

Article

Novel Mathematical Model of Breast Cancer Diagnostics Using an Associative Pattern Classification

Raúl Santiago-Montero ¹, Humberto Sossa ^{2,3}, David A. Gutiérrez-Hernández ^{1,*}, Víctor Zamudio ¹, Ignacio Hernández-Bautista ⁴ and Sergio Valadez-Godínez ⁵

¹ Tecnológico Nacional de México/Instituto Tecnológico de León, León 37290, Guanajuato, Mexico; rsantiago066@gmail.com (R.S.-M.); v.zamudio@itleon.edu.mx (V.Z.)

² Instituto Politécnico Nacional (CIC), CD de México 07738, Mexico; hsossa@cic.ipn.mx

³ Tecnológico de Monterrey, Campus Guadalajara, Zapopan 45138, Jalisco, Mexico

⁴ Cátedra-CONACyT-Tecnológico Nacional de México/I.T. León, León 37290, Guanajuato, Mexico; ihernandezb@conacyt.mx

⁵ Universidad Humani Mundial, Campus San Francisco del Rincón, San Francisco del Rincón 37378, Guanajuato, Mexico; svaladezg@gmail.com

* Correspondence: david.gutierrez@itleon.edu.mx

Received: 27 January 2020; Accepted: 18 February 2020; Published: 1 March 2020



Abstract: Breast cancer is a disease that has emerged as the second leading cause of cancer deaths in women worldwide. The annual mortality rate is estimated to continue growing. Cancer detection at an early stage could significantly reduce breast cancer death rates long-term. Many investigators have studied different breast diagnostic approaches, such as mammography, magnetic resonance imaging, ultrasound, computerized tomography, positron emission tomography and biopsy. However, these techniques have limitations, such as being expensive, time consuming and not suitable for women of all ages. Proposing techniques that support the effective medical diagnosis of this disease has undoubtedly become a priority for the government, for health institutions and for civil society in general. In this paper, an associative pattern classifier (APC) was used for the diagnosis of breast cancer. The rate of efficiency obtained on the Wisconsin breast cancer database was 97.31%. The APC's performance was compared with the performance of a support vector machine (SVM) model, back-propagation neural networks, C4.5, naive Bayes, k-nearest neighbor (k-NN) and minimum distance classifiers. According to our results, the APC performed best. The algorithm of the APC was written and executed in a JAVA platform, as well as the experimental and comparativeness between algorithms.

Keywords: breast cancer detection (BCD); associative memory (AM); associative processing (AP); neural network (NN); pattern recognition (PR)

1. Introduction

Breast cancer is a disease in which a highly malignant type of tumor originates in breast cells. A tumor is an abnormal mass of body tissue. Tumors can be cancerous (malignant) or non-cancerous (benign). In general, tumors occur when cells divide and multiply excessively in the body. Normally, the body controls the division and growth of cells. New cells are created to replace old ones or to perform new functions. Cells that are damaged or are no longer needed die to give way to healthy replacement cells. If the balance of cell division and death is disturbed, a tumor may form. Breast cancer can be of the invasive or non-invasive type, and can occur in both men and women, although in men it is a hundred times less common than in women [1]. The risk factors for developing breast

cancer are many. The most important factor is related to gender, followed by age, obesity, physical activity, diet, alcohol consumption [2] and vitamin D concentration. Although vitamin D has emerged as a potentially important determinant of breast cancer, information is still scarce. Some studies show that it can be a risk factor [3–8], while others have shown that it is not [9–12]. To date, the exact reasons for breast cancer development are unknown [1].

Worldwide, every twenty seconds a new case of breast cancer is diagnosed. Only 10% of the cases are detected at initial stages [13]. Breast cancer is the second leading cause of death in women and this number is increasing [14]. For example, in terms of U.S.A. statistics, about 1 in 8 U.S.A. women (about 12%) will develop invasive breast cancer over the course of her lifetime. In 2020, an estimated 276,480 new cases of invasive breast cancer are expected to be diagnosed in women in the U.S.A., along with 48,530 new cases of non-invasive (in situ) breast cancer. About 2620 new cases of invasive breast cancer are expected to be diagnosed in men in 2020. A man's lifetime risk of breast cancer is about 1 in 883. About 42,170 women in the U.S.A. are expected to die in 2020 from breast cancer. Death rates have been steady in women under 50 since 2007 but have continued to drop in women over 50. The overall death rate from breast cancer decreased 1.3% per year from 2013 to 2017. These decreases are thought to be the result of treatment advances and earlier detection through screening [15–18].

A successful diagnosis in the early stages of breast cancer allows for better treatment, thereby increasing the probability of the person's survival. The cost of breast cancer treatment is high, especially at advanced stages of the disease due to the late diagnosis [19,20].

Mammography is the most commonly used method for the diagnosis and detection of breast cancer but has several disadvantages [21]. One disadvantage is that up to 20% of false negative results are obtained from the tests. Also, false positive results are directly dependent on the radiologist's opinion. There is also a risk of over-diagnosis which results in an excess of treatment. Mammograms require a small amount of radiation exposure that, if done repeatedly, could provoke cancer [22–24].

Another widely used method for the diagnosis of cancer is fine needle aspiration cytology (FNAC) [25]. The procedure of this method consists of extracting, through a needle, a sample of blood from the area affected by the cancer, and then analyzing it under a microscope. Then, according to the different characteristics of the cells, the specialist must decide whether the cancer cells are malignant or benign. However, this decision is not easy to make, and you usually choose to get a second opinion. In addition, computer information processing takes time, which leads to a demanding computational expense.

The area of computer science that is used to make an automatic classification is pattern recognition. Two of the main tasks in pattern recognition are classification and prediction. Among the most current and widespread techniques are artificial neural networks. These techniques are inspired by the behavior of biological neurons, simulating the process of their learning process. This computational model requires a set of descriptions of the classes or types to classify. The set of descriptions should be labeled to generalize the classification process [26–31].

Many methods for the diagnosis of breast cancer have been described in the literature. In [32], for example, the authors introduce a method based on associative memories for medical diagnosis including the diagnosis of breast cancer. In [33], the authors present a comparative study between several training methods of neural networks with the same objective: diagnosis of breast cancer. In [34], the authors describe another algorithm with the objective of combining a set of association rules and an artificial neural network. In [35], the researchers describe two methods, analyzed in artificial neural networks, for the diagnosis and prognosis of breast cancer. In [36], the authors combine neural networks and decision trees to solve the same problem. In summary, in [37], the researchers propose an evolutionary algorithm applied to the diagnosis of breast cancer.

In this article, we describe a classification method and use a set of numerical descriptions of patients with and without cancer. This process could help a specialist make decisions about the diagnosis of breast cancer or bi-class classification tasks in general. The simplicity of APC operations allows rapid classification to be applied to massive databases or applied in real-time processes. In

addition, it does not require any prior processing of the database to extract important features. The classifier does not need to be trained with an extensive or balanced database. As we will see, a few samples (less than 10%) are sufficient to obtain a well-trained classifier with good results. Noise tolerance is another notable feature of the classifier. The algorithm generates two decision regions where, up to now, the most distorted versions of a given pattern are classified without any problem, provided they do not fall into the neutral zone generated by the APC.

2. Theoretical Description

Classes are natural states of objects associated with concepts [29]. We will use the letter m to define the number of classes denoted as $\{c_i \in \Omega | i = 1, 2, \dots, m\}$, where Ω is the set of all classes, known as the interpretation space. Features by which objects are characterized are known as the space representation. The goal of supervised classification is to find an inductive hypothesis in the representation space that corresponds to the structure of the interpretation space [38]. In other words, the goal is to find a pattern classifier algorithm that allows the division of the interpretation space into different regions, so that the set of known patterns can be separated in the n -dimensional space and unknown patterns can be classified. It has been shown that this can be done using associative memories. Associative memories allow pattern classification by associating them with a class or a region.

2.1. Associative Memories

An associative memory is a single layered neural network that allows researchers to map input patterns \mathbf{x}^k to output patterns \mathbf{y}^k , such that each pattern \mathbf{x}^k is associated with a pattern and \mathbf{y}^k [39]. The formulas $\mathbf{x}^k \in \mathbf{X}^n \forall k \in \{1, 2, \dots, p\}$, $\mathbf{y}^k \in \mathbf{Y}^m \forall k \in \{1, 2, \dots, p\}$, and k are an index that represents a specific pair of associated patterns: n and m are the dimensionality of \mathbf{x}^k and \mathbf{y}^k , respectively; p is the cardinality of the set of patterns; and \mathbf{X} and \mathbf{Y} are any two sets. An associative memory M can be represented as follows:

$$\mathbf{x}^k \rightarrow M \rightarrow \mathbf{y}^k \quad (1)$$

Memory M is a correlation matrix of the p associations [40], whose fundamental set of associations is represented as:

$$S = \{(\mathbf{x}^k, \mathbf{y}^k) | k = 1, 2, \dots, p\} \quad (2)$$

During the learning process of memory M , each pair $(\mathbf{x}^k, \mathbf{y}^k) \in S(\mathbf{x}^k, \mathbf{y}^k) \in S$ is presented with the associative memory. During the recovery process, an input pattern \mathbf{x}^w is presented with the input of the already trained memory M . If $\mathbf{x}^k = \mathbf{x}^w$ for all $k \in \{1, 2, \dots, p\}$, then the associative memory operates in an auto-associative way; otherwise, if, for at least one k , $\mathbf{x}^k \neq \mathbf{x}^w$, then the memory operates in hetero-associative way [41].

2.2. Associative Classification of Patterns

In [42–44], the authors propose an APC that combines the learning association rule of Anderson–Kohonen–Nakano’s linear associator (LA) [41,45,46] and the recovery rule of the Lernmatrix (LM) [47,48]. An APC has two advantages over an LA and an LM: (1) An APC classifier allows operation with real-valued vectors, eliminating the disadvantage of the Lernmatrix classifier that operate only with binary-valued vectors; and (2) APCs remove the orthogonality restriction on the fundamental set S of the linear associator [49], as well as the restriction that the number p of patterns of the fundamental set is small with respect to the dimension n of the input patterns \mathbf{x}^k [50,51]. It is worth mentioning that the minimum size for the training set at which an APC’s performance is stable is about 10% the size of the class with the smallest number of instances [52]. The following are given:

1. A fundamental set of associations:

$$S = \{(\mathbf{x}^k, \mathbf{y}^k) | k = 1, 2, \dots, p\} \quad (3)$$

where $\mathbf{x}^k \in \mathbb{R}^n$ is the set of input patterns, $\mathbf{y}^k \in \{0, 1\}^m$ is the set of output patterns, n is the dimension of \mathbf{x}^k , m is the dimension of \mathbf{y}^k , and p is the cardinality of S .

- The class $c \in \{1, 2, \dots, m\}$ to which each input pattern \mathbf{x}^k belongs is defined as:

$$\mathbf{y}_j^k = \begin{cases} 1 & \text{for } j = c \\ 0 & \text{for } j = 1, 2, \dots, c-1, c+1, \dots, m \end{cases} \quad \forall k \in \{1, 2, \dots, p\} \quad (4)$$

The steps for learning the APC are as follows:

- Compute the average vector as

$$\bar{\mathbf{x}} = \frac{1}{p} \sum_{k=1}^p \mathbf{x}^k \quad (5)$$

- Translate all the patterns of the fundamental set with respect to the mean vector as

$$\mathbf{x}_t^k = \mathbf{x}^k - \bar{\mathbf{x}} \quad (6)$$

- Build matrix M as

$$M = \sum_{k=1}^p \mathbf{y}^k [\mathbf{x}_t^k]^t \quad (7)$$

For recovery by means of the APC, the below steps should be followed (given the key pattern $\mathbf{x}^\omega \in \mathbb{R}^n$).

- Translate \mathbf{x}^ω as

$$\mathbf{x}_t^\omega = \mathbf{x}^\omega - \bar{\mathbf{x}} \quad (8)$$

Perform the following product

$$\mathbf{z}^\omega = M \mathbf{x}_t^\omega \quad (9)$$

- Compute the components of class vector $\mathbf{y}^\omega \in \{0, 1\}^m$ as

$$\mathbf{y}_j^\omega = \begin{cases} 1 & \text{if } \mathbf{z}_j^\omega = \mathbf{V}_{h=1}^p \mathbf{z}_h^\omega \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

Finally, find the index class to which $\mathbf{x}^\omega \in \mathbb{R}^n$ belongs as the position j in vector \mathbf{y}_j^ω , where $\mathbf{y}_j^\omega = 1$.

2.3. Numerical Example

To understand the operation of the APC a numerical example is given next. Suppose we are given the following set of associations:

$$\mathbf{x}^1 = \begin{pmatrix} 6 \\ 5 \\ 2 \end{pmatrix}, \mathbf{y}^1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}; \quad \mathbf{x}^2 = \begin{pmatrix} -4 \\ 11 \\ -8 \end{pmatrix}, \mathbf{y}^2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

In this case $p = 2$, $n = 3$, and $m = 2$.

Construction of the association matrix is according to the discussed material.

1. Computation of the average vector is

$$\bar{\mathbf{x}} = \frac{1}{p} \sum_{k=1}^p \mathbf{x}^k = \frac{1}{2} \left[\begin{pmatrix} 6 \\ 5 \\ 2 \end{pmatrix} + \begin{pmatrix} -4 \\ 11 \\ -8 \end{pmatrix} \right] = \begin{pmatrix} 1 \\ 8 \\ -3 \end{pmatrix}$$

2. Translation of the input patterns is

$$\mathbf{x}_t^1 = \mathbf{x}^1 - \bar{\mathbf{x}} = \begin{pmatrix} 6 \\ 5 \\ 2 \end{pmatrix} - \begin{pmatrix} 1 \\ 8 \\ -3 \end{pmatrix} = \begin{pmatrix} 5 \\ -3 \\ 5 \end{pmatrix}, \quad \mathbf{x}_t^2 = \mathbf{x}^2 - \bar{\mathbf{x}} = \begin{pmatrix} -4 \\ 11 \\ -8 \end{pmatrix} - \begin{pmatrix} 1 \\ 8 \\ -3 \end{pmatrix} = \begin{pmatrix} 5 \\ -3 \\ 5 \end{pmatrix}.$$

3. Construction of matrix M is

$$M = \sum_{k=1}^p \mathbf{y}^k [\mathbf{x}_t^k]^t = \begin{pmatrix} 1 \\ 0 \end{pmatrix} (5 \ -3 \ 5) + \begin{pmatrix} 0 \\ 1 \end{pmatrix} (-5 \ 3 \ -5) = \begin{pmatrix} 5 & -3 & 5 \\ -5 & 3 & -5 \end{pmatrix}$$

Classification of the input pattern is $\mathbf{x} = \begin{pmatrix} 6 \\ 5 \\ 2 \end{pmatrix}$. We can see that this vector is a non-distorted version of vector \mathbf{x}^1 .

1. Translation of the vector is

$$\mathbf{x}_t^1 = \mathbf{x}^1 - \bar{\mathbf{x}} = \begin{pmatrix} 6 \\ 5 \\ 2 \end{pmatrix} - \begin{pmatrix} 1 \\ 8 \\ -3 \end{pmatrix} = \begin{pmatrix} 5 \\ -3 \\ 5 \end{pmatrix}$$

2. The product is gotten by

$$\mathbf{z}^\omega = M \mathbf{x}_t^\omega = \begin{pmatrix} 5 & -3 & 5 \\ -5 & 3 & -5 \end{pmatrix} \begin{pmatrix} 5 \\ -3 \\ 5 \end{pmatrix} = \begin{pmatrix} 59 \\ -59 \end{pmatrix}$$

3. Computation of the class vector \mathbf{y}^ω is

$$\mathbf{y}^\omega = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

4. The index class of vector \mathbf{x} is found according to the above discussion. Vector $\mathbf{x} = \begin{pmatrix} 6 \\ 5 \\ 2 \end{pmatrix}$ should be classified into class number one.

Suppose we are now given a distorted version of the first vector as follows: $\mathbf{x} = \begin{pmatrix} 4 \\ 7 \\ -1 \end{pmatrix}$. Let us find again the class in which this vector should be put.

1. Translation of the vector is

$$\mathbf{x}_t^2 = \mathbf{x}^2 - \bar{\mathbf{x}} = \begin{pmatrix} 4 \\ 7 \\ -1 \end{pmatrix} - \begin{pmatrix} 1 \\ 8 \\ -3 \end{pmatrix} = \begin{pmatrix} 3 \\ -1 \\ 2 \end{pmatrix}$$

2. The product is gotten by

$$\mathbf{z}^\omega = M\mathbf{x}_t^\omega = \begin{pmatrix} 5 & -3 & 5 \\ -5 & 3 & -5 \end{pmatrix} \begin{pmatrix} 3 \\ -1 \\ 2 \end{pmatrix} = \begin{pmatrix} 28 \\ -28 \end{pmatrix}$$

3. Computation of the class vector \mathbf{y}^ω is

$$\mathbf{y}^\omega = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

4. The index of the class of vector \mathbf{x} is found according to the above discussion. Vector $\mathbf{x} = \begin{pmatrix} 4 \\ 7 \\ -1 \end{pmatrix}$ should be classified into class number one.

From this very simple example, note the case of two class problems.

1. Learning and translation of the two input vectors provokes that they become the negative of each other. This is $\mathbf{x}^2 = -\mathbf{x}^1$. Due to the fact that the output vectors for the two classes are orthogonal, matrix M will be composed of \mathbf{x}^1 and its negative. This is $M = \begin{pmatrix} \mathbf{x}^1 \\ -\mathbf{x}^1 \end{pmatrix}$. Note that between the two vectors there is a *neutral position*; this corresponds to vector $\mathbf{x} = (0 \ 0 \ \dots \ 0)^t$.
2. Classification of a non-distorted version of any of the input vectors' translation provokes that it is first transformed to its translated original version. Multiplication of the association matrix M will always give a maximum value at the index class of the input vector.
3. Classification of a distorted version of any of the input vectors' translation provokes that it be first moved to one of the translated original versions. The moved vector could appear on one side or the other side of its corresponding translated original version. While the noise added to the input vector does not cause that its translated version does not surpasses the neutral position, the input vector will always be correctly classified. Of course, if translation of the input vector produces $\mathbf{x}_t^\omega = (0 \ 0 \ \dots \ 0)^t$, then the class of the vector cannot be found because $\mathbf{z}^\omega = M\mathbf{x}_t^\omega = (0 \ 0 \ \dots \ 0)^t$.

Next, we discuss the details of the database used to test the performance of the APC. We also give a few words about the set of classifiers with which the APC is compared.

2.4. Wisconsin Breast Cancer Database

This database was compiled by Dr. William H. Wolberg at the hospitals of the University of Wisconsin, Madison [53]. We obtained the database from the pattern recognition database repository of the University of California, Irvine (UCI) [54]. It is a compilation of breast tumor cases compiled from 1989 to 1990 by FNAC. It contains 699 instances of which 458 (65.5%) belong to the class "benign" and 241 (34.5%) to the class "malignant". Each event consists of 9 cytological features: (1) clump thickness, indicating grouping of cancer cells in multilayer; (2) uniformity of cell size, indicating metastasis to lymph nodes; (3) uniformity of cell shapes, identifying cancerous cells of varying size; (4) marginal adhesion, suggesting loss of adhesion, i.e., a sign of malignancy but the cancerous cells lose this property so this retention of adhesion is an indication of malignancy; (5) single epithelial cell size (SECS), if the SECS become larger, it may be a malignant cell; (6) bare nuclei, without cytoplasm coating, found in benign tumors; (7) bland chromatin, usually found in benign cells; (8) normal nucleoli, generally very small in benign cells; (9) mitoses, the process in cell division by which the nucleus divides.

Table 1 shows the range of values for each feature, valued on a scale of 1 to 10, with 1 being the closest to "benign" and 10 being the most anaplastic [53,55]. Moreover, the mean and standard

deviation of each cytological characteristic is included. The classes that form the database, “benign” and “malignant” are not linearly separable.

Table 1. Details of the attributes of the Wisconsin database.

#	Description	Type	Range of Values	Mean	Standard Deviation
1	Clump thickness	Numeric	1–10	4.418	2.816
2	Uniformity of cell size	Numeric	1–10	3.134	3.051
3	Uniformity of cell shapes	Numeric	1–10	3.207	2.972
4	Marginal adhesion	Numeric	1–10	2.807	2.855
5	Single epithelial cell size	Numeric	1–10	3.216	2.214
6	Bare nuclei	Numeric	1–10	3.545	3.644
7	Bland chromatin	Numeric	1–10	3.438	2.438
8	Normal nucleoli	Numeric	1–10	2.867	3.054
9	Mitoses	Numeric	1–10	1.589	1.715
10	Class	Nominal	Benign, Malignant		
Class Distribution			Benign: 458 (65.5%) Malignant: 241 (34.5%)		
Number of missing values			16		
Number of instances			699		

Before becoming publicly available, the dataset had 701 points. In January of 1989, after being revised, 2 instances from Group 1 were considered inconsistent and were removed from the dataset. Two more revisions occurred before the actual state of the dataset, both aimed to substitute values from zero to one, so the value range of the features is 1–10.

The data can be considered ‘noise-free’ [28] and has 16 missing values, which are the bare nuclei for 16 different instances, from Group 1 to 6. Table 1 is a summary of the state of the dataset used in this paper.

2.5. Minimum Distance Classifier

The minimum distance classifier (MDC) determines a given pattern belongs to a class by finding the nearest class in which the pattern can be put. This is normally done by computing a distance to each class representative [27–29].

2.6. Naïve Bayes

The naive Bayes (NB) is a classification algorithm that makes use of Bayes’ theories. The NB classifier assumes that the presence (or absence) of a particular class feature is not related to the presence (or absence) of any other feature, given the class variable [28,56,57].

2.7. K-Nearest Neighbor Classifier

The k-NN (k-nearest neighbors) [58] is a kind of minimum distance classifier, where for each class a sample is taken to establish to which class the pattern should be assigned. It is called nearest neighbor because the feature vector whose distance is less than the distance of the remaining vectors in the space of samples will determine the class in which the input vector should be put.

2.8. Back-Propagation

The back-propagation algorithm (BP) allows adjusting the weights of a neural network (NN) with the aim of finding a hyperplane or a set of hyperplanes that divide the interpretation space into different regions. This algorithm uses the gradient descent method to minimize the square error

between the network's output and the desired output [38]. Depending on the problem to be solved, the NN is configured with a number of connections, layers, input neurons and output neurons.

2.9. Support Vector Machine (SVM)

An SVM calculates a set of hyperplanes of separation in a high dimensional space. Hyperplanes have a maximum separation distance to the points (support vector) closest to them [59].

2.10. C4.5

This algorithm generates a decision tree for classification. It is based on its predecessor: the ID3 algorithm [28].

2.11. Comparison

A comparison among the six classifiers described in the previous section (MDC, NB, k-NN (k = 1, 2 and 3), BP, SVM and C4.5), and the ACP classifier was performed. We used the Wisconsin breast cancer database without removing the main attributes. Six experiments were conducted based on holdout validation as well as a validation experiment with 10 folds, both in a stratified manner. Experiments for holdout were 1%–99%, 10%–90%, 30%–70%, 50%–50%, and 70%–30% of the training-test, respectively. Each test for holdout was repeated one hundred times; an average classification performance was obtained. For 10-fold cross validation, ten repetitions were done, and an average performance was obtained.

3. Experimental Results

Table 2 presents a summary of the holdout validation tests. One can see that in all cases the APC obtained the best classification performance. Its performance increases as the number of training patterns is incremented.

Table 2. Summary of holdout classification.

Training-Test	Classifier								
	ACP	MDC	NB	1-NN	2-NN	3-NN	BP	SVM	C4.5
1%–99%	96.39	94.33	92.23	95.44	88.78	91.25	66.44	94.00	93.87
10%–90%	97.16	96.06	95.60	94.96	93.52	95.69	95.52	96.20	91.93
30%–70%	97.14	95.98	96.06	95.44	94.25	96.27	96.08	96.56	93.87
50%–50%	97.31	96.05	96.15	95.62	94.84	96.55	96.37	96.75	94.32
70%–30%	97.31	96.31	96.36	95.65	95.12	96.84	96.59	96.93	94.68

Table 3 shows a summary of the 10-fold test where APC obtained 97.13% correct classification.

Table 3. Summary of classification results for 10-fold cross validation.

Validation	Classifier								
	ACP	MDC	NB	1-NN	2-NN	3-NN	BP	SVM	C4.5
10-Folds	97.13	95.91	96.07	95.28	94.81	96.60	96.40	96.62	95.01

Figure 1 shows the algorithm that has been used in this paper.


```

public class Aprendizaje {
    public static void memoryObt(double [][] mPatterns, double [][] classes, double [][] memoryC){
//number of classes == number of patterns
        for(int row_mPatterns=0; row_mPatterns<mPatterns.length; row_mPatterns++){
            //multiplied each feature pattern by each value class
            for(int column_mclasses=0; column_mclasses<classes[0].length; column_mclasses++){
                for(int column_mPatterns=0; column_mPatterns<mPatterns[0].length; column_mPatterns++){
                    memoryC[column_mclasses][column_mPatterns]
=memoryC[column_mclasses][column_mPatterns]+
                        mPatterns[row_mPatterns][column_mPatterns] * classes[row_mPatterns][column_mclasses];
                }
            }
        }
    }
}

//-----
public static void medianVectorCal(double [][] mPatterns, double [] median_vector){
//this method obtain the median vector of the training vector
    int num_rows= mPatterns.length;
    int num_columns=mPatterns[0].length;
    for(int row=0; row<num_rows; row++){
        for(int column=0; column<num_columns; column++){
            median_vector[column]=median_vector[column]+mPatterns[row][column];
//
                System.out.println("value pattern "+ row + " "+ column + " = "+
median_vector[column]+column);
            }//for column
        }//for row
        for(int column=0; column<median_vector.length; column++){
            median_vector[column]=median_vector[column]/(double)num_rows;
        }//for column
    }//medianVectorCal
}

//-----
public static void traslationVector(double [][] mPatterns, double [] median_vector){
    int num_rows= mPatterns.length;
    int num_columns=mPatterns[0].length;
    for(int row=0; row<num_rows; row++){
        for(int column=0; column<num_columns; column++){
            mPatterns[row][column]=mPatterns[row][column]-median_vector[column];
            System.out.println("value pattern "+ row + " "+ column + " = "+ mPatterns[row][column]);
        }//for column
    }//for row
}
}

```

Figure 1. Algorithm employed for the APC classification.

4. Conclusions

The APC classifier is a one-shot machine learning technique with low computational cost and high efficiency, with bi-class classification problems. Our technique learns with a low number of instances in each class and it is not necessary for the database to be balanced.

The APC classifier is a simple and easy to implement method that makes use of associative memories for training and testing. The diagnosis of breast cancer by associative pattern classification

results in a simple and effective tool that can assist a user to make decisions concerning the prediction of breast cancer. Some methodologies that have been proposed in the literature need to extract the important features prior to training. The technique proposed in this paper does not require such a procedure. Our technique performed better than that of several well-known classifiers: support vector machines, the C4.5 algorithm based on decision trees, the naive Bayes, k-NN, and minimum distance.

Since the APC is an efficient classification algorithm in bi-class databases, it is suitable to be implemented in mobile applications can be used in short-term online diagnosis and support the process of mass population analysis. In turn, this research offers the possibility that, in the near future, software that accompanies the doctor can be developed so that, once the sample is obtained, it is characterized and included in the database being used. This would allow the specialist to obtain classification parameters on the sample and diagnose, depending on its classification, a particular breast cancer situation. This would be done quickly, backed by computer science and previously verified algorithms that offer certainty and support for decision making.

Author Contributions: Conceptualization, R.S.-M. and S.V.-G.; formal analysis, H.S. and I.H.-B.; investigation, R.S.-M., H.S., V.Z. and I.H.-B.; methodology, R.S.-M., H.S. and V.Z.; project administration, D.A.G.-H.; resources, D.A.G.-H. and S.V.-G.; software, I.H.-B. and S.V.-G.; supervision, V.Z.; validation, D.A.G.-H.; writing—original draft, H.S.; writing—review and editing, D.A.G.-H. All authors have read and agreed to the published version of the manuscript.

Acknowledgments: The authors would like to acknowledge the support obtained from CONACYT (under funding 65, Fronteras de la Ciencia), IPN (under fundings numbers 20190007 and 20200630) and TecNM-ITL.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Seiler, A.; Chen, M.A.; Brown, R.L.; Fagundes, C.P. Obesity, dietary factors, nutrition, and breast cancer risk. *Curr. Breast Cancer Rep.* **2018**, *10*, 14–27. [[CrossRef](#)] [[PubMed](#)]
- Spei, M.E.; Samoli, E.; Bravi, F.; La Vecchia, C.; Bamia, C.; Benetou, V. Physical activity in breast cancer survivors: A systematic review and meta-analysis on overall and breast cancer survival. *Breast* **2019**, *44*, 144–152. [[CrossRef](#)] [[PubMed](#)]
- Welsh, J. Vitamin D and prevention of breast cancer 1. *Acta Pharmacol. Sin.* **2007**, *28*, 1373–1382. [[CrossRef](#)] [[PubMed](#)]
- Holick, M.F. Vitamin D deficiency. *N. Engl. J. Med.* **2007**, *357*, 266–281. [[CrossRef](#)] [[PubMed](#)]
- Cui, Y.; Rohan, T.E. Vitamin D, calcium, and breast cancer risk: A review. *Cancer Epidemiol. Prev. Biomark.* **2006**, *15*, 1427–1437. [[CrossRef](#)] [[PubMed](#)]
- Yao, S.; Kwan, M.L.; Ergas, I.J.; Roh, J.M.; Cheng, T.D.; Hong, C.C.; McCann, S.E.; Tang, L.; Davis, W.; Liu, S.; et al. Association of serum level of vitamin D at diagnosis with breast cancer survival: A case-cohort analysis in the pathways study. *JAMA Oncol.* **2017**, *3*, 351–357. [[CrossRef](#)]
- Andrade, F.O.; Hilakivi-Clarke, L. Nutrition and Breast Cancer Prevention. *Nutr. Cancer Prev.: Mol. Mech. Dietary Recommend.* **2019**, *21*, 368.
- Ratnadiwakara, M.; Rooke, M.; Ohms, S.J.; French, H.J.; Williams, R.B.H.; Li, R.W.; Zhang, D.; Lucas, R.M.; Blackburn, A.C. The SuprMam1 breast cancer susceptibility locus disrupts the vitamin D/calcium/parathyroid hormone pathway and alters bone structure in congenic mice. *J. Steroid Biochem. Mol. Biol.* **2019**, *188*, 48–58. [[CrossRef](#)]
- Shin, M.H.; Holmes, M.D.; Hankinson, S.E.; Wu, K.; Colditz, G.A.; Willett, W.C. Intake of dairy products, calcium, and vitamin D and risk of breast cancer. *J. Natl. Cancer Inst.* **2002**, *94*, 1301–1310. [[CrossRef](#)]
- Lin, J.; Manson, J.E.; Lee, I.M.; Cook, N.R.; Buring, J.E.; Zhang, S.M. Intakes of calcium and vitamin D and breast cancer risk in women. *Arch. Intern. Med.* **2007**, *167*, 1050–1059. [[CrossRef](#)]
- Anders, C.K.; Johnson, R.; Litton, J.; Phillips, M.; Bleyer, A. Breast cancer before age 40 years. In *Seminars in Oncology*; WB Saunders: London, UK, 2009; Volume 36, pp. 237–249.
- Song, D.; Deng, Y.; Liu, K.; Zhou, L.; Li, N.; Zheng, Y.; Hao, Q.; Yang, S.; Wu, Y.; Zhai, Z.; et al. Vitamin D intake, blood vitamin D levels, and the risk of breast cancer: A dose-response meta-analysis of observational studies. *Aging (Albany NY)* **2019**, *11*, 12708. [[CrossRef](#)] [[PubMed](#)]

13. Pisu, M.; Schoenberger, Y.M.; Herbey, I.; Brown-Galvan, A.; Liang, M.I.; Riggs, K.; Meneses, K. Perspectives on conversations about costs of cancer care of breast cancer survivors and cancer center staff: A qualitative study. *Ann. Intern. Med.* **2019**, *170* (Suppl. 9), S54–S61. [CrossRef] [PubMed]
14. Chang, L.; Weiner, L.S.; Hartman, S.J.; Horvath, S.; Jeste, D.; Mischel, P.S.; Kado, D.M. Breast cancer treatment and its effects on aging. *J. Geriatr. Oncol.* **2019**, *10*, 346–355. [CrossRef] [PubMed]
15. American Cancer Society. How Common Is Breast Cancer? Available online: <https://www.cancer.org/cancer/breast-cancer/about/how-common-is-breast-cancer.html> (accessed on 13 January 2020).
16. American Cancer Society. Cancer Facts & Figures 2020. Available online: <https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2020/cancer-facts-and-figures-2020.pdf> (accessed on 14 January 2020).
17. National Cancer Institute. BRCA Mutations: Cancer Risk and Genetic Testing. Available online: <https://www.cancer.gov/about-cancer/causes-prevention/genetics/brca-fact-sheet> (accessed on 30 January 2018).
18. American Cancer Society. Breast Cancer Risk Factors You Cannot Change. Available online: <http://www.cancer.org/cancer/breast-cancer/risk-and-prevention/breast-cancer-risk-factors-you-cannot-change.html> (accessed on 11 September 2019).
19. Knaul, F.M.; Nigenda, G.; Lozano, R.; Arreola-Ornelas, H.; Langer, A.; Frenk, J. Breast cancer in Mexico: An urgent priority. *Salud Pública de México* **2009**, *51* (Suppl. 2), S335–S344. [CrossRef] [PubMed]
20. Mohar, A.; Bargalló, E.; Ramírez, M.T.; Lara, F.; Beltrán-Ortega, A. Available resources for breast cancer treatment in Mexico. *Salud Pública de México* **2009**, *51* (Suppl. 2), S263–S269.
21. National Cancer Institute. 2012. Available online: <http://www.cancer.gov> (accessed on 17 December 2019).
22. Preston, D.L.; Mattsson, A.; Holmberg, E.; Shore, R.; Hildreth, N.G.; Boice, J.D., Jr. Radiation effects on breast cancer risk: A pooled analysis of eight cohorts. *Radiat. Res.* **2002**, *158*, 220–235. [CrossRef]
23. BEIR, V. *Committee to Assess Health Risks from Exposure to Low Levels of Ionizing Radiation*; National Research Council: Washington, DC, USA, 2006.
24. Miglioretti, D.L.; Lange, J.; van den Broek, J.J.; Lee, C.I.; van Ravesteyn, N.T.; Ritley, D.; Kerlikowske, K.; Fenton, J.J.; Melnikow, J.; de Koning, H.J.; et al. Radiation-induced breast cancer incidence and mortality from digital mammography screening: A modeling study. *Ann. Intern. Med.* **2016**, *164*, 205–214. [CrossRef]
25. Lamb, J.; Anderson, T.J.; Dixon, M.J.; Levack, P.A. Role of fine needle aspiration cytology in breast cancer screening. *J. Clin. Pathol.* **1987**, *40*, 705–709. [CrossRef]
26. Fukunaga, K. *Introduction to Statistical Pattern Recognition, 2nd ed*; Academic Press Professional, Inc.: San Diego, CA, USA, 1990.
27. Friedman, M.; Kandel, A. *Introduction to Pattern Recognition: Statistical, Structural, Neural and Fuzzy Logic Approaches, Series in Machine Perception and Artificial Intelligence*; World Scientific Publishing Company: London, UK, 1999.
28. Duda, R.; Hart, P.; Stork, D. *Pattern Classification, 2nd ed.*; Wiley Interscience: New York, NY, USA, 2001.
29. Marques de Sá, J.P. *Pattern Recognition, Concepts, Methods, and Applications*; Springer-Verlag: New York, NY, USA, 2002.
30. Bishop, C.M. *Pattern Recognition and Machine Learning (Information Science and Statistics)*; Springer: New York, NY, USA, 2006.
31. Theodoridis, S.; Koutroumbas, K. *Pattern Recognition, 4th ed.*; Academic Press: San Diego, CA, USA, 2009.
32. Aldape-Pérez, M.; Yañez-Marquez, C.; Camacho-Nieto, O.; Argüelles-Cruz, A.J. An associative memory approach to medical decision support systems. *Comput. Methods Programs Biomed.* **2011**, *106*, 287–307. [CrossRef]
33. Paulin, F.; Santhakumaran, A. Classification of Breast cancer by comparing Back propagation training algorithms. *Int. J. Comput. Sci. Eng.* **2011**, *3*, 327–332.
34. Karabatak, M.; Ince, M.C. An expert system for detection of breast cancer based on association rules and neural network. *Expert Syst. Appl.* **2011**, *36*, 3465–3469. [CrossRef]
35. 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. *Oncol. Rep.* **2006**, *15*, 975–981. [CrossRef] [PubMed]
36. Jerez-Aragonés, J.M.; Gómez-Ruiz, J.A.; Ramos-Jiménez, G.; Muñoz-Pérez, J.; Alba-Conejo, E. A combined neural network and decision trees model for prognosis of breast cancer relapse. *Artif. Intell. Med.* **2003**, *27*, 45–63. [CrossRef]

37. Peña-Reyes, C.A.; Sipper, M. Applying Fuzzy CoCo to Breast Cancer Diagnosis. *Proc. Congr. Evol. Comput.* **2000**, *2*, 1168–1175.
38. Mitchell, T.M. *Machine Learning*; McGraw-Hill: New York, NY, USA, 1997.
39. Palm, G.; Schwenker, F.; Sommer, F.T.; Strey, A. Neural Associative Memories. *Biol. Cyber.* **1993**, *36*, 39.
40. Prasad, B.D.C.N.; Krishna Prasad, P.E.S.N.; Yeruva, S.; Sita Rama Murty, P. A Study on Associative Neural Memories. *Int. J. Adv. Comput. Sci. Appl.* **2011**, *1*, 124–133.
41. Kohonen, T. Correlation Matrix Memories. *IEEE Trans. Comput.* **1972**, *21*, 353–359. [[CrossRef](#)]
42. Santiago-Montero, R.; Yañez-Marquez, C.; Diaz-de-Leon, J.L. Hybrid pattern classifier based on Steinbuch's Lernmatrix and Anderson-Kohonen's Linear Associator. *Res. Comput. Sci.: Pattern Recogn. Adv. Perspect.* **2002**, *1*, 449–460.
43. Santiago-Montero, R.; Yañez-Marquez, C.; Diaz-de-Leon, J.L. *Associative Pattern Classifier: Theoretical Advances, Technical Report*; Computing Science Center-IPN: CD de México, Mexico, 2002.
44. Santiago-Montero, R. Hybrid Pattern Classifier Based on Steinbuch's Lernmatrix and Anderson-Kohonen's Linear Associator. Master's Thesis, Computing Science Center-IPN: CD de México, Mexico, 2003.
45. Anderson, J.A. A simple neural network generating an interactive memory. *Math. Biosci.* **1972**, *14*, 197–220. [[CrossRef](#)]
46. Nakano, K. Associatron-A model for associative memory. *IEEE Trans. Syst. Man Cyber.* **1972**, *3*, 380–388. [[CrossRef](#)]
47. Steinbuch, K. Die Lernmatrix. *Kybernetik* **1961**, *1*, 36–45. [[CrossRef](#)]
48. Steinbuch, K.; Frank, H. Nichtdigitale Lernmatrizen als Perzeptoren. *Kybernetik* **1963**, *1*, 117–124. [[CrossRef](#)] [[PubMed](#)]
49. Hassoun, M.H. *Associative Neural Memories: Theory and Implementation*; Oxford University Press: New York, NY, USA, 1993.
50. Rosenfeld, E.; Anderson, J.A. *Neurocomputing: Foundations of Research*; Anderson, J.A., Rosenfeld, E., Eds.; MIT Press: Cambridge, MA, USA, 1988.
51. Hassoun, M.H. *Fundamentals of Artificial Neural Networks*, 1st ed.; MIT Press: Cambridge, MA, USA, 1995.
52. Soria-Alcaraz, J.A.; Santiago-Montero, R.; Carpio, M. One criterion for the selection of the cardinality of learning set used by the Associative Pattern Classifier. In Proceedings of the 2010 IEEE Electronics, Robotics and Automotive Mechanics Conference, Washington, DC, USA, 28 September–1 October 2010; pp. 80–84.
53. Wolberg, W.H.; Mangasarian, O.L. Multisurface Method of Pattern Separation for Medical Diagnosis Applied to Breast Cytology. *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 9193–9196. [[CrossRef](#)] [[PubMed](#)]
54. Frank, A.; Asuncion, A. {UCI} Machine Learning Repository. 2010. Available online: <http://archive.ics.uci.edu/ml> (accessed on 14 August 2019).
55. Dheeru, D.; Taniskidou, E.K. UCI Machine Learning Repository. Available online: <https://archive.ics.uci.edu/ml/index.php> (accessed on 28 January 2017).
56. Rish, I. An empirical study of the naive Bayes classifier. *IJCAI Workshop Empirical Methods Artif. Intell.* **2001**, *3*, 41–46.
57. Zhang, H. *The Optimality of Naive Bayes*; AAAI Press: Palo Alto, CA, USA, 2004.
58. Patrick, E.; Fischer, F.P., III. A generalized k-nearest neighbor rule. *Inf. Control* **1970**, *16*, 128–152. [[CrossRef](#)]
59. Abe, S. *Support Vector Machines for Pattern Classification (Advances in Pattern Recognition)*; Springer-Verlag New York, Inc.: New York, NY, USA, 2005.

