



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

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

Editorial label ECORFAN: 607-8695

BECORFAN Control Number: 2020-05

BECORFAN Classification (2020): 111220-0005

Pages: 12

RNA: 03-2010-032610115700-14

**ECORFAN-México, S.C.**

143 – 50 Itzopan Street  
La Florida, Ecatepec Municipality  
Mexico State, 55120 Zipcode  
Phone: +52 1 55 6159 2296  
Skype: ecorfan-mexico.s.c.  
E-mail: contacto@ecorfan.org  
Facebook: ECORFAN-México S. C.

Twitter: @EcorfanC

[www.ecorfan.org](http://www.ecorfan.org)

**Holdings**

Mexico	Colombia	Guatemala
Bolivia	Cameroon	Democratic
Spain	El Salvador	Republic
Ecuador	Taiwan	of Congo
Peru	Paraguay	Nicaragua

# Resumen

El presente 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 con cáncer de mama, basadas en un modelo generado con técnicas de minería de datos.

Introducción

Aplicaciones existentes

Metodología

Resultados

Conclusiones

Referencias

# Introducción

- El 16% de los cánceres en mujeres en el mundo corresponde al cáncer de mama (OMS). En México 22.56 habitantes por cada 100,000 tienen cáncer de mama, siendo la primera causa de muerte en mujeres mayores de 25 años (SS).
- Para la estadificación del cáncer se utiliza el sistema TNM de la AJCC en el cual se utilizan 3 criterios para definir la etapa del cáncer: el tamaño del tumor (T), los ganglios linfáticos afectados (N) y la presencia o no de metástasis (M).
- La minería de datos en el área médica ha apoyado en la toma de decisiones, desde la virtualización inteligente hasta el control de calidad.

# Aplicaciones existentes



Home About Predict Predict Tool Contact Legal Change Language



We recommend that patients use this tool in consultation with their doctor.

<https://breast.predict.nhs.uk/tool>

Reset

Predict is not designed to be used in all cases. [Click here for more details.](#)  
If you are unsure of any inputs or outputs, click on the **i** buttons for more information.

DCIS or LCIS only? **i** Yes No

Age at diagnosis **i** - [ ] +  
Age must be between 25 and 85

Post Menopausal? **i** Yes No Unknown

ER status **i** Positive Negative

HER2 status **i** Positive Negative Unknown

Ki-67 status **i** Positive Negative Unknown  
Positive means more than 10%

Invasive tumour size (mm) **i** - [ ] +  
If there was more than one tumour, enter the size of the largest tumour. If neo-adjuvant th the size before neo-adjuvant therapy.

Tumour grade **i** 1 2 3

Detected by **i** Screening Symptoms Unknown

Positive nodes **i** - [ ] +

Micrometastases only **i** Yes No Unknown  
Enabled when positive nodes is 1.

## Adjuvant! Online

### Adjuvant! Online

Decision making tools for health care professionals

El propósito de **Adjuvant! Online** es asistir a los profesionales de la salud y pacientes con cáncer en la evaluación de los riesgos y beneficios de seguir un tratamiento adyuvante adicional (adyuvante: por lo general la quimioterapia, terapia hormonal, o ambas cosas) después de la cirugía.

El objetivo es ayudar a los profesionales de la salud hacer estimaciones del riesgo de resultados negativos (el cáncer o la mortalidad relacionada con la recaída) sin terapia adyuvante sistémica, las estimaciones de la reducción de estos riesgos asociados a la terapia, y los riesgos de efectos secundarios de la terapia. Estas estimaciones se basan en la información introducida sobre los pacientes y sus tumores (por ejemplo, edad del paciente, tamaño tumoral, afectación ganglionar, grado histológico, etc) Estas estimaciones se pueden exportar a hojas impresas en formatos de texto y gráfica útiles para ser incluidas en la historia clínica del paciente.

Debido a la complejidad de la interpretación de algunos de las entradas de información (las ambigüedades sobre el tamaño del tumor, márgenes, etc), la información debe ser registrada por un profesional de la salud con experiencia en oncología (cáncer de la medicina).

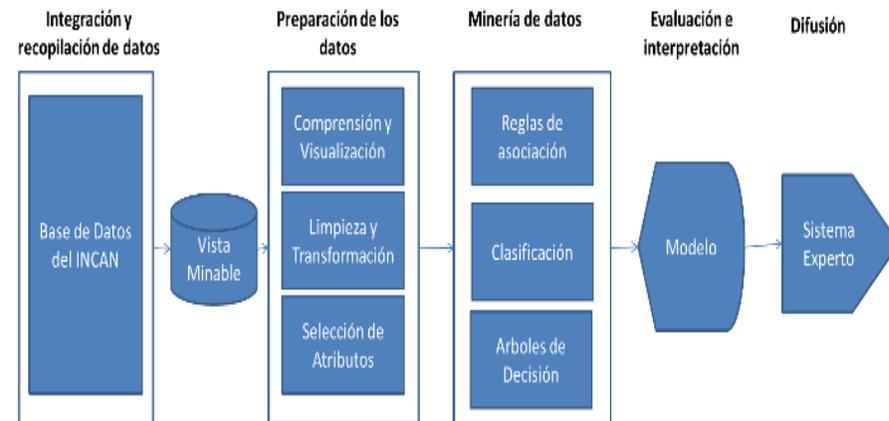
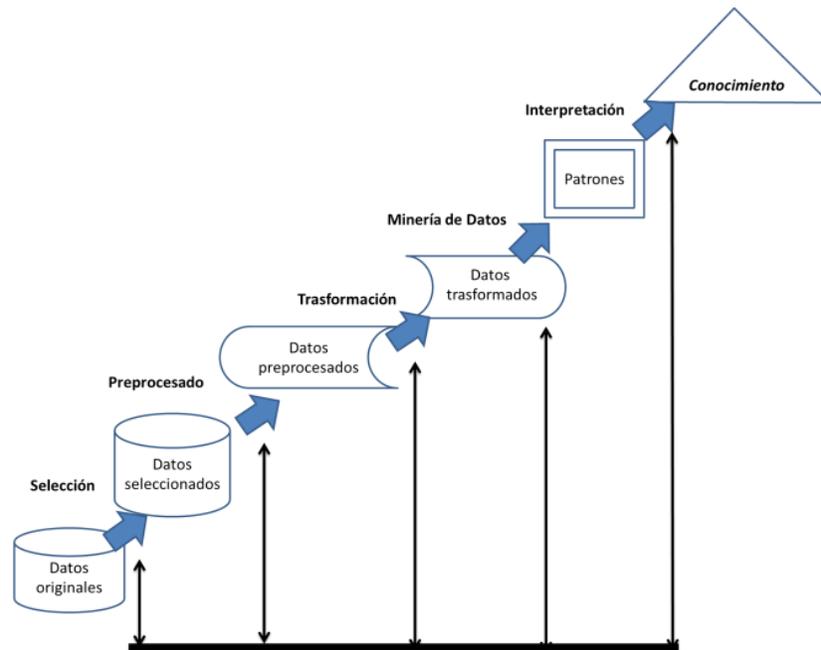
El servidor proporciona tanto la posibilidad de hacer consultas online a través de la red después de haber completado un breve registro o la posibilidad de descargar un programa informático que puede ser instalado también en Palm y Pocket Pcs. El programa hace estimaciones de riesgo para neoplasias de mama, colon, y pulmón.

URL: <https://www.adjuvantonline.com/>

Referencias: Ravdin PM, Siminoff LA, Davis GJ, Mercer MB, Hewlett J, Gerson N, Parker HL. Computer program to assist in making decisions about adjuvant therapy for women with early breast cancer. [J Clin Oncol. 2001 Feb 15;19\(4\):980-91.](#)

# Metodología

La metodología empleada para la obtención de conocimiento en base de datos es la KDD (Knowledge Discovery in Databases), que transforma información de nivel inferior a mayor conocimiento en el nivel superior.

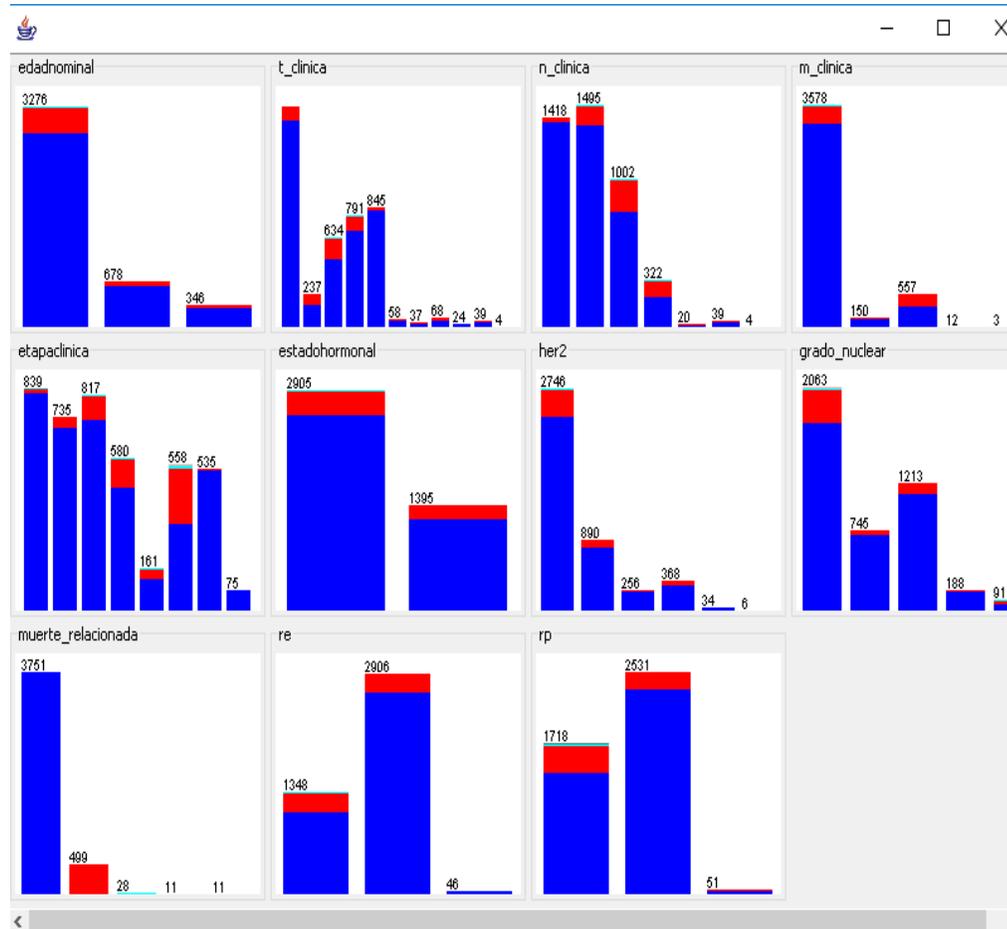


# Fase 1. Integración de los datos

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 y un cohorte proporcionado por INCMSZ de 473 pacientes de 2011 a 2018.

Atributo	Descripción del atributo
Registro	Número de registro del paciente
Edad	Edad del paciente en años
clínica	Tamaño tumoral
n_clinica	Numero de ganglios
m_clinica	Metástasis a distancia
etapa_clinica	Etapa clínica
estado_hormonal	Estado hormonal
receptores_estrogenos	Receptores de estrógenos H_score
receptores_progesterona	Receptores de progesterona H_score
her2	Cantidad de proteína HER2
ki67	Células positivas encontradas.
grado_nuclear	Grado del tumor
fecha_diagnostico	Fecha que se diagnosticó cáncer
fecha_reurrencia	Fecha de recurrencia
fecha_ultima_consulta	Fecha de la última consulta
fecha_recaida	Fecha de recaída de la paciente por cáncer
muerte_relacionada	Muerte/viva
fecha_muerte	Fecha de muerte

# Fase 2. Preparación de los datos



Prioridad de atributos (mayor a menor)	Ranking- ReliefAttributeEval	GreedySteeptwise - Distribución en folds (Indica el porcentaje de impacto).
1	Anios_vividos	Edad_nominal -100%
2	Estado_vital	Etaaclinica – 80%
3	Grado_nuclear	Estadohormonal – 80%
4	T_clinica	t_clinica - 70%
5	Receptores_progesterona	her2 – 70%
6	N_clinica	anios_vividos – 70%
7	M_clinica	grado_nuclear – 50%
8	Receptores_estrogeno	n_clinica – 40%
9	Estadohormonal	receptores_estrogenos – 10%
10	Edadnominal	m_clinica – 0%
11	Her2	receptores_progesterona – 0%
12	Etaaclinica	estado_vital - 0%

# Fase 3: Técnicas de minería de datos

Técnicas de Clasificación. Permiten construir modelos que permitan predecir la categoría de cada una de las instancias, es decir permiten determinar las características de las pacientes que sobrevivieron al cáncer y las que no.

```
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)
```

```
Classifier output

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

J48 pruned tree
-----

m_clinica = 1
| grado_nuclear = -1: 1 (22.0/8.0)
| 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 (159.0/55.0)
| | | | | | | | | | 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)
```

# Método Kaplan-Meier

***Función de Supervivencia:***

$$S(t) = \prod_{j|t_i \leq t} \left(1 - \frac{\text{fallecidos}_j}{\text{sobrevivientes}_j}\right)$$

meses_vividos	Mueren	Viven	expuestos	fallecidos	ptramo	p
0	6		4,300	6	0.998605	0.998605
1	17		4,294	17	0.996041	0.994651
2	13		4,277	13	0.99696	0.991628
3	11		4,264	11	0.99742	0.98907
4	11	6	4,253	11	0.997414	0.986512
5	11	55	4,236	11	0.997403	0.98395
6	9	51	4,170	9	0.997842	0.981826
7	8	34	4,110	8	0.998054	0.979915
8	14	62	4,068	14	0.996559	0.976543
9	10	54	3,992	10	0.997495	0.974097
10	6	48	3,928	6	0.998473	0.972609
....	....	...	....	...	...	...
89		22	33	0	1	0.793225
90		9	11	0	1	0.793225
94		1	2	0	1	0.793225
98		1	1	0	1	0.793225

# Resultados

**Inicio** **Predecir** **Análisis de Supervivencia** **Acerca de.**

### Escribe los datos solicitados:

Edad:

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

Ganglios Linfáticos Regionales:  
 N0  N1  N2  N3  Nx  Desconocido

Etapas clínicas:  
 0  I  IIa  IIb  IIIa  IIIb  IIIc  IV  Desconocido

Her2:  
 0  1+  2+  3+  Desconocido

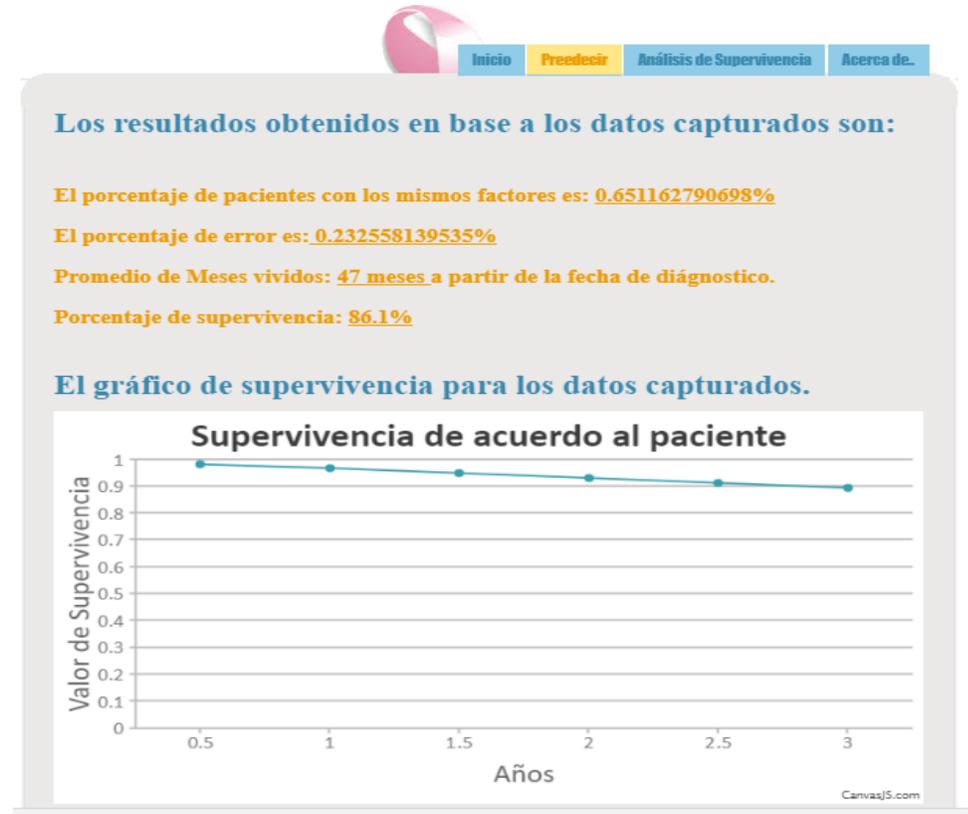
Grado nuclear:  
 Bajo  Intermedio  Alto  Desconocido

Metastasis a distancia:  
 M0  M1  Mx  Desconocido

Receptor de estrógeno:  
 Positivo  Negativo

Receptor de progesterona:  
 Positivo  Negativo

**Predecir** **Limpiar valores**



# Conclusiones

- De acuerdo a los expertos médicos, este es el primer sistema experto que permite calcular el pronóstico según las variables clínico-patológicas de pacientes mexicanas.
- Las aplicaciones del sistema pueden ser múltiples en la práctica clínica habitual, la educación y en la adopción de políticas públicas para el cáncer de mama en México.
- Actualmente se está trabajando en un modelo predictivo de beneficio de tratamiento oncológico basado también en un sistema que aplique técnicas de inteligencia artificial.

# Referencias

- 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
- 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.
- 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.
- 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.
- 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



**ECORFAN®**

© ECORFAN-Mexico, S.C.

No part of this document covered by the Federal Copyright Law may be reproduced, transmitted or used in any form or medium, whether graphic, electronic or mechanical, including but not limited to the following: Citations in articles and comments Bibliographical, compilation of radio or electronic journalistic data. For the effects of articles 13, 162,163 fraction I, 164 fraction I, 168, 169,209 fraction III and other relative of the Federal Law of Copyright. Violations: Be forced to prosecute under Mexican copyright law. The use of general descriptive names, registered names, trademarks, in this publication do not imply, uniformly in the absence of a specific statement, that such names are exempt from the relevant protector in laws and regulations of Mexico and therefore free for General use of the international scientific community. BECORFAN is part of the media of ECORFAN-Mexico, S.C., E: 94-443.F: 008- ([www.ecorfan.org/](http://www.ecorfan.org/) booklets)