

ARTÍCULO ACEPTADO

Usos de la Inteligencia Artificial para el diagnóstico temprano del cáncer de la mama

José Luis Llaguno Roque, Rocío Erandi Barrientos Martínez, Tania Romo González y Héctor Gabriel Acosta Mesa

Introducción

El cáncer de mama es una de las patologías más frecuentes a nivel mundial y una de las principales causas de muerte en la edad adulta. Se caracteriza por la propagación acelerada y descontrolada de células del epitelio glandular de la mama. Son células sanas que aumentaron su capacidad reproductiva, proliferando y multiplicándose hasta constituir un tumor, el cuál puede ser maligno o benigno dependiendo de sus características [1].

La Organización Mundial de la Salud (OMS) reportó en 2020 2.3 millones de mujeres diagnosticadas con cáncer de mama, de las cuales fallecieron 685,000 en todo el mundo. La buena noticia es que al evitar hábitos pocos saludables (como obesidad, consumo excesivo de alcohol, tabaco ó mala alimentación) y factores de riesgo [2], aunado a la aplicación de estrategias de prevención basadas en investigaciones científicas, es posible reducir entre un 30 % y 50 % la incidencia de esta patología [3].

En México, la historia no es diferente, ya que es la segunda causa de muerte, solo por detrás de las enfermedades cardiovasculares y diabetes, llegando a convertirse en un problema de salud pública y un serio desafío para el sistema de salud mexicano. Las estadísticas indican que en 2020 fallecieron 7821 mujeres por cáncer de mama, lo cual representa el 8 % de todos los fallecimientos por tumores malignos.

Las pruebas tradicionales que se utilizan para detectar el cáncer de mama son: exploración clínica de la mama, ecografía, mastografía, y toma de biopsia [4]. Estas pruebas en general son invasivas, subjetivas, caras, y en la mayoría de los casos dolorosas, por lo que resultan ineficaces en la detección temprana del cáncer y en la disminución de la mortalidad [3]. La biopsia, es de las pruebas más certeras, ya que se extrae una muestra de células mamarias, las cuales se envían a un laboratorio para su análisis, para determinar si las células son o no cancerosas.

Desafortunadamente en la mayoría de los casos, la detección del cáncer de mama en México se realiza en las últimas etapas de la enfermedad, lo cual dificulta la posibilidad de realizar un tratamiento eficaz, y por en-

de exitoso; es por esto que la esperanza de vida de las pacientes mexicanas es muy baja en relación con países desarrollados [5]. Ésta es la razón por la cual es necesario generar métodos diagnósticos más tempranos, eficaces y menos invasivos.

La Inteligencia Artificial (IA) puede apoyar en esta tarea tan compleja, a través de técnicas de aprendizaje automático, el cual es un campo de la IA que a través de algoritmos dota a las computadoras de capacidad para identificar patrones en conjuntos de datos para producir conocimiento que apoye a la toma de decisiones, mediante la construcción de modelos matemáticos. Este aprendizaje permite a las computadoras aprender a realizar tareas específicas de forma autónoma, sin haber sido explícitamente programadas para ello, mejorando su desempeño a partir de la experiencia adquirida a través de ejemplos.

En el área de la medicina, la IA ha asistido en procesos como la prevención y el diagnóstico del cáncer de mama usando técnicas de aprendizaje automático y visión por computadora. La mayoría de las aportaciones se han hecho en el procesamiento de imágenes de ultrasonido, termográficas y mastográficas, para la identificación de patrones que permiten clasificar a mujeres que presentan patología mamaria de aquellas que no lo presentan [6].

Inteligencia Artificial y el diagnóstico temprano de cáncer de mama

Investigaciones recientes han mostrado cómo la presencia de auto-anticuerpos puede funcionar como un buen método de diagnóstico temprano de cáncer de mama. Éstas se basan en la premisa de que las células tumorales presentan en su superficie antígenos que al ser detectados por el sistema inmunológico desencadenan una respuesta antitumoral. Dentro de esta respuesta se encuentra la producción de auto-anticuerpos, los cuales pueden ser utilizados como una señal carcinogénica antes que se manifieste la tumoración en el organismo y sin ser invasivo a diferencia de los métodos tradicionales [7].

En medicina, la IA ha asistido en procesos como la prevención y el diagnóstico del cáncer de mama.

En México, Romo-González et al. [8] desarrollaron un método que corrobora la presencia de auto-anticuerpos contra células tumorales en ductos específicos de la mama, lo que permite distinguir a las mujeres con y sin patología mamaria. Esto a través del análisis de bandas expresadas en imágenes unidimensionales, de la reacción de los auto-anticuerpos con antígenos en una cierta línea tumoral. Las bandas se obtuvieron mediante la técnica de laboratorio Western Blot, que detecta una proteína específica de una muestra de sangre y el riesgo al paciente es mínimo. Si bien los resultados son alentadores, el análisis de las imágenes es complejo, subjetivo y tardado, con tiempos de espera de hasta un mes para lograr crear la base de datos binaria (1 presencia, 0 ausencia de proteína), ya que el análisis de las imágenes requiere que el experto alinee las bandas de las tiras de cada paciente con la ayuda de algún software comercial, pero la identificación final de las bandas y su posición depende del ojo del experto, por lo tanto la identificación de estos patrones de bandeo requiere de herramientas más precisas y automáticas.

En el área de IA, Sánchez-Silva et al. [9] propuso un sistema semiautomático para el análisis de bandas de proteínas, mediante la clasificación de patrones representados como series de tiempo. Para la clasificación usó el algoritmo K-vecinos más cercano (K-NN) con distancias de similitud Euclidiana, Mahalanobis y Correlación, obteniendo un 66.40% de precisión en la clasificación con tres clases (saludable, patología benigna de mama, cáncer de mama) y un 86.36% de precisión en la clasificación con dos clases (saludable, cáncer de mama). Dichos porcentajes fueron similares a los reportados con el método tradicional utilizado por el experto, sin embargo, la representación continua por series de tiempo es semiautomática y evita la subjetividad y complejidad del proceso manual de binarización de la base de datos. Aún es necesario mejorar el método y automatizarlo mediante el desarrollo de un software que permita analizar cualquier imagen de patrones de bandeo de una manera fácil y rápida.

Metodología

En la Figura 1 se propone una metodología que permitirá a partir de imágenes de patrones de bandeo, evaluar el algoritmo de aprendizaje supervisado K-NN, con una distancia de similitud llamada DTW, la cual mide la similitud entre series de tiempo de distintas longitudes, esperando mejorar la precisión de la clasificación de datos obtenidos de la presencia de auto-anticuerpos a antígenos tumorales. Se tendrán dos escenarios, el primero con tres clases de series de tiempo (saludable, patología benigna de mama, cáncer de mama) y el segundo con dos clases (saludable, cáncer de mama).

Los detalles de cada proceso de la metodología son los siguientes:

Imágenes de Western blot. Se han utilizado 149 imágenes, 50 corresponden a pacientes con cáncer de mama, 50 a pacientes con patología benigna de mama y 49 a pacientes saludables. Las imágenes han sido proporcionadas por el área de Biología y Salud integral del Instituto de Investigaciones Biológicas de la Universidad Veracruzana, siguiendo las normas éticas y con el correspondiente consentimiento informado de las participantes. El protocolo ha sido revisado y aprobado por el Comité de Ética en Investigación del Hospital General de México Dr. Eduardo Liceaga” (DI/12/111/03/064). El estudio se ajusta al Código de Ética de la Asociación Médica Mundial (Declaración de Helsinki), impreso en el British Medical Journal (18 de julio de 1964).

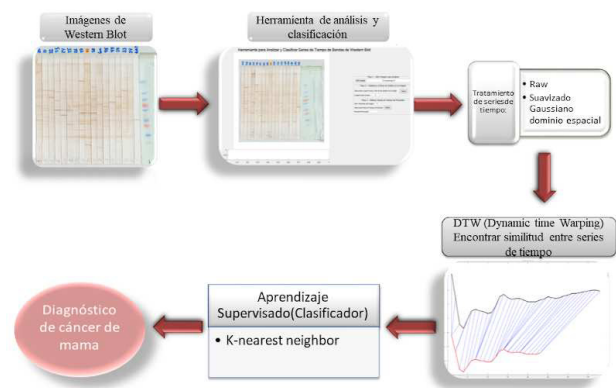


Figura 1. Metodología.

Herramienta de análisis y clasificación. Se ha desarrollado en el lenguaje de programación “Matlab”, para obtener las series de tiempo de datos crudos (Raw).

Obtener series de tiempo. Las series de tiempo se han obtenido delimitando el área en donde se encuentran los patrones de bandeo en las imágenes. Estas se representan a partir de la variación de las tonalidades de los píxeles de las bandas. Ver Figura 2.

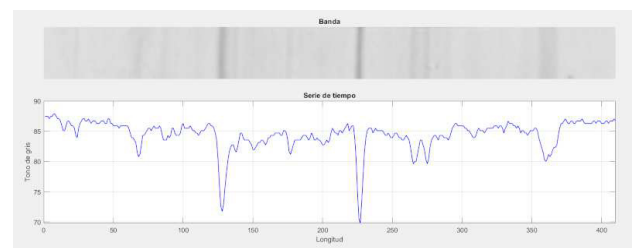


Figura 2. Serie de Tiempo que se obtiene de las imágenes al promediar los diferentes valores de los renglones que pertenecen a cada columna, por lo que de una imagen de N (filas) x M (columnas) se obtiene una serie de tiempo de 1 x M.

Medida de similitud entre series de tiempo. Se ha utilizado una técnica de inteligencia artificial que permite encontrar una medida de similitud entre series de tiempo de diferentes longitudes y desfasadas en el tiempo. La búsqueda de la medida de similitud ha sido restringida por una ventana con valor de 0.06 (6%), 0.10 (10%), 0.20 (20%), que corresponde al porcentaje de la longitud de la serie de tiempo con mayor longitud.

Tratamiento de series de tiempo. Se han utilizado las series de tiempo en dos formas: 1) con valores crudos y 2) con suavizado gaussiano en el dominio del espacio, donde se han usado técnicas de visión computacional, ya que corresponde a un filtro pasa bajas que permite mejorar la calidad de una imagen mediante el realce de ciertas características, así como atenuar el ruido. En la Figura 3 se muestran un ejemplo de ambas series.

Clasificación de series de tiempo. Se ha evaluado el desempeño del algoritmo de aprendizaje supervisado K-NN. Para los experimentos propuestos, el valor de K ha sido de 1, 3, 5, 7, 10.

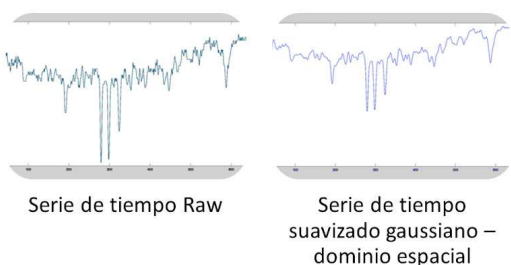


Figura 3. Ejemplos del procesamiento de las series de tiempo.

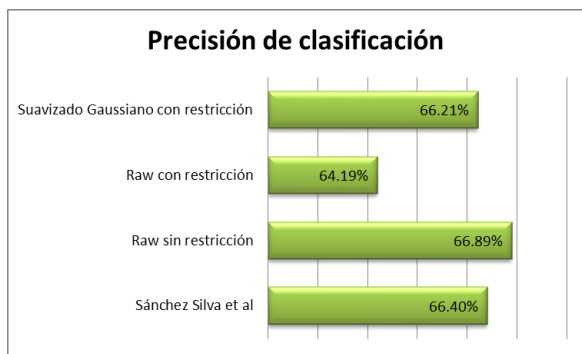


Figura 4. Resultados de clasificación de series de tiempo con tres clases.

Resultados

Se han llevado a cabo dos experimentos a partir de la metodología propuesta. En la Figura 4 se muestran los resultados de los experimentos que se realizaron con tres clases. La mejor precisión de clasificación que se obtuvo

fue de 66.88% con las series de tiempo con datos crudos (RAW), con K=3 en el clasificador K-NN y la distancia de similitud sin ventana de restricción.

En la Figura 5 se muestran los resultados de los experimentos que se realizaron con dos clases. La mejor precisión de clasificación que se obtuvo fue de 88.77% con las series de tiempo con datos crudos, con K=3 en el clasificador K-NN y la distancia de similitud sin ventana de restricción.

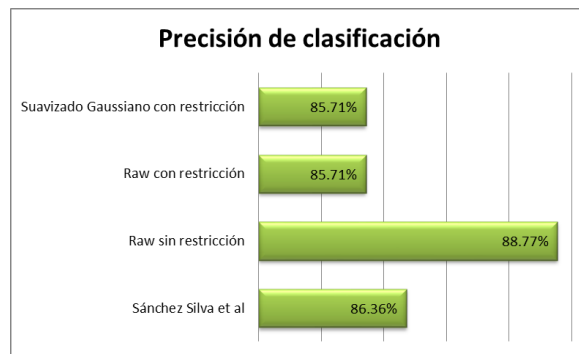


Figura 5. Resultados de clasificación de series de tiempo con dos clases.

Conclusiones

El cáncer de mama es uno de los mayores desafíos en el sistema de salud mundial. La detección temprana del cáncer de mama mediante la respuesta del sistema inmunológico a antígenos tumorales, permitirá tener un mejor pronóstico en el tratamiento, así como el descenso en la mortalidad de las pacientes.

Los resultados obtenidos en los experimentos preliminares para la discriminación de las series de tiempo mediante el algoritmo de clasificación K-NN, con la distancia de similitud DTW, sugieren un ligero aumento en la precisión de clasificación con respecto a lo reportado en [8], sin embargo, los valores no son concluyentes en estos momentos, toda vez que es necesario aumentar la casuística de estudio para obtener tendencias estadísticamente significativas.

En cuanto a la metodología propuesta, ha sido posible explorar técnicas de aprendizaje automático y visión por computadora que han permitido obtener las medidas de similitud de las series de tiempo sin necesidad de utilizar la transformación geométrica de escalamiento, así como también atenuar el ruido de las series de tiempo con el uso del filtro gaussiano, con la finalidad de mejorar el desempeño de la precisión de clasificación.

Si bien esta metodología es susceptible de ser mejorada, y más experimentación es necesaria para poder aplicarla y así obtener un beneficio social, aporta avances y conocimiento que pueden servir de base para futuras investigaciones con mejores resultados.*

REFERENCIAS

1. Collado, R. (2011). Hablemos del cáncer de mama con Roche. Recuperado el 13 de Junio de 2021, de https://www.seom.org/seomcms/images/stories/recursos/infopublico/publicaciones/HABLEMOS_CANCER_MAMA.pdf.
2. Carvalho-Fernández, A.F., Mesquita-Melo, E., de-Almeida-Araújo, I.M. y Figueiredo-Carvalho, Z.M. (2005). Aspectos culturales en el proceso de padecer cáncer de mama. Avances en Enfermería. Recuperado el 8 de Junio del 2022, de <https://revistas.unal.edu.co/index.php/avenferm/article/view/37559>.
3. Chávarri-Guerra, Y., Villarreal-Garza, C., et al. (2012). Breast cancer in Mexico: a growing challenge to health and the health system. *Lancet Oncol*, 13(8), 335-343.
4. División de Excelencia Clínica. (2011). Diagnóstico y Tratamiento de la Patología Mamaria Benigna en Primer y Segundo Nivel de Atención. IMSS, ISBN 978-607-8270-73-6, 71.
5. Brandan, M.E. (2006). Detección del Cáncer de Mama: Estado de la Mamografía en México. *Revista del Instituto Nacional de Cancerología*, 27(3), 16.
6. Zhang, M. et al. (2021). Background Parenchymal Enhancement on Breast MRI as a Prognostic Surrogate: Correlation With Breast Cancer Oncotype Dx Score. *Frontiers in Oncology*, 10, 1-7.
7. Piura, E. y Piura, B. (2011). Autoantibodies to Tailor-Made Panels of Tumor-Associated Antigens in Breast Carcinoma. *Journal of Oncology*, 2011, 1-7.
8. Romo-González, T. et al. (2015). The Network of Antigen-Antibody Reactions in Adult Women with Breast Cancer or Benign Breast Pathology or without Breast Pathology. *PLoS ONE*, 10(3), 1-21.
9. Sánchez-Silva, D.M., Acosta-Mesa, H.G. y Romo-González, T. (2018). Semi-Automatic Analysis for Unidimensional Immunoblot Images to Discriminate Breast Cancer Cases Using Time Series Data Mining. *International Journal of Pattern Recognition and Artificial Intelligence*, 32(1), 1-20.

SOBRE LOS AUTORES



José Luis Llaguno Roque es estudiante de Doctorado en Biología Integrativa del Instituto de Investigaciones Biológicas de la Universidad Veracruzana. Maestro en Inteligencia Artificial y Licenciado en Informática por la Universidad Veracruzana. Sus intereses son visión computacional y aprendizaje automático.



Rocío Erandi Barrientos Martínez es Maestra en Inteligencia Artificial y Licenciada en Informática por la Universidad Veracruzana. Técnico-Académico del Instituto de Investigaciones en Inteligencia Artificial de la Universidad Veracruzana. Áreas de Interés: Aprendizaje automático y visión por computadora.



Tania Romo González de la Parra es Química Farmacéutica Bióloga por la Universidad Veracruzana. Doctorado en Ciencias Biomédicas por la UNAM y Posdoctorado en Cáncer de Mama en el Departamento de Inmunología del Instituto de Investigaciones Biomédicas de la UNAM. Investigadora en el Instituto de Investigaciones Biológicas de la Universidad Veracruzana en donde desarrolla diversos proyectos de investigación en las líneas 1) Psiconeuroinmunología, salud y bienestar y 2) estilos de vida y salud.



Héctor Gabriel Acosta Mesa es Doctor en Inteligencia Artificial por la Universidad de Sheffield, Inglaterra. Maestro en Inteligencia Artificial por la Universidad Veracruzana. Ingeniero en Sistemas Computacionales por el Instituto Tecnológico de Veracruz. Profesor-Investigador de tiempo completo del Instituto de Investigaciones en Inteligencia Artificial de la Universidad Veracruzana. Miembro del Sistema Nacional de Investigadores del Consejo Nacional de Ciencia y Tecnología. Áreas de interés: Aprendizaje automático, visión por computadora y sus aplicaciones.