Application of Artificial Intelligence in the prediction of breast cancer survival in Mexican women

Aplicación de la Inteligencia artificial en la predicción de la supervivencia de cáncer de mama en mujeres mexicanas

MELO-MORIN, Julia Patricia†*, AHUMADA-CERVANTES, María de los Ángeles and SANTANA-ESPARZA, Gil

Instituto Tecnológico Superior de Pánuco, División de Ingeniería en Sistemas Computacionales

ID 1st Author: *Julia Patricia, Melo-Morin* / **ORC ID**: 0000-0001-7145-2344, **Researcher ID Thomson**: I-3412-2018, **CVU CONACYT ID**: 248185T

ID 1st Coauthor: *María de los Ángeles, Ahumada-Cervantes* / **ORC ID**: 0000-0001-8164-2889, **Researcher ID Thomson**: ABE-2558-2020, **CVU CONACYT ID**: 825136

ID 2nd Coauthor: *Gil, Santana-Esparza* / **ORC ID**: 0000-0002-1124-4006, **CVU CONACYT ID**: 399382

DOI: 10.35429/EJDRC.2020.10.6.6.13 Received March 28, 2020; Accepted June 25, 2020

Abstract

In Mexico, the leading cause of death caused by malignant tumors in women is breast cancer and the general survival of five years treated in facilities of the Public Health System is between 75 and 80%. There are applications that determine the survival of patients with breast cancer, based on the use of drugs that are not prescribed in Mexico, so cancer specialists cannot consider the information offered by these programs for decision-making with patients Mexican. This article describes the development of an expert system that, by applying artificial intelligence techniques, allows the evaluation and prediction of patient survival, based on a model generated with data mining techniques. Rules were obtained from the patterns obtained with data collected from patients with breast cancer since 2006. The development of the system is governed by the Knowledge Discovery from Databases (KDD) methodology, supported by the WEKA tool for modeling data mining techniques. There is a data warehouse of 4,773 women with breast cancer provided by two tertiary hospitals in Mexico City: an INCan cohort of 4,300 patients diagnosed from 2006 to 2013 with a median follow-up of 40.5 months of survival and an INCMSZ-provided cohort of 473 patients from 2011 to 2018 with a median of 39 months.

Survival, Breast cancer, Data mining

Resumen

En México la primera causa de muerte causada por tumores malignos en las mujeres, es el cáncer de mama y la supervivencia general de cinco años tratadas en instalaciones del Sistema de Salud Pública es entre el 75 y 80%. Hay aplicaciones que determinan la supervivencia a las pacientes con cáncer de mama, basándose en el uso de fármacos que en México no son recetados, por lo que los especialistas en cancerología no pueden considerar la información ofrecida por dichos programas para la toma de decisiones con pacientes mexicanas. Este artículo describe el desarrollo de un sistema experto que aplicando técnicas de inteligencia artificial permita la evaluación y realice la predicción de la supervivencia de las pacientes, basadas en un modelo generado con técnicas de minería de datos. Se obtuvieron reglas de los patrones obtenidos con los datos recopilados de pacientes con cáncer de mama desde el año 2006. El desarrollo del sistema se rige por la metodología de Descubrimiento de Conocimiento a partir de Bases de Datos (KDD), apoyado en la herramienta WEKA para el modelado de las técnicas de minería de datos. Se cuenta con un almacén de datos de 4773 mujeres con cáncer de mama proporcionada por dos hospitales terciarios de la Ciudad de México: un cohorte del INCan de 4300 pacientes diagnosticadas desde el 2006 a 2013 con una mediana de seguimiento de 40.5 meses de supervivencia y un cohorte proporcionado por INCMSZ de 473 pacientes de 2011 a 2018 con una mediana de 39 meses.

Supervivencia, Cáncer de mama, Minería de datos

Citation: MELO-MORIN, Julia Patricia, AHUMADA-CERVANTES, María de los Ángeles and SANTANA-ESPARZA, Gil. Application of Artificial Intelligence in the prediction of breast cancer survival in Mexican women. ECORFAN Journal-Democratic Republic of Congo. 2020, 6-10: 6-13

* Correspondence to Author (email: patricia.melo@itspanuco.edu.mx)
† Researcher contributing first author.

Introduction

The World Health Organization indicates that 16% of cancers in women in the world correspond to breast cancer, with 1,671,149 new cases diagnosed annually, with 521,907 deaths annually (World Health Organization (WHO), 2017). In Mexico, 22.56 inhabitants per 100,000 have breast cancer, being the leading cause of death in women over 25 years of age, and the 2020 prognosis indicates that 16,500 new cases were presented. Lozano R, et al (2008), indicates that almost all cases of breast cancer are identified, but only 10% in stage 1 of the same, so, if the cancer is identified in a timely manner and adequate treatment is offered , it is curable.

In order for a doctor to formulate an adequate treatment scheme, it is necessary to know the stage of the cancer. For cancer staging, the AJCC TNM system is used in which 3 criteria are used to define the stage of the cancer: tumor size (T), affected lymph nodes (N) and the presence or absence of metastases (M), together with the stage of the cancer there are other factors that affect the prognosis such as the type of cancer, presence of estrogen and progesterone receptors, biological subtype, presence of the HER2 receptor, how fast the tumor grows, general conditions of the patient, among others. These clinical and pathological variables help the physician to define the patient's prognosis more precisely and to make more prudent decisions about treatment.

The application of data mining in the medical area has supported decision-making; transforming volumes of data, experience, knowledge and wisdom improving the health service offered. It is important for a medical specialist to have tools that help make decisions about the disease, to strengthen effective communication between the doctor and the patient, since it is essential in any treatment (OMS), the medical personnel being key in the treatment and communication in the patient so that they commit to treatment and thus control the disease (Montero JE, 2020).

This article describes some research that applies data mining techniques to cancer data warehouses. Some applications developed in other countries that predict cancer survival are also mentioned.

The methodology used in the research carried out is explained, as well as the results obtained.

State of the art

There are many studies related to the identification of patterns that affect breast cancer by applying different artificial intelligence techniques.

Piñeros et al (2008), characterized the sociodemographic factors of Colombian women with breast cancer, using a database of the District Health Department. A significant association was found with educational level, with a higher proportion of cancers in early stages in women with higher educational levels. A descriptive statistical analysis was applied with ANOVA tests; Associations between variables of a categorical type by means of the chi-square test or Fisher's exact test.

Timaran R. & Yépez M. (2016), used the database of the population cancer registry of the municipality of Pasto in Colombia and extracted survival patterns applying decision tree techniques in women diagnosed with cervical cancer since 1998 to 2007. The result obtained is that the patient is a survivor if the life span is greater than 37 months from the date of diagnosis and using association techniques, they determined the factors that affect survival.

An article by Reparaz et al (2008) used the data mining cluster task to 206 records treated with prostate branchytherapy, to characterize and classify the population of patients with prostate cancer and obtained as a result that 83% of the patients, had a successful treatment.

Hernández & Lorente (2009), applied Weka's software to the Wisconsin Breast Cancer Database data warehouse, with 699 instances and 11 attributes, to classify a tumor as benign or malignant, and determined that all attributes affect to some extent the classification, that is, the lower the attribute value, the greater the probability that the tumor is benign.

Martínez et al (2008), applied Bayesian Networks for the classification of medical data in the processing of databases related to medical conditions such as bed cancer, cancer tumors, diabetes and hepatitis, in order to determine if Bayesian Networks are a effective and reliable tool in decision-making in the medical field and become an assistant in medical diagnosis and treatment. The databases that were used were taken from the repository of data of the University of California of 286 patients for breast cancer. The classification algorithms used were Naive Bayes, Tan, Hill-Climber and K2, resulting in that for all algorithms the classification percentages of the variables are above 70% accuracy.

Camacho C. (2014), used digital image processing and their analysis; using a Heuristic method based on Data Mining to extract essential information from mammographic images and transform them into patterns, to later classify them into subgroups of patterns for the formation of families through homogeneity and maximization of coincidence indexes, in order to facilitate the timely diagnosis of breast cancer.

Molero-Castillo et al (2012) with the database of the Surveillance, Epidemiology and Final Results Program of the National Cancer Institute (NCI) in the United States, characterized patients of Hispanic origin with breast cancer, applying data series.

Existing applications

Adjuvant Online (AO) is a tool that helps physicians in making adjuvant treatment decisions for patients with stage 1 and 2 breast cancer. The program assesses the risk of relapse and death within a 10-year period to individual patients, as well as the benefit that a specific adjuvant treatment provides.

The risk estimates calculated by AO are based on observations of the 10-year overall survival of women between the ages of 20 and 79 years diagnosed with breast cancer between the years 1988 and 1992 in the United States and are part of the SEER database (Surveillance, Epidemiology, End-results), which covers 14% of the US population.

The treatment estimate is obtained by calculating the risk of negative outcomes (death or relapse) of a patient and multiplied by the proportion of negative events that an adjuvant treatment is known to prevent.

The url of the official site of the application is www.adjuvantonline.com (Adjuvant, Inc., 2016).

Predict is an online application that uses a mathematical model to support patients and doctors in deciding the ideal chemotherapy or hormone therapy treatment after breast cancer surgery. This site does not store patient information, it only requests characteristics of certain cancer factors.

Version 2.0 of Predict was developed using 5694 breast cancer registries for women in England from 1999-2003 and has been tested with women with breast cancer worldwide.

The project was developed by the Cambridge Cancer Unit, the Cambridge University Department of Oncology and the Eastern Cancer Registry and Information Center.

The url of the official site of the application is www.predictnhs.uk (Public Health Engleand and Cambridge University, sn).

Methodology

Artificial intelligence in the medical area has been very useful in the detection of different types of diseases. The application of data mining is essential in the detection and diagnosis of cancer to obtain patterns in large volumes of data (Timaran R. & Yépez M., 2016).

The methodology used to obtain knowledge in a database is the KDD (Knowledge Discovery in Databases), which transforms information from a lower level to greater knowledge at the higher level Hernández et al, 2005, Mitra, 2003). Figure 1 indicates each of the stages.

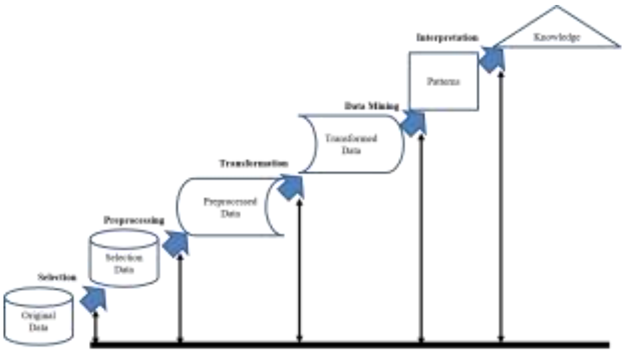


Figure 1 Stages for the knowledge extraction process. *Source: Based on Fayyad et al, (1996) and Gómez, (1998)*

To obtain the survival analysis of a patient, it is necessary to study the time variable and the relationship that exists with other variables (Abraira, 1996). To calculate the survival time, the start date and the end date of the follow-up are defined, and it is the time elapsed between both. The termination date may be because the patient died or contact with her was lost, which is why it is considered censored (Aguayo C. & Lora M., 2007). To determine survival curves, the Kaplan-Meier method is applied, which is the accumulated probability of survival over time. Formula 1 describes the survival formula that describes whether an individual or event occurs in a time equal to or greater than t.

$$S(t) = \prod_{j|ti\leq ti}(1-\frac{deceasedj}{survivorsj}) \qquad (1)$$

The risk function h (t), also called the instantaneous mortality rate, is the probability density function of the event at t, it indicates the probability that an individual dies in that unit of time. Formula 2 indicates the risk function.

$$h(t) = \frac{\text{number of deaths in t / time unit}}{\text{number of survivors in t}} \qquad (2)$$

To determine the factors that affect breast cancer survival in Mexican women, the KDD methodology was followed for the extraction of knowledge, as shown in Figure 2.

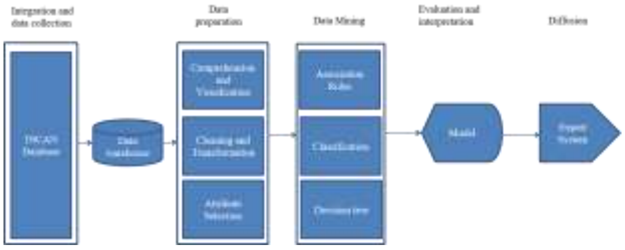


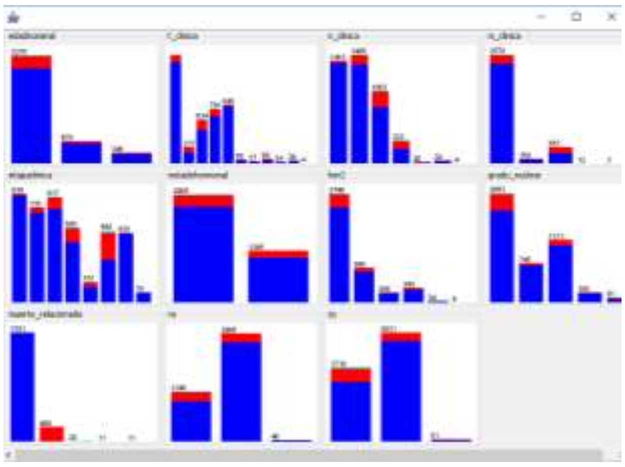
Figure 2 Stages for the knowledge extraction process. *Source: Self Made*

The data warehouse is composed of 18 different attributes, which are described in Table 1.

#	Attribute name	Description	Values	Meaning
1	Nominal_Age	Patient age	1	Age <40
			2	Age> = 40 and Age <= 69
			3	Age> = 70
2	t_clinica	Primary tumor	-1	"Null or empty"
			0	"T0"
			1	"T1"
			2	"T2"
			3	"T3"
			4	"T4a"
			5	"T4b"
			6	"T4c"
			7	"T4d"
			8	"Tx"
			99	"Unknown"
3	n_clinica	Regional lymph nodes	-1	"Null or empty"
			0	"N0"
			1	"N1"
			2	"N2"
			3	"N3"
			4	"Nx"
			99	"Unknown"
4	Hormonal_state	Hormonal Status	1	Pre menopause
			2	(Age <50)
5	M_clinica	Distant metastasis	-1	Postmenopause
			0	(Age> = 50)
			1	"Null or empty"
			2	“Mx”
			99	"Unknown"
6	clinical_stage	Clinical stage	0	"0"
			1	"I"
			2	"IIa"
			3	"IIb"
			4	"IIIa"
			5	"IIIb"
			6	"IIIc"
			7	"IV"
7	estrogen_receptors	estrogen_receptors	0	"Null or empty"
			1	Negative
			2	Weakly positive
			3	Moderately positive
			4	Intensely positive
8	progesterone_receptors	H_score progesterone receptors	0	"Null or empty"
			1	Negative
			2	Weakly positive
			3	Moderately positive
			4	Intensely positive
9	her2	HER2	-1	"Null or empty"
			0	"0"
			1	"1+"
			2	"2+"
			3	"3+"
			99	"Unknown"
10	Years_Viv	Years lived after diagnosis	(Death_date-diagnostic_date) / 365	
11	nuclear_degree	Nuclear grade	0	"Low"
			1	"Intermediate"
			2	"High"
			99	"Unknown"
12	death_related	Death / alive	0	Another cause
			1	For cancer
			99	Unknown.
			2	Death not established
			3	Still alive
13	Vital_state	Actual state	3	"Live"
			1	"Dead"
			2	"Losses"

Table 1 Attributes of the mineable view
Source: Self Made

Different data mining techniques were applied, using Weka software, considering the data warehouse. Graph 1 shows the relationship of each of the attributes with the class to be analyzed (alive / dead). The blue color represents living patients and the red color represents dead patients.



Graphic 1 Attribute relationship with the dead / alive main class
Source: Weka Software, Own Execution

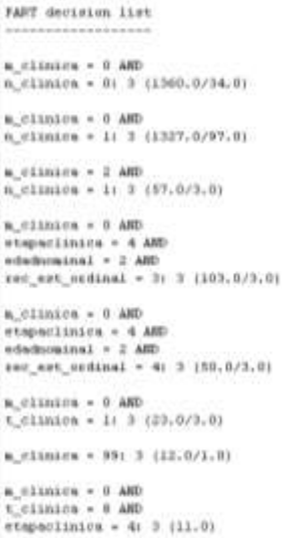
Table 2 shows a comparison between the different classification algorithms used, indicating the instances correctly and incorrectly classified, as well as the Kappa statistic that measures the coincidence of the prediction with the real class. The last column of the Table indicates the result of the level of error generated by the model after applying the algorithm.

Algorithm	Well-classified instances (%)	Misclassified Instances (%)	Kappa statistic	Absolut mistake
One R	87%	13%	0.0069	0.0516
Decision Table	87%	13%	0.0806	0.0873
Part	86%	14%	0.1356	0.0779
Prism	97%	2%	0.9208	0.0067
Ridor	87%	12%	0	0.0905
J48	87%	13%	0.1087	0.0803
BFTree	87%	13%	0.0113	0.0863
Naive Bayes	83%	17%	0.3336	0.0829

Table 2 Results of the classification algorithms
Source: Self Made

When applying the classification algorithms, a model was generated with different rules that determine the survival of Mexican women, describing characteristics in each of the attributes, as well as the percentage of certainty of each one of them, determining the average life span of the group of patients that was classified in each of the generated rules.

Figures 3 and 4 show the generation of the rules by some data mining algorithms, indicating the values in the different attributes and drawing conclusions from them.

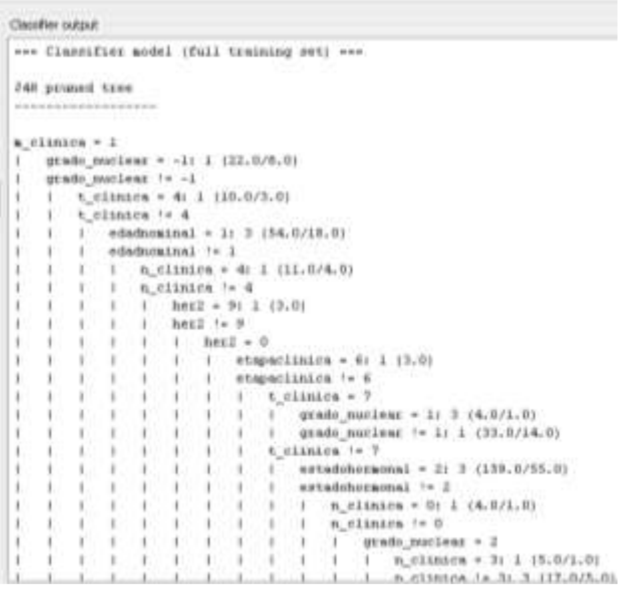


```
PART decision list
------------------

m_clinica = 0 AND
n_clinica = 0: 3 (1360.0/34.0)

m_clinica = 0 AND
n_clinica = 1: 3 (1327.0/97.0)

m_clinica = 2 AND
n_clinica = 1: 3 (57.0/3.0)

m_clinica = 0 AND
etapaclinica = 4 AND
edadnominal = 2 AND
rec_est_ordinal = 3: 3 (103.0/3.0)

m_clinica = 0 AND
etapaclinica = 4 AND
edadnominal = 2 AND
rec_est_ordinal = 4: 3 (50.0/3.0)

m_clinica = 0 AND
t_clinica = 1: 3 (23.0/3.0)

m_clinica = 99: 3 (12.0/1.0)

m_clinica = 0 AND
t_clinica = 8 AND
etapaclinica = 4: 3 (11.0)
```

Figure 3 Rules generated by the Part algorithm
Source: Weka Software, Own Execution

The rules have the form "If ... (conditions to be met) then ... (class to which it belongs), for example:

- Rule 1. If the patient has distant M0 metastases and N1 regional lymph nodes, then she is classified as class 3 (alive). This rule is verified by 1,327 patients, with 97 errors.
- Rule 2. If the patient has distant M0 metastases and N0 regional lymph nodes, then she is classified as class 3 (alive). Verifying the rule 1360 patients, with 34 errors.



```
=== Classifier model (full training set) ===

J48 pruned tree
------------------

m_clinica = 1
|   grado_nuclear = -1: 1 (22.0/6.0)
|   grado_nuclear != -1
|   |   t_clinica = 4: 1 (10.0/3.0)
|   |   t_clinica != 4
|   |   |   edadnominal = 1: 3 (54.0/18.0)
|   |   |   edadnominal != 1
|   |   |   |   n_clinica = 4: 1 (11.0/4.0)
|   |   |   |   n_clinica != 4
|   |   |   |   |   her2 = 9: 1 (3.0)
|   |   |   |   |   her2 != 9
|   |   |   |   |   |   her2 = 0
|   |   |   |   |   |   |   etapaclinica = 6: 1 (3.0)
|   |   |   |   |   |   |   etapaclinica != 6
|   |   |   |   |   |   |   |   t_clinica = 7
|   |   |   |   |   |   |   |   |   grado_nuclear = 1: 3 (4.0/1.0)
|   |   |   |   |   |   |   |   |   grado_nuclear != 1: 1 (33.0/14.0)
|   |   |   |   |   |   |   |   t_clinica != 7
|   |   |   |   |   |   |   |   |   estadohormonal = 2: 3 (139.0/55.0)
|   |   |   |   |   |   |   |   |   estadohormonal != 2
|   |   |   |   |   |   |   |   |   |   n_clinica = 0: 1 (4.0/1.0)
|   |   |   |   |   |   |   |   |   |   n_clinica != 0
|   |   |   |   |   |   |   |   |   |   |   grado_nuclear = 2
|   |   |   |   |   |   |   |   |   |   |   |   n_clinica = 3: 1 (5.0/1.0)
|   |   |   |   |   |   |   |   |   |   |   |   n_clinica != 3: 3 (17.0/5.0)
```

Figure 4 Rules generated by the J48 algorithm
Source: Weka Software, own execution

Of the rules with the greatest force in the live / dead class, the main attributes for the patient to remain alive according to the model are considered to be: distant metastases N0 and regional lymph nodes N0 or N1. Similarly for the dead patient, the attributes are primary tumor T4b or T4d, and high nuclear grade.

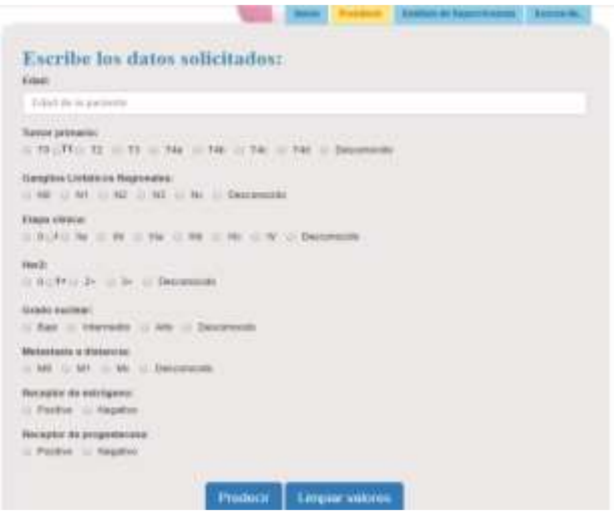
Results

A computer system was implemented that applies the rules obtained in the data mining model, with the attributes identified as significant in Mexican patients, and the survival time. By providing new data, the computer system determines the survival time so that medical specialists can determine the appropriate treatment for the patient. The results of applying the algorithms to identify significant attributes are shown in Table 3.

Attribute priority (highest to lowest)	Ranking-ReliefAttributeEval	GreedySteepwise - Distribution in folds (Indicates the percentage of impact).
1	Years_lived	Nominal_Age -100%
2	Vital_state	Clinical stage - 80%
3	Nuclear_grade	Hormonal status - 80%
4	T_clinica	t_clinica - 70%
5	Progesterone_receptors	her2 - 70%
6	N_clinica	years_lived - 70%
7	M_clinica	
8	Estrogen_receptors	nuclear_degree - 50%
9	Hormonal state	n_clinica - 40%
10	Nominal age	estrogen_receptors - 10%
11	Her2	m_clinica - 0%
12	Clinical stage	progestereone_receptors - 0%

Table 3 Attributes selected by the different algorithms. *Source: Self Made*

The most significant algorithms were considered in the system carried out, as shown in Figure 5.



Escribe los datos solicitados:

Edad:

Edad de la paciente

Tumor primario:
○ T0 ○ T1 ○ T2 ○ T3 ○ T4a ○ T4b ○ T4c ○ T4d ○ Desconocido

Ganglios Linfáticos Regionales:
○ N0 ○ N1 ○ N2 ○ N3 ○ Nx ○ Desconocido

Etapa clínica:
○ 0 ○ I ○ IIa ○ IIb ○ IIIa ○ IIIb ○ IIIc ○ IV ○ Desconocido

Her2:
○ 0 ○ 1+ ○ 2+ ○ 3+ ○ Desconocido

Grado nuclear:
○ Bajo ○ Intermedio ○ Alto ○ Desconocido

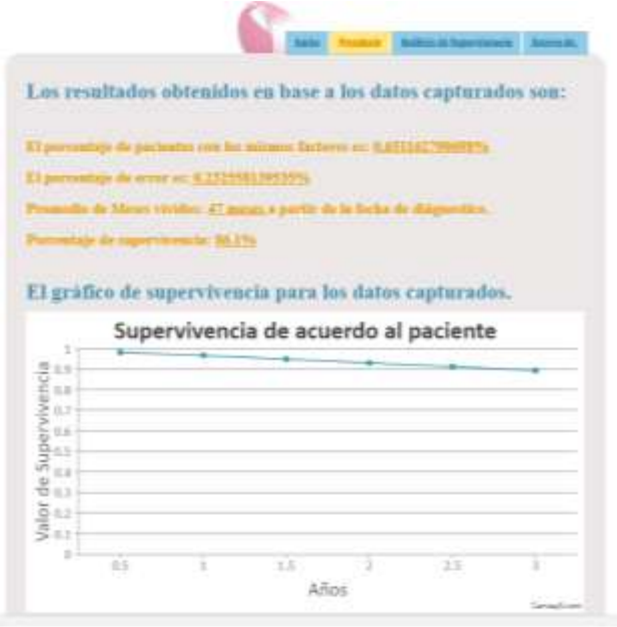
Metástasis a distancia:
○ M0 ○ M1 ○ Mx ○ Desconocido

Receptor de estrógeno:
○ Positivo ○ Negativo

Receptor de progesterona:
○ Positivo ○ Negativo

Figure 5 Capture form
Source: Application execution.

After accepting the data in the system, pressing the Predict button and taking the knowledge base, conclusions are obtained and the survival time of the patient is determined, supported by the established rules, as indicated in Figure 6.



Los resultados obtenidos en base a los datos capturados son:

El porcentaje de pacientes con los mismos factores es: 0.651162790698%

El porcentaje de error es: 0.232558139535%

Promedio de Meses vividos: 47 meses a partir de la fecha de diagnóstico.

Porcentaje de supervivencia: 86.1%

El gráfico de supervivencia para los datos capturados.

Figure 6 Survival results of the patient. Source: Application execution.

Acknowledgments

I am grateful to Dra. Alejandra Armengol Alonso, from the Department of Hematology and Oncology at the National Institute of Medical Sciences and Nutrition (INCAN), in Mexico City, who provided the data for the research, as well as validated the results obtained.

Conclusions

According to medical experts, this is the first expert system that allows calculating the prognosis according to the clinical-pathological variables of Mexican patients and it is of great relevance to know the survival of a Mexican patient with breast cancer in the public health system with access to essential treatments. The applications of the system can be multiple in routine clinical practice, education, and in the adoption of public policies for breast cancer in Mexico. Currently, work is being done on a predictive model for the benefit of cancer treatment, also based on a system that applies artificial intelligence techniques.

References

Aguayo Canela M, Lora Monge E (2007). Cómo hacer "paso a paso" un Análisis de Supervivencia con SPSS para Windows. Servicio de Medicina Interna. Consultado on-line en septiembre 17, 2017 en DocuWeb fabis.org

Adjuvant Inv. (2016). Adjuvant. Sitio on-line www.adjuvantonline.com

Abraira, V. & Pérez de Vargas A. (1996). Métodos Mutivariantes en Bioestadística. Ed. Centro de Estudios Ramos Areces. Consultado on line en septiembre 13, 2017 en http://www.hrc.es/bioest/

Camacho C., S. S. (2014). Método Heurístico para el Diagnóstico de Cáncer de Mama basado en Minería de Datos. Revista del Postgrado en Informática, 97.

Fayyad, U.; Piatetsky-Shapiro G. et al (1996). From Data Mining to Knowledge Discovery in Databases. Artificial Intelligence Magazine.

Hernández J., Ramírez M.& Ferri C. (2005). Introducción a la Minería de Datos. Ed. Pearson.

Hernández M., & Lorente R. (2009). Minera de datos aplicada a la detección de Cáncer de Mama. Universidad Carlos III, Madrid España.

Martínez, R. E. B., Ramírez, N. C., Mesa, H. G. A., Rabatte, I., Suárez, P. P. L., Trejo, M. D. C. G., & Morales, M. S. L. B. (2008). Evaluación del Potencial de Redes Bayesianas en la Clasificación en Datos Médicos. Revista Médica de la Universidad Veracruzana, 8(1), 33-37.

Mitra S. & Acharya T. (2003). Data mining: multimedia, soft computing and bioinformatics. John Wiley & Sons.

Molero-Castillo, G. G., González, Y. C., & Campaña, M. E. M. (2012). Sld 161 Caracterización y análisis de la base de datos de cáncer de mama SEER-DB. Informática y salud 2013.

Montero, J. E. (2020). La nota hipermedia como estrategia didáctica y de comunicación entre médico y paciente.

Organización Mundial de la Salud. (2004). Carga Mundial de Morbilidad. Consultado en abril 19 en http://apps.who.int/gho/data/

Organización Mundial de la Salud. (2017). Cáncer de mama: prevención y control. Consultado en abril 19 en http://www.who.int/topics/cancer/breastcancer/es/index1.html

Piñeros, M., Sánchez, R., Cendales, R., Perry, F., Ocampo, R., & García, Ó. A. (2008). Características sociodemográficas, clínicas y de la atención de mujeres con cáncer de mama en Bogotá. Rev colomb cancerol, 12(4), 181-190.

Public Health Engleand and Cambridge University (sn). Predict. Sitio on-line www.predictnhs.uk

Reparaz, D., Merlino, H., Rancan, C., Rodríguez, D., Britos, P. V., & García Martínez, R. (2008). Determinación de la Eficacia de la Braquiterapia en Tratamiento de Cáncer Basada en Minería de Datos. In X Workshop de Investigadores en Ciencias de la Computación.

Secretaria de Salud (2017). Cáncer de mama. Consultado en abril 20 en http://cnegsr.salud.gob.mx/contenidos/Programas_de_Accion/CancerdelaMujer/InfEstad.html

Scheffer,T. (2001). Finding association rules that trade support optimally against confidence.In L.de Raedt and A.Siebes, editors, Proceedings of the Fifth European Conference on Principles of Data Mining and Knowledge Discovery, Freiburg, Germany. Berlin: Springer-Verlag, pp. 424–435.

Timarán-Pereira, R & Yépez-Chamorro, M.C. (2016). Caracterización de la supervivencia de mujeres con cáncer invasivo de cuello uterino usando minería de datos. Revista de Investigación, Desarrollo e Innovación,7(1), 127-139. doi: 10.19053/20278306.v7.n1.2016.4315

Timarán-Pereira, S. R., Hernández-Arteaga, I., Caicedo-Zambrano, S. J., Hidalgo-Troya, A. y Alvarado Pérez, J. C. (2016). El proceso de descubrimiento de conocimiento en bases de datos. En Descubrimiento de patrones de desempeño académico con árboles de decisión en las competencias genéricas de la formación profesional (pp. 63-86). Bogotá: Ediciones Universidad Cooperativa de Colombia. doi: http://dx.doi.org/10.16925/9789587600490