GUIDELINE

Using a Genetic-Fuzzy Algorithm as a Computer Aided Breast Cancer Diagnostic Tool

Abir Alharbi[1]*, F Tchier[1], MM Rashidi[2]

Abstract

Computer-aided diagnosis of breast cancer is an important medical approach. In this research paper, we focus on combining two major methodologies, namely fuzzy base systems and the evolutionary genetic algorithms and on applying them to the Saudi Arabian breast cancer diagnosis database, to aid physicians in obtaining an early-computerized diagnosis and hence prevent the development of cancer through identification and removal or treatment of premalignant abnormalities; early detection can also improve survival and decrease mortality by detecting cancer at an early stage when treatment is more effective. Our hybrid algorithm, the genetic-fuzzy algorithm, has produced optimized systems that attain high classification performance, with simple and readily interpreted rules and with a good degree of confidence.

Keywords: Fuzzy systems - genetic algorithms - optimization methods - breast cancer - computer aided diagnosis

Asian Pac J Cancer Prev, **17 (7)**, 3651-3658

Introduction

In medical science, diagnosis of a disease is very complicated, and many tests must be done on patients to obtain a near accurate diagnosis. This has given rise to computerized diagnostic tools, intended to aid the physician in making primary medical decisions and hence an early diagnosis. A major area for such computerized tools is in the domain of cancer diagnosis. Specifically breast cancer, since the physician needs to know early on whether the patient under examination exhibits the symptoms of a benign, or a malignant case. The computerized assisted diagnostic tools should attain the highest possible performance, which means they must classify correctly benign or malignant cases with a good degree of confidence. Moreover, it would be desirable for such diagnostic systems to be well interpreted by the physicians.

In this research paper, we combine two methodologies, namely the fuzzy systems and the genetic algorithms to automatically produce automated systems for breast cancer early diagnosis. The major advantage of fuzzy systems is the simple interpretation; however, finding good fuzzy systems is a hard task. This is where genetic algorithm contributes, helping us in optimizing a computer production of the fuzzy systems, based on a database of training cases. There are several different examples of the application of fuzzy systems and evolutionary algorithms in the medical domain, such as applying to the Wisconsin Breast Cancer Diagnosis Data (WBCD) in USA (Andreas et al., 1999), or applying them on pathogenesis of acute sore throat conditions in humans (Carmona et al., 2015), or combining with wavelets as in (Nguyen et al., 2015) or with neural networks as in (Rashidi et al., 2011). In our paper, we present the genetic-fuzzy algorithm, which we developed for the Saudi Arabian breast cancer database consisting of 260 patients collected nationwide. Our aim is to aid physicians in obtaining an early-computerized diagnosis and hence prevent the development of cancer through identification and removal or treatment of premalignant abnormalities; early detection can also improve survival and decrease mortality by detecting cancer at an early stage when treatment is more effective.

In the next sections, we provide a brief overview of fuzzy systems and genetic algorithms respectively. Then, we describe the genetic-fuzzy approach, the Saudi breast cancer data, and the relative parameters settings. In the last sections, we discuss the results of our best-evolved systems, and finally we present our concluding remarks and future work.

Fuzzy Systems

Fuzzy logic is a form of many-valued logic, which deals with reasoning that is approximate rather than fixed and exact. Compared to traditional binary sets (where variables may take on true or false values) fuzzy logic variables may have a truth-value that ranges in degree between 0 and 1. Fuzzy logic has been extended to handle the concept of partial truth, where the truth-value may range between completely true and completely false. Hence, Fuzzy logic is a computational method

[1]*Mathematics Department, King Saud University, Riyadh, Saudi Arabia,* [2]*Department of Mechanical Engineering, Tongji University, Shanghai,China *For correspondence: abir@ksu.edu.sa.*

manipulating information in a way that resembles human logical reasoning processes (Yager and Filev, 1994; Yager and Zadeh, 1994). A fuzzy variable is characterized by its fuzzy variable and the membership functions of these variables.

Figure 1 shows an example of a fuzzy variable with two possible values labelled Low and High, and orthogonal membership functions, to guaranty the sum of all membership functions at any point is one. The plot shows degree of membership with input values, m and n defining the start point and the length of membership function edges, respectively.

A fuzzy inference system is a rule-based system that uses fuzzy logic, rather than Boolean logic (Zadeh, 1965). The structure includes four main components a fuzzifier, translating crisp (real valued) inputs into fuzzy values, an inference engine applying a fuzzy reasoning mechanism to obtain a fuzzy output, a defuzzifier, translating the output back into a crisp value; and a knowledge base, containing both an ensemble of fuzzy rules, and a group of connection membership functions as seen in Figure 2. Moreover, the decision making process is performed in the inference engine using the rules contained in the rule base. These fuzzy rules define the input and output of the fuzzy variables. A fuzzy rule has the form [if antecedent then consequent], where the antecedent is a fuzzy-logic expression composed of one or more simple fuzzy expressions connected by fuzzy operators, and the consequent is an expression that assigns fuzzy values to the output variables. The inference engine performance the learning phase where it evaluates all the rules in the rule base and combines the weighted consequents of all relevant rules into a single fuzzy set using the aggregation operation (Mendel, 1995). An example of a fuzzy rule in our case would be: if (v1 is Low) and (v2 is Low) then (output is benign; where v1 and v2 are variables given in the data set.

Using the direct fuzzy model with knowledge from a human expert, the fuzzy modelling identifies the parameters of a fuzzy inference system so that a desired decision can be made. This task is difficult when the problem space is complicated and very large; thus, it motivates us to apply genetic algorithms to this space and produce optimum fuzzy models. In the literature, there are several approaches to fuzzy modelling based on neural networks (Jang and Sun, 1995), genetic algorithms (Alander, 1997; Cordon et al., 1997; Heider and Drabe, 1997), and linear programming (Mangasarianet al.,1994). Selection of relevant variables and adequate rules is critical for obtaining a good accurate classification system. One of the major problems in fuzzy modelling is that the amount of computation grows exponentially with the number of variables. The parameters of fuzzy inference systems can be classified into four categories, logical, structural, connective, and operational. In fuzzy modelling, logical parameters are usually predefined from experience in the problem settings. Typical choices for the reasoning mechanism are Mamdani-type, Takagi-Sugeno-Kang, and singleton-type (Vourimaa, 1994). Common fuzzy operators are min, max, product, probabilistic and sum (Tchier, 2013). The most common membership functions are triangular, trapezoidal, and bell-shaped. For defuzzification, the (COA) and the mean of maxima (MOM) methods are the mostly used (Mendel, 1995;Tchier, 2014).

Genetic Algorithms

A genetic algorithm (GA) is a search heuristic that mimics the process of natural selection. GA is used to generate useful solutions to optimization and search problems. GA belong to the larger class of evolutionary algorithms (EA), which generate solutions to optimization problems using techniques inspired by natural evolution, such as inheritance, mutation, selection, and crossover (Koza, 1992).GA are usually applied to spaces which are too large to be exhaustively searched and they have many applications in bioinformatics, medical (Michalewicz, 1996), science, engineering (Rashidi et al., 2011), economics, manufacturing, computational mathematics (Alharbi et al., 2007), and many other fields.

The genetic algorithm method is an iterative procedure that involves a population representing the search space for solutions to the problem, as individuals, each one represented by a finite string of symbols, called the genome. The basic genetic algorithm proceeds as follows: an initial population of individuals is generated at random or heuristically. In every evolutionary step (gene rationstep), the individuals in the current population are decoded and evaluated according to a fitness function that describes the optimization problem in the search space. To form a new population (the next generation), individuals are selected according to their fitness. Many selection procedures are available, one of the simplest being fitness-proportionate selection, where individuals are selected with a probability proportional to their relative fitness. This ensures that the expected number of times an individual is chosen is approximately proportional to its relative performance in the population. Thus, high-fitness individuals stand a better chance to reproduce and bring new individuals to the population, while low-fitness will not.

New individuals are introduced into the population by genetic operators called crossover and mutation. Crossover is performed with probability between two selected individuals (parents) exchanging parts of their genomes to form two new individuals (offspring's). The mutation operator prevents premature convergence to local optima by randomly sampling new points in the search space; it is performed by flipping bits at random, with some small probability. GA is a stochastic iterative processes, which is not necessarily guaranteed to converge, and the stopping condition may be specified as a maximal number of generations or a chosen level of the fitness.

Genetic-Fuzzy Algorithms

Since evolutionary algorithms are used to search large complex, search spaces and are able to give optimal and near-optimal solutions on numerous diverse problems, therefore genetic-fuzzy algorithms can be considered as a modelling optimization process where the parameters of a fuzzy system constitute the search space as seen in Figure 2. Many researchers investigated the application of evolutionary techniques in the domain of fuzzy modelling (Kovalerchuk et al., 1997; Muthukrishnan, 2014), where the tuning of fuzzy inference systems involved in control tasks were done by genetic algorithms. Evolutionary fuzzy modelling has been applied to many domains, branching into many areas as chemistry, telecommunications (Heider and Drabe, 1997; Herrara et al., 1995), biology (Lee and Takagi, 1993), geophysics and medicine (Andres et al., 1999; Carmona et al., 2015). The evolutionary algorithm can be used to tune the knowledge contained in the fuzzy system by finding membership function values. An initial fuzzy system is defined by an expert. Then, the membership function values are encoded in a genome, and an evolutionary algorithm is used to find systems with high performance. Evolution often overcomes the local-minima problem seen in other gradient descent-based optimization methods. Artificial evolution can be applied in different stages of the fuzzy parameters search depending on several conditions like the availability of a priori knowledge, the size of the parameter, and the availability and completeness of input/output data. These types of fuzzy parameters whcih can be used to define targets for evolutionary fuzzy modelling are: structural parameters, connective parameters, and operational parameters.

In many cases, the available information about the system is composed almost exclusively of input/output data, and specific knowledge make up the system structure. In such a case, evolution has to deal with the simultaneous design of rules, membership functions, and structural parameters. Structure learning permits to specify other criteria related to the interpretability of the system, such as the number of membership functions and the number of rules. While, the strong interdependency among the parameters involved in this form of learning may slow down the convergence of the genetic algorithm. Both connective and structural parameters modelling are viewed as rule base learning processes with different levels of complexity. In the evolutionary algorithm applications, the main approaches for evolving such rule systems are the Michigan approach, the Pittsburgh approach (Alander, 1997), and the iterative rule learning approach (Karr, 1991).

In the Michigan approach, each individual represents a single rule, and the entire population represents the fuzzy inference system. Since several rules participate in the inference process, the rules are in constant competition for the best action to be proposed, and cooperate to form an efficient fuzzy system. In the Pittsburgh approach, the evolutionary algorithm maintains a population of candidate fuzzy systems, each individual representing an entire fuzzy system. Selection and genetic operators produce new generations of fuzzy systems. This approach allows including additional optimization criteria in the fitness function, thus affording the implementation of multi-objective optimization. The main disadvantage of this approach is its computational cost, since a population of a complete fuzzy system has to be evaluated each generation.

Saudi Breast Cancer Data

Breast cancer is known as one of the most common cancers types affecting the female population. It is one of the major causes of death among women and a true emergency for health care systems of industrialized countries. It is one of the major causes of death among women and a true emergency for health care systems of industrialized countries. One of the epidemiological studies conducted by (Al-Diab et al., 2013) reported that the incidence of breast cancer in Saudi Arabia was 19.8% of all the female cancers detected in Saudi Arabia (El-Akkadal et al., 1986). Top researchers in the field such as GLOBOCAN project (Ferlay et al., 2013) have shown that breast cancer is the second most common malignancy for women in Saudi Arabia in 2012 (Figure 3). Nevertheless, there is a paucity of detailed published epidemiologic data. An earlier report according to Saudi National Cancer Registry mentioned an increasing proportion of breast cancer among women of different ages from 10.2% in 2000 to 24.3% in 2012 (Al Diab et al., 2013). The presence of a breast mass is an alert, but it does not always indicate a malignant cancer. Fine needle aspiration (FNA)2 of breast masses is a cost-effective, non-traumatic, and mostly non- invasive diagnostic test that obtains information needed to evaluate malignancy. The medical diagnosis data of breast cancer used in this study is from patients in Saudi Arabia (AlDiab, et al., 2013). The database is similar to the WBCD database of the University of Wisconsin Hospital (Merez and Murphy, 1996), where diagnosis of breast masses is based solely on an FNA test Nine visually assessed characteristics of an FNA sample considered relevant for diagnosis are identified, and were assigned an integer value between 1 and 10. The diagnostics in the database were done by specialists in the field, and the database itself consists of 260 cases, with each entry representing the classification for a patient with eleven entries: (patient number, v1, v2, v3,.., v9, Diagnostic: Benign or Malignant). The nine measured variables are as follows: v1 is clump thickness, v2 is uniformity of cell size, v3 is uniformity of cell shape, v4 is marginal adhesion, v5 is single epithelial cell size, v6 is bare nuclei, v7 is bland chromatin, v8 is normal nucleoli and v9 is mitosis.

Basically, an initial fuzzy rule base is defined by an expert, for example a fuzzy rule can be given as: if [v1 *is Low*] and [v7 *is Low*] then (output is benign). The genetic algorithm then fine-tunes the membership functions, i.e. the m and n values defining Low and High (Figure 1). The genetic algorithm is also used to find either the rule consequents, or other subset rules to be included in the rule base. As the membership functions are fixed this approach lacks the flexibility to modify substantially the system behavior. One of the major disadvantages of knowledge tuning is its dependency on the initial setting of the knowledge base. Furthermore, as the number of variables and membership functions increases, large dimensionality decreases the system's performance. Evolutionary structure learning is done by encoding within the genome an entire fuzzy system using the Pittsburgh approach. The fuzzy system computes a continuous appraisal value of the malignancy of a case, based on the input values. According to the fuzzy system's output, the threshold unit then outputs a benign or malignant diagnostic. In order to evolve the fuzzy model we must set some preliminary parameters in the fuzzy system itself and in the genetic algorithm encoding.

Genetic-Fuzzy Parameters

All previous knowledge about the problem and about the existent rule-based models gives us valuable information for our choices of fuzzy parameters. It has been shown in previous work that systems with no more than four rules obtain high performance (Andreas et al., 1999) and there is no need for a higher number of rules. Moreover, small number of variable is associated with benign cases, and the higher-valued variables are associated with malignancy. Each variable should have semantic meaning and the fuzzy set should clearly define a range that describes it. Any value belongs to at least one fuzzy set (Low, High, or both); no value lies outside the range of all sets. Since all the labels have semantic meaning, for each label, at least one element of the space should have a membership value equal to one. Hence, a Low membership value of 0.8 entails a High membership value of 0.2, and for each element, the sum of all its membership values should be equal to one. The parameter settings are set as in the following.

A.The fuzzy system parameters:

i). Logical parameters:

Reasoning mechanism: singleton-type fuzzy system (output membership are real values).

Fuzzy operators: min and max.

Input membership function type: orthogonal, trapezoidal.

Defuzzification method: weighted average.

ii). Structural parameters:

Relevant variables: specified by the genetic algorithm.

Number of input membership functions: two, denoted Low and High.

Number of output membership functions: two singletons for the benign and malignant diagnostic cases.

Number of rules: specified by the user between 1 and 4, found by the genetic algorithm.

Antecedents of rules: found by the genetic algorithm.

Consequent of rules: the algorithm finds rules for the benign diagnostic; the malignant diagnostic is an else condition.

Rule weights: the learning is done by letting active rules have a weigh of value 1, and the else condition has a weight of 0.25.

Input membership function values: found by the genetic algorithm

Output membership function values: following the database provided, we used a value of 2 for benign and 4 for malignant.

B. The genetic algorithm system parameters:

We applied the Pittsburgh-style-structure learning, namely, using a genetic algorithm to search for three parameters, the genome, input membership function values, and antecedents of rules are: *i*). Membership function parameters: Nine variables (v1- v9) each with two parameters m and n, defining the start point and the length of the membership function, respectively. *ii*). Antecedents: The i-th rule has the form: if (v1 is M^1i) ...and (v2 is M^9i) then (output is benign), where M_{ji} represents the membership function, which can take on the values: 1 for Low, 2 for High, or 0 for Other.

To evolve the fuzzy inference system, we used a genetic algorithm with a fixed population size of 50 individuals, with the length of each genome depending on the number of rules (a three rule has genome with 45 bits). The algorithm terminates when the maximum number of generations is reached at 300, or when the increase in fitness of the best individual over five successive generations falls below a certain threshold, set at 2×10^{-6}. Our fitness function Fis set to the classification performance, computed as the percentage of cases correctly classified, given by

$$F=Fr-\alpha\ Fc \qquad (1)$$

where $\alpha = 0.1$, Fr, the ratio of correctly diagnosed cases, which is the most important measure of performance, and Fc measures the confidence, penalizing systems with large number of low appraisal value cases i.e., cases that are diagnosed with low confidence. The crossover between the two chosen parents genome is done at a single point randomly chosen with probability 0.8 to produce the new generation offspring. The selection operator of parent's genome is set to the stochastic uniform selection method, and the mutation done on the new offspring has probability 0.01. Hence, the experiment starts by finding from a population of 50 genomes of length 45, where the first 18 bits represent the parameters of the membership functions (m, n) of each Vi and the remaining 27 bits are the output function M_{ji} for each Vi in the three rule base system showing Low or High or irrelevant. Table 1 shows the parameters encoding to form a single individual is genome. The GA runs throughout the generations to find the best genome in this population. The best genome is the one, which classifies correctly the largest number of the 260 cases given in the data set. After all 300 generations (repeated 50 times), the genetic algorithm finds the optimum genome; hence, it finds the best diagnostic system.

Result

The solution scheme we present for the Saudi breast cancer database diagnosis consists of a fuzzy system and a threshold unit. The fuzzy system computes a continuous appraisal value of the malignancy of a case, based on the input values. The threshold unit then outputs a benign or malignant diagnostic according to the fuzzy system's output. In order to evolve the fuzzy model we must set the fuzzy system parameters and the genetic algorithm encoding according to parameter settings discussed earlier. Table 1 shows the parameters encoding, forming a single individual genome. Figure 4 shows an example of a sample genome structure with its interpretation from a single rule fuzzy system, with 27 parameters: where m1=1, n1=5, m2=2, n2=3, …m9=1, n9=4, and membership functions : M_1^1=1, M_2^1=0, M_3^1=1, M_4^1=0, M_5^1=2, M_6^1= M_7^1= M_8^1= M_9^1=0.

The evolutionary experiments performed fall into three learning categories, in accordance with the data partitioning into two distinct sets: training set and testing set. The three experimental categories are: *i*). Training set contains all 260 cases of the database, while the testing set is empty. *ii*). Training set contains 75% of the data cases, and the testing set contains the remaining 25% of the cases. *iii*). Training set contains 50% of the database cases and the testing set contains the remaining 50% of the cases.

In the last two categories, the choice of training-set cases is done randomly, and is performed at the outset of every evolutionary run. The number of rules per system was fixed at the beginning, to be between one and four, i.e. evolution seeks a system with a given number of rules. Fifty evolutionary runs were performed, all of which found systems whose classification performance is above 95% of the cases in the dataset correctly diagnosed. MATLAB Genetic Toolbox (Matlab Toolboxes, 2015) was modified to implement the genetic-fuzzy algorithm and to generate the graphs of the results. CPU time is efficient since it takes average 4.8 minutes to go through 300 generations with a 2.4 GHz Intel Core i5 processor.

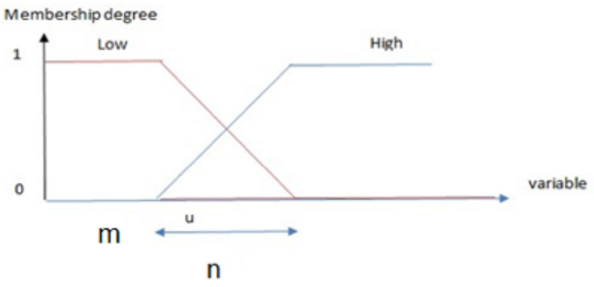

Figure 1. Example of a Fuzzy Variable with Values High and Low

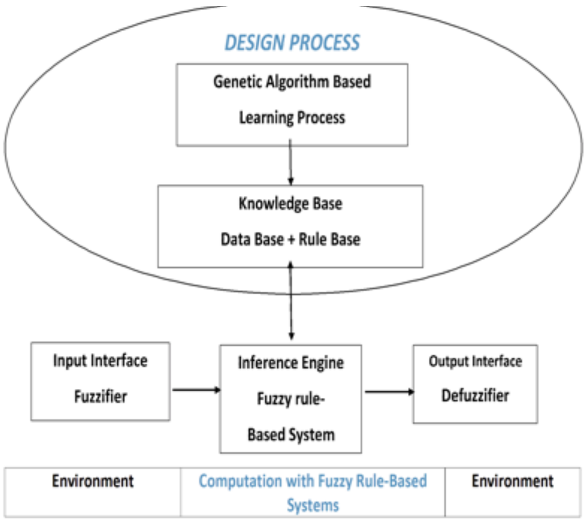

Figure 2. Basic Structure of a Genetic-fuzzy System

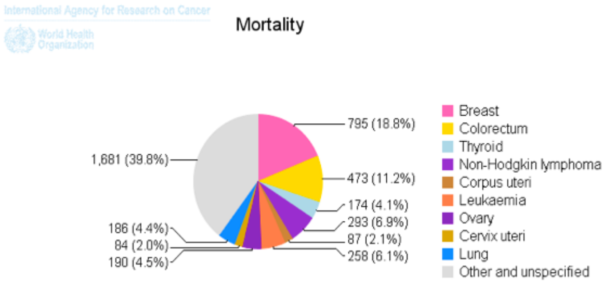

Figure 3. Percentages of Several Types of Cancers Detected in Saudi Arabia in 2012

Figure 5, consists of the best diagnostic system with three rules (45 parameters). Taking into account the performance classification rate this system is the top one over all 50 evolutionary runs. It obtained 98.3% correct classification rate over the benign cases, 96.2% correct classification rate over the malignant cases, and an overall classification rate of 97.33%. Table 2 presents the average performance obtained by the genetic algorithm with this system over all 50 evolutionary runs, divided according to the three experimental categories. The performance value denotes the percentage of cases correctly classified. Three such performance values are shown: the performance over the training set; the performance over the test set; and the overall performance on the entire database.

Even though our work is on a different dataset our proposed fuzzy three rule system described in this paper performs very well and reaches comparable results similar to work done on the WBCD dataset(Andres et al., 1999; Setiono, 1996) in terms of both performance and simplicity of rule as seen in Table 3. It is worth noting that

	V1	V2	V3	V4	V5	V6	V7	V8	V9
m	1	2	1	4	6	2	2	3	1
n	5	3	2	7	7	4	8	1	4

Figure 4. Example of a Genome Structure and Interpretation for a Single Rule Evolved Fuzzy System. Rule: if (v1 is Low) and (v3 is Low) and (v5 High) then (output is benign) else (output is malignant)

	V1	V2	V3	V4	V5	V6	V7	V8	V9
m	2	5	8	4	6	3	4	5	4
n	5	3	1	2	1	6	3	2	1

Figure 5. The best evolved fuzzy diagnostic system with three rules. It exhibits an overall classification rate of 97.33%. Rule 1 : if (v3 is Low) and (v7 is Low) and (v8 is Low) and (v9 is Low) then (output is benign); Rule 2 : if (v1 is Low) and (v2 is Low) and (v4 is Low)and (v5 is High) and (v9 is Low) then (output is benign); Rule 3 : if (v1 is Low) and (v4 is Low) and (v6 is Low) and (v8 is Low) then (output is benign)else (output is malignant)

	V1	V2	V3	V4	V5	V6	V7	V8	V9
m	1	1	3	8	6	2	1	3	
n	5	8	1	1	1	6	8	1	

Figure 6. The Best Two Rule Fuzzy Diagnostic System with Overall Classification Rate of 97.03. Rule1 : if (v2 is Low) and (v3 is Low) then (output is benign); Rule 2: if (v2 is Low) and (v5 is Low) and (v6 is Low)and (v8 is Low) then (output is benign)else (output is malignant)

	V1	V2	V3	V4	V5	V6	V7	V8	V9
m	4	2	5	6	6	2	4	3	6
n	3	5	3	1	2	3	3	1	5

Figure 7. The Best Fuzzy Diagnostic System with four Rules. It exhibits an overall classification rate of 96.67%. Rule 1: if (v3 is Low) and (v7 is Low) and (v8 is Low) and (v9 is Low) then (output is benign); Rule 2: if (v1 is Low) and (v2 is Low) and (v4 is High)and (v5 is High) and (v9 is Low) then (output is benign); Rule 3: if (v1 is Low) and (v7 is Low) and (v6 is Low)and (v8 is Low) then (output is benign); Rule 4: if (v3 is Low) and (v2 is Low) and (v4 is High) and (v9 is Low) then (output is benign)else (output is malignant)

Table 1. Parameter Encoding of Genome

	Values	Bits	Total bits
m	1 to 8	3	27
n	1 to 8	3	27
M	0 to 2	2	18*number of rules

Table 2. Results Divided According to the three Experimental Categories for the Best three Rules Diagnostic System

	Performance		
	Training set	Test set	Overall
Training/test 100/0	-	-	97.33%
Training/test 75/25	98.30%	96.21%	97.25%
Training/test 50/50	97.50%	96.61%	97.05%

Table 3. Comparing Over All Results for a three Rule Base System in Our Work and with other Approaches in the Literature with a Different Dataset and Similar Methodology

Research approach	This work	Andres et al.	Setiono
Performance	97.33 %	97.80 %	97.14 %

Table 4. Results of Overall Classification Performance for all Fuzzy Diagnostic Systems with Rules 1-4

Rules per-system	Best system (%)	Average (%)
1	96.19	96.8
2	97.03	96.7
3	97.33	97
4	96.67	96.7

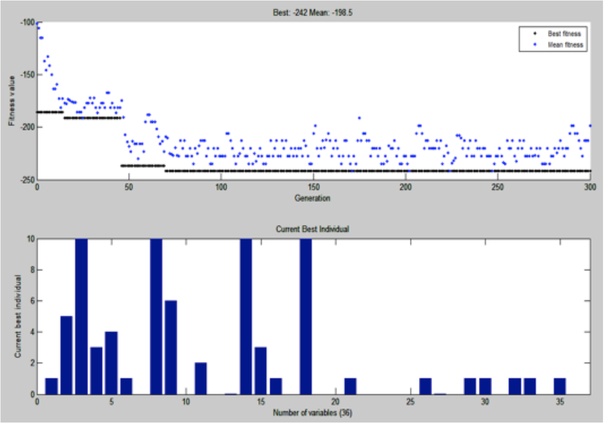

Figure 8. Plots of the Best Fitness Value over the Generations and the Current Best Individual of all 36 Variables in the Two Rule

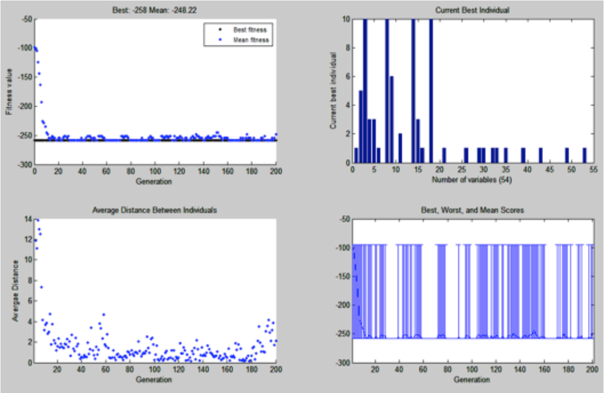

Figure 9. Plots of the Best Fitness Value over the Generations and the Current Best Individual, Average Distance between Individuals, and the Selection Function for Best Parent in a Four Rule Fuzzy Diagnostic System

	V1	V2	V3	V4	V5	V6	V7	V8	V9
m	4	2	5	6	6	2	4	3	6
n	3	5	3	1	2	3	3	1	5

Figure 10. The Best Fuzzy Diagnostic System with One Rule. It exhibits an overall classification rate of 96.19%. Rule 1 : if (v2 is Low) and (v5 is Low) and (v6 is Low) and (v8 is Low) then (output is benign) else (output is malignant)

Andres et al. had 699 cases in the WBCD dataset from patients in USA and they used a different fitness function denoted F=Fc-0.05Fv-0.01Fe, such that Fc, the number of correctly diagnosed cases, Fv measures the linguistic integrity (interpretability), and Fe adds selection pressure towards systems with low quadratic error. Moreover, Setiono used an application of the neural networks that involves Boolean rule bases extracted from trained neural networks on the WBCD dataset. As we can see in the results shown in Table 3, the classification performance values obtained by the two other papers with the similar methodology but different database, are looking very close in terms of accuracy and in time efficiency, even though in our work here on the Saudi database we used a simpler fitness function which depended only on the number of correctly classified cases and the confidence of the diagnosis and we only have 260 cases in our database.

Figure 6 shows a diagnostic system with two rules, which obtained 97% correct classification rate in the benign cases, 97.06% correct classification rate over the malignant cases, and an overall classification rate of 97.03%. Figure 7 gives the diagnostic system with four rules. It obtains 96.55% correct classification rate in the benign cases, 96.8% correct classification rate over the malignant cases, and an overall classification rate of 96.67%. Figure 8 shows the plot of the best fitness value over the generations and the current best individual of all 36 variables in a two rule fuzzy diagnostic system. Figure 9 shows the plots of the best in fitness and best individual for the four-rule system, also the distance between individuals. Finally, Figure 10 delineates the best one-rule system found through our evolutionary approach. It obtains 96.17% correct classification rate in the benign cases, 96.22% correct classification rate

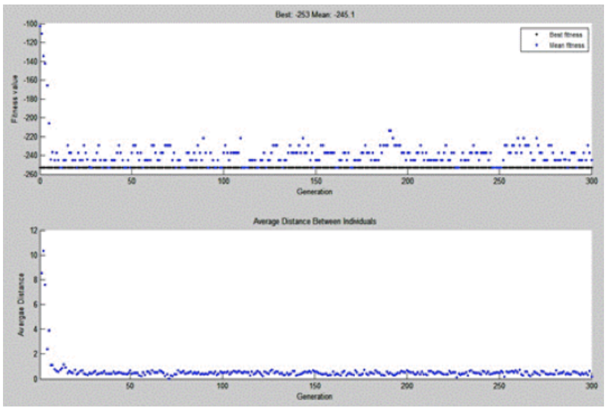

Figure 11. Plots of the Best Fitness Value over the Generations, and Average Distance Between Individuals for the Evolved Fuzzy One Rule Diagnostic System

over the malignant cases, and an overall classification rate of 96.19%. Figure 11 shows the best fitness and distance between individuals with this system. We have performed 40 evolutionary runs in every system, the results of which are summarized in Table 4. Results are divided into four classes, in accordance with the number of rules-per-system, with 10 runs per class; shown are the resulting best systems as well as the average per class. In this comparison, we can see the three-rule system achieves a higher percentage of correct diagnosed cases than the other systems.

After completing the fuzzification phase it is time for the inference engine to compute the truth value of each rule (see Figure 2), by applying the fuzzy 'and' operator to combine the antecedent clauses in a fuzzy manner. This results in the output truth-value, which is a continuous value representing the rule's degree of activation. Thus, a rule is not just activated, but it is activated to a certain degree represented by a value between 0 and 1. The inference engine now goes on to apply the aggregation operator and combining the continuous rule activation values to produce a fuzzy output with a set truth-value. The defuzzifier then works to produce the final continuous value of the fuzzy inference system; this latter value is the value that is passed on to the threshold unit. For our best three rule fuzzy system given in Figure 5 we calculate the membership values for each 260 patients and with the "and" function we get the appraisal value in the range [3,5]. We chose to place the threshold value at 3, with inferior values classified as benign, and superior values classified as malignant. Thus, if a case in the database scores a value of 2.6 and that is classified as benign, it is close to the threshold 3 so its confidence is considered low. This demonstrates a prime advantage of fuzzy systems, which is the ability to find an output not only in binary form: benign, malignant, but also with a measure representing the system's confidence in its output. Our three-rule system has computed intermediate values [2.5, 3.5] for only 23 cases; which means that in these cases the system is less confident about the output, but for the remaining 237 cases it has diagnosed it with high confidence. This is considered a very good diagnosis result compared with other computerized diagnosis systems published in previous work on the WCBD dataset.

Discussion

In this paper, we applied a combined genetic-fuzzy algorithm to the Saudi breast cancer diagnosis database. Our evolved computerized systems exhibit both high classification performances with a high confidence measure; and with a few simple rules, that are easily interpretable. Our results suggest that the genetic-fuzzy approach is highly effective on medical diagnosis problems; in fact, our best three rule fuzzy system calculated a diagnosis for all 260 patients with a 97.33% accuracy and a confidence of 91%. This demonstrates a prime advantage of fuzzy systems, which is the ability to find an output not only in binary form: benign, malignant, but also with a measure representing the system's confidence in the output.

Our future work will involve applying the genetic-fuzzy approach to other complex Cancer diagnosis problems such as Prostate or lung cancer diagnosis, which should help physicians detect it at an early stage. We will also try alternative fuzzy logic approaches such as Neuro-Fuzzy networks or Fuzzy Petri with evolutionary methods. In addition, we will explore another promising area combining evolutionary algorithms with neural networks such as adaptive neuro-fuzzy inference systems and evolved principle component analysis neural networks to develop other computerized diagnosis tools.

Acknowledgements

This research project was supported by a grant from the "Research Centre of the Female Scientific and Medical Colleges", Deanship of Scientific Research, King Saud University.

References

Alander JT (1997). An indexed bibliography of genetic algorithms with fuzzy logic Fuzzy. Fuzzy evolutionary computation. Springer, USA, 299-318.

AlDiab R, Qureshi S, AlSaleh KA, et al (2013). Studies on the methods of diagnosis and biomarkers used in the early detection of breast cancer in the kingdom of saudi Arabia. *World J Med Sci*, **5**, 72-88.

Al Diab R, Qureshi S, Khalid A, et al (2013). Review on breast cancer in the kingdom of Saudi Arabia. *Middle East J Scientific Res*, **14**, 532-43.

Alharbi A, Rand W, Rolio R, et al (2007), Understanding the semantics of genetic algorithms in dynamic environments; a case study using the shaky ladder hyperplane-defined functions, workshop on evolutionary algorithms in stochastic and dynamic environments, incorporated in evo conferences Valencia, Spain.

Andres C, Reyes P, Sipper M, et al (1999). A genetic-fuzzy approach to breast cancer diagnosis. Artificial Intelligence Med, **17**, 131-55.

Carmona J, Ruiz-Rodado V, del Jesus MJ, et al (2015). A fuzzy genetic programming-based algorithm for subgroup discovery and the application to one problem of pathogenesis of acute sore throat conditions in humans. *Informat Sci*,

298, 180-97.

Cordon O, Herrera F, Lozano M, et al (1997). On the combination of fuzzy logic and evolutionary computation: a short review and bibliography. *Fuzzy Evolutionary Computation*, **1**, 33-56.

Dennis B, Muthukrishnan S (2014). GFS: Adaptive Genetic Fuzzy System for medical data classification, Applied Soft Computing, Elsevier, **25**, 242-52.

El-Akkad SM, Amer M.H, Lin GS, et al (1986). Pattern of cancer in Saudi Arabia. *King Faisal Specialist Hospital Cancer J*, **58**, 1172-8.

Heider H, Drabe T,(1997).Fuzzy system design with a cascaded genetic algorithm. *IEEE Int Conference Evolutionary Computat*, **1**, 585-8.

Ferlay J, Soerjomataram I, Ervik M, et al (2012). cancer incidence and mortality worldwide: iarc cancer base no. 11, lyon, france: international agency for research on cancer; 2013.

Herrera F, Lozano M, Verdegay JL, et al (1995). Generating fuzzy rules from examples using genetic algorithms. Fuzzy Logic and Soft Computing. *World Scientific*, **1**, 11-20.

Jang JR, Sun CT (1995). Neuro-fuzzy modeling and control. *Proceedings of the IEEE*, **83**, 378-406.

Karr CL (1991). Genetic algorithms for fuzzy controllers, *A I Expert*, **6**, 26-33.

Kovalerchuk B, Triantaphyllou E, Ruiz JF, et al (1997). Fuzzy logic in computer-aided breast cancer diagnosis. *Artificial Intelligence Medical*, **11**, 75-85.

Koza J R, (1992), Genetic Programming, USA, MIT Press.

Lee M A, Takagi H, (1993). Integrating design stages of fuzzy systems using genetic algorithms. *IEEE International Conference on Fuzzy Systems*, **1**, 612-7.

Mangasarian OL, Street WN, Wolberg WH, et al (1994). Breast cancer diagnosis and prognosis via linear programming. *Mathematical Programming Technical Report*, **94**, 94-10.

Matlab Tool Box Guide retrieved Jan 2015 from http://www.mathworks.com/products/global-optimization/features.html#genetic-algorithm-solver.

Mendel J M, (1995). Fuzzy logic systems for engineering: a tutorial. *Proceedings of the IEEE*, **83**, 345-377.

Merz CJ, Murphy PM (1996). UCI repository of machine learning-databases. http://www.ics.uci.edu/~mlearn/MLR repository.

Michalewicz Z, (1996).Genetic Algorithms Data Structures, Evolution Programs, 3rd edition, Berlin, Springer-Verlag.

Nguyen T, Khosravi A, Creighton D, et al (2015). Classification of healthcare data using genetic fuzzy logic system and wavelets. *Expert Systems with Applications*, **42**, 2184-97.

Rashidi M M, Anwar O, Beg A, Basiriparsa, Nazari F, et al (2011). Analysis and optimization of a trans critical power cycle with regenerator using artificial neural networks and genetic algorithms, proceedings of the institution of mechanical engineers. Part A: *J Power Energy*, **225**, 701-17.

Setiono R, (1996). Extracting rules from pruned neural networks for breast cancer diagnosis. Artificial Intelligence in Medicine, 8, 37-51.

Tchier F (2014). Relational demonic fuzzy refinement. *J Applied Mathematics*, **2014**, 1-17.

Tchier F (2013). Fuzzy demonic refinement. international conference on basic and applied sciences regional annual fundamental science symposium 2013, Johor, Malaysia.

Vuorimaa P (1994). Fuzzy self-organizing map. *Fuzzy Sets Systems*, **66**, 223-31.

Yager R R, Filev D P, (1994). Essentials of Fuzzy Modeling and Control, Canada, Wiley.

Yager RR, Zadeh LA (1994). Fuzzy sets neural networks and soft computing. New York, Van Nostrand Reinhold.

Zadeh LA (1965). Fuzzy sets. *Information Control*, **8**, 338-53.