

Desarrollo de una base de datos de imágenes y su protocolo para futuro entrenamiento de IAs especializadas en la detección de melanomas

Chávez Domínguez, Samantha

2024

<https://hdl.handle.net/20.500.11777/6138>

<http://repositorio.iberopuebla.mx/licencia.pdf>

Desarrollo de una base de datos de imágenes y su protocolo para futuro entrenamiento de IAs especializadas en la detección de Melanomas

Chávez Domínguez Samantha (onceavo semestre en Ingeniería Biomédica)¹, Díaz Victoriano Reyes de Jesús (décimo semestre en Ingeniería Biomédica)¹, Reyes González Julieta (noveno semestre en Ingeniería Biomédica)¹, Saldaña de los Santos Rosario (onceavo semestre en Ingeniería Biomédica)¹, Merino Arroyo Carlos Miguel (profesor responsable)¹, Grande Barreto Jonás (profesor asesor)^{1,2}, Moreno Hernández Ana (profesor asesor)¹ Abraham Castro Ponce (médico asesor)³, Diego Alberto Juárez Salazar (colaborador)³.

¹Universidad Iberoamericana Puebla, San Andrés Cholula, Puebla, México, ²Universidad Politécnica Metropolitana de Puebla, Puebla, México y ³Centro Oncológico Integral, Puebla, México.

Palabras clave: Base de datos, Detección, Dermatoscopia, Etiquetado, Melanoma.

***Autor Corresponsal:** rosario.saldana@iberopuebla.mx

Introducción

El melanoma es un cáncer de piel agresivo con alto riesgo de metástasis. La detección temprana, basada en el método "ABCDE" (Asimetría, Borde, Color, Diámetro y Evolución), es crucial para mejorar el pronóstico [1]. Este proyecto busca facilitar la detección automática del melanoma mediante un protocolo que permite a los médicos subir y evaluar imágenes de lunares según los criterios "ABCDE", brindando una orientación inicial sin sustituir la consulta médica. Además, se ha creado una base de datos para almacenar y organizar estas evaluaciones, útil para el análisis posterior y futuros estudios. [2]

Metodología

En colaboración con el COI, se seleccionó la base de datos ISIC (*International Skin Imaging Collaboration*) del año 2017, la cuál es de uso público para investigadores interesados en el tema, está compuesta por 2000 imágenes de personas con tonos de piel más claros, provenientes de Norteamérica, debido a la falta de una base de datos mexicana abierta al público, las imágenes usadas, están captadas en alta resolución con dermatoscopio y un archivo .CSV que clasifica las lesiones en melanoma, queratosis o sano. [1] La metodología ABCD, ampliamente investigada para la detección de melanomas, se empleó para analizar las imágenes, omitiendo la "E" (Evolución) por requerir seguimiento clínico. [2] Esta metodología evalúa la asimetría (A), borde (B), color (C) y diámetro (D) de las lesiones. Los melanomas suelen ser irregulares y asimétricos, presentan bordes dentados, muestran múltiples colores (marrón, negro, rojo, blanco, azul) y superan los 6 mm de diámetro. Para la segmentación de las imágenes, se implementó el modelo SAM2, un sistema de META configurado para generar máscaras de la lesión y eliminar áreas irrelevantes. [3] El código desarrollado del proyecto, se ejecutó en la tarjeta gráfica Nvidia RTX A6000, optimizando el procesamiento en paralelo con CUDA, esencial para el modelo SAM2, cabe recalcar que el modelo fue implementado en el código, haciendo modificaciones en el modelo para que se base únicamente en las características de los lunares. [4] Para la extracción de características, se implementaron algoritmos que emulan criterios médicos. K-Means identificó colores dominantes y su porcentaje en la imagen, el cual se usó principalmente por practicidad. Otros algoritmos calcularon área, perímetro, asimetría (con *Intersection over Union*), irregularidad de bordes (comparando perímetro real e ideal), el índice de Simpson (uniformidad de colores) y la desviación estándar de los colores. Estos parámetros permiten un análisis más preciso de las lesiones. Las métricas usadas en el proyecto, fueron basadas en una investigación realizada por la Facultad de Ingenierías del Campus Guindy en India, siendo revisadas previamente por los investigadores clínicos del COI. [1] Posteriormente se realizó el protocolo con los pasos a seguir para la implementación de imágenes a la base de datos.

1. Depilar la zona sobre y alrededor de la lesión, esto para evitar la interferencia de las vellosidades capilares que pueden ser interpretadas por el algoritmo como el objeto de interés a analizar o generar falsos objetos al crear bordes falsos debido a su distribución.
2. Limpiar la zona y asegurarse de que no haya objetos en el área focal debido a que la principal función del algoritmo SAM2 es encontrar y segmentar objetos dentro de la zona de fotografía y a pesar de que la configuración está dada para detectar lunares puede confundirse si hay más figuras que hagan contraste dentro de la imagen
3. Utilizar un dispositivo de captura digital que permita ajustar el brillo y el tiempo de obturación. El médico o captador que este tomando la fotografía debe cuidar que haya una homogeneidad en la distribución de la iluminación de la imagen para evitar una saturación que impida la detección de los colores reales de la imagen y

- también para evitar la pérdida de bordes debido a la difuminación que genera el reflejo excesivo de la piel al lente.
4. Evitar la generación de viñetas o contornos alrededor de la imagen, ya que complican el procesamiento. La generación de contornos en la imagen, así como de viñetas crea un conflicto en el algoritmo que al encontrar esto como un borde definido, lo toma como una entidad a segmentar, ignorando las configuraciones de detección de lunares y tomando como objeto a segmentar todo lo que se encuentre dentro de la viñeta o contorno.
 5. Si el dispositivo no permite evitar los contornos, recortar la imagen después de tomarla. En caso de que el modelo del dermatoscopio sea una apertura inferior a la apertura del lente y genere un efecto de microscopio, recortar solo la parte interna evitando los bordes para no generar objetos falsos.
 6. Si persisten los contornos, recortar la imagen como se indicó en el paso anterior. En caso de que tras el primer recorte de imagen no haya sido suficiente para eliminar todos los contornos, hacer otro recorte a la imagen.
 7. Guardar la imagen en formato .jpg o .png, en alta definición (HD) y relación 16:9. La calidad HD establece una resolución de 1280p x 720p, esto se establece para representar la mayor calidad posible de valores del lunar en una imagen y el mantener una relación 16:9 permite estandarizar los procesos de análisis de imagen. En caso de que la imagen no cumpla con estos requisitos tras realizar los puntos 4-6, descartar imagen.
 8. No utilizar el flash del dispositivo, solo la luz del dermatoscopio, para evitar puntos blancos o saturación. El uso del flash propio del dispositivo de captura puede generar conflictos debido a que los dermatoscopios cuentan con luz propia, también el uso de un flash externo puede generar reflejos en el lente que son captados por las cámaras y al ser puntos muy brillantes, el algoritmo pueden confundirlo como el área de interés a analizar.
 9. Al ingresar la imagen, debe incluirse la edad, el sexo biológico y el diagnóstico confirmado por biopsia, indicando solo si es maligno o benigno.

Análisis de Resultados

Debido a las características del funcionamiento de nuestro algoritmo se procesaron solo 690 imágenes de los 2000 originales. El análisis de los resultados se basó en la segmentación de una lesión en la piel usando el modelo SAM2, donde la lesión se destaca en color morado para facilitar su análisis (Fig.1a, anexo). Posteriormente, se muestra un recorte de la zona de interés sin el fondo de la piel, simplificando la imagen para enfocar el análisis en las características específicas de la lesión (Fig. 1b, anexo). Se presenta también una máscara binaria junto con un desglose de los colores dominantes y sus porcentajes, lo que ayuda a identificar patrones de color asociados con variaciones en la pigmentación (Fig.1c, anexo). Finalmente, se muestra una tabla con datos cuantitativos sobre la lesión segmentada, incluyendo color dominante, porcentaje de cada color en los planos RGB, área, perímetro, diámetro, asimetría, irregularidad, índice de Simpson y desviación estándar. Esto son cruciales para analizar la forma, tamaño y patrones de color de la lesión, contribuyendo a evaluar la posible malignidad en casos sospechosos de melanoma, por último se realizó la validación clínica del sistema de detección de melanoma se llevó a cabo con una muestra representativa de 60 imágenes (Fig.2a, anexo), bajo la supervisión del médico asesor del COI, quien evaluó visualmente las mismas para verificar los resultados dados por el algoritmo, con base a su criterio médico, así mismo asignó una calificación binaria (0 o 1) a cada variable ABCD, indicando la presencia o ausencia de indicios de melanoma (Fig.2b, anexo). En las gráficas de dispersión, se correlacionaron los valores del algoritmo con las características de las lesiones y la calificación médica, que indica si son o no candidatos a melanoma. Cada punto representa un valor del algoritmo, clasificado como melanoma (1) o no (0) según el médico. Sin embargo, aún no se puede estandarizar, debido a que los valores están distribuidos de manera uniforme en las gráficas, lo que dificulta observar diferencias claras entre ambas categorías (Fig.3a y Fig. 3b, anexo).

Conclusiones

Este proyecto, en colaboración con el Centro Oncológico Integral (COI), ha logrado desarrollar una base de datos de imágenes de melanomas, clasificadas según el método "ABCDE del melanoma", con el fin de entrenar modelos de inteligencia artificial (IA) que mejoren la detección temprana de esta enfermedad. Utilizando imágenes de alta resolución y algoritmos avanzados como SAM2, se ha creado un protocolo que permite a los médicos evaluar lunares sospechosos de manera inicial, sin reemplazar la consulta médica. Aunque sigue en fase de desarrollo, la herramienta promete ser útil para médicos generales y dermatólogos, ayudando a identificar pacientes con mayor riesgo de melanoma. Los resultados preliminares son alentadores y sugieren que esta tecnología podría mejorar la precisión en la detección temprana, especialmente en comunidades con recursos limitados y áreas rurales. En las próximas mejoras del proyecto, se tiene contemplado ampliar la base de datos e incluir más imágenes de personas con piel más oscura para asegurar su representatividad, así como perfeccionar los algoritmos con herramientas complementarias como la textura y características morfológicas avanzadas. Este enfoque ampliado contribuirá a la mejora del diagnóstico temprano y facilitará el tratamiento del melanoma en el futuro.

Anexo de imágenes

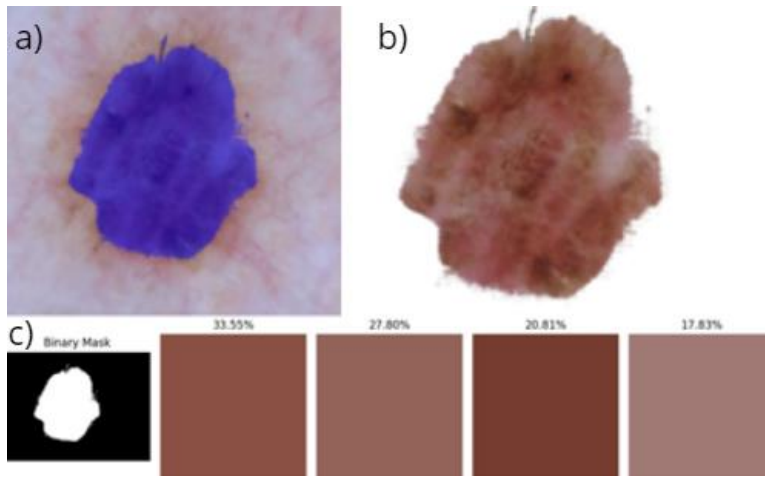


Fig. 1 (a) Segmentación de la zona de interés, (b) Parte de la lesión de interés, (c) Máscara binaria desplegada con colores identificados y porcentaje correspondiente a cada uno de ellos

Tabla de Comparación:

	Nombre Archivo Pequeño	Número de Fila	Asimetría_Code	Bordes_Code	Color_Code	Diámetro_Code
0	ISIC_0000003_cropped_mask.png	703	1.00	1.47	0.73	333.69
1	ISIC_0000013_cropped_mask.png	287	0.58	1.43	0.64	902.29
2	ISIC_0000016_cropped_mask.png	329	0.23	1.39	0.74	660.74
3	ISIC_0000018_cropped_mask.png	412	0.44	1.32	0.68	752.69
4	ISIC_0000019_cropped_mask.png	549	0.37	1.73	0.75	412.99
5	ISIC_0000021_cropped_mask.png	661	0.87	1.61	0.70	858.02
6	ISIC_0000022_cropped_mask.png	320	0.33	1.38	0.64	1137.37

	Nombre Archivo Pequeño	ASIMETRÍA	BORDES	COLOR	DIÁMETRO
0	ISIC_0000003_cropped_mask.png	1	1	1	1
1	ISIC_0000013_cropped_mask.png	0	1	1	0

Fig. 2 (a) Datos guardados de las características de la imagen segmentada dentro del data set, (b) Clasificación de características por parte del área clínica

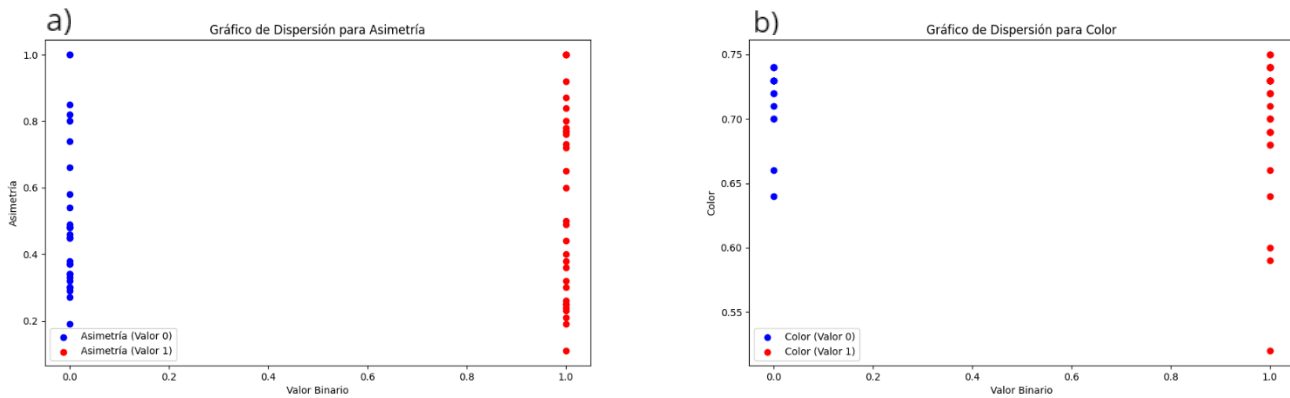


Fig. 3 (a) Gráfica de dispersión valores de asimetría, (b) Gráfica de dispersión valores de color

Referencias

1. H. Tsao et al., “Early detection of melanoma: Reviewing the ABCDEs”, *J. Amer. Acad. Dermatol.*, vol. 72, n.º 4, pp. 717–723, abril de 2015. Accedido el 14 de noviembre de 2024. [En línea]. Disponible: <https://doi.org/10.1016/j.jaad.2015.01.025>
2. C. K. Viknesh, P. N. Kumar, R. Seetharaman y D. Anitha, “Detection and Classification of Melanoma Skin Cancer Using Image Processing Technique”, *Diagnostics*, vol. 13, n.º 21, p. 3313, octubre de 2023. Accedido el 14 de noviembre de 2024. [En línea]. Disponible: <https://doi.org/10.3390/diagnostics13213313>
3. C. Chen et al., “MA-SAM: Modality-agnostic SAM adaptation for 3D medical image segmentation”, *Med. Image Anal.*, p. 103310, agosto de 2024. Accedido el 14 de noviembre de 2024. [En línea]. Disponible: <https://doi.org/10.1016/j.media.2024.103310>
4. S. Khanmohammadi, N. Adibeig y S. Shanehbandy, “An improved overlapping k-means clustering method for medical applications”, *Expert Syst. with Appl.*, vol. 67, pp. 12–18, enero de 2017. Accedido el 14 de noviembre de 2024. [En línea]. Disponible: <https://doi.org/10.1016/j.eswa.2016.09.025>