BioMedical Engineering OnLine

Fuzzy method for pre-diagnosis of breast cancer from the Fine Needle Aspirate analysis

Gláucia RMA Sizilio[1*], Cicília RM Leite[2], Ana MG Guerreiro[1] and Adrião D Dória Neto[1]

* Correspondence: glauciasizilio@gmail.com
[1]Department of Computer Engineering and Automation - DCA and Department of Biomedical Engineering, Federal University of Rio Grande do Norte (UFRN), Caixa Postal 1524 – Campus Universitário - UFRN/CT/DCA, Rio Grande do Norte, CEP 59072-970, Brazil
Full list of author information is available at the end of the article

Abstract

Background: Across the globe, breast cancer is one of the leading causes of death among women and, currently, Fine Needle Aspirate (FNA) with visual interpretation is the easiest and fastest biopsy technique for the diagnosis of this deadly disease. Unfortunately, the ability of this method to diagnose cancer correctly when the disease is present varies greatly, from 65% to 98%. This article introduces a method to assist in the diagnosis and second opinion of breast cancer from the analysis of descriptors extracted from smears of breast mass obtained by FNA, with the use of computational intelligence resources - in this case, fuzzy logic.

Methods: For data acquisition of FNA, the Wisconsin Diagnostic Breast Cancer Data (WDBC), from the University of California at Irvine (UCI) Machine Learning Repository, available on the internet through the UCI domain was used. The knowledge acquisition process was carried out by the extraction and analysis of numerical data of the WDBC and by interviews and discussions with medical experts. The PDM-FNA-Fuzzy was developed in four steps: 1) Fuzzification Stage; 2) Rules Base; 3) Inference Stage; and 4) Defuzzification Stage. Performance cross-validation was used in the tests, with three databases with gold pattern clinical cases randomly extracted from the WDBC. The final validation was held by medical specialists in pathology, mastology and general practice, and with gold pattern clinical cases, i.e. with known and clinically confirmed diagnosis.

Results: The Fuzzy Method developed provides breast cancer pre-diagnosis with 98.59% sensitivity (correct pre-diagnosis of malignancies); and 85.43% specificity (correct pre-diagnosis of benign cases). Due to the high sensitivity presented, these results are considered satisfactory, both by the opinion of medical specialists in the aforementioned areas and by comparison with other studies involving breast cancer diagnosis using FNA.

Conclusions: This paper presents an intelligent method to assist in the diagnosis and second opinion of breast cancer, using a fuzzy method capable of processing and sorting data extracted from smears of breast mass obtained by FNA, with satisfactory levels of sensitivity and specificity. The main contribution of the proposed method is the reduction of the variation hit of malignant cases when compared to visual interpretation currently applied in the diagnosis by FNA. While the MPD-FNA-Fuzzy features stable sensitivity at 98.59%, visual interpretation diagnosis provides a sensitivity variation from 65% to 98% (this track showing sensitivity levels below those considered satisfactory by medical specialists). Note that this method will be used in an Intelligent Virtual Environment to assist the decision-making (IVEMI), which amplifies its contribution.

Keywords: Computational intelligence, Fuzzy logic, Fine needle aspirate, Decision support system, Breast cancer diagnosis, Telediagnosis

Background

Breast cancer is one of the leading causes of death among women worldwide and it is confirmed that early detection and accurate diagnosis of this disease can ensure long-term patient survival [1]. According to the World Health Organisation [2], about one third of the costs of cancer treatment can be reduced if cases are detected and treated early.

On the other hand, aiming to provide greater security, reliability and robustness to services and procedures, mainly when dealing with human lives, healthcare processes are increasingly becoming computerized. A growing area of research relates to the use of techniques from Computational Intelligence (CI) applied to the processing of information necessary for the medical diagnosis. We can cite as examples, [3-15].

This paper presents a method to assist in breast cancer diagnosis from the analysis of descriptors extracted from smears of breast mass obtained by FNA (Fine Needle Aspirate), incorporating features of computational intelligence (in this case, fuzzy logic) and inserted into a collaborative telediagnosis environment (called IVEMI - [16]).

Diagnosis of breast cancer and FNA

A carcinogen breast tumor is a breast mass that is growing abnormally and uncontrolled. There are three popular methods for breast cancer diagnosis: mammography; FNA with visual interpretation; and surgical biopsy [17]. The ability of these methods to diagnose cancer correctly when the disease is present is: mammogram - from 68% to 79%; FNA with visual-interpretation - from 65% to 98%; and surgical biopsy - 100%. [18]. It is noted that: mammography lacks sensitivity; the sensitivity of FNA with visual interpretation varies greatly (as a result of the visual interpretation); and although surgical biopsy is accurate it is also a very intrusive, time-consuming and expensive method [19].

FNA, which has been widely accepted in the approach to investigating mammary lesions, is the easiest and fastest biopsy technique to be performed, being a percutaneous procedure (through the skin) in which the specialist physician uses a thin needle (which varies from 0.6 to 0.8 mm) and a syringe to take samples of fluid from a breast cyst or remove clusters of cells in a solid mass. The needle is inserted into the skin toward the lesion, with the objective of collecting cells for further evaluation of their morphology, quantity and distribution through cytological examination.

The genetic material extracted from the breast by FNA is usually sent to a Pathology laboratory for examination by pathologists (doctors specialized in disease diagnosis through lab testing), who perform the analysis identifying the cells' characteristics from observing, under a microscope, smears made with this material on sheets of glass and stained using special techniques.

Computational intelligence and fuzzy logic

Computational Intelligence (CI) enables, through intelligent techniques some of them inspired by nature, the development of intelligent systems that imitate aspects of human behaviour, such as: learning, perception, reasoning, evolution and adaptation [20]. Some examples of Computational Intelligence techniques are: Artificial Neural Networks, biological neuron-inspired technique [14,15]; Evolutionary Computation, inspired by biological evolution [12]; Expert Systems, inspired by inference process [11]; and Fuzzy Logic, inspired by language processing.

The fuzzy systems theory is a formal approach that aims to address the modelling, representation, reasoning and the inaccurate information procedure as a troubleshooting strategy [21].

Introduced in 1965 [22], the fuzzy set theory is a tool to model the imprecision and ambiguity that arises in complex systems [22,23], and it was created from the combination of the concepts of classical logic and groupings of Łukasiewicz [24] defining degrees of relevance.

A fuzzy set differs from a classic set to assign to each element a value in the unit interval [0, 1]. Specifically, a fuzzy set is defined as a function A of a set x, called universe of discourse, to [0, 1]. The function A is referred to as a membership function, and the value A (x) represents the degree of relevance – or compatibility – of the element x with the concept represented by all the fuzzy set. Thus, the fuzzy logic proposed by Zadeh [22,23] provides a mathematical model for the processing of inaccurate or vague information and concepts, intending to make computers carry out inferences as people.

The fuzzy processing is generally composed of: Rules Base (provided by specialists or extracted from numerical data); Fuzzification Stage (it activates the rules from a set of precise entries); Inference Stage (determines how rules are enabled); Defuzzification Stage (it provides precise output, generating a fuzzy set of output), as illustrated in Figure 1.

Methods

This fuzzy method that assists in the diagnosis and second opinion of breast cancer, called Pre-Diagnosis Module FNA-Fuzzy (PDM-FNA-Fuzzy), was developed through
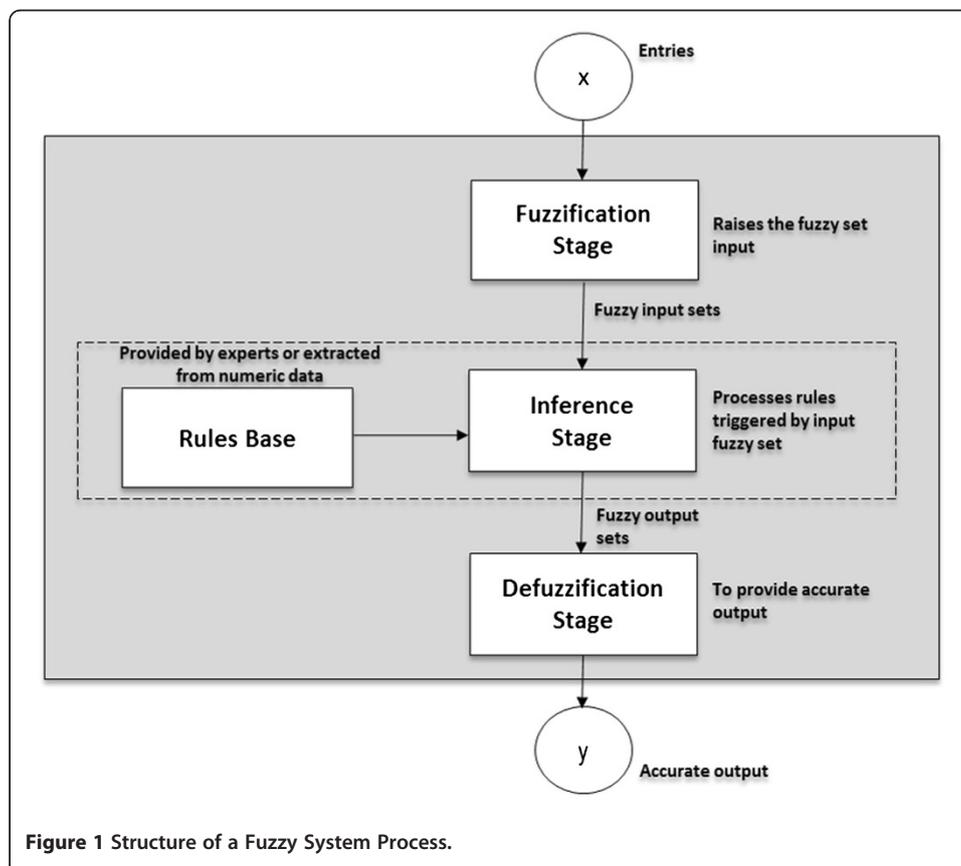

Figure 1 Structure of a Fuzzy System Process.

the analysis of extracted smears from breast mass obtained by FNA. The PDM-FNA-Fuzzy is inserted into the Intelligent Virtual Environment of Medical Interaction (IVEMI), which is a virtual environment (architecture shown in Figure 2) from which the doctor responsible can trigger, requesting pre-diagnosis or second opinion on clinical cases with suspected breast cancer and whose patient has undergone the examination of FNA.

Data acquisition

For FNA data acquisition, the Wisconsin Diagnostic Breast Cancer Data (WDBC), of UCI Machine Learning Repository, available on the internet by the domain of University of California Irvine [25] was used. The WDBC is a public database, consisting of a gold pattern data set[1], i.e. with confirmation of malignant and benign diagnosis.

The selected database, WDBC, was created in 1993, and presents 569 records of patients with known diagnosis (357 cases being benign and 212 cases malignant) and uses material (smears) collected by FNA, transformed into a digital image from which the main parameters (descriptors) were extracted.

For viewing and manipulating data from WDBC we used MATLAB (MathWorks – student version).

Pre-processing of data

Samples (small droplets of viscous liquid) were collected from aspiration of breast mass with thin needle, spread on glass blades slides, stained (aiming to highlight the cell nuclei) and digitized [26]. Examples of captured images are presented in Figure 3. Then, in the process of digital evaluation performed by Street et al. [26], the exact location of
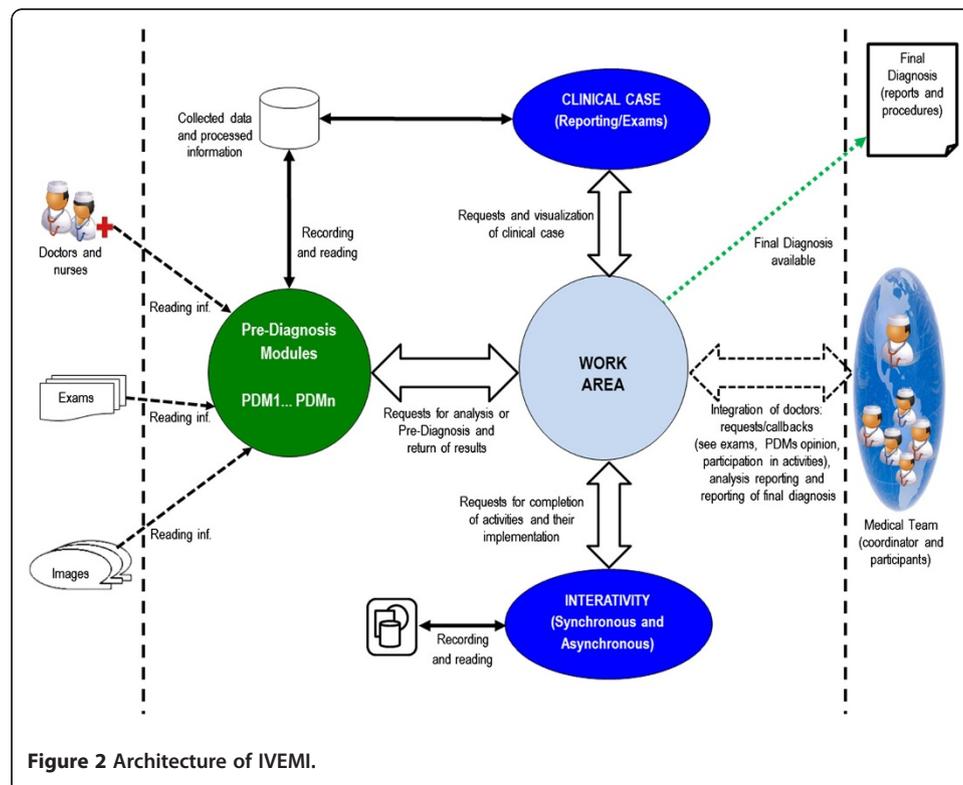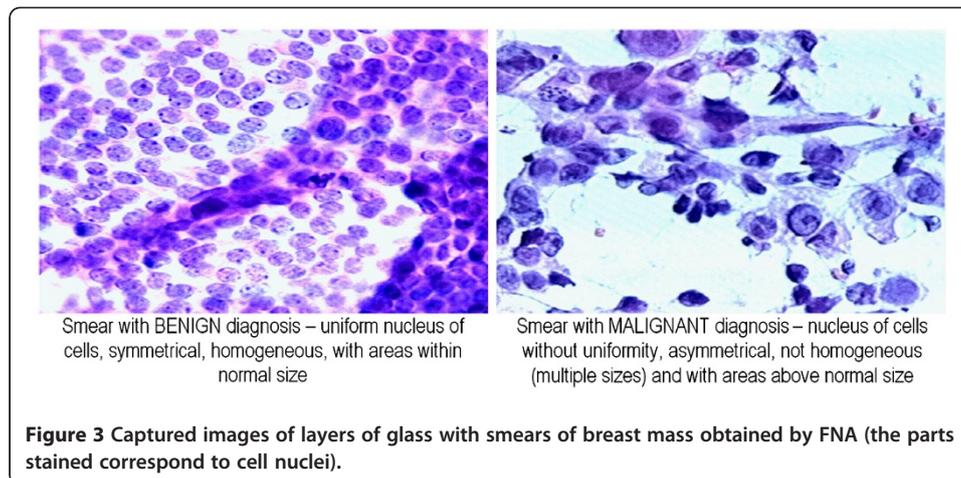

Figure 2 Architecture of IVEMI.

Figure 3 Captured images of layers of glass with smears of breast mass obtained by FNA (the parts stained correspond to cell nuclei).

each cell nucleus was specified, and the morphometric analysis of the cell nuclei, extracting characteristics such as size, shape and texture was complete.
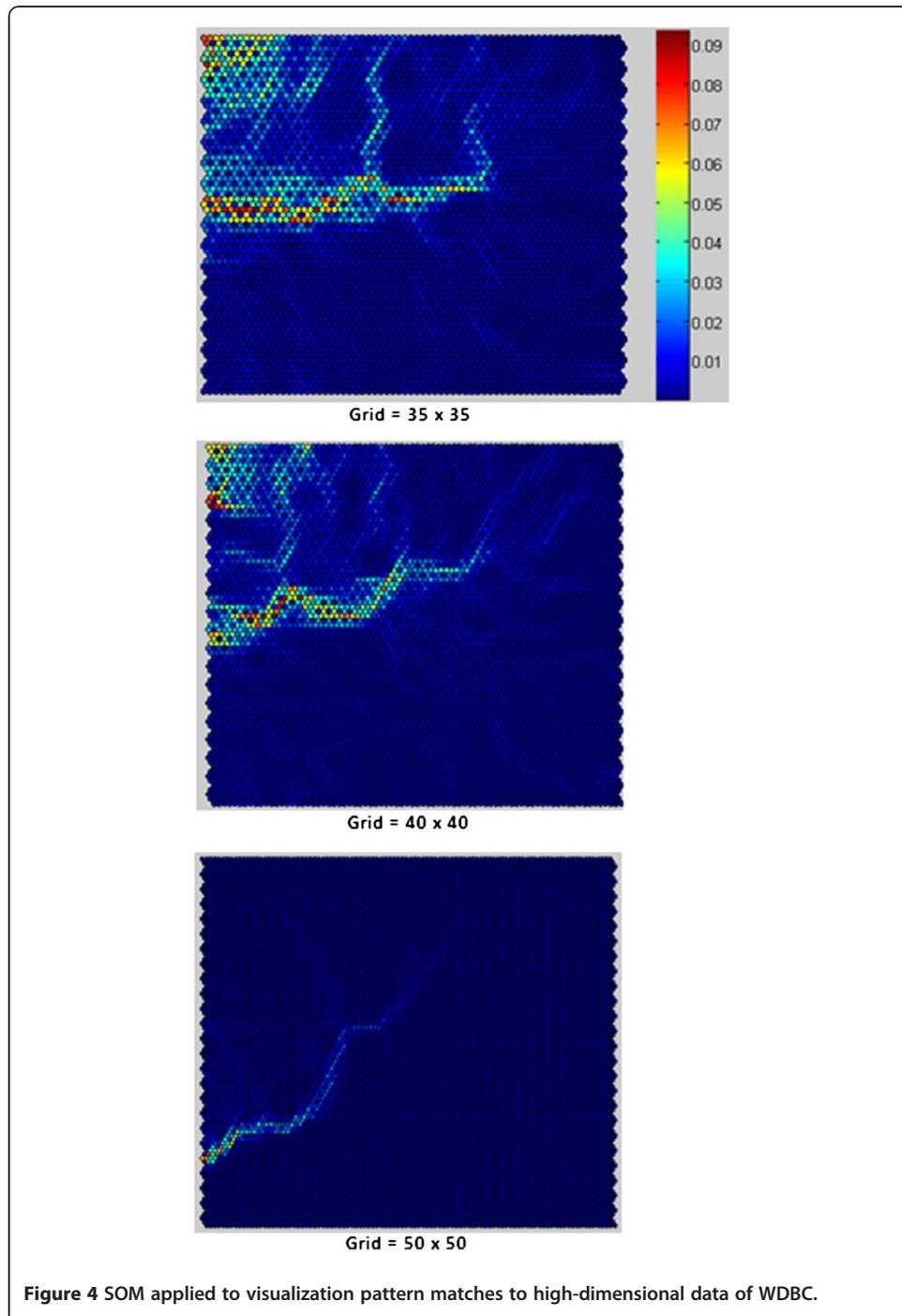
In addition to the code for the identification and diagnosis (gold pattern)[1], each record of WDBC presents 10 descriptors (related to the cell nucleus and modelled such that the highest values are associated with malignancy): radius; texture; perimeter; area; smoothness; compactness; concavity; concave points; symmetry; and fractal dimension. We must point out that the mean value, the extreme value and the standard error of each descriptor were calculated for each image, resulting in a total of 30 (thirty) resources for each case in the study.

The knowledge acquisition process was accomplished in two ways: (i) extraction and analysis of numerical data of WDBC, considering the same as gold pattern (i.e. with diagnosis confirmed)[1], and (ii) interviews and discussions with medical experts[2] (of pathology, general practitioner and mastology areas) who provided technical support and followed the development of this Pre-Diagnosis Intelligent Method.

To reduce the dimensionality of the problem and optimize the processing tests, the PCA[3] technique in WDBC was applied (average values), once having verified that the descriptors with higher energy rates are, in decreasing order: area, perimeter, texture, and radius. The experiment was repeated for the extreme values, obtaining the same result, and the analyses carried out were confirmed with medical specialists (of pathology, general practitioner and mastology areas).

Aimed at the collation and visualization of high-dimensional data, after normalization of descriptors, the SOM[4] algorithm was applied with: linear initialization; hexagonal topology; gaussian neighborhood function; neighborhood radius equal to 1; in a 10-dimensional space of characteristics (descriptors); and with variations in the size of the grid and the amount of iterations of the algorithm batch type. The results of the application of SOM were viewed using the Unified distance Matrix (U-Matrix) [27], a technique that presents weight-distance relationships between neighboring neurons of output space (distance between the units of the grid), showing a separation between the groups (classes of patterns). In the U-Matrix, relations between the neighboring neurons are seen on the surface U (x, y) as "valleys" and "mountains". Valleys - the topographic relief - correspond to the regions of neurons that are similar, while mountains, i.e. relatively high values in the U-matrix, reflect the dissimilarity between neighboring

neurons and may be associated with borders of groups of neurons [28]. The topological order property of the SOM, as shown in the right side scale in Figure 4, in the U-Matrix is as folows: a dark blue color represents that the distance between the nodes (units) is small ("valleys") and thus classes of patterns exist; the light blue, green and yellow indicate an average distance between the nodes (beginning of the "mountains"); the orange and red represent that there is a great distance between the nodes ("mountains"), i.e. they are gaps that serve to separate the classes. The application of the SOM for visualization of high-

Figure 4 SOM applied to visualization pattern matches to high-dimensional data of WDBC.

dimensional data (Figure 4) showed that it is not possible to obtain distinct groups (classes of patterns), indicating the existence of nebulous data groupings in the WDBC, which justifies the use of Fuzzy logic in the method developed.

Parallel to the application of PCA and SOM and, mainly, through preliminary analysis of the WDBC descriptors and related images performed along with medical specialists (in pathology, general practice and mastology), were: a) the extracted (selected) descriptors that was more relevant to the diagnosis of breast cancer from the analysis of cell nuclei of smears obtained by FNA, the most relevant being, the "area", the "perimeter" and the "texture"; b) the discarded descriptors, "fractal dimension", "compactness" and "concavity", because they are not actually used in medical practice for pathological analysis; and c) newly generated descriptors, in order to translate the method evaluations normally carried out by pathologists and that were not directly presented in WDBC.

Among the newly generated descriptors, those that presented a significant influence on the results "improvement" were the descriptors: "uniformity", difference between the radius extreme value and the radius mean value, representing whether the cellular nuclei have similar or highly variable sizes; and "homogeneity", difference between the extreme value of symmetry and the mean value of symmetry, representing whether the cellular nuclei have similar or highly variable symmetries.

Complementing the analysis of the numerical data, the minimum and maximum parameters for each diagnosis known (benign and malignant) of each descriptor used in the development of PDM-FNA-Fuzzy (as shown in Table 1) were detected, excluding the descriptors discarded by medical specialists. Results showed the existence of fuzzy intervals for all descriptors (benign GPD values are within the range of the malignant GPD and vice versa), not being linearly possible to diagnose breast mass as benign or malignant.

Processing and classification of data – Fuzzy Method

Before the proposed problem involving various fuzzy situations and considering the literature studied, it was found that the strategy of applying fuzzy logic could bring greater benefits (like expert knowledge acquisition, rules base generation, process

Table 1 Minimum and maximum parameters for each diagnosis (benign and malignant) of each descriptor

DESCRIPTOR	UNIT**	GPD* BENIGN		GPD* MALIGNANT	
		Minimum Value	Maximum Value	Minimum Value	Maximum Value
Area	μm^2	143.5	992.1	361.6	2501
Perimeter	µm	43.79	114.6	71.9	188.5
Texture	dimensionless	9.71	33.81	10.38	39.28
Radius	µm	6.981	17.85	10.95	28.11
Smoothness	µm	0.05263	0.1634	0.07371	0.1447
Concave Points	quantity	0	0.08534	0.02031	0.2012
Simetry	µm	0.106	0.2743	0.1308	0.304
Uniformity	µm	0.248	3.09	0.65	11.76
Homogeneity	µm	0.0184	0.2278	0.0295	0.4041

* GPD = Gold Pattern Diagnosis.
** µm (micrometer) = 1 x 10^{-6}m.

automation and pre diagnosis greater precision) and satisfactory results, in addition to dealing with modelling, representation, the reasoning and the inaccurate information procedure as a troubleshooting strategy.

Thus, the implementation of the intervention and control actions in the intelligent method developed, uses fuzzy logic since it enables to capture the experts' knowledge, as well as the appropriate treatment to fuzzy situations inherent in the problem classifying smears from breast mass obtained by FNA.

The algorithm developed to assist the creation of fuzzy system applied to the medical field is presented below.

Algorithm: establishment of fuzzy system applied to the medical area

Step 1:   Definition

- -> Identify the problem

Step 2:   Medical knowledge acquisition

- -> Obtain technical information from one or more medical specialists
- -> Extract data and information from gold pattern databases (with diagnosis confirmed)
- -> Obtain information in technical literature available

Step 3:   Fuzzification stage

- -> Define entry membership functions and their fuzzy rules

Step 4:   Rules base

- -> Define fuzzy rules covering all possibilities

Step 5:   Inference Stage

- -> Reporting observations to fuzzy sets
- -> Evaluate each case for all fuzzy rules
- -> Combine the information from the defined fuzzy rules

Step 6:   Defuzzification stage

- -> Define membership functions and output sets
- -> Define the defuzzification function

Step 7:   Results verification

- -> Ask results are satisfactory?
- If answer = "No"
- -> Return to Step 2

- If answer = "Yes"
- Finalize

This way, the definition of Fuzzy Method to assist in the diagnosis of breast cancer and its stages (Fuzzification Stage, Rules Base, Inference Stage and Defuzzification Stage) are listed below and instantiated through the system implemented.

PDM-FNA-Fuzzy Definition

Pre-Diagnosis Module FNA-Fuzzy performs the analysis of extracted descriptors of smears from breast mass obtained by FNA, considering the parameters that indicate malignant and benign diagnosis and the fuzzy rules base defined, responsible for inferences in the set of entries, generating pre-diagnosis, malignant or benign, to assist the diagnosis of breast cancer made by the doctor.

Experiments were carried out with all possible combinations of descriptors listed in Table 1, i.e. in addition to fuzzy methods for each descriptor, models have been developed for all groups of two, three, four, and so on up to the limit of nine descriptors, taking into account that the descriptors correspond to the input variables of fuzzy method. Within the set AREA, PERIMETER, UNIFORMITY and HOMOGENEITY produced the best results, the PDM-FNA-Fuzzy in question uses these four descriptors, with fuzzy method as described below.
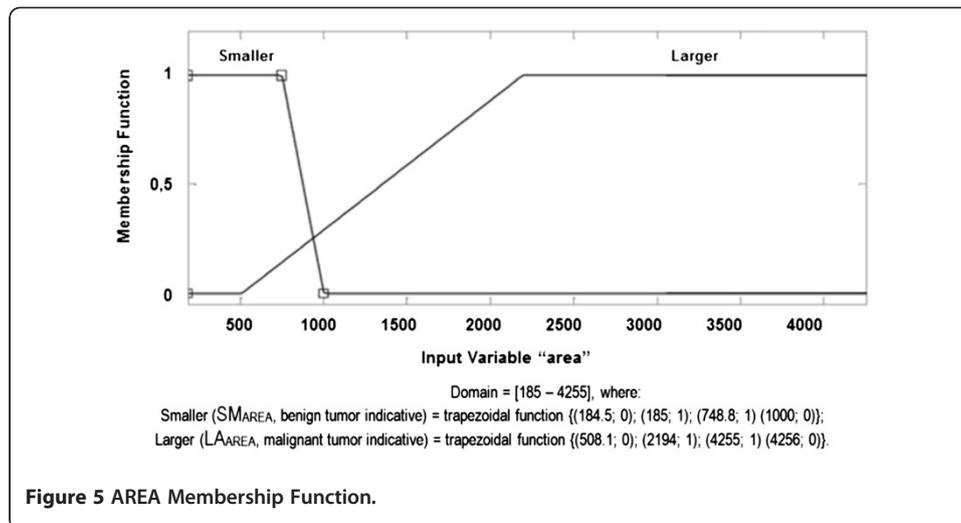
Fuzzification Stage

At this stage the input variables have been defined, identifying to which fuzzy sets they belong, assigning the respective degree to each relevance. The fuzzy sets, represented by the membership functions, were adjusted by heuristics, on the universe of discourse in order to improve the results achieved. Thus, before the creation of the fuzzy system, it was necessary to define the membership functions (fuzzy sets) used both at the fuzzification and defuzzification stages. The entries of PDM-FNA-Fuzzy in question are the descriptors AREA, PERIMETER, UNIFORMITY and HOMOGE NENEITY that have been defined through the membership functions described below:

a. Area membership function (AREA): considering a domain of [185 − 4255], this membership function is composed of "Smaller AREA" and "Larger AREA", in linguistic terms SM_{AREA} and LA_{AREA}, respectively, representing the tracks, according to the fuzzy set below and illustrated in Figure 5.

AREA fuzzy set:

Smaller AREA $(SM_{AREA}) \leq 100 \rightarrow SM_{AREA} = \{(184.5; 0), (185; 1), (748.8; 1), (100; 0)\}$;
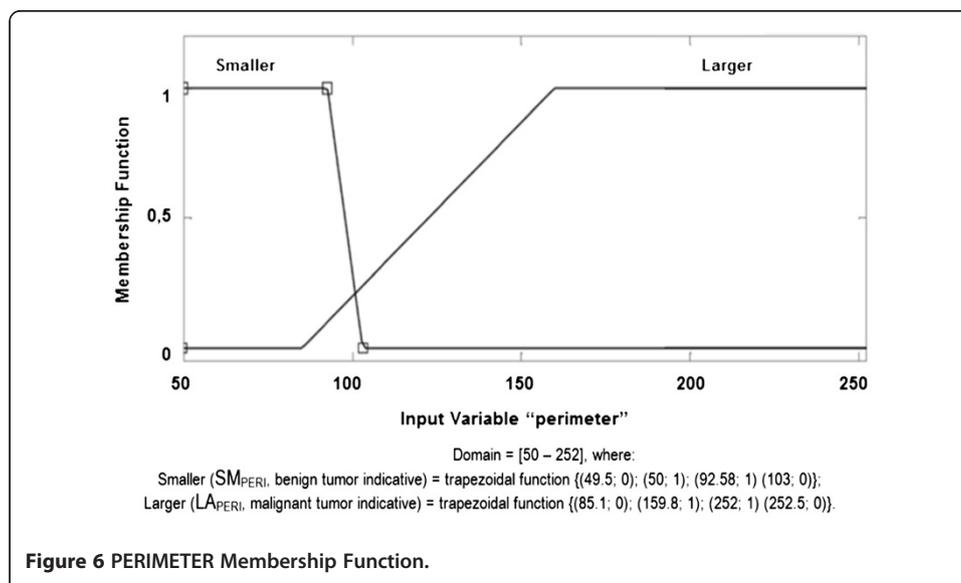Lager AREA $(LA_{AREA}) \geq 508.1 \rightarrow LA_{AREA} = \{(508.1; 0), (219; 1), (4255; 1), (4256; 0)\}$.

b. Perimeter membership function (PERI): considering a domain of [50 − 252], this membership function is composed of "Smaller PERI" and "Larger PERI", in linguistic terms SM_{PERI} and LA_{PERI}, respectively, representing the tracks, according to the fuzzy set below and illustrated in Figure 6.
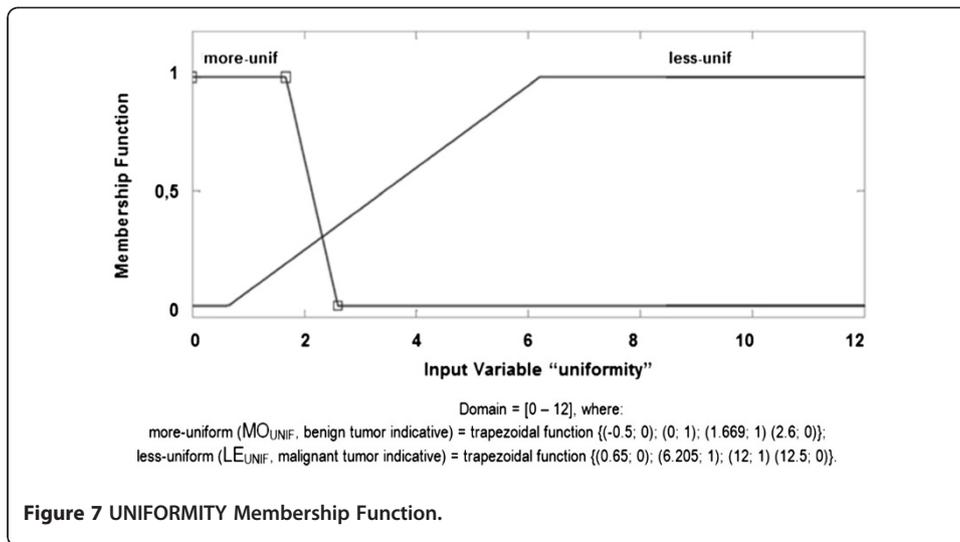
Figure 5 AREA Membership Function.

PERIMETER fuzzy set:

Smaller PERI $(SM_{PERI}) \leq 103 \rightarrow SM_{PERI} = \{(49.5; 0), (50; 1), (92.58; 1), (103; 0)\}$;
Larger PERI $(LA_{PERI}) \geq 85.1 \rightarrow LA_{PERI} = \{(85.1; 0), (159.8; 1), (252; 1), (252.5; 0)\}$.

c. Uniformity membership function (UNIF): considering a domain of $[0 - 12]$, this membership function is composed of "More UNIF" and "Less UNIF", linguistically represented as MO_{UNIF} and LE_{UNIF}, respectively, representing the tracks, according to the fuzzy set below and illustrated in Figure 7. It is important to note that, for this descriptor, more UNIF is associated with lower values (i.e. smaller values in this descriptor indicate there is more uniformity among the cellular nuclei and indicate a benign diagnosis) and less UNIF is associated to larger values (i.e. the larger values in this descriptor indicate there is less uniformity and indicate malignant diagnosis).
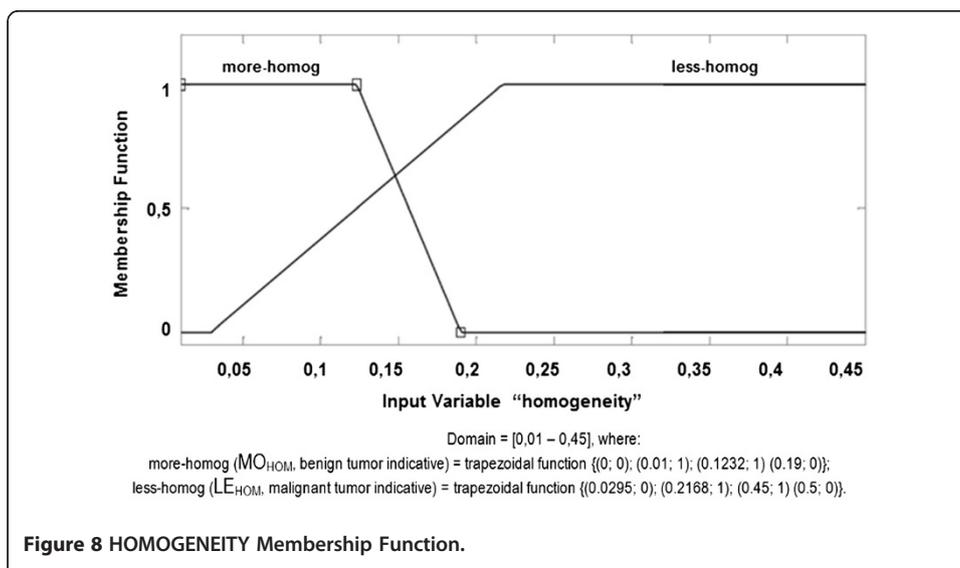

Figure 6 PERIMETER Membership Function.

Domain = [0 – 12], where:
more-uniform (MO$_{UNIF}$, benign tumor indicative) = trapezoidal function {(-0.5; 0); (0; 1); (1.669; 1) (2.6; 0)};
less-uniform (LE$_{UNIF}$, malignant tumor indicative) = trapezoidal function {(0.65; 0); (6.205; 1); (12; 1) (12.5; 0)}.

Figure 7 UNIFORMITY Membership Function.

UNIFORMITY fuzzy set:

More UNIF (MO$_{UNIF}$)≤ 2.6→MO$_{UNIF}$ = {(−0.5; 0), (0; 1), (1.669; 1), (2.6; 0)};
Less UNIF (LE$_{UNIF}$)≥0.65→LE$_{UNIF}$ = {(0.65; 0), (6.205; 1), (12; 1), (12.5; 0)}.

d. Homogeneity membership function (HOM): considering a domain of [0.01 − 0.45], this membership function is composed of "More HOM" and "Less HOM", in linguistic terms MO$_{HOM}$ and LE$_{HOM}$, respectively, representing the tracks, according to the fuzzy set below and illustrated in Figure 8. It is important to note that, for this descriptor, more HOM is associated with lower values (i.e. smaller values in this descriptor indicate there is more homogeneity among the cellular nuclei and indicate a benign diagnosis) and less HOM is linked to larger values (i.e. larger values in this descriptor indicate there is less homogeneity and indicate a diagnosis of malignancy).

Domain = [0,01 – 0.45], where:
more-homog (MO$_{HOM}$, benign tumor indicative) = trapezoidal function {(0; 0); (0.01; 1); (0.1232; 1) (0.19; 0)};
less-homog (LE$_{HOM}$, malignant tumor indicative) = trapezoidal function {(0.0295; 0); (0.2168; 1); (0.45; 1) (0.5; 0)}.

Figure 8 HOMOGENEITY Membership Function.

HOMOGENEITY fuzzy set:

More HOM $(MO_{HOM}) \leq 0.19 \rightarrow MO_{HOM} = \{(0;0), (0.01;1), (0.1232;1), (0.19;0)\}$;
Less HOM $(LE_{HOM}) \geq 0.0295 \rightarrow LE_{HOM} = \{(0.0295;0), (0.2168;1), (0.45;1), (0.5;0)\}$.

The membership functions were built using the direct method, having been confirmed by the medical experts (in pathology, mastology and general practice) the parameters extracted from WDBC, covering all data of the membership functions (values that represent each function and the degree of relevance, within the function, of each one of them) in order to set them explicitly. There are several membership functions that can be used at this fuzzification stage. All functions available in Matlab were applied (trials and tests) on the fuzzy system concerned, noting that the trapezoidal function was the one that presented the best results in PDM-FNA-Fuzzy, by best representing the functions according to the context.

Rules Base – Fuzzy Rules definition

The rules base was assembled with the following structure: IF <premises>THEN <conclusion>. For the definition of rules of PDM-FNA-Fuzzy to assist in the diagnosis of breast cancer in question, it is possible to standardize the following structure:

$R : \{R_1, R_2, R_3, \ldots, R_n\} \rightarrow$ Set of rules;
$DESC : \{DESC_1, DESC_2, DESC_3, \ldots, DESC_n,\} \rightarrow$ Set of descriptors
  (representing the set of entries);
$P : \{B(\downarrow), Undef(\downarrow\uparrow), M(\uparrow)\} \rightarrow$ Parameterization of the descriptor's situation
  (Benign, Undefined and Malignant);
$D : \{D_1, D_2, D_3, \ldots, D_n\} \rightarrow$ Set of diagnosis possibilities.

Rules definition:

$< (R_1, R_2, R_3, \ldots, R_n,) >:$ **IF** $< (DESC1, DESC2, DESC3, \ldots, DESCn,) >$
$< \{ B(\downarrow), Undef(\downarrow\uparrow), M(\uparrow)\} > $ **and/or** $< (DESC1, DESC2, DESC3, \ldots, DESCn,) >$
$< \{ B(\downarrow), Undef(\downarrow\uparrow), M(\uparrow)\} >$ **and/or** ... **THEN** $< D_1, D_2, D_3, \ldots, D_n >$

Consequently, 16 (sixteen) rules were defined for PDM-FNA-Fuzzy object of this study, using 4 (four) descriptors and with 3 (three) possibilities of pre-diagnosis <results>. To exemplify, below some of the rules:

Rule 1: If smaller AREA and smaller PERI and more UNIF (descriptor with lower value) and more HOM (descriptor with smaller value) then benign diagnosis.
  if AREA↓and PERI↓and UNIF↓and HOM↓then B

Rule13: If larger AREA and larger PERI and more UNIF (descriptor with lower value) and more HOM (descriptor with smaller value) then diagnosis undefined.
  if AREA↑**and** PERI↑**and** UNIF↓**and** HOM↓**then** Undef

Rule 16: If larger AREA and larger PERI and less UNIF (understand descriptor with larger value) and less HOM (descriptor with larger value) then malignant diagnosis.
  if AREA↑**and** PERI↑**and** UNIF↑**and** HOM↑**then** M

We must point out that the rules defined (16 rules) cover all possible combinations of inputs and outputs of the proposed issue and that the consistency of the rules was examined in order to avoid contradictions. The rules base, presented in the Table 2, was developed from the analysis of numerical data and multiple meetings, discussions and interviews with medical experts from the fields of pathology, mastology, and general practitioners.

It should be noted that, following the medical practice, the procedure taken for undefined (Undef) pre-diagnosis (results) are referred to as the situation of suspected malignant tumour, which indicates a biopsy procedure(similar to the malignant pre-diagnosis), i.e., if in doubt the patient is referred for a biopsy. Thus, from the classification point of view for having a biopsy or not, the PDM-FNA-Fuzzy can be seen as a binary classifier with the record of 2.1% of cases classified as undefined and thus regarded as malignant.

Inference stage

In this stage, the entries were analysed to generate the fuzzy output set with its respective compatibility degree. The PDM-FNA-Fuzzy developed used the fuzzy model proposed by Mamdani [29], in which the activation function of each rule is enabled and the system of inference determines the degree of compatibility of the rules premise contained in the rules base. After this, it determines which rules are enabled and applies them to the output membership function, remaining just linking all output nebulous sets activated (and their respective degrees of compatibility) into a single Output Set (OS). This OS represents all results (diagnosis) that are acceptable for the input set, each with its compatibility level. Each case was also assessed, at this stage, for all fuzzy rules and the combination of information was carried out from the rules already defined in the Rules Base.

Table 2 Rules base

Rule number	Rule specification
1	**if** AREA ↓ **and** PERI ↓ **and** UNIF ↓ **and** HOM ↓ **then** B
2	**if** AREA ↓ **and** PERI ↓ **and** UNIF ↓ **and** HOM ↑ **then** Undef
3	**if** AREA ↓ **and** PERI ↓ **and** UNIF ↑ **and** HOM ↓ **then** Undef
4	**if** AREA ↓ **and** PERI ↓ **and** UNIF ↑ **and** HOM ↑ **then** Undef
5	**if** AREA ↓ **and** PERI ↑ **and** UNIF ↓ **and** HOM ↓ **then** Undef
6	**if** AREA ↓ **and** PERI ↑ **and** UNIF ↓ **and** HOM ↑ **then** Undef
7	**if** AREA ↓ **and** PERI ↑ **and** UNIF ↑ **and** HOM ↓ **then** Undef
8	**if** AREA ↓ **and** PERI ↑ **and** UNIF ↑ **and** HOM ↑ **then** Undef
9	**if** AREA ↑ **and** PERI ↓ **and** UNIF ↓ **and** HOM ↓ **then** Undef
10	**if** AREA ↑ **and** PERI ↓ **and** UNIF ↓ **and** HOM ↑ **then** Undef
11	**if** AREA ↑ **and** PERI ↓ **and** UNIF ↑ **and** HOM ↓ **then** Undef
12	**if** AREA ↑ **and** PERI ↓ **and** UNIF ↑ **and** HOM ↑ **then** Undef
13	**if** AREA ↑ **and** PERI ↑ **and** UNIF ↓ **and** HOM ↓ **then** Undef
14	**if** AREA ↑ **and** PERI ↑ **and** UNIF ↓ **and** HOM ↑ **then** Undef
15	**if** AREA ↑ **and** PERI ↑ **and** UNIF ↑ **and** HOM ↓ **then** Undef
16	**if** AREA ↑ **and** PERI ↑ **and** UNIF ↑ **and** HOM ↑ **then** M

Defuzzification stage

This stage was used to generate a single numeric value, from all possible values contained in the fuzzy set obtained in the inference stage, to generate the diagnosis. As a diagnosis resulting from the relations and variability of the descriptors AREA, PERIMETER, UNIFORMITY and HOMOGENEITY, the function centroid (which presented the best results) and the domain $[0 - 1]$ was adopted for defuzzification.

The Pre-Diagnosis (PD) membership function, Defuzzification, is composed of "Benign", "Undefined" and "Malignant", represented linguistically as B_{PD}, $Undef_{PD}$ and M_{PD}, respectively, representing the tracks $[\leq 0.5; 0.5 - 0.6;$ and $\geq 0.6]$, as output set below and illustrated in Figure 9.

Output Set (OS):

$$PD \text{ benign } (B_{PD}) \leq 0.5 \rightarrow B_{PD} = \{(-0.5; 0), (0; 1), (0.4; 1), (0.5; 0)\};$$
$$PD \text{ undefined } (Undef_{PD}) \geq 0.5 \text{ e} \leq 0.6 \rightarrow Undef_{PD} = \{(0.5; 0), (0.55; 1), (0.6; 0)\};$$
$$PD \text{ malignant } (M_{PD}) \geq 0.6 \rightarrow M_{PD} = \{(0.6; 0), (0.7; 1), (1; 1), (1.5; 0)\}.$$
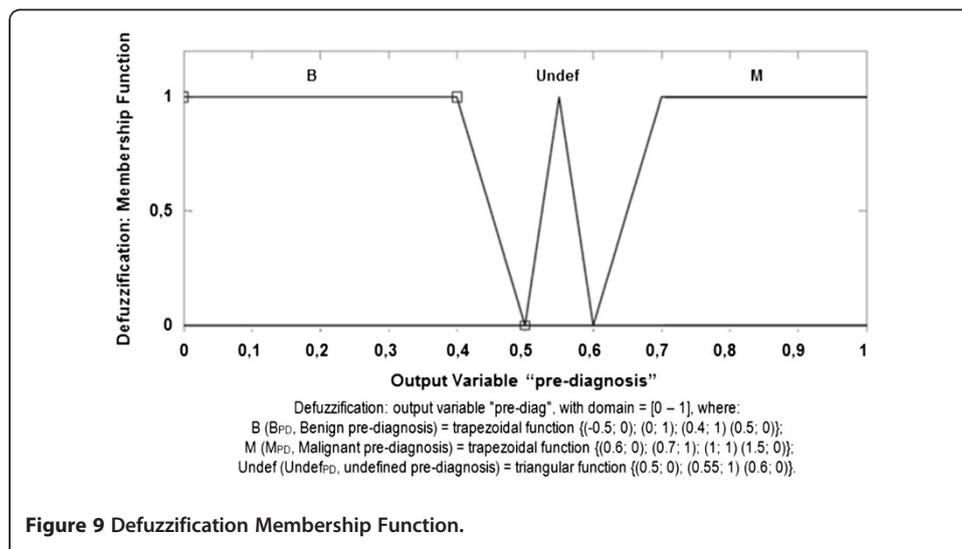
Post-processing

In post-processing, the result, in the form of malignant or benign pre-diagnosis, is stored on the server and made available on the screen by means of IVEMI (on the desktop discussion of clinical case), both for the doctor who requested the pre-diagnosis or second opinion, as to the other users with access permission to the respective clinical case.

Validation

Testing of the PDM-FNA-Fuzzy, the object of this study, was carried out using the MATLAB R2010a (student version), due to the tools available in this application to the development of models and the rapid visualization of the results obtained in the fuzzy system.

The PDM-FNA-Fuzzy developed to assist in the diagnosis of breast cancer performs the interaction between the descriptors AREA, PERIMETER, UNIFORMITY and HOMOGENEITY (extracted from smears obtained by FNA), operated by inference

Figure 9 Defuzzification Membership Function.

rules of the system expert in fuzzy logic, triggering classification and assistance in medical diagnosis actions.

All the tests were initially carried out using the WDBC, even the identification of the main characteristics of the fuzzy system, such as:

- the identification of the set of descriptors that provide the best results, called "best input set" (BIS);
- identification of the best set of rules (BSR); and
- the definition of what membership functions, which parameters and what defuzzification functions are most suitable for use with the BIS and the BSR.

The membership functions and their respective fuzzy final sets of each descriptor used, AREA, PERIMETER, UNIFORMITY and HOMOGENEITY, are presented, respectively, in Figures 5, 6, 7 and 8.

The validation of the rules base was held in conjunction with medical professionals (in pathology, general practice and mastology), considering the fuzzy set indicators of both malignant and benign diagnosis. As a consequent action of the descriptors' relations and variability the domain $[0 - 1]$, representing the tracks $[< 0.5; 0.5 - 0.6; > 0.6]$, was adopted to defuzzification, which is represented in linguistic terms as "Benign", "Undefined" and "Malignant", respectively, as presented in Figure 9.

After this phase, cross-validation was used for testing, in order to fine-tune the parameters of the membership functions of PDM-FNA-Fuzzy. Therefore, three databases were generated, each of them with 150 (one hundred and fifty) gold pattern clinical cases randomly extracted from WDBC.

The validation of PDM-FNA-Fuzzy was performed using a database with 100 (one hundred) gold pattern clinical cases (i.e. diagnosis known and confirmed), randomly extracted from WDBC.

We must point out that the validations of both the knowledge gained and the results achieved were performed during the development of PDM-FNA-Fuzzy and also, in the final instance, by medical specialists in the areas of pathology, mastology and general practice.

Results

By means of the experiments performed, it was found that the input set that featured the best results has the following characteristics:

a) fuzzy system: Mamdani;
b) membership functions of the entry set: trapezoidal;
c) input set composed of 4 variables (descriptors), with the following fuzzy sets:
c.1)

$$\text{AREA com SM}_{\text{AREA}} = \{(184.5; 0), (185; 1), (748.8; 1), (1000; 0)\}$$
$$\text{and LA}_{\text{AREA}} = \{(508.1; 0), (2194; 1), (4255; 1), (4256; 0)\};$$

c.2)

$$\text{PERIMETER with L}_{\text{PERI}} = \{(49.5; 0), (50; 1), (92.58; 1), (103; 0)\}$$
$$\text{and LA}_{\text{PERI}} = \{(85.1; 0), (159.8; 1), (252; 1), (252.5; 0)\};$$

c.3.)

UNIFORMITY with $MO_{UNIF} = \{(-0.5; 0), (0; 1), (1.669; 1), (2.6; 0)\}$
and $LE_{UNIF} = \{(0.65; 0), (6.205; 1), (12; 1), (12.5; 0)\}$; and

c.4.)

HOMOGENEITY with $MO_{HOM} = \{(0; 0), (0.01; 1), (0.1232; 1), (0.19; 0)\}$
and $LE_{HOM} = \{(0.0295; 0), (0.2168; 1), (0.45; 1), (0.5; 0)\}$;

d) rules base: 16 rules;
e) membership functions of the output set:
e.1)

trapezoidal for classification**Benign**, being B_{PD}
$= \{(-0.5; 0), (0; 1), (0.4; 1), (0.5; 0)\}$;

e.2)

trapezoidal for classification**Malignant**, being M_{PD}
$= \{(0.6; 0), (0.7; 1), (1; 1), (1.5; 0)\}$; and

e.3)

triangular for classification**Undefined**, being $Undef_{PD}$
$= \{(0.5; 0), (0.55; 1), (0.6; 0)\}$;

f) defuzzification: Centroid function;
g) output variable: 1 (result = pre-diagnosis).

The best result achieved is shown in the Diagnostic Test Assessment Matrix presented in Table 3, as well as in the Matrix of Confusion presented in Table 4.

It is noted in the diagnostic test assessment matrix (Table 3), that the PDM-FNA-Fuzzy developed presents: 98.59% sensitivity, which is the ability of a diagnostic test to identify the real positive in individuals truly ill, meaning a satisfactory percentage of hits in the pre-diagnosis of malignancies; and 85.43% specificity, which is the ability of a diagnostic test to identify the real negative in individuals truly healthy, corresponding to the correct pre-diagnosis of benign cases.

We must point out that, in the laboratory examination (biopsy) of smears obtained by FNA for identification of breast cancer, it is more important to get good results in sensitivity than in specificity ([30-32]). Subsequently, among the tests performed during the development of PDM-FNA-Fuzzy to assist in the diagnosis of breast cancer, there were several

Table 3 Diagnostic test of assessment matrix of PDM-FNA-Fuzzy developed to assist in the diagnosis of breast cancer

Diagnostic test Assessment			
	GOLD PATTERN DIAGNOSIS		
FUZZY-FNA	Malignant (%)	Benign (%)	**TOTAL**
Malignant (%)	36.73	9.14	45.87
Benign (%)	0.53	53.60	54.13
TOTAL	37.26	62.74	100.00
Sensitivity = 98.59%	**Specificity = 85.43%**		

Table 4 Confusion matrix of the diagnostic test of PDM-FNA-Fuzzy developed to assist the diagnosis of breast cancer

Confusion matrix		
	GOLD PATTERN	
FUZZY-FNA	Malignant	Benign
Malignant	**0.99**	0.15
Benign	0.01	**0.85**

with satisfactory results as well, but they were not selected as the best solution, having been discarded, as, for example, the test sets A, B and C, presented below.

The tests of set A were conducted from the best input set, with changes in nebulous sets (parameters) of the membership functions. In Table 5, the results of sensitivity and specificity of the same are presented. Notably test A.1 presents 99.06% sensitivity, however the medical experts found the specificity of 64.15% unsatisfactory. The tests A.8 and A.10 presented the same sensitivity of PDM-FNA-Fuzzy developed (98.59%), but lower specificity (84.31% and 84.87%, respectively). The other tests presented sensitivity less than 98.59% and thus were discarded.

The tests of set B were conducted from the best input set, with changes in the types of membership function of the input set and, consequently, in their nebulous set

Table 5 Comparison of the tests presented in "TEST SET A" (changes were realized in the fuzzy sets of membership functions)

Tests	Sensitivity (%)	Specificity (%)
PDM-FNA-Fuzzy developed	**98.59**	**85.43**
Test A.1 [1]	99.06	64.15
Test A.2 [2]	92.92	90.48
Test A.3 [3]	98.11	70.87
Test A.4 [4]	93.87	89.92
Test A.5 [5]	96.23	88.80
Test A.6 [6]	97.17	87.39
Test A.7 [7]	97.64	86.83
Test A.8 [8]	98.59	84.31
Test A.9 [9]	98.11	86.55
Test A.10 [10]	98.59	84.87

(1) changes in Test A.1: SM_{AREA} = {(184.5; 0), (185; 1), (749; 1), (800; 0)} e SM_{PERI} = {(49.5; 0), (50; 1), (92.6; 1), (95; 0)} e MO_{UNIF} = {(-0.5; 0), (0; 1), (1.67; 1), (1.87; 0)} e MO_{HOM} = {(0; 0), (0.01; 1), (0.123; 1), (0.143; 0)}.
(2) changes in Test A.2: SM_{AREA} = {(184.5; 0), (185; 1), (748.8; 1), (800; 0)} e SM_{PERI} = {(49.5; 0), (50; 1), (92.58; 1), (127.1; 0)} e MO_{UNIF} = {(-0.5; 0), (0; 1), (1.669; 1), (3.09; 0)} e MO_{HOM} = {(0; 0), (0.01; 1), (0.1232; 1), (0.2278; 0)}.
(3) changes in Test A.3: SM_{AREA} = {(184.5; 0), (185; 1), (748.8; 1), (800; 0)} e SM_{PERI} = {(49.5; 0), (50; 1), (92.58; 1), (95; 0)} e MO_{UNIF} = {(-0.5; 0), (0; 1), (1.669; 1), (3.09; 0)} e MO_{HOM} = {(0; 0), (0.01; 1), (0.1232; 1), (0.2278; 0)}.
(4) changes in Test A.4: SM_{AREA} = {(184.5; 0), (185; 1), (748.8; 1), (800; 0)} e SM_{PERI} = {(49.5; 0), (50; 1), (92.58; 1), (110; 0)} e MO_{UNIF} = {(-0.5; 0), (0; 1), (1.669; 1), (3.09; 0)} e MO_{HOM} = {(0; 0), (0.01; 1), (0.1232; 1), (0.2278; 0)}.
(5) changes in Test A.5: SM_{AREA} = {(184.5; 0), (185; 1), (748.8; 1), (800; 0)} e SM_{PERI} = {(49.5; 0), (50; 1), (92.58; 1), (106; 0)} e MO_{UNIF} = {(-0.5; 0), (0; 1), (1.669; 1), (3.09; 0)} e MO_{HOM} = {(0; 0), (0.01; 1), (0.1232; 1), (0.2278; 0)}.
(6) changes in Test A.6: SM_{AREA} = {(184.5; 0), (185; 1), (748.8; 1), (800; 0)} e SM_{PERI} = {(49.5; 0), (50; 1), (92.58; 1), (106; 0)} e MO_{UNIF} = {(-0.5; 0), (0; 1), (1.669; 1), (3.09; 0)} e MO_{HOM} = {(0; 0), (0.01; 1), (0.1232; 1), (0.18; 0)}.
(7) changes in Test A.7: SM_{AREA} = {(184.5; 0), (185; 1), (748.8; 1), (800; 0)} e SM_{PERI} = {(49.5; 0), (50; 1), (92.58; 1), (106; 0)} e MO_{UNIF} = {(-0.5; 0), (0; 1), (1.669; 1), (2.5; 0)} e MO_{HOM} = {(0; 0), (0.01; 1), (0.1232; 1), (0.18; 0)}.
(8) changes in Test A.8: SM_{AREA} = {(184.5; 0), (185; 1), (748.8; 1), (800; 0)} e MO_{UNIF} = {(-0.5; 0), (0; 1), (1.669; 1), (2.5; 0)} e MO_{HOM} = {(0; 0), (0.01; 1), (0.1232; 1), (0.18; 0)}.
(9) changes in Test A.9: SM_{PERI} = {(49.5; 0), (50; 1), (92.58; 1), (103.5; 0)}.
(10) changes in Test A.10: SM_{PERI} = {(49.5; 0), (50; 1), (92.58; 1), (102.3; 0)}.

Table 6 Comparison of the tests presented in "TEST SET B" (changes were realized in the membership functions of the entry set and its fuzzy sets)

Tests	Type of membership function (after adjustments in fuzzy sets)	Sensitivity (%)	Specificity (%)
PDM-FNA-Fuzzy developed	**trapezoidal[1]**	**98.59**	**85.43**
Test B.1	triangular[2]	98.59	83.47
Test B.2	gaussian2[3]	98.11	84.31
Test B.3	dsigmoidal[4]	98.11	84.59
Test B.4	polinomial zero[5]	98.59	82.91

(1) trapezoidal - function with straight lines with a flat top, resembling a truncated triangle.
(2) triangular - function with straight lines, in the form of a triangle.
(3) gaussiana2 - composed of two different gaussian curves.
(4) dsigmoidal - created from the difference between two sigmoidais functions.
(5) polinomial zero – asymmetric polynomial function, being zero at both ends, with an increase in the middle.

(parameters). In Table 6, the results of sensitivity and specificity of the same are presented. Notably the tests B.1 and B.4 showed the same sensitivity that the PDM-FNA-Fuzzy developed (98.59%), but lower specificity (84.47% and 82.91%, respectively). The other tests showed sensitivity less than 98.59%, having been discarded.

The C set were performed from the best input set, with changes only in the defuzzification function. Presented in Table 7, are the results of sensitivity and specificity of the same. It is worthy to note that all of the tests presented the same sensitivity that the PDM-FNA-Fuzzy developed (98.59%), but lower specificity.

Thus, the results achieved by the PDM-FNA-Fuzzy, the object of this study, were considered satisfactory by medical specialists (in pathology, general practice and mastology), mainly for their high sensitivity (malignant cases hit) presented, as can be seen in Table 3.

The sensitivity of 98.59% presented by MPD-FNA-Fuzzy is at the same level of prominence of other works using the same dataset with other techniques such as, for example, [11] and [14], using Probabilistic Neural Network-PNN with 31-568-2 topology. Although other works, for example, [11], [12] [14], [15], are more accurate than MPD-FNA-Fuzzy, they use ten descriptors, while the MPD-FNA-Fuzzy uses only four descriptors, two of which are extracted indirectly from WDBC, which simplifies the model and streamlines processing.

Table 7 Comparison of tests presented in "TEST SET C" (changes were realized in the defuzzification functions)

Tests	Defuzzification function	Sensitivity (%)	Specificity (%)
PDM-FNA-Fuzzy developed	**centroid[1]**	**98.59**	**85.43**
Test C.1	bisector[2]	98.59	83.47
Test C.2	mom[3]	98.59	77.59
Test C.3	lom[4]	98.59	73.67
Test C.4	som[5]	98.59	77.59

(1) centroid - calculates the output set (OS) area center generated in the inference stage and determines its projection on the x-axis, that is the control output value.
(2) bisector - exact position that splits the output set into two equal areas.
(3) mom (Middle of Maximum) - it performs the arithmetic mean of all maximum values of the OS.
(4) lom (Largest of Maximum) - considers the greatest among all the maximum values of the OS.
(5) som (Smallest of Maximum) - considers the lowest among all the maximum values of the OS.

Conclusions

This work presented an intelligent method to assist the diagnosis and second opinion of breast cancer, using a fuzzy method capable of processing and sorting data (descriptors) extracted from smears of breast mass obtained by FNA.

Processing, testing and validation using fuzzy method were carried out by medical specialists using the gold pattern database, i.e. with real data and real and verified diagnosis.

The main contributions of this paper are:

- specification and implementation of fuzzy method (MPD-FNA-Fuzzy) that meets the requirements to assist breast cancer diagnosis, carried out using the analysis of real data and contact with experts;
- reduction of malignant cases variation hit when compared to visual interpretation currently applied in the diagnosis by FNA. While the MPD-FNA-Fuzzy features stable sensitivity in 98.59%, visual interpretation diagnosis provides a sensitivity variation from 65% to 98%, this track showing sensitivity levels below those considered satisfactory by medical specialists;
- the use of intelligent systems techniques, more specifically, fuzzy logic, to assist the diagnosis and second opinion of breast cancer from smears of FNA;
- development of a Pre-Diagnosis Method that can be embedded into a virtual environment of medical interaction;
- detection of the main descriptors of WDBC to assist the diagnosis of breast cancer;
- creation, from the WDBC, of new important descriptors to assist the breast cancer diagnosis: UNIFORMITY and HOMOGENEITY;
- definition of algorithm for fuzzy system development applied to the medical field.

Endnotes

[1]Gold pattern means that the true diagnosis is known and confirmed for each clinical case. In the case of WDBC, malignant diagnoses were confirmed by surgical biopsy and benign diagnosis by subsequent periodic medical examinations.

[2]Onofre Lopes Hospital (UFRN); Graduate Program in Health Sciences (UFRN); Promater Hospital; e Oncology and Mastology Clinic of Natal/RN.

[3]PCA (Principal Component Analysis) is a linear projection technique that performs statistical analysis of correlation between parameters, reducing the dimensionality of the problem [33].

[4]SOM (Self Organizing Map), also known as Kohonen self-organizing maps, that have the ability to form mappings that preserve the topology between input and output spaces [33].

Author details
[1]Department of Computer Engineering and Automation - DCA and Department of Biomedical Engineering, Federal University of Rio Grande do Norte (UFRN), Caixa Postal 1524 – Campus Universitário - UFRN/CT/DCA, Rio Grande do Norte, CEP 59072-970, Brazil. [2]Department of Informatic - State University of Rio Grande do Norte (UERN), Av. Campus Universitário, Rio Grande do NorteBR 110, Km 46CEP 59625-620, Brazil.

References
1. Chen H-L, Yang B, Liu J, Liu D-Y: **A support vector machine classifier with rough set-based feature selection for breast cancer diagnosis.** *Expert Systems with Applications: An International Journal* July, 2011, **38**(7):9014–9022. doi:10.1016/j.eswa.2011.01.120.
2. WHO Disease and injury country estimates: *World Health Organization, 2009*. URL: http://www.who.int/healthinfo/global_burden_disease/en/. (disponível na Internet, capturado em 08/03/2011).
3. Sriraam N, Eswaran C: **Performance Evaluation of Neural Network and Linear Predictors for Near-Lossless Compression of EEG Signals.** *Information Technology in Biomedicine, IEEE Transactions on. USA* 2008, **12**:87–93.
4. Soares HB, Dória Neto AD, Carvalho MAG: **An Intelligent System for Detection and Analysis of Skin Cancer based on Wavelet Transform and Support Vector Machine.** *In: XVIII SIBGRAPI, 2005, Natal. Proc. of XVIII SIBGRAPI* 2005.
5. Rogal S Jr, Parais EC, Kaestner CAA, Figueredo MV, Beckert Neto A: *Grouping of Cardiac Arrhythmias Using ART2*. Workshop. Uberlândia, MG, Brasil: I Algorithms and Data Mining; 2005.
6. Leite CRM, Sizilio GRMA, Dória Neto AD, Valentim RAM, Guerreiro AMGA: **Fuzzy Model for Processing and Monitoring Vital Signs in ICU Patients.** *BioMedical Engineering Online (Online)* 2011, **10**:68. doi:10.1186/1475-925X-10-68.
7. Jara AJ, Blaya FJ, Zamora MA, Skarmeta A: **An ontology and rule based intelligent information system to detect and predict myocardial diseases. Information Technology and Applications in Biomedicine, 2009.** *ITAB 2009, 9th International Conference on. Larnaca, Chipre* 2009:1–6.
8. Koutsojannis C, Nabil E, Tsimara M, Hatzilygeroudis I: *Using Machine Learning Techniques to Improve the Behaviour of a Medical Decision Support System for Prostate Diseases*. ISDA '09. Ninth International Conference on. Pisa, Italy: Intelligent Systems Design and Applications, 2009; 2009:341–346.
9. Barakat N, Bradley AP, Barakat MNH: **Intelligible Support Vector Machines for Diagnosis of Diabetes Mellitus. Information Technology in Biomedicine, IEEE Transactions on.** *USA:* July 2010, **14**:1114–1120.
10. Sewak M, Vaidya P, Chan C-C, Duan Z-H: **SVM Approach to Breast Cancer Classification. Computer and Computational Sciences, 2007. IMSCCS 2007.** *Second International Multi-Symposiums* 2007, **13-15**:32–37. doi:10.1109/IMSCCS.2007.46.
11. Anagnostopoulos I, Anagnostopoulos C, Vergados D, Rouskas A, Kormentzas G: **The Wisconsin Breast Cancer Problem: Diagnosis and TTR/DFS Time Prognosis Using Probabilistic and Generalised Regression Information Classifiers.** *Oncology Reports, special issue Computational Analysis and Decision Support Systems in Oncology* 2006, **15**:975–981.
12. Mohamed MA, Hegazy AE-F, Badr AA: **Evolutionary Fuzzy ARTMAP Approach for Breast Cancer Diagnosis.** *International Journal of Computer Science and Network Security* April, 2011, **11**(4):77–84.
13. Liu B, Abbass HA, McKay B: **Classification Rule Discovery with ant Colony Optimisation.** *IEEE Computational Intelligence Bulletin* February, 2004, **3**(1):31–35.
14. Anagnostopoulos I, Maglogiannis I: **Neural Network-Based Diagnostic and Prognostic Estimations in Breast Cancer Microscopic Instances.** *Medical and Biological Engineering and Computing Journal* 2006, **44**(9):773–784. doi:10.1007/s11517-006-0079-4.
15. Aruna S, Rajagopalan SPA, Nandakishore LV: **An Empirical Comparasion of Supervised Learning Algorithms in Disease Detection.** *International Journal of Information Technology Convergence and Services – IJITCS* 2011, **1**:81–92.
16. Sizilio GRMA, Leire CRM, Guerreiro AMG, Dória Neto AD: **Ambiente de Telediagnóstico Colaborativo Utilizando Plafaforma Inteligente de Auxílio à Tomada de Decisão.** *Revista Brasileira de Engenharia Biomédia* 2011, **27**:1–12. ISSN 1517-3151.
17. Monfair F: *Essentials of Diagnostic Breast Pathology: A Practical Approach*. Berlin: Springer-Verlag Publishing; 2007.
18. Rakha EA, Ellis IO: **An overview of assessment of prognostic and predictive factors in breast cancer needle core biopsy specimens.** *J Clin Pathol.* 2007, **60**(12):1300–1306. PMCID: PMC2095575. Copyright 2007 The BMJ Publishing Group and the Association of Clinical Pathologists. Published online 2007 July 14. doi: 10.1136/JCP.2006.045377.
19. Street WN: *Xcyt. A system for remote cytological diagnosis and prognosis of breast cancer*. Soft Computing Techniques in Breast Cancer Prognosis and Diagnosis, in press. Boca Raton, FL: CRC Press: In: L. C. Jain (ed.); 1999.
20. Engelbrech AP: *Computational Intelligence: An Introduction*. Chichester, UK: 2nd ed. John Wiley and Sons; 2007.
21. Dubois D, Prade H: **Fuzzy Sets and Fuzzy Systems.** *Theory and Applications, Academic Press* 1980.
22. Zadeh LA: **Fuzzy sets.** *Information and Control* 1965, **8**(3):338–353.

23. Zadeh LA: *Fuzzy sets and information granularity*. North-Holland Publishing Co.: Amsterdam: In Advances in Fuzzy Set Theory and Applications, M. M. Gupta, R. K. Ragade and R. R. Yager editors, 3–18; 1979.
24. Lukasiewicz J: **O logice trójwartościowej (in Polish). Ruch filozoficzny 5:170–171. English translation: On three-valued logic, in L. Borkowski (ed.).** *Selected works by Jan Lukasiewicz, North–Holland, Amsterdam* 1970:87–88. ISBN 0-7204-2252-3.
25. Frank A, Asuncion A: *UCI Machine Learning Repository. University of California, School of Information and Computer Science: Irvine, CA, 2010.* URL: http://archive.ics.uci.edu/ml (disponível na I nternet, capturado em 08/12/2010).
26. Street WN, Wolberg WH, Mangasarian OL: **Nuclear feature extraction for breast tumor diagnosis. IS&T/SPIE 1993 International Symposium on Electronic Imaging.** *Science and Technology* 1993, **1905**:861–870.
27. Ultsch A: **Self-Organizing Neural Networks for Visualization and Classification**. In *Information and Classification*. Edited by Opitz O, *et al*. Berlin: Springer; 1993:307–313.
28. Costa JAF, Andrade Netto ML: **Segmentação de mapas auto-organizáveis com espaço de saída 3-D.** *Sba Controle & Automação [online]*. 2007, **18**(2):150–162. doi:ISSN 0103-1759. http://dx.doi.org/10.1590/S0103-17592007000200002.
29. Mamdani EH: **Application of fuzzy algorithms for control of simple dynamic plant.** *Procedings of IEEE* 1974, **121**(12):1585–1588.
30. European Communities. European guidelines for quality assurance in breast cancer screening and diagnosis: **Fourth Edition. European Comission.** *Luxembourg: Office for Official Publications of the European Communities* 2006, ISBN 92-79-01258-4.
31. Armitage P, Berry G: *Statistical Methods in Medical Research*. Blackwell: 3.ed. Londres; 1994.
32. Daniel WW: *Biostatistics: A foundation for Analysis in the Health Sciences*. New York: Wiley: 6.ed; 1995.
33. Haykin S: *Neural Networks: a comprehensive fundation*. 2nd edition. Prentice Hall: USA; 1999.