Designing Breast Cancer Diagnostic Systems via a Hybrid Fuzzy-Genetic Methodology

Carlos Andrés Peña-Reyes and Moshe Sipper*

January 7, 1999

Abstract

The automatic diagnosis of breast cancer is an important, real-world medical problem. In this paper we focus on the Wisconsin breast cancer diagnosis (WBCD) problem, combining two methodologies—fuzzy systems and evolutionary algorithms—so as to automatically produce diagnostic systems. We find that our fuzzy-genetic approach produces systems exhibiting the highest classification performance shown to date, and which are also (human-)interpretable.
Keywords: Fuzzy systems; Genetic algorithms; Breast cancer diagnosis

1 Introduction

A major class of problems in medical science involves the diagnosis of disease, based upon various tests performed upon the patient. When several tests are involved, the ultimate diagnosis may be difficult to obtain, even for a medical expert. This has given rise, over the past few decades, to computerized diagnostic tools, intended to aid the physician in making sense out of the welter of data. A prime target for such computerized tools is in the domain of cancer diagnosis. Specifically, where breast cancer is concerned, the treating physician is interested in ascertaining whether the patient under examination exhibits the symptoms of a benign case, or whether her case is a malignant one.

A good computerized diagnostic tool should possess two characteristics, which are often in conflict. First, the tool must attain the highest possible *performance*, i.e., diagnose the presented cases correctly as being either *benign* or *malignant*. Moreover, it would be highly desirable to be in possession of a so-called *degree of confidence*: the system not only provides a binary diagnosis (benign or malignant), but also outputs a numeric value which represents the degree to which the system is confident about its response. Second, it would be highly beneficial for such a diagnostic system to be human-friendly, exhibiting so-called *interpretability*. This means that the physician is not faced with a black box that simply spouts answers (albeit correct) with no explanation; rather, we would like for the system to provide some insight as to *how* it derives its outputs.

In this paper we combine two methodologies—fuzzy systems and evolutionary algorithms—so as to automatically produce systems for breast cancer diagnosis. The major advantage of fuzzy systems is that they favor interpretability, however, finding good fuzzy systems can be quite an arduous task. This is where evolutionary algorithms step in, enabling the automatic production of fuzzy systems, based on a database of training cases.

Fuzzy modeling is the task of identifying the parameters of a fuzzy inference system so that a desired behavior is attained. The parameters of fuzzy inference systems can be classified into four categories (Table 1): logical, structural, connective, and operational. Generally speaking, this order also represents their relative influence on performance, from most influential (logical) to least influential (operational).

Evolutionary algorithms are used to search large, and often complex, search spaces. They have proven worthwhile on numerous diverse problems, able to find near-optimal solutions given an adequate performance (fitness) measure. Fuzzy modeling can be considered as an optimization process where part or all of the parameters of a fuzzy system constitute the search space. Works investigating the application of evolutionary techniques in the domain of fuzzy modeling had first appeared about a decade ago, focusing mainly on control problems. Evolutionary fuzzy modeling has since been applied in an ever-growing number of domains, branching into areas as diverse as chemistry, medicine, telecommunications, biology, and geophysics.

*The authors are with the Logic Systems Laboratory, Swiss Federal Institute of Technology, IN-Ecublens, CH-1015 Lausanne, Switzerland (e-mail: Carlos.Pena@di.epfl.ch, Moshe.Sipper@epfl.ch).

Table 1: Parameter classification of fuzzy inference systems.

Class	Parameters
Logical	Reasoning mechanism
	Fuzzy operators
	Membership function types
	Defuzzification method
Structural	Relevant variables
	Number of membership functions
	Number of rules
Connective	Antecedents of rules
	Consequents of rules
	Rule weights
Operational	Membership function values

Both connective and structural parameters modeling can be viewed as rule base learning processes with different levels of complexity. In the evolutionary algorithm community there are two major approaches for evolving such rule systems: the Michigan approach and the Pittsburgh approach [1]. A more recent method has been proposed specifically for fuzzy modeling: the iterative rule learning approach [2].

In Section 2 we describe the Wisconsin breast cancer diagnosis (WBCD) problem, which is the focus of our interest in this paper. Section 3 then describes our particular evolutionary approach to the WBCD problem. In Section 4 we delineate our results, followed by concluding remarks in Section 5.

2 The Wisconsin breast cancer diagnosis problem

In this section we present the medical diagnosis problem which is the object of our study. Breast cancer is the most common cancer among women, excluding skin cancer. The presence of a breast mass[1] is an alert sign, 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 Wisconsin breast cancer diagnosis (WBCD) database [3] is the result of the efforts made at the University of Wisconsin Hospital for accurately diagnosing breast masses based solely on an FNA test [4]. Nine visually assessed characteristics of an FNA sample considered relevant for diagnosis were identified, and assigned an integer value between 1 and 10. The measured variables are as follows:

1. Clump Thickness (v_1);
2. Uniformity of Cell Size (v_2);
3. Uniformity of Cell Shape (v_3);
4. Marginal Adhesion (v_4);
5. Single Epithelial Cell Size (v_5);
6. Bare Nuclei (v_6);
7. Bland Chromatin (v_7);
8. Normal Nucleoli (v_8);
9. Mitosis (v_9).

The diagnostics in the WBCD database were furnished by specialists in the field. The database itself contains 683 cases, with each entry representing the classification for a certain ensemble of measured values:

case	v_1	v_2	v_3	$\cdots$	v_9	*diagnostic*
1	5	1	1	$\cdots$	1	*benign*
2	5	4	4	$\cdots$	1	*benign*
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
683	4	8	8	$\cdots$	1	*malignant*

[1] Most breast cancers are detected as a lump or mass on the breast, by self-examination, by mammography, or by both.

[2] Fine needle aspiration is an outpatient procedure that involves using a small-gauge needle to extract fluid directly from a breast mass.

Note that the diagnostics do not provide any information about the degree of benignity or malignancy.

There are several studies based on this database. Bennet and Mangasarian [5] used linear programming techniques, obtaining a 99.6% classification rate on 487 cases (the reduced database available at the time). However, their solution exhibits little understandability, i.e., diagnostic decisions are essentially black boxes, with no explanation as to how they were attained. With increased interpretability in mind as a prime objective, a number of researchers have applied the method of extracting Boolean rules from neural networks [6–8]. Their results are encouraging, exhibiting both good performance and a reduced number of rules and relevant input variables. Nevertheless, these systems use Boolean rules and are not capable of furnishing the user with a measure of confidence for the decision made. Our preliminary work on the evolution of fuzzy rules showed that it is possible to obtain high performance, coupled with interpretability and a confidence measure [9].

3 Evolving fuzzy systems for the WBCD problem

The solution scheme we propose for the WBCD problem is depicted in Figure 1. It 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.

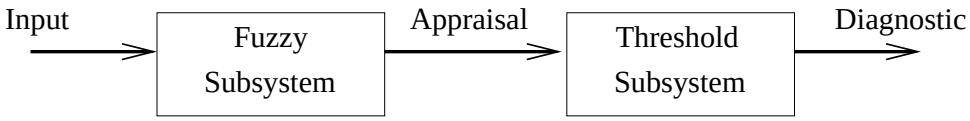

Figure 1: Proposed diagnosis system.

In order to evolve the fuzzy model we must make some preliminary decisions about the fuzzy system itself and about the genetic algorithm encoding. In this section we describe our choices, followed in the next section by a presentation of our results.

3.1 Fuzzy system parameters

Previous knowledge about the WBCD problem and about some of the extant rule-based models represents valuable information to be used for our choice of fuzzy parameters. When defining our setup we took into consideration the following three results, described in previous works: (1) small number of rules [6,9]; (2) small number of variables [7–9]; and (3) monotonicity of the input variables [9].

Some fuzzy models forgo interpretability in the interest of improved performance. Where medical diagnosis is concerned, interpretability—also called linguistic integrity—is the major advantage of fuzzy systems. This motivated us to take into account the following five semantic criteria, defining constraints on the fuzzy parameters [10,11]: (1) distinguishability; (2) justifiable number of elements; (3) coverage; (4) normalization; and (5) orthogonality.

Referring to Table 1, and taking into account the above criteria, we delineate below the fuzzy system setup:

- Logical parameters: singleton-type fuzzy systems; min-max fuzzy operators; orthogonal, trapezoidal input membership functions; weighted-average defuzzification.
- Structural parameters: two input membership functions (*Low* and *High*); two output singletons (*benign* and *malignant*); a user-configurable number of rules (based on our previous results [9], we limited the number of rules to be between 1–5). The relevant variables are one of the genetic algorithm's objectives.
- Connective parameters: the antecedents of the rules are searched by the genetic algorithm. The algorithm finds rules for the *benign* diagnostic; the *malignant* diagnostic is an `else` condition. Active rules have a weight of value 1, and the `else` condition has a weight of 0.25.
- Operational parameters: the input membership function values are to be found by the genetic algorithm. For the output singletons we used the values 2 and 4, for *benign* and *malignant*, respectively.

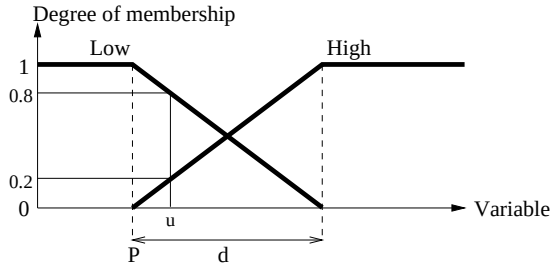

Figure 2: Example of a fuzzy variable with two possible fuzzy values labeled **Low** and **High**, and orthogonal membership functions, plotted above as degree of membership versus input values. P and d define the start point and the length of membership function edges, respectively. The orthogonality condition means that the sum of all membership functions at any point is one. In the figure, an example value u is assigned the membership values $\mu_{Low}(u) = 0.8$ and $\mu_{High}(u) = 0.2$ (as can be seen $\mu_{Low}(u) + \mu_{High}(u) = 1$).

3.2 The genetic algorithm

We apply Pittsburgh-style structure learning [1,12], using a genetic algorithm to search for three parameters. The genome, encoding relevant variables, input membership function values, and antecedents of rules, is constructed as follows:

- Membership function parameters. There are nine variables (v_1 – v_9), each with two parameters P and d, defining the start point and the length of the membership function edges, respectively (Figure 2).
- Antecedents. The i-th rule has the form:

  if (v_1 **is** A_1^i) **and ... and** (v_9 **is** A_9^i) **then** (*output* **is** *benign*),

  where A_j^i represents the membership function applicable to variable v_j. A_j^i can take on the values: 1 (*Low*), 2 (*High*), or 0 or 3 (*Other*).
- Relevant variables are searched for implicitly by letting the algorithm choose non-existent membership functions as valid antecedents; in such a case the respective variable is considered irrelevant.

Table 2 delineates the parameters encoding, which together form a single individual's genome. Figure 3 shows a sample genome.

Table 2: Parameters encoding of an individual's genome. Total genome length is $54 + 18N_r$, where N_r denotes the number of rules.

Parameter	Values	Bits	Quantity	Total bits
P	[1-8]	3	9	27
d	[1-8]	3	9	27
A	[0-3]	2	$9N_r$	$18N_r$

To evolve the fuzzy inference system, we applied a standard genetic algorithm [1] with a fixed population size of 200 individuals, and fitness-proportionate selection. The algorithm terminates when the maximum number of generations, G_{max}, is reached (we set $G_{max} = 2000 + 500 \times N_r$, i.e., dependent on the number of rules used in the run), or when the increase in fitness of the best individual over five successive generations falls below a certain threshold (in our experiments we used threshold values between 2×10^{-7} and 4×10^{-6}).

Our fitness function combines three criteria: (1) F_c: classification performance, computed as the percentage of cases correctly classified; (2) F_e: the quadratic difference between the continuous appraisal value (in the range $[2, 4]$) and the correct discrete diagnosis given by the WBCD database (either 2 or 4); and (3) F_v: the average number of variables per active rule. The fitness function is given by $F = F_c - \alpha F_v - \beta F_e$, where $\alpha = 0.05$ and $\beta = 0.01$ (these latter values were derived empirically). F_c, the percentage of correctly diagnosed cases, is the most important measure of performance. F_v measures

P_1	d_1	P_2	d_2	P_3	d_3	P_4	d_4	P_5	d_5	P_6	d_6	P_7	d_7	P_8	d_8	P_9	d_9	...
4	3	1	5	2	7	1	1	1	6	3	7	4	6	7	1	1	5	

...	A_1^1	A_2^1	A_3^1	A_4^1	A_5^1	A_6^1	A_7^1	A_8^1	A_9^1
	0	1	3	3	2	3	1	3	1

(a)

Database

	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9
P	4	1	2	1	1	3	4	7	1
d	3	5	7	1	6	7	6	1	5

Rule base

Rule 1 **if** (v_2 **is** *Low*) **and** (v_5 **is** *High*) **and** (v_7 **is** *Low*) **and** (v_9 **is** *Low*) **then** (*output* **is** *benign*)

Default **else** (*output* **is** *malignant*)

(b)

Figure 3: Example of a genome for a single-rule system. (a) Genome encoding. The first 18 positions encode the parameters P and d for the nine variables v_1–v_9. The rest encode the membership functions applicable to the nine antecedents of each rule. (b) Interpretation. Database and rule base of the single-rule system encoded by (a). The parameters P and d are interpreted as illustrated in Figure 2.

the linguistic integrity (interpretability), penalizing systems with a large number of variables per rule (on average). F_e adds selection pressure towards systems with low quadratic error.

4 Results

This section describes selected results obtained when applying the methodology described in Section 3 (the full description is given in [12]). We first delineate in Subsection 4.1 the success statistics relating to the evolutionary algorithm. Then, in Subsection 4.2, we describe in full two evolved fuzzy systems that exemplify our approach. Finally, in Subsection 4.3, we discuss the issue of obtaining a confidence measure of the system's output, going beyond a mere binary, benign-malignant classification.

4.1 The genetic algorithm...

A total of 120 evolutionary runs were performed, all of which found systems whose classification performance exceeds 94.5%. In particular, considering the best individual per run (i.e., the evolved system with the highest classification success rate), 78 runs led to a fuzzy system whose performance exceeds 96.5%, and of these, 8 runs found systems whose performance exceeds 97.5%.

Table 3 compares our best systems with the top systems obtained by four other rule-based diagnostic approaches. The first three approaches—those of Setiono [6], Setiono and Liu [7], and Taha and Ghosh [8]—involve Boolean rule bases extracted from trained neural networks; the last approach is our own previous work [9]. The evolved fuzzy systems described in this paper can be seen to surpass those obtained by these four previous approaches in terms of both performance and simplicity of rules. As shown in Table 3, we obtained the highest-performance systems for all five rule-base sizes, i.e., from one-rule systems all the way up to five-rule systems. Not only is high performance exhibited, but, moreover, our fuzzy approach enables the introduction of a confidence measure of the diagnostic decision (see Subsection 4.3). In contrast, the Boolean rule-based systems [6–8] provide but a single binary value, indicating whether the case in question is benign or malignant. Compared with our previous work [9], the current approach not only improves performance, but also obtains systems with less antecedents per rule (which are thus more easily comprehensible).

4.2 ...and the fuzzy systems it discovered

We next describe two of our top-performance systems, which serve to exemplify the solutions found by our evolutionary approach. The first system, delineated in Figure 4, consists of three rules (note that the `else` condition is not counted as an active rule). Taking into account all three criteria of performance—classification rate, number of rules per system, and average number of variables per rule—this system

Table 3: Comparison of the best systems evolved by our approach with the top systems obtained by four other rule-based diagnostic approaches. The first three approaches—those of Setiono [6], Setiono and Liu [7], and Taha and Ghosh [8]—involve Boolean rule bases extracted from trained neural networks; the last approach is our own previous work [9]. Shown below are the classification performance values of the top systems obtained by these approaches, along with the average number of variables-per-rule given in parentheses. Results are divided into five classes, in accordance with the number of rules-per-system, going from one-rule systems to five-rule ones.

Rules-per-system	Setiono [6]	Setiono and Liu [7]	Taha and Ghosh [8]	Peña and Sipper [9]	This work
1	95.42% (2)	–	–	96.35% (3)	97.07% (4)
2	–	–	–	96.65% (7)	97.36% (3)
3	97.14% (4)	97.21% (4)	–	–	97.80% (4.7)
4	–	–	–	–	97.80% (4.8)
5	–	–	96.19% (1.8)[a]	–	97.51% (3.4)

[a] Note that Taha and Ghosh [8] obtained slightly better results for the five-rules case by directly using their trained neural networks, rather than the extracted rule-based systems. Herein, our interest lies with these latter, rule-based systems.

Database

	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9
P	3	5	2	2	8	1	4	5	4
d	5	2	1	2	4	7	3	5	2

Rule base

Rule 1	**if** (v_3 **is** *Low*) **and** (v_7 **is** *Low*) **and** (v_8 **is** *Low*) **and** (v_9 **is** *Low*) **then** (*output* **is** *benign*)
Rule 2	**if** (v_1 **is** *Low*) **and** (v_2 **is** *Low*) **and** (v_3 **is** *High*) **and** (v_4 **is** *Low*) **and** (v_5 **is** *High*) **and** (v_9 **is** *Low*) **then** (*output* **is** *benign*)
Rule 3	**if** (v_1 **is** *Low*) **and** (v_4 **is** *Low*) **and** (v_6 **is** *Low*) **and** (v_8 **is** *Low*) **then** (*output* **is** *benign*)
Default	**else** (*output* **is** *malignant*)

Figure 4: The best evolved, fuzzy diagnostic system with three rules. It exhibits an overall classification rate of 97.8%, and an average of 4.7 variables per rule.

can be considered the top one over all 120 evolutionary runs. It obtains an overall classification rate (i.e., over the entire database) of 97.8%.

A thorough test of this three-rule system revealed that the second rule (Figure 4) never fires, i.e., it is triggered by none of the input cases. Thus, it can be eliminated altogether from the rule base, resulting in a two-rule system (also reducing the average number of variables-per-rule from 4.7 to 4).

Finally, Figure 5 delineates the best one-rule system found through our evolutionary approach. It obtains an overall classification rate of 97.07%.

4.3 Diagnostic confidence

Up until now we have been using the evolved fuzzy systems to ultimately produce a binary classification value—*benign* or *malignant*—with no finer gradations. Going back to Figure 1, we note that the diagnostic system comprises in fact two subsystems: the first subsystem consists of the (evolved) fuzzy system, which, upon presentation of an input (in our case, a WBCD database entry) proceeds to produce a *continuous* appraisal value; this value is then passed along to the second subsystem—the threshold unit—which produces the final binary output (*benign* or *malignant*). The first subsystem (the fuzzy system) is the one evolved in our approach. The threshold subsystem simply outputs *malignant* if the appraisal value is below a fixed threshold value, and outputs *benign* otherwise. The threshold value is assigned by the user based on knowledge of the problem at hand.

The appraisal value can accompany the final output of the diagnostic system, serving as a confidence measure. This demonstrates a prime advantage of fuzzy systems, namely, the ability to output not only a (binary) classification, but also a measure representing the system's confidence in its output. For example,

Database

	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9
P	4	4				2		2	
d	3	1				5		7	

Rule base

Rule 1 **if** (v_1 **is** *Low*) **and** (v_2 **is** *Low*) **and** (v_6 **is** *Low*) **and** (v_8 **is** *Low*) **then** (*output* **is** *benign*)

Default **else** (*output* **is** *malignant*)

Figure 5: The best evolved, fuzzy diagnostic system with one rule. It exhibits an overall classification rate of 97.07%, and a rule with 4 variables.

the three-rule system of Figure 4 computes intermediate appraisal values (between, say, 2.4 and 3.6) for 39 cases; these might thus be considered the cases for which we are somewhat less confident about the output.

5 Concluding remarks

In this paper we applied a combined fuzzy-genetic approach to the Wisconsin breast cancer diagnosis problem. Our evolved systems exhibit both characteristics outlined in Section 1: first, they attain high classification performance (the best shown to date), with the possibility of attributing a confidence measure to the output diagnosis; second, the resulting systems involve a few simple rules, and are therefore interpretable.

Our experience to date suggests that the fuzzy-genetic approach is highly effective where such medical diagnosis problems are concerned. We are currently pursuing two avenues of research: (1) application of the fuzzy-genetic approach to more complex diagnosis problems; and (2) improving and expanding upon the methodology presented herein (e.g., by making use of recent advances in evolutionary computation). Our underlying goal is to provide an approach for automatically producing high-performance, interpretable systems for real-world diagnosis problems.

References

[1] Z. Michalewicz, *Genetic Algorithms + Data Structures = Evolution Programs*, Springer-Verlag, Heidelberg, third edition, 1996.

[2] F. Herrera, M. Lozano, and J. L. Verdegay, "Generating fuzzy rules from examples using genetic algorithms," in *Fuzzy Logic and Soft Computing*, B. Bouchon-Meunier, R. R. Yager, and L. A. Zadeh, Eds., pp. 11–20. World Scientific, 1995.

[3] C. J. Merz and P. M. Murphy, "UCI repository of machine learning databases," http://www.ics.uci.edu/~mlearn/MLRepository.html, 1996.

[4] O. L. Mangasarian, R. Setiono, and W.-H Goldberg, "Pattern recognition via linear programming: Theory and application to medical diagnosis," in *Large-Scale Numerical Optimization*, T. F. Coleman and Y. Li, Eds., pp. 22–31. SIAM, 1990.

[5] K. P. Bennett and O. L. Mangasarian, "Neural network training via linear programming," in *Advances in Optimization and Parallel Computing*, P. M. Pardalos, Ed., pp. 56–57. Elsevier Science, 1992.

[6] R. Setiono, "Extracting rules from pruned neural networks for breast cancer diagnosis," *Artificial Intelligence in Medicine*, pp. 37–51, 1996.

[7] R. Setiono and H. Liu, "Symbolic representation of neural networks," *Computer*, vol. 29, no. 3, pp. 71–77, March 1996.

[8] I. Taha and J. Ghosh, "Evaluation and ordering of rules extracted from feedforward networks," in *Proceedings of the IEEE International Conference on Neural Networks*, 1997, pp. 221–226.

[9] C. A. Peña-Reyes and M. Sipper, "Evolving fuzzy rules for breast cancer diagnosis," in *Proceedings of 1998 International Symposium on Nonlinear Theory and Applications (NOLTA'98)*, Lausanne, 1998, vol. 2, pp. 369–372, Presses Polytechniques et Universitaires Romandes.

[10] J. J. Espinosa and J. Vandewalle, "Constructing fuzzy models with linguistic integrity," Submitted to IEEE Transactions on Fuzzy Systems.

[11] W. Pedrycz and J. Valente de Oliveira, "Optimization of fuzzy models," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 26, no. 4, pp. 627–636, August 1996.

[12] C. A. Peña-Reyes and M. Sipper, "A fuzzy-genetic approach to breast cancer diagnosis," Submitted to Artificial Intelligence in Medicine.