAS method	88.14	0.98285e+007	3.9395e+005	88.37	8.7095e+006
Fuzzy MAS method	73.49	1.1069e+007	0.95186e+006	83.72	8.3867e+006

In fuzzy-probabilistic MAS method, LR and HR women who must pay respectively 3.9395e+005 and 0.98285e+007 for their premium are more eager to purchase premium since the proposed premium is significantly lower in comparison with fuzzy MAS where the corresponding premium rates are 0.95186e+006 and 1.1069e+007 respectively.

Consequently, due to proceeding analysis and comparisons, the insurance company can determine the fuzzy-probabilistic MAS method is more accurate for breast cancer risk assessment and premium assignment in a certain population, where the method ensures the bilateral benefits of the insurance company as well as insured women.

Additionally the insurance company can determine its treatment, prevention and screening services based on the fixed HR and LR insurance premiums that are proposed by the fuzzy-probabilistic MAS.

It should be mentioned that the proposed fuzzy-probabilistic multi-agent system assumes $FII = 5$ in its computations, where decreasing it to $FII = 3$ leads to the increase of LR and HR premiums and consequently decreases the number of insurance purchasers and increases the general revenue of insurance company based on Eq. (26). In contrast, increasing *FII* to 7 leads to the decrease of premium rates and general revenue of insurance company, while attracting more insurance purchasers. This is while the success average of the proposed method in all these three choices of *FII* is similar.

6. Conclusions and future work

Here we propose a general paradigm for distributed risk assessment and insurance premium assignment under uncertain conditions by integrating the capabilities of fuzzy probability and multi-agent systems synergistically. Through this paper we show how fuzzy logic can be benefited to model soft data. We demonstrate that multi agent systems offer an attractive approach for distributed computing problems through decentralization, cooperation and result sharing with a higher reliability and flexibility. In the particular case of breast cancer development, the proposed fuzzy-probabilistic multi-agent system employs fuzzy computing and probabilistic computing to assess fuzzy risks and fuzzy probabilities to model uncertain probabilities. Also the proposed approach for fuzzy probabilities computation is especially helpful when we want to model and combine imprecise probabilities derived from linguistic data and statistical data. System can also estimate the breast cancer premium based on the calculated risk which can be offered by insurance companies for preventive and screening facilities. Finally we compare the results of the proposed fuzzy-probabilistic MAS with those of the fuzzy MAS. The results indicate the superiority of the proposed method in the case of breast cancer risk assessment and premium assignment. The system can also be improved as a social or private health system for assessing BC risk and offering preventive advices.

Regardless of the economic aspects, the proposed system can be employed as a private health system for any woman or as a public health system for any expert. The system can be improved by offering different preventive advices for any user about appropriate life style, health habits, screening and preventive methods.

The research can be extended by augmenting the available database by supplying the healthy women factors and re-evaluating the performance of the system. The available risk factors can be completed by considering breast density and other BC single-nucleotide polymorphisms (SNPs) factors. Finally, the proposed approach can be applied for other diseases.

References

[1] Zadeh LA. Possibility theory and soft data analysis. In: Cobb L, Thrall RM, editors. mathematical frontiers of the social and policy sciences. Boulder, Colorado: Westview Press; 1981. p. 69–129.
[2] Zadeh LA. Soft computing and fuzzy logic. IEEE Softw 1994;11(6):48–56.
[3] Zadeh LA. Fuzzy probabilities. Inf Process Manage 1984;20(3):363–72.
[4] Buckley JJ. Fuzzy probabilities, new approach and applications. Springer; 2005.
[5] Zadeh LA. Fuzzy sets. Inf Control 1965;8(3):338–53.
[6] Tatari F, Akbarzadeh-T M-R. Intensity assessment of pedestrian collisions in city of Mashhad based on fuzzy probabilities. IFSA2009, Portugal; 2009.
[7] Dunyak JP, Wunsch D. Fuzzy probability for system reliability. In: Proceedings of the 37th IEEE conference on decision & control Tampa; 1998.
[8] Dunyak JP et al. A theory of independent fuzzy probability for system reliability. In: IEEE transactions on fuzzy systems, vol. 7, no. 2; 1999.
[9] Kenarangui R. Event-tree analysis by fuzzy probability. IEEE Trans Reliab 1991;40(1).
[10] Zhou J. Reliability assessment method for pressure piping containing circumferential defects based on fuzzy probability. Int J Press Vessels Pip 2005;82:669–78.
[11] Hüllermeier Karimi E. Risk assessment system of natural hazards: a new approach based on fuzzy probability. Fuzzy Sets Syst 2007;158:987–99.
[12] Yu D, Park WS. Combination and evaluation of expert opinions characterized in terms of fuzzy probabilities. Ann Nucl Energy 2000;27:713–26.
[13] Zadeh LA. Toward a perception-based theory of probabilistic reasoning with imprecise probabilities. J Stat Plann Infer 2002;105:233–64.
[14] Wang G, Mao S, He Y. Optimal decision fusion with fuzzy priori probabilities. In: Proceedings of IBP; 1998.
[15] Wang G, Mao S, He Y. Distributed detection fusion with fuzzy priori probabilities and fuzzy cost function. In: IEEE international radar conference; 2000.
[16] Ralescu L, Ralescu DA, Yamakata Yoko. Inference by aggregation of evidence with applications to fuzzy probabilities. Inf Sci 2007;177:378–87.
[17] Huang Z, Gedeon TD. Information retrieval estimation via fuzzy probability. In: World automation congress (WAC); 2006.
[18] Buckley JJ et al. Fuzzy mathematics in economics and engineering. Heidelberg, Germany: Physica-Verlag; 2002.
[19] Kim J-H, Vadakkepat P. Multi-agent systems: a survey from the robot-soccer perspective. Intell Autom Soft Comput 2000;6(1):3–17.
[20] Wooldridge M. An introduction to multi agent systems. Willey; 2002.
[21] Layde PM, Webster LA, Baughman AL, Wingo PA, Rubin GL, Ory HW, The Center and Steroid Hormone Study Group. The independent associations of parity, age at first full term pregnancy and duration of breast feeding with risk of breast cancer. J Clin Epidemiol 1989;42(IO):963–73 [Printed in Great Britain].
[22] Tatari F, Akbarzadeh-T M-R, Azemi A. Agent-based centralized fuzzy Kalman filtering for uncertain stochastic estimation, ICSCCW, Cyprus; 2009.
[23] Corchado Juan M, Bajo J, de Paz Y, Tapia DI. Intelligent environment for monitoring Alzheimer patients, agent technology for health care. Decision Supp Syst 2008;44:382–96.
[24] Laleci GB, Dogac A, Olduz M, Tasyurt I, Yuksel M, Okcan A. SAPHIRE: a multi-agent system for remote healthcare monitoring through computerized clinical guidelines, SAPHIRE project by the Scienti_c and Technical Research Council of Turkey, Project No.: EEEAG 105E133; 2007.
[25] Jun Su C. Mobile multi-agent based, distributed information platform (MADIP) for wide-area e-health monitoring. Comput Ind 2008;59:55–68.
[26] Uckun S. Intelligent systems in patient monitoring and therapy management, a survey of research projects; 1993.
[27] Zhao X, Yuan S, Yu Z, Ye W, Cao J. Designing strategy for multiagent system based large structural health monitoring. Expert Syst Appl 2008;34:1154–68.
[28] González-Vélez H, Mier M, Julià-Sapé M, Arvanitis TN, García-Gómez JM, Robles M et al. Health agents: distributed multi-agent brain tumor diagnosis and prognosis. Springer Science + Business Media, LLC; 2007.
[29] Klüver J, Stoica C, Unland R. Multi-agent system and soft computing technologies as a fruitful symbiosis for reliable medical diagnoses. In: BISC

FLINT-CIBI international joint workshop on soft computing for internet and bioinformatics. US: University of Berkeley; 2003 [December 15–19].
[30] Laszlo BI. Cooperative medical diagnoses elaboration by physicians and artificial agents, emergent properties in natural and artificial, complex systems; 2007.
[31] Unland R, Ulieru M. Swarm Intelligence and the Holonic Paradigm: a promising symbiosis for a medical diagnostic system, KES, vol. 2; 2005.
[32] Nguyen MT, Fuhrer P, Pasquier-Rocha J. Enhancing e-health information systems with agent technology. 1700 Fribourg, Switzerland: University of Fribourg, Department of Computer Science; 2008.
[33] Tian J, Tianfield H. A multi-agent approach to the design of an E-medicine system. Berlin Heidelberg: Springer-Verlag; 2003.
[34] Ulieru M, Hadzic M, Chang E. Soft computing agents for e-Health in application to the research and control of unknown diseases. Inf Sci 2006;176:1190–214.
[35] Coltell O, Arregui M, Falomir Z, Corella D, Ordovas JM. Multi-agent approaches for biomedical informatics applications in the health framework: genomic epidemiology and nutrigenomics. Bio Info Genomica subgroup of the systems Integration and Re-Engineering Research Group (IRIS). Universitat Jaume I. Avda. Vicent Sos Baynat, s/n. 12071-Castellón, Spain; 2005.
[36] Linard C, Ponc N, Fontenille D, Lambin EF. A multi-agent simulation to assess the risk of malaria re-emergence in southern France. Ecol Model 2009;220:160–74.
[37] Edwards QT, Palomares MR. Assessment of risk for breast cancer utilizing history & quantitative models in primary care. J Nurse Pract JNP; 2008. p. 361.
[38] Orgun B, Vub J. HL7 ontology and mobile agents for interoperability in heterogeneous medical information systems. Comput Biol Med 2006;36:817–36.
[39] Sutton D, Patkar V. Common KADS analysis and description of a knowledge based system for the assessment of breast cancer. Expert Syst Appl 2009;36:2411–23.
[40] Fletcher SW, Hayes DF, Dizon DS. <http://www.uptodate.com/contents/patient-information-risk-factors-for-breast-cancer> (accessed 2010).
[41] Hulka BS, Moorman PG. Reprint of breast cancer: hormones and other risk factors. Maturitas 2008;61(1–2):203–13.
[42] Mackay J, Schulz P, Rubinelli S, Pithers A. Online patient education and risk assessment: project OPERA from cancer backup putting inherited breast cancer risk information into context using argumentation theory. Patient Educ Couns 2007;67:261–6.
[43] Park K, Chang SJ, Chang Kim H, Cheol z Park E, Sook Lee E, Mo Nam C. Big gap between risk perception for breast cancer and risk factors: nationwide survey in Korea. Patient Education and Counseling; 2009.
[44] Chen SJ, Chen SM. A new method for handling multi criteria fuzzy decision-making problems using FN-IOWA operators. Cyber Syst 2003;34(2):109–97.
[45] Bellcross C. Approaches to applying breast cancer risk prediction models in clinical practice. Oncol Genet 2009;6(8).
[46] Dipiro JT, Tablert RL et al. Pharmacotherapy: a pathophysiologic approach. 5th ed. New York: Mc Graw-Hill; 2002. p. 2223–51.
[47] Devita VT, Hellman S, Rosenberg SA. Cancer: principles and practice of oncology. 7th ed. Philadelphia: Lippincott Williams and Wilkins; 2005. p. 1415–25.
[48] Carmichael AS, Sami JM. Dixon, breast cancer risk among the survivors of atomic bomb and patients exposed to therapeutic ionizing radiation. EJSO 2003;29:475–9.
[49] Santen RJ, Boyd NF, Chlebowski RT, et al. Critical assessment of new risk factors for breast cancer: considerations for development of an improved risk prediction model. Endocr Relat Cancer 2007;14:169–87.
[50] Bassett LW, Hendrick RE, Bassford T, Butler PF, Carter D et al. Quality determinants of mammography. Rockville, MD: US Department of Health and Human Services, Agency for Health Care Policy and Research; 1994.
[51] Gammon MD, John EM, Britton JA. Recreational and occupational physical activities and risk of breast cancer. J Natl Cancer Inst 1998;90:100–17.
[52] Bell J. The new genetics in clinical practice. BMJ 1998;316:618–20.
[53] Beaudet AL. ASHG presidential address: making genomic medicine a reality. Am J Hum Genet 1998;64(1999):1–13.
[54] Lichtenstein P, Holm NV, Verkasalo PK, et al. Environmental and heritable factors in the causation of cancer-analyses of cohorts of twins from Sweden, Denmark and Finland. N Engl J Med 2000;343:78–85.
[55] Peto J, Mack TM. High constant incidence in twins and other relatives of women with breast cancer. Nat Genet 2000;26:411–4.
[56] Easton DF, Pooley KA, Dunning AM, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 2007;447:1087–93.
[57] Hunter DJ, Kraft P, Jacobs P, et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. Nat Genet 2007;39:870–4.
[58] Stacey SN, Manolescu A, Sulem P, et al. Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. Nat Genet 2007;39:865–9.
[59] Turgut S, Yaren A, Kursunluoglu R, Turgut G. MDR1 C3435T polymorphism in patients with breast cancer. Arch Med Res 2007;38:539–44.
[60] Pharoah P, Antoniou AC, Easton DF, Ponder B. Polygenes, risk prediction, and targeted prevention of breast cancer. New Engl J Med; 2008 [June 26].
[61] Gail MH, Brinton LA, Byar DP, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. J Natl Cancer Inst 1989;81:1879–86.
[62] Claus EB, Risch N, Thompson WD. Autosomal dominant inheritance of early-onset breast cancer: implications for risk prediction. Cancer 1994;73:643–51.
[63] Tyrer J, Duffy SW, Cuzick J. A breast cancer prediction model incorporating familial and personal risk factors. Stat Med 2004;23:1111–30.
[64] Chen SJ, Chen SM. Fuzzy risk analysis based on similarity of generalized fuzzy numbers. IEEE Trans Fuzzy Syst 2003;11(1):45–56.
[65] Cheng CH. A new approach for ranking fuzzy numbers by distance method. Fuzzy Sets Syst 1998;95(3):307–17.
[66] Murakami S, Maeda S, Imamura S. Fuzzy decision analysis on the development of a centralized regional energy control system. In: Proceedings of the IFAC symposium on fuzzy information, knowledge representation and decision analysis, Tokyo, Japan; 1983. p. 363–8.
[67] Yager S. On a general class of fuzzy connectives. Fuzzy Sets Syst 1980;4(3):235–42.
[68] Yong D, Qi L. A TOPSIS-based centroid-index ranking method of fuzzy numbers and its application in decision-making. Cyber Syst 2005;36(6):581–95.
[69] Chen S-M, Wang C-H. Fuzzy risk analysis based on ranking fuzzy numbers using a-cuts, belief features and signal/noise ratios. Expert Syst Appl 2009;36:5576–81.
[70] Chen LH, Lu HW. An approximate approach for ranking fuzzy numbers based on left and right dominance. Comput Math Appl 2001;41(12):1589–602.
[71] Lu HW, Wang CB. An index for ranking fuzzy numbers by belief feature. Inf Manage Sci 2005;16(3):57–70.
[72] Chen LH, Lu HW. The preference order of fuzzy numbers. Comput Math Appl 2002;44(10):1455–65.
[73] Matiaa F, Jiméneza A, Al-Hadithia BM, Rodriguez-Losada D, Gallna R. The fuzzy Kalman filter: state estimation using possibilistic techniques. Fuzzy Sets Syst 2006;157:2145–70.
[74] Armstrong K, Weber B, FitzGerald G, Hershey JC, Pauly MV, Lemaire J, et al. Life insurance and breast cancer risk assessment: adverse selection. Genet Test Decision Discrim Am J Med Genet 2003;120A:359–64.
[75] Lemaire J, Subramanian K, Armstrong K, Asch DA. Pricing term insurance in the presence of family history of breast or ovarian cancer. North Am Actuar J 2000;4(2).
[76] Ward E, Halpern M, Schrag N, Cokkinides V, DeSantis C, Bandi P, et al. Association of insurance with cancer care utilization and outcomes. CA Cancer J Clin 2008;58:9–31.
[77] Mousavi S-M, Montazeri A, Mohagheghi M-A, Mousavi-J A, Harirchi I, Najafi M, et al. Breast cancer in Iran: an epidemiological review. Breast J 2007;13(4):383–91.