



## Application of the Lernmatrix tau[9] to the classification of patterns in medical datasets

Patricio Reyes León<sup>1</sup>, Julio César Salgado Ramírez<sup>1</sup>, José Luis Velázquez Rodríguez<sup>2</sup>

<sup>1</sup>Universidad Politécnica de Pachuca, México, [patricioreyesleon@micorreo.upp.edu.mx](mailto:patricioreyesleon@micorreo.upp.edu.mx), [csalgado@upp.edu.mx](mailto:csalgado@upp.edu.mx)

<sup>2</sup>Instituto Politécnico Nacional, Centro de Investigación en Computación, CDMX, México, [joseluisvelazquezro@gmail.com](mailto:joseluisvelazquezro@gmail.com)

### ABSTRACT

In health area, doctors are busy finding useful tools to pre-diagnose chronic diseases. Chronic diseases have specific behaviors that can be detected before a patient enters critical phases. If a chronic disease is diagnosed early, the patient is highly likely to have a good quality of life, which is why this research work was motivated. By health sector protocols it is not easy to obtain information from patients with some type of chronic disease, for this reason, this work makes use of databases with real information that has been collected by the scientific community, information that has been specifically treated to do research with pattern classification algorithms (also called intelligent computing algorithms). The information is disease-classified, has entering patterns and test patterns for intelligent computing algorithms to use it and measure their performance. The databases used in this work are KEEL and ICU.

This work will introduce how the new associative pattern classification algorithm Lernmatrix tau[9](LM- $\tau$ [9]) behaves before 9 of the most commonly used classification algorithms in the state of the art. The results will show that this intelligent computing algorithm, despite its simplicity, is an algorithm that should be taken into account in the pre-diagnosis of in chronic diseases.

**Key words:** pre-diagnosis, chronic diseases, computational intelligence, intelligent pattern classification.

### 1. INTRODUCTION

According to the World Health Organization, chronic diseases such as cancer, diabetes, hypertension appear by environmental factors or are inherited [1-4]. Diseases like these have motivated researchers to focus on minimizing the negative effects of chronic diseases. An important example of this effort and of global interest is that researchers have made use of intelligent algorithms to make pre-diagnosis of such diseases. For health professionals, a pre-diagnosis increases a patient's life chances because they can be given adequate treatment [5]. Pre-diagnosis can provide useful information for the development of a chronic diseases cure [6]. As noted,

chronic diseases, which are of social impact to the population, is a research topic in the global scientific community [7].

Breast cancer is the most common cancer in women around the world. In Mexico, breast cancer is the second leading cause of death in women over the age of 20. Worldwide, is detected of breast cancer in a woman each 9 minutes and more than 60,000 14-year-old women have this type of cancer [8]. It is clear that chronic diseases are a world health concern.

Some researchers around the world have used intelligent computing algorithms as a possible useful diagnostic tool for chronic diseases with important results, however, some algorithms are like a "black box" that makes impossible to determine which patterns were misclassified and why [9], this reason does not demerit them. The algorithms used as classifiers for chronic diseases in the literature are: the Bayesian classifier [10], the K-NN [11], decision trees [12], neural networks, including deep learning [13-15], support vector machines (SVM) [16] and associative classification algorithms [17-30].

This work uses LM- $\tau$ [9] Lernmatrix associative classifier [31] as the basis for the classification of chronic diseases. LM- $\tau$ [9] is an improved algorithm of Karl Steinbuch's original model [32]. The results obtained will be compared with the most important intelligent computing algorithms offered by the WEKA platform [33].

Lujan Garcia and colleagues present their research in a learning method that classifies automatically transfer images chest X-ray to classify patients with pneumonia getting good results in performance metrics reported [34].

In a recent investigation on COVID-19, Lujan García presents a convolutional neural network with pre-trained weights, as a classification model for patients with COVID-19, pneumonia or who are healthy. Successfully classifying patients with SARS-CoV-2 and pneumonia [35].

This paper is structured as follows: in section 2 it shows some state-of-the-art works related to medical diagnosis, based on algorithmic solutions. Sections 3 and 4 describe the disease data banks and intelligent computing algorithms to be used. Section 5 sets out the experimental results of this research.

Finally, section 6 includes the conclusions and outlines some ideas for future work.

## 2. RELATED WORKS

Carrying out a medical pre-diagnosis of chronic diseases is a methodology accepted by the scientific community in the area of medicine, to face the growing increase in chronic diseases that affect the population, thus promoting an improvement in the quality of life of patients who, suffer or are predisposed to suffer some type of chronic disease. It is for this reason that various investigations have been proposed based on the use of algorithms to carry out this pre-diagnosis. Some of the more relevant research is briefly mentioned below.

In the investigation of [7], from Fazekeas, seasonal time series were used to perform the periodicity analysis of childhood lymphoid leukemia in Hungary, performing the analysis both for the total number of patients and for the data series divided by the mean, finding in this study a certain periodicity in the dates on which the diagnosis was made. Noting differences in the peaks of the time series, most of them appearing in winter, mainly in the younger group of patients.

In an approach to liver diseases, Abdar [6] in his research used data mining techniques to identify liver diseases using a new decision tree model, in combination with an impulse algorithm, obtaining a yield of 93.75% of accuracy, showing a marked improvement with traditional decision tree models, and they also show in their research that women are more likely to develop liver disease than men.

In [36] Mungle proposes for his research related to breast cancer, to use a hybrid clustering algorithm, k-means to quantify the proliferative index of cancer cells based on the Ki-67 nucleus count by means of images of stained cells with K-67. showing good results in their research.

Whereas Padmavathy proposes to use a neuro-fuzzy adaptive inference system (ANFIS) for the detection of breast cancer through images, achieving a performance of 98% in accuracy, as a performance metric of the classification model [37].

Recently González-Patiño and collaborators used segmented mammographic images, applying metaheuristics, to diagnose breast cancer in a timely manner. As a result, they observed an improvement in the error rate, compared to some classical methods such as the Otsu method [38].

Regarding prostate cancer, Guidi proposes using images to detect prostate cancer, identifying patients who could receive adaptive treatment, comparing two types of radiotherapy, which are: administered and planned. In this way, she was able to present in her research a tool capable of classifying patients with morphological variations and predicting

possible problems caused by the treatment administered or planned [39].

In the research proposed by Chang [40], a decision support web system was developed considering the performance of sensitivity, as well as previous decisions for the pre-diagnosis of chronic diseases, taking as a reference the Bayes theorem and expert opinions to perform decision making.

As an alternative, Vyas in his research refers that to understand a disease it is necessary to understand the molecular mechanisms, such as the amount of protein-protein interactions, focusing his efforts on diabetes mellitus, Vyas and his collaborators, proposed in his document [41] a model that is based on SVM, for the classification of structural and genomic fingerprints of this disease.

Golub mentions in his research [42] the importance of seeking a new approach to classify new types of cancer, for which he proposes in his study the use of genetic expressions of DNA microarrays, with the aim of classifying cancer in acute leukemias, be successful in identifying new types of leukemia cancer.

Instead, Polat focuses his research on chronic kidney diseases, showing that to obtain good classification performance it is necessary to use the correct feature selection algorithms, in order to reduce the dimensionality of the data sets. Therefore, he proposed to use an SVM model for the diagnosis, in conjunction with dimensionality reduction techniques [43].

## 3. DISEASE DATASETS

Below is a brief description of the disease databases used in this research during the experimental phase. These data sets were obtained from the KEEL repositories [44] and the University of California at Irvine (UCI) [45], these repositories have a compendium of data sets related to diseases that humans suffer from among other topics of interest scientific. In this research, the databases of:

Wisconsin (wsc)- Donated by Dr. William H. Wolberg of the University of Wisconsin and contains clinical cases of patients who underwent surgery for breast cancer. With a total of 683 patterns described by 9 and divided into 2 classes depending on whether the tumor was malignant or benign.

The Thyroid dataset was donated by the Garavan Institute in Australia to the UCI repository. It contains a total of 215 patient patterns, whether or not they present hyperthyroidism or hypothyroidism. This dataset was subsequently divided by the KEEL repository, from which the new-thyroid1 (nwt1) and new-thyroid2 (nwt2) data sets were obtained, used in this research to classify patients with hyperthyroidism and hypothyroidism respectively.

Haberman (hbrm) is a dataset donated by Tjen-Sien Lim in 1999 to the UCI repository, and contains case studies from the Billings Hospital of the University of Chicago, on patients who underwent surgery for breast cancer, with a total of 306 patterns described by 3 numerical attributes, with two classes for the classification of patients who survived or not surgery, within 5 years after having undergone it.

Spectfheart is a dataset from the UCI repository donated by the Lukasz A. Kurgan University of Colorado in Denver that contains 267 patterns described by 44 numerical attributes in order to detect cardiac abnormalities, according to the information from SPECT computed tomography images, in addition of the dataset classes are distributed in normal and abnormal.

Ecoli is a dataset from the UCI repository donated by KenetaNakai from the Institute of Biology and Molecular and Cellular at the University of Osaka for the localization of proteins in bacteria. Subsequently, the KEEL repository performed a division of the ecoli-0-3-4-7\_vs\_5-6 dataset which contains 257 patterns described by 7 attributes and two classes.

The Cleveland-0\_vs\_4 dataset (cve) belongs to the KEEL repository, taking data from the original ICU heart disease dataset, it has 177 patterns, 13 attributes and two classes, table 1 shows the above described.

**Table 1.** Datasets description

Dataset	Attributes	Patterns	Classes
Wisconsin	9	683	2
newt-thyroid1	5	215	2
newt-thyroid2	5	215	2
Haberman	3	306	2
Spectfheart	44	267	2
ecoli-0-3-4-7_vs_5-6	7	257	2
cleveland-0_vs_4	13	177	2

**4. CLASSIFIERS**

Subsection 4.1 will briefly explain 9 classification algorithms offered by the WEKA platform [33], algorithms that are the most commonly used in the state of the art and that were used in the experimental phase for the classification of chronic disease patterns.

Subsection 4. 2 will describe the new associative pattern classifier LM-τ[9] [17, 31], which is a particular case of the new minimalist machine learning paradigm [46].

**4.1 WEKA platform algorithms**

WEKA Platform Selected Pattern Classification Algorithms:

1. MultilayerPerceptron(MLP): MLP is an artificial neural network made up of multiple layers, its design attempts to solve classification problems with classes that are not linearly separable. MLP is considered by the scientific community to be an excellent pattern classifier, along with SVM is the "enemy to overcome" in all comparative studies. The MLP consists mainly of three layers: the input layer, the hidden layer and the output layer: In the latter layer are the neurons whose output values correspond to the class label [15-17].
2. RandomTree (RT): This algorithm is a pattern classification algorithm that consists of building a decision tree randomly [47].
3. RandomForest (RF): This classifier is a random forest algorithm, i.e., it is a combination of decision trees. It generates multiple trees randomly, and each decision tree casts a unit vote for the most popular class and that way it is possible to classify an input pattern [48].
4. NaiveBayes (NB): It is a probabilistic classifier based the Bayes naive theorem. This classifier considers independent all its attributes from the point of view probabilistic, contrary to what normally happens in the real world [10].
5. Instance Based 1 (IB1): This is the 1-NN classifier, which assigns a test pattern the class to which its nearest neighbor belongs [49].
6. IB3: It is classifier 3-NN. IB3 assigns to a test pattern the class that results by vote in the three nearest neighbors.
7. SMO: It is a support vector machine (SVM) algorithm. The algorithm for its operation uses function optimization and so-called support vectors. This type of algorithm tries to find a hyperplane that best separates the classes, and if it is not achieved, it uses the kernel trick to transform the patterns to a larger dimension space than the original. A kernel can be a linear, polynomial, radial-based, or sigmoid function [16].
8. J48: It is a decision tree classification algorithm. It is one of the most used in pattern classification. J48 is the name given in WEKA to a decision tree type derived from the old ID3. Decision trees are appreciated because they are explainable, they are based on graph theory and allow to display in structured form how instances of a dataset are

classified. The structure contains a root node at the top of the tree, and the intermediate nodes called sheets that correspond to the attributes. Classes are displayed at the bottom of the tree [12].

9. Logistic (LR): This algorithm is logistic regression, which is a statistical technique of machine learning. It acts as inputs actual values and makes a prediction about the probability that the input belongs to a particular class. This probability is calculated based on a sigmoid function whose expression involves the exponential function [50].

**4.2The New Algorithm AssociativeLM-τ[9]**

According to [32], in the learning phase of the original Lernmatrix model (LM), each binary input pattern  $\mathbf{x}^\mu$  of  $n$ -dimension is associated with an output  $\mathbf{y}^\mu$  pattern of type one-hot, so that if the input pattern belongs to class  $k$ , the  $\mathbf{y}^\mu$   $k$ -th component has value one, while all other components have zero value.

The learning phase creates an  $M$  array of zeros, with  $p$  rows and  $n$  columns, and for each pair in the learning set the  $M$  array updates its input values according to the following Steinbuch rule:

$$\Delta m_{ij} = \begin{cases} +\varepsilon \text{ si } y_i^\mu = 1 = x_j^\mu \\ -\varepsilon \text{ si } y_i^\mu = 0 \text{ y } x_j^\mu = 1 \\ 0 \text{ otherwise} \end{cases} \quad (1)$$

The only restriction for the value of  $\varepsilon$  is that it be positive. Therefore, it is valid to choose the value of  $\varepsilon$  as 1. In all the examples in this paper, we will use the value  $\varepsilon = 1$ .

If  $\mathbf{x}^\omega$  it is an  $n$ -dimensional pattern whose class is unknown, it is possible to estimate the class of that pattern in the recovery phase of the Lernmatrix model. To do this, the  $M$  matrix operates with the  $\mathbf{x}^\omega$  pattern to try to get the corresponding  $\mathbf{y}^\omega$  pattern. Because it is a  $\mathbf{y}^\omega$  one-hot  $p$ -dimensional pattern, the class information is in the nonzero component. This process is expressed like as:

$$y_i^\omega = \begin{cases} 1 \text{ si } \sum_{i=1}^n m_{ij} x_i^\omega = \text{MAX} \left[ \sum_{j=1}^n m_{hj} x_j^\omega \right] \\ 0 \text{ otherwise} \end{cases} \quad (2)$$

The class obtained is not always correct, because the LM has a problem called saturation that prevents the correct recovery of all one-shot patterns. For this reason, the LM is not competitive with classifiers reported in the state of the art due its saturation problem.

The new LM-τ[9]algorithm significantly improves the results of the original LM[44]. The LM for its functionality uses to

the original expressions (1) and (2) of the LM, the code Johnson-Möbius and the τ[9] transform.

The Johnson-Möbius code converts an array of real numbers into a set of binary strings with a very simple structure. Johnson-Möbius works like this:

1. Minimum value is obtained from a list of numbers.
2. Each number of the list is added with the negative of the minimum value obtained, in order to make it zero.
3. The new list of numbers is multiplied by an appropriate power of 10 for removing decimals.
4. Maximum value is obtained of the numbers in the list generated by step 3. Maximum number obtained indicates the number of bits to use for each number of the list. Each number in the list in step 3 indicates the number of 1 that will define it.

The following example illustrates the code Johnson-Möbius:

It has this list of 5 real numbers: 1.7, -0.1, 1.9, 0.2 and 0.6. is obtained the minimum value which is -0.1, the -0.1 is multiplied by -1 getting 0.1. Then, 0.1 is added with all numbers, generating the next list: 1.8, 0, 2.0, 3 and 7. The binary representation of this last list is shown in Table 2.

**Tabla 2.** Johnson-Möbius

1.7	00111111111111111111
-0.1	00000000000000000000
1.9	11111111111111111111
0.2	00000000000000000111
0.6	00000000000011111111

The LM-τ[9] transforms a binary digit into a binary digit pair as follows:

$$\begin{aligned} \tau^{[9]}(1) &= \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ \tau^{[9]}(0) &= \begin{pmatrix} 0 \\ 1 \end{pmatrix} \end{aligned} \quad (3)$$

The τ[9] transform, although simple, makes the LM significantly improve its performance, transforming the LM into a new competitive classification algorithm in the state of the art.

Like any associative memory, LM-τ[9] consists of two phases: learning phase and recall phase, which are defined as:

**Learning phase**

1. Apply the Johnson-Möbius code to  $x^\omega$
2. Apply the τ[9] transform to all  $x^\omega$  components obtained from the previous step.

3. Associate  $x^\omega$  with each transformed in step 2 a one-hot output pattern.
4. Apply the expression (1) to generate M.

**Recovery phase**

1. Apply the Johnson-Möbius code to  $x^\mu$ , where  $x^\mu$  is an unknown pattern
2. Apply  $\tau[9]$  to all components of the  $x^\mu$  transformed pattern obtained from the previous step.
3. Apply the expression (2) to get  $y^\omega$ .

**5. RESULTS**

The experimentation framework, to determine the performance of the algorithms, took into account the following elements: the **10** algorithms described in subsection 4.1, 4.2 and a computer with Windows 10 operating system, Intel (R) Core processor (TM) i5-7300HQ CPU at 2.4 GHz and 8 GB RAM.

**5.1 Validation Methods**

This research is based on supervised learning, which implies, define a set of learning patterns and a set of test patterns, using the database described in section 3. Both sets of patterns must form a partition. A partition is a method of validation, and the methods that stand out in the state of art are: bootstrap, hold-out, leave-one-out and, the most commonly used, k-cross-fold-validation [46]. When k-cross-fold-validation is applied, the most popular value of k is 10. However, because some of the data banks used in this investigation are unbalanced, it is recommended that K=5; that is, 5-cross-fold-validation method is recommended [51].

**5.2 Performance Measurement**

It is common for patterns to be grouped into two classes in disease databases. The Wisconsin database groups 683 patterns into two classes, malignant tumors and benign tumors. In the medical field, when someone has a disease, that person is said to be positive for that disease and otherwise it is a negative case. For the Wisconsin database, malignant tumors are positive class patterns, and benign tumors are negative class patterns.

When applying a classifier to a disease database with two classes, the result of presenting the classifier with a test pattern has the following possibilities: a). The test pattern is positive, and the result given by the classifier is positive, it is TRUE POSITIVE (TP), b). The test pattern is positive, and the result given by the classifier is negative, it is FALSE NEGATIVE (FN), c). The test pattern is negative, and the result given by the classifier is positive. is FALSE POSITIVE (FP) and d). The test pattern is negative, and the result given by the classifier is negative, it is TRUE NEGATIVE (TN).

The confusion matrix is a matrix array that results from applying a classifier to all test patterns. Figure 1 shows the confusion matrix.

		Prediction	
		Positive	Negative
Observation	Positive	TP	FP
	Negative	FN	TN

**Fig. 1.**Confusion matrix

TP and TN correspond to cases where the classifier succeeded, and FP and FN correspond to cases where the classifier made mistakes. Ideally, FP and FN must be zero.

Two performance measures are derived from the confusion matrix, which are Sensitivity and Specificity that are appropriate for unbalanced databases. Sensitivity is the number of patterns that the algorithm classified as positive divided by the total positive patterns contained in the database, as shown in (4). Specificity is the number of patterns that the algorithm classified as negative divided by the total negative patterns contained in the database, as shown in (5).

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{4}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{5}$$

The performance measure called Balanced Accuracy is defined based on (4) and (5), this measure is appropriate for unbalanced databases. Balanced Accuracy is defined in (6).

$$BA = \frac{\text{Sensitivity} + \text{Specificity}}{2} \tag{6}$$

If there are no classification errors the result is 1 (100%), that is, when FP and FN are zero (which is very rare). Depending on the next percentages will indicate whether BA is a good measure or not. Above 0.90 is excellent, between 0.81 and 0.90 is good, between 0.60 and 0.80 is regular and values less than 0.6 are very bad result.

**5.3 Results and Discussion**

This section presents the results of the experiments, using the 5-fold-cross-validation and BA validation method as a performance measure for the 10 classifiers. In addition, the best percentages of each database have been emphasized with boldface. All BA results are with two decimal places. Table 3 shows the above described.

**Table 3.** Results

Algorithm	Databases						
	wsc	nwt1	nwt2	hbrm	spec	ecoli	clv
NB	0.96	0.98	0.98	0.57	<b>0.77</b>	0.93	0.79
SVM	<b>0.97</b>	0.77	0.75	0.50	0.50	0.93	0.87
LR	0.96	0.96	0.96	0.54	0.62	0.92	0.67
MLP	0.95	0.95	0.95	0.58	0.66	0.92	0.76
J48	0.94	0.94	0.94	0.57	0.60	0.87	0.70
RT	0.92	0.98	0.91	0.58	0.65	0.76	0.66
RF	0.96	0.95	0.92	0.55	0.60	0.93	0.67
IB1	0.94	0.97	0.98	<b>0.64</b>	0.59	0.86	0.67
IB3	0.96	0.96	0.92	0.55	0.58	0.89	0.81
LM- $\tau$ [9]	0.96	<b>0.99</b>	<b>0.99</b>	0.52	0.62	<b>0.96</b>	<b>0.90</b>

It is observed that only 4 of the 10 pattern classification algorithms were the first place and these were: NB in the spec database, SVM in the wsc database, IB1 in the hbrm database and the LM- $\tau$ [9] was first in four databases: nwt1, nwt2,ecoli and clv. In addition, the LM- $\tau$ [9] was second place in wscand clv datasets, fourth place in spec database and penultimate place in hbrm database.The above results were obtained by BA measurement.

However, if only the sick is taken into account, i.e., the positive cases, because in reality those are the cases that interest doctors, so, the Sensitivity is the appropriate performance measure. So that LM- $\tau$ [9] is third place with a value of 0.31, thus, it is a competitive algorithm in this issue.

As noted, LM- $\tau$ [9], this new algorithm, despite its simplicity is a competitive algorithm compared to the classification algorithms most commonly used in the classification of chronic diseases.

**6. CONCLUSION AND FUTURE WORK**

This article presented the results of applying 10 pattern classification algorithms in chronic disease databases, KEEL and ICU databases. The databases of chronic diseases were proteins in bacteria, heart diseases ,breast cancer, heart diseases, hyperthyroidism and hypothyroidism. This experiment aimed to study the performance of these algorithms applied to the pre-diagnosis of some chronic diseases, diseases that seriously impact humans.

The best 9 classifiers chosen among the state of the art were compared with, a new algorithm in the state of the art, the associative pattern classifier LM- $\tau$ [9]. The results obtained from the experiments show the superiority of SVM performance, of the IB1 algorithm performance and, notably, of the new LM- $\tau$ [9] performance. The results obtained confirm, with a high degree of certainty, the usefulness of the classifiers used in this work in the pre-diagnosis of chronic diseases.

As future works, it is proposed to compare these results with the performances of deep learning algorithms and apply the 10 algorithms used in this work in other disease databases of different nature, as well as in databases in other areas of human activity, where intelligent computing has shown areas of scientific interest [52-88].

**ACKNOWLEDGEMENT**

The authors want to thank the Universidad Politécnica de Pachuca, the Instituto Politécnico Nacional, Mexico (Secretaría Académica, EST, CIDETEC, COFAA), the CONACYT and SNI for their support in the development of this work.

**REFERENCES**

1. González-Patiño, D., Villuendas-Rey, Y., Argüelles-Cruz, A. J., Camacho-Nieto, O., & Yáñez-Márquez, C. **AISAC: An Artificial Immune System for Associative Classification Applied to Breast Cancer Detection.**Applied Sciences, Vol. 10, no. 2, p. 515, Jan. 2020.
2. Uriarte-Arcia, A.V., I. López-Yáñez, and C. Yáñez-Márquez,**One-hot vector hybrid associative classifier for medical data classification.**PloS one, Vol. 9, no. 4, p. e95715, Apr. 2014.
3. Aldape-Pérez, M., Yáñez-Márquez, C., Camacho-Nieto, O., López-Yáñez, I., & Argüelles-Cruz, A. J, **Collaborative learning based on associative models: Application to pattern classification in medical datasets.** Computers in Human Behavior, Vol. 51, pp. 771–779, Oct. 2015.
4. Villuendas-Rey, Y., Alanis-Tamez, M. D., Rey-Benguría, C., Yáñez-Márquez, C., & Nieto, O. C, **Medical Diagnosis of Chronic Diseases Based on a Novel Computational Intelligence Algorithm.**J. UCS, vol. 24, no 6, p. 775-796, Jun. 2018.
5. Hay, S. I., Abajobir, A. A., Abate, K. H., Abbafati, C., Abbas, K. M., Abd-Allah, F., ... and Aboyans, V. **Global, regional, and national disability-adjusted life-years (DALYs) for 333 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016.** *The Lancet*, vol. 390, no 10100, pp. 1260-1344, Sep. 2017.

6. Abdar, M., Zomorodi-Moghadam, M., Das, R. & Ting, I. H. **Performance analysis of classification algorithms on early detection of liver disease.** Expert Systems with Applications, Vol. 67, pp. 239-251, Jan. 2017.
7. M. Fazekas. **Analysing Data of Childhood Acute Lymphoid Leukaemia by Seasonal Time Series Methods.** J. UCS, Vol. 12, no 9, pp. 1190-1195, Aug. 2006.
8. Castrezana Campos, M. D. R. **Geografía del cáncer de mama en México.** Investigaciones geográficas, no 93, pp. 0-0, Aug. 2017.
9. Rijo, R., Silva, C., Pereira, L., Gonçalves, D., and Agostinho, M. **Decision Support System to Diagnosis and Classification of Epilepsy in Children.** J. UCS, Vol. 20, no 6, pp. 907-923, Feb. 2014.
10. Bhargavi, P., & Jyothi, S. **Applying naive bayes data mining technique for classification of agricultural land soils.** International journal of computer science and network security, Vol. 9, no 8, pp. 117-122, Aug. 2009.
11. Cover, T., and Hart, P. **Nearest neighbor pattern classification.** IEEE transactions on information theory, vol. 13, no 1, pp. 21-27, May. 1967.
12. Quinlan, J. R. **Decision trees and decision-making.** IEEE Transactions on Systems, Man, and Cybernetics, Vol. 20, no 2, pp. 339-346, Mar. 1990.
13. McCulloch, W. S., & Pitts, W. **A logical calculus of the ideas immanent in nervous activity.** The bulletin of mathematical biophysics, Vol. 5, no 4, pp. 115-133. 1943.
14. Rosenblatt, F. **The perceptron: a probabilistic model for information storage and organization in the brain.** Psychological review, Vol. 65, no 6, p. 386 1958.
15. Rumelhart, D. E., Hinton, G. E., & Williams, R. J. **Learning internal representations by error propagation.** Published by Bradford books/MIT press, Cambridge MA. 1986.
16. Cortes, C., & Vapnik, V. **Support-vector networks.** Machine learning, Vol. 20, no 3, pp. 273-297 1995.
17. Yáñez-Márquez, C., López-Yáñez, I., Aldape-Pérez, M., Camacho-Nieto, O., Argüelles-Cruz, A. J., & Villuendas-Rey, Y. **Theoretical Foundations for the Alpha-Beta Associative Memories: 10 Years of Derived Extensions, Models, and Applications.** Neural Processing Letters, Vol. 48, no 2, pp. 811-847. Dec. 2018.
18. López-Yáñez, I., Yáñez-Márquez, C., Camacho-Nieto, O., Aldape-Pérez, M., & Argüelles-Cruz, A. J. **Collaborative learning in postgraduate level courses.** Computers in Human Behavior, Vol. 51, pp. 938-944, 2015.
19. López-Yáñez, I., Sheremetov, L., & Yáñez-Márquez, C. **A novel associative model for time series data mining.** Pattern Recognition Letters, Vol. 41, pp. 23-33. 2014.
20. Serrano-Silva, Y. O., Villuendas-Rey, Y., & Yáñez-Márquez, C. **Automatic feature weighting for improving financial Decision Support Systems.** Decision Support Systems, Vol. 107, pp. 78-8. Jan. 2018.
21. Lytras, M. D., Mathkour, H., Abdalla, H. I., Yáñez-Márquez, C., & De Pablos, P. O. **The Social Media in Academia and Education Research R-evolutions and a Paradox: Advanced Next Generation Social Learning Innovation.** J. UCS, Vol. 20, no 15, pp. 1987-1994 Dec. 2014.
22. Acevedo-Mosqueda, M. E., Yáñez-Márquez, C., & López-Yáñez, I. **Alpha-Beta bidirectional associative memories: theory and applications.** Neural Processing Letters, vol. 26, no 1, pp. 1-40. Jun. 2007.
23. García-Floriano, A., Ferreira-Santiago, Á., Camacho-Nieto, O., & Yáñez-Márquez, C. **A machine learning approach to medical image classification: Detecting age-related macular degeneration in fundus images.** Computers & Electrical Engineering, vol. 75, pp. 218-229. Nov. 2019.
24. Villuendas-Rey, Y., Rey-Benguría, C. F., Ferreira-Santiago, Á., Camacho-Nieto, O., & Yáñez-Márquez, C. **The naive associative classifier (NAC): a novel, simple, transparent, and accurate classification model evaluated on financial data.** Neurocomputing, Vol. 265, pp. 105-115. Mar. 2017.
25. Godínez, I. R., López-Yáñez, I., & Yáñez-Márquez, C. **Classifying patterns in bioinformatics databases by using Alpha-Beta associative memories.** In *Biomedical Data and Applications* Springer, Vol 224 pp. 187-210, 2009.
26. Moreno-Moreno, P., & Yáñez-Márquez, C. **The new informatics technologies in education debate.** In *World Summit on Knowledge Society* Springer, pp. 291-296. Sep. 2008.
27. Ramírez-Rubio, R., Aldape-Pérez, M., Yáñez-Márquez, C., López-Yáñez, I., & Camacho-Nieto, O. **Pattern classification using smallest normalized difference associative memory.** Vol. 93, pp. 104-112. Feb. 2017.
28. Uriarte-Arcia, A. V., López-Yáñez, I., Yáñez-Márquez, C., Gama, J., & Camacho-Nieto, O. **Data stream classification based on the gamma classifier.** Mathematical Problems in Engineering, Vol. 2015. Jul. 2015
29. García-Floriano, A., López-Martín, C., Yáñez-Márquez, C., & Abran, A. **Support vector regression for predicting software enhancement effort.** Information and Software Technology, Vol. 26, no 1, p. 1-40. Jan. 2018.
30. Acevedo, M. E., Yáñez-Márquez, C., & Acevedo, M. A. **Associative models for storing and retrieving concept lattices.** Mathematical Problems in Engineering, **Associative models for storing and retrieving concept lattices** Vol 2010 Jun 2010..
31. Velázquez-Rodríguez, J. L., Villuendas-Rey, Y., Camacho-Nieto, O., & Yáñez-Márquez, C. **A Novel and Simple Mathematical Transform Improves the Performance of Lernmatrix in Pattern**

- Classification.** Mathematics, Vol. 8, no 5, p. 732. May. 2020.
32. Steinbuch, K. **Die lernmatrix.** Kybernetik, vol. 1, no 1, pp. 36-45. 1961.
  33. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. **The WEKA data mining software: an update.** ACM SIGKDD explorations newsletter, 1vol. 11, no 1, pp. 10-18.2009.
  34. Luján-García, J. E., Yáñez-Márquez, C., Villuendas-Rey, Y., & Camacho-Nieto, O. **A Transfer Learning Method for Pneumonia Classification and Visualization.** Applied Sciences, Vol. 10, no 8, p. 2908. Apr. 2020.
  35. Luján-García, J. E., Moreno-Ibarra, M. A., Villuendas-Rey, Y., & Yáñez-Márquez, C. **Fast COVID-19 and Pneumonia Classification Using Chest X-ray Images.** Mathematics. Vol. 8, no 9, p. 1423. Aug. 2020.
  36. Mungle, T., Tewary, S., Arun, I., Basak, B., Agarwal, S., Ahmed, R. & Chakraborty, C. **Automated characterization and counting of Ki-67 protein for breast cancer prognosis: A quantitative immunohistochemistry approach.** Computer Methods and Programs in Biomedicine, Vol. 139, p. 149-161. Nov. 2017.
  37. Padmavathy, T. V., Vimalkumar, M. N., & Bhargava, D. S. **Adaptive clustering based breast cancer detection with ANFIS classifier using mammographic images.** Cluster Computing, vol. 22, no 6, pp. 13975-13984, Feb.2019.
  38. González-Patiño, D., Villuendas-Rey, Y., Argüelles-Cruz, A. J., & Karray, F. **A novel bio-inspired method for early diagnosis of breast cancer through mammographic image analysis.** Applied Sciences, Vol. 9, no 21, p. 4492. Oct. 2019.
  39. Guidi, G., Maffei, N., Vecchi, C., Gottardi, G., Ciarmatori, A., Mistretta, G. M., ... & Costi, T. **Expert system classifier for adaptive radiation therapy in prostate cancer.** Australasian physical & engineering sciences in medicine, Vol. 40, no 2, p. 337-348. Feb. 2017.
  40. Chang, C. C., Cheng, C. S., & Huang, Y. S. **A Web-Based Decision Support System for Chronic Diseases.** J. UCS, vol. 12, no 1, pp. 115-125. Jan. 2006.
  41. Vyas, R., Bapat, S., Jain, E., Karthikeyan, M., Tambe, S., & Kulkarni, B. D. **Building and analysis of protein-protein interactions related to diabetes mellitus using support vector machine, biomedical text mining and network analysis.** Computational biology and chemistry, Vol. 65, pp. 37-44. Sep. 2016.
  42. Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P. & Bloomfield, C. D. **Molecular classification of cancer: class discovery and class prediction by gene expression monitoring.** science, Vol. 286, no 5439, p. 531-537. Oct. 1999.
  43. Polat, H., Mehr, H. D., & Cetin, A. **Diagnosis of chronic kidney disease based on support vector machine by feature selection methods.** Journal of medical systems, Vol. 41, no 4, p. 55. Feb. 2017.
  44. Alcalá-Fdez, J., Fernández, A., Luengo, J., Derrac, J., García, S., Sánchez, L., & Herrera, F. **Keel data-mining software tool: data set repository, integration of algorithms and experimental analysis framework.** Journal of Multiple-Valued Logic & Soft Computing, Vol. 17, pp. 255–287. Apr. 2011.
  45. Dua, D. and Graff, C. **UCI Machine Learning Repository** [<http://archive.ics.uci.edu/ml>]. Irvine, CA: University of California, School of Information and Computer Science. 2019.
  46. Yáñez-Márquez, C. **Toward the Bleaching of the Black Boxes: Minimalist Machine Learning.** IT Professional, Vol. 22, no 4, pp. 51-56. Jul. 2020.
  47. Kalmegh, S. **Analysis of weka data mining algorithm reptree, simple cart and randomtree for classification of indian news.** International Journal of Innovative Science, Engineering & Technology, Vol. 2, no 2, pp. 438-446. 2015.
  48. Zhang, L., & Suganthan, P. N. **Random forests with ensemble of feature spaces.** Pattern Recognition, Vol. 47, no 10, p. 3429-3437. 2014.
  49. de Haro-García, A., Cerruela-García, G., & García-Pedrajas, N. **Instance selection based on boosting for instance-based learners.** Pattern Recognition, 96, 106959. 2019.
  50. Bootkrajang, J., & Kabán, A. **Learning kernel logistic regression in the presence of class label noise.** Pattern Recognition, Vol. 47, no 11, p. 3641-3655. May. 2014.
  51. López, V., Fernández, A., García, S., Palade, V., & Herrera, F. **An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics.** Information sciences, Vol. 250, p. 113-141. Jul. 2013.
  52. Y. Villuendas-Rey, C. Rey-Benguría, M. Lytras, C. Yáñez-Márquez, and O. Camacho-Nieto. **Simultaneous instance and feature selection for improving prediction in special education data,** *Program*, vol. 51, pp. 278-297, 2017.
  53. I. Roman-Godinez, I. Lopez-Yanez, and C. Yanez-Marquez. **A new classifier based on associative memories,** in *2006 15th International Conference on Computing*, 2006, pp. 55-59.
  54. E. Guzman, S. Alvarado, O. Pogrebnyak, and C. Yanez. **Image recognition processor based on morphological associative memories,** in *Electronics, Robotics and Automotive Mechanics Conference (CERMA 2007)*, 2007, pp. 260-265.
  55. C. Yáñez-Márquez, M. E. Cruz-Meza, F. A. Sánchez-Garfias, and I. López-Yáñez. **Using alpha-beta associative memories to learn and recall RGB images,** in *International Symposium on Neural Networks*, 2007, pp. 828-833.
  56. E. Guzmán, O. Pogrebnyak, C. Yañez, and L. P. S. Fernandez. **Morphological transform for image**



- compression**, *EURASIP Journal on advances in signal processing*, vol. 2008, p. 426580, 2008.
57. I. J. Rudas, I. Z. Batyrshin, A. H. Zavala, O. C. Nieto, L. Horváth, and L. V. Vargas. **Generators of fuzzy operations for hardware implementation of fuzzy systems**, in *Mexican International Conference on Artificial Intelligence*, 2008, pp. 710-719.
  58. I. J. Rudas, I. Z. Batyrshin, A. H. Zavala, O. C. Nieto, and L. V. Vargas. **Digital fuzzy parametric conjunctions for hardware implementation of fuzzy systems**, in *2009 IEEE International Conference on Computational Cybernetics (ICCC)*, 2009, pp. 157-166.
  59. A. H. Zavala, I. Z. Batyrshin, I. J. Rudas, L. V. Vargas, and O. C. Nieto. **Parametric operations for digital hardware implementation of fuzzy systems**, in *Mexican International Conference on Artificial Intelligence*, 2009, pp. 432-443.
  60. A. H. Zavala, O. C. Nieto, I. Batyrshin, and L. V. Vargas. **VLSI Implementation of a Module for Realization of Basic t-norms on Fuzzy Hardware**, in *2009 IEEE International Conference on Fuzzy Systems*, 2009, pp. 655-659.
  61. S. J. Lopez, O. C. Nieto, and J. I. C. Oria. **Non-parametric modeling of uncertain hyperbolic partial differential equations using pseudo-high order sliding mode observers**, *International Journal of Innovative Computing, Information and Control*, vol. 8, pp. 1501-1521, 2012.
  62. A. H. Zavala, I. Z. Batyrshin, O. C. Nieto, and O. Castillo. **Conjunction and disjunction operations for digital fuzzy hardware**, *Applied Soft Computing*, vol. 13, pp. 3248-3258, 2013.
  63. I. Salgado, O. Camacho, C. Yáñez, and I. Chairez. **Proportional derivative fuzzy control supplied with second order sliding mode differentiation**, *Engineering Applications of Artificial Intelligence*, vol. 35, pp. 84-94, 2014.
  64. I. Salgado, I. Chairez, O. Camacho, and C. Yáñez. **Super-twisting sliding mode differentiation for improving PD controllers performance of second order systems**, *ISA transactions*, vol. 53, pp. 1096-1106, 2014.
  65. C. Yáñez-Márquez, I. López-Yáñez, and G. d. I. L. S. Morales. **Analysis and prediction of air quality data with the gamma classifier**, in *Iberoamerican Congress on Pattern Recognition*, 2008, pp. 651-658.
  66. I. López-Yáñez, A. J. Argüelles-Cruz, O. Camacho-Nieto, and C. Yáñez-Márquez. **Pollutants time-series prediction using the Gamma classifier**, *International Journal of Computational Intelligence Systems*, vol. 4, pp. 680-711, 2011.
  67. L. Cleofas-Sánchez, V. García, R. Martín-Félez, R. Valdovinos, J. S. Sánchez, and O. Camacho-Nieto. **Hybrid associative memories for imbalanced data classification: an experimental study**, in *Mexican Conference on Pattern Recognition*, 2013, pp. 325-334.
  68. Y. Villuendas-Rey, M. M. Garcia-Lorenzo, and R. Bello. **Support Rough Sets for decision-making**, in *Fourth International Workshop on Knowledge Discovery, Knowledge Management and Decision Support*, 2013.
  69. Y. Villuendas-Rey and M. M. Garcia-Lorenzo. **Attribute and case selection for nn classifier through rough sets and naturally inspired algorithms**, *Computación y Sistemas*, vol. 18, pp. 295-311, 2014.
  70. C. Yáñez-Márquez, M. Aldape-Perez, I. López-Yáñez, and O. Camacho-Nieto. **Emerging computational tools: Impact on engineering education and computer science learning**, *International Journal of Engineering Education*, pp. 533-542, 2014.
  71. A. Ferreira-Santiago, C. Yanez-Marquez, I. Lopez-Yanez, O. Camacho-Nieto, M. Aldape-Perez, and A. J. Arguelles-Cruz. **Enhancing engineering education through link prediction in social networks**, *International Journal of Engineering Education*, pp. 1566-1578, 2016.
  72. S. Cerón-Figueroa, I. López-Yáñez, W. Alhalabi, O. Camacho-Nieto, Y. Villuendas-Rey, M. Aldape-Pérez, *et al.* **Instance-based ontology matching for e-learning material using an associative pattern classifier**, *Computers in Human Behavior*, vol. 69, pp. 218-225, 2017.
  73. A. García-Florian, A. Ferreira-Santiago, C. Yáñez-Márquez, O. Camacho-Nieto, M. Aldape-Pérez, and Y. Villuendas-Rey. **Social Web Content Enhancement in a Distance Learning Environment: Intelligent Metadata Generation for Resources**, *International Review of Research in Open and Distributed Learning*, vol. 18, pp. 161-176, 2017.
  74. S. Ortiz-Ángeles, Y. Villuendas-Rey, I. López-Yáñez, O. C. Nieto, and C. Yáñez-Márquez. **Electoral Preferences Prediction of the YouGov Social Network Users Based on Computational Intelligence Algorithms**, *J. UCS*, vol. 23, pp. 304-326, 2017.
  75. J. A. Antón-Vargas, Y. Villuendas-Rey, C. Yáñez-Márquez, I. López-Yáñez, and O. Camacho-Nieto. **Improving the performance of an associative classifier by Gamma rough sets based instance selection**, *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 32, p. 1860009, 2018.
  75. E. Barroso, Y. Villuendas, and C. Yanez. **Bio-inspired algorithms for improving mixed and incomplete data clustering**, *IEEE Latin America Transactions*, vol. 16, pp. 2248-2253, 2018.
  76. D. González-Patiño, Y. Villuendas-Rey, and A. J. Argüelles-Cruz. **The potential use of bioinspired algorithms applied in the segmentation of mammograms**, 2018.
  77. Y. Villuendas-Rey. **Maximal similarity granular rough sets for mixed and incomplete information systems**, *Soft Computing*, vol. 23, pp. 4617-4631, 2019.
  78. Y. Villuendas-Rey, J. A. Hernández-Castaño, O. Camacho-Nieto, C. Yáñez-Márquez, and I. López-Yáñez. **NACOD: A Naïve Associative**

- Classifier for Online Data**, *IEEE Access*, vol. 7, pp. 117761-117767, 2019.
79. Y. Villuendas-Rey, C. Yáñez-Márquez, J. A. Antón-Vargas, and I. López-Yáñez. **An Extension of the Gamma Associative Classifier for Dealing With Hybrid Data**, *IEEE Access*, vol. 7, pp. 64198-64205, 2019.
80. P. Moreno-Moreno, C. Yanez-Marquez, and O. A. Moreno-Franco. **The new informatics technologies in education debate**, *International Journal of Technology Enhanced Learning*, vol. 1, pp. 327-341, 2009.
81. C. Yanez-Marquez, I. Lopez-Yanez, O. Camacho-Nieto, and A. J. Arguelles-Cruz. **BDD-based algorithm for the minimum spanning tree in wireless ad-hoc network routing**, *IEEE Latin America Transactions*, vol. 11, pp. 600-601, 2013.
82. J. A. Hernández-Castaño, Y. Villuendas-Rey, O. Camacho-Nieto, and C. Yáñez-Márquez. **Experimental platform for intelligent computing (EPIC)**, *Computación y Sistemas*, vol. 22, pp. 245-253, 2018.
83. I. Salgado, C. Yáñez, O. Camacho, and I. Chairez. **Adaptive control of discrete-time nonlinear systems by recurrent neural networks in quasi-sliding mode like regime**. *International Journal of Adaptive Control and Signal Processing*, Vol. 31, no. 1, pp. 83-96, 2017.
84. C. Lopez-Martin, C. Yanez-Marquez, and A. Gutierrez-Tornes. **A fuzzy logic model based upon reused and new & changed code for software development effort estimation at personal level**. In *Proc. IEEE 2006 15th International Conference on Computing*, 2006, pp. 298-303.
85. H. Calvo and A. Gelbukh. **Unsupervised learning of ontology-linked selectional preferences**, In *Iberoamerican Congress on Pattern Recognition*. Springer, Berlin, Heidelberg pp. 418-424, 2015.
86. A. Gelbukh and H. Calvo. **Applications**, *Studies in Computational Intelligence*, Springer, 765, pp. 69-83, 2018.
87. Khalid Jasim M., and al- kaabneh K. **Caries Detection in Panoramic Dental X-ray Images Using Support Vector Machine**. *International Journal of Science and Applied Information Technology*, Vol. 8 no. 6, pp. 40-44. Nov. 2019.
88. Abualigah L., Naser A., Shehab M., and Alshinwan M. **Review of Optimization Methods in Medical Imaging**. *International Journal of Science and Applied Information Technology*, Vol. 8 no. 6, pp. 36-39. Nov. 2019.