

Procesamiento de mamografías para detección de tumores

Carmen Jimena Santizo Monterroso, Dr. –Ing. José Leal
Departamento de Ingeniería Biomédica
san20449@uvg.edu.gt, jaleal@uvg.edu.gt

Resumen

La mamografía supone ser el estándar de oro en la detección temprana del cáncer de mama, incrementando las posibilidades de supervivencia. Sin embargo, en algunos casos la realización de esta implica incomodidad, dolor y sensibilidad. Además, el análisis y la interpretación de éstas se basa en una inspección visual con una valoración subjetiva y cualitativa; lo que implica posibles falsos positivos o negativos. Por otra parte, cuando las mamas están conformadas mayormente por tejido glandular, se dan enmascaramientos por la propia densidad del seno.¹ Por ende, se desarrolló con MATLAB una herramienta de apoyo para el diagnóstico y detección temprana de cánceres mamarios, capaz de identificar lesiones ocultas no palpables y en estadios precoces; que en el contexto médico requerirían estudios complementarios molestos. Con el uso de las características radiómicas del seno y Machine Learning (ML), se clasifican las masas correctamente entre benignas y malignas; incrementando las posibilidades de supervivencia.

En el mundo

2da. causa de mortalidad por cáncer²
1 nuevo caso cada 18 seg

En Guatemala

> 2,244 casos identificados en el 2023²

Objetivos

Objetivo General

- Crear un algoritmo de procesamiento de imágenes médicas mamográficas, capaz de diferenciar tumores mamarios entre benignos y malignos.

Objetivos Específicos

- Desarrollar algoritmo de preprocesamiento de imágenes médicas obtenidas de un repositorio, reduciendo los artefactos de la adquisición y estandarizando las imágenes para análisis posterior.
- Identificar los mejores métodos para segmentar las regiones de interés en las imágenes preprocesadas a partir de las características radiómicas del seno, para la detección de masas mamarias.
- Entrenar algoritmos de ML con imágenes médicas obtenidas de repositorios para la detección y diferenciación automática entre tumores benignos y malignos.
- Obtener imágenes médicas de pacientes guatemaltecas y validar el funcionamiento del algoritmo en colaboración de expertos en radiología.

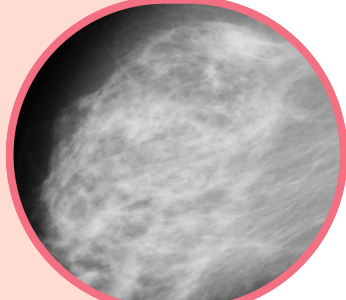
Métodos

El procesamiento de mamografías se realiza en MATLAB con un banco de 322 mamografías reales (MIAS, UK) y datos asociados por medio de un informe. El flujo de trabajo realiza ajustes a las mamografías y resalta patrones de calcificaciones.



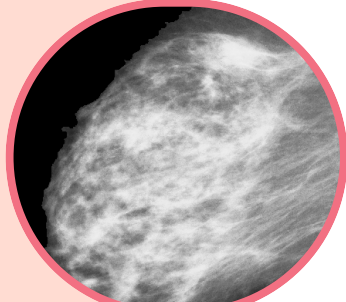
Importación

- Visualización, carga y almacenamiento de imágenes.
- Extracción de datos e información asociada.
- Importación de mamografías de forma automática.



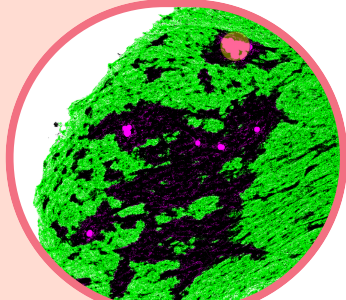
Pre procesamiento

- Reducción de cualquier artefacto de adquisición.
- Eliminación de fondo `imclose()` y ruido `medfilt()`.
- Normalización de intensidad y contraste `histeq()`.



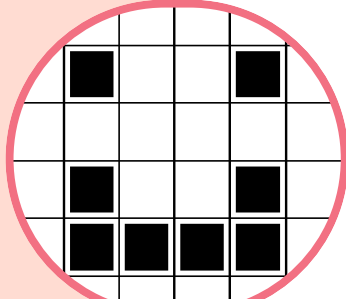
Segmentación

- Limpieza de bordes `imclearborder()`.
- 2 bandas de colores (**lesión** y **tejido sano**).
- Regiones de interés (**ROI**) para validar hallazgos.



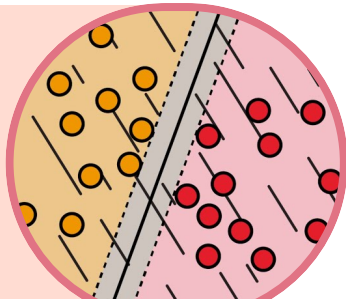
Post procesamiento

- Extracción de características `extractLBPFeatures()` y propiedades de las mamografías `regionprops()`.
- Recopilación de información y patrones para modelo.



Clasificación

- Etiquetas para modelo de ML: 1-Benigno y 2-Maligno.
- Entrenamiento modelo SVM (20% training y 80% test).
- Matriz de confusión para evaluar rendimiento.



Resultados

Tras 15 versiones de algoritmos, se logran observar tejidos de distintas densidades y un realce de zonas con calcificaciones. La segmentación del músculo pectoral se realiza manualmente para cada estudio, no se logra automatizar el proceso. Se obtienen 7 modelos de ML supervisado, siendo el de SVM con mayor precisión para una clasificación en dos clases (benigno y maligno).

Métricas de Rendimiento para el Modelo Support Vector Machine (SVM)

- 98% de precisión, para predecir que un tumor es de tipo maligno.
- 86% de sensibilidad, alta capacidad para identificar casos positivos reales.
- 50% de especificidad, baja capacidad para identificar casos negativos.

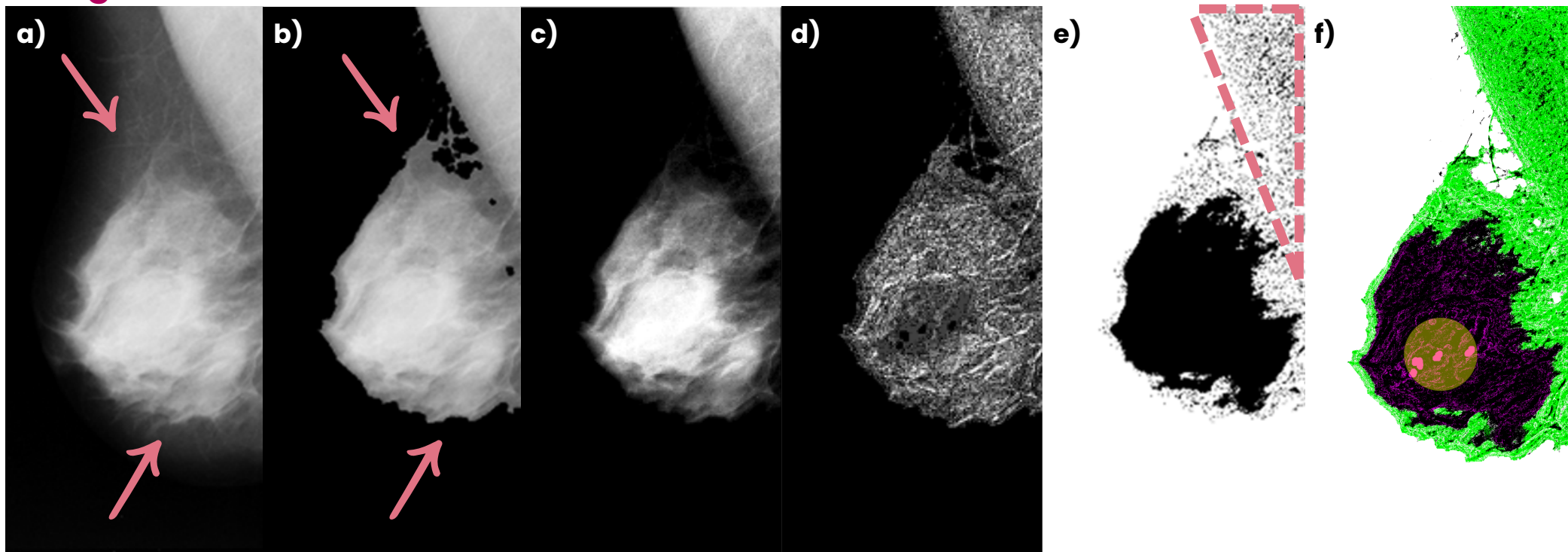
	+	-
+	53	1
-	9	1

Precisión General

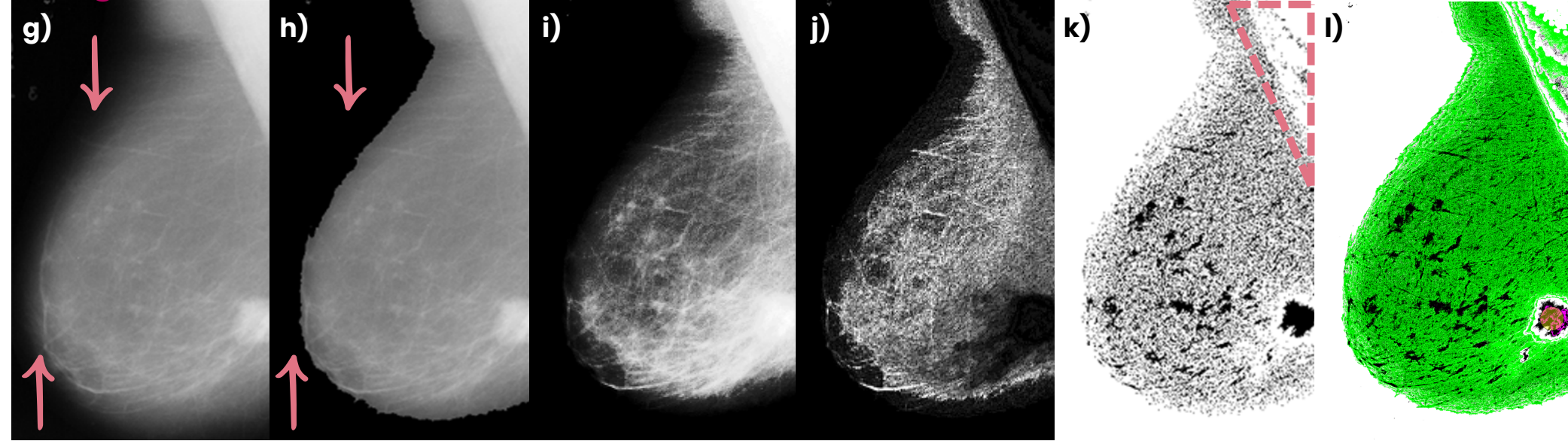
84 %

- Alta exactitud para predecir etiquetas correctas de casos desconocidos.
- 84 de cada 100 predicciones realizadas por el modelo son acertadas, tanto en casos benignos (negativos) como malignos (positivos).

Benigno



Maligno



(a) (g) Mamografía cruda. (b) (h) Menor distorsión, mejora de textura y bordes circunscritos. (c) (i) Mejora del contraste. (d) (j) Corrección de iluminación e intensidad. (e) (k) Segmentación del músculo pectoral. (f) (l) Áreas con **presencia de calcificaciones**, áreas con **tejido propio de la mama sin lesiones** y **ROI** (x, y).

Conclusiones

- El algoritmo creado es una herramienta de apoyo para la detección temprana del cáncer de mama, proporcionando información adicional para efectuar diagnósticos en conjunto con la opinión médica y variables no implícitas.
- El flujo de trabajo permite obtener imágenes de mayor calidad y sin artefactos, con una segmentación correcta entre tejido sano y posibles calcificaciones.
- En general todos los casos se segregan correctamente, con un mayor alcance en casos con mamas con un patrón denso y en su mayoría tejido glandular, que enmascaran las lesiones y que no son posible apreciar por inspección visual. Se hallan lesiones adicionales no reportadas por el informe de la base de datos.
- La segmentación del músculo pectoral repercute en la extracción de características, ya que su histología representa una posible área de calcificación.
- El modelo SVM detecta y clasifica correctamente tumores malignos y benignos a partir de *features* extraídas, con un 84% de precisión para el conjunto de prueba.
- Se predice eficientemente la clasificación de mamografías desconocidas de pacientes guatemaltecas, en su mayoría con lesiones halladas correctamente.

Trabajo Futuro

En el ámbito de diagnóstico médico, se prevé que el algoritmo sea empleado como una herramienta para el tamizaje temprano de cánceres mamarios, especialmente en estadios precoces y en casos con mamas densas. Sin embargo, aunque las métricas de rendimiento parecen aceptables en otros escenarios, en el contexto clínico no es suficiente para garantizar seguridad. Esto representa un 16% riesgo de categorizar un cáncer maligno cuando es benigno y viceversa. Por ende, es necesario aplicar *deep learning* (CNN), un modelo que aprende a partir del reconocimiento de patrones visuales de manera similar a las neuronas de un cerebro biológico. A través de CNN, evaluar una clasificación según el sistema BI-RADS (7 categorías de 0 a 6), catalogando a 1 como normal o benigno y a 6 como altamente sospechoso o maligno; con un mayor volumen de mamografías y variables asociadas del paciente.

Referencias

- OMS, (2024), Cáncer de mama.
- Kihn-Alarcón, et al., (2022), Breast cancer in young women in Guatemala.

Agradecimientos

Dr. Aldo Mario Dardón Aguilar del Hospital El Pilar.

