

UNIVERSIDAD CATÓLICA SANTO TORIBIO DE MOGROVEJO
FACULTAD DE INGENIERÍA
ESCUELA DE INGENIERÍA DE SISTEMAS Y COMPUTACIÓN



**ARTÍCULO DE REVISIÓN: SISTEMAS PREDICTIVOS DEL CÁNCER
DE MAMA BASADOS EN REDES NEURONALES ARTIFICIALES**

**TRABAJO DE INVESTIGACIÓN PARA OPTAR EL GRADO ACADÉMICO DE
BACHILLER EN INGENIERÍA DE SISTEMAS Y COMPUTACIÓN**

AUTOR

SERGIO ALEXANDER MONDRAGON SILVA

ASESOR

KARLA CECILIA REYES BURGOS

<https://orcid.org/0000-0003-3520-5076>

Chiclayo, 2022

INFORME DE ORIGINALIDAD

25%

INDICE DE SIMILITUD

23%

FUENTES DE INTERNET

18%

PUBLICACIONES

13%

TRABAJOS DEL
ESTUDIANTE

FUENTES PRIMARIAS

1	www.researchgate.net Fuente de Internet	3%
2	sedici.unlp.edu.ar Fuente de Internet	1%
3	Badar ul Islam, Shams Forruque Ahmed. "Short-Term Electrical Load Demand Forecasting Based on LSTM and RNN Deep Neural Networks", Mathematical Problems in Engineering, 2022 Publicación	1%
4	latamt.ieeeer9.org Fuente de Internet	1%
5	ao.um5s.ac.ma Fuente de Internet	1%
6	downloads.hindawi.com Fuente de Internet	1%
7	Mekhriniso Abdukodirova, Salma Abdullah, Abeer Alsadoon, P.W.C. Prasad. "Deep learning for ovarian follicle (OF) classification	1%

Índice

Resumen	4
Abstract	5
Introducción	6
Metodología	6
Resultados	10
Conclusiones	12
Referencias	13

Resumen

El presente artículo buscará revisar los artículos de investigaciones realizadas sobre de sistemas predictivos del cáncer de mamá basados en redes neuronales con el fin de identificar tendencias tanto en la construcción de estos sistemas como en los países que las desarrollan. Por tal motivo, se realizará una búsqueda en las bases de datos a las que se obtuvo acceso, se procederá a plantear filtros y criterios de exclusión e inclusión para lograr obtener artículos con un estándar definido. Como resultado de lo anterior, se obtuvieron doce trabajos de investigación de los cuales se pudo afirmar que el mayor número de investigaciones provienen de la India e Inglaterra observándose una gran ausencia de los países latinoamericanos. Además, se logró tener un listado de las principales características que son utilizadas en el entrenamiento de las redes neuronales aplicadas en los diferentes sistemas informáticos. Finalmente, se espera que este artículo sirva como referencia e inspiración para investigar e implementar software predictivo en el área de salud.

Palabras clave: sistemas predictivos, redes neuronales artificiales, cáncer de mama.

Abstract

This paper's goal is to carry out a systematic review of breast cancer predictive systems based on neural networks in order to identify trends both in the construction of these systems and in the countries that develop them. For this reason, a search will be carried out in the databases of authoritative academic resources, filters and exclusion and inclusion criteria will be proposed to obtain articles with a defined standard. As a result of the above, twelve research papers were obtained from which it could be affirmed that the greatest number of investigations come from India and England, observing a great absence from Latin American countries. In addition, it was possible to have a list of the main characteristics that are used in the training of the neural networks applied in the different computer systems. Finally, this article is expected to serve as a reference and inspiration to investigate and implement predictive software in the area of health.

Keywords: predictive systems, artificial neural networks, breast cancer.

Introducción

El objetivo del presente trabajo es realizar una revisión sistemática de la literatura sobre estudios relacionados con sistemas inteligentes, basados en redes neuronales artificiales, utilizados en la predicción de cáncer de mama para identificar la tendencia durante los últimos años.

Actualmente es común ver diferentes trabajos relacionados con la predicción de eventos ya sean meteorológicos, económicos o médicos utilizando la inteligencia artificial. Por lo tanto, es necesario entender lo que existe detrás de esta tecnología y por qué muchas de estas aplicaciones están basadas en redes neuronales artificiales. Así pues, el siguiente trabajo es impulsado por el afán de entender esta nueva tecnología para servir como referencia en futuras investigaciones.

Por otro lado, la importancia de este trabajo también está relacionado con el pronto diagnóstico del cáncer de mama. Según la Dirección General de Epidemiología (DGE) [13], durante el año 2017 se registraron 9.6 % de pacientes afectados con cáncer de mama (en relación con el total de casos de cáncer detectados), así el cáncer de mama se convierte en el 4to tipo de cáncer más común entre hombres y mujeres. Por lo tanto, es apropiado una detección temprana y precisa del cáncer para no lamentar sus fatales consecuencias.

Metodología

En esta sección se presentará el procedimiento de búsqueda presentando las bases de datos utilizadas, los términos de búsqueda y los criterios de inclusión y exclusión.

Preguntas de investigación

Con el fin de guiar la investigación se plantearon preguntas respecto a investigaciones de los últimos 3 años que abordan el tema de redes neuronales para la predicción de cáncer de mama:

- o¿De qué países provienen la mayor cantidad de publicaciones?
- o¿Cuáles son las características más utilizadas como datos de entrada al diseñar la red neuronal?

Planteamiento de la búsqueda

Las bases de datos utilizadas para la búsqueda fueron las siguientes:

1. ProQuest
2. ScienceDirect

Se decidió buscar utilizando los siguientes términos de búsquedas con conectores lógicos utilizados en ambas bases de datos:

- i. T1: “neural network” AND “breast cancer prediction system”
- ii. T2: "artificial neural network" AND ti("prediction") AND ti ("breast cancer") NOT diabetes

Además, se utilizaron los siguientes criterios de búsqueda:

TABLA I
CRITERIOS DE INCLUSIÓN Y EXCLUSIÓN

INCLUSIÓN	EXCLUSIÓN
-Se buscará en artículos, revistas, documentos de conferencias existentes en las bases de datos descritas y confiables. -Publicaciones que contengan las principales características de predicción utilizadas.	-Artículos pequeños, es decir, artículos con 4 páginas o menos sin una fuente confiable. -Publicaciones sin fundamentar las características de predicción utilizadas. -Publicaciones únicamente de carácter médico y no informático. -Publicaciones no referentes a cáncer de mama.

Ejecución de la búsqueda

Se comenzó por buscar los términos de búsqueda en las bases de datos propuestas sin utilizar ningún filtro, con el fin de mostrar la cantidad de información existente desde una vista general. En ProQuest, se encontraron 2 artículos con el término de búsqueda T1 y 19 artículos con el término de búsqueda T2. En ScienceDirect, se encontraron 3 artículos con el término de búsqueda T1 y 54 artículos con el término de búsqueda T2. Esta información se encuentra plasmada en la Tabla 2 y 3.

TABLA 2
RESULTADOS DE PRIMERA BÚSQUEDA CON EL TÉRMINO DE BÚSQUEDA T1.

Base de datos	"neural network" AND "breast cancer prediction system"
ProQuest	2
ScienceDirect	3
Total	5

TABLA 3
RESULTADOS DE PRIMERA BÚSQUEDA CON EL TÉRMINO DE BÚSQUEDA T2.

Base de datos	"artificial neural network" AND ti("prediction") AND ti("breast cancer") NOT diabetes
ProQuest	19
ScienceDirect	54
Total	73

En la segunda búsqueda en ambas bases de datos se utilizaron como filtros el año de publicación entre el 2017 y 2019. En ProQuest se activó la casilla de evaluados por expertos y se buscaron las publicaciones con el tipo: Artículo principal, artículo, capítulo de libro, conferencia, disertación tesis, documento de referencia, estudio de caso, informe general, informe, informe de estadísticas, informe técnico, libro, ponencia, transcripción; así, se obtuvo 1 publicación utilizando el término de búsqueda T1 y 5 publicaciones utilizando como término de búsqueda a T2. En ScienceDirect se aplicaron los filtros mencionados y se buscaron las publicaciones con el tipo: research articles, book chapters, conference abstracts, software publication; y se obtuvieron 2 publicaciones utilizando el término de búsqueda T1 y 13 publicaciones utilizando el término de búsqueda T2. Esta información se encuentra resumida en las tablas 4 y 5.

TABLA 4
RESULTADOS DE LA SEGUNDA BÚSQUEDA CON EL TÉRMINO DE BÚSQUEDA T1.

Base de datos	“neural network” AND “breast cancer prediction system”
ProQuest	1
ScienceDirect	2
Total	3

TABLA 5
RESULTADOS DE LA SEGUNDA BÚSQUEDA CON EL TÉRMINO DE BÚSQUEDA T2.

Base de datos	"artificial neural network" AND ti("prediction") AND ti("breast cancer") NOT diabetes
ProQuest	6
ScienceDirect	13
Total	19

Por último, se aplicaron los criterios de inclusión y exclusión sobre los artículos filtrados de ambas bases de datos. De ProQuest se obtuvo 1 publicación con el término de búsqueda T1 y 2 publicaciones con el término de búsqueda T2. De ScienceDirect Se obtuvo 1 publicación con el término de búsqueda T1 y 5 publicaciones con el término de búsqueda T2. Los dos últimos criterios de exclusión fueron los más utilizados, puesto que existen muchos artículos relacionados con el tema, pero de temática únicamente médica sin mencionar algún programa informático. Del mismo modo, existieron publicaciones de sistemas predictivos, pero sobre cánceres diferentes al cáncer de mama o relacionados con otras enfermedades. El detalle de las publicaciones obtenidas se encuentra en las tablas 6 y 7.

TABLA 6
RESULTADOS DE APLICAR CRITERIOS DE INCLUSIÓN Y EXCLUSIÓN AL TÉRMINO DE BÚSQUEDA 1.

Base de datos	“neural network” AND “breast cancer prediction system”
ProQuest	[1]

ScienceDirect	[11] [12]
Total	3

TABLA 7

RESULTADOS DE APLICAR CRITERIOS DE INCLUSIÓN Y EXCLUSIÓN AL TÉRMINO DE BÚSQUEDA 2.

Base de datos	"artificial neural network" AND ti("prediction") AND ti("breast cancer") NOT diabetes
ProQuest	[2] [3] [4] [5] [6]
ScienceDirect	[7] [8] [9] [10]
Total	9

Resultados

Se procede a analizar los trabajos de la tabla 6 y 7. Ubicando el país donde se publicó la investigación y realizando un conteo de estos, podemos mostrar el número de investigaciones por país. Así, la India e Inglaterra son los países con mayor número de publicaciones en el tema planteado como se puede observar en la tabla 8.

TABLA 8

LISTADO DE PUBLICACIONES Y PAÍSES.

Países	Bases de datos		
	ProQuest	ScienceDirect	Total
India	[1]	[10] [12]	3
Inglaterra	[4] [5] [6]	-	3
Australia	-	[7]	1
EEUU	[3]	-	1
Egipto	-	[11]	1
Korea	-	[8]	1
Marruecos	-	[9]	1
Suecia	[2]	-	1
			12

Además, se extrajeron las características más utilizadas como datos de entrenamiento de la RNA. De este modo, se obtuvo que las características más utilizadas como datos de entrada son: edad, tamaño de tumor, imágenes y la radiación recibida(radioterapia). La información anteriormente nombrada queda detallada en la tabla 9.

TABLA 9
LISTADO DE LAS CARACTERÍSTICAS MÁS RECURRENTES UTILIZADAS COMO DATOS DE ENTRADA

Características	Bases de datos		
	ProQuest	ScienceDirect	Total
Age	[1] [2] [5] [6]	[7] [11]	3
Tumor Size	[1] [2] [6]	[7] [11]	1
Images	[3]	[8] [9] [10]	1
Radiation- chemotherapy	[1] [2] [6]	[7]	1
Stage of cancer	-	[8]	1
Grade	[2]	-	1
Race	[1]	[7] [11]	3
Lmph Node Involvement	[1] [6]	[7]	3
Marital status	[1] [5]	[7]	3
Primary site	[1] [2]	[11]	3
Radiation	[1]	[7]	2
Endocrine therapy	[1]	[7]	2
Histological Type	[1]	[7]	2
Behavior Code	[6] [2]	-	2
Number of nodes examined	[1]	[11]	2
Number of primaries	[1]	[7]	2
Clinical extensión of tumor	[1]	[7]	2
Gender	[5]	[7]	2
DNA	[4]	-	1
Cells features	-	[12]	1

Conclusiones

Con el presente artículo de revisión se puede concluir que la mayor cantidad de investigaciones sobre el tema de sistemas inteligentes para la predicción del cáncer de mama se encuentran en la India e Inglaterra. Así, queda en evidencia que la India tiene un continuo avance tecnológico en distintas áreas incluyendo en este tema en específico inclusive tiene mayor número de publicaciones que Estados Unidos.

Por otro lado, la característica más utilizada en los sistemas de predicción del cáncer de mama son la edad, el tamaño del tumor, imágenes de las ecografías y la radiación recibida por radioterapia. Entonces, esto nos indica que los sistemas predictivos pueden ser entrenados con una amplia gama de características y que existen unas más predominantes que otras.

También se puede concluir que en América Latina no existen publicaciones referentes al tema planteado lo que nos demuestra una escasa atención por un problema muy común y fatal si no se detecta a tiempo. Se espera que este artículo de revisión sirva para futuras investigaciones y despierte el interés en sistemas predictivos aplicados en el área de salud.

Referencias

[1] A. F. Nafees and Ritika, «A study on early prevention and detection of breast cancer using three-machine learning techniques,» *International Journal of Advanced Research in Computer Science*, vol. 9, pp. 37-42, 2018. [En línea]. Disponible en: <https://search.proquest.com/docview/2101240999?accountid=37610>

[2] C. Forsare, M. Bak, A. Falck, D. Grabau, F. Killander, P. Malmström, L. Rydén, O. Stal, M. Sundqvist, P. Bendajl and M. Ferno, "Non-linear transformations of age at diagnosis, tumor size, and number of positive lymph nodes in prediction of clinical outcome in breast cancer," *BMC Cancer*, vol. 18, 2018. [En línea]. Disponible en: <https://search.proquest.com/docview/2158422286?accountid=37610>.

[3] C. Selvi and M. Suganthi, "A Novel Enhanced Gray Scale Adaptive Method for Prediction of Breast Cancer," *J. Med. Syst.*, vol. 42, (11), pp. 1-8, 2018. [En línea]. Disponible en: <https://search.proquest.com/docview/2115641114?accountid=37610>.

[4] E. Manzanarez, D. Flores, E. Gutiérrez. Cervantes and P. Juárez, "Model based on GA and DNN for prediction of mRNA-Smad7 expression regulated by miRNAs in breast cancer," *Theoretical Biology and Medical Modeling*, vol. 15, 2018. [En línea]. Disponible en: <https://search.proquest.com/docview/2168586687?accountid=37610>.

[5] M. Chang, V. Viassolo, N. Probst, P. Chappuis and M. Katopodi, "Machine learning techniques for personalized breast cancer risk prediction: comparison with the BCRAT and BOADICEA models," *Breast Cancer Research*, vol. 21, 2019. [En línea]. Disponible en: <https://search.proquest.com/docview/2260267156?accountid=37610>.

[6] M. M. Bakre, C. Ramkumar, A. Kuman, C. Basavaraj, C. Prakash, L. Buturovic, L. Madhav, N. Naidu, P. R. S. Somashekar, S. Gupta, D. Chandra and M. Pegram, "Clinical validation of an immunohistochemistry-based CanAssist-Breast test for distant recurrence prediction in hormone receptor-positive breast cancer patients," *Cancer Medicine*, vol. 8, (4), pp. 1755-1764, 2019. [En línea]. Disponible en: <https://search.proquest.com/docview/2266308241?accountid=37610>.

[7] J. Yang, K. Than, M. Hagenbuchner and N. Shukla, "Breast cancer data analysis for survivability studies and prediction," *Computer Methods and Programs in Biomedicine*, vol. 155, pp. 199 – 208, 2018. [En línea]. Disponible en:

<http://www.sciencedirect.com/science/article/pii/S0169260717307551>

[8] D. My Vo, N. Nguyen and S. Lee, "Classification of breast cancer histology images using incremental boosting convolution networks," *Information Sciences*, vol. 482, pp. 123-138, 2019. [En línea]. Disponible en:

<http://www.sciencedirect.com/science/article/pii/S0020025518310466>

[9] W. Cherif," Optimization of K-NN algorithm by clustering and reliability coefficients: application to breast-cancer diagnosis," *Procedia Computer Science*,vol. [En línea]. Disponible en:

<http://www.sciencedirect.com/science/article/pii/S1877050918301376>

[10] I. Kumar, J. Virmani, H. Bhadauria, M. Panda, "Chapter 13 - Classification of Breast Density Patterns Using PNN, NFC, and SVM,Classifiers," *Soft Computing Based Medical Image Analysis*,Academic Press, pp 223-243,2018. [En línea]. Disponible en: <http://www.sciencedirect.com/science/article/pii/B9780128130872000129>

[11] S. Khodary M. Hamouda, M. Wahed, R. Abo and K. Riad, "Robust breast cancer prediction system based on rough set theory at National Cancer Institute of Egypt," *Computer Methods and Programs in Biomedicine*, vol. 153, pp. 259-268, 2018. [En línea]. Disponible en: <http://www.sciencedirect.com/science/article/pii/S0169260716314717>

[12]M. Kumari, V. Singh," Breast Cancer Prediction system," *Procedia Computer Science*, vol 132, pp. 371-376,2018. [En línea]. Disponible en: <http://www.sciencedirect.com/science/article/pii/S1877050918309323>

[13]" Boletín Empidemiológico Del Perú, "Ministerio de Salud del Perú,Perú, vol. 27, 2018.[En línea]. Disponible en: <https://www.dge.gob.pe/portal/docs/vigilancia/boletines/g2018/31.pdf>